

長期追跡健康診断データにおける遺伝・環境要因を用いた生活習慣病リスク決定法

牛田 泰徳¹・加藤 竜司¹・谷村 大輔²・井澤 英夫²・安井 健二^{3,4}・高瀬 智和⁵
吉田 安子^{5,6}・川瀬 三雄⁵・吉田 勉^{5,7}・室原 豊明²・本多 裕之^{1,6*}

¹名古屋大学工学研究科, ²名古屋大学大学院医学系研究科, ³日本ガイシ健康保険組合, ⁴あおやまクリニック,
⁵日本ガイシ株式会社, ⁶名古屋大学予防早期医療創成センター, ⁷名城大学薬学部

(2010年8月5日受付 2010年10月1日受理)

Determination of combinational genetic and environmental risk factors of lifestyle-related disease by using health check-up data obtained from long-term follow-up

Yasunori Ushida¹, Ryuji Kato¹, Daisuke Tanimura², Hideo Izawa², Kenji Yasui^{3,4}, Tomokazu Takase⁵, Yasuko Yoshida^{5,6}, Mitsuo Kawase⁵, Tsutomu Yoshida⁷, Toyooki Murohara², and Hiroyuki Honda^{1,6*} (*School of Engineering, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8603*¹; *Nagoya University School of Medicine, Tsurumaicho, Showa-ku, Nagoya 466-8550*²; *NGK Health Insurance Society, Sudacho, Mizuho-ku, Nagoya 467-8530*³; *Aoyama Clinic, Sakae 3-7-13, Naka-ku, Nagoya 460-0008*⁴; *NGK Insulators, Ltd., Sudacho, Mizuho-ku, Nagoya 467-8530*⁵; *MEXT Innovative Research Center for Preventative Medical Engineering, Furocho, Chikusa-ku, Nagoya 464-8601*⁶; *Faculty of Pharmacy, Meijo University, Yagotoyama 150, Tenpaku-ku, Nagoya 468-8503*⁷)
Seibutsu-kogaku **88**: 562-569, 2010.

Metabolic syndrome or lifestyle-related diseases develop as a result of the interaction between various genetic factors and environmental factors. Based on the health check-up data collected during a long-term follow-up (at least 7 years), we categorized a large sample population (n = 2061 subjects; men = 87%) into 3 groups (case: subjects who developed metabolic syndrome during follow-up; supercontrol: subjects free of lifestyle-related risk components; control: subjects with clinical components similar to those observed in the case subjects before follow-up). A bioinformatics approach was employed to determine the combinational genetic and environmental factors. Two types of prediction datasets were constructed to determine the predictive risk factors to discriminate between (1) case and supercontrol and between (2) case and control groups. By using logistic regression analysis, we found 25 novel risk factor combinations including 66 single nucleotide polymorphisms (SNPs) and 6 environmental factors. Moreover, to search risk factor combinations with high prediction accuracy, we used our Criterion of Detecting Personal Group (CDPG) in this study. We found that the combination of *ADIPOR1* (rs1539355) with an environment factor (smoking) was the most significant predictor of metabolic syndrome. Such risk factor combinations, and not genetic risk factors alone, could help to identify the need to modify life style for prevention of metabolic syndrome.

[**Key words:** health check-up, metabolic syndrome, lifestyle-related disease, risk factor combination, single nucleotide polymorphism]

メタボリック症候群 (metabolic syndrome, MetS) は、中心性肥満やインスリン抵抗性、脂質異常、高血圧などを含むさまざまな代謝異常が個人に集積した状態であり、心血管疾患や2型糖尿病の前兆である¹⁾。これらの生

活習慣病の発症には、遺伝要因と環境要因が複雑に相互作用しており²⁾、その要因を特定することは難しいといわれている。これまで男女を分けた解析³⁾、年齢や性別など環境要因によって補正した回帰分析⁴⁾を行うことに

*連絡先 E-mail: honda@nubio.nagoya-u.ac.jp

より、遺伝要因である一塩基多型 (single nucleotide polymorphism, SNP) に関しては生活習慣病関連のリスク要因が数多く同定されている。堀田ら⁴⁾は高度な肥満の症例908例、健常者1495例でロジスティック回帰分析を行い、*INSIG2* (rs7566605) というSNPがオッズ比1.61で日本人の肥満と関連することを同定した。これらの回帰分析では、遺伝子の関連性のみを探索している。しかしながら、MetSの解明には、遺伝要因と環境要因の両方の関係を考慮することも有効になると考えられる。

本研究では、既知の生活習慣病関連66SNPsと喫煙、飲酒など6環境因子の相互作用解析に、ロジスティック回帰分析、および著者らが開発した Criterion of Detecting Personal Group (CDPG)⁵⁾を用い、今回新しくMetSに関連する遺伝要因と環境要因を網羅的に組み合わせ解析することで、リスク要因の探索を目指した。これまで、リスク要因の探索には、ある時点でMetSに該当している群と該当していない群の疾患関連因子を比較するケース・コントロール研究が主に行われていたが⁶⁾、本研究では、企業の健康保険組合が管理する従業員2061名の過去7年以上にわたる健康診断データが解析できるチャンスを得た。本解析データは、NGK Study (NCT00408824) として、Clinicaltrials.gov⁷⁾に臨床研究登録されており、本人の個性に当たるSNPデータの入手を含めた同意が得られている。SNPデータが入手可能な大規模長期追跡健康診断データの解析は、日本において前例のないものである。リスク要因探索において、遺伝因子は変化しない素因であるため評価しやすいが、環境因子は生活スタイルや年齢で変化しやすいため、正当性を評価することが困難であることが問題とされていた。しかし、長期間の追跡調査によってコホートの健康状態の履歴 (ログ) とMetS該当の有無が明確となれば、遺伝要因だけでなく環

境要因に関しても正確に評価できると考えられ、通常の後向き試験では不可能であった、追跡開始時の環境要因のバイアスを考慮に入れた、汎化性の高いリスク要因を特定できると考えられた。このような観点から、追跡開始時と追跡終了時のMetS該当の有無を考慮したコホート分類を行って解析対象を構築し、遺伝要因と環境要因、両方の組み合わせによるリスク要因の探索、およびMetSの早期予防診断に重要なリスク要因を探索した。

解析方法

解析スキーム 解析モデル構築スキームの概要をFig. 1に示す。本研究では、日本ガイシの健康診断受診者 (2061例、男性:1803例、女性:258例) の長期追跡データを解析対象とした。まず、1999年以前から2006年まで健康診断記録のある対象1458例を抽出し解析コホートとした。1998年以前の臨床検査項目には、欠損値も多く見られるので、臨床検査項目の値が確実にそろった初年度 (1999年) から臨床検査項目がある受診者を対象とした。コホート中の、1999年以前の追跡開始時と2006年の腹囲 (追跡開始時はBMI)、血圧値、血糖値、中性脂肪、HDL コレステロール値の臨床検査項目値を調査することで、追跡開始時はMetSではなく、追跡期間中に新たにMetSに該当すると診断された症例78例 (Case)、追跡期間中、これらの臨床検査項目値が正常であり、MetSにはまったく該当しない健常者188例 (Supercontrol)、追跡開始時点ではCaseとこれらの臨床検査項目値が類似したが、追跡期間中にMetSに該当しないと診断された症例76例 (Control) を3つの群として選抜した。MetS関連リスク要因の探索では、CaseとSupercontrolの比較 (解析対象1)、CaseとControlの比較 (解析対象2) という2つの解析対象を構築し、リスク

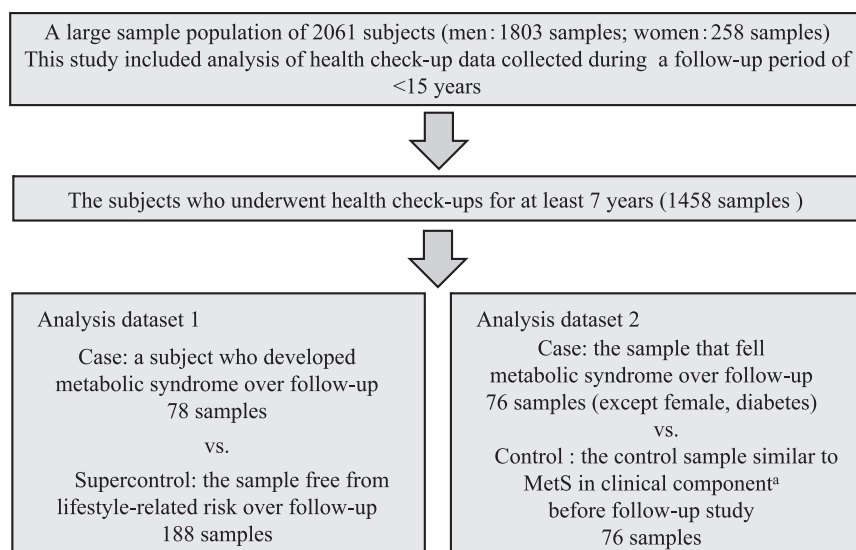


Fig. 1. Analysis model of metabolic syndrome. ^aClinical test components: age, BMI, systolic blood pressure, diastolic blood pressure, total cholesterol, triglycerides, HDL cholesterol, glucose.

要因を探索した。本解析対象は大規模な長期追跡健康診断データを用いて作成され、SNPのリスク要因まで解析できる、これまでとは比較できないほど貴重な対象である。

解析項目 ロジスティック回帰分析では、リスク要因の項目として、生活習慣病と関連報告のある66SNPs、および喫煙、飲酒、高血圧該当、糖尿病該当、高脂血症該当、高尿酸血症該当の6環境因子を用いた。SNPはmajor (AA), hetero (Aa), minor (aa) という3つのアレルタイプを、dominant (AA vs. Aa + aa), recessive (AA + Aa vs. aa) によって識別した。CDPG解析では、66SNPsと喫煙、飲酒、高尿酸血症該当の3環境因子をリスク要因項目とした。この3環境因子は、高血圧、糖尿病、高脂血症とは違い、MetSの判断基準には直接該当しないため、喫煙がリスク要因であれば、喫煙を控えるといったように、リスク要因と選択された場合にアドバイス可能な項目である。

解析手法 最初に、Case, Supercontrol (or Control) という2群ラベルを従属変数、1SNPを独立変数としたロジスティック回帰分析により、 P value < 0.05 の基準でMetS関連SNPを選択した。次に、これまでの解析では行われなかったSNPと環境因子の網羅的組み合わせ解析を行った。まず、1SNP + 1環境因子を独立変数とし、計792通り (66SNPs × 2 (dominant or recessive) × 6環境因子) の組み合わせに対して網羅的にロジスティック回帰分析を行い、SNP、環境因子とも P value < 0.05 となるMetSに関連するリスク要因を選択した。そして最後に、1SNP + 1環境因子を独立変数とし、CDPGによって、計396通り (66SNPs × 2 (dominant or recessive) × 3環境因子) の中から、MetSの予測に必要なリスク要因を探索した。CDPGは、式(1)で表わされ、複数のリスク要因から、患者の推定に必要なリスク要因のみを探索する基準である⁵⁾。本研究では、ロジスティック回帰分析(独立変数: 1SNP+1環境因子)で選択されたリスク要因から、患者がリスク要因を持ち、健常者がリスク要因を持たない要因を順番に選択した。

$$CDPG = \max \left\{ \frac{N_{Case}^m}{N_{Case}} - \frac{N_{Control}^m}{N_{Control}} \right\} \quad \text{式 (1)}$$

N_{Case} : Caseの症例数 $N_{Control}$: Controlの症例数

N_{Case}^m : m個のリスク要因のうち1つ以上に該当するCase症例数

$N_{Control}^m$: m個のリスク要因のうち1つ以上に該当するControl症例数

CDPGは発症未知者が患者になるか推定する有効な手法であり、ある発症未知者が選択したリスク要因いずれかをもった場合、患者と推定し、当てはまるリスク要因を発症要因と推定する。一方、いずれのリスク要因にも当てはまらない場合、健常者と推定することが可能になる。

以上のように、ロジスティック回帰分析(独立変数: 1SNP+1環境因子)とCDPG解析を組み合わせることに

より、これまでの解析では行われなかったSNPと環境因子の網羅的組み合わせ解析を実施し、リスク要因を探索した。

結 果

全コホートの評価 まず著者らは、全コホート情報を統計的に評価することで、解析コホートとしての妥当性を検証した。最長15年に及ぶ一般企業の健康診断受診者(2061例、男性: 1803例、女性: 258例)のうち、過去7年以上の診断記録のある対象が1458例で、そのうち現在(2006年)MetSと診断される人は151例(平均年齢44.5歳)、その割合は10.4% (=151例/1458例)であった。厚生労働省2004年国民健康・栄養調査の概要に基づく、日本におけるMetSの年齢別割合は、「30歳代で4.0%、40歳代で10.3%、50歳代で14.2% (男女の平均値)」であり、本コホートデータにおけるMetSの割合10.4%は、40歳代の全国平均10.3%とほぼ同程度であった。また、1999年以前の追跡開始時と2006年の臨床検査項目値から、7年以上の追跡期間中に新たにMetSに該当した受診者は78例、5.3% (=78例/1458例)であった。全国平均においても、解析データの追跡期間と類似した30歳代から40歳代の10年間の経年変化でMetS割合が6.3% (30歳代4.0%→40歳代10.3%)増加しており、追跡期間中のMetS増加割合も全国平均と同程度であった。以上のことから本解析コホートは、バイアスのない解析コホートであるといえる。

解析コホート評価 著者らは、全コホートデータ中からCase vs. Supercontrolの解析対象、Case vs. Controlの解析対象を選択した。そして、リスク要因を特定する前に、これらの解析対象が今回の長期追跡調査データの利点であるバイアスやノイズの少ないデータであるか、追跡調査開始時の臨床検査項目を検討した(Table 1)。Case vs. Supercontrolの比較(解析対象1)では、飲酒、糖尿病、年齢項目を除いて、解析対象間に差が見られた(P value < 0.05)。一方、Case vs. Controlの比較(解析対象2)では、高脂血症を除いて、解析対象間に差は見られなかった。このことから、Case vs. Supercontrolという対象群は、MetSという症状をより正確に診断するためのリスク項目を探索するのに有用であり、Case vs. Controlという対象群は、MetSを予知するためのリスク要因を探索するのに有用であるといえる。

ロジスティック回帰分析(独立変数: 1SNP) 解析対象1、リスク要因Case vs. Supercontrolおよび、解析対象2、Case vs. Controlの比較対象群において、生活習慣病との関連が示唆される遺伝要因(66SNPs)に関して、1SNPずつを独立変数としたロジスティック回帰分析を行った。解析対象1では、66SNPs中7SNPsで、解析対象2では、5SNPsでMetSとの有意な関連性が見られた(Table 2a)。解析対象1では、中でもADIPOR1(rs1539355) dominant

Table 1. Clinical characteristics of the case, supercontrol and control groups (Characteristics before the follow-up study).

	Analysis dataset 1					Analysis dataset 2				
	Case (N=78)		Supercontrol (N=188)		<i>P</i> value ^a	Case (N=76)		Control (N=76)		<i>P</i> value ^a
	N	%	N	%		N	%	N	%	
Sex, Male	77	98.72	152	80.85	<0.001	76	100.00	76	100.00	1
Smoking	52	66.67	82	43.62	<0.001	50	65.79	44	57.89	0.241
Drinking	64	82.05	146	77.66	0.405	62	81.58	55	72.37	0.173
Hypertension	17	21.79	0	0.00	<0.001	17	22.37	12	15.79	0.409
Diabetes	1	1.28	0	0.00	0.293	0	0.00	0	0.00	1
Hyperlipemia	25	32.05	0	0.00	<0.001	23	30.26	36	47.37	0.045
Hyperuricemia	10	12.82	2	1.06	<0.001	10	13.16	9	11.84	1

	Analysis dataset 1					Analysis dataset 2				
	Case (N=78)		Supercontrol (N=188)		<i>P</i> value ^b	Case (N=76)		Control (N=76)		<i>P</i> value ^b
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Age (year)	31.5	7.5	30.1	4.7	0.156	31.3	7.6	33.1	4.7	0.084
Height (cm)	171.8	5.8	168.7	7.1	<0.001	172.2	5.4	170.6	5.9	0.098
Weight (kg)	68.0	5.9	57.9	7.3	<0.001	68.2	5.8	66.6	6.5	0.115
BMI (kg/m ²)	23.0	1.4	20.3	1.9	<0.001	23.0	1.4	22.8	1.5	0.538
Systolic blood pressure (mmHg)	129.9	12.4	115.8	10.9	<0.001	130.0	12.5	127.5	11.2	0.203
Diastolic blood pressure (mmHg)	78.3	9.2	69.3	6.8	<0.001	78.5	9.3	76.2	5.8	0.064
Total cholesterol (mg/dl)	188.7	29.7	164.9	21.3	<0.001	187.0	28.5	192.6	33.2	0.302
Triglycerides (mg/dl)	147.9	74.7	72.0	34.2	<0.001	148.1	75.9	138.0	66.1	0.428
HDL cholesterol (mg/dl)	48.5	11.5	57.8	9.7	<0.001	48.2	11.0	50.7	7.7	0.154
Glucose (mg/dl)	92.2	9.7	87.9	7.9	0.006	92.1	9.8	94.6	7.7	0.132

^a *P* values were analyzed using Fisher's exact probability test.^b *P* values were analyzed using Welch's *t* test.

において最も有意な関連が示された（オッズ比 = 2.65, *P* value = 0.007）。オッズ比は、リスク要因に該当する場合に、MetSに該当するリスクが何倍増加するかを示した指標である。また*APOA1* (rs11216158), *APOC3* (rs2854117) は、解析対象1, 2の両方で有意な関連が示された。この解析結果は、これまでの解析では不可能であった、大規模な長期追跡健康診断データを用いて作成された、より信頼性の高い解析結果である。

ロジスティック回帰分析（独立変数：1SNP+1環境因子） 次に、これまでの解析では行われなかった SNP と環境因子の網羅的組み合わせ解析を行った。1SNP + 1環境因子を独立変数とし、網羅的にロジスティック回帰分析を行ったところ、解析対象1で25の組み合わせがリスク要因として選択された（Table 2b）。Table 2b, SNP RISK は、ロジスティック回帰分析（独立変数：1SNP）により SNP がリスクであった場合のオッズ比、environmental factor RISK は、ロジスティック回帰分析（独立変数：1環境因子）により環境因子がリスクであった場合のオッズ比、SNP & environmental factor RISK は、ロジスティック回帰分析（独立変数：1SNP+1環境因子）により SNP と環境因子の両方ともがリスクであった場合のオッズ比を示す。*LEPR* (rs1137110) は、高血圧該当、高脂血症該当、高尿酸血症該当、喫煙との4組み合わせが選択された。*ADIPOR1* (rs1539355), *APOA1* (rs11216158), *APOC3* (rs2854117) は、高血圧該当、

高尿酸血症該当、喫煙との3組み合わせが選択された。また、解析対象2において、組み合わせリスク要因は選択されなかった。

CDPGを用いたMetSの予測 最後に、上述の2つの解析対象群に対してCDPGを用いることで、MetSの予測に必要なリスク要因を探索した。

解析対象1では、ロジスティック回帰分析（独立変数：1SNP+1環境因子）で選択された組み合わせリスク要因に対して、CDPG解析を行ったところ、*ADIPOR1* (rs1539355), AAのジェノタイプと喫煙の組み合わせがMetSの推定に必要な唯一のリスク要因として選択され、このリスク要因をもつ人をCaseとして推定し、持たない人をSupercontrolとして推定した場合、正答率は70.34%であった（Fig. 2）。*ADIPOR1* (rs1539355) と喫煙の両方ともリスク要因を保有するグループ（A）とそれ以外のグループ（B）に分けた場合、AはBよりもMetSに該当するリスク（オッズ比）は4.21倍と高く、このリスク要因との関連性の検出力は、有意水準 $\alpha = 2.53 \times 10^{-4}$ (=0.05/66SNPs/3環境因子)、CaseとSupercontrolのサンプル比を本解析と同様、Case 76例、Supercontrol 187例とした場合、検出力は93.23%であった。

また、解析対象1において*ADIPOR1* (rs1539355) と喫煙のリスク要因保有と追跡前、追跡後のBMI変化の関係を調査した（Fig. 3）。*ADIPOR1* (rs1539355), 喫煙の両方ともリスク要因を保有するグループ（SNP &

Table 2. The selected risk factors with logistic regression analysis.

(a) Follow-up study results for 66 SNPs (includes all signals with $P < 0.05$).

Analysis dataset 1					RISK frequency(%)		Logistic regression analysis	
SNP identifier	Gene	Model	RISK genotype	NO RISK genotype	Case (N=78)	Supercontrol (N=188)	Odds ratio (95%CI)	P value
rs1539355	<i>ADIPOR1</i>	dominant	AA	AG + GG	86	70	2.65 (1.30-5.39)	0.007
rs7539542	<i>ADIPOR1</i>	dominant	GG	CG + CC	74	61	1.84 (1.02-3.31)	0.042
rs11216158	<i>APOA1</i>	dominant	CT + TT	CC	42	28	1.87 (1.08-3.24)	0.026
rs662799	<i>APOA5</i>	dominant	AG + GG	AA	62	48	1.74 (1.02-2.99)	0.043
rs662799	<i>APOA5</i>	recessive	GG	AA + AG	17	8	2.31 (1.04-5.11)	0.039
rs2854117	<i>APOC3</i>	recessive	CC + CT	TT	86	72	2.39 (1.17-4.88)	0.016
rs7965413	<i>VWF</i>	dominant	CT + TT	CC	74	62	1.80 (1.00-3.24)	0.050
rs1137100	<i>LEPR</i>	dominant	GG	AG + AA	72	55	2.18 (1.21-3.91)	0.009

Analysis dataset 2					RISK frequency (%)		Logistic regression analysis	
SNP identifier	Gene	Model	RISK genotype	NO RISK genotype	Case (N=76)	Control (N=76)	Odds ratio (95%CI)	P value
rs388915	<i>AGTR1</i>	dominant	AA	AG + GG	89	76	2.64 (1.07-6.51)	0.035
rs5186	<i>AGTR1</i>	dominant	AC + CC	AA	26	13	2.36 (1.02-5.45)	0.045
rs11216158	<i>APOA1</i>	dominant	CT + TT	CC	42	25	2.18 (1.09-4.35)	0.027
rs2854117	<i>APOC3</i>	recessive	CC + CT	TT	86	70	2.56 (1.15-5.73)	0.022
rs12953	<i>PECAM1</i>	dominant	CC	CT + TT	28	11	3.25 (1.33-7.89)	0.009

(b) Follow-up study results for combinations of 66 SNPs with 6 environmental factors (includes all signals with SNP $P < 0.05$ and environmental factor $P < 0.05$).

Analysis dataset 1					Odds ratio (95%CI)		
SNP identifier	Gene	RISK genotype	NO RISK genotype	Environmental factor	SNP RISK	Environmental factor RISK	SNP & environmental factor RISK
rs1539355	<i>ADIPOR1</i>	AA	AG + GG	Hypertension	2.65 (1.30-5.39)	13.3 (4.29-41.0)	35.8 (5.3-241)
rs662799	<i>APOA5</i>	GG	AA + AG	Hypertension	2.31 (1.04-5.11)	13.3 (4.29-41.0)	33.0 (4.6-238.7)
rs6141	<i>THPO</i>	CT + TT	CC	Hypertension	1.53 (0.85-2.77)	13.3 (4.29-41.0)	34.5 (5.5-218.8)
rs7539542	<i>ADIPOR1</i>	GG	GC + CC	Hypertension	1.84 (1.02-3.31)	13.3 (4.29-41.0)	32.6 (5.4-198.4)
rs11216158	<i>APOA1</i>	CT + TT	CC	Hypertension	1.87 (1.08-3.24)	13.3 (4.29-41.0)	30.8 (5.4-174.7)
rs2854117	<i>APOC3</i>	CC + CT	TT	Hypertension	2.39 (1.17-4.88)	13.3 (4.29-41.0)	38.2 (5.5-264.7)
rs1137100	<i>LEPR</i>	GG	GA + AA	Hypertension	2.18 (1.21-3.91)	13.3 (4.29-41.0)	29.0 (4.9-174)
rs1255998	<i>ESR2</i>	GG	GC + CC	Hyperlipemia	1.55 (0.88-2.71)	26.4 (9.28-75.3)	120 (16-898)
rs179983	<i>NOS3</i>	GT + TT	GG	Hyperlipemia	0.83 (0.39-1.75)	26.4 (9.28-75.3)	96.0 (10.7-858.7)
rs1799883	<i>FABP2</i>	CT + TT	CC	Hyperlipemia	1.26 (0.74-2.17)	26.4 (9.28-75.3)	92.5 (11.8-724.1)
rs1137100	<i>LEPR</i>	GG	GA + AA	Hyperlipemia	2.18 (1.21-3.91)	26.4 (9.28-75.3)	67.0 (9.0-500.5)
rs1539355	<i>ADIPOR1</i>	AA	AG + GG	Hyperuricemia	2.65 (1.30-5.39)	13.7 (2.92-64.0)	30.3 (3.1-295.2)
rs662799	<i>APOA5</i>	AG + GG	AA	Hyperuricemia	1.74 (1.02-2.99)	13.7 (2.92-64.0)	30.0 (3.6-251.8)
rs11216158	<i>APOA1</i>	CT + TT	CC	Hyperuricemia	1.87 (1.08-3.24)	13.7 (2.92-64.0)	27.8 (3.3-232.2)
rs2854117	<i>APOC3</i>	CC + CT	TT	Hyperuricemia	2.39 (1.17-4.88)	13.7 (2.92-64.0)	32.9 (3.3-327.2)
rs7965413	<i>VWF</i>	CT + TT	CC	Hyperuricemia	1.80 (1.00-3.24)	13.7 (2.92-64.0)	26.7 (3.1-233.2)
rs1137100	<i>LEPR</i>	GG	GA + AA	Hyperuricemia	2.18 (1.21-3.91)	13.7 (2.92-64.0)	47.9 (3.3-703.9)
rs1539355	<i>ADIPOR1</i>	AA	AG + GG	Smoking	2.65 (1.30-5.39)	2.77 (1.58-4.87)	11.5 (3.0-43.2)
rs662799	<i>APOA5</i>	GG	AA + AG	Smoking	2.31 (1.04-5.11)	2.77 (1.58-4.87)	7.11 (1.72-29.35)
rs1255998	<i>ESR2</i>	GG	CG + CC	Smoking	1.55 (0.88-2.71)	2.77 (1.58-4.87)	5.64 (1.76-18.10)
rs1255998	<i>ESR2</i>	GG + CG	CC	Smoking	2.08 (0.99-4.38)	2.77 (1.58-4.87)	6.24 (1.65-23.64)
rs7539542	<i>ADIPOR1</i>	GG	CG + CC	Smoking	1.84 (1.02-3.31)	2.77 (1.58-4.87)	6.38 (1.95-20.94)
rs11216158	<i>APOA1</i>	CT + TT	CC	Smoking	1.87 (1.08-3.24)	2.77 (1.58-4.87)	6.08 (1.93-19.16)
rs2854117	<i>APOC3</i>	CC + CT	TT	Smoking	2.39 (1.17-4.88)	2.77 (1.58-4.87)	6.05 (1.61-22.66)
rs1137100	<i>LEPR</i>	GG	AG + AA	Smoking	2.18 (1.21-3.91)	2.77 (1.58-4.87)	5.67 (1.73-18.52)

Smoking RISK) は、SNP のみのリスク要因を保有するグループ (SNP RISK)、喫煙のみのリスク要因を保有するグループ (Smoking RISK)、リスク要因をまったくもたないグループ (NO RISK) と比べて、追跡前後で明らかな BMI の上昇が見られた (追跡前 BMI 平均値 = 21.84

[kg/m²], 追跡後 BMI 平均値 = 23.41 [kg/m²], P value = 3.24×10^{-5}).

一方、解析対象 2 では、ロジスティック回帰分析 (独立変数: 1SNP+1 環境因子) と同様、MetS の予測に必要なリスク要因は選択されなかった。

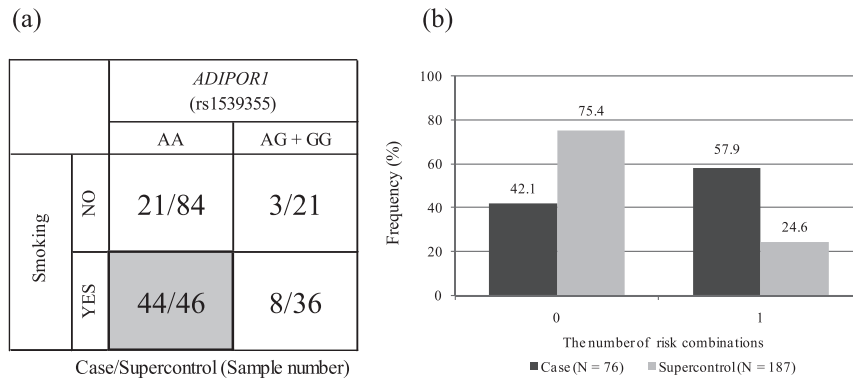


Fig. 2. Results of the CDPG combination analysis (analysis dataset 1). (a) Risk factor combinations selected by CDPG. (b) Accuracy of the prediction of metabolic syndrome by using CDPG. The prediction accuracy in analysis dataset 1 was 70.34%. Prediction accuracy is the percentage of correctly predicted numbers (MetS numbers: samples that have assigned to more than 1 risk factor by CDPG; Non-MetS numbers: samples that have not been assigned to any risk factor) in the total sample numbers.

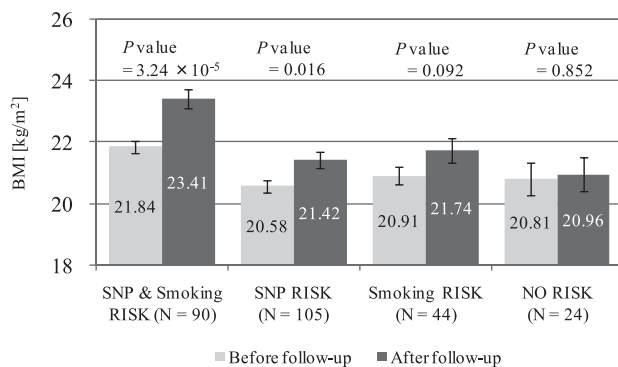


Fig. 3. Comparison of BMI variation in each risk group (analysis dataset 1). *P* values were analyzed using Welch's *t* test. The error bar indicates SEM.

解析結果のまとめ 解析結果の要約図を Fig. 4 に示す。これまでの解析では不可能であった、SNP データを含む大規模な長期追跡健康診断データを用いて、まず独立変数を 1 SNP としてロジスティック回帰分析を行うことで、MetS に特に関連する SNP として *ADIPOR1* (rs1539355), *APOA1* (rs11216158), *APOC3* (rs2854117) が選択された (Stage 1)。次に、これまでの解析では行われなかった、SNP と環境因子を網羅的に組み合わせることでロジスティック回帰分析を行うことにより、解析対象 1 から 25 の組み合わせリスク要因を選択した (Stage 2)。最後に、CDPG 解析を行うことで、*ADIPOR1* (rs1539355) と喫煙の組み合わせが MetS の推定に最も有用なリスク要因として選択された (Stage 3)。

Sample			
Analysis dataset 1	Case: 78 samples	Vs.	Supercontrol: 188 samples
Analysis dataset 2	Case: 76 samples	Vs.	Supercontrol: 76 samples
Stage. 1 Logistic regression analysis			
Independent variable: 1 SNP			
Three SNPs (<i>ADIPOR1</i> [rs1539355], <i>APOA1</i> [rs11216158], <i>APOC3</i> [rs2854117]) were strongly associated with MetS.			
Stage. 2 Multiple logistic regression analysis			
Independent variable: 1 SNP and 1 environmental factor			
Novel 25 combinations of risk factor were selected (analysis dataset 1).			
Stage.3 CDPG analysis			
The combination of <i>ADIPOR1</i> (rs1539355) with environment factor (smoking) was selected as a most significant predictor of metabolic syndrome (analysis dataset 1).			

Fig. 4. Summary of the analysis results.

考 察

長期追跡データの重要性 本研究の解析対象は2061例と非常に大規模、かつ7年以上の追跡調査結果というデータであり、MetS該当頻度も通常と同等であった。このため本解析から得られたリスク要因は、十分に他の解析対象にも応用できるものである。ロジスティック回帰分析、CDPGを用いたリスク要因の探索において、本研究では長期間の追跡調査データという利点を生かし、Case vs. Supercontrolの比較および、Case vs. Controlの比較のための解析対象を構築することができた。従来のケース・コントロール研究や大きなサンプル集団のメタ解析⁸⁾では、多くの場合比較する対象群そのものに年齢、BMI、収縮期血圧、拡張期血圧、中性脂肪、HDLなどの臨床検査項目値においてすでに大きな有意差がある。しかしながら本研究では、Caseとして追跡前はMetSでなく、追跡後にMetSに該当した症例を用いており、従来の解析対象群に含まれる背景的なバイアスやノイズを大きく低減できていると考えられる。また、解析対象2では、臨床検査項目値のほぼすべてにおいて有意差がない、すなわち背景因子のほぼ同等な対象集団を比較することができている。SNPデータが入手可能な大規模長期追跡健康診断データの解析は、日本において前例のないものであり、このようなモデルから得られたリスク要因は非常に汎化性・精度の高いものであることが期待できる。

また、今回のコホートの特徴は企業従業員であるという点が挙げられる。一般市民を対象にしたコホート研究では、定期的な追跡調査が非常に困難で脱落者が生じやすい。しかし、企業従業員をコホートとする場合、企業内の移動は考えられるものの、退職までは確実にデータが取得できる可能性が高い。このため、企業従業員のデータはコホート研究にはきわめて有用であると言える。

遺伝・環境要因の網羅的な組み合わせ解析の重要性

本研究では、これまでの解析では行われなかったSNPと環境因子の網羅的な組み合わせ解析を行うことで、*ADIPOR1* (rs1539355)と喫煙など、組み合わせとして25のMetSリスク要因を選択できた。これまでの報告では、(1) SNPがMetSと関連する、(2) 喫煙、飲酒、高血圧該当、糖尿病該当、高脂血症該当、高尿酸血症該当などの環境要因がMetSと関連する、という報告はされているが、今回のようなSNP、環境要因の組み合わせリスク要因としての報告は少ない。また網羅的な組み合わせ(独立変数：1SNP+1環境因子)に対して、ロジスティック回帰分析、CDPG解析を行うことで、たとえば*ADIPOR1* (rs1539355)は、6環境要因の中でも特に喫煙との組み合わせでリスクが高いことを発見した。このように、SNPがどの環境因子と組み合わせると最もリスクが高くなるか評価することは、これまで行われてこなかったことであり、後で述べる早期予防診断の可能性も考えると、本

研究における網羅的な組み合わせ解析は非常に重要であるといえる。

Apolipoprotein遺伝子の影響 本研究のロジスティック回帰分析(独立変数：1SNP)において、*APOA1* (Apolipoprotein A-I, rs11216158)、*APOC3* (Apolipoprotein C-III, rs2854117)が、解析対象1、解析対象2の両方からMetSの関連SNPとして選択された。*APOA1/C3/A4/A5*遺伝子は染色体11番領域に集合しており、密接に関連している。今回選択された*APOA1*はHDLの主要タンパク質として、組織から肝臓までのコレステロール流出を促進する役割を担い、*APOA1*の減少は冠動脈疾患や糖尿病に影響を与える。超低密度リポタンパク質(VLDL)タンパク質である*APOC3*は、リポタンパク質リパーゼと肝臓リパーゼを阻害し、トリグリセリドの豊富な粒子の異化を遅らせるといわれ、*APOC3*の増大は高中性脂肪血症を発生させる。また、*APOA1/C3/A4/A5*遺伝子上に存在しているSNPが中性脂肪濃度の上昇、LDL粒子サイズと関連することも知られている⁹⁾。このことから、*APOA1* (rs11216158)、*APOC3* (rs2854117)は、MetSの中でも腹囲、中性脂肪やHDLといった脂質代謝に関連した因子として選択されたのではないかと考えられる。

Adiponectinと喫煙の関係 Adiponectinは2型糖尿病や肥満、心血管疾患などで低下する唯一の脂肪由来のサイトカインであり¹⁰⁾、Adiponectin濃度とMetSの関連報告¹¹⁾、Adiponectinをコードしている*ADIPOQ*のSNPとAdiponectin濃度、2型糖尿病との関連報告があり¹²⁾、本研究で選択された*ADIPOR1* (Adiponectin receptor 1, rs1539355)がMetSと関連している可能性が十分考えられる。また、喫煙者と非喫煙者の比較、マウスへのニコチンと酸化ストレス投与実験によって、喫煙によりAdiponectin濃度が減少することが報告されている¹³⁾。検出力をみても、*ADIPOR1* (rs1539355)と喫煙のMetSとの関連を評価するのに十分なものであった。このことから、本研究解析対象1の組み合わせ解析で選択された*ADIPOR1* (rs1539355)と喫煙の組み合わせリスク要因の信頼性は高いものであるといえる。

遺伝・環境要因の組み合わせによる早期予防診断の可能性 本研究において、これまでの解析では行われなかった、SNPと環境因子を網羅的に組み合わせ、ロジスティック回帰分析(独立変数：1SNP+1環境因子)、CDPG解析を用いることで、遺伝要因と環境要因の組み合わせリスクの可視化を実現し、複数の組み合わせ要因を提案した(Table 2b, Fig. 2a)。MetSや生活習慣病は遺伝要因、環境要因が複雑に相互作用して発症するものであるため、このような両要因の組み合わせを知ることが、より正確な疾患の原因を突き止めることにつながる。また、MetSの治療改善法として生活習慣の指導が行われることが多いが、具体的でない生活全般の改善は非常にストレスが大きく、長く続かないことが多い。しかし、

本研究結果のように、生活習慣の改善項目を具体的に個々の遺伝要因と関連して指導することで、ストレスの少ない持続性の高い治療効果が期待できると考えられる。たとえば、本研究結果からは、*ADIPOR1* (rs1539355) がAAという遺伝要因を持つ人には、環境要因の中でも特に喫煙を控えるアドバイスが可能であると言える。また、本研究結果からは選択されなかったが、高尿酸血症に該当する場合には、食事をコントロールするようにアドバイスする、飲酒に該当する場合には、飲酒を控えるようアドバイスすることも可能である。さらに本解析データのように長期の追跡調査データがある場合には、追跡調査後の臨床検査項目の変化 (Fig. 3) を示すような指導ができれば生活習慣改善に対する動機付けの向上につながると期待できる。

今後の課題 本研究はコホート解析であるため、レトロスペクティブ解析で 사용되는症例と比べて、追跡期間中に新しくMetSに該当した症例が少なかった。解析対象2の組み合わせ解析でリスク要因が一つも選択されなかった要因として、症例数の少なさがあげられる。今後、本研究で得られたリスク要因の正当性の立証および新たなリスク要因の探索には、別の大型コホート研究での確認検証および更なる追跡により、症例が増えることが必要になるであろう。

また、本研究で遺伝要因の探索候補とした66SNPsは、生活習慣病との関連がこれまでの研究で示唆されたSNPに限定しており、ゲノムワイド関連解析¹⁴⁾と比べると対象SNP数は少ない。解析に用いた環境因子項目も、喫煙、飲酒、高血圧該当、糖尿病該当、高脂血症該当、高尿酸血症のみである。このため、さらなる遺伝因子、運動習慣などの環境因子の解析が、新しいリスク組み合わせの探索のために重要であろう。

要 約

メタボリック症候群などの生活習慣病は、遺伝要因と環境要因が複雑に相互作用して発症するといわれている。本研究では、最長15年に及ぶ一般企業の健康診断受診者 (2061例、男性:1803例、女性:258例) から過去7年以上の診断記録のある1458例の追跡データを解析対象とし、これまでの解析では行われなかった、遺伝因子と環境因子の網羅的な組み合わせ解析により、25のメタボリック症候群関連リスク要因を選択し、中でも*ADIPOR1* (rs1539355) と喫煙の組み合わせが最も有力なメタボリック症候群予測因子として選ばれた。本研究の解析対象は、過去7年以上の診断記録のある対象が1458例で、企業従業員という安定した長期の追跡調査データであるためMetSに該当した患者だけでなく、MetSに該当しつつある患者予備軍に関係するリスク要因を探索することができた。本研究のように遺伝因子と

環境因子とのリスク要因の組み合わせを発見することは、個々の遺伝背景に応じた限定した生活習慣改善項目を指導することが可能となるため、本解析で示すような多因子疾患の早期予防診断手法が今後有効になると期待される。

本研究は、文部科学省の平成18年度科学技術振興調整費による「先端融合イノベーション創出拠点の形成」プログラムの一環として実施された「分析・診断医工学による予防早期医療の創成」プロジェクトの研究成果の一部を取りまとめたものである。

文 献

- 1) Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., Gordon, D. J., Krauss, R. M., Savage, P. J., Smith, S. C. Jr., Spertus, J. A., and Costa, F.: *Circulation*, **112**, 2735–2752 (2005).
- 2) Willet, W. C.: *Science*, **296**, 695–698 (2002).
- 3) Yamada, Y., Izawa, H., Ichihara, S., Takatsu, F., Ishihara, H., Hirayama, H., Sone, T., Tanaka, M., and Yokota, M.: *N. Engl. J. Med.*, **347**, 1916–1923 (2002).
- 4) Hotta, K., Nakamura, M., Nakata, Y., Matsuo, T., Kamo-hara, S., Kotani, K., Komatsu, R., Itoh, N., Mineo, I., Wada, J., Masuzaki, H., Yoneda, M., Nakajima, A., Miyazaki, S., Tokunaga, K., Kawamoto, M., Funahashi, T., Hamaguchi, K., Yamada, K., Hanafusa, T., Oikawa, S., Yoshimatsu, H., Nakao, K., Sakata, T., Matsuzawa, Y., Tanaka, K., Kamatani, N., and Nakamura, Y.: *J. Hum. Genet.*, **53**, 857–862 (2008).
- 5) Tomita, Y., Asano, H., Izawa, H., Yokota, M., Kobayashi, T., and Honda, H.: *IPSJ Digital Courier*, **2**, 691–709 (2006).
- 6) McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N.: *Nat. Rev. Genet.*, **9**, 356–369 (2008).
- 7) <http://clinicaltrials.gov/ct2/show/NCT00408824>
- 8) Ye, Z., Liu, E. H., Higgins, J. P., Keavney, B. D., Lowe, G. D., Collins, R., and Danesh, J.: *Lancet*, **367**, 651–658 (2006).
- 9) Mar, R., Pajukanta, P., Allayee, H., Groenendijk, M., Dallinga-Thie, G., Krauss, R. M., Sinsheimer, J. S., Cantor, R. M., de Bruin, T. W., and Lusis, A. J.: *Circ. Res.*, **94**, 993–999 (2004).
- 10) Yaturu, S., Bridges, J. F., and Subba Reddy, D. R.: *Med. Sci. Monit.*, **12**, CR17–CR20 (2006).
- 11) Wang, J., Li, H., Franco, O. H., Yu, Z., Liu, Y., and Lin, X.: *Obesity*, **16**, 172–178 (2008).
- 12) Hivert, M. F., Manning, A. K., McAteer, J. B., Florez, J. C., Dupuis, J., Fox, C. S., O'Donnell, C. J., Cupples, L. A., and Meigs, J. B.: *Diabetes*, **57**, 3353–3359 (2008).
- 13) Iwashima, Y., Katsuya, T., Ishikawa, K., Kida, I., Ohishi, M., Horio, T., Ouchi, N., Ohashi, K., Kihara, S., Funahashi, T., Rakugi, H., and Ogihara, T.: *Hypertension*, **45**, 1094–1100 (2005).
- 14) The Wellcome Trust Case Control Consortium.: *Nature*, **447**, 661–678 (2007).