

## がん関連遺伝子の SNP 探索

太田 力<sup>1)</sup>

1) 常葉大学保健医療学部理学療法学科

## 要 旨

飲食物、薬、タバコ煙などに含まれる化学物質、酸化ストレス、複製エラー、紫外線、放射線等による DNA 障害は、DNA の突然変異を誘発する。DNA の突然変異は遺伝情報の不安定化、ひいては、がん化、老化、遺伝病などの発症の原因になっていることが知られている。ヒトは、遺伝情報の安定を保ちながら生体を守るために化学物質の解毒システムや DNA 障害を修復する多様な修復システムを獲得してきた。これらのシステムの機能は人によって異なる。つまり、異なる体質をもって生まれる。この体質に影響を与えている一つが SNP（一塩基多型）である。我々は、解毒システムや DNA 修復システムの機能の差異が、がん発症や抗がん剤の感受性などと関連しているのではないかと考えた。そこで、がん患者 100 名と健常者ボランティア 20 名から採取した血液からゲノム DNA を抽出し、既存のがん関連因子、DNA 修復酵素、解毒酵素の発現を制御している主な転写因子を合わせた 283 遺伝子の exon 領域内の塩基配列の解析を行った。その結果、アミノ酸変異を引き起こす 607 個の cSNP を含む 1952 個の SNP を見出した。

キーワード：SNP, DNA 修復酵素, 解毒酵素, exon

## はじめに

背景：多くの人のゲノム配列を比較すると 1000 塩基に 1 塩基の割合で多様性（異なる塩基）があることがわかってきた。この違いを SNP（single nucleotide polymorphism：SNP）という。SNP の数は 300 万個以上あることが推定され、この一塩基の違いは目や髪の毛の色、背の高さ、顔つきなどの個性となって現れている。この個性の中には、病気の成り易さ・成り難さなど多くの疾病との関わりもあることが予想されており、SNP と

病気に関連を突きとめる研究が盛んに行われている。国立がん研究センターでも、がんに関わる SNP を見出すことになった。2000 年代前期は、限られた遺伝子の SNP が発見されているに過ぎなかった。さらに、そのほとんどは欧米人で高い頻度で見つかる SNP であったため、日本人では非常に低い頻度で見つかる SNP が多く含まれていた。従って、日本人の病気との相関解析に利用できる SNP の数は少なかった。そこで、我々は大規模なゲノム解析を行うことで、日本人の

SNP を探索することを考えた。

着想の経緯：塩基配列の解析は、キャピラリー塩基配列解析装置（キャピラリーシーケンサー）を用いて行うため、我々の研究室での解析能力は1日に200万塩基（2 Mb）程度であった。従って、我々にはヒトの全ゲノム配列（31億塩基：3.1 Gb = 3100 Mb）を解読することは不可能であった。SNP の定義がヒト集団の中での1%以上とされていたため、解析サンプル数は最低でも日本人100名が必要であった。我々は安全を期して120名の解析を行うことにした。SNP 探索は通常健常者を用いて解析されていたが、我々は敢えてがん患者を用いて解析を行なった。その理由は、SNP の中には健常者には低頻度（1%程未満）で見つかるが、がん患者では高頻度（1%以上）で見つかるものがある可能性を考えた。つまり、健常者のみの解析ではがんに関連するSNPの見落としがあるのではないかと考えた。そこで、100名のがん患

者（肺がん、胃がん、大腸がん、食道がん、乳がん、白血病など）と20名の健常者ボランティアのサンプルを解析することにした。解析サンプル数が決まったことで、次に解析遺伝子の数の算出を行なった。酵素の機能に影響を与えるSNP、つまりアミノ酸変異を引き起こすSNPが最もがんに関連があると考えた。そこで、遺伝子のタンパク質をコードしている exon 領域内のSNP探索を行うことにした。1つの遺伝子でタンパク質をコードしている exon の平均的な長さは150塩基程で、平均的な個数は10個程度であったので、我々の予算で解析できる遺伝子の数は約300と計算された。既存のがん関連因子（遺伝子変異によってがん発症に関与することが既に報告されていた因子）の約100遺伝子に加え、がんと関係が予想される遺伝子のリストアップを行った。がん研究センター研究所では、以前から動物モデルを用いた化学発がんの研究が盛んに行われており、化学物質による発がん過程に解毒酵素（代謝酵素）と

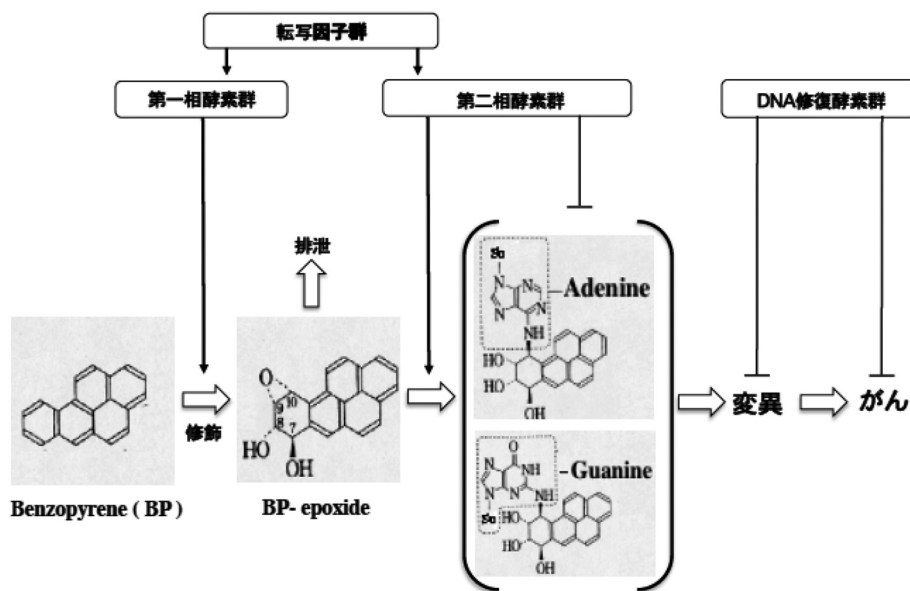


図1. 解毒酵素とDNA修復酵素による発がん抑制過程

化学物質ベンツピレン（Benzopyrene; BP）は、第一相解毒酵素群によって修飾（エポキシ化）され、発がん性を示すBP-epoxideとなる。BP-epoxideは第二相解毒酵素群が抱合（結合）することで親水性となり、細胞外へ排泄される。しかし、排泄されずに残ったBP-epoxideはDNA（アデニンやグアニン）に結合し、遺伝子変異を誘発する。この遺伝子変異ががん発症を誘発する。多くの解毒酵素群は毒物の侵入時に特定の転写因子群によって大量産生されることから、これらの転写因子群も解毒酵素群やDNA修復酵素群と共にがん抑制に関与していることがわかる。

DNA 修復酵素の関与が知られていた(図 1).

そこで、解毒酵素と DNA 修復酵素の遺伝子リストを作成したところ、主な解毒酵素は 134 遺伝子、DNA 修復酵素は 151 遺伝子あることがわかった。しかし、このプロジェクトの前年から、中村祐輔先生（現がんプレシジョン医療プロジェクト理事）の率いる東京大学のグループが、日本人（健常者）の解毒酵素を中心とした遺伝子の SNP 探索を開始していたため、解毒酵素は候補遺伝子から除外した。解毒酵素の多くが特定の転写因子によって制御されていることが知られていたもので、我々は解毒酵素の発現を制御している主な転写因子（32 遺伝子）を解析リストに加えた。これで、既存のがん関連因子の 100 遺伝子、DNA 修復酵素の約 151 遺伝子、解毒酵素の発現を制御している主な転写因子の 32 遺伝子を合わせた 283 遺伝子が SNP 探索のためのゲノム解析の遺伝子として絞り込まれた。

#### 解析方法

我々は、説明文書・同意文書により同意を得た、国立がん研究センターに来院した患者 100 名、および、がんには罹患していない健常者ボランティア 20 名の血液を収集した。各サンプルからゲノム DNA を抽出後、各遺伝子のアミノ酸をコードしている exon 領域を

挟んだ intron 領域内に設定したプライマーを用いて、多チャンネルキャピラリー型塩基配列解析装置（ABI 3700）によって塩基配列を解読した。解読した塩基配列は、Genbank の登録塩基配列を標準配列として SNP 探索ソフト POLYPHRED を用いて SNP 抽出、SNP 分類を行った。SNP のアレル頻度は各 SNP の存在する割合を%として計算し、頻度の低い SNP の割合で表示した。

#### 結 果

解析した 120 名の塩基配列を相互に比較しながら、異なる塩基を SNP として抽出した。この解析によって、日本人における 1952 個の SNP を見出すことに成功した（表 1）。また、アレル頻度を調べてみると、5%未満の SNP が 1009 個あり全体の 52%を占めていることがわかった。この結果から、低頻度の SNP が非常に多いことが判明した。

なお、SNP は見出される領域やアミノ酸変異の有無で以下のように 5 種類に分類されている。①タンパク質をコードしていないイントロン領域内に見つかった SNP(intron SNP: iSNP)、②タンパク質をコードしているエクソン領域内に見つかったものであるが、アミノ酸に変化を与えない SNP(silent SNP: sSNP)、③タンパク質をコードしているエクソン領域内に見つかったものであり、

表 1. SNP の数とアレル頻度

アレル頻度(X; %)	$x \leq 10$	$10 < x \leq 20$	$20 < x \leq 30$	$30 < x \leq 40$	$40 < x \leq 50$	合計
cSNP	469	52	33	36	27	607
sSNP	493	75	46	46	50	710
iSNP	293	77	76	92	97	635
( junction SNP )	( 5 )	( 1 )	( 2 )	( 0 )	( 0 )	( 5 )
( In & del SNP )	( 87 )	( 10 )	( 10 )	( 2 )	( 10 )	( 74 )
合計	1255	204	145	174	174	1952

なお、( ) で示した全ての Junction SNP と In & del SNP は iSNP に含まれていた。

且つアミノ酸変異を起こす SNP (coding SNP: cSNP), ④スプライシング配列内の SNP でスプライシングの異常を誘発する可能性のある SNP(Junction SNP), ⑤塩基が挿入あるいは欠失していた SNP (Ins & Del SNP) である。これらの SNP の中で、我々が最も注目したアミノ酸変異を引き起こす SNP である cSNP は 607 個見出すことができた。Ins & Del SNP は 119 個あったが、全て intron 領域内で見つかったため、アミノ酸変異を引き起こすものではなかった。失っていた SNP (Ins & Del SNP) である。これらの SNP の中で、我々が最も注目したアミノ酸変異を引き起こす SNP である cSNP は 607 個見出すことができた。Ins & Del SNP は 119 個あったが、全て intron 領域内で見つかったため、アミノ酸変異を引き起こすものではなかった。

## 考 察

我々は、120 名の 283 遺伝子の塩基配列の解読から、日本人における 1952 個の SNP を見出すことに成功した。これらの SNP の中で、exon 領域内で見つかった数は 1317 個であった。ヒトの遺伝子の総数は 23,000 程度と推定されているので、1.3%の遺伝子を解析したことになる。この解析結果を外挿して計算すると、ヒトの 23,000 遺伝子のアミノ酸をコードしている exon 領域内の SNP 数は約 10 万個と推定される。この中で、アミノ酸変異を起こす cSNP は 4.6 万個と推定される。従って、この 4.6 万個の SNP は、遺伝子の機能に影響を与える可能性のある SNP と考えられる。また、解読した intron 領域内の塩基数は exon 領域内の 10%程度と少なかったが、iSNP は 635 個と全体の 33%を占めていた。このことは、iSNP の大半が遺伝子の機能に影響を与えない SNP であるため、淘汰されずにヒトのゲノム内に残ったと推測される。

我々は、日本人における 1952 個の SNP を見出すことに成功したが、SNP を見つただけでは、どの SNP ががんに関連するものかわからない。我々は、全ての SNP 解析の完了を待たずに、早期に得られた比較的アレル頻度の高い (10%以上の) cSNP を用いて、1002 名の非小細胞肺がんの患者と 685 名の健常者コントロールで相関解析を行なった。その結果、TP53, POLI, REV1およびLIG4 の 4 つの DNA 修復因子の SNP が非小細胞肺がんの易罹患性と相関することがわかった<sup>1)</sup>。また、251 名の小細胞肺がんの患者と 685 名の健常者コントロールで相関解析を行なったところ、DNA 修復因子 MTH1 の SNP が小細胞肺がんの易罹患性と相関することがわかった<sup>2)</sup>。さらに、1542 名の乳がん患者と 1462 名の健常者コントロールで相関解析を行なったところ、DNA 修復因子 NBS1 の SNP が乳がんの易罹患性と相関することがわかった<sup>3)</sup>。

今回見出した SNP はアレル頻度が低いものが多かったため、相関解析に利用できる SNP は限定されてしまった。アレル頻度が 5%未満の SNP を相関解析に利用するには数千人以上のがん患者や健常者コントロールが必要となる。従って、低頻度の SNP ががんに関連があるのか調べるには、公共データベースなどに公開されているがん患者や健常者の塩基配列情報を用いて相関解析を行う必要がある。さらに、相関解析でがんに関連が見つかった SNP をがんの予防や治療に応用するには、これらの SNP が、どのようにがんに関連しているのか、その分子機構を調べることが重要であり、今後の課題でもある。

## 謝 辞

遺伝子解析は国立がん研究センターで行われたものである。特に、疾病ゲノムセンターのスタッフの皆様に感謝する。

## 引用文献

- 1) Sakiyama T, Kohno T, et al: Association of Amino Acid Substitution Polymorphisms in DNA Repair Genes, TP53, POLI, REV1 and LIG4, with Lung Cancer Risk. *Int. J Cancer* 114:730-737,2005.
- 2) Kohno T, Sakiyama T, et al: Association of polymorphisms in the MTH1 gene with small cell lung carcinoma risk. *Carcinogenesis* 27:2448-2454,2006.
- 3) Yamamoto Y, Miyamoto M, et al: A Rare Polymorphic Variant of *NBS1* Reduces DNA Repair Activity and Elevates Chromosomal Instability. *Cancer Res* 74:3707-3715,2014.