

図6 Low Resolution Typing のための HLA-DRB1 のグループ分け。RG1: DRB1*0101, RG2: DRB1*1501,1502,1602, RG36: DRB1*0301,1301,1302,1402,1403,1406, RG4: DRB1*0401, 0403, 0404, 0405, 0406, 0407, 0410, RG56: DRB1*1101, 1307, RG512: DRB1*1201, 1202, RG7: DRB1*0701, RG8: DRB1*0802, 0803, 0809, RG9: DRB1*0901, RG10: DRB1*1001.

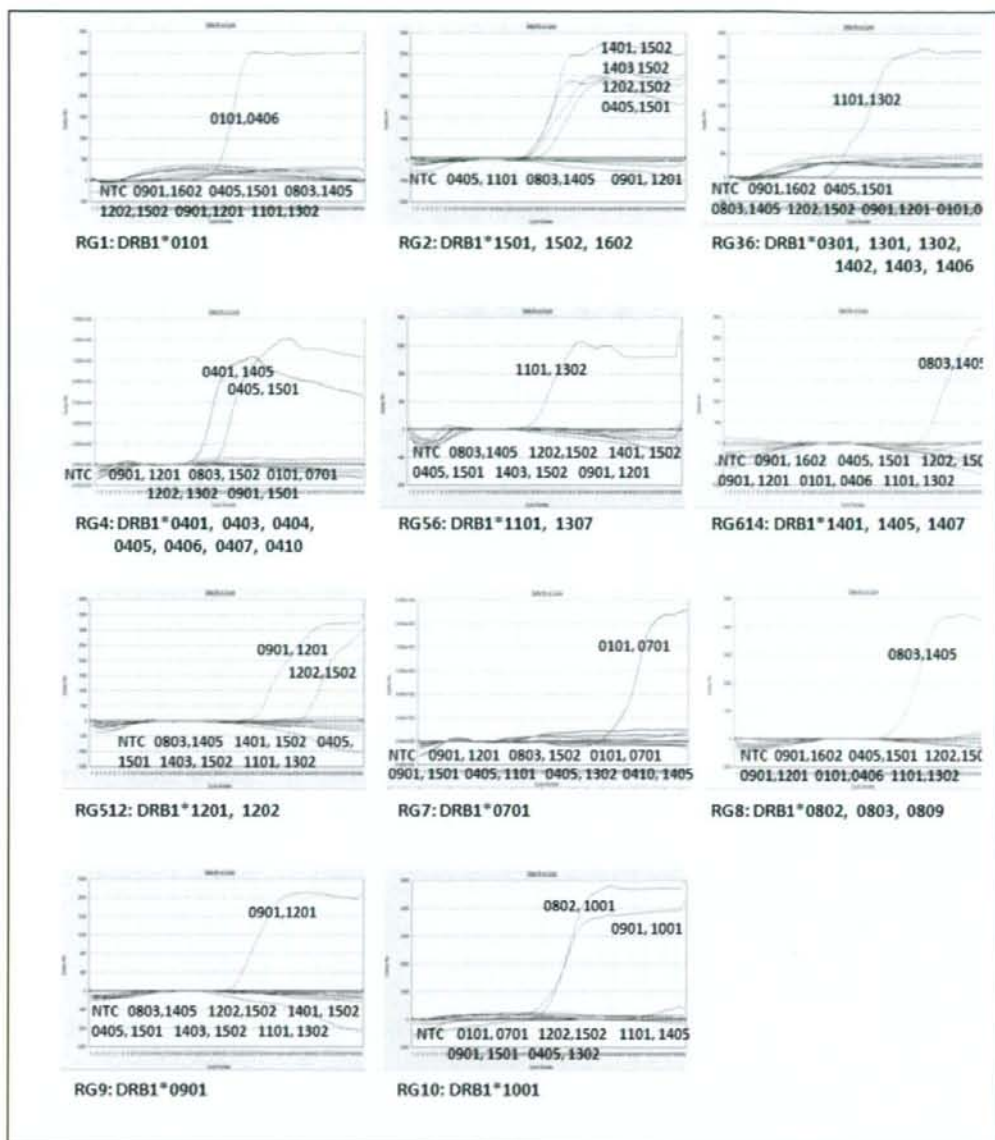


図7 DRB1のLow Resolution Typing：増幅曲線

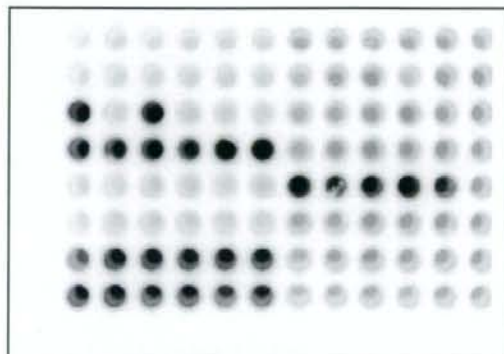


図8 LAMP 反応の SYBR Green I による可視化
LAMP による増幅終了後、SYBR Green I を滴下し発色させた。
陽性はグリーン、陰性はオレンジになるが、これはフィルター
をかけて撮影した。

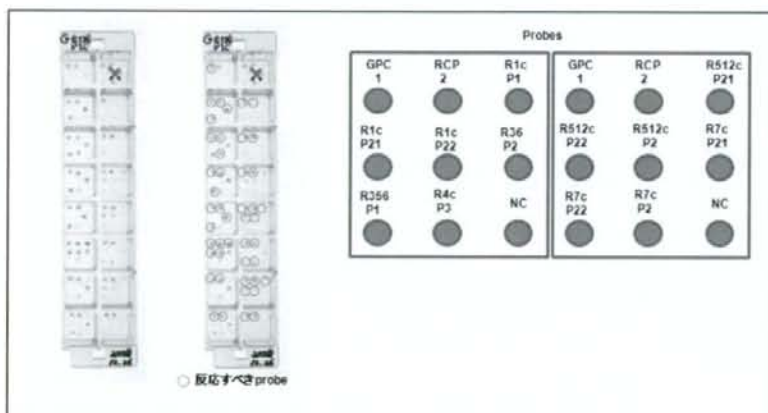


図9 プラスチック基板上的でのLAMP反応。各種プローブを固定した各well
中、biotin-dUTP 存在下でLAMP 反応を行い、その後 Avidin-Ap と
BCIP/NBT 溶液により発色させた。

III. 研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年	ページ
秋山康一、井ノ上逸朗	SNP による疾患遺伝子探査：その方法と成果	井村裕夫	実践ゲノムの最前線	六然社	東京	2009	180-190
井ノ上逸朗、田嶋敦、成田暁、安野勝史、秋山康一、高橋朋子、崔泰林	SNP 解析入門	井ノ上逸朗	SNP 解析入門	ダイナコム	千葉	2009	1-183

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Kulski JK, Shigenari A, Shiina T, Hosomichi K, Yawata M, Inoko H.	HLA-A allele associations with viral MER9-LTR nucleotide sequences at two distinct loci within the MHC alpha block.	Immunogenetics			in press
Meguro A, Ota M, Katsuyama Y, Oka A, Ohno S, Inoko H, Mizuki N.	Association of the toll-like receptor 4 gene polymorphisms with Behcet's disease.	Ann Rheum Dis	67	725-7	2008
Akiyama M, Yatsu K, Ota M, Katsuyama Y, Kashiwagi K, Mabuchi F, Iijima H, Kawase K, Yamamoto T, Nakamura M, Negi A, Sagara T, Kumagai N, Nishida T, Inatani M, Tamihara H, Ohno S, Inoko H, Mizuki N.	Microsatellite analysis of the GLC1B locus on chromosome 2 points to NCK2 as a new candidate gene for normal tension glaucoma.	Br J Ophthalmol	92	1293-6	2008
Kimura T, Kobayashi T, Munkhbat B, Oyungereel G, Bilegtsaikhan T, Anar D, Jambaldorj J, Munkhsaikhan S, Munkhtuvshin N, Hayashi H, Oka A, Inoue I, Inoko H.	Genome-wide association analysis with selective genotyping identifies candidate loci for adult height at 8q21.13 and 15q22.33-q23 in Mongolians.	Hum Genet	123	655-60	2008
Ohtsuka M, Inoko H, Kulski JK, Yoshimura S.	Major histocompatibility complex (Mhc) class Ib gene duplications, organization and expression patterns in mouse strain C57BL/6.	BMC Genomics	9	178	2008

Sekigawa T, Tajima A, Hasegawa T, Hasegawa Y, Inoue H, Sano Y, Matsune S, Kurono Y, Inoue I	Gene-expression profiles in human nasal polyp tissues and identification of genetic susceptibility in aspirin intolerant asthma	Clin Exp Allergy			in press
Murakami Y, Aly HH, Tajima A, Inoue I, Shimotohno K	Regulation of the hepatitis C virus genome replication by miR-199a	J Hepatol	50	453-460	2009
Bilguvar K, Yasuno K, Niemela M, Ruigrok YM, Fraunberg M, Duijn CM, Berg LH, Mane S, Mason C, Choi M, Gaaal E, Bayri Y, Kolb L, Arlier Z, Ravuri S, Ronkainen A, Tajima A, Laakso A, Hata A, Kasuya H, Koivisto T, Rinne J, Ohman J, Breteler MMB, Wijmenga C, State MW, Rinkel GJE, Hernesniemi J, Jaaskelainen JE, Palotie A, Inoue I, Lifton RP, Gunel M	Susceptibility loci for intracranial aneurysm in Europe and Japanese populations	Nat Genet	40	1472-7	2008
Nishida N, Koike A, Tajima A, Ogasawara Y, Ishibashi Y, Uehara Y, Inoue I, Tokunaga K	Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals	BMC Genomics	9	431	2008
Krischek B, Kasuya H, Tajima A, Akagawa H, Sasaki T, Yoneyama T, Ujiiie H, Kubo O, Bonin M, Takakura K, Hori T, Inoue I	Network-based gene expression analysis of intracranial aneurysm tissue reveals role of antigen presenting cells	Neuroscience	154	1398-1407	2008
Uno Y, Suzuki Y, Wakaguri H, Sakamoto Y, Sano H, Osada N, Hashimoto K, Sugano S, Inoue I	Analysis of expressed sequence tags from liver in cynomolgus monkey (<i>Macaca fascicularis</i>): A systematic identification of drug-metabolizing enzyme genes	FEBS Let	582	351-358	2008
Sasahara A, Kasuya H, Krischek B, Tajima A, Onda H, Sasaki T, Akagawa H, Hori T, Inoue I	Gene expression in a canine basilar artery vasospasm model: a genome-wide network-based analysis	Neurosurg Rev	31	283-290	2008

Bae JS, Cheong HS, Kim JO, Lee SO, Kim EM, Lee HW, Kim S, Kim JW, Cui T, Inoue I, Shin HD	Identification of SNP markers for common CNV regions and association analysis of risk of subarachnoid aneurismal hemorrhage in Japanese population	Biochem Biophys Res Commun	373	593-596	2008
Osada N, Hashimoto K, Kameoka Y, Hirata M, Tamura R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I	Large-scale analysis of <i>Macaca fascicularis</i> transcripts and inference of genetic divergence between <i>M. fascicularis</i> and <i>M. mulatta</i>	BMC Genomics	9	90	2008
Saigo K, Yoshida K, Ikeda R, Sakamoto Y, Murakami Y, Urashima T, Asano T, Kenmochi T, Inoue I	Integration of hepatitis B virus DNA into the MPP (mixed lineage leukemia) 4 gene and rearrangements of MLL4 in human hepatocellular carcinoma cells	Hum Mutat	i29	703-708	2008
Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, Inoue I	Genome-wide expression analyses of testes of non-obstructive azoospermia patients demonstrate a specific gene expression profile and implicate ART3 in genetic susceptibility	PLoS Genet	4	e26	2008
Nakaoka H, Gaillard C, Fujinaka K, Watanabe N, Ito M, Kawada K, Ibi T, Sasae Y, Sasaki Y	The use of link provider data to improve national genetic evaluation across weakly connected subpopulations	J Anim Sci	87	62-71	2009
Nakaoka H, Gaillard C, Ibi T, Sasae Y, Sasaki Y	Adjusting for heterogeneity of variance for carcass traits affects single and multiple trait selections in genetic evaluation of Japanese Black cattle	Anim Sci J	79	645-654	2008

IV. 研究成果の刊行物・別刷

SNP 解析入門

〇〇大学〇〇科
バイオ太郎

1. 初めて統計遺伝解析に接する研究者のための解説.....	5
1.1 はじめに.....	5
1.2 どのような検体収集をおこなうか.....	6
1.3 統計学的検定の前提.....	7
1.4 アソシエーション・スタディにおける統計遺伝解析.....	8
1.5 有意差があるというのはどういうこと.....	12
1.6 得られた有意差をどのように評価したらいい?.....	13
1.7 Hardy-Weinberg 平衡と乖離した場合の統計遺伝解析.....	14
1.8 さいごに.....	16
2. SNP を用いた疾患感受性遺伝子同定.....	17
2.1 はじめに.....	17
2.2 SNP を用いた疾患感受性遺伝子マッピングの概略.....	18
2.2.1 研究対象とする「多因子疾患」の遺伝性について.....	18
2.2.2 疾患感受性遺伝子マッピングのための研究デザイン.....	20
2.3 タイピングする SNP について.....	26
2.4 SNP による遺伝子マッピング法.....	32
2.5 エビデンスに基づく治療・診断に向けての課題.....	34
2.6 おわりに.....	36
3. SNP による集団遺伝学的研究.....	37
3.1 はじめに.....	37
3.2 集団遺伝学とは.....	38
3.2.1 遺伝的浮動、自然選択、そして「中立的」な突然変異.....	39
3.2.2 集団の有効な大きさ、塩基多様度、そして most recent common ancestor (MRCA).....	42
3.3 疾患関連研究としての集団遺伝学の意義.....	46
3.4 全ゲノム SNP スキャンからの疾患感受性候補遺伝子の抽出.....	50
3.5 SNP と遺伝子発現量多型との関わり.....	54
3.6 おわりに.....	56
4. SNP による連鎖解析.....	57
4.1 はじめに.....	57
4.2 連鎖と組換え.....	58
4.3 連鎖解析とは.....	60
4.4 SNP と連鎖解析.....	65
4.5 SNPLINK—高密度 SNP 連鎖解析用 Perl スクリプト.....	68
4.6 おわりに.....	71
5. 多重検定についての考え方および解決策.....	72

5.1 はじめに.....	72
5.2 Bonferroni の補正法.....	73
5.3 False discovery rate (FDR) を制御する手法.....	77
5.4 パーミュテーションテスト (並べ替え検定)	81
5.5 各手法のまとめ.....	83
5.6 おわりに.....	85
6. 遺伝子間相互作用の検出法.....	86
6.1 はじめに.....	86
6.2 遺伝子間相互作用とその検出法.....	87
6.3 ノンパラメトリック法.....	90
6.3.1 組み合わせ法.....	90
6.3.2 再帰分割法.....	93
6.3.3 セット関連解析法.....	97
6.4 ニューラルネットワーク.....	100
6.5 グラフィカルモデリング.....	103
6.6 遺伝子間相互作用解析法の課題と今後の展望.....	106
7. 遺伝子コピー数変異 (CNV) と疾患.....	107
7.1 CNV とは.....	107
7.2 CNV の成り立ち.....	110
7.3 CNV と疾患.....	116
7.4 CNV の検出方法.....	120
7.5 利用可能な CNV データベース.....	123
7.6 最後に.....	125
8. 高密度 SNP タイピングアレイを用いたゲノム全域解析による疾患遺伝子解析.....	126
8.1 はじめに.....	126
8.2 高密度 SNP タイピングアレイ実験.....	127
8.2.1 ターゲット配列の増幅、末端ラベル.....	128
8.2.2 アレイとターゲット配列のハイブリダイゼーション.....	128
8.2.3 アレイの洗浄とイメージスキャン.....	129
8.3 データ解析 (SNP タイピング)	130
8.4 データ解析 (DNA コピー数変異解析)	132
8.5 まとめ.....	134
9. BeadChip を用いたゲノム全域解析による疾患遺伝子解析.....	135
9.1 はじめに.....	135
9.2 BeadChip の原理.....	136
9.3 BeadChip 実験.....	139

9.3.1	ターゲット配列の増幅、断片化	139
9.3.2	アレイとターゲット配列のハイブリダイゼーション	140
9.3.3	アレイの洗浄とイメージスキャン	140
9.4	データ解析 (SNP タイピング)	141
9.5	データ解析 (遺伝子コピー数変異/ヘテロ接合欠失)	143
9.6	まとめ	147
10.	これだけは知っておきたい SNP 解析のための統計学	148
10.1	はじめに	148
10.2	推定と仮説検定	149
10.2.1	推定	149
10.2.2	仮説検定	151
10.3	ベイズ推定	157
10.4	おわりに	161
<用語説明>		162
<参考文献>		169

1. 初めて統計遺伝解析に接する研究者のための解説

1.1 はじめに

疾患遺伝子研究に初めて接する方を対象にこの章をくみたい。大方は医学、薬学を含めた生命科学をバックグラウンドにされていることだろう。きちんと統計学を学ぶ機会のなかった方がほとんどであり、統計学に拒否感までもっている方もいるかもしれない。この章は、統計学を学んできた方はまったく対象としていないことを断っておく。分子生物学者からよくいわれることがある、曰く「統計学は信じない、数字のトリックだろう」、そのような誤解を解きつつ、疾患遺伝子研究における統計学について説明したい。最近、医療の分野では患者とのコミュニケーションスキルが問われている。患者に伝えるべきことの多くは統計学に基づく確率的な事象なのに、医師が統計学を理解していないという現実のため、確率に基づく結果だと認識していない患者にうまく伝えることができていないともいえるのではないだろうか。統計遺伝学的手法に関しては SNPalyze などのソフトウェアを用いるとして、基本的な考え方についてまとめる。参考文献に示した「[数学いらずの医科統計学](#)」は理解しやすく初めて統計学に接する研究者にお薦めの手引書である。

1.2 どのような検体収集をおこなうか

患者と対照で有意に差のある要因を検出する目的でアソシエーション・スタディがおこなわれる。疫学的研究では、年齢、性別等、なるべく患者、対照の背景を揃えた上である要因の関与を比較検討する。遺伝要因を検討する際にも同様の考えに基づいてもいいのだが、小さな効果しかない遺伝要因を検出するためには戦略が必要である。さまざまな要因の積み重ねで発症にいたると仮定すると、遺伝要因の関与なしに加齢とともに環境要因の積み重ねで発症する患者もいる。すなわち非遺伝要因の積み重ねが大きい高齢発症の患者では遺伝要因の関与が小さいことが予想される。実際に家族歴を有する患者の発症年齢は家族歴のない患者より若いことが一般的である。一方、対照はその疾患の非罹患者が望ましい。罹患頻度が高い common disease は、若年者ではまだ罹患していなくても将来罹患する可能性がある。そうすると若年者対照は除いた方がよく高齢非罹患者が理想的である。できれば疾患家族歴のない対照が望ましい。現実的には対照として、罹患の有無を問わず単に日本人一般集団を用いることもある。その場合、患者数の5倍ほどの多くの対照例があることが望ましい。実際のところ、患者収集に関しては疾患ごとに異なる考えがあるだろう。特に臨床情報収集に関しては疾患ごとに考慮すべきである。また遺伝背景の強い患者を対象とすると遺伝要因の検出には効率的と考えられるが、患者検体数を増やすことと相容れないところがある。高齢でないこと、家族歴を有すること、が優先されよう。

患者： まず対象疾患を有すること、そして家族歴を有すること、高齢でないこと、を優先して収集

対照： 対象疾患を有しないこと、家族歴がないこと、高齢であること、を優先して収集、場合に応じ日本人一般集団を対照とする

1.3 統計学的検定の前提

疾患遺伝子研究に限らず有意差検定をおこなうにあたり、いくつかの基本的前提がある。

- 1) サンプルは母集団からランダムに抽出されていること
 - 2) 検定する形質（病気か否か）以外の遺伝背景は同一である
 - 3) 有意差を得るための十分な検体数であること
-
- 1) 実際的には判定は困難と思われる。前節で述べたよう、患者集団を母集団として考えると、その中から条件を加えて検体収集することもあるので（例えば高齢者の高血圧は遺伝要因が小さいと考えられ除外するといったことである）、1) を満たしていることがいいとはいえないかもしれない。
 - 2) に関してはよくいわれるのが階層化の問題である。よくいわれる例え話がある。サンフランシスコベイエリアにおいて、箸の使い方のうまさとは HLA との関連を調べた。結果、HLA-A1 タイプと箸をうまく使える形質が関連していた。HLA-A1 は chopstick gene だった？もちろんそんなことはなく、サンフランシスコでのスタディ集団は中国系と白人集団が混在しており、中国系集団では HLA-A1 頻度が高いことによりこのような結論となった。いうまでもなく集団中に混在していた中国系集団が箸をうまく使う。遺伝的背景が均一な集団内での比較が重要である。ゲノム全域にわたり数万個の SNP タイピングデータがあると多次元尺度法などにより集団の階層化の確認ができる (Purcell S, et al. 2007; Price AL et al. 2006)。
 - 3) の検体数に関しては遺伝背景を同一とする良質な検体なら多い方がいいに決まっている。特に多くの SNP を解析する場合、その数に応じた補正が必要で（多重検定の補正）、強い有意差を得るには十分な検体数を必要とする。また他の研究成果を追試する場合に、偽陰性を避けるため、十分な検体数を確保しなければならない。具体的な検体数の考え方については、2 章を参考にされたい。

1.4 アソシエーション・スタディにおける統計遺伝解析

疾患遺伝子研究において、ある遺伝子（変異、多型）が病気の原因となっているかを知りたい。そのためには、その疾患にしか存在しない変異があると原因と考えやすい。言い換えると、ある変異をもっていると必ず病気になるという事例であり、通常はメンデルの遺伝法則に従うだろう。

耳あか遺伝子の例を示す。耳あかには乾性と湿性タイプがあり、この形質は遺伝的に規定されていることが知られていた。長崎大学の新川らにより、連鎖解析とポジショナル・クローニングにより原因遺伝子が同定された(Yoshiura et al. 2006)。関与する遺伝子は多剤耐性に関与する ABCG11 であり、薬剤を排出するトランスポーターの役割が考えられている。SNP rs17822931 が原因となっており、マイナーアレル A をホモ接合体で有すると乾性となることが見いだされた。論文で示されている日本人 126 名の遺伝子型の分布を表に示す（表 1.1）。

表 1.1 日本人 126 名の遺伝子型の分布

rs17822931	GG	GA	AA
Dry	0	1	87
Wet	4	34	0

ひとりの例外を除いて乾性タイプでは遺伝子型 AA となっており、常染色体劣性遺伝を示すことがわかる。実際に AA のホモ接合体を有する人は例外なしに乾性タイプとなる。この場合、見て明らかで統計的な解析は必要としないだろう。ちなみにひとりの例外が存在するが、エクソン 29 に 27-bp の欠損を有することが示されており、rs17822931 の G アレル上に変異が存在したとすると、遺伝子変異がホモ接合体で存在することとなる。

例外を除いて、疾患リスクとならない GG + GA をまとめ 2×2 分割表を作成し統計解析を試みてみよう（表 1.2）。

表 1.2 2×2 分割表

rs17822931	GG + GA	AA
Dry	0	87
Wet	38	0

セルに 0 があるので、Fisher's exact test により検討すると $P = 2.2 \times 10^{-16}$ と算出される。明らかに耳あかが乾性であることと AA 遺伝子型が関連していることがわかるだろう。

さて多因子疾患の場合、ひとつの遺伝子のみでは成因を説明できない。そもそも遺伝要因のみでなく、環境要因、遺伝子間相互作用、環境要因との相互作用によって発症にいたると考えなければならない。ひとつひとつの要因は大きいものではないので、その関与については統計学的な検定を要する。

最初に本態性高血圧の解析例を示す。本態性高血圧は典型的な common disease であり、複雑な遺伝要因が存在すると予想されている。我々は候補遺伝子解析からアンジオテンシノーゲン遺伝子(AGT)を感受性遺伝子と特定することができた。原因多型とされる AGT プロモーター多型(-6G/A)とタイピング結果を以下に示す。(表 1.3)

表 1.3 AGT プロモーター多型(-6G/A)とタイピング結果

-6 G/A	G	A
高血圧患者	509 (53.4%)	445 (46.6%)
対照	448 (61.6%)	280 (38.4%)

患者、対照のアレルカウントと頻度が示されている。頻度を見ると A アレル (adenine type) が患者で高いことが見て取れる。ただし A アレルを有すると必ず高血圧になるわけでない。実際、対照でも 38.4% が A アレルを有している。統計学的有意差検定が求められる。アレルカウントを患者と対照で比較し、患者で A アレルが有意に多いと疾患に関連していることとなり、統計解析は通常カイ二乗検

定が用いられる。ここでは結果のみ示す。 $\chi^2=11.0$, 自由度 1、 $P=0.0008$ という結果で統計的に有意と判定された。

ここからカイ検定について若干詳細な説明を加えたい。まず遺伝型が患者、対照で以下のように分布しているとしよう (表 1.4、表 1.5)。

表 1.4 患者と対象の遺伝子型の分布

		患者	対照
遺伝子型	AA	n_1	m_1
	Aa	n_2	m_2
	aa	n_3	m_3
縦計		n	m

表 1.5 患者と対象のアレル数

		罹患者	対照	横計	
アレル	A	N_a	N_b	R_1	$N_a=2n_1+n_2$, $N_b=2m_1+m_2$
	a	N_c	N_d	R_2	$N_c=n_2+2n_3$, $N_d=m_2+2m_3$
縦計		$2n$	$2m$	N	

ホモ接合体では、ふたつの染色体が例えば A アレルとなっているので、 $2n_1$ となる。 2×2 分割表でカイ二乗検定をおこなう。それぞれのセルが独立であるかどうかの検定である。

カイ二乗検定は以下で表される。

$$\chi^2 = \frac{(\text{観察値} - \text{期待値})^2}{\text{期待値}} \quad (\text{式 1})$$

各セルのアレル頻度に独立であるとすると、

N_a の期待度数 E_a は $2nR_1/N$

N_b の期待度数 E_b は $2mR_1/N$

N_c の期待度数 E_c は $2nR_2/N$

Nd の期待度数 E_d は $2mR_2/N$ となる。

独立性の検定は以下となる。

$$\chi^2 = \frac{(Na-E_a)^2}{E_a} + \frac{(Nb-E_b)^2}{E_b} + \frac{(Nc-E_c)^2}{E_c} + \frac{(Nd-E_d)^2}{E_d} \quad (\text{式 2})$$

期待値を a, b, c, d, N で変換すると、

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+b)(c+d)(a+c)(b+d)} \quad (\text{式 3})$$

と計算すればいい。ここでは自由度 1 の χ^2 値の対応表から P 値を得る。もちろんこんな計算を実際やることはなく、大方どのような統計ソフトでも算出できる。 χ^2 の基本式(式 1)からわかるよう、 χ^2 値は数に比例する。そのままの頻度差を保てば、検体数が 2 倍になれば χ^2 値も 2 倍になる。

アソシエーション・スタディではカイ検定で有意差検定をおこなう。アレル頻度での検定では 2×2 分割表を作成し自由度 1 で算出(検体数 $\times 2$ のアレル数となる)。遺伝子型では 2×3 分割表を作成し自由度 2 で算出。