

## 臨床ゲノム研究の現状と Pharmacogenomics (PGx) のガイドラインの整備

鎌谷直之

## Current Status of Researches in Genomic Medicine and the Guidelines for Pharmacogenomics (PGx)

Naoyuki KAMATANI

*Center for Genomic Medicine (CGM), RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*

(Received September 13, 2010)

Since the whole human genome sequence has become available, the methods to search for genes of diseases or drugs responses (traits) have changed dramatically. The former approach designated as “candidate gene approach” is now dominated by “genome-wide approach”. In the former approach, researchers search for the genes based on the functions using biochemistry and molecular biology; however, in the latter approach, the genes are searched for by the genetic and statistical methods. Initially, monogenic diseases were the targets of the researches; however, polygenic diseases and drug responses have become the targets. Parametric linkage analysis was quite useful for identifying responsible genes for monogenic diseases. Genome-wide association study (GWAS) has been introduced for the identification of the genes for polygenic diseases and drug responses. GWAS was first introduced from 2002 to 2004 in Center for Genomic Medicine, RIKEN but has expanded rapidly to other countries including US, Europe and Asian countries from 2005. In Nature Genetics journal, about half of the articles published recently include the data from GWAS studies. Both qualitative and quantitative traits have been analyzed by GWAS. Qualitative traits include diseases and drug responses and quantitative traits include physical measures and clinical laboratory test values. Recent reports about the association between drug responses and genes have clarified many important pharmacogenomic associations. For these data to be analyzed efficiently and used appropriately; however, guidelines for researches and clinical tests concerning pharmacogenomics (PGx) are necessary. “Guideline for pharmacogenomic test” was issued in 2009 and, in addition, an extended guideline covering various fields is now being discussed.

**Key words**—genome-wide association study (GWAS); linkage analysis; pharmacogenomics; guideline; statistical genetics

## 1. はじめに

ヒトゲノムを始め、多くの種的全ゲノム配列が明らかになった今、医学生物学研究は大きく変わろうとしている。世代を伝わる情報の全体が判明したことにより、その設計図を基にした解析が始まった。設計図を基にデータの取得が機械化され、データのサイズが爆発的に増えた。得られた膨大なデータを基にコンピュータを用いた研究が盛んに行われるようになった。

膨大なデータ取得は、まず DNA シークエンサー

によってなされ、引き続き遺伝型解析機によって行われた。最近では大規模並列シークエンサーが次々に発売され、データのサイズは幾何学的に増えている。対象となる形質も、最初は単一遺伝子疾患の研究が急速に進み、引き続き多因子疾患の研究が行われるようになった。さらには、薬物反応性に関する遺伝子の検索も盛んに行われている。

このような大規模なデータとコンピュータ、及びデータを解析するアルゴリズムの発達により華々しい成果が得られるようになった。まず、連鎖解析技術の進歩により、多くの単一遺伝子病の原因遺伝子が解明された。さらに、ゲノムワイド関連解析 (genome-wide association study: GWAS) により多くの多因子疾患に関連する遺伝子が解明されつつある。薬物反応性に関連する遺伝子も次々に明らかに

理化学研究所ゲノム医科学研究センター (〒230-0045 神奈川県横浜市鶴見区末広町1丁目7番22号)

e-mail: kamatani@msb.biglobe.ne.jp

本総説は、日本薬学会第130年会シンポジウム S13 で発表したものを中心に記述したものである。

されている。また、大規模並列シークエンサーと高速コンピュータにより、さらに進歩の速度を増しているように見える。

## 2. 疾患と薬物反応性の遺伝的要因の発見の歴史

疾患を予防し、あるいは改善するためには疾患に関連する要因を知る必要がある。疾患の要因として遺伝と環境が考えられる。個々の疾患や個々の患者により、遺伝と環境の要因の相対的重要性は異なると考えられる (Fig. 1)。単一遺伝子型 (メンデル型) の遺伝病などでは、遺伝的要因のみで疾患が発症し、外傷などでは環境要因がほとんどである。単一遺伝子疾患の場合、要因の解析は比較的容易である。その理由は浸透率が 100% の遺伝病の場合、遺伝型 (genotype) と疾患が一对一に対応するからである。

これまでの手法は、疾患に関連がある遺伝子をまず予測し、その上でその産物の機能的変化を示すと言う手法であった。これらの手法は、「候補遺伝子アプローチ」という手法に分類できる。それに比べ、最近大きな力を発揮している手法が「ゲノムワイドアプローチ」である。後者の手法の利点は、疾患の病態から候補となる遺伝子を絞る必要がないという点である。そして、後者の手法が可能になった

のは、ヒトゲノムの全貌が明らかになったからである。

ただし、機能を手掛かりに研究を進める候補遺伝子アプローチが非常に理解し易いの 비해、統計学を駆使するゲノムワイドアプローチは、特に日本人研究者に理解し難いように思われる。その理由の 1 つは日本の多くの医学生物学研究者が生化学、細胞生物学、分子生物学の面からは十分な教育を受け、経験もあるにもかかわらず、ゲノムワイドアプローチに必要な遺伝学、統計学、情報学の教育や経験に乏しいからである。

ゲノムワイドアプローチには、単一遺伝子形質の検索に適合した連鎖解析と、多因子形質の検索に適合した GWAS がある。それぞれについて解説する。

## 3. 単一遺伝子形質とパラメトリック連鎖解析

単一遺伝子形質の原因遺伝子の検索にパラメトリック連鎖解析が威力を発揮している。遺伝継承法則は、遺伝継承に関連する出来事の詳細を正確に記載するものである。これを用いれば、家系図に含まれる、親族関係、マーカー座位の遺伝型、個人の表現型などの観察データの確率を正確に記載することが可能である。その確率を表す式 (確率関数) には未知のパラメータが 1 つ存在し、それは表現型に関係

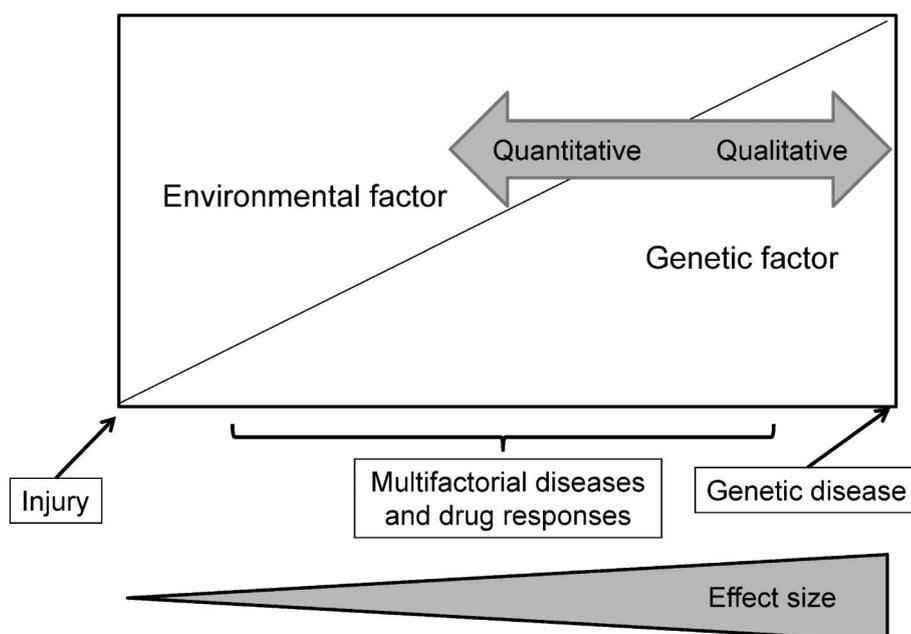


Fig. 1. Effects of Genetic and Environmental Factors on Diseases

There are both genetic and environmental factors for diseases, and the proportions of the factors vary between both diseases and individuals. An example of a disease with only genetic factors is a genetic disease, and that with only environmental factors is an injury. The strength of the effect of each gene on the disease is expressed as an effect size. Except for genetic diseases that are located at the right end, the effects of genes on diseases should be studied quantitatively rather than qualitatively.

する座位の位置である。したがって、その位置は、特定の固定されたマーカーと表現型関連座位（罹患状態座位）の間の組換え割合をパラメータとすることにより表される（アレル頻度や浸透率などは、別に与えられる）。

パラメトリック連鎖解析では、このような確率関数を最尤法により解析する。パラメトリック連鎖解析を最初に提案した研究者は Fisher である。彼は 1922 年にショウジョウバエのデータを題材にして、最尤法を用いた組換え割合の推定法を発表した。<sup>1)</sup>

連鎖解析にはパラメトリック連鎖解析のほかに、ノンパラメトリック連鎖解析がある。前者が遺伝病など、効果サイズの極めて大きな形質を対象とするのに比較し、後者は多因子疾患など、効果サイズの小さな形質を対象とする。前者では探索座位は「原因座位」と呼べるものであるが、後者では「関連座位」と呼んだ方が適当である。前者は Fig. 1 では、最も右に位置し、後者は右と左の中間に位置する。ノンパラメトリック連鎖解析では遺伝継承法則を直接使わず、間接的に用いる。ノンパラメトリック連鎖解析の代表は、罹患同胞対解析である。すなわち、同じ表現型を持った同胞の家系を多数集め解析を行う。

パラメトリック連鎖解析は単一遺伝子病の原因座位の解明に極めて大きな威力を発揮した。原因が不明であった多くの単一遺伝子病の原因が明らかになった。しかしノンパラメトリック連鎖解析は多因子疾患の関連座位の解明に期待されたほどの威力を発揮していない。

#### 4. 多因子形質とゲノムワイド関連解析 (GWAS)

単一遺伝子形質に比較し、多因子形質の要因を検索することは困難を伴う。多数の要因が関与しているため、1つの要因の関与の程度は低い (Fig. 1)。そのため、その要因があれば疾患、なければ疾患ではない、というような明確な一対一対応を期待することは不可能である。そのため、その要因を持っている方が持っていないより病気になり易い、というような量的比較が中心となる。

連鎖解析の中でもノンパラメトリック連鎖解析は多因子形質の検索のための手法である。しかし、ノンパラメトリック連鎖解析が大きな威力を発揮できない理由は、陽性領域の広さにある（約 10 cM 程

度であり、1 Mbp ほどの長さに相当）。ノンパラメトリック連鎖解析に代わって多因子疾患の関連座位の探索法として出現したのがゲノムワイド関連解析 (GWAS: genome-wide association study) である。

GWAS は症例の集団と、対照 (コントロール) の集団を別々に集め、その 2 つの集団の間でアレル頻度、あるいは遺伝型頻度が異なるかどうかを検討する方法である。おおよその手順を Fig. 2 に示す。メンデルの優劣の法則は遺伝型の下で、特定の確率で (浸透率)、特定の表現型をとることを主張する。もし、遺伝型により表現型をとる確率が異なれば、異なった表現型の集団の遺伝型の割合は異なっているはずである。したがって、GWAS もまた、メンデルの法則を間接的に用いた解析法と言える。

問題は全ゲノムをカバーするのにどれくらい密にマーカーを配置すればよいかである。一般にゲノム上には連鎖不平衡の強い領域と比較的弱い領域がある。強い領域は数 kb-数 10 kb (場合によっては 100 kb 以上) の範囲で連続しており、これを連鎖不平衡ブロック (あるいはハプロタイプブロック) という。連鎖不平衡ブロック内の SNP は比較的少数の SNP により代表され、これをタグ SNP (tagging SNP) と呼ぶ。三人種 (アフリカ系、ヨーロッパ系、東洋系) の連鎖不平衡ブロック構造とタグ SNP は HapMap プロジェクト (理化学研究所遺伝子多型研究センターが 24% のデータに貢献) で明らかにされ公表されている (<http://www.hapmap.org/>) (東洋人は日本人と漢中国人が半々)。このプロジェクトによって得られたタグ SNP が全ゲノム関連解析のための 30 万-100 万 SNP セットに用いられている (Illumina, Affymetrix)。

2002-2004 年にかけて、理研で世界最初に開始された GWAS は、<sup>2,3)</sup> 最初のうちは手法に関して強い批判を受けたものの、2007 年より、大規模タイピングシステムの商品化により世界的に広がった。<sup>4)</sup> 今や Nature Genetics 誌の論文の 45% が GWAS を用いたものとなっている (Fig. 3)。最近の GWAS では一人当たり 50 万-100 万の SNP 遺伝型を基に、数百人、数千人のデータを解析するという、かつてない規模の解析がなされる。このような大規模な解析の有効性を支えているのが前述の遺伝的データにおける確率の正確さである。遺伝以外のデータ解析

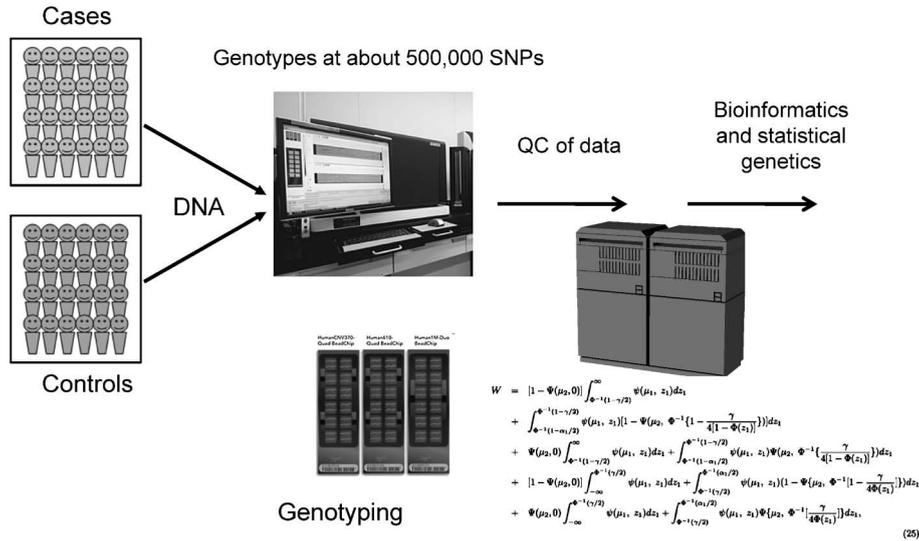


Fig. 2. Schematic Drawing of the Procedures for GWAS (Genome-wide Association Study)

DNA is collected from many cases and controls, and the genotypes at about 500000 SNPs are obtained. Subsequently, the data are analyzed by algorithms based on statistical genetics, and the SNPs significantly associated with a disease are searched for. The distinction between cases and controls does not exist for quantitative traits.

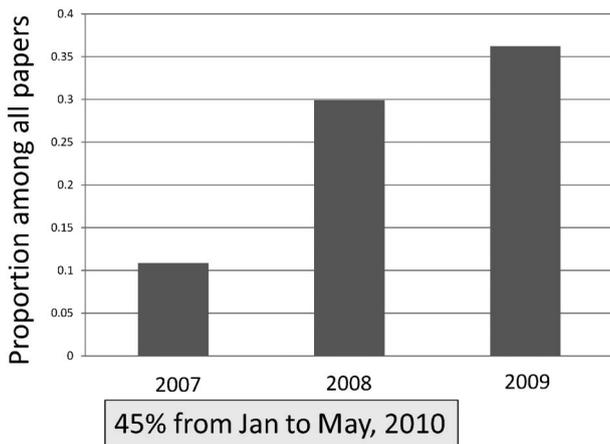


Fig. 3. Change in the Proportion of GWAS-related Papers in Nature Genetics

The number of papers using the data from GWAS is rapidly increasing.

ではこれほどの信頼できる結果は得られないであろう。

5. 質的形質と量的形質の関連

質的な対象と量的な対象は異なる。日本語を話すわれわれは、英語圏の人々に比べ、この感覚が乏しい。それは数えられる名詞と数えられない名詞の違いがないことにも起因すると考える。逆に、もともとわれわれにそのような区別をする習慣がなく、そのため対象物を質的、量的に区別する言語が発達しなかったのかもしれない。日本が純粋数学には強いのに、応用数学の分野で比較的弱いのも、このよう

な背景があるのかもしれない。

ゲノムの多様性に基づく個体の多様性の表現を「表現型」という。表現型は質的 (qualitative) 表現型と量的 (quantitative) 表現型に分類される。例えば、糖尿病のようなものは質的表現型であり、血糖値や身長は量的表現型である。

Fisher は、1918 年の重要な論文で、<sup>5)</sup> 量的形質が多数の要因の足し算で表されるモデル (相加的モデル) を提唱した。このようなモデルを相加的ポリジーンモデルと呼ぶ。そして、量的形質が特定の閾値を超えると (あるいは閾値より低くなると) 特定の表現型をとる (罹患する) というモデルを提唱することにより、量的形質と質的形質を統合することに成功した。実は、Fisher の先駆的な相加的モデルを基に、分散分析、多変数回帰、多変数ロジスティック回帰などの様々な統計学的手法が発達してきたのである。したがって、そのような統計学的モデルが、最近になって得られるようになった膨大な遺伝型データにぴったり当てはまるのはむしろ当然のことなのである。

多変数回帰モデルを用いた変量の補正、標準正規化、逆分散法を用いたメタ解析などの手法を用いて、身長などの身体所見、血液学や生化学などの臨床検査所見を対象にした量的形質の GWAS が多数発表されるようになった。<sup>6)</sup>

例えば、Fig. 4 にわれわれの研究による、量的形

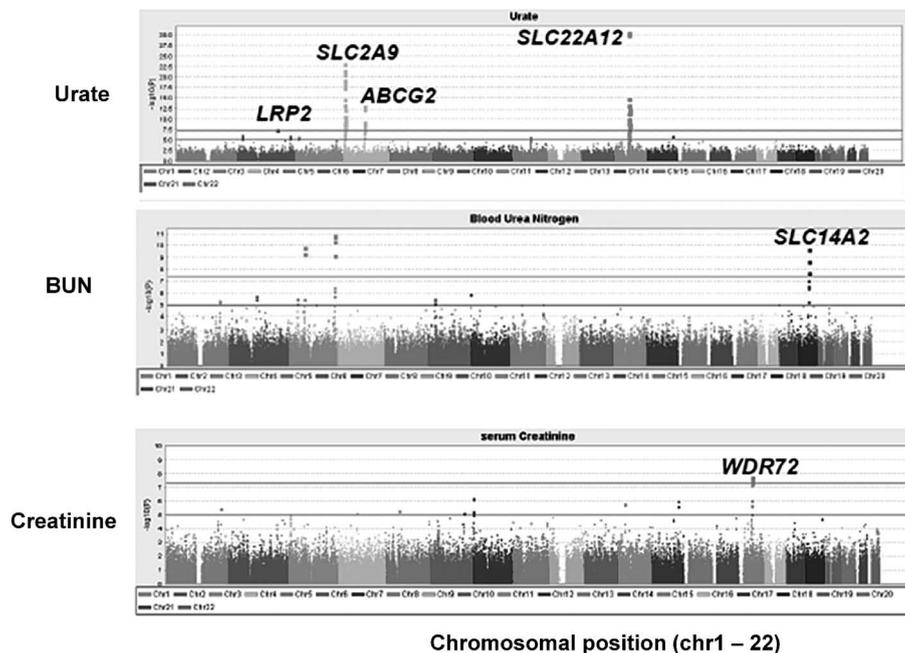


Fig. 4. Part of Manhattan Plots Indicating the Results of GWAS Targeted at Clinical Laboratory Data as Phenotypes

Clinical laboratory data are interpreted as phenotypes and GWAS was performed using the data from about 14700 Japan Biobank samples. Only the results about urate, BUN and creatinine are shown. These plots are called Manhattan plots that indicate  $-\log P$  values for the associations between SNP genotypes and phenotypes for about 500000 SNPs arranged in order from chr 1 to 22 (or X chromosome) from left to right. In the analysis, various covariates were used for the adjustment and meta-analysis was performed using the inverse variance method. The red horizontal line shows the genome-wide significance level of  $5 \times 10^{-8}$ , and the plots higher than this level are considered to be significant (From Ref. 6).

質を対象にした GWAS の結果を示す。<sup>6)</sup> われわれは最近、日本バイオバンクの 14700 人より得られた臨床検査の値を量的表現型として、多数の臨床検査値に関連する遺伝子を同定した。<sup>6)</sup> そのうち、46 個が新規の関連を示すものであった。例えば、白血球数には *MHC*, *CSF3* などが、赤血球数やヘモグロビンには *ABO* などが、血小板数には *MHC*, *THPO*, *GP1BA* などの多型が関連することがわかった。

さらに、BUN には *SLC14A2* の多様性が関連する (Fig. 4)。*SLC14A1* は赤血球膜表面の尿素受容体、*SLC14A2* は腎尿細管/睾丸の尿素受容体をコードしているので、これは極めて妥当な関連である。クレアチニンに関連するのは *WDR72* であり (Fig. 4)、これは、常染色体性劣性 Hypomaturation amelogenesis imperfecta の原因遺伝子である。また、*HbA1c* の濃度に関連する可能性が GWAS で示されている。配列からは、囊胞輸送に関連する、プロペラ機能を持つなどが示唆されている。しかし、なぜこの遺伝子が血清クレアチニンと関連するのか、全く不明である。

## 6. ゲノム薬理学に関するガイドライン

GWAS はゲノム薬理学の分野でも威力を発揮しつつある。スタチンによる重症筋障害に関係する *SLCO1B1* 遺伝子,<sup>7)</sup> ウイルス性肝炎のインターフェロン治療に関連する *IL28B* 遺伝子<sup>8)</sup> などである。また、GWAS 以外でも薬物反応性に関する遺伝子の発見が相ついでいる。

例えば、カルマバゼピン、アロプリノールなどは Stevens-Johnson 症候群の頻度が高いことで知られているが、中国漢人においては *HLA-B\*1502*,<sup>9)</sup> *HLA-B\*5801*<sup>10)</sup> アレルの関与が極めて大きいことがわかった。また、アバカビルによる重症副作用には *HLA-B\*5701*<sup>11)</sup> の関与が大きい。そのため、米国や日本でそれらの薬物の添付文書に遺伝子と副作用の関連の情報が記載されるようになった。

このようにゲノム薬理学が新薬開発や薬物使用において重要な役割を果たし始めると、それを研究や臨床に応用するためのガイドラインが必要になってきた。日本人類遺伝学会、日本臨床検査医学会、日本臨床検査標準協議会 (JCCLS) は共同で「ファーマコゲノミクス検査の運用指針」を発表した (2009 年 11 月改訂, <http://jshg.jp/news/data/pgx.pdf>).

しかし、このガイドラインは日常診療における生殖細胞系列のファーマコゲノミクス検査に限ったガイドラインである。現在、さらに広い範囲をカバーするためのガイドライン制定に向け、日本人類遺伝学会、日本臨床検査医学会、日本臨床薬理学会、日本TDM学会、日本臨床検査標準協議会（オブザーバー）で協議中である。

## 7. おわりに

分子生物学、細胞生物学はこれまで主流を占めた候補遺伝子アプローチによる形質関連遺伝子を探索するために最も重要な研究分野であった。これに比べ、最近出現したゲノムワイドアプローチは候補遺伝子を定めず、遺伝学的、統計的に関連遺伝子を探索する手法である。後者には応用数学、情報学、統計学などの知識や技術が必要であり、世界的にこの分野の研究者を多く養成することが医学生物学で重要になりつつある。その点から考え、日本は、欧米は言うに及ばず、アジアの国々の後塵を拝しているというのが現状であり、憂慮すべき状況である。

## REFERENCES

- 1) Fisher R. A., *Am. Nat.*, **56**, 406–411 (1922).
- 2) Ozaki K., Ohnishi Y., Iida A., Sekine A., Yamada R., Tsunoda T., Sato H., Sato H., Hori M., Nakamura Y., Tanaka T., *Nat. Genet.*, **32**, 650–654 (2002).
- 3) Suzuki A., Yamada R., Chang X., Tokuhiko S., Sawada T., Suzuki M., Nagasaki M., Nakayama-Hamada M., Kawaida R., Ono M., Ohtsuki M., Furukawa H., Yoshino S., Yukioka M., Tohma S., Matsubara T., Wakitani S., Teshima R., Nishioka Y., Sekine A., Iida A., Takahashi A., Tsunoda T., Nakamura Y., Yamamoto K., *Nat. Genet.*, **34**, 395–402 (2003).
- 4) The Wellcome Trust Case Control Consortium, *Nature*, **447**, 661–678 (2007).
- 5) Fisher R. A., *Trans. Roy. Soc. Edinburgh*, **52**, 399–433 (1918).
- 6) Kamatani Y., Matsuda K., Okada Y., Kubo M., Hosono N., Daigo Y., Nakamura Y., Kamatani N., *Nat. Genet.*, **42**, 210–215 (2010).
- 7) Link E., Parish S., Armitage J., Bowman L., Heath S., Matsuda F., Gut I., Lathrop M., Collins R., *N. Engl. J. Med.*, **359**, 789–799 (2008).
- 8) Tanaka Y., Nishida N., Sugiyama M., Kurosaki M., Matsuura K., Sakamoto N., Nakagawa M., Korenaga M., Hino K., Hige S., Ito Y., Mita E., Tanaka E., Mochida S., Murawaki Y., Honda M., Sakai A., Hiasa Y., Nishiguchi S., Koike A., Sakaida I., Imamura M., Ito K., Yano K., Masaki N., Sugauchi F., Izumi N., Tokunaga K., Mizokami M., *Nat. Genet.*, **41**, 1105–1109 (2009).
- 9) Chung W. H., Hung S. I., Hong H. S., Hsieh M. S., Yang L. C., Ho H. C., Wu J. Y., Chen Y. T., *Nature*, **428**, 486 (2004).
- 10) Hung S.-I., Chung W.-H., Liou L.-B., Chu C.-C., Lin M., Huang H.-P., Lin Y.-L., Lan J.-L., Yang L.-C., Hong H.-S., Chen M.-J., Lai P.-C., Wu M.-S., Chu C.-Y., Wang K.-H., Chen C.-H., Fann C. S. J., Wu J.-Y., Chen Y.-T., *Proc. Natl. Acad. Sci. USA*, **102**, 4134–4139 (2005).
- 11) Mallal S., Phillips E., Carosi G., Molina J. M., Workman C., Tomazic J., Jägel-Guedes E., Rugina S., Kozyrev O., Cid J. F., Hay P., Nolan D., Hughes S., Hughes A., Ryan S., Fitch N., Thorborn D., Benbow A.; PREDICT-1 Study Team, *N. Engl. J. Med.*, **358**, 568–579 (2008).