

総 説

1 塩基多型解析とその応用

東京女子医科大学 医学部 法医学

王 秀玲・澤口 聡子

(受理 平成 17 年 5 月 9 日)

Analysis of Single Nucleotide Polymorphisms and Its Application

Xiuling WANG and Toshiko SAWAGUCHI

Department of Legal Medicine, Tokyo Women's Medical University, School of Medicine

Single nucleotide polymorphisms (SNPs) are places along the chromosomes where the genetic code tends to vary from one person to another by just a single base. They are estimated to occur about once every 1,000 bases along the 3-billion-base human genome. SNPs are an increasingly important tool for genetic and biological research. SNPs analysis is becoming increasingly important for studies of drug resistance, evolution, and molecular epidemiology in mycobacterium tuberculosis, human immunodeficiency virus, and other organisms. Although current genomic databases contain information on several million SNPs and the numbers are growing at a very fast rate, the true value of a SNP in this context is a function of the quality of the annotations that characterizes it. The most common application of SNPs is in association studies that look for a statistically significant association between SNP alleles and phenotypes, in order to pinpoint candidate causative genes. Data derived from analysis of SNPs are being applied in many diverse fields, from medical studies of disease mechanisms and individual drug response, to population genetics for tracking migration and mixing of ancestral groups and also in forensic science for the identification of human remains and identification of individuals from bodily samples.

Key words: SNPs, genome association, genetic diagnosis, disease mechanisms, drug resistance

はじめに

SNPs (single nucleotide polymorphisms) はゲノム上に最も多く存在する単純な多型である¹⁾。個人間の SNPs (1 塩基多型) はヒトゲノムで約 300~500 塩基対 (base pair: bp) に 1 つあり、ゲノム全体で約 300 万個あるといわれている²⁾。個人間での DNA 塩基配列の 1 塩基の違いにより各個人の体質などが異なってくる。1996 年 Neil Risch らによる SNPs を用いた関連分析 (transmission disequilibrium test: TDT) 法によって多因子疾患遺伝子をマッピングすることが可能である事実が示されてから、SNPs に対する注目がにわかに高まった³⁾。

近年、SNPs の研究は全世界で非常に大きな規模で進行しつつあり、ヒトゲノムの全解読によって多数の SNPs 塩基配列が解明され、SNPs マーカーの開発とその医療・産業・社会への応用は急速な広が

りを見せている。これらの SNPs を用いた集団の遺伝学的解析のため国際的な HapMap 計画 (ゲノム全域で SNPs のハプロタイプを決定する計画) が進められ、大量の集団サンプルを収集する計画はイギリス、アメリカ、カナダなどいくつかの国で進行している。

日本では、現代の生活習慣病を中心に、30 万人計画と呼ばれる患者サンプルの収集プロジェクトが進められている⁴⁾。特に、SNPs には、直接的に個人の疾患感受性や薬剤応答性に寄与し得る可能性、また間接的に疾患感受性遺伝子を同定するツールとなりうる可能性があり、注目が集まっている。SNPs マーカーは、病気に関連する遺伝子を探索するために非常に有用な道具であり、多数の SNPs を用いた関連分析によって糖尿病や高血圧、老人性痴呆、免疫疾患、癌などの多因子疾患の要因遺伝子を同定できる

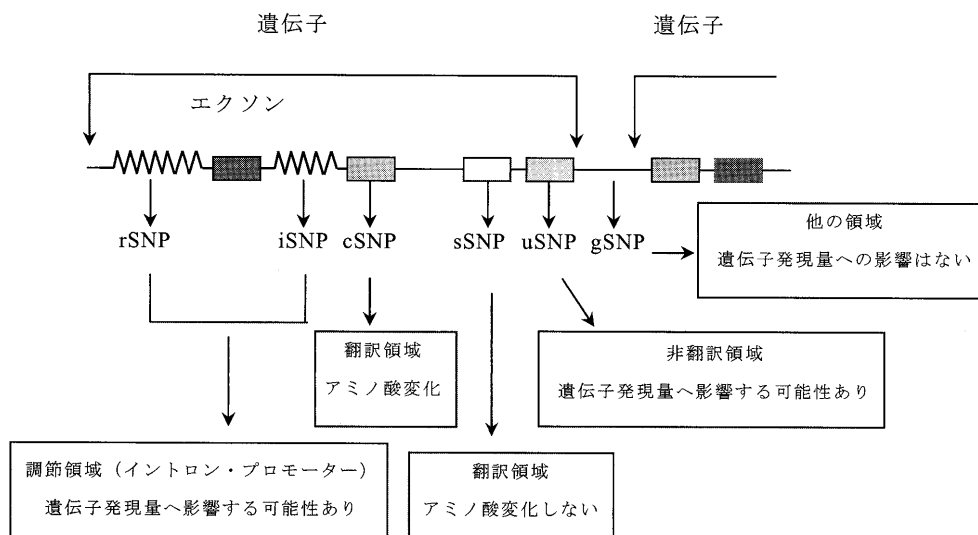


図1 SNPsの分類

可能性が示されている。

SNPs解析は、各個人に合わせて最適な医療を提供できるオーダーメイド医療（個人化医療）の実現のために、疾患のメカニズムの解明が不可欠である。また、SNPs解析に必要な鋳型DNAの鎖長は100bp前後と非常に短いため、陳旧人骨や血痕、毛髪、精液斑、唾液斑、ホルマリン固定組織などを断片化したDNAからも効率よく型判定が可能であり、法医鑑識科学において有用な検査法になることが示唆されてきた。

1. SNPsの定義と分類

SNPsは個人間でのDNA塩基配列の1塩基の違いによる多型を意味している。ヒトゲノムの配列決定の過程で、個体間にはゲノム配列の個人差が数百万箇所見出された。個人差の大半は1塩基の違いであり、その違いが集団中に1%以上の頻度で見られる塩基配列の変異が存在する場合をSNPsと呼ぶ。

例えば、アルコール感受性の個人差は、アルデヒド脱水酵素遺伝子の1塩基の相違により、アルコールが飲めるヒトの遺伝子の一部(TACTGAAGTGAAC)とアルコールが飲めないヒトの遺伝子の一部(TACTAAAGTGAAC)を比較すると、飲めるヒトのG(guanine; グアニン)の1つが、飲めないヒトではA(adenine; アデニン)に変わっている、これは1塩基多型(SNPs)の1例である。

SNPの分類は、そのゲノム上に存在する位置によって分類することができる(図1)。一般的に遺伝子の機能に関係する可能性のあるrSNP(regulatory SNP)、iSNP(intronic SNP)およびuSNP(untrans-

lated SNP)を合わせたSNPをcSNP(coding SNP)と呼び、その他の遺伝子の機能に関係ないゲノムSNPをgSNP(genome SNP)と呼ぶ⁵⁾。SNPsの研究にはcDNAをターゲットにしたcSNPs(single nucleotide polymorphisms on cDNA)とゲノムDNAをターゲットにしたSNPがある。

cSNPsは、蛋白のアミノ酸配列に変化を起こし、その蛋白の機能に変化を起こすようなSNPを見つけるため有効である。それに対してゲノムレベルでのSNPsの検索は、mRNAの塩基配列に反映されないイントロンのSNPsや遺伝子発現を調節しているエンハンサーやサイレンサー配列など、遺伝子から離れたところにあり、cDNAに相当しない場所のSNPsを見つけるのに有効である。

2. SNPs解析の意義

SNPs研究と解析の実際について図2に示す。SNPs解析に関する研究は、SNPsの同定とSNPsのタイピングという大きな2つの段階に分けられる。

SNPsの同定は、新しいSNPsを発見し、ゲノム上にその位置をマッピングする作業で、膨大な数のSNPsが同定される。ゲノム上マッピングされたSNPsは、疾患や薬剤代謝に関連する遺伝子の特定に役立ち、SNPsデータベースの情報を基に、ゲノムワイドな遺伝子同定を効率よく行うことができる。

SNPsタイピングでは、マッピングとは異なり、多数の検体に対して特定のSNPsを迅速に解析すると共に、SNPsの位置にある塩基の決定を正確に判断することが可能である。SNPsは多型マーカーとし

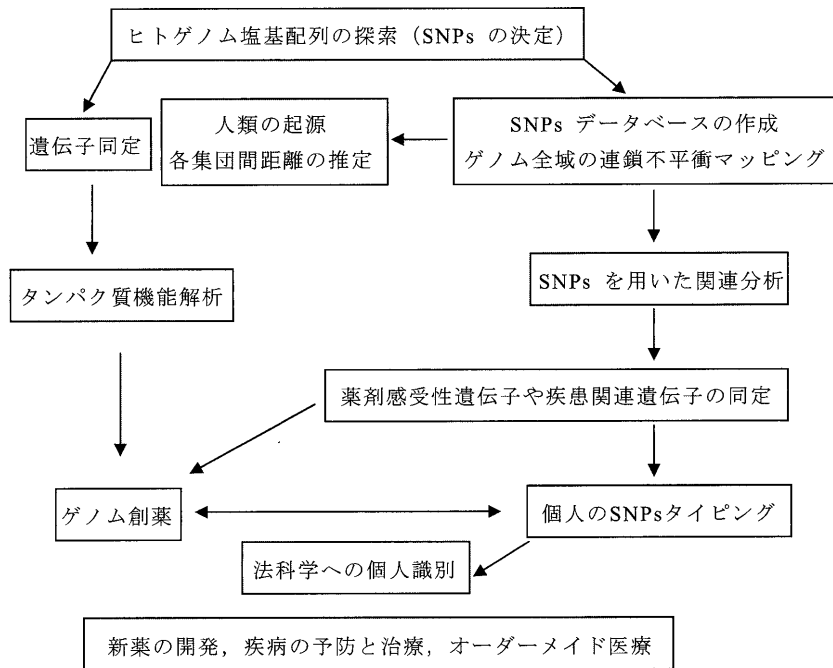


図2 SNPs 研究と解析の実際

て、病気の発症や増悪などに関連する遺伝子を見つけるために非常に有用な道具である。

SNPs は、遺伝子発現に影響を及ぼすものであり、タンパク質のアミノ酸を変化させてその働きに影響を及ぼし、遺伝子（産物）の質的異常・量的調節異常につながる場合には、SNPs を病気のリスク診断や薬剤の使い分け、診断に利用することができる。SNPs は病気に関する遺伝子を探す際の判断根拠として、病気に罹りやすさの判断や薬剤に対する応答性および副作用の違いなどの解明に役割を果たす。

人類遺伝学的な見地からは、その集団に特有な SNPs を比較することで系統樹を作成し、民族固有の SNPs を比較して民族の起源や広がり歴史などを推測することができる。

SNPs は同種個体間で最も多く観察される遺伝子多型である。1つの SNPs 座位では、2種類のアリルがあるが、数座位の SNPs マーカーを同時に用いて連鎖不平衡に基づくハプロタイプ (haprotype) を構築することにより多型性を高めることができる。この理由から、多数の SNPs 解析により、法医学鑑識科学の個人識別や血縁関係の判断などへの応用に有効であることが示されている。

3. SNPs の応用

1) 臨床医学への SNPs の応用

SNPs が医学へ積極的に応用された理由は、ヒトゲノム配列がほぼ解明され、すでに世界の共有のも

のとなったことによる。現在の疾患ゲノム研究における最大のテーマは、多因子疾患関連遺伝子である。多因子疾患発症には、遺伝要因、環境要因などが関与しているが、それぞれの疾患における遺伝要因の関与の程度は異なり、また、個人によっても異なる。臨床医学の分野では、単一遺伝子病や染色体異常の遺伝子変異 (mutation) から、多因子病の発病促進遺伝子である遺伝子変異 (variation) の時代に移行しつつある。

疾患感受性遺伝子は、連鎖不平衡に依存した高密度 SNPs マッピングに基づく多因子疾患の原因遺伝子を同定することができる (図3)。現代の生活習慣病として2型糖尿病原因遺伝子の解析は、さまざまな集団、人種で施行され、疾患感受性遺伝子座がいくつかの成績で一致していた。第1染色体長腕、第20染色体短、長腕はいくつかの民族で認められている遺伝子座であり、民族を超えて共通の疾患感受性遺伝子の存在が強く疑われている^{6)~8)}。

2型糖尿病における原因遺伝子の解析では、連鎖不平衡マッピングを用いて疾患感受性遺伝子カルパイン10 (calpain-10) が同定された^{9)~11)}。カルパイン10の同定で遺伝子的にわかったことは、まずいくつかの感受性アレルが存在すること、あるハプロタイプの組合せは発症リスクを上げるが、あるものは下がること、イントロンにある iSNP が転写レベルに関連することなどである。SNP-43 がカルパイン10遺

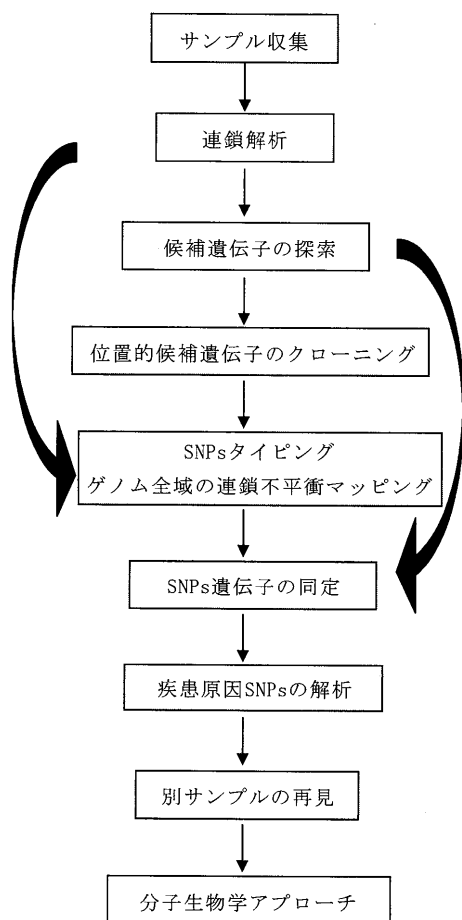


図3 疾患感受性遺伝子の同定

伝子の転写・発現に影響を与えることにより、末梢ではインスリン抵抗性を引き起こすことが判明した⁶⁾⁷⁾¹²⁾。

SNPsを用いて糖尿病や高血圧、老人性痴呆、免疫疾患、癌などの多因子疾患の要因遺伝子が同定でき、個人の遺伝子多型による薬剤感受性の差を決定することで、オーダーメイド医療を行うことも可能となる。さらに、病気の罹りやすいSNPsを検査して病気の予防に貢献する。

2) ゲノム創薬へのSNPsの応用

薬剤感受性に関連する遺伝子の解析やタンパク質機能の解析によりゲノム情報を基にした新しい薬品づくり（ゲノム創薬）が可能になり、同定された遺伝子について各個人がどのような遺伝子型を持っているかの判定とあわせて、今まで画一的に行われていた病気に対する治療法を個人にあったものへと変化させる際の判断根拠を得ることができるようになると考えられる。

例えばSNPs情報を基に創薬を行う場合、欧米人のSNPs情報から作られた薬をそのまま日本人に適

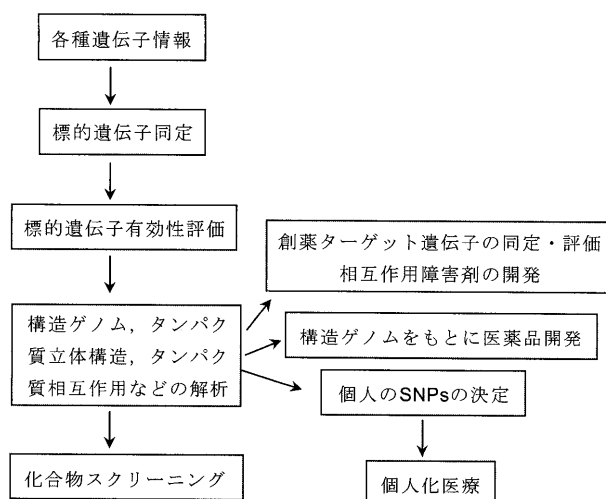


図4 ゲノム創薬の流れ

応すると、副作用の頻度や薬効が予想されたものと異なる場合が生じ得る。つまり我々にとっては、日本人特有なSNPsの遺伝子頻度の情報を知ることは意義のあることである。

ゲノム創薬では、標的遺伝子の同定（target identification）と有効性評価（target validation）をいかに網羅的、効率的に行うことが目的となる。ゲノム創薬の流れを図4に示した。標的遺伝子からの医薬品開発には、標的遺伝子の有効性が評価されれば、遺伝子をコードするタンパク質を発現・獲得し、コンビナトリアルケミストリーによるタンパク質に特異的に結合する化合物の効率的な合成とスクリーニング、タンパク質—タンパク質相互作用の解析、タンパク質の構造、SNPsに基づいた医薬品の設計、構造ゲノム（structural genomics）からのアプローチも可能になる⁹⁾。個人間のSNPs遺伝子は、薬物代謝酵素の一塩基多型により薬剤の毒性が違ってくるため、SNPsを活かして臨床試験における被験者の選定、医薬品の有効性の向上、副作用の低減などにも利用されている。

3) 遺伝学へのSNPsの応用

集団遺伝学の進展により、ヒト集団内におけるSNPsの維持機構や全ゲノムのハプロタイプ、連鎖不平衡（linkage disequilibrium）構造、進化生物学的意義などが明らかになり、世界中で精力的に多因子疾患の遺伝子解析が進行中である。

同じ染色体で近接する2つのアレルを考えると、両方の変異を有する集団内におけるハプロタイプ頻度が、単純に各アレルの積では与えられない。これは、それぞれのアレルがランダムに存在してい

るのではなく関連して存在しているからである。このように、2つ以上のアリルがランダムでなく関連している状態のことを連鎖不平衡という。ある世代で集団中の個体に SNPs 変異が生じたとすると、遺伝的浮動により SNPs は頻度を増す。経過した世代が多くなるほど、組換えにより、祖先型のハプロタイプを保持したゲノム領域は小さくなる。集団中で祖先型の保たれたゲノム領域が連鎖不平衡を示す範囲となる。

ヒトゲノムにおける連鎖不平衡の構造は、各集団によって異なることが明らかになってきた。Gabrielらの解析において、アジア集団（モンゴロイド）とヨーロッパ集団（コーカソイド）ではブロック（block）構造は平均 22kb、アフリカ大陸在住集団とアフリカンアメリカン集団（ネグロイド）では平均 11kb の大きさで、そのブロック内に存在する数種の共通（common）ハプロタイプは 3~5 種であったと報告された¹³⁾。現在観察されているブロック内のハプロタイプは、コーカソイド集団と日本人集団（モンゴロイド）でほぼ同一であり、コーカソイド集団、モンゴロイド集団で認められる共通ハプロタイプはネグロイド集団においても大部分認められる¹⁴⁾。ヒト集団に認められるハプロタイプ構造は、集団の起源と歴史を反映している。

これらの知見は、疾患感受性遺伝子同定という医学的インパクトだけでなく、ヒト集団の進化学、集団遺伝学、人類遺伝学、考古学、人口統計学にも大きなインパクトを与えることが期待されている。

4) 法医鑑識学への SNPs の応用

法医学的試料からの検査において、従来の VNTR（variable number of tandem repeat）多型や STR（short tandem repeat）多型の検査のみならず、SNPs による個人識別の遺伝マーカーの開発、検査法の確立は現在の法医鑑識科学の中で重要な位置を占めている¹⁵⁾¹⁶⁾。法医実務における鑑定検査は、通常の遺伝子検査とは異なり、検査対象となるサンプル（白骨、屍蠟化死体、腐敗組織、微量な残骸、毛髪、爪、精液斑、血痕など）の状況が千差万別である。SNPs の検査は、100bp あるいは 1ng の DNA 断片であれば、遺伝子を解析することができるが、VNTR 多型や STR 多型などを用いた超微量かつ断片化した DNA 試料から解析をすることが困難な場合は、SNPs 検査法を取り入れることが必要と示された¹⁷⁾。

親子鑑定の実例を図 5 に示した。本事例は、擬父

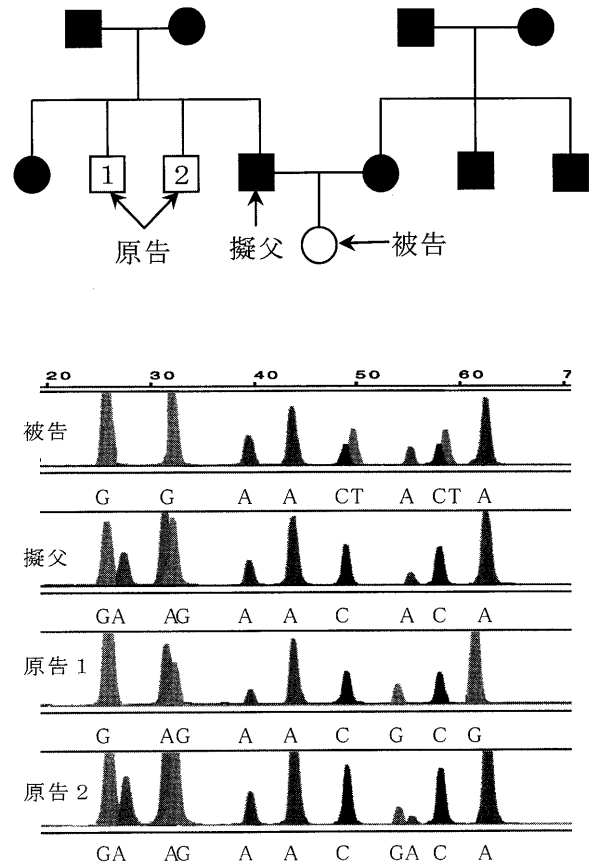


図 5 SNPs 解析によるホルマリン固定肝臓組織からの親子鑑定

上：家系図

下：被告（血液）、擬父（ホルマリン固定臓器）および原告 1, 2（血液）の 8 ローカス（ss4947490, ss4974689, ss6658727, ss5013903, ss4974676, ss4974729, ss4019224, ss4974915）の SNPs 解析の結果。8 ローカスのプライマーを SNPs（塩基置換）部位の直前に 3' 末端がくるようにデザインして 5' 側に poly dT を設定した。マルチプレックス 1 塩基のプライマー伸長反応は、SnaPshot Multiplex Kit を用いてラベリング用の PCR 増幅を行い、蛍光ラベルされた PCR 産物を ABI PRISM 310 Genetic Analyzer によりキャピラリー電気泳動で分離した。分離された電気泳動のパターンは GeneScan Software を用いて解析し、蛍光の色と移動度の違いで型判定を行った。

と母の両者が死亡している摘出し否認請求事件であり、死亡している擬父の試料が、10 年間保存されたホルマリン固定した肝臓組織であった。長期間ホルマリンで固定された組織からの DNA は断片化され、通常の VNTR 多型および STR 多型を用いて検査が得られた結果は、解釈が非常に困難であった。SNPs の解析法を用いて長期間ホルマリン固定された組織からは、擬父の遺伝子型の判定ができ、父権肯定確率は 8 ローカス（ss4947490, ss4974689,

ss6658727, ss5013903, ss4974676, ss4974729, ss4019224, ss4974915) SNPsにより0.9780であった。この値はEssen-Möllerの解釈に従うと「非常に父らしい」という範疇に入る値であり、死亡者は摘出子の父と判定して差し支えないものと考えられた。

4. SNPsの展望

21世紀はバイオテクノロジーの世紀であり、生命の持つ機能に重きがおかれる。現在、各国の研究機関においてSNPsの探索・解析が進められており、それぞれデータベース化がなされている。また、インターネット上でネグロイド、コーカソイド、モンゴロイドの3大集団のSNPsハプロタイプマップが登録され、遺伝子に関する疾患情報へのリンクなどのアノテーションもダウンロードが可能となっている。

そのような大規模な研究により多数のSNPsと薬物への反応性、副作用の関係がわかるなら、個人の遺伝子を調べ、それに基づいて薬物を選択したり、薬物濃度・量を調整したりすることも可能になると考えられる。また、個々のSNPsの集団における頻度や医学的な意義などが明らかにされれば、次のステップとして、個人の配列をタイピングしてそこからSNPsの情報を引き出し、法科学の個人識別や疾患の治療などに応用するという段階に入ると考えられる。

おわりに

最近、従来の集団遺伝学(population genetics)は、体系的ゲノムサイエンス(genomic science)と統合された概念の集団遺伝学として提唱され、疾患関連研究におけるその有用性が検討されている。SNPs関連遺伝子の解析は、遺伝学、臨床医学、薬剤開発など多方面の立場から、非常に重要である。さらには、人類の起源やさまざまな人類集団の歴史が明らかとなり、環境とゲノムの相互作用についての深い理解が得られるに違いない。今後、大量のSNPsマーカーの開発、データベースの充実、分析の簡便化・自動化などの技術の進歩とともに、精度の高い科学的遺伝子診断・治療が可能と期待される。

文 献

- 1) The International SNP Map Working Group: A

map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **15**: 928-933, 2001

- 2) Nollou P, Wagener C: Methods for detection of point mutations: performance and quality assessment. *Clin Chem* **43**: 1114-1128, 1997
- 3) Risch N, Merikanges K: The future of genetic studies of complex human diseases. *Science* **273**: 1516-1517, 1996
- 4) 榎 佳之: ヒトゲノム解読完了. *Mol Med* **41**: 2-11, 2004
- 5) 中村祐輔: 「ポストシーケンズのゲノム科学1 SNP 遺伝子多型の戦略—ゲノムの多様性と21世紀のオーダーメイド医療」, 中山書店, 東京 (2002)
- 6) Elbein SC, Chu W, Ren Q et al: Role of calpain-10 gene variants in familial type 2 diabetes in Caucasians. *J Clin Endocrinol Metab* **87**: 650-654, 2002
- 7) Baier LJ, Permana PA, Yang X et al: A calpain-10 gene polymorphism is associated with reduced muscle mRNA levels and insulin resistance. *J Clin Invest* **106**: R69-R73, 2000
- 8) Abbott A: Structures by numbers. *Nature* **408**: 130-132, 2000
- 9) Wiltshire S, Hattersley AT, Hitman GA et al: A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the diabetes UK Warren 2 repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* **69**: 553-559, 2001
- 10) Lindgren CM, Mahtani MM, Widen E et al: Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am J Hum Genet* **70**: 509-516, 2002
- 11) Mori Y, Otabe S, Dina C et al: Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* **51**: 1247-1255, 2002
- 12) Tsai HJ, Sun G, Weeks DE et al: Type 2 diabetes and three calpain-10 gene polymorphisms in Samoans: no evidence of association. *Am J Hum Genet* **69**: 1236-1244, 2001
- 13) Gabriel SB, Schaffner SF, Nguyen H et al: The structure of haplotype blocks in the human genome. *Science* **296**: 2225-2229, 2002
- 14) 坂上拓郎, 井ノ上逸朗: ヒトゲノムのハプロタイプ構造と関連解析への展望. *Mol Med* **40**: 36-42, 2003
- 15) 王 秀玲, 澤口聡子, 西澤悦子ほか: X染色体上SNPs解析. *DNA多型* **11**: 242-245, 2003
- 16) 王 秀玲, 澤口聡子, 青木杏江ほか: 日本人におけるX染色体上SNPs解析. *日法医誌* **57** (1): 77, 2003
- 17) 王 秀玲, 澤口聡子: SNPs解析とその親子鑑定への応用. *DNA多型* **12**: 7-10, 2004