Risk factor analysis on lifestyle-related diseases by exhaustive combinational analysis of genetic and health-check data

遺伝・健診データの網羅的組み合わせ解析による生活習慣病関連要因の探索

Yasunori USHIDA

牛田　泰徳

Department of Biotechnology, School of Engineering,

Nagoya University, Japan

名古屋大学大学院　工学研究科　化学・生物工学専攻

Contents

Chapter 1

General introduction

Clinical studies on several lifestyle-related diseases have been performed. In general, clinical studies were mainly divided into cross-sectional study and cohort study.

In cross-sectional study, study groups investigate data collected at a defined time. In the case-control study, the structure of dataset consists of case subjects who developed certain disease vs. control subjects who did not develop that disease, and the analysis was performed to assess the causes of disease. Such cross-sectional study may involve lifestyle-related data, including environmental factor such as smoking, drinking, exercise or diet experience in the past. Many cross-sectional studies were conducted using large amounts of data consisting of a large number of people, and many risk factors on several lifestyle-related diseases have been identified. For example, it was reported that adiponectin levels are associated with the development of metabolic syndrome (MetS) [1]. Although cross-sectional study has the advantage in identification of risk factors, there also have difficulties in recalling past events.

On the other hand, in cohort study, study groups use the subjects who are selected

before the appearance of the disease. Then, study groups follow a group of people who do not have the disease for a period of time in comparison to those who develop this disease. Cohort studies between exposure and disease strongly aid in studying causal associations by reducing recall error. Although prospective cohort studies are expensive to conduct and need long-term follow-up, there is great benefit of yielding the most reliable results in observational epidemiology. Framingham heart study is the most famous cohort study [2], and several findings have been reported from this study. For instance, the risk of cardiovascular disease has been predicted on nine standard risk factors [3].

As risk factors of lifestyle disease, we must consider not only clinical characteristics and environmental factors but also genetic factors [4]. Genetic polymorphism is the variant that appears in at least 1% of a population and it is considered that these polymorphisms affect several lifestyle-related diseases. In the 1980s through 1990s, repeat polymorphisms and insert/delete polymorphisms were mainly studied. By investigating polymorphisms, various polymorphisms about genetic disorders such as Cystic Fibrosis [4] and Huntington's disease [5] were identified. However, the identification of polymorphisms about lifestyle-related diseases was difficult because of these complex causes. In 1992, Cambien F. *et al*. showed that an insert/delete polymorphism located in the gene encoding angiotensin-converting enzyme seems to be a potent risk factor of lifestyle-related diseases for the first time by searching 610 case subjects vs. 733 control subjects [6]. This report has created the research trend to identify various genetic polymorphisms associated with lifestyle-related diseases. The human genome project, which started in 1990 to analyze

human whole genes [7] and the SNP (single nucleotide polymorphism) Consortium which was established in 1999 to produce a public resource of SNPs [8], have identified an enormous number of SNPs. SNP can be transformed from a single base sequence (A, T, C or G) to binary data (0 or 1). The discovery of these SNPs enabled us a number of associated studies. These studies focused on the identification of candidate genes which were reported to associate with lifestyle-related diseases and led to the discovery of genetic risk factors [9-12]. Then, the International HapMap Project started in 2003 to determine the common patterns of DNA sequence variation in the human genome [13]. The effort of the Project to build a haplotype map has provided the framework of human DNA sequence variation, and led the development for simultaneous genotyping technology of over 500,000 SNPs on a few arrays (Affymetrix GeneChip 500K Mapping Array Set). The technology of large SNP typing array was applied to lifestyle-related diseases. The most famous genome-wide association study was reported by the Wellcome Trust Case Control Consortium in 2007 [14]. They conducted genome-wide association study (using 500K SNP typing array) undertaken in a very large number of people, which has examined 2,000 individuals for each of 7 major lifestyle-related diseases and a shared set of 3,000 controls. These cross-sectional case-control studies have identified 24 independent association signals in 6 diseases such as coronary artery disease and type 2 diabetes. In recent years, using genome-wide SNP array, gene-environmental interaction analysis has been conducted. For example, genome-wide gene-environment study has identified glutamate receptor gene *GRIN2A* as Parkinson's disease modifier gene via interaction with coffee [15]. Genome-wide association analyses of esophageal squamous cell carcinoma have also identified that drinkers with both of *ADH1B* and *ALDH2* SNPs

had a four-fold increase in the disease risk [16]. However, such environmental factors, which are used in interaction analysis, are specific to each disease of interest.

These studies have considerably surpassed early expectations, reproducibly identifying hundreds of variants or interactions in lifestyle-related diseases, but they have explained only a small proportion of estimated heritability [17]. Many explanations for this missing heritability have been suggested, including much larger numbers of variants of smaller effect yet to be found; rarer variants (possibly with larger effects) that are poorly detected by available genotyping arrays that focus on variants present in 1% or more of the population; low power to detect gene–gene interactions or gene-environmental interactions; and inadequate accounting for shared environment among relatives [18]. In this thesis, not only interactions between SNPs and specific environmental factors but also exhaustive combinations of genetic and environmental factors or clinical characteristics were investigated to search new combinational risk factors. Health check-up data from two large cohort studies was used to identify exact risk factors associated with the development of MetS from various lifestyle-related diseases.

In this thesis, both statistical analysis and machine learning methods to evaluate exhaustive combinations of genetic and environmental factors were conducted. In statistical analysis, each causal risk factor associated with diseases was identified by evaluating both the effect size such as odds ratio or hazard ratio which shows the strength of a phenomenon and the *P* value which shows the reliability of a phenomenon. In fact, regression analyses such as logistic regression analysis [19] and Cox proportional hazards regression [20] have been mainly performed to search the

risk factors. Not only statistical analysis enables us to identify risk factors but also the statistical analytical model can predict the development of various lifestyle-related diseases. Meanwhile, in machine learning, the data is learned to extract potential risk factors and the optimal model is constructed by developing the algorithm. Although the fact which the machine learning is available to extract risk factors is the same as statistical analysis, the former focuses on constructing better prediction model. In the analysis of clinical studies, the prediction between case subjects and control subjects is conducted by using constructed prediction model [21]. In fact, machine learning methods such as Criterion of Detecting Personal Group (CDPG) and BagPART has been developed by us [22-23]. In recent years, VARCLUS and matroid methods were implemented on a metabolic syndrome dataset to analyze the structure of metabolic risk factors [24], and two-step multifactor dimensionality reduction analysis has been applied for the gene-gene interaction analysis to bipolar disorder [25]. In previous research, the combinational modeling between physical environmental factors and human sensory evaluations in housing science was performed by mean of FNN (Fuzzy Neural Network) [26]. FNN has been applied in several fields and shown considerable flexibility in modeling of such complex phenomena as biochemical engineering processes [27–28], food science [29], protein structural science [30], and peptide science [31–32]. Both statistical analysis and machine learning are important to search risk factors associated with lifestyle-related diseases. However, the fact that common diseases are caused by complex interactions of multiple genetic and environmental factors ends up making the different solution approach between medical field and engineering field. In medical field, it is primarily important to identify what are the risk factors for lifestyle-related diseases. In other words, the data was analyzed

for every single factor. Many studies tend to investigate the strength of individual factor and the reliability of the causal dependence interpreted by the factor. Therefore, statistical analysis is more suitable for this problem solution. In engineering field, on the other hand, the concept of creative design and manufacturing has strong impact. Many studies tend to build models using potential risk factors (usually several risk factors) and distinguish case conditions from control conditions. Therefore, machine learning is more suitable for this approach.

There are differences in approach between engineering field and medical field. In this thesis, clinical study data was analyzed by using both statistical analysis and machine learning methods. Active collaboration between engineering and medicine will be promoted by identifying the same risk factors from statistical analysis and machine learning.

In this thesis, exhaustive combinational analyses of genetic and environmental factors or clinical characteristics were conducted.

In Chapter 2, exhaustive combinational analyses of genetic and environmental factors were conducted to identify risk factors of MetS. This study included 78 Case subjects who developed MetS over follow-up and 188 Supercontrol subjects who were free from lifestyle-related risk over follow-up from among 1458 Japanese employee volunteers who had long-term follow-up (at least 7 years). Firstly, multiple logistic regression analysis involving one SNP and one environmental factor as independent variable was performed to evaluate total 792 exhaustive combinations (66SNPs × 2 (dominant or recessive) × 6 environmental factors) and searched risk combinations. Moreover, CDPG in forward selection was performed to select reliable combinational

risk factors that MetS samples had and Supercontrol or control samples didn't have. In addition, the author investigated the amount of BMI change during follow-up with or without risk factors. In conclusion, the combination of *ADIPOR1* (rs1539355) with an environment factor (smoking) was identified as the most significant predictor of MetS.

In Chapter 3, exhaustive combinational analyses of clinical characteristics were conducted to find risk factors of MetS. Health check-up data from two studies was analyzed to identify the same combinational risk factors. A original study included 77 case subjects who developed MetS during the follow-up and 152 healthy control subjects who were free of lifestyle-related risk components from 1803 Japanese employee volunteers, which are also performed as in Chapter 2 (only male). A replication study included 2196 case subjects and 2196 healthy control subjects from other 31343 Japanese male employees who had long-term follow-up (8 years). Firstly, FNN was performed to search for combinational risk factors of clinical characteristics. Moreover, logistic regression analysis was performed to evaluate the reliability of these combinational risk factors statistically. In conclusion, a combination of an elevated level of γ-glutamyltranspeptidase (γ-GTP) and an elevated white blood cell (WBC) count was identified as the most significant risk factor associated with MetS. The findings that the γ-GTP level is significantly associated with habitual drinking of alcohol and that the WBC count is significantly associated with habitual smoking were also shown.

In this thesis, newly identified combinational risk factors were linked to environmental factor such as smoking or drinking. Clarifying environmental factors for preventing lifestyle diseases related to combinational risk factors can help decrease the

stress and increase the effectiveness of interventions and recommendations. For example, advising a person who has certain risk factor to especially refrain from smoking among several other environmental factors is effective. The results obtained from reliable studies would be the aide to instruct a personalized novel diagnostic and therapeutic method. The study on searching combinational risk factors is considered to contribute to the improvement of lifestyle-related diseases or the daily care of health.

References

1 Matsuzawa, Y., Funahashi, T., Kihara, S., and Shimomura, I.: **Adiponectin and metabolic syndrome.** *Arterioscler Thromb Vasc Biol*, **24,** 29-33 (2004).

2 http://www.framinghamheartstudy.org/index.html

3 Pencina, M. J., D'Agostino, R. B. Sr, Larson, M. G., Massaro, J. M., and Vasan, R. S.: **Predicting the 30-year risk of cardiovascular disease: the framingham heart study.** *Circulation*, **119,** 3078-3084 (2009).

4 Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., Drumm, M. L., Lannuzzi, M. C., Collins, F. S., and Tsui, L. C.: **Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.** *Science*, **245,** 1066-1073 (1989).

5 The Huntington's Disease Collaborative Research Group (HDCRG).: **A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.** *Cell*, **72,** 971-983 (1993).

6 King, R. A., Rotter, J. I., and Motulsky, A. G.: **The Genetic Basis of Common Diseases Vol. 20** (eds. Motulsky, A.G., Harper, P. S., Scriver, C. and Bobrow, M.), Oxford Univ. Press, Oxford, (1992).

7 Cambien, F., Poirier, O., Lecerf, L., Evans, A., Cambou, J. P., Arveiler, D., Luc, G., Bard, J. M., Bara, L., Ricard, S., Tiret, L., Amouyel, P., Alhenc-Gelas, F., and Soubrier, F.: **Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction.** *Nature*, **359,** 641-644 (1992).

8 Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie W. R., and Venter, J. C.: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science*, **252,** 1651-1656 (1991).

9 Thorisson, G. A. and Stein, L. D.: **The SNP Consortium website: past, present and future.** *Nucleic Acids Res*, **31,** 124-127 (2003).

10 Altshuler, D., Hirschhorn, J. N., Klannemark, M., Lindgren, C. M., Vohl, M. C., Nemesh, J., Lane, C. R., Schaffner, S. F., Bolk, S., Brewer, C., Tuomi, T., Gaudet, D., Hudson, T. J., Daly, M., Groop, L., and Lander, E. S.: **The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.** *Nat Genet*, **26,** 76–80 (2000).

11 Yamada, Y., Izawa, H., Ichihara, S., Takatsu, F., Ishihara, H., Hirayama, H., Sone, T., Tanaka, M., and Yokota, M.: **Prediction of the risk of myocardial infarction from polymorphisms in candidate genes.** *N Engl J Med.*, **347**, 1916-1923 (2002).

12 Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y., and Tanaka, T.: **Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction.** *Nat Genet*, **32,** 650–654 (2002).

13 The International HapMap Consortium.: **The International HapMap Project.** *Nature*, **426,** 789-796 (2003).

14 The Wellcome Trust Case Control Consortium.: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.**

Nature, **447,** 661-678 (2007).

15 Hamza, T. H., Chen, H., Hill-Burns, E. M., Rhodes, S. L., Montimurro, J., Kay, D. M., Tenesa, A., Kusel, V. I., Sheehan, P., Eaaswarkhanth, M., Yearout, D., Samii, A., Roberts, J. W., Agarwal, P., Bordelon, Y., Park, Y., Wang, L., Gao, J., Vance, J. M., Kendler, K. S., Bacanu, S. A., Scott, W. K., Ritz, B., Nutt, J., Factor, S. A., Zabetian, C. P., and Payami, H.: **Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee.** *PLoS Genet*, **7,** e1002237 (2011).

16 Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., Hu, Z., He, Z., Jia, W., Abnet, C. C., Liang, L., Hu, N., Miao, X., Zhou, Y., Liu, Z., Zhan, Q., Liu, Y., Qiao, Y., Zhou, Y., Jin, G., Guo, C., Lu, C., Yang, H., Fu, J., Yu, D., Freedman, N. D., Ding, T., Tan, W., Goldstein, A. M., Wu, T., Shen, H., Ke, Y., Zeng, Y., Chanock, S. J., Taylor, P. R., and Lin, D.: **Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions.** *Nat Genet*, **44,** 1090-1097 (2012).

17 Maher, B.: **Personal genomes: The case of the missing heritability.** *Nature*, **456,** 18-21 (2008).

18 Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A., and

Visscher, P. M.: **Finding the missing heritability of complex diseases.** *Nature*, **461,** 747-753 (2009).

19 Lemeshow, S. and Hosmer, D. W. Jr.: **A review of goodness of fit statistics for use in the development of logistic regression models.** *Am J Epidemiol*, **115,** 92-106 (1982).

20 Cox, D. R.: **Regression models and life tables.** *J R Stat Soc B*, **34,** 187-220 (1972).

21 Takahashi, H., Masuda, K., Ando, T., Kobayashi, T., and Honda, H.: **Prognostic predictor with multiple fuzzy neural models using expression profiles from DNA microarray for metastases of breast cancer.** *J Biosci Bioeng*, **98,** 193-199 (2004).

22 Tomita, Y., Asano, H., Izawa, H., Yokota, M., Kobayashi, T., and Honda, H.: **Classification method for predicting the development of myocardial infarction by using the interaction between genetic and environmental factors.** *IPSJ Digital Courier*, **2**, 691-709 (2006).

23 Kawamura, T., Takahashi, H., and Honda, H.: **Proposal of newly gene filtering method, BagPART, for gene expression analysis with small samples.** *J Biosci Bioeng*, **105,** 81-84 (2008).

24 Woolston, A., Tu, Y. K., Baxter, P. D., and Gilthorpe, M. S.: **A comparison of different approaches to unravel the latent structure within metabolic syndrome.** *PLoS One*, **7,** e34410 (2012).

25 Oh, S., Lee, J., Kwon, M. S., Weir, B., Ha, K., and Park, T.: **A novel method to identify high order gene-gene interactions in genome-wide association studies: gene-based MDR.** *BMC Bioinformatics*, **13(Suppl 9),** S5 (2012).

26 Ushida, Y., Kato, R., Tanimura, D., Izawa, H., Yasui, K., Takase, T., Yoshida, Y., Kawase, M., Yoshida, T., Murohara, T., and Honda, H.: **Determination of combinational genetic and environmental risk factors of lifestyle-related disease by using health check-up data obtained from long-term follow-up.** *Seibutsu Kogaku Kaishi*, **88,** 562–569 (2010).

27 Hanai, T., Katayama, A., Honda, H., and Kobayashi, T.: **Automatic fuzzy modeling for *Ginjo sake* brewing process using fuzzy neural networks**. *J Chem Eng Jpn*, **30,** 94–100 (1997).

28 Tomida, S., Hanai, T., Honda, H., and Kobayashi, T.: **Construction of COD simulation model for activated sludge process by recursive fuzzy neural network**. *J Chem Eng Japan*, **34,** 369–375 (2001).

29 Tominaga, O., Ito, F., Hanai, T., Honda, H., and Kobayashi, T.: **Sensory modeling of coffee with a fuzzy neural network.** *J Food Sci*, **67,** 363–368 (2002).

30 Kato, R., Nakano, H., Konishi, H., Kato, K., Koga, Y., Yamane, T., Kobayashi, T., and Honda, H.: **Novel strategy for protein exploration: High-throughput screening assisted with fuzzy neural network.** *J Mol Biol*, **351,** 683–692 (2005).

31 Kaga, C., Okochi, M., Tomita, Y., Kato, R., and Honda, H.: **Computationally assisted screening and design of cell-interactive peptides by a cell-based assay using peptide arrays and a fuzzy neural network algorithm.** *BioTechniques*, **44,** 393–402 (2008).

32 Takeshita, T., Okochi, M., Kato, R., Kaga, C., Tomita, Y., Nagaoka, S., and Honda, H.: **Screening of peptides with a high affinity to bile acids using**

peptide arrays and a computational analysis. *J Biosci Bioeng*, **112,** 92–97 (2011).

Chapter 2

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

2.1. Introduction

Metabolic syndrome (MetS) is characterized by a clustering of metabolic abnormalities including central obesity, insulin resistance, dyslipidemia, and hypertension. It has been identified as a frequent cause of the development of cardiovascular disease and type 2 diabetes mellitus [1]. Since these multifactorial diseases are caused by complex interactions between genetic and environmental factors [2], it is difficult to clarify causal relationships. Previously, stratification analysis by sex [3] and regression analysis adjusted for environmental factors such as age and sex [4] identified several single nucleotide polymorphisms (SNPs) associated with

lifestyle-related diseases. Hotta *et al.* performed logistic regression analysis using 908 severely obese subjects and 1495 normal-weight control subjects and identified that rs7566605 (*INSIG2*) is significantly associated with obesity in Japanese subjects (odds ratio = 1.61) [4]. It should be noted that these regression analyses examined only genetic associations. However, it is also very important to determine both genetic and environmental risk factors associated with MetS.

In the present study, we selected 66 SNPs selected from candidate genes potentially associated with lifestyle diseases. We also selected 6 environmental factors including smoking and drinking. We analyzed gene–environmental interactions by logistic regression analysis and our previously proposed method, the Criterion of Detecting Personal Group (CDPG) method [5]. By performing an exhaustive combinational analysis of genetic and environmental factors, we aimed to find risk factor combinations associated with MetS. So far, case–control studies comparing subjects who have MetS with those who do not are mainly used to identify risk factors [6]. Our study was based on the long-term follow-up (at least 7 years) health check-up data of 2061 volunteers who were employees managed by a company's health insurance society. This data was registered as a clinical study called the NGK study (see Clinicaltrials.gov, identifier NCT00408824 [7]). Informed consent including genome information was obtained from all participants. This was the first analysis of a large sample of follow-up health check-up data containing SNP information in Japan, making the results invaluable. When searching for risk factors of lifestyle diseases, genetic factors are easy to evaluate because they are static. On the other hand, environmental factors are difficult to evaluate properly, because they can change easily according to lifestyle or age. However, our long-term follow-up data clarified the

association between health status records and the judgment of MetS, which is impossible in normal case–control studies. This enabled us to detect more versatile risk factors while considering environmental factor bias. Ultimately, our study aimed to identify reliable combined genetic and environmental risk factors associated with MetS by constructing a suitable analysis model considering the judgment of MetS before and after follow-up. In addition, we aimed to clarify which personalized lifestyle improvement factors correspond to personal genetic factors.

2.2. Methods

2.2.1. Analytical model

The analytical model of our study is shown in Figure 1. Our study involved volunteers who were employees of NGK Insulators Ltd. (n = 2061, 1803 males, 258 females). First, as cohort data, we extracted the data of 1458 subjects who had long-term follow-up health check-up data from before 1999 to 2006. Since clinical characteristic data before 1998 were incomplete, we used complete clinical characteristic data after 1999. Clinical characteristics including body mass index (BMI), systolic blood pressure, diastolic blood pressure, fasting plasma glucose, total cholesterol, triglycerides, and HDL-cholesterol in 1999 (when the cohort study started) were analyzed. In 2006, when the cohort study finished, we investigated the clinical characteristics of waist circumference, systolic blood pressure, diastolic blood pressure, fasting plasma glucose, total cholesterol, triglycerides, and HDL-cholesterol. Then, subjects were divided into 3 categories: subjects who developed MetS over the follow-up were classified as Case group; subjects free from lifestyle-related risks over the follow-up were classified as Supercontrol group; and subjects with clinical characteristics similar to those of MetS before the follow-up study were classified as Control group. We selected 78, 188, and 76 Case, Supercontrol, and Control subjects, respectively. Moreover, we constructed 2 datasets for analysis for selected risk factors: (1) Case group vs. Supercontrol group and (2) Case group vs. Control group. These datasets were extracted from large health check-up data obtained from long-term follow-up and were invaluable for searching not only for environmental risk factors, but SNP risk factors as well.

2.2.2. Analysis of parameters

Logistic regression analysis picked up 66 SNPs selected from candidate genes potentially associated with lifestyle diseases as well as 6 environmental factors as risk factor parameters. The 6 environmental factors included smoking, drinking, hypertension, diabetes mellitus, hyperlipemia, and hyperuricemia. We categorized genotypes as major (AA), hetero (Aa), and minor (aa) into 2 models: the dominant model (AA vs. Aa+aa) and the recessive model (AA+Aa vs. aa). CDPG analysis picked up 66 SNPs and 3 environmental factors including smoking, drinking, and hyperuricemia as risk factors. These 3 environmental factors are not used as criteria for MetS; thus, they differ from hypertension, diabetes mellitus, and hyperlipemia. Therefore, we considered these environmental factors to be useful for providing advice to patients; for example, advising a person to cease smoking if they have smoking as a risk factor.

2.2.3. Analytical methods

We conducted 3 analyses. The first analysis was a logistic regression analysis involving one SNP as the independent variable and the definition of Case or Supercontrol group (or Control group) as the dependent variable. The level of significance was set at $P < 0.05$. Moreover, we performed an exhaustive combinational analysis of genetic and environmental factors. This combinational analysis was conducted using 2 methods. The second method was a multiple logistic regression analysis involving one SNP and one environmental factor as independent variables. We evaluated a total of 792 exhaustive combinations (66 SNPs × 2 [dominant/recessive] × 6 environmental factors) and selected combinational risk factors of MetS when both

the SNP and environmental factor were significant. In third method, we searched for combinational risk factors of MetS by performing CDPG involving one SNP and one environmental factor as the independent variable and evaluated a total of 396 exhaustive combinations (66 SNPs × 2 [dominant/recessive] × 3 environmental factors). CDPG is defined as equipment 1 and is used to select the minimum number of risk factors required to classify optimum development patterns from several risk factors [5]. In the present study, from the risk factors selected by logistic regression analysis (independent variable: one SNP and one environmental factor), we selected combinational risk factors in forward selection that MetS subjects had but Supercontrol and Control subjects did not.

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

N_{Case} : Number of Cases N_{Control} : Number of Controls

N_{Case}^{m} : Number of Cases with more than 1 risk factor

N_{Control}^{m} : Number of Controls with more than 1 risk factor

CDPG analysis is an effective method for predicting whether a subject will become a Case or Control. If a subject has one of the risk factors, the subject is predicted to become a Case. On the other hand, if a person does not have any risk factors, CDPG analysis predicts the person will become a Control.

In summary, we conducted an exhaustive combinational analysis of genetic and environmental factors using logistic regression analysis (independent variable: one SNP and one environmental factor) and CDPG analysis. In addition, we searched for combinational risk factors.

2.3. Results

2.3.1. Study population

First, we investigated the validity of our study population by statistically evaluating all data. Based on the health check-up data from 2061 Japanese employee volunteers (n = 2061 subjects, 1803 males, 258 females), we selected 1458 subjects who had long-term follow-up data (at least 7 years). In 2006, 151 subjects (average age: 44.5 years) were diagnosed with MetS and the percentage was 10.4% (= 151 subjects / 1458 subjects). In the outline of results from the 2004 National Health and Nutrition Examination Survey in Ministry of Health, Labour, and Welfare, 4.0%, 10.3%, and 14.2% of subjects of both sexes in Japan in their 30s, 40s, and 50s, respectively, were diagnosed with MetS. The national average of 10.3% for subjects in their 40s is almost the same as in the present study (10.4%). Moreover, by investigating the clinical characteristics in 1999 (before follow-up) and 2006 (after follow-up), we extracted 78 Case subjects that developed MetS over follow-up. The percentage of Case subjects was 5.3% (= 78 subjects / 1458 subjects). In the national average, the percentage of Case subjects increased 6.3% from subjects in their 30s to 40s (4.0% to 10.3%), similar to the present study. Therefore, our study population exhibited reliability without any bias.

2.3.2. Clinical characteristics of analysis datasets

We constructed 2 analysis datasets from out study cohort: (1) Case group vs. Supercontrol group and (2) Case group vs. Control group. Before searching for risk

factors, we investigated the clinical characteristics before follow-up in order to confirm the reliability of our analysis datasets and that they had acceptable levels of bias and noise (Table 1). In analysis dataset 1, almost all characteristics except drinking, diabetes mellitus, and age showed statistical significance ($P < 0.05$). On the other hand, in analysis dataset 2, only hyperlipemia showed statistical significance. These results demonstrate that analysis dataset 1 is useful for searching for risk factors associated with the exact diagnosis of MetS and that analysis dataset 2 is effective for searching for risk factors associated with the prognosis of MetS.

2.3.3. Logistic regression analysis (independent variable: one SNP)

We performed logistic regression analysis using analysis datasets 1 and 2 with one SNP as the independent variable. A total of 66 SNPs selected from candidate genes potentially associated with lifestyle diseases were analyzed. Of them, 7 and 5 SNPs in analysis datasets 1 and 2 were associated with MetS, respectively (Table 2a). In analysis dataset 1, *ADIPOR1* (rs1539355) dominant exhibited the most significant association (odds ratio = 2.65, $P = 0.007$). The odds ratio indicates the proportional change in risk associated with a given risk factor. Moreover, *APOA1* (rs11216158) and *APOC3* (rs2854117) showed statistical significance in both analysis datasets. These analyses were based on large health check-up data obtained from the long-term follow-up; such an analysis has not been conducted in Japan until now. Therefore, these results are reliable.

2.3.4. Logistic regression analysis (independent variable: one SNP and one environmental factor)

Next, we performed an exhaustive combinational analysis of genetic and environmental factors. We conducted a multiple logistic regression analysis involving one SNP and one environmental factor as the independent variable. In analysis dataset 1, 25 risk combinations were selected (Table 2b). The odds ratio of SNP RISK indicates the proportional change in risk associated with SNP risk, the odds ratio of environmental RISK indicates the proportional change in risk associated with environmental risk, and the odds ratio of SNP & environmental RISK indicates the proportional change in risk associated with both SNP and environmental risk. *LEPR* (rs1137110) was selected 4 times in combinations with hypertension, hyperlipemia, hyperuricemia, and smoking. *ADIPOR1* (rs1539355), *APOA1* (rs11216158), and *APOC3* (rs2854117) were selected 3 times in combinations with hypertension, hyperuricemia, and smoking. Meanwhile, no combinations of risk factors were significant in analysis dataset 2.

2.3.5. Predicting MetS using CDPG

Finally, we used CDPG analysis on the 2 datasets to search for essential risk factors to predict MetS. In analysis dataset 1, we used the risk combinations selected by logistic regression analysis (independent variable: one SNP and one environmental factor). The combination of *ADIPOR1* (rs1539355) AA genotype with an environment factor (smoking) was the only essential risk factor for the prediction of MetS. If the subjects that had and did not have this risk combination were predicted to be Case subjects and Supercontrol subjects, respectively, the accuracy was 70.34% (Figure. 2). The odds ratio for group A subjects who had the risk combination of *ADIPOR1* (rs1539355) and smoking versus group B subjects who did not have this risk

combination was 4.21; that is, those in the former group are 4.21 times as likely to have MetS than those in the latter group. The statistical power of this risk combination with MetS was 93.23% when the significance level is $\alpha = 2.53 \times 10^{-4}$ (= 0.05 / 66 SNPs / 3 environmental factors) and the population proportion equals to the number of 76 Case subjects and 187 Supercontrol subjects.

Moreover, we investigated the change in BMI during the follow-up in subjects with and without *ADIPOR1* (rs1539355) and smoking risk combination in analysis dataset 1 (Figure 3). BMI clearly increased in the subjects who had both *ADIPOR1* (rs1539355) and smoking risk factors (SNP & Smoking RISK) compared with the subjects who had only the SNP risk factor (SNP RISK), only the smoking risk factor (Smoking RISK), and no risk factors (NO RISK) (mean BMI before follow-up = 21.84 kg/m^2, after follow-up = 23.41 kg/m^2, $P = 3.24 \times 10^{-5}$).

Meanwhile, in analysis dataset 2, there was no combination of risk factors similar to those found in the logistic regression analysis (independent variable: one SNP and one environmental factor).

2.3.6. Summary of results

The results are summarized in Figure 4. We used large health check-up data obtained from the long-term follow-up of subjects that included SNP information. First, logistic regression analysis of single SNPs revealed 3 SNPs (*ADIPOR1* [rs1539355], *APOA1* [rs11216158], and *APOC3* [rs2854117]) strongly associated with MetS (Stage 1). Second, exhaustive multiple logistic regression analysis of genetic and environmental factors revealed 25 novel risk factor combinations in analysis dataset 1 (Stage 2). Finally, CDPG analysis revealed that the combination of *ADIPOR1*

(rs1539355) with environment factor (smoking) was the most significant predictor of MetS in analysis dataset 1 (Stage 3).

2.4. Discussion

2.4.1. Importance of long-term follow-up data

The present study was based on large (2061 subjects) and long-term follow-up (at least 7 years) data, and the prevalence of MetS was the same as reported previously. Therefore, the risk factors found in the present study are applicable to other studies. In logistic regression analysis and CDPG analysis, by leveraging the strength of long-term follow-up data, we constructed 2 analysis datasets: (1) Case group vs. Supercontrol group and (2) Case group vs. Control group. In previous case–control studies and large meta-analyses [8], Case subjects were almost always different from Control subjects with respect to clinical characteristics such as age, BMI, systolic arterial pressure, diastolic blood pressure, triglycerides, and HDL-cholesterol. However, we classified subjects who developed MetS over follow-up as Case subjects. This was considered to greatly reduce background bias and noise, which impacted previous studies. In analysis dataset 2, almost no clinical characteristics showed significance. These results indicate that the Case subjects and Control subjects had similar backgrounds. The present study is the first analysis of large health check-up data obtained from long-term follow-up including SNP information in Japan. The risk factors obtained from these analysis datasets appears to be versatile and extremely precise.

It should be noted that our cohort comprised volunteers from a single company. The long-term follow-up of a cohort of general citizens is difficult because of subjects dropping out. However, although transfers likely occurred in a cohort study of

company employees, follow-up data may be obtained until retirement. Thus, the health check-up data of company employees is very useful in cohort studies.

2.4.2. Importance of exhaustive combinational analysis of genetic and environmental factors

In the present study, the exhaustive combinational analysis of genetic and environmental factors selected 25 risk factor combinations for MetS, including the combination of *ADIPOR1* (rs1539355) and smoking. Previous studies report that certain SNPs are associated with MetS and that environmental factors including smoking, drinking, hypertension, diabetes mellitus, hyperlipemia, and hyperuricemia are associated with MetS. However, there are few results regarding combinations of genetic and environmental risk factors. Moreover, exhaustive combinational analysis of logistic regression analysis (independent variable: one SNP and one environmental factor) and CDPG analysis revealed that smoking is more strongly associated with MetS than any other environmental factor in combination with *ADIPOR1* (rs1539355). Thus, the environmental factor most strongly associated with lifestyle diseases in combination with SNPs was not determined. Considering the possibility of early preventive and diagnostic methods of lifestyle-related diseases described later on, the exhaustive combinational analysis in the present study may be very important.

2.4.3. Effect of the apolipoprotein gene

The logistic regression analysis (independent variable: single SNP) revealed that *APOA1* (Apolipoprotein A-I, rs11216158) and *APOC3* (Apolipoprotein C-III, rs2854117) were associated with MetS in both analysis datasets 1 and 2. The

APOA1/C3/A4/A5 gene cluster is on chromosome 11. These genes also interact with each other. *APOA1*, which was selected in the present study, encodes the major protein of HDL-cholesterol and promotes cholesterol efflux from tissues to the liver for excretion. Defects in *APOA1* are associated with coronary artery disease and diabetes mellitus. *APOC3* encodes a protein component of very-low-density lipoprotein and inhibits lipoprotein lipase and hepatic lipase. *APOC3* is considered to inhibit the translation of triglyceride-rich particles. An increase in *APOC3* induces the development of hypertriglyceridemia. Moreover, SNPs in the *APOA1/C3/A4/A5* gene cluster region are reported to increase plasma triglyceride levels and be associated with LDL-cholesterol particle size [9]. These findings suggest *APOA1* (rs11216158) and *APOC3* (rs2854117) may be related to lipid metabolism risk factors such as abdominal circumference, triglycerides, and HDL-cholesterol in MetS.

2.4.4. Association between adiponectin genes and smoking

Adiponectin is the only adipose-derived cytokine decreased in individuals with obesity, type 2 diabetes, and coronary artery disease [10]. It is reported that adiponectin levels are associated with the development of MetS [11]. Furthermore, SNPs in the *ADIPOQ* gene region, which encodes adiponectin, are reported to be associated with adiponectin levels and type 2 diabetes [12]. These previous results suggest *ADIPOR1* (Adiponectin receptor 1, rs1539355), which was selected in the present study, is associated with the development of MetS. Furthermore, an epidemiological survey comparing smokers and never-smokers and a study inducing nicotine or oxidative stress in 3T3-L1 adipocytes demonstrate that a smoking habit is associated with lower adiponectin levels [13]. The statistical power of the present analysis was sufficient to

clarify the combination of *ADIPOR1* (rs1539355) and smoking is associated with the development of MetS. Therefore, the combination of *ADIPOR1* (rs1539355) and smoking selected in analysis dataset 1 is reliable risk factor.

2.4.5. Early preventive and diagnostic methods using combinations of genetic and environmental factors

In the present study, we performed logistic regression analysis (independent variable: one SNP and one environmental factor) and CDPG analysis. This combinatorial analysis has never been conducted previously. We analyzed risk factor combinations of genetic and environmental factors (Figure 2a) and found several risk factor combinations related to MetS (Table 2b). MetS and lifestyle-related diseases develop as a result of complex interactions between various genetic and environmental factors. Therefore, the findings of these risk factor combinations help identify the causes of diseases more precisely. Although several recommendations for lifestyle-related diseases are given to prevent the development of MetS, proposing recommendations that are not objective create substantial stress and noncompliance in patients. However, clarifying factors for preventing lifestyle diseases related to personal genetic factors, which was the objective of the present study, can help decrease the stress and increase the effectiveness of interventions and recommendations. For example, from the present findings, we could advise a person who has the genetic risk factor *ADIPOR1* (rs1539355) AA genotype to especially refrain from smoking among several other environmental factors. Moreover, although the present study does not provide specific advice regarding genetic risk factors, we could instruct a person with hyperuricemia to their control diet or a person who have

drinking habit to refrain drinking for example. Furthermore, when long-term follow-up data are used such as in the present study, instructional methods that show changes in clinical characteristics during follow-up (e.g., Figure 3) may be useful for motivating patients to make positive lifestyle changes.

2.4.6. Future research topics

In the present cohort study, few subjects developed MetS over the follow-up compared with retrospective studies. Furthermore, no risk factor combinations were selected from analysis dataset 2, partly because of the small number of Case subjects. Therefore, a long-term follow-up or other large cohort studies may be required to confirm the risk factor combination found in the present study and identify other risk factor combination.

Moreover, the present study used only 66 SNPs selected from candidate genes potentially associated with lifestyle diseases. This is a small number of SNPs compared with those used in genome-wide association studies [14]. Furthermore, only 6 environmental factors were analyzed in the present study: smoking, drinking, hypertension, diabetes mellitus, hyperlipemia, and hyperuricemia. Therefore, analyses including other genetic and environmental factors such as exercise are essential in order to find new risk factor combinations.

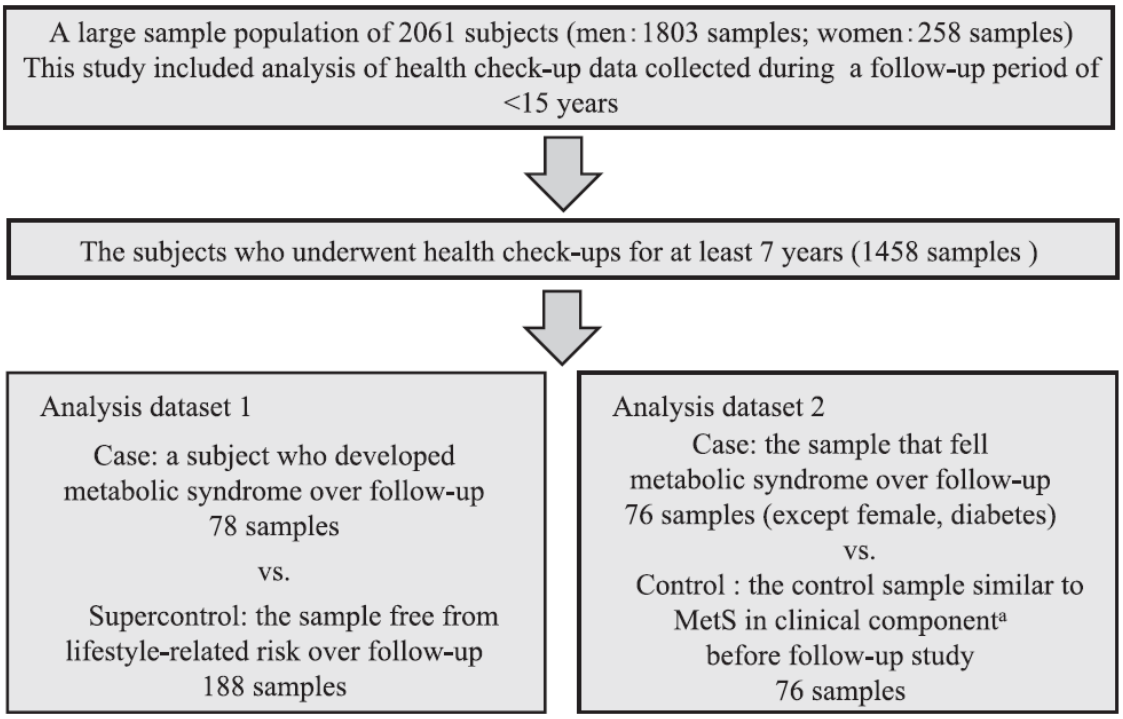

Figure 1 Analysis model of metabolic syndrome

[a]Clinical test components : age, BMI, systolic blood pressure, diastolic blood pressure, total cholesterol, triglycerides, HDL cholesterol, glucose

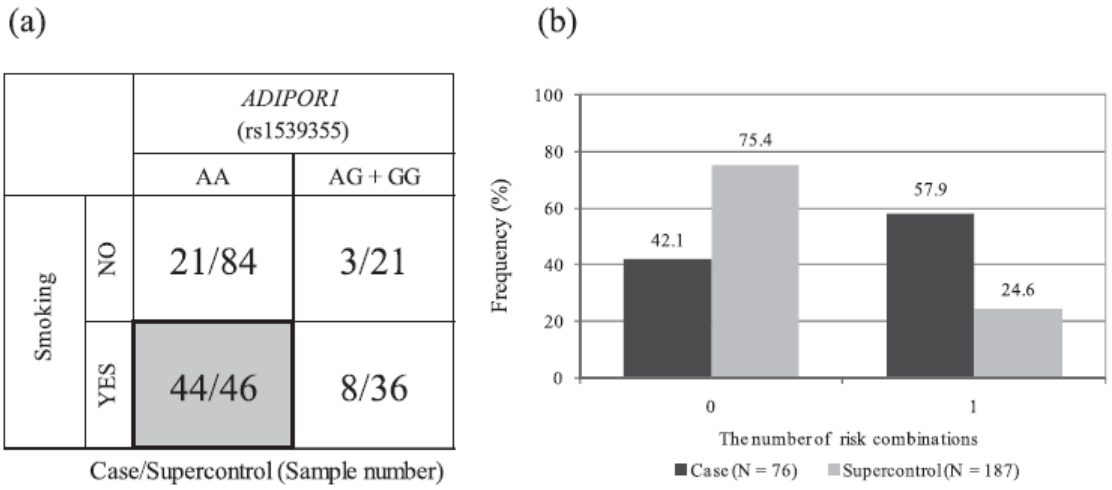

Figure 2 Results of the CDPG combination analysis (analysis dataset 1)

(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: subjects that have assigned to more than 1 risk factor by CDPG; Non-MetS numbers: subjects that have not been assigned to any risk factor) in the total subjects.

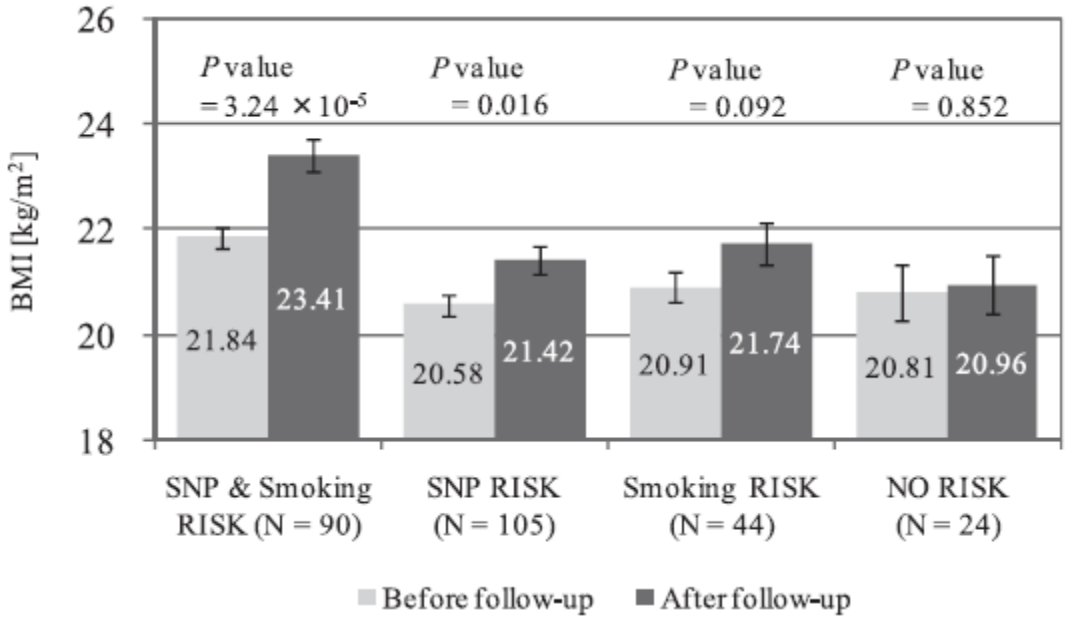

Figure 3 Comparison of BMI variation in each risk group (analysis dataset 1)

P values were analyzed using Welch's *t* test. The error bar indicates SEM.

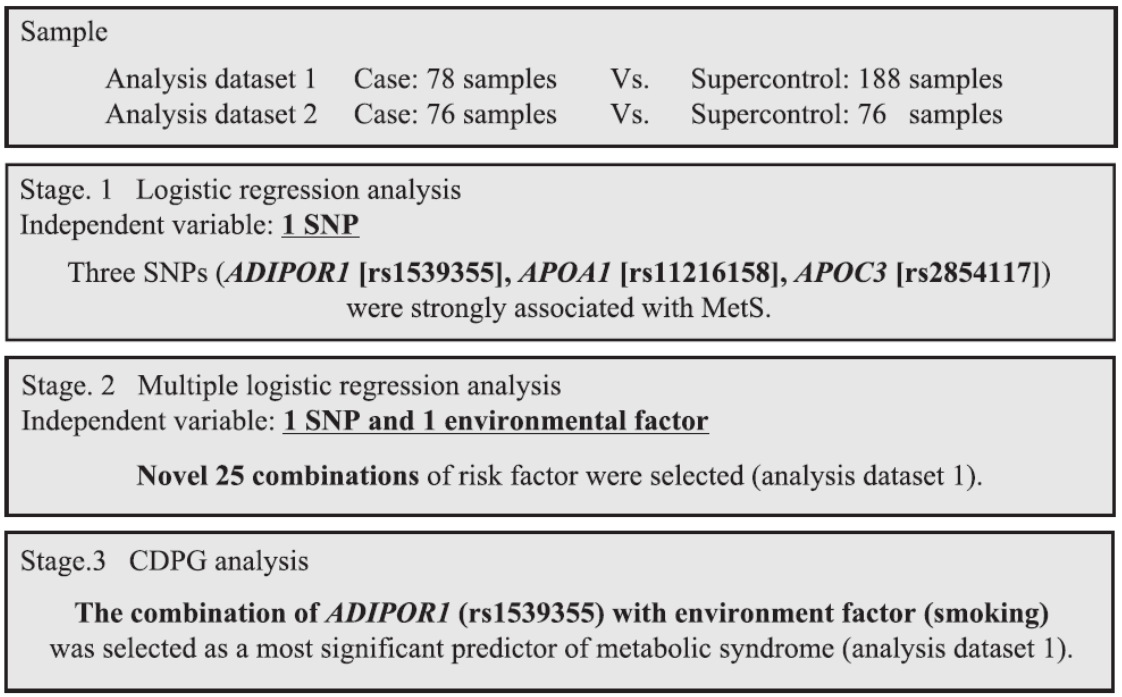

Figure 4 Summary of the analysis results

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)			Case (N=76)		Control (N=76)		
	N	%	N	%	*P* value[a]	N	%	N	%	*P* value[a]
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)			Case (N=76)		Control (N=76)		
	Mean	SD	Mean	SD	*P* value[b]	Mean	SD	Mean	SD	*P* value[b]
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

Table 2 The selected risk factors with logistic regression analysis

(a)Follow-up study results for 66 SNPs (includes all signal 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	*ADIPOR1*	dominant	AA	AG+GG	86	70	2.65(1.30-5.39)	0.007
rs7539542	*ADIPOR1*	dominant	GG	CG+CC	74	61	1.84(1.02-3.31)	0.042
rs11216158	*APOA1*	dominant	CT+TT	CC	42	28	1.87(1.08-3.24)	0.026
rs662799	*APOA5*	dominant	AG+GG	AA	62	48	1.74(1.02-2.99)	0.043
rs662799	*APOA5*	recessive	GG	AA+AG	17	8	2.31(1.04-5.11)	0.039
rs2854117	*APOC3*	recessive	CC+CT	TT	86	72	2.39(1.17-4.88)	0.016
rs7965413	*VWF*	dominant	CT+TT	CC	74	62	1.80(1.00-3.24)	0.050
rs1137100	*LEPR*	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	NORISK genotype	Case (N=76)	Control (N=76)	Odds Ratio (95%CI)	*P* value
rs388915	*AGTR1*	dominant	AA	AG+GG	89	76	2.64(1.07-6.51)	0.035
rs5186	*AGTR1*	dominant	AC+CC	AA	26	13	2.36(1.02-5.45)	0.045
rs11216158	*APOA1*	dominant	CT+TT	CC	42	25	2.18(1.09-4.35)	0.027
rs2854117	*APOC3*	recessive	CC+CT	TT	86	70	2.56(1.15-5.73)	0.022
rs12953	*PECAM1*	dominant	CC	CT+TT	28	11	3.25(1.33-7.89)	0.009

(b) Follow-up study results for combination 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	*ADIPOR1*	AA	AG+GG	Hypertension	2.65(1.30-5.39)	13.3(4.29-41.0)	35.8(5.3-241)
rs662799	*APOA5*	GG	AA+AG	Hypertension	2.31(1.04-5.11)	13.3(4.29-41.0)	33.0(4.6-238.7)
rs6141	*THPO*	CT+TT	CC	Hypertension	1.53(0.85-2.77)	13.3(4.29-41.0)	34.5(5.5-218.8)
rs7539542	*ADIPOR1*	GG	GC+CC	Hypertension	1.84(1.02-3.31)	13.3(4.29-41.0)	32.6(5.4-198.4)
rs11216158	*APOA1*	CT+TT	CC	Hypertension	1.87(1.08-3.24)	13.3(4.29-41.0)	30.8(5.4-174.7)
rs2854117	*APOC3*	CC+CT	TT	Hypertension	2.39(1.17-4.88)	13.3(4.29-41.0)	38.2(5.5-264.7)
rs1137100	*LEPR*	GG	GA+AA	Hypertension	2.18(1.21-3.91)	13.3(4.29-41.0)	29.0(4.9-174)
rs1255998	*ESR2*	GG	GC+CC	Hyperlipemia	1.55(0.88-2.71)	26.4(9.28-75.3)	120(16-898)
rs1799983	*NOS3*	GT+TT	GG	Hyperlipemia	0.83(0.39-1.75)	26.4(9.28-75.3)	96.0(10.7-858.7)
rs1799883	*FABP2*	CT+TT	CC	Hyperlipemia	1.26(0.74-2.17)	26.4(9.28-75.3)	92.5(11.8-724.1)
rs1137100	*LEPR*	GG	GA+AA	Hyperlipemia	2.18(1.21-3.91)	26.4(9.28-75.3)	67.0(9.0-500.5)
rs1539355	*ADIPOR1*	AA	AG+GG	Hyperuricemia	2.65(1.30-5.39)	13.7(2.92-64.0)	30.3(3.1-295.2)
rs662799	*APOA5*	AG+GG	AA	Hyperuricemia	1.74(1.02-2.99)	13.7(2.92-64.0)	30.0(3.6-251.8)
rs11216158	*APOA1*	CT+TT	CC	Hyperuricemia	1.87(1.08-3.24)	13.7(2.92-64.0)	27.8(3.3-232.2)
rs2854117	*APOC3*	CC+CT	TT	Hyperuricemia	2.39(1.17-4.88)	13.7(2.92-64.0)	32.9(3.3-327.2)
rs7965413	*VWF*	CT+TT	CC	Hyperuricemia	1.80(1.00-3.24)	13.7(2.92-64.0)	26.7(3.1-233.2)
rs1137100	*LEPR*	GG	GA+AA	Hyperuricemia	2.18(1.21-3.91)	13.7(2.92-64.0)	47.9(3.3-703.9)
rs1539355	*ADIPOR1*	AA	AG+GG	Smoking	2.65(1.30-5.39)	2.77(1.58-4.87)	11.5(3.0-43.2)
rs662799	*APOA5*	GG	AA+AG	Smoking	2.31(1.04-5.11)	2.77(1.58-4.87)	7.11(1.72-29.35)
rs1255998	*ESR2*	GG	CG+CC	Smoking	1.55(0.88-2.71)	2.77(1.58-4.87)	5.64(1.76-18.10)
rs1255998	*ESR2*	GG+CG	CC	Smoking	2.08(0.99-4.38)	2.77(1.58-4.87)	6.24(1.65-23.64)
rs7539542	*ADIPOR1*	GG	CG+CC	Smoking	1.84(1.02-3.31)	2.77(1.58-4.87)	6.38(1.95-20.94)
rs11216158	*APOA1*	CT+TT	CC	Smoking	1.87(1.08-3.24)	2.77(1.58-4.87)	6.08(1.93-19.16)
rs2854117	*APOC3*	CC+CT	TT	Smoking	2.39(1.17-4.88)	2.77(1.58-4.87)	6.05(1.61-22.66)
rs1137100	*LEPR*	GG	AG+AA	Smoking	2.18(1.21-3.91)	2.77(1.58-4.87)	5.67(1.73-18.52)

2.5. Summary

Metabolic syndrome or lifestyle-related diseases develop as a result of the interaction between various genetic factors and environmental factors. The present study, based on the health check-up data of Japanese volunteers who were employees of a company (n = 2061 subjects, 1803 males, 258 females), we selected 1458 subjects who had long-term follow-up data (at least 7 years); thus, the subjects were stable. Exhaustive combinational analyses of genetic and environmental factors, which combinatorial analysis has never been conducted in previously, revealed 25 risk combinations. Moreover, we identified that the combination of *ADIPOR1* (rs1539355) with an environment factor (smoking) was the most significant predictor of metabolic syndrome. Present study had 1458 subjects who had long-term follow-up (at least 7 years) and they were stable data because of company employee volunteers. This enabled us to find risk factors associated with not only MetS but also potential MetS sufferers. Detecting combined genetic and environmental risk factors could help provide personalized lifestyle improvement factors that correspond to personal genetic factors. Such personalized early preventive and diagnostic methods for lifestyle-related diseases might be useful in the near future.

2.6. References

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.: **Diagnosis and management of the metabolic syndrome: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement.** *Circulation*, **112,** 2735-2752 (2005).
2. Willet, W. C.: **Balancing life-style and genomics research for disease prevention.** *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.: **Prediction of the risk of myocardial infarction from polymorphisms in candidate genes.** *N Engl J Med*, **347,** 1916-1923 (2002).
4. Hotta, K., Nakamura, M., Nakata, Y., Matsuo, T., Kamohara, 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.: **INSIG2 gene rs7566605 polymorphism is associated with severe obesity in Japanese.** *J Hum Genet*, **53,** 857-862 (2008).
5. Tomita, Y., Asano, H., Izawa, H., Yokota, M., Kobayashi, T., and Honda, H.: **Classification method for predicting the development of myocardial infarction by using the interaction between genetic and environmental**

factors. *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.: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges.** *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.: **Seven haemostatic gene polymorphisms in coronary disease: meta-analysis of 66,155 cases and 91,307 controls.** *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.: **Association of the APOLIPOPROTEIN A1/C3/A4/A5 gene cluster with triglyceride levels and LDL particle size in familial combined hyperlipidemia.** *Circ Res*, **94,** 993-999 (2004).
10. Yaturu, S., Bridges, J. F., and Subba Reddy, D. R.: **Decreased levels of plasma adiponectin in prediabetes, Type 2 diabetes and coronary artery disease.** *Med Sci Monit*, **12,** CR17-CR20 (2006).
11. Wang, J., Li, H., Franco, O. H., Yu, Z., Liu, Y., and Lin, X.: **Adiponectin and Metabolic Syndrome in Middle-aged and Elderly Chinese.** *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.: **Common variants in the adiponectin gene (ADIPOQ) associated with plasma adiponectin**

levels, type 2 diabetes, and diabetes-related quantitative traits: the Framingham Offspring Study. *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.: **Association of hypoadiponectinemia with smoking habit in men.** *Hypertension*, **45,** 1094-1100 (2005).

14. The Wellcome Trust Case Control Consortium.: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature*, **447,** 661-678 (2007).

Chapter 3

Combinational risk factors of metabolic syndrome identified by fuzzy neural network analysis of health-check data

3.1. Introduction

Metabolic syndrome (MetS) is characterized by a clustering of metabolic abnormalities, including glucose intolerance, insulin resistance, central obesity, dyslipidemia, and hypertension, and it has been identified as a frequent cause to the development of cardiovascular disease [1]. The prevalence of MetS in Japan has increased over recent decades as a result of changes in diet and physical activity [2]. To investigate the relationship between diet or physical activity and risk marker plays effective roles in finding the most suitable lifestyle factor to improve developing MetS. It is useful for proposing a personalized diagnostic and therapeutic method. There is also an urgent need to establish an appropriate and sensitive screening marker to identify individuals at a high risk of developing MetS, thereby preventing a further

increase in its incidence. So far, indices such as the low-density lipoprotein (LDL) to high-density lipoprotein (HDL) ratio (L/H) [3] or the ratio of adiponectin to homeostasis model assessment–insulin resistance (adiponectin/HOMA-IR ratio) [4] have been proposed as combinational risk factors. We have also reported that a combination of adiponectin receptor 1 (ADIPOR1; rs1539355) with an environmental factor (smoking habit) is suitable as a combinational risk factor for MetS [5]. There is, however, a need to identify new combinational risk factors.

In this study, we used a fuzzy neural network (FNN) in a bioinformatics approach to search for complex risk characteristics. Hirose et al. predicted the prevalence of MetS using artificial neural network [6]. The FNN is one of artificial neural network models that have been used in medical research as a powerful tool for the accurate detection of causal relationships [7–10]. FNN analysis has two main advantages. The first is its ability to select parameters on the basis of a parameter-increase method to permit the identification of the most influential parameters in the data. FNN analysis has the same predictable ability as multiple logistic regression. The second is its ability to extract predictive rules called fuzzy rule that can predict objective properties to reproduce the results.

So far, FNN has shown considerable flexibility in modeling of such complex phenomena as biochemical engineering processes (modeling of links between process valuables and process outputs) [11–12], food science (modeling of links between chemical components and sensory evaluation) [13], protein structural science (modeling of links between amino acid sequences and enzyme function) [14], housing science (modeling of links between physical environmental factors and human sensory evaluations) [15], and peptide science (modeling of links between peptide sequences

and peptide functions) [16–17]. We therefore conjectured that FNN might serve as a suitable method for identifying specific characteristics that affect the pathogenesis of MetS.

The present study had two chief merits. The first was that the studies were based on subjects receiving health check-ups rather than on clinical patients; this has the advantage that periodical health examination is free of model bias, so that our results apply to the general population. The second merit was the high quality of our data because the relevant studies involved large numbers of subjects who were followed over a long time (at least seven years).

Overall, the aim of our studies was to identify reliable combinational risk factors associated with MetS by using an FNN and to contribute to the prevention of MetS by mitigating the identified risk combination.

3.2. Methods

3.2.1. Study design

To identify a significant combination of factors associated with MetS, we performed a two-stage study. The clinical characteristics before study start are summarized in Tables 1 and 2. In the original study, we selected 77 case subjects and 152 healthy control subjects from among 1803 Japanese male employees [5]. This longitudinal study was conducted by using health check-up data collected during a long-term follow-up (at least seven years). A replication study involved 2196 case subjects and 2196 healthy control subjects from among another 31343 other Japanese male employees. This study was also a longitudinal one and was conducted over eight years. All studies were performed according to the guidelines of the Declaration of Helsinki. Informed consent was obtained from all participants, and the studies were approved by Nagoya University School of Medicine. In both studies, we used clinical data before follow-up to predict the cause of MetS.

3.2.2. Definitions of case, healthy control and normal control

We used the criteria proposed by the Japan Society for the Study of Obesity (JASSO) [18] to identify subjects with MetS and supercontrol subjects.

1) Obesity: Waist circumference ≥85 cm in men or body-mass index (BMI) ≥25 kg/m^2 if the waist circumference was not measured.

2) Raised blood pressure: systolic blood pressure ≥130 mmHg and/or diastolic blood pressure ≥85 mmHg.

3) Dyslipidemia: triglyceride ≥150 mg/dL and/or HDL cholesterol < 40 mg/dL

4) Raised fasting glucose: fasting glucose ≥110 mg/dL

Subjects were classified as suffering from MetS if they were obese and they showed any two of the other three criteria. Subjects who were free of any of the risk components were classified as supercontrol. Then we defined case, healthy control and normal control according to the criteria below.

Case: Subjects who developed MetS during the follow-up.

Healthy control: Subjects who were remained as supercontrol during the follow-up.

Normal control: Subjects who weren't MetS during follow-up.

Subjects with antihypertensive, lipid-lowering and anti-diabetic agents were excluded from analysis.

3.2.3. Measurements

The baseline health examination performed before follow-up included physical measurements, serum biochemical measurements, urine measurements, medication use and a questionnaire. Physical measurements of height, weight and body mass index were measured in the fasting state. Blood samples were obtained from subjects in the fasted condition for serum biochemical measurements. After the subject had rested for 10 min in sitting position, 14ml of blood were collected from the antecubital vein into tubes containing EDTA. After blood samples were sent to the clinical laboratory testing company, biochemical measurements were determined according to standard laboratory procedures. Biochemical measurements collected in this study include;

(1) Lipids: total cholesterol, triglyceride and HDL-cholesterol.

(2) Carbohydrate: glucose.

(3) Non-protein nitrogenous compounds: urea nitrogen, creatinine and uric acid.

(4) Serum enzymes: γ-glutamyltranspeptidase (γ-GTP), glutamic-oxyacetic transaminase (GOT), glutamic-pyvuvic transaminase (GPT).

(5) Hematology: red blood cells (RBC), hemoglobin, hematocrit and white blood cells (WBC).

Urine samples were also collected in the morning. After urine samples were sent to clinical laboratory testing company, urine uribilinogen, urine protein, urine sugar and urinary occult blood were measured. Medication use was assessed by the examining physicians. Drinking habit and smoking habit were collected by standard questionnaire. The questionnaire asked about the frequency of alcohol consumption on a weekly basis and smoking habit (never, past or current smoker). Drinking habit was defined as the subject who drank once a week and more. Smoking habit was defined as past or current smoker. In replication study, exercise habit was also divided into four categories by the time of exercise per week; exercising every day, exercising twice or more a week, exercising once a week and no-exercising. The aim of our studies was to identify risk factors from routine health check-up parameters generally measured. Therefore, the well-defined risk factors such as insulin weren't measured in our study.

3.2.4. FNN analysis

We conducted an FNN analysis to identify any significant combinations that are associated with MetS. The procedure for constructing the model is shown schematically in Figure 1. This model has two inputs x_1 and x_2, one output y^*, and two membership functions f_{low} and f_{high} in each premise. FNN has three kinds of connection weights: W_c, W_g, and W_f [19]. The connection weights W_c and W_g determine the positions and gradients of the sigmoid functions; these decide the grade of each

membership function f_{low} or f_{high} by means of the formula shown below,

$$f(x) = 1/[1+exp\{- W_g(x + W_c)\}] \quad (1)$$

where x is input value and $f(x)$ is the product of the grade of membership function. The products of the grades are fed to the next unit Π. W_f is the weight of each production rule and decides the output y^* by means of the sum of the connection weights W_f and Π. In our original study, for input data we used 16 clinical characteristics that were not directly related to MetS criteria (Table 3). In our replication study, we used the two clinical characteristics that were identified as a result of the original study. All datasets were randomly arranged, divided equally into five datasets, and subjected to a fivefold cross-validation (CV) by using four datasets as training data and one dataset for validation. Through this fivefold CV, the combination of two input parameters that provided the best predictive accuracy as an average throughout the CV was selected by means of the forward-selection method. The accuracy was calculated as shown below, and the model with the highest accuracy was selected as the best combination.

Accuracy (%) = {(the number of correct estimation for training data) / (the number of training data)} × 1/3 + {(the number of correct estimation for test data) / (the number of test data)} × 2/3

Here, we judged that a correct estimation was achieved if the output signal y^* from the model was more than 0 for a case subject and less than 0 for a healthy control subject; otherwise, the estimation was judged to be incorrect. We compared the accuracy in FNN analysis with those in multiple linear regression and multiple logistic regression.

3.2.5. Statistical analysis

To ensure that the risk combination selected by FNN analysis was statistically

reliable, we performed logistic regression analysis involving a selected characteristic as an independent variable and the definition of “case subject” or “healthy control subject” as a dependent variable. Study estimates were adjusted for age, drinking habit, smoking habit and the components of MetS including BMI, systolic blood pressure, diastolic blood pressure, triglyceride, HDL-cholesterol and fasting plasma glucose. In replication study, we added exercise habit for adjustment. In addition, using correlation coefficient and correlation ratio, we tested the association between the selected characteristic and other clinical characteristics. A characteristic was considered statistically significant at a P value of less than 0.05. All statistical analyses were performed with R software (Version 2.13.1, http://www.r-project.org/).

3.3. Results

3.3.1. Clinical characteristics

The clinical characteristics before study start in the original study and in the replication study are listed in Tables 1 and 2, respectively. In both studies, the weight, BMI, systolic blood pressure, diastolic blood pressure, serum total cholesterol, serum triglyceride, and fasting plasma glucose were significantly higher in the case subjects than in the healthy control subjects, whereas serum HDL−cholesterol was significantly lower in the case subjects. In the replication study, the case subjects were significantly taller than the healthy control subjects.

3.3.2. FNN analysis (original study)

By means of the FNN analysis of health check-up data before study start from the original study (Table 4), we identified a combination of the γ-glutamyltranspeptidase (γ-GTP) level and the white blood cell (WBC) count as being indicative of MetS. The FNN analysis had a high accuracy of 77.4% compared with the baseline of 63.7% calculated by the null method that estimates all subjects to be case subjects or healthy control subjects. This accuracy in FNN analysis was similar to an accuracy of 75.8% in multiple linear regression and an accuracy of 75.8% in multiple logistic regression. This combination of parameters, which showed the best predictive accuracy, is illustrated as a fuzzy rule in Figure 2A. In this matrix, most case subjects were classified in the high γ-GTP level (≥26.9 IU/L) and high WBC count (≥5.83 $\times 10^3$ cells/μL) group, whereas most healthy control subjects were classified in the low

γ-GTP level (< 26.9 IU/L) and low WBC count (< 5.83 $\times 10^3$ cells/μL) group. The numbers of case subjects and healthy control subjects are shown in the upper line of each cell of the matrix and the weights required to yield a case of MetS are shown in the lower line of each cell of the matrix. The matrix for the high γ-GTP level and high WBC count showed a high weight of 1.07, which corresponds to a significant factor for a case of MetS. This trend was also shown by a scatter plot of γ-GTP level versus WBC count (Figure 3A).

3.3.3. FNN analysis (replication study)

The replication study confirmed the association between MetS development and a combination of a high γ-GTP level and a high WBC count. The FNN analysis showed a high accuracy of 67.1% compared with a baseline of 50.0%. This accuracy was similar to an accuracy of 66.1% in multiple linear regression and an accuracy of 68.5% in multiple logistic regression (Table 4). In the fuzzy rule, most case subjects were classified as showing a combination of a high γ-GTP level (≥30.2 IU/L) and a high WBC count (≥6.64 $\times 10^3$ cells/μL), which corresponded to a high weighting of 1.08 (Figure 2B). This fuzzy rule was also visualized as a scatter plot (Figure 3B).

3.3.4. Statistical verification of the γ-GTP level and WBC count as an indicator of MetS

Table 5 shows the differences in the γ-GTP level and the WBC count between the case subjects and the healthy control subjects. The difference between γ-GTP level and MetS was significant after adjusting for age, drinking habit, smoking habit and the components of MetS (original study: $P = 0.014$, replication study: $P = 1.71 \times 10^{-5}$,

combined study: $P = 3.11\times10^{-6}$). This significant difference remained significant after adjusting for exercising habit in replication study ($P = 1.69\times10^{-5}$). Although the difference between WBC and MetS was weak after adjusting for age, drinking habit, smoking habit and the components of MetS in replication study, the association remained significant in combined study (original study: $P = 0.002$, replication study: $P = 0.107$, combined study: $P = 0.031$). These results showed that the γ-GTP level and the WBC count, as selected by FNN analysis, together form a statistically reliable indicator.

To explain the difference between the γ-GTP levels in the original study (mean in healthy control subjects = 16.3 IU/L) and those in the replication study (mean in healthy control subjects = 27.3 IU/L), we compared the clinical characteristics of the participants in the two studies. We found that age had a significant effect ($P < 1.0 \times 10^{-99}$), so we investigated the correlation coefficient between age and γ-GTP levels. A scatter plot of age versus γ-GTP levels showed that the difference in the mean γ-GTP level was due to age (Figure 4).

Finally, we calculated correlation ratio for the original study (Table 6). The γ-GTP level was significantly associated with habitual drinking of alcohol ($P = 1.41 \times 10^{-2}$), but not with habitual smoking ($P = 0.406$). On the other hand, the WBC count was significantly associated with habitual smoking ($P = 1.18 \times 10^{-5}$), but not with habitual drinking ($P = 0.695$). The same tendency was found in the replication study (Table 7).

3.4. Discussion

In our study, we used an FNN as a computational method to analyze complex characteristics. The FNN analysis is a powerful machine-learning method for detecting, with maximal accuracy, significant combinations of characteristics that are associated with a particular attribute. By using an FNN, we identified that a combination of the γ-GTP level and the WBC count is a characteristic that is associated with MetS. As shown in Table 4, the accuracy in FNN analysis was similar to the accuracy in multiple linear regression and the accuracy in multiple logistic regression. However, FNN analysis also has an ability to visualize the risk of the high γ-GTP level and high WBC count group easily using fuzzy rule as Figure 2. The FNN also lacked statistical significance, so we reexamined selected characteristics by means of statistical analysis with suitable adjustments. The statistical results confirmed that the γ-GTP level and the WBC count are significant factors in MetS, confirming that the FNN has good predictive powers and is suitable for use in practical applications.

We excluded the characteristics included in judgments of metabolic syndrome. We firstly conducted FNN analysis including the components of metabolic syndrome, selecting a combination of triglycerides and WBC. However, we thought these components directly affect the prevalence of MetS. We aimed to search latent risk factors using remaining 16 clinical and laboratory characteristics. We showed that an elevated γ-GTP level and an elevated WBC count are combinational risk factors for MetS. γ-GTP is a marker of fatty liver disease and γ-GTP levels have been found to be associated with the prevalence of MetS in previous East Asian studies [20–22]. An

increase in levels of liver enzymes may be related to excess deposition of fat in the liver. The WBC count is a marker of systemic inflammation and it has also been found to be associated with the prevalence of MetS in a previous study [23]. The WBC count is controlled by cytokines, especially interleukin-6 and interleukin-8 [24], and WBCs play a major role in inflammatory processes and in defending the body against infectious disorders. In addition, a previous study has shown that the mean WBC count increases with an increase in serum γ-GTP [25]; this implied that elevated γ-GTP levels might reflect subclinical inflammation. This result from our FNN method may show that this combination has a synergistic effect.

Our study also showed that habitual drinking is related to an elevated level of γ-GTP, in agreement with a previous study [26]. However, we also showed that there is no significant association between habitual drinking and MetS. Similarly, we showed that habitual smoking is linked to an increase in the WBC count. However, the association between habitual smoking and MetS was low significant. This tendency was the same that found in a previous study [27]. Although γ-GTP levels and WBC counts are generally included in blood tests performed during periodic health examinations, these characteristics have seldom been considered in risk assessment. Our result could be useful in a personalized risk-prevention method, advising people with elevated γ-GTP level and an elevated WBC count to improve their diet and physical activity.

In this study, completely healthy people were used as controls. We also conducted logistic regression analysis between case subjects who developed MetS during the follow-up and normal control subjects who weren't MetS during the follow-up. In original study, including 77 case subjects and 597 normal control subjects, the

significant association between γ-GTP and MetS wasn't observed (*P* value = 0.213). On the other hand, the significant association between WBC and MetS was observed (*P* value = 4.76×10^{-4}). In replication study, including 2196 case subjects and 27246 normal control subjects, the significant association between γ-GTP and MetS was observed (*P* value = 5.00×10^{-21}). The significant association between WBC and MetS was also observed (*P* value = 1.41×10^{-26}). Although our study couldn't find significant association between γ-GTP and MetS in original study partly because of small sample, we showed significant associations in replication study consisted of large subjects. This result suggests that both of the association between γ-GTP and MetS and the association between WBC and MetS may be derived from the difference between those who will develop the overt clinical picture of metabolic syndrome and those who will develop some of its components without fulfilling the criteria for its diagnosis.

Our present study has several limitations however. First, for clinical data before the follow-up, we used a modified definition of MetS involving the BMI instead of the waist circumference; however, several studies have shown that the BMI is an equally effective characteristic as the waist circumference and it has been adopted for analyses of the association between the adiponection gene and metabolic traits, including MetS [28]. Secondly, we analyzed data from male subjects exclusively. This was because only one woman showed an indication of MetS among 2061 Japanese company employees in our original study. Although the potential bias was minimized by adjusting for age, drinking habit, and smoking habit, our findings may have limited value in the case of women. Thirdly, we could not subdivide drinking habit and smoking habit quantitatively. As in a previous study [26], the strength of risk of MetS may be related to the drinking status or the smoking status.

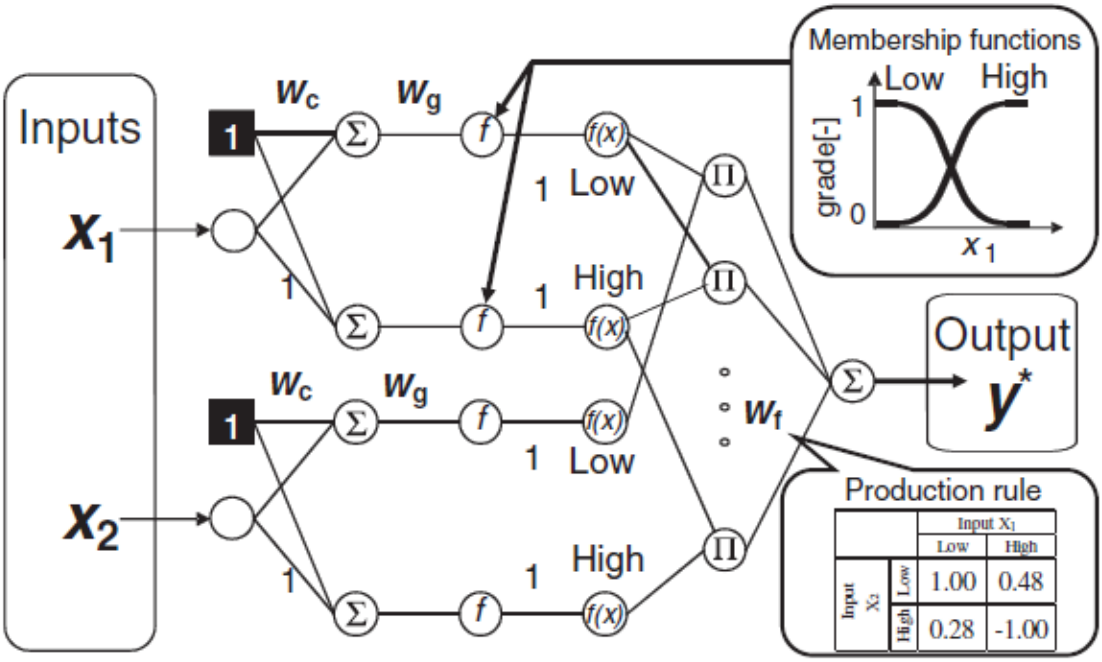

Figure 1 Fuzzy neural network (FNN) model (two inputs, one output)

The most effective combination of input characteristic contributing to MetS was identified by the use of parameter-increasing method.

A

		γ-GTP (IU/L)	
		< 26.93	≧ 26.93
WBC ($\times 10^3$cells/μL)	< 5.83	9 / 47 (-1.22)	6 / 7 (-0.41)
	≧ 5.83	12 / 22 (-0.32)	18 / 3 (1.07)

Case/Healthy control

B

		γ-GTP (IU/L)	
		< 30.24	≧ 30.24
WBC ($\times 10^3$cells/μL)	< 6.64	300 / 990 (-1.31)	731 / 509 (-0.02)
	≧ 6.64	298 / 427 (-0.68)	867 / 270 (1.08)

Case/Healthy control

Figure 2 Fuzzy rule

The fuzzy rule visualizes the risk combination identified by FNN analysis. The numbers of case subjects and healthy control subjects are shown in the upper line of each cell of the matrix, and the weights required to yield a case of MetS are shown in the lower line of each cell of the matrix. In both studies, most case subjects were classified as showing high levels of γ-GTP and high WBC counts, giving a high weight of 1.07 or 1.08, which means a significant factor for a case of MetS. (A): Original study. (B): Replication study.

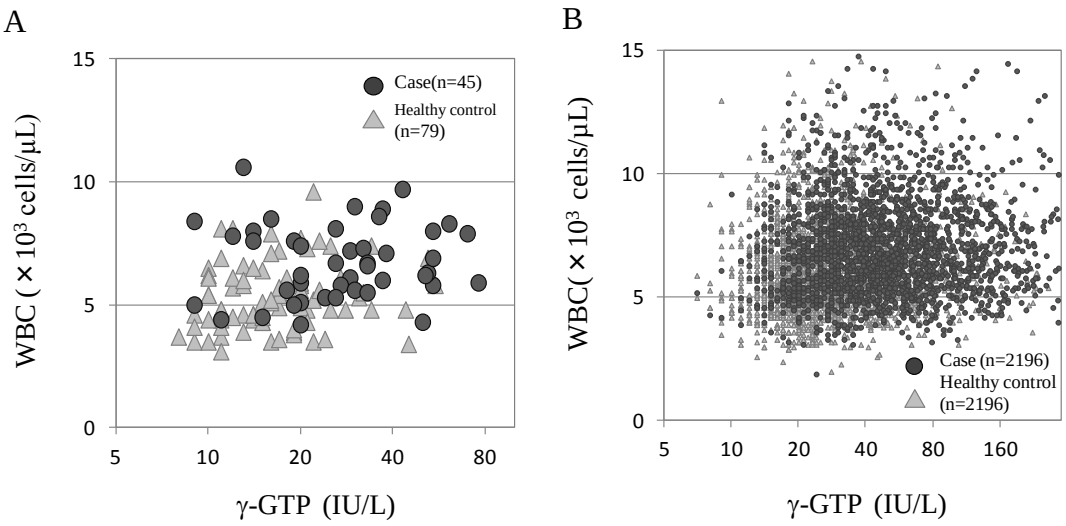

Figure 3 Scatter plots of γ-GTP level versus WBC count

Scatter plots of γ-GTP level versus WBC count show that a combination of an elevated γ-GTP level and an elevated WBC count is associated with MetS. (A): Original study. (B): Replication study.

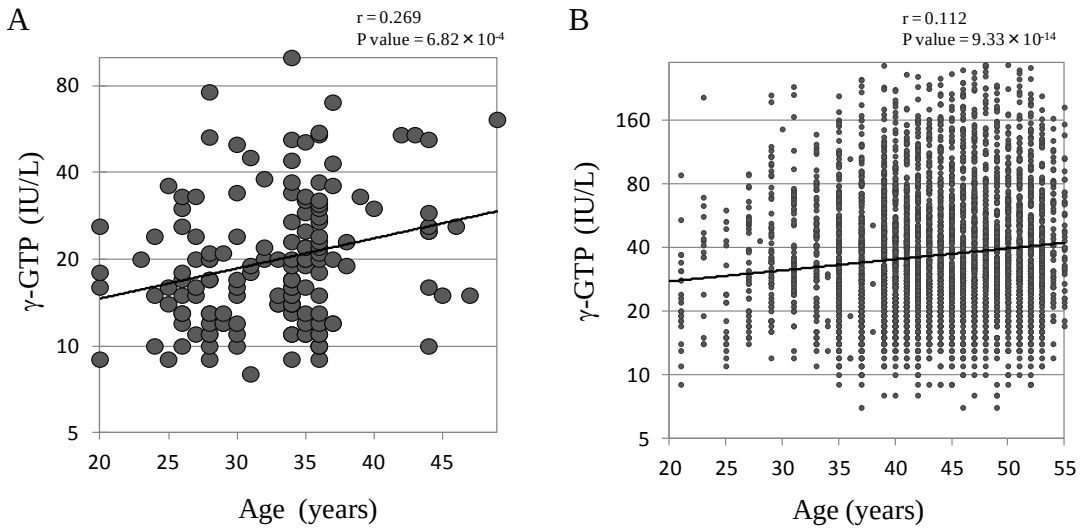

Figure 4　Scatter plots of age versus γ-GTP level

Scatter plots of age versus γ-GTP level show that the difference in the γ-GTP level between the original study and the replication study was due to the age of the subjects. (A): Original study. (B): Replication study.

Table 1 Characteristics of original study

Characteristic	Case		Healthy control		*P* value
	n	mean ±SD or n (%)	n	mean ±SD or n (%)	
Male n (%)	77	77 (100)	152	152 (100)	1.000
Age (years)	77	31.4±7.5	152	30.6±4.6	0.299
Height (cm)	77	172.1±5.4	152	170.9±5.5	0.125
Weight (kg)	77	68.2±5.8	152	59.9±6.3	8.28×10^{-19}
BMI (kg/m^2)	77	23.0±1.4	152	20.5±1.9	1.58×10^{-21}
Systolic blood pressure (mmHg)	75	130.0±12.5	152	116.9±11.2	7.45×10^{-14}
Diastolic blood pressure (mmHg)	75	78.4±9.2	152	69.8±7.0	1.92×10^{-13}
Serum total cholesterol (mg/dl)	57	187.5±28.5	99	166.0±21.5	3.37×10^{-7}
Serum triglycerides (mg/dl)	57	166.0±148.7	98	74.1±34.9	2.74×10^{-8}
Serum HDL-cholesterol (mg/dl)	54	48.0±11.0	90	57.1±9.8	9.14×10^{-7}
Fasting plasma glucose (mg/dl)	53	92.2±9.7	85	88.2±7.9	9.60×10^{-3}
Alcohol habit n (%)	76	63 (82.9)	152	131 (86.2)	0.513
Smoking habit n (%)	75	51 (68.0)	151	81 (53.6)	3.94×10^{-2}

Data are mean ± SD or n (%) unless noted otherwise.

Differences in characteristics between case and healthy control subjects were evaluated by linear regression analysis.

Table 2 Characteristics of replication study

Characteristic	Case		Healthy control		*P* value
	n	mean ±SD or n (%)	n	mean ±SD or n (%)	
Male n (%)	2196	2196 (100)	2196	2196 (100)	1.000
Age (years)	2196	43.5±7.7	2196	43.4±5.4	0.519
Height (cm)	2196	170.3±5.7	2196	169±5.8	1.58×10^{-13}
Weight (kg)	2196	72.4±9.1	2196	60.4±6.7	$< 1.0\times10^{-99}$
BMI (kg/m^2)	2196	24.9±2.7	2196	21.1±1.9	$< 1.0\times10^{-99}$
Systolic blood pressure (mmHg)	2194	125±13.7	2196	110.4±9.4	$< 1.0\times10^{-99}$
Diastolic blood pressure (mmHg)	2195	79.9±10.0	2196	69.3±7.6	$< 1.0\times10^{-99}$
Serum total cholesterol (mg/dl)	2196	204.9±34.8	2196	186.7±30.4	3.10×10^{-73}
Serum triglycerides (mg/dl)	2196	161.4±115.6	2196	76.1±28.0	$< 1.0\times10^{-99}$
Serum HDL-cholesterol (mg/dl)	2196	55.0±13.3	2196	67.6±14.8	$< 1.0\times10^{-99}$
Fasting plasma glucose (mg/dl)	2196	99.7±19.9	2196	90.1±7.4	1.28×10^{-95}
Alcohol habit n (%)	2194	1693 (77.2)	2195	1704 (77.6)	0.712
Smoking habit n (%)	2195	1427 (65.0)	2195	1195 (54.4)	8.17×10^{-13}

Data are mean ± SD or n (%) unless noted otherwise.

Differences in characteristics between case and healthy control subjects were evaluated by linear regression analysis.

Table 3 Sixteen input characteristics for the FNN analysis

Input number			
1	Smoking habit	**9**	Hematocrit (%)
2	Blood urea nitrogen (mg/dl)	**10**	RBC (million cells/μL)
3	Creatinine (mg/dl)	**11**	WBC (cells/μL)
4	Uric Acid (mg/dl)	**12**	Urine urobilinogen (%)
5	γ-GTP (IU/L)	**13**	Urine protein (%)
6	Hemoglobin (g/dl)	**14**	Urine sugar (%)
7	GOT (IU/L)	**15**	Urinary occult blood (%)
8	GPT (IU/L)	**16**	Alcohol habit

Table 4 Inputs selected by FNN

Study	Input number	Baseline(%)	Accuracy (%)			number of subject		Input characteristic	
			FNN	multiple linear regression	multiple logistic regression	Case	Healthy control	1 input	2 inputs
Original Study	1 input	63.46	73.88	69.23	73.72	57	99	γ-GTP (IU/L)	-
	2 inputs	63.71	77.43	75.81	75.81	45	79	γ-GTP (IU/L)	WBC (cells/μL)
Replication Study	1 input	50.00	64.96	64.18	66.44	2196	2196	γ-GTP (IU/L)	-
	2 inputs	50.00	67.14	66.10	68.53	2196	2196	γ-GTP (IU/L)	WBC (cells/μL)

Table 5 Statistical analysis of characteristics selected by FNN

Study	Characteristic	model 1		model 2[b]		model 3[c]		model 4[d]	
		OR (95%CI)	*P* value	OR (95%CI)	*P* value	OR (95%CI)	*P* value	OR (95%CI)	*P* value
Original Study	γ-GTP (doubling)[a]	4.71 (2.63-8.41)	1.69×10^{-7}	5.98 (3.12-11.5)	7.25×10^{-8}	4.06 (1.33-12.4)	0.014	-	
	WBC (1000 cells/μL)	1.83 (1.39-2.41)	1.62×10^{-5}	1.94 (1.41-2.65)	3.86×10^{-5}	2.69 (1.44-5.02)	0.002		
Replication Study	γ-GTP (doubling)[a]	2.64 (2.45-2.85)	$< 1.0\times10^{-99}$	2.84 (2.62-3.08)	$< 1.0\times10^{-99}$	1.32 (1.17-1.51)	1.71×10^{-5}	1.33 (1.17-1.51)	1.69×10^{-5}
	WBC (1000 cells/μL)	1.31 (1.26-1.35)	2.73×10^{-51}	1.30 (1.25-1.35)	5.06×10^{-42}	1.05 (0.99-1.12)	0.107	1.06 (0.99-1.12)	0.094
Combined Study[e]	γ-GTP (doubling)[a]	2.65 (2.46-2.86)	$< 1.0\times10^{-99}$	2.86 (2.64-3.10)	$< 1.0\times10^{-99}$	1.35 (1.19-1.53)	3.11×10^{-6}	-	
	WBC (1000 cells/μL)	1.32 (1.27-1.36)	1.28×10^{-55}	1.31 (1.26-1.36)	1.62×10^{-45}	1.07 (1.01-1.14)	0.031		

Odds ratio (OR) with 95% confidence interval (CI) indicates the proportional change in risk associated with each increase by the amount indicated in parentheses.

[a] geometric mean (2.5-97.5%) was used because of skewed distriburions.

Differences in characteristics between case and healthy control subjects were evaluated by logistic regression analysis.

[b] Differences were evaluated by logistic regression analysis with adjustment for age, drinking habit and smoking habit.

[c] Differences were evaluated by logistic regression analysis with adjustment for age, drinking habit, smoking habit and the components of MetS.

[d] Differences were evaluated by logistic regression analysis with adjustment for age, drinking habit, smoking habit, the components of MetS and exercise habit.

[e] In combined study, the difference of study was added in adjustment.

Table 6 Correlation ratios in original study

Characteristic 1	Characteristic 2	n	η	*P* vlaue
γ-GTP (IU/L)	Alcohol habit	155	0.197	1.41×10^{-2}
γ-GTP (IU/L)	Smoking habit	153	0.068	0.406
WBC (cells/μL)	Alcohol habit	137	-0.034	0.695
WBC (cells/μL)	Smoking habit	135	0.369	1.08×10^{-5}

Table 7 Correlation ratios in replication study

Characteristic 1	Characteristic 2	n	η	*P* vlaue
γ-GTP (IU/L)	Alcohol habit	4389	0.232	$< 1.0\times10^{-99}$
γ-GTP (IU/L)	Smoking habit	4390	0.068	7.01×10^{-6}
WBC (cells/μL)	Alcohol habit	4389	-0.074	9.74×10^{-7}
WBC (cells/μL)	Smoking habit	4390	0.386	$< 1.0\times10^{-99}$

3.5. Summary

We have shown that the combination of the γ-GTP level and the WBC count is the most significant risk factor associated with MetS. By using a statistical analysis adjusted by age, drinking habit, smoking habit and the components of MetS, we confirmed that the FNN analysis method is suitable for identifying combinations of factors associated with the risk of lifestyle diseases. Our results may be useful in providing a novel personalized diagnostic and therapeutic method, depending on the individual subject's γ-GTP level and WBC count.

3.6. References

1. Robert, H. E., Scott, M. G., and Paul, Z. Z.: **The metabolic syndrome.** *Lancet*, **365,** 1415–1428 (2005).
2. McCurry, J.: **Japanese people warned to curb unhealthy lifestyles: Health experts urge a return to dietary basics to prevent future health problems.** *Lancet*, **363,** 1126 (2004).
3. Panagiotakos, D. B., Pitsavos, C., Skoumas, J., Chrysohoou, C., Toutouza, M., Stefanadis, C. I., and Toutouzas, P. K.: **Importance of LDL/HDL cholesterol ratio as a predictor for coronary heart disease events in patients with heterozygous familial hypercholesterolaemia: A 15-year follow-up (1987–2002).** *Curr Med Res Opin*, **19,** 89–94 (2003).
4. Nakatochi, M., Miyata, S., Tanimura, D., Izawa, H., Asano, H., Murase, Y., Kato, R., Ichihara, S., Naruse, K., Matsubara, T., Honda, H., and Yokota, M.: **The ratio of adiponectin to homeostasis model assessment of insulin resistance is a powerful index of each component of metabolic syndrome in an aged Japanese population: Results from the KING study.** *Diabetes Res Clin Pract*, **92,** e61–e65 (2011).
5. Ushida, Y., Kato, R., Tanimura, D., Izawa, H., Yasui, K., Takase, T., Yoshida, Y., Kawase, M., Yoshida, T., Murohara, T., and Honda, H.: **Determination of combinational genetic and environmental risk factors of lifestyle-related disease by using health check-up data obtained from long-term follow-up.** *Seibutsu Kogaku Kaishi*, **88,** 562–569 (2010).
6. Hirose, H., Takayama, T., Hozawa, S., Hibi, T., and Saito, I.: **Prediction of**

metabolic syndrome using artificial neural network system based on clinical data including insulin resistance index and serum adiponectin. *Comput BiolMed*, **41,** 1051-1056 (2011).

7. Honda, H. and Kobayashi, T.: **Fuzzy control of bioprocess.** *J Biosci Bioeng*, **89,** 401–408 (2000).

8. Hibino, S., Hanai, T., Nagata, E., Matsubara, M., Fukagawa, K., Shirataki, T., Honda, H., and Kobayashi, T.: **Fuzzy neural network model for assessment of Alzheimer-type dementia.** *J Chem Eng Jpn*, **34,** 936–942 (2001).

9. Ando, T., Suguro, M., Hanai, T., Kobayashi, T., Honda, H., and Seto M: **Fuzzy neural network applied to gene expression profiling for prognosis of diffuse large B-cell lymphoma.** *Jpn J Cancer Res*, **93,** 1207–1212 (2002).

10. Ando, T., Suguro, M., Kobayashi, T., Seto, M., and Honda, H.: **Multiple fuzzy neural network system for outcome prediction and classification of 220 lymphoma patients on the basis of molecular profiling.** *Cancer Sci*, **94**, 906–913 (2003).

11. Hanai, T., Katayama, A., Honda, H., and Kobayashi, T.: **Automatic fuzzy modeling for *Ginjo sake* brewing process using fuzzy neural networks**. *J Chem Eng Jpn*, **30,** 94–100 (1997).

12. Tomida, S., Hanai, T., Honda, H., and Kobayashi, T.: **Construction of COD simulation model for activated sludge process by recursive fuzzy neural network**. *J Chem Eng Japan*, **34,** 369–375 (2001).

13. Tominaga, O., Ito, F., Hanai, T., Honda, H., and Kobayashi, T.: **Sensory modeling of coffee with a fuzzy neural network.** *J Food Sci*, **67,** 363–368 (2002).

14. Kato, R., Nakano, H., Konishi, H., Kato, K., Koga, Y., Yamane, T., Kobayashi, T., and Honda, H.: **Novel strategy for protein exploration: High-throughput screening assisted with fuzzy neural network.** *J Mol Biol*, **351,** 683–692 (2005).

15. Ushida, Y., Kato, R., Morimoto, T., and Honda, H.: **Detection of physical environmental factors on comfortableness of housing.** *Trans Jpn Soc Kansei Eng*, **9,** 97–102 (2009).

16. Kaga, C., Okochi, M., Tomita, Y., Kato, R., and Honda, H.: **Computationally assisted screening and design of cell-interactive peptides by a cell-based assay using peptide arrays and a fuzzy neural network algorithm.** *BioTechniques*, **44,** 393–402 (2008).

17. Takeshita, T., Okochi, M., Kato, R., Kaga, C., Tomita, Y., Nagaoka, S., and Honda, H.: **Screening of peptides with a high affinity to bile acids using peptide arrays and a computational analysis.** *J Biosci Bioeng*, **112,** 92–97 (2011).

18. The Examination Committee of the Criteria for Metabolic Syndrome in Japan.: **Definition and criteria of the metabolic syndrome in Japan.** *Japan Intern Med*, **94,** 188–201 (2005).

19. Horikawa, S., Furuhashi, T., and Uchikawa, Y.: **On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm.** *IEEE Trans Neural Netw*, **3,** 801–806 (1992).

20. Oda, E., Kawai, R., Watanabe, K., and Sukumaran, V.: **Prevalence of metabolic syndrome increases with the increase in blood levels of gamma-glutamyltransferase and alanine aminotransferase in Japanese**

men and women. *Intern Med*, **48,** 1343–1350 (2009).

21. Ryu, S., Chang, Y., Woo, H. Y., Yoo, S. H., Choi, N. K., Lee, W. Y., Kim, I., and Song, J.: **Longitudinal increase in gamma-glutamyltransferase within the reference interval predicts metabolic syndrome in middle-aged Korean men.** *Metabolism*, **59,** 683–689 (2010).

22. Xu, Y., Bi, Y. F., Xu, M., Huang, Y., Lu, W. Y., Gu, Y. F., Ning, G., and Li, X. Y.: **Cross-sectional and longitudinal association of serum alanine aminotransaminase and gamma-glutamyltransferase with metabolic syndrome in middle-aged and elderly Chinese people.** *J Diabetes*, **3,** 38–47 (2011).

23. Lao, X. Q., Neil, Thomas, G., Jiang, C., Zhang, W., Adab, P., Lam, T. H., and Cheng, K. K.: **White blood cell count and the metabolic syndrome in older Chinese: The Guangzhou Biobank cohort study.** *Atherosclerosis*, **201,** 418–424 (2008).

24. van Oostrom, A. J., Sijmonsma, T. P., Verseyden, C., Jansen, E. H., de Koning, E. J., Rabelink, T. J., and Castro, Cabezas, M.: **Postprandial recruitment of neutrophils may contribute to endothelial dysfunction.** *J Lipid Res*, **44,** 576–583 (2003).

25. Nakanishi, N., Suzuki, K., and Tatara, K.: **Serum γ-glutamyltransferase and risk of metabolic syndrome and type 2 diabetes in middle-aged Japanese men.** *Diabetes Care*, **27,** 1427–1432 (2004).

26. Haren, M. T., Li, M., Petkov, J., and McDermott, R. A.: **Alcohol, metabolic risk and elevated serum gamma-glutamyl transferase (GGT) in indigenous Australians.** *BMC Public Health*, **10,** 454 (2010).

27. Ishizaka, N., Ishizaka, Y., Toda, E., Nagai, R., and Yamakado, M.: **Association between cigarette smoking, white blood cell count, and metabolic syndrome as defined by the Japanese criteria.** *Intern Med*, **46,** 1167–1170 (2007).

28. Tanimura, D., Shibata, R., Izawa, H., Hirashiki, A., Asano, H., Murase, Y., Miyata, S., Nakatochi, M., Ouchi, N., Ichihara, S., Yasui, K., Yoshida, T., Naruse, K., Matsubara, T., and Yokota, M.: **Relation of a common variant of the adiponectin gene to serum adiponectin concentration and metabolic traits in an aged Japanese population.** *Eur J Hum Genet*, **19,** 262–269 (2011).

Chapter 4

Concluding remarks and Future works

The prevention of lifestyle-related diseases is the most attractive theme these days. In fact, several instructions such as dieting or exercising are conduced. However, metabolic syndrome or lifestyle-related diseases can be caused by complex interactive associations among uncertain factors such as genetic or environmental factor. As a result, it might be impossible to instruct to improve all risk factors. Therefore, exhaustive analysis may be needed to a personalized novel diagnostic and therapeutic method. In this thesis, exhaustive combinational analyses were developed to contribute to their solutions.

In the Chapter 1, a general introduction covering the importance of finding combinational risk factors were discussed. In fact, to find combinational risk factors may be lead to motivation of lifestyle diseases improvement. And considering to these backgrounds, the objective and the strategy of this thesis were described.

In the Chapter 2, exhaustive combinational analyses of genetic and environmental

factors were performed, and the importance of exhaustive combinational analyses was shown. This study had 1458 subjects who had long-term follow-up (at least 7 years) and they were stable data because of company employees. Exhaustive method of logistic regression analysis revealed 25 risk factor combinations. Exhaustive method of Criterion of Detecting Personal Group (CDPG) revealed that the combination of *ADIPOR1* (rs1539355) with an environment factor (smoking) was the most significant predictor of metabolic syndrome. Moreover, the investigation of the change of BMI during follow-up showed that BMI obviously increased in the group which had both *ADIPOR1* (rs1539355) and smoking risk factors. From our study finding, the advice that the person who had genetic risk factor of *ADIPOR1* (rs1539355) AA genotype instructs to refrain especially smoking from several environmental factors will be effective.

In the Chapter 3, combinational clinical characteristics were searched by established methods of Fuzzy neural network (FNN) and logistic regression analysis. In this study, two study's clinical characteristics were analyzed. Our original study included 77 case subjects who developed metabolic syndrome during the follow-up and 152 healthy control subjects who were free of lifestyle-related risk components from among 1803 Japanese male employees. Our replication study included 2196 case subjects and 2196 healthy control subjects. Both studies identified that the combination of the γ-GTP level and the white blood cells (WBC) count is the most significant risk factor associated with metabolic syndrome. This combination was selected by FNN and confirmed by logistic regression analysis adjusting age, drinking habit, smoking habit and the components of metabolic syndrome. Our study also showed that habitual

drinking is related to an elevated level of γ-GTP and that habitual smoking is linked to an increase in the WBC count. Our result could be useful in a personalized risk-prevention method, advising people with elevated γ-GTP level and an elevated WBC count to improve their diet and physical activity.

Thus, the concept of combinational risks using genetic and environmental factors or clinical characteristics is confirmed by exhaustive combinational methods. And the finding of combinational risk factors would be the aide to instruct a personalized novel diagnostic and therapeutic method.

In the future work, we have planning exhaustive combinational analyses of not only genetic factors, environmental factors and clinical characteristics but also daily meal data. We also have planning to analysis other lifestyle-related diseases such as chronic kidney disease. Lifestyle-related diseases cause complex interaction, so the analysis of daily meal data may contribute to find novel combinational risk factors leading to personalized novel diagnostic and therapeutic methods.

In fact, the quantity of clinical check-up data, SNP data and daily meal data is more and more increasing. I hope that exhaustive combinational analysis using statistic, machine learning and constructing database helps to contribute to the improvement of life-style related diseases.

List of publications

List of publications for dissertation

(1) **Yasunori Ushida**, Ryuji Kato, Tomoko Morimoto, Hiroyuki Honda: Detection of physical environmental factors on comfortableness of housing. *Transactions of Japan Society of Kansei Engineering*, **9,** 97–102 (2009).

(2) **Yasunori Ushida**, Ryuji Kato, Daisuke Tanimura, Hideo Izawa, Kenji Yasui, Tomokazu Takase, Yasuko Yoshida, Mitsuo Kawase, Tsutomu Yoshida, Toyoaki Murohara and Hiroyuki Honda: Determination of combinational genetic and environmental risk factors of lifestyle-related disease by using health check-up data obtained from long-term follow-up. *Seibutsu Kogaku Kaishi*, **88,** 562–569 (2010).

(3) **Yasunori Ushida**, Ryuji Kato, Kosuke Niwa, Daisuke Tanimura, Hideo Izawa, Kenji Yasui, Tomokazu Takase, Yasuko Yoshida, Mitsuo Kawase, Tsutomu Yoshida, Toyoaki Murohara and Hiroyuki Honda: Combinational risk factors of metabolic syndrome identified by fuzzy neural network analysis of health-check data. *BMC Medical Informatics and Decision Making*, **12,** 80 (2012).

Conference

Domestic 7 times

(The domestic conferences related this thesis are described below)

(1) **Yasunori Ushida**, Yasuyuki Tomita, Masahiro Nakatochi, Ryuji Kato, Mitsuhiro Yokota and Hiroyuki Honda: Searching of gene polymorphism risk-factor in lifestyle-related disease by using stratification method of environment factor. *The Society of Chemical Engineers 73th Annual Meeting*, Hamamatsu, Japan, March, 2008

(2) **Yasunori Ushida**, Hideo Izawa, Daisuke Tanimura, Toyoaki Murohara, Fumie Nagayoshi, Ryuji Kato and Hiroyuki Honda: Search of combinational genetic and environmental risk factors in Metabolic Syndrome by using gene polymorphism. *The Society of Chemical Engineers 74th Annual Meeting*, Yokohama, Japan, March, 2009

(3) **Yasunori Ushida**, Ryuji Kato and Hiroyuki Honda: Importance of gene polymorphism in lifestyle-related disease risk-factor research. *The Society of Chemical Engineers 43th Autumn Meeting*, Nagoya, Japan, September, 2011

Acknowledgments

This study has been carried out in Bioprocess Engineering laboratory, Department of Biotechnology, School of Engineering, Nagoya University, Japan.

First of all, the author expresses his sincerest gratitude to his supervisor, Professor Hiroyuki Honda, Department of Biotechnology, School of Engineering, Nagoya University, for his supreme leading, support and assistance for this research. The author has been given a multiplicity of opportunities and encouragements to grow as a researcher by Professor Honda.

The author expresses deep gratefulness to Professor Shinji Iijima and Associate Professor Mina Okochi, Department of Biotechnology, School of Engineering, Nagoya University, Associate Professor Ryuji Kato, Department of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, and Professor Masakazu Jimbo, Department of Computer Science and Mathematical Informatics, Graduate School of Information Science, Nagoya University, for their thoughtful advices to this research.

The author expresses deep appreciation to Assistant Professor Kei Kanie, Department of Basic Medicinal Sciences, Graduate School of Pharmaceutical Sciences, Nagoya University, for their precise leading, thoughtful advice, and many encouragements.

The author also expresses deep appreciation to Professor Mitsuhiro Yokota, Department of Cardiovascular of Genome Science, School of Dentistry, Aichi-Gakuin University, Dr. Hideo Izawa (presently with Department of Cardiology, School of Medicine, Fujita Health University), Dr. Hiroyuki Asano, Dr. Daisuke Tanimura, Dr. Seiko Miyata (presently with Department of Biomedical Sciences, School of Life and Health Science, Chubu University), and Dr. Mayo Sukegawa, for their thoughtful advices and many encouragements. Especially, the author's genetic and medical knowledge was inspired and strongly led by Professor Yokota's advice.

The author would like to express deep appreciation to Dr. Masahiro Nakatochi, Department of Biotechnology, School of Engineering, Nagoya University, for his precise leading and a multiplicity of his thoughtful advices and encouragements. This research could never be completed without his helpful suggestions and precise thesis selection.

The author also expresses his sincere thanks to his collaborators, Dr. Tomoko Morimoto, for providing the study data and her thoughtful advices in List of publications for dissertation (1).

The author also expresses his sincere thanks to his collaborators, Dr. Toyoaki Murohara, Dr. Yoshida Tsutomu, Dr. Kenji Yasui, Dr. Yoshida Yasuko, Dr. Kawase Mitsuo, Mr. Kosuke Niwa, and Mr. Tomokazu Takase, for providing the study data, their thoughtful advices and many encouragements in Chapter 2-3.

The author appreciates deeply to the member of Bioinformatics team in Bioprocess Engineering laboratory (Dr. Yasuyuki Tomita, Dr. Masahiro Nakatochi, Mr. Takashi Kawamura, and Mr. Syun Kawai) for their earnest cooperation, helpful discussions, and precious friendships. Many thanks are due to all the members of Bioprocess Engineering laboratory for their creative discussions, warmhearted assistance, and valuable friendships. Especially, the author thanks to Dr. Hiroshi Ito, Mr. Tomoya Sugita, Mr. Hiroto Sasaki, Mr. Fumitaka Kumazawa, Mr. Hironobu Koike, Mr. Kenji Kojima, Mr. Takeshi Sugawara, Ms. Toshimi Sugiura, Mr. Toshikazu Takeshita, and Mr. Yoshihide Nagura, for sharing precious experiences and time in the laboratory together.

Finally, the author is expressing the deepest appreciation to his parents, family members, and all of people who have supported his life for devoting everything to the doctor course.

March, 2013, Yasunori Ushida