

図1 マイクロアレイ/シーケンサーによるデータ産生量と遺伝学的解析手法の変遷
 国際プロジェクトの進展がこれらの技術革新を後押ししていることがみてとれる。(文献6より許可を得て
 改変のうえ転載)

の分類がある。また、疾患を形質として考えた場合、集団においてそのバリエーションのアレル(対立遺伝子)を保有している割合(アレル頻度)の違いによって、関連する疾患やバリエーションの性質も異なる(表1)。例えば、メンデル遺伝病の原因バリエーション(privateバリエーション)の場合、集団におけるアレル頻度は非常に稀であるが、一つの遺伝子のエクソン上から疾患への影響力(指標としてオッズ比などがある)が大きいバリエーションが同定されることが多い(表1)。一方、多因子疾患のバリエーション(主としてcommonバリエーション)は集団におけるアレル頻度が高い反面、関連するバリエーションは数十~数百個にもおよび個々のバリエーションが疾患におよぼす影響力は微小である(表1)。例えば、遺伝要因の影響

が強いとされている1型糖尿病ですら、現在までに40箇所以上の染色体領域が関連することが報告されている⁸⁾。従って、疾患の原因バリエーションを同定するためには、それぞれの疾患を規定するバリエーションの特徴に従った研究手法を選択することが重要である。

遺伝マーカーと解析手法

疾患の遺伝要因を探索する研究には、広大なゲノム上に点在する特徴的な配列を遺伝標識(マーカー)として活用してきた(図2)。例えば、メンデル遺伝病の原因遺伝子探索では、疾患に罹患している血縁者を多く含む大家系の検体を準備して、反復配列であるマイクロサテライトなどの遺伝マーカーと連鎖して代々継承されて

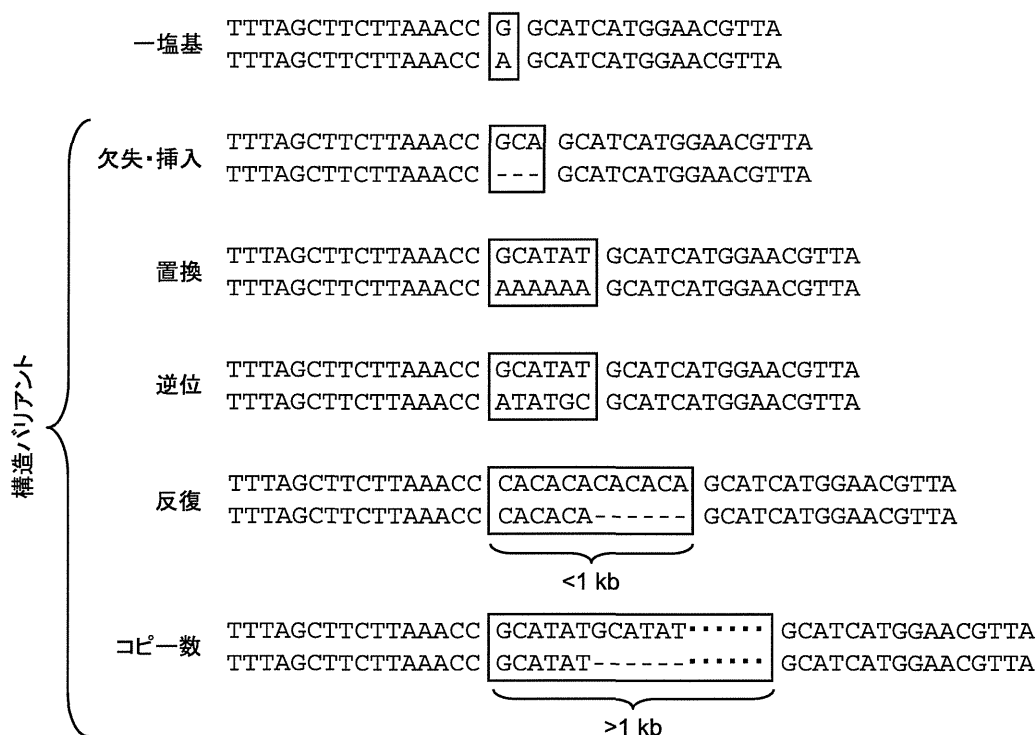


図2 バリエントの構成による分類

一塩基のみが異なる一塩基バリエントとそれ以外の構造バリエントに大別される。いずれも疾患の遺伝要因をゲノム上から探索する際に用いる遺伝マーカーとして利用されている。

表1 集団におけるバリエントの頻度による分類と疾患との関連性

バリエントの種類	集団における アレル頻度	関連疾患 (有病率)	バリエントの個数 (ゲノム上の主な位置)	個々のバリエント の疾患への影響力	影響の種類
common (ありふれた)	≥1%*	多因子疾患 (高)	数十～数百個 (遺伝子外)	小	発現調節
rare (稀な)	0.1%<頻度<1%	多因子疾患 (中)	数個～数十個 (遺伝子外、遺伝子上)	中	発現調節 アミノ酸置換
private (極端にまれな)	<0.1%	メンデル遺伝病 (低)	単一 (エキソン上)	大	アミノ酸置換

*アレル頻度が1%以上のバリエントをかつては“多型”と呼んでいた。

いるゲノム領域を絞り込んでいく連鎖解析の手法が用いられてきた⁹⁾。この方法では、特定されたゲノム領域に存在する多数の遺伝子の一つ一つを入念に解析することによって、最終的に疾患の責任遺伝子に存在するエキソン上のバリエント(表1)に辿り着くことも多い。1980年代後半からは、連鎖解析によって数々のメンデル遺伝病に関連するバリエントが同定されていった(図1)。また、かつてはある種の連鎖解

析を多因子疾患に関連するゲノム領域を探索する方法としても応用することがあった。しかし、有病率などに偏りのあることが予想される家系を用いた解析結果と一般集団における多因子疾患の罹患者を対象とした解析結果とでは必ずしも結果が一致しないことが最近明らかにされている⁹⁾。

一方、家系を用いずに一般集団から形質の異なる2群を抽出して両群で頻度に差のあるバリ

アントを検出する遺伝学的解析手法をケース・コントロール関連解析（相関解析）という。相関解析の一つに、遺伝マーカーを用いることなく従来の医学・生物学的知識から疾患の病態との関係が推測される遺伝子（候補遺伝子）を選択して解析する候補遺伝子解析がある¹⁰⁾（図1）。候補遺伝子解析では、サンガー法によるダイレクトシーケンスなどによって取得した遺伝子のエキソンの配列情報をもとに相関解析を実施し、疾患に関連するバリエントを同定する。しかし、多因子疾患を対象としたこれまでの研究では、統計学的な検出力が不足している小規模集団での解析例が多く、再現性の高い結果が得られていない。これは、解析したい遺伝子の数が多かたり、活性中心などの興味深いドメインをコードしているエキソンが長かたりすることで、コストや労力の許容量に応じて検体数を減らさなければならないことに起因している。また、多因子疾患のバリエントは疾患に対する個々の影響力が小さいため、真に関連しているバリエントを同定するためには検体数を増やす必要がある上に、そもそも多くのバリエントはエキソン上からは同定されない（表1）。従って、多因子疾患のバリエント探索には候補遺伝子解析では限界があり、遺伝子のエキソン以外のイントロンや上流に存在する転写調節配列だけでなく、遺伝子外のゲノム配列も含めた網羅的な解析を可能にする技術や遺伝マーカーの開発が以前から期待されていた。

候補遺伝子から全ゲノムへ

ヒトゲノム計画の後継の国際プロジェクトとして、2002年から国際ハップマップ計画がスタートした（図1）。本プロジェクトは、白人とアフリカ人それぞれ30組のトリオ（父、母、子）とアジア人として非血縁者の日本人と中国人それぞれ45人ずつ、総計3人種・270人分について人種ごとのゲノム全体のハプロタイプ（遺伝的に連鎖していることが統計的に算出された複数のバリエントの組み合わせ）の地図を構築するために発足した¹¹⁾。ハプロタイプ情報を得るためには、人種ごとにゲノム上の一塩基多型

（single-nucleotide polymorphism, SNP；一塩基バリエント（表1）の内、アレル頻度が1%以上のバリエントを特にSNPと呼ぶ）を高密度に決定（ジェノタイピング）し、個々のSNPのアレル頻度情報を取得する必要がある。そのためには、網羅的なジェノタイピング技術の開発が必須であり、本プロジェクトが原動力となりDNAマイクロアレイの技術が急速に進歩した（図1）。

マイクロアレイの基盤（チップ）もしくはビーズ上には、配列の中央にSNPが配置されている何十万個ものオリゴヌクレオチドプローブが固層化されており、その一つ一つと相補鎖を形成（ハイブリダイズ）するヒトゲノム由来のDNA断片を検出する。これは、1検体あたり十万レーン以上のサザンブロットング解析が3日程度で完結してしまうことに相当する画期的な技術である。莫大な資金と労力を投じたハップマップ計画では、膨大なジェノタイプデータの取得を通じて各SNPの人種別頻度情報および近傍SNPどうしのハプロタイプ情報を学術的成果として取得することができた¹²⁾。加えて、技術的成果としてアレイ上のプローブの高密度化およびハプロタイプ情報に基づく“タグSNP（領域を代表するSNP）”の選別によるプローブの質の向上が達成できた¹²⁾。その結果、アレイあたりの解析コストの劇的な低下がもたらされると共に、これらの成果は全ゲノム上にまんべんなく配置されたSNPを遺伝マーカーとして活用する相関解析、いわゆるゲノムワイド関連解析（genome-wide association study, GWAS）へと発展した（図1）。

GWAS がもたらしたもの

GWASが確立されて以来、現在までに様々な多因子疾患に関連する何千ものバリエントが同定されている¹³⁾。2007年にはサイエンス誌の“Breakthrough of the Year”にも選ばれ、当初その勢いは人類遺伝学における“gold rush”とさえ言われた¹⁴⁾。では、我々は果たして黄金の山を得ることができたのだろうか？

GWASの成功例の先駆けとして、2005年に

報告された加齢黄斑変性 (age-related macular degeneration, AMD) が挙げられる¹⁵⁾。加齢黄斑変性は、網膜中央の黄斑が加齢とともに変性する多因子疾患であり、米国では失明原因の第1位の疾患である。この報告では、約10万個のプロープが搭載されている初期のアレイ (現在は100万~250万プロープのアレイが主流; 図1) を用いて、白人のケース96例とコントロール50例という現在では桁違いに小規模な検体群を用いたGWASが実施された。その結果、ケース群に有意なマーカーSNPの同定を足掛かりに、補体系の調節因子の一つであるH因子遺伝子 (complement factor H, *CFH*) のエキソン上からアミノ酸置換を伴うSNPを発見することに成功した。その後、*CFH*とは異なる複数の補体系因子上のSNPもAMDと関連していることが次々に明らかにされた¹⁶⁾¹⁷⁾ ことから、網膜における補体系と加齢黄斑変性の発症との関連性を解明する興味深い課題が残されている。また、これらのSNPはいずれもオッズ比が高い (>2.0) ことから、複数のSNPのリスクアレルの有無の情報を組み合わせることによって、AMDの発症リスクを予測するゲノム診断への応用が検討されている¹⁸⁾。このように、多因子疾患のGWASによって遺伝子上からオッズ比の高いバリエーションが同定された場合、疾患の発症機序を解明する基礎研究が進展するだけでなく、疾患の発症を予測する新しい診断法の開発へと発展する可能性を秘めている。

しかし、GWASに寄せられた過度な期待に反して、現在までに同定されているバリエーションの大多数は個々の疾患の発症機序の解明やゲノム診断の実現に直結していない。同定されたマーカーSNPやそれと連鎖しているSNPの疾患に対する影響力が小さい (オッズ比1.2~1.5程度) 上に、そもそも多くのSNPが近傍に遺伝子の存在しない“gene desert (遺伝子砂漠)”から同定されたからである⁷⁾。例えば、冠動脈疾患¹⁹⁾と2型糖尿病²⁰⁾の両者に関連することで注目された染色体領域 (9p21.3) には *CDKN2A* や *CDKN2B* などの遺伝子が存在するものの、バリエーションが同定された領域はこれらの遺伝子から130 kb以上

離れている遺伝子砂漠である。一部のバリエーションは *CDKN2B* の下流に位置する *CDKN2B-AS1* 上から同定されたが、*CDKN2B-AS1* は非コード遺伝子 (RNAには転写されるもののタンパク質には翻訳されない遺伝子。RNAとして何らかの機能を果たしていると考えられている) であり、その機能は不明である。興味深いことに、本領域は冠動脈疾患と2型糖尿病の他にも虚血性脳卒中、大動脈瘤、神経膠腫、悪性黒色腫など複数の多因子疾患と関連することが明らかにされている²¹⁾。加えて、我々が実施した原発開放隅角緑内障 (primary open-angle glaucoma, POAG) のGWAS²²⁾²³⁾ においても、POAG患者に特有なバリエーションが9p21.3の *CDKN2B-AS1* 上から同定された。これらのバリエーションは、他施設での検証や複数施設のデータを統合したメタ解析によって確かにそれぞれの疾患に関連していることが証明されている。従って、GWASによって少なくとも多因子疾患に関連する染色体領域は絞り込めたものの、統計学的に疾患に関連付けられたこれらのバリエーションが個々の疾患の病態にどのように関わっているのか、その分子機序を解明する難題は今もなお残されているのが現状である。

遺伝子砂漠の生物学的意義

遺伝子砂漠には統計学的に疾患に関連付けられるバリエーションが存在する以上、生物学的に何らかの意味のある塩基配列が潜在し、その配列上のバリエーションの有無が疾患の病態に関与していることが推測される²⁴⁾ (図3)。現時点での生物学的知識から我々が考慮すべきゲノム配列は、遠隔遺伝子の発現を調節しているエンハンサーやタンパク質に結合してその活性を制御する各種RNAをコードする配列 (図3A) あるいはメチル化などのエピジェネティックマーク (図3B) が挙げられる。そして、これらの配列の有無とバリエーションとの関連性を検証するためには、GWASで同定された染色体領域全体にわたる高精度な塩基配列情報の取得が不可欠である。

一度に大量かつカバー率の高い (=高精度な)

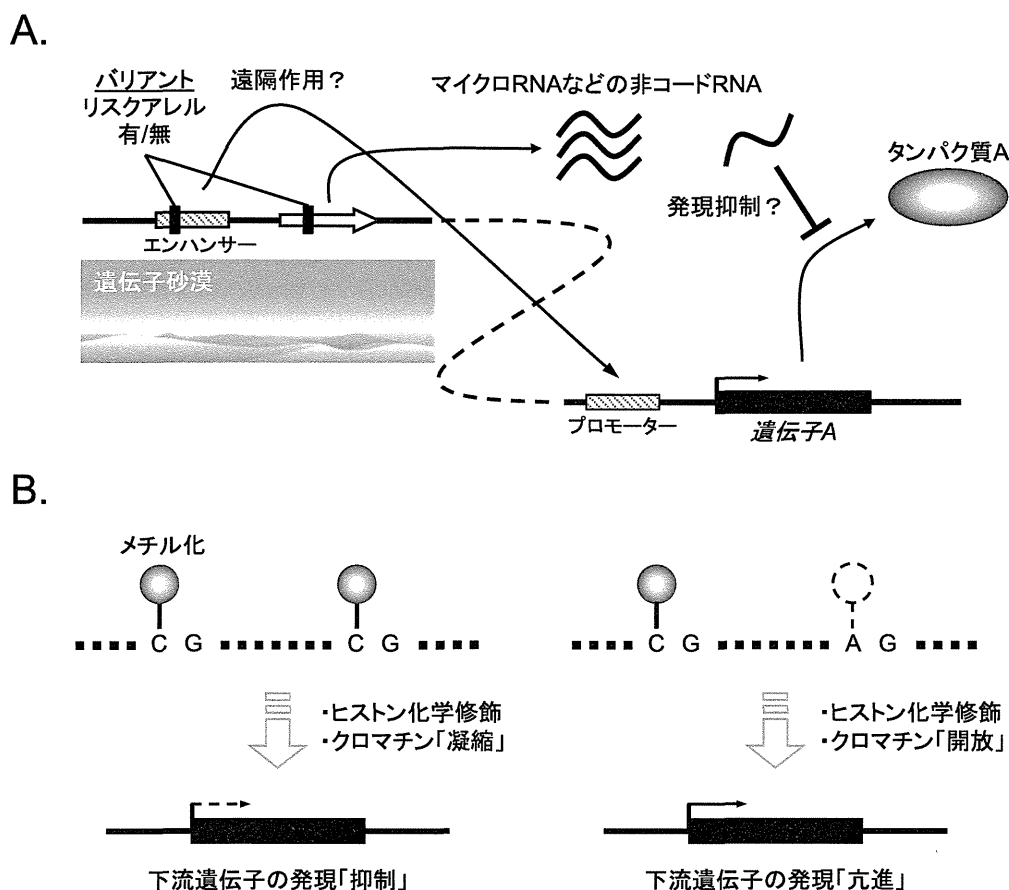


図3 遺伝子砂漠のバリエントの機能

(A) エンハンサーなどの調節配列や非コードRNA配列上にあるバリエントのリスクアレルの有無によって、遠隔遺伝子やタンパク質の発現量に変化している可能性がある。(B) エピジェネティックな修飾を受ける塩基がバリエントであった場合、下流遺伝子の発現量に影響をおよぼす可能性がある。(文献24より許可を得て改変のうえ転載)

シーケンスデータを産生することが可能な次世代シーケンサー (next-generation sequencer, NGS) が開発された (図1) ことによって、広範囲にわたるゲノム領域を詳細に解析することが可能になってきた。NGSでは、バリエントだけでなくその前後も含めた配列情報をこれ以上無い解像度 (塩基配列) で取得できるので、理論的には遺伝マーカーを用いることなくヒトゲノム上の全バリエントを検出することが可能である。現在、国際プロジェクトとして1000人ゲノム計画²⁵⁾ が進行しており (図1)、1000人分のヒトゲノムシーケンスデータの取得を目指している。その過程で、ハップマップ計画でアレイにより取得したバリエントのジェノタイプ

データの誤りを指摘したり、NGSのデータ産生量や解析能力の更なる進歩に貢献したりする波及効果もたらされている。しかし、NGSといえども一般施設においてヒト全ゲノムの高精度なシーケンスデータを多検体から取得することは検体処理能力およびコストの両面から依然難しい。そこで、シーケンサーに供するゲノムDNAの前処理として、染色体領域や構成成分 (遺伝子のエクソンなど) をあらかじめ特異的かつ効率的に捕獲する技術 (標的配列濃縮法) の開発と改良が盛んに行われている²⁶⁾。現在世界中のGWAS研究者は、同定された疾患関連領域を適切な標的配列濃縮法によって捕獲し、NGSを用いて領域全体の高精度な塩基配列情報を取

得ることによって、GWASでは明かされなかった多因子疾患の分子機序の本質に迫ろうと試行錯誤している。

最近、GWASを起点とする多因子疾患の分子機序の解明に向けた展開として、上述した冠動脈疾患に関連する9p21.3領域をNGSによって精査した例が報告された²⁷⁾。本研究では、バリエーションのリスクアレル保有者／非保有者（総計50人分）について当該領域の塩基配列をNGSで決定後、1) 領域内の全バリエーションの抽出、2) 配列予測プログラムによるエンハンサーの推定、3) NGSを用いたChromosome conformation capture (3C)法（核内でエンハンサーとプロモーターが調節タンパク質を介して物理的に会合している位置を塩基配列として捉える方法）によるエンハンサーの標的配列の同定、を実施した。

まず、配列予測の結果から9p21.3は全ゲノム中の遺伝子砂漠の中でもエンハンサーが2番目に高密度に存在することが判明した。従って、本領域が多数の遠隔遺伝子の発現を調節するハブ（中継地点）の役割を担っていることが示唆された。この結果は、GWASによってまったく異なる病態を示す複数の多因子疾患に関連するバリエーションが、実は同一領域の異なるエンハンサー上に位置している可能性を示した重要な知見である。加えて、*CDKN2B*のアンチセンスRNAが転写産物として検出される*CDKN2B-ASI*の配列中にもエンハンサーがコードされていることから、同一配列がRNAとDNAの異なる次元における機能をコードしていることを示唆する興味深い知見でもある。

次に、本研究では血管内皮細胞株を用いた*in vitro*の実験系で冠動脈疾患に関連するバリエーションを含むエンハンサー配列と転写因子(STAT1)との結合活性がリスクアレルの有無によって異なることを示した。更には、このエンハンサーが約950 kbも離れた*IFNA21*領域を始め、複数の遺伝子領域と会合していることを3C法で明らかにした。以上の結果は、統計学的に疾患と関連付けられたバリエーションの生物学的意義を初めて分子レベルで説明することができた画期的

な成果である。今後、本研究例のようにNGSを活用しながらGWASで同定された多数のバリエーションと疾患の発症機序との関連性を解明する糸口が得られることが期待される。

おわりに

数百個にもおよぶことが想定されている多因子疾患に関連する遺伝要因(バリエーション)(表1)をすべて洗い出すことは、精密な臨床診断により明確にケース群とコントロール群に分配された検体を多数(時に数万検体)準備することさえできれば可能になってきている。これまでGWASで汎用されてきた50万~100万プローブのアレイでは、ハップマップ計画の情報をもとに白人集団におけるアレル頻度が5%以上のcommonバリエーションが搭載されていたため、日本人にとってはそもそもバリエーションでない約3~4割に相当するジェノタイプデータを破棄していた²²⁾²³⁾²⁸⁾。しかし、最近主流の250万プローブのアレイでは、アジア人固有のバリエーションも搭載可能であり、commonバリエーションだけでなく1000人ゲノム計画の情報が反映されたアレル頻度が1%未満のエキソン上のrareバリエーションも搭載されている。加えて、NGSによるエクソーム解析が普及してきたことから、メンデル遺伝病に限らず多因子疾患に関連する未知のバリエーションがエクソン上から同定される例も増えてきている。従って、多因子疾患に特有なバリエーションを人種やアレル頻度によらず網羅的に検出する技術的な環境は整ってきている。

一方、最近では疾患群をいくつかのグループに亜分類し、それぞれに固有なバリエーションを同定する試みも盛んに行われている²⁹⁾。例えば、我々が実施したPOAGのGWASにおいても、POAG群を眼圧の高い群と低い群(正常眼圧緑内障)の病型別に分類して解析したところ、*CDKN2B-ASI*上のバリエーションは正常眼圧緑内障患者にのみ関連するバリエーションであることが明らかになった²³⁾。すなわち、解析対象疾患を特徴的な形質によって細分化することで、疾患の発症に関連する遺伝要因だけでなく、疾患の病型、進行の速度や方向性などを規定する因子

が同定できる可能性がある。今後は、GWASを起点とした疾患の発症に関連する遺伝要因の探索から原因バリエーションの生化学的な機能解析に至る横軸方向への広がりだけでなく、疾患それ自体を構成する複雑な遺伝的背景を追究する縦軸方向への深化の両方が推進されることで、多因子疾患の本質を理解するゲノム医科学研究が急速に発展していくことが予想される。

多因子疾患に関連するバリエーションが同定され尽くした後は、遺伝子を中心とした従来のシグナル伝達系に加えて、ゲノム上の全バリエーションとそれらの機能を加味したネットワークが描けるであろう。そして、9p21.3領域のように複

数の多因子疾患が共有しているハブとなるような領域が整理されて浮かび上がってくるに違いない。しかし、多因子疾患の真の病態を理解するためには、遺伝要因を統合的に理解するだけでは不十分である。環境要因との長期間にわたる（＝加齢に伴う）微弱な相互作用が疾患に与えるエピジェネティックな影響を解明するためには従来の分子生物学的手法では限界があり、多因子疾患の病態を反映したモデル生物の開発などの更なる技術革新が今後の大きな課題である。

開示すべき利益相反状態はない。

文 献

- 1) Lander ES, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860-921.
- 2) Venter JC, et al. The sequence of the human genome. *Science* 2001; 291: 1304-1351.
- 3) Gardiner K. Human genome organization. *Curr Opin Genet Dev* 1995; 5: 315-322.
- 4) Claverie JM. Gene number. What if there are only 30,000 human genes? *Science* 2001; 291: 1255-1257.
- 5) Carninci P, et al. The transcriptional landscape of the mammalian genome. *Science* 2005; 309: 1559-1563.
- 6) 池田陽子, 中野正和, 森 和彦. 緑内障の遺伝子. *臨床眼科* 2012; 29: 355-357.
- 7) Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 2009; 10: 241-251.
- 8) Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SF, Polychronakos C, Hakonarson H. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet* 2011; 7: e1002293.
- 9) 池田陽子, 中野正和, 田代 啓, 森 和彦, 木下 茂. 緑内障の検査診断学 3. 遺伝子診断 2011; 53: 207-220.
- 10) Singer JB. Candidate gene association analysis. *Methods Mol Biol* 2009; 573: 223-230.
- 11) The International HapMap Project. *Nature* 2003; 426: 789-796.
- 12) Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; 449: 851-861.
- 13) Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? *J Hum Genet* 2010; 55: 195-206.
- 14) Topol EJ, Murray SS, Frazer KA. The genomics gold rush. *JAMA* 2007; 298: 218-221.
- 15) Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; 308: 385-389.
- 16) Gold B, Merriam JE, Zernant J, Hancox LS, Taiber AJ, Gehrs K, Cramer K, Neel J, Bergeron J, Barile GR, Smith RT; AMD Genetics Clinical Study Group, Hageman GS, Dean M, Allikmets R. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006; 38: 458-462.
- 17) Yates JR, Sepp T, Matharu BK, Khan JC, Thurlby DA, Shahid H, Clayton DG, Hayward C, Morgan J, Wright AF, Armbrecht AM, Dhillon B, Deary IJ, Redmond E, Bird AC, Moore AT; Genetic Factors in AMD Study Group. Complement C3 variant and the risk of age-related macular degeneration. *N Engl J Med* 2007; 357: 553-561.
- 18) Maller J, George S, Purcell S, Fagerness J, Altshuler

- D, Daly MJ, Seddon JM. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006; 38: 1055-1059.
- 19) Samani NJ, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; 357: 443-453.
- 20) Scott LJ, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; 316: 1341-1345.
- 21) Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* 2010; 6: e1000899.
- 22) Nakano M, Ikeda Y, Taniguchi T, Yagi T, Fuwa M, Omi N, Tokuda Y, Tanaka M, Yoshii K, Kageyama M, Naruse S, Matsuda A, Mori K, Kinoshita S, Tashiro K. Three susceptible loci associated with primary open-angle glaucoma identified by genome-wide association study in a Japanese population. *Proc Natl Acad Sci USA* 2009; 106: 12838-12842.
- 23) Nakano M, Ikeda Y, Tokuda Y, Fuwa M, Omi N, Ueno M, Imai K, Adachi H, Kageyama M, Mori K, Kinoshita S, Tashiro K. Common variants in CDKN2B-AS1 associated with optic-nerve vulnerability of glaucoma identified by genome-wide association studies in Japanese. *PLoS One* 2012; 7: e33389.
- 24) 中野正和, 池田陽子, 森 和彦. 緑内障 Genome-Wide Association Study 最新の知見 : 2. 次世代シーケンサーをいかに活用するか. *あたらしい眼科* 2012; 29: 355-357.
- 25) Kuehn BM. 1000 Genomes Project promises closer look at variation in human genome. *JAMA* 2008; 300: 2715.
- 26) 中野正和, 田代 啓. 大規模シーケンサー解析用ヒトゲノム標的配列濃縮法. *実験医学* 2010; 28: 3147-3153.
- 27) Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* 2011; 470: 264-268.
- 28) Ueta M, Sotozono C, Nakano M, Taniguchi T, Yagi T, Tokuda Y, Fuwa M, Inatomi T, Yokoi N, Tashiro K, Kinoshita S. Association between prostaglandin E receptor 3 polymorphisms and Stevens-Johnson syndrome identified by means of a genome-wide association study. *J Allergy Clin Immunol* 2010; 126: 1218-1225 e10.
- 29) Nakano M, and Tashiro K. Association studies getting broader: a commentary on 'A polymorphism of the POLG2 gene is genetically associated with the invasiveness of urinary bladder cancer in Japanese males'. *J Hum Genet* 2011; 56: 550-551.

