

「テラーメイド医療を目指したゲノム情報活用基盤技術」
平成 15 年度採択研究代表者

井ノ上 逸朗

(東京大学医科学研究所 客員助教授)

「sub-common disease の感受性遺伝子同定と個人型易罹患性診断への応用」

1. 研究実施の概要

SNP データベースは急速に充実し HapMap 計画も終了するなど、SNP 解析による個人差医療のインフラは整備されている。疾患感受性、薬剤感受性に関与する SNP は個人差医療の重要な鍵を握る。しかしながら、1 0 0 0 万におよぶ SNP すべてのタイピングは時間、労力、費用の面から現実的でない。また遺伝背景の弱い疾患では感受性遺伝子を遺伝統計学的に同定できても、その生物学的機能関与が捉えにくい。また個々人の将来予測に応用することは現時点では非常に困難である。本研究は、比較的遺伝背景の強い、脳動脈瘤、後縦靭帯骨化症、家族性甲状腺癌といった sub-common disease を対象とし、ゲノム全域連鎖解析でマップされた領域から効率の良いポジショナル（候補）クローニングをおこない、疾患感受性遺伝子を同定する。そして遺伝要因、生活習慣、環境要因の関与を統合した量的形質モデルを構築し、集団から個人への諸因子の関与を明らかにする。そして個人型易罹患性診断への応用を目指し、テラーメイド医療の実現を目指す。

比較的遺伝背景の強い sub-common disease ではあるが、遺伝様式、浸透度などパラメーターの事前決定はできないので、ノンパラメトリック連鎖解析をおこなう。脳動脈瘤において 1 1 4 罹患同胞対でのゲノム全域スクリーニングで 5、7、1 4 番染色体に連鎖を認めた。後縦靭帯骨化症においては 1 4 1 対のスクリーニングで 2 1 番染色体に連鎖を確認できている。また連鎖領域からの感受性遺伝子同定にも成功した。家族性甲状腺癌は 3 1 0 家系が登録されており、現時点で 9 2 罹患同胞対の収集ができ、ゲノム全域スクリーニングにより 3、6 番染色体に連鎖を認めている。ノンパラメトリック連鎖解析ではおよそ 1 0 cM におよぶ連鎖を認め、領域には 1 0 0 個以上の遺伝子が存在するので連鎖不平衡解析によるピンポイントが必要となる。そのためには効率的な SNP タイピング確立、統計解析が求められる。SNP タイピングはハイスループット法である TaqMan assay-on-demand を用いる。連鎖不平衡、Permutation 法を加えたハプロタイプ解析により疾患感受性遺伝子を同定する。すでに必要なプログラム (SNPAlyze) はダイナコムと共同で開発し、膨大な SNP 情報を迅速に処理できる。

これまでの SNP での遺伝疫学研究は集団での易罹患性（リスク）を示しており、個人が

どの程度病気に罹りやすいかにはまったくアプローチされていなかった。SNP データ、生活習慣、臨床検査データを組み合わせ、閾値モデルと variant component model を組み合わせた新規モデルを構築することにより個人の疾患予測をおこなう未来型診断システムを提案する。特に個人の易罹患性は連続的な変数を示すという量的形質とそれがある閾値を越えたとき発症するという閾値モデルの考えを基礎におき、そこに variance component model を展開することで遺伝情報、生活習慣、臨床データなどを組み込んだ易罹患性モデルを構築し、将来予測へ繋げる。

2. 研究実施内容

脳動脈瘤感受性遺伝子同定

脳動脈瘤罹患同胞対連鎖解析によりゲノム全域でマッピングされた染色体領域からのポジショナル・クローニングを目指す。全国レベルで収集した 114 罹患同胞対を用いたゲノム全域連鎖解析により、3ヶ所に候補領域を絞り（5、7、14番染色体）、7q11-21 で最大ロッド値 3.22 を得た。

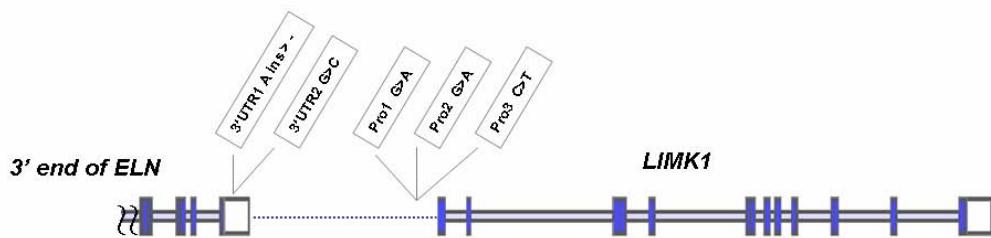
7q11-21 での連鎖ピークは D7S2472 で認めており、この領域の体系的な SNP スクリーニングを試みた。領域 26 遺伝子、188 SNPs を用いてゲノム領域での連鎖不平衡マッピングをおこなった。ゲノム領域からのスクリーニングでエラスチン遺伝子および隣接する LIM kinase1 遺伝子領域に有意差を有する SNP を検出でき、それらはふたつの遺伝子にまたがる連鎖不平衡ブロックに存在していた。またゲノム領域でのウインドウスキャンを用いたハプロタイプ解析により、まったく同一領域に有意差を得ることができた。エラスチンは血管弹性纖維の主要構成成分であり、当然脳動脈瘤の候補遺伝子と考えられた。そこで、患者 404 例、対照 458 例を用いた領域の詳細な連鎖不平衡解析により、5 SNPs で $P < 0.001$ の有意差を認めた。もっとも強い有意差を ELN 3' UTR (+659) SNP で認め、 $P = 0.000002$ であった（表 1）。また韓国人検体（患者 195 例、対照 250 例）で同じ検討をおこない、低い有意差ながら関連を示す結果を得ている（ $P = 0.027$ ）。

ヒト臍帯動脈から得た培養平滑筋細胞を用いて、遺伝子型に伴う機能解析をおこなった。ELN 3' UTR (+502) A insertion 多型そして LIMK1 promoter (-187) 多型の存在により、それぞれの転写産物が減少することが real time RT-PCR で示された（図 1）。すなわちこれらは機能的多型であり、脳動脈瘤発症と関与している可能性が強く示唆された。機能的関与をより詳細に検討するために、ルシフェラーゼベクターを用いた in vitro 実験を試みた。ルシフェラーゼ下流に ELN 3' UTR 配列を組み込むと、ELN 3' UTR (+502) A insertion の存在によりルシフェラーゼ活性の減少を認め、mRNA の不安定化を生じたと予測できた。また、EMSA により RNA 結合蛋白の存在を検出している。LIMK1 プロモーター領域をルシフェラーゼ上流に組み込んだ結果、プロモータ活性の低下を確認している。さらにハプロタイプ解析により、これらの多型はひとつのハプロタイプ上に存在し、もっとも有意差の強かった ELN 3' UTR (+659) SNP が at-risk ハプロタイプのタグとなっていることが判明

した。すなわちエラスチン、LIMK1 の両方の同時発現低下があると脳動脈瘤を発生しやすいという新規メカニズムを提唱できた。最近エラスチンはシグナル分子として平滑筋細胞へ作用し、その下流には LIM kinase 1 が位置し、アクチンフィラメント結合を調節することも明らかになった。それらが、相互作用しながら脳動脈瘤発生に関与しているという発症機構が考えられた。

さらには、他の連鎖領域、5、14番染色体についてもスクリーニングしているところで、14番染色体については日本人、韓国人集団で有意差を認める遺伝子同定ができる。現時点では機能解析による遺伝子多型の関与について検討中である。

表1 Elastin と LIM kinase 1 遺伝子の脳動脈瘤への関与



Locus	Nucleotide Variation	NCBI dbSNP ID	Frequency		χ^2	P	Permutation P ^c	OR (95% c.i.)
			Japanese IA (n = 404)	non-IA (n = 458)				
<i>ELN</i> ^a								
3'UTR(+502)	->A ins.	Non-dbsNP	13.8% (111/804)	10.4% (94/906)	4.75	0.029	0.029	1.38 (1.03-1.89)
3'UTR(+659)	G>C	rs8326	7.4% (59/792)	2.5% (23/912)	22.47	0.000002	0.000001	3.11 (1.96-5.45)
<i>LIMK1</i> ^b								
Pro.(-961)	G>A	rs6460071	13.3% (107/804)	7.4% (66/896)	16.37	0.000052	0.000030	1.93 (1.41-2.68)
Pro.(-428)	G>A	rs710968	18.6% (149/802)	12.4% (107/862)	12.13	0.00050	0.00038	1.61 (1.23-2.08)
Pro.(-187)	C>T	Non-dbsNP	13.4% (106/792)	7.3% (65/894)	17.22	0.000033	0.000019	1.97 (1.39-2.69)

^aNumbers in parentheses indicate the nucleotide position from the stop codon

^bNumbers in parentheses indicate the nucleotide position from the first nucleotide of exon 1.

^cPermutation P values were generated after 1,000,000 iterations.

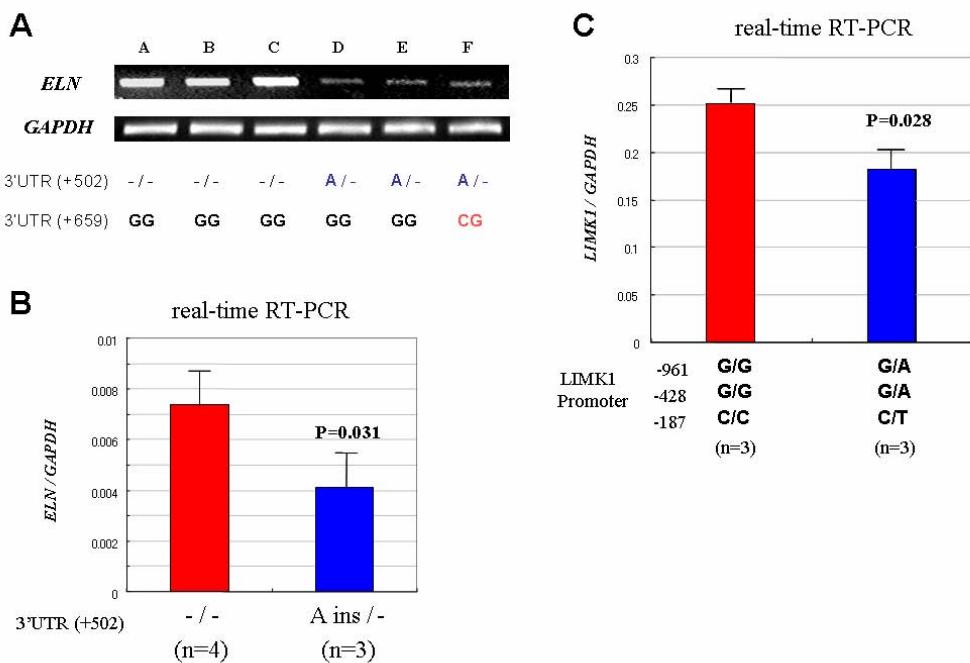


図1 Expressions of ELN mRNA (A and B) and LIML1 mRNA (C) in human SMC from umbilical artery according to genotypes of the 3'UTR and the promoter region, respectively.

家族性甲状腺癌のゲノム全域連鎖解析

野口病院（内野眞也外科部長）において 172 家系、260 サンプル（内、罹患同胞対は 79 家系 92 同胞対）の家族性甲状腺癌サンプル及び健常者 106 サンプルの収集がなされた。72 家系、150 サンプル、81 ペアを用いてゲノム全域罹患同胞対連鎖解析を行った。連鎖解析には、SNPs タイピング用 DNA マイクロアレイ (GeneChip Human Mapping Xba 50K Array (Affymetrix 社製)) 及び GeneSpring GT (Agilent 社製)(Multipoint nonparametric, Genetic Map; deCODE)を使用した。この結果、3 番、6 番染色体に連鎖（それぞれ Z score 3.25, 3.13）を認めた（図 2）。連鎖領域内の遺伝子について遺伝子機能、発現情報などから候補遺伝子を選択し、144 家系サンプルを用いて PCR 産物の直接シークエンス法によるコード領域の変異スクリーニングを行ったところ、受容体型チロシンキナーゼ-RAS シグナル伝達系上の候補遺伝子 A 内に家族性甲状腺癌患者検体でミスセンス変異が集積していることを見出した（表 2）。同定されたミスセンス変異のうち新規のものは、健常者 106 サンプルにおいて検出されず家系内で甲状腺癌との同時分離を認めるものが多く、発癌と関連する遺伝子であることが強く示唆された。

今後さらに候補遺伝子 A について、FNMTCA 家系検体での変異解析を継続し、甲状腺癌患者との関連解析を行う。

Chromosome 6

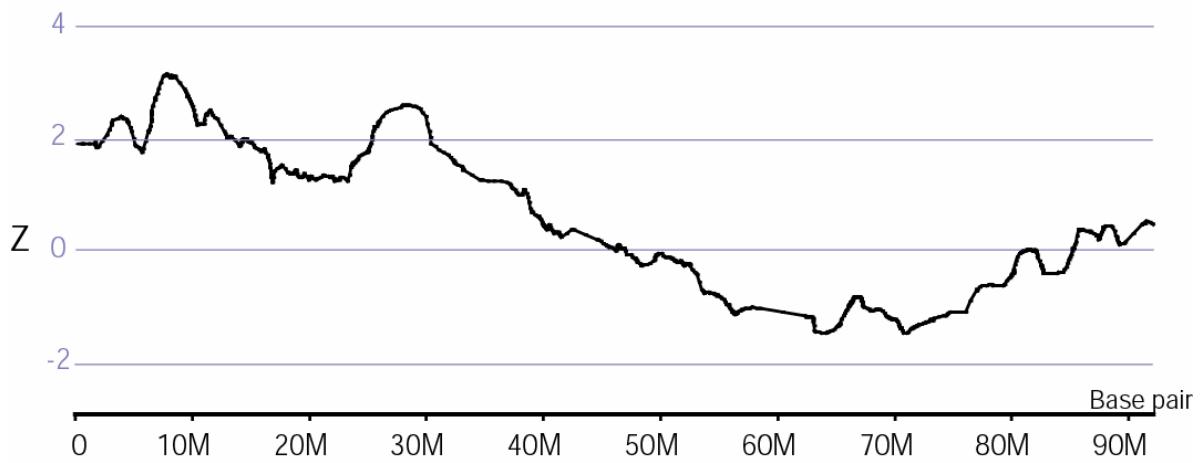


図 2 Multipoint nonparametric Z score of chromosome 6 for FNMTC ASPs with thyroid carcinoma

表 2 FNMTC 家系サンプルで同定された遺伝子 A 内変異リスト

Base_Change	Type	Effect	Aa_Change
49G>A	Sub	missense	G17R
120T>C	Sub	silent	-
141C>T	Sub	silent	-
346G>A	Sub	missense	G116R
393G>A	Sub	silent	-
748A>C	Sub	missense	K250Q
765A>G	Sub	silent	-
1019T>C	Sub	missense	M340T
1112C>T	Sub	missense	S371L
1454G>T	Sub	missense	G485V
1556C>G	Sub	missense	P519R
1672T>C	Sub	missense	C558R
1680C>G	Sub	silent	-
1943G>C	Sub	missense	G648A
1988A>G	Sub	missense	E663G
2233C>T	Sub	missense	R745C
2295C>T	Sub	silent	-
2544A>G	Sub	silent	-
2617G>A	Sub	missense	D873N
2784C>T	Sub	silent	-

Variance component-Threshold モデルを用いた診断および予測プログラムの開発

Common disease といわれる多因子疾患は、単一因子のみによって支配されるものではなく、遺伝的要因の他、生活習慣や環境的因素が、相互かつ複雑に絡み合いながら、その liability (易罹患性) を決定しており、その複雑さは、これらの疾患の発症メカニズムの解明を困難なものにしている。特に個々人に合わせたテーラーメイド医療を確立するにあたり、遺伝要因と環境要因を相容れた上で個人予測を試みることが鍵となる。現在でも、遺伝子多型情報に加え、生活習慣についての問診データや、血糖値、血圧などの臨床検査データをもとに、易罹患性に影響を及ぼす因子の特定を行ない、その結果を疾患発症予測に還元するシステム構築に関する議論はなされつつある。しかしながらそのほとんどはロジスティック重回帰分析など従来の統計学的手法を用いたものである。そのため、複雑な高次元相互作用の検出には、未知パラメータの数が指数関数的に増加する問題があり、非現実的な検体数を必要とすることが多く、また各要因間に存在する相関も正しく考慮されずに計算が行なわれている可能性が高い。比較的大きな効果を有する一部の遺伝子多型が特定されつつあるものの、解決すべき問題は依然として多く残されたままである。

本研究チームでは、閾値モデルおよび variance component (VC) モデルを併用した新たな解析手法の開発を進めている。閾値モデルにおいては、量的形質と同様、正規分布などにしたがう連続的な変異を示す変数としての易罹患性が存在し、この値がある閾値を超えた場合に疾患を発症するという仮定がなされる。これにより、ロジスティック重回帰分析同様、発症もしくは非発症の二値変数である表現型のデータに、量的形質で一般的に用いられる線形モデルの理論が取り入れられることとなる。従来の遺伝モデルが、複数の因子の組み合わせによって生じるサブクラスの効果を、各々独立したパラメータとして扱っているのに対し、VC モデルにおいては、ある一つの分散成分によってもたらされるランダムな変量効果として扱う。その基本形および対数尤度関数 $\ln L$ は各々以下に示す通りであり、本手法では、マルコフ連鎖モンテカルロ (MCMC) アルゴリズムを用いて、この対数尤度関数を最大化する方向へ、遺伝分散や各被験者の liability といったパラメータの最適解を検索していく。

$$y = Xb + Zv + e$$

$$\ln L = -0.5*|V| - 0.5*(y - Xb)' V^{-1} (y - Xb) \quad (\text{ただし, } V = ZGZ' \sigma_v^2 + I)$$

VC モデルを想定することにより、未知パラメータの数の増加が抑えられ、易罹患性に寄与する高次元相互作用の解析が、現実的なサイズのデータで可能となり、要因間の相関を考慮する必要性も生じない。またここで、各 SNP における多型情報をもとに作成した遺伝子関係行列 G で相互に関連付けて各サブクラスの効果の計算を行なうことも、本手法の特長である。 G の要素 g_{ij} は、 $\sum_{k=1}^N u_{ij}^k w_k$ で計算される。ここで、 N はサブクラスを定義する SNP 数、 u_{ij}^k はサブクラス i と j において、SNP k の遺伝子型が一致しているか否かを示すバイナリ変数 (一致していれば 1、いなければ 0) 、 w_k は SNP k における重み付け係数を表す。こ

れにより、異なるサブクラス間において、各 SNP が共通した遺伝子型を示した場合に生じる似通いが考慮され、より多くの情報に基づいた、信頼度の高い結果が期待される。さらに、サブクラス数同様、因子の組み合わせのパターンは、その数に伴って指数関数的に増加する。したがって、データを解析する際、発症因子の候補となる項目が多岐に渡る場合、膨大な数の計算が必要となる。そこで、reversible jump マルコフ連鎖モンテカルロ (RJMCMC) 法を組み合わせることで、計算の効率化を図った。これにより、易罹患性に寄与する高次元相互作用をもたらす因子の組み合わせを迅速かつ的確に同定することが可能となる。

本研究により、遺伝子間のみならず、生活習慣などの因子との間に生じる複雑な高次元相互作用も広く的確に同定することが可能となれば、従来の手法では解明し得ない、多因子疾患の発症メカニズムの全容を明らかにする一助となると考えられる。また本手法は、量的形質の解析にも用いることが可能である。

現在までに、ツールの基本形を完成させ、シミュレーション実験による検討を行なっており、ここでは、8SNP について 640 名の被験者の記録が存在し、そのうち 4、および 7 の 2SNP が相互作用を持つとした場合（ケース 1）と、3、4、および 7 の 3SNP が相互作用を持つとした場合（ケース 2）の 2 条件についての結果を示す。次頁の表に示された通り、設定条件と一致した結果が得られており、このことは本手法の妥当性、および今後に向けた可能性を示唆するものと言える。

今後は、これまでの結果をさらに発展させ、今後の生活習慣の変化や加齢などの影響も考慮しつつ、将来の疾患発症の可能性を時系列的に予測するための統合的なシステムを確立する。

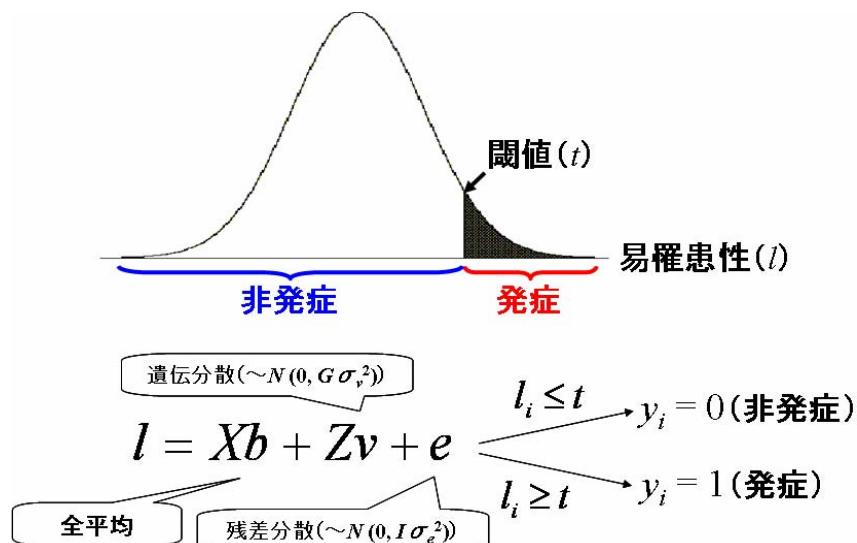


図 3. Variance component-Threshold モデルの概要

$$g_{ij} = \sum_{k=1}^N u_{ij}^k w_k$$

Subject	SNP1	SNP2	SNP3		1	2	3	4
1	A/A	C/C	C/T	→	1	1.00	0.65	0.40
2	A/A	C/A	C/T		2	1.00	0.40	0.05
3	A/A	A/A	C/C		3	Sym.	1.00	0.25
4	A/A	C/C	C/T		4			1.00
5	A/G	C/A	C/C					

$(w_1 = 0.40, w_2 = 0.05, w_3 = 0.25)$

$g_{12} = 1 * 0.40 + 0 * 0.05 + 1 * 0.25 = 0.65$

図4. 遺伝子型関係行列

表3. 遺伝モデルの事後分布

ケース 1		ケース 2	
SNPs	Intensity	SNPs	Intensity
4,7	0.4659	3,4,7	0.6704
4,7,8	0.0525	4,7	0.0540
1,5,8	0.0495	3,4	0.0471
1,3,4	0.0393	4,6,8	0.0238
2,4,7	0.0385	2,4,7	0.0171
他の モデル	0.3543	他の モデル	0.1876

(各 10 回反復演算)

表4 A. シミュレーション実験において設定した各遺伝子型の浸透率とその推定値
(ケース 1 の場合)

SNP7		SNP4			0.3837 (0.0968)	0.3154 (0.0557)	0.2758 (0.0560)
		M/M	M/m	m/m			
M/M	0.4013	0.3085	0.2266				
					0.2480 (0.0561)	0.3258 (0.0484)	0.3951 (0.0708)
					0.1561 (0.0511)	0.3140 (0.0504)	0.5203 (0.0723)

(10 回反復演算、カッコ内の数値は推定値の標準偏差を表す)

表4 B. シミュレーション実験において設定した各遺伝子型の浸透率とその推定値
(ケース2の場合)

SNP4	SNP3									
	M/M			M/m			m/m			
	M/M	M/m	m/m	M/M	M/m	m/m	M/M	M/m	m/m	
SNP7	M/M	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	
	M/m	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	
	m/m	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	0.3085	
		0.2823 (0.0546)	0.3311 (0.0366)	0.3314 (0.0681)	0.2896 (0.0718)	0.2916 (0.0501)	0.3338 (0.0707)	0.3105 (0.0984)	0.2857 (0.0787)	0.3426 (0.0719)
		0.2979 (0.0437)	0.2969 (0.0558)	0.3311 (0.0782)	0.1321 (0.0395)	0.3117 (0.0439)	0.6009 (0.0640)	0.3109 (0.0702)	0.3111 (0.0537)	0.3095 (0.0559)
		0.2616 (0.0618)	0.2852 (0.0562)	0.3338 (0.0563)	0.3082 (0.0609)	0.3244 (0.0740)	0.3137 (0.0464)	0.3276 (0.0577)	0.2731 (0.0595)	0.3473 (0.0714)

(10回反復演算、カッコ内の数値は推定値の標準偏差を表す)

3. 研究実施体制

「感受性遺伝子同定」グループ

①研究分担グループ長：井ノ上 逸朗（東京大学、客員助教授）

②研究項目：脳動脈瘤感受性遺伝子同定

後縦靭帯骨化症感受性遺伝子同定

マイクロアレイ解析による疾患病態解明

将来予測プログラム開発

「臨床データを用いた将来予測プログラム開発」グループ

①研究分担グループ長：羽田 明（千葉大学、教授）

②研究項目：検診臨床データの収集および解析

将来予測プログラムの作成

「甲状腺癌のゲノム全塲連鎖解析」グループ

①研究分担グループ長：小島 俊男（理化学研究所、チームリーダー）

②研究項目：家族性甲状腺癌のゲノム全塲連鎖解析および疾患遺伝子同定

4. 主な研究成果の発表（論文発表および特許出願）

(1) 論文（原著論文）発表

- Krischek B, Kasuya H, Akagawa H, Tajima A, Narita A, Onda H, Hori T, Inoue I Can endothelial nitric oxide synthase gene polymorphisms identify intracranial aneurysms that are more prone to rupture in Japanese patients? *J Neurosurg*, in press.
- Akagawa H, Tajima A, Sakamoto Y, Krischek B, Yoneyama T, Kasuya H, Onda H, Hori T, Kubota M, Machida T, Saeki N, Hata A, Hashiguchi K, Kimura E, Kim C-J, Yang T-K,

- Lee J-Y, Kimm K, Inoue I. A haplotype spanning two genes, *ELN* and *LIMK1*, decreases their transcripts and confers susceptibility to intracranial aneurysms. *Hum Mol Genet*, in press
- Krischek B, Inoue I. Genetics of intracranial aneurysm. *J Hum Genet*, in press
 - Inoue I, Ikeda R, Tsukahara S. PLZF and TSG-6 identified by gene expression analysis play roles in the pathogenesis of OPLL. *J Pharmacol Sci*, in press.
 - Niikura Y, Tabata Y, Tajima A, Inoue I, Arai K, Watanabe S. Zebrafish numb homologue: phylogenetic evolution and involvement in regulation of left-right asymmetry. *Mech Dev* in press
 - Krischek B, Narita A, Akagawa H, Kasuya H, Tajima A, Onda H, Yoneyama T, Hori T, Inoue I. Is there any evidence for linkage on chromosome 17cen in affected Japanese sib-pairs with an intracranial aneurysm? *J Hum Genet*, in press.
 - Nakajima T, Wooding S, Satta Y, Jinnai N, Goto S, Hayasaka I, Saitou N, Guan-Jun J, Tokunaga K, Jorde LB, Emi M, Inoue I. Evidence for natural selection in the HAVCR1 gene: high degree of amino-acid variability in the mucin domain of human HAVCR1 protein. *Genes Immun* 6, 398–406, 2005.
 - Krischek B, Kasuya H, Inoue I. Collagen Morphology Is Not Associated With the Ala549Pro Polymorphism of the COL1A2 Gene. *Stroke* 36:2068–69, 2005.
 - Tsukahara S, Miyazawa N, Akagawa H, Forejtova S, Pavelka K, Tanaka T, Toh S, Tajima A, Akiyama I, Inoue I. COL6A1, the candidate gene for ossification of posterior longitudinal ligament, is associated with diffuse idiopathic skeletal hyperostosis in Japanese. *Spine*, 30, 2321–2324, 2005.
 - Koizumi H, Hashimoto K, Shimizu E, Mashimo Y, Hata A, Iyo M. Further analysis of microsatellite marker in the BDNF gene. *Am J Med Genet Neuropsychia Genet*, 135B, 103, 2005.
 - Lin L, Ozaki T, Takada Y, Kageyama H, Nakamura Y, Hata A, Zhang J-H, Simonds WF, Nakagawara A, Koseki H. tpors, a p53 and topoisomerase I-binding RING finger protein, is a coactivator of p53 in growth suppression induced by DNA damage. *Oncogene* 24, 3385–3396, 2005
 - Kobashi G, Ohta K, Shido K, Hata A, Yamada H, Minakami H, Fujimoto S, Kondo K, Tamashiro H. The Egogram is a Potent, Independent Risk Factor for Hypertension in Pregnancy in Japanese Women. *Seminar Thromb Hemostat*.31, 302–306, 2005
 - Kobashi G, Hata A, Shido K, Ohta K, Yamada H, Kato EH, Minakami H, Fujimoto S, Kondo K, Tamashiro H. Insertion/deletion polymorphism of angiotensin-converting enzyme gene and preeclampsia in Japanese. *Seminar Thromb Hemostat* 31, 346–350, 2005

- Hiraide A, Imazeki F, Yokosuka O, Kanda T, Kojima H, Fukui K, Suzuki Y, Hata A, Saisho H. Fas polymorphisms influence susceptibility to autoimmune hepatitis. *Am J Gastroenterol* 100, 1322–1329, 2005
- Suzuki H, Komiya A, Kamiya N, Inamoto T, Kawamura K, Miura J, Suzuki N, Nakatsu H, Hata A, Ichikawa T. Development of a nomogram to predict probability of positive initial prostate biopsy among Japanese patients. *Urology* 67, 131–136, 2006

(2) 特許出願

H17年度出願件数： 0件 (CREST研究期間累積件数： 1件)