

# 癌とゲノム科学

P. Andrew Futreal\*, Arek Kasprzyk†, Ewan Birney†, James C. Mullikin‡, Richard Wooster\*§ & Michael R. Stratton\*§

\* Cancer Genome Project and ‡ Informatics Division, Sanger Centre, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK

† EBI-EMBL, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK

§ Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK

腫瘍発生の原因遺伝子を同定することは癌研究の中心的目標である。我々は、ヒトゲノム概要配列から推測されるタンパク質を対象に、既知の腫瘍抑制遺伝子に対する側系遺伝子(paralogue)を探索した。しかし、新規遺伝子は確認できなかった。次に、癌細胞のゲノム配列とヒトゲノム概要配列とを比較し、癌遺伝子の配列変化を直接検出できるかどうかを検討した。融合癌遺伝子(染色体転座によって生じ、両末端がゲノム上の異なる座位に検出される)由来のキメラと思われる転写産物は、正常組織と癌組織の双方で同程度に検出された。これは、偽陽性の検出レベルが、かなり高いことを示している。今回の我々の実験は、現時点では癌細胞から得られているDNA配列が、量的にも質的にも限られていることを強調する結果となった。

癌はすべて、DNA塩基配列の異常によって引き起こされる。ヒト細胞のDNAは、一生の間、たえず変異原物質にさらされ、複製の誤りを起こすので、個々の細胞にはDNA塩基配列のわずかな変化が少しずつ蓄積する。これらの体細胞変異の1つが、ある決定的な遺伝子の機能を変化させると、その変異が生じた細胞の増殖を有利にすることがある。すると、この細胞から派生した細胞群(クローン)が出現し、さらにこれに関連した遺伝子群を標的とする変異と、その結果生じるクローン拡張の波が引き続き、周囲の組織に侵入して転移する細胞をつくりだす。癌は最もありふれた遺伝性疾患である。欧米諸国では3人に1人が癌になり、5人に1人が癌で死んでいる<sup>1</sup>。

これまで、約30個の劣性癌遺伝子(腫瘍抑制遺伝子)と100個以上の優性癌遺伝子が同定されている。癌遺伝子を同定するために従来とられていた最も成功率の高い方法は、地図作成戦略を用いて遺伝子の位置をゲノム地図上の狭い領域にしぼり込み、次にその領域中の候補遺伝子を選別して癌の症例に見られる変異を選択することだった。しかし、この戦略には限界がある。地図作成の情報は不明瞭でまぎらわしい場合があり、ゲノム上に明白な「目印」を残さない癌遺伝子もあるため、その地図を用いて簡単に遺伝子の位置を決められないものもある(たとえば、1個のアミノ酸が変わるだけの塩基置換によって活性化される優性癌遺伝子)。このような変異を起こした遺伝子は、実際のところ従来の検出法では捕らえられない。これらの遺伝子を同定するには、癌に関連した生物学的情報に基づいて有力な候補遺伝子を選択する方法がとられてきた。

ゲノム塩基配列は、残された癌遺伝子の同定にどのくらい役立つだろうか。可能性としては、既知の癌遺伝子の側系遺伝子(paralogue)を探索し、膨大な候補遺伝子の一覧表をつくれればよい。そこで我々は、ヒトゲノム<sup>2</sup>に書き込まれたタンパク質のうち現時点で「本物らしい」と考えられるタンパク質群の中から、既知の劣性癌遺伝子(腫瘍抑制遺伝子)の側系遺伝子を探索した(補完資料参照;表1)。我々は、そのような遺伝子族の大部分を検索したが、採用した検索条件(最少50個のアミノ酸配列のうち50%以上が同一であること)では新規遺伝子の明白な証拠は見つからなかった。ほかの多数の遺伝子族では、新しい側系遺伝子が見つかる<sup>2</sup>のに癌遺伝子の側系遺伝子が見つからないのは、タンパク質の調べ方が不完全だったのかもしれない。この結果を確認するため、予測される遺伝子と読み枠の可能性をすべて考慮し、癌遺伝子の1つ、TP53の側系遺伝子を概要ヒトゲノム配列全体の中から徹底的に探索した<sup>3</sup>。この

追加探索法でも、未知の側系遺伝子は見つからなかった。

新規の側系遺伝子が見つからないのは、これらの遺伝子族が生物学的にも医学的にも重要なため、遺伝子の大部分がすでに見つかっているからかもしれない。しかし、ゲノムの塩基配列未決定領域に新たな側系遺伝子が隠れていることもありうるし、遺伝子予測が不完全なために検出されなかったのかもしれない。また、側系遺伝子の検出自体が、使われる検索基準に依存する。検索条件の厳密度を低くすると予想どおり、側系遺伝子に的中する数が事実増加したが、その遺伝子の多くは生物学的に疑わしいものだった。

このやり方は、有望な癌遺伝子候補の探索に本当に有効なのだろうか。変異を起こした癌遺伝子は、ある種の遺伝子族(たとえば、情報伝達キナーゼ群やGDP結合タンパク質群)にいくつも見られるが、これらの遺伝子族は大きく、ほとんどの場合、そのうちのごくわずかな遺伝子だけが癌に関与している。実際、癌遺伝子は、構造と機能の多様性が著しく、たとえば、既知の劣性癌遺伝子が他の癌遺伝子と高い相同性をもつことはまずない。これらの癌遺伝子が指令するタンパク質は多種多様な生物学的機能や生化学的機能に関係している。しかも、重要な癌関連遺伝子の類縁遺伝子の多く(たとえば、TP53と塩基配列類似性を示すp73遺伝子とp63遺伝子<sup>4</sup>)は、癌において変異するとの報告はない(ただし、遺伝子発現の変化により変わる可能性はある)。したがって、過去の経験の影響をあまり受けなければ、癌を誘発する変異についてもっとわかるかもしれない。構造や推定される機能が何であろうと、ともかくあらゆる遺伝子あるいはタンパク質を候補と考えて探究する努力を続けるべきである。

既知の癌遺伝子に類似したゲノム塩基配列を探るだけでは、癌遺伝子を見つける機会を逃すかもしれない。今回得られた概要配列は、遺伝子の塩基配列のほかに、遺伝子の編成、遺伝子構造および染色体に沿った順序に関する情報も含んでいる。癌には、ゲノムの崩壊と解体という特徴がある。もしかすると、癌細胞におけるゲノムの構造変化を検出するための鋳型として、ゲノム塩基配列を使えるかもしれない。

そのためには、今回の概要配列(最終的には完成配列)を癌細胞ゲノム由来の該当塩基配列と比較する必要がある。しかし、利用できる癌ゲノム塩基配列はほとんどなく、あっても不完全なものしかない。癌ゲノム由来の塩基配列で最大のものは、腫瘍組織から作成したcDNAライブラリのクローンを部分的に調べる計画(たとえば、癌ゲノム解析計画 Cancer Genome

Anatomy Project: CGAP [http://www.ncbi.nlm.nih.gov/CGAP/]に由来する。原理的には、これらの癌ゲノム配列を今回の概要配列と比較し、腫瘍抑制遺伝子の不活性化や優性癌遺伝子の活性化をもたらす可能性のある体細胞塩基置換および小さな挿入や欠失を探してみることも考えられる。しかし、実際はこれらのデータベースに含まれる塩基配列は比較的わずかしかなかく、しかも、どの種類の腫瘍についても転写産物の一部分しか調べられていない。そのうえ、(癌を引き起こす変異はタンパク質のコード領域に集まっているのに)利用できる塩基配列はおもに遺伝子の非翻訳領域である。たとえ意味のあるまれな体細胞変異が検出されたとしても、配列決定の誤り(cDNA ライブラリにもゲノム塩基配列にもある)の中に埋もれているか、あるいは無意味な多型の森の中に隠れているだろう。

我々は、今回得られた概要配列とともにこれらのcDNAライブラリの塩基配列を利用し、癌に見られる別の種類の変化を

探してみた。癌遺伝子の活性化をもたらす遺伝子再編成は、染色体転座の結果として生じることがある。この種の異常は、白血球、リンパ腫および肉腫ではよく見られ<sup>5,6</sup>、キメラ転写産物(染色体切断点の両側の転写される遺伝子部分からできた融合遺伝子の産物)がつくられることが多い(ただし、染色体転座が完全な転写産物の調節異常をもたらすだけの例もある)。興味をそそられることに、この遺伝子再編成による発癌様式は成人に多い上皮癌ではあまり報告されていない。破壊されて複雑になった上皮癌細胞ゲノムの中に再編成が隠れているのか、あるいは単に上皮癌には再編成が存在しないのかは、近い将来にわかるかもしれない。

我々は、そのような遺伝子再編成の証拠を探すため、(正常組織標本と腫瘍組織標本から作成したcDNA ライブラリ由来の)CGAPプログラムの塩基配列をすべて入手し、その中から2つの塩基配列が得られていたクローン(正常組織由来215,889クローンと癌組織由来25,446クローン)を選択した。

表1 側系遺伝子の探索に用いた劣性癌遺伝子

遺伝子記号	アクセッション番号	遺伝子名	同定された側系遺伝子*	癌症候群	(生殖系列および/または体細胞変異が検出された)癌の種類
APC	P25054	大腸腺腫症遺伝子	APC様遺伝子	家族性大腸ポリポシス	大腸癌, 膵臓癌, 類腺腫, 肝芽細胞腫
BRCA1	P38398	家族性乳癌/卵巣癌遺伝子1	なし	遺伝性乳癌/卵巣癌	遺伝性乳癌/卵巣癌
BRCA2	P51587	家族性乳癌/卵巣癌遺伝子2	なし	遺伝性乳癌/卵巣癌	遺伝性乳癌/卵巣癌
CDH1	P12830	カドヘリン1遺伝子	N カドヘリン, P カドヘリン, R カドヘリン	家族性胃癌	小葉性乳癌
CDKN2A	P42771	サイクリン依存性キナーゼインヒビター2A(p16)遺伝子	なし	皮膚悪性黒色腫2	黒色腫, その他
CYLD	CAB93533	家族性円柱腫症遺伝子	なし	家族性円柱腫症	円柱腫
EP300	Q09472	300-kD E1A結合タンパク質遺伝子	Rubinstein-Taybi 遺伝子	該当なし	大腸癌, 乳癌, 膵臓癌
EXT1	Q16394	多発性外骨腫症1型遺伝子	EXT類似1	多発性外骨腫症1型	外骨腫症, 骨肉腫
EXT2	Q93063	多発性外骨腫症2型遺伝子	EXT類似3	多発性外骨腫症2型	外骨腫症, 骨肉腫
CDKN1C	P49918	サイクリン依存性キナーゼインヒビター1C遺伝子	なし	ベックウィズ・ビーデマン症候群	ウィルムス腫瘍, 横紋筋肉腫
STK11	Q15831	セリン/トレオニンキナーゼ11遺伝子	なし	ポイツ・ジェガーズ症候群	空腸過誤腫, 卵巣腫瘍, 睾丸癌(精巣癌), 膵臓癌
MAP2K4	P45985	MAPキナーゼキナーゼ4遺伝子	MAPキナーゼ遺伝子群	該当なし	膵臓癌, 乳癌, 結腸癌
MEN1	O00255	多発性内分泌腫瘍症1型遺伝子	なし	多発性内分泌腫瘍症1型	副甲状腺/下垂体腺腫, 島細胞癌, カルチノイド腫瘍
MLH1	P40692	大腸菌MutL相同遺伝子	なし	家族性非腺腫性大腸癌	大腸癌, 子宮内膜癌, 卵巣癌
MSH2	P43246	大腸菌MutS相同遺伝子2	なし	家族性非腺腫性大腸癌	大腸癌, 子宮内膜癌, 卵巣癌
NF1	P21359	神経線維腫症1型遺伝子	なし	神経線維腫症1型	神経線維腫, 神経膠腫
NF2	P35240	神経線維腫症2型遺伝子	モエシン	神経線維腫症2型	髄膜腫, 聴神経腫
PRKAR1A	P10644	プロテインキナーゼAタイプ1-α調節サブユニット遺伝子	なし	カーニー複合体	粘液腫, 内分泌腫瘍
PTCH	Q13635	ショウジョウバエの <i>patched</i> 遺伝子の相同遺伝子	パッチド2	母斑様基底細胞腫症候群	基底細胞腫, 髄芽細胞腫
PTEN	O00633	ホスファターゼおよびテンシン相同遺伝子	なし	カウデン症候群	過誤腫, 神経膠腫, 前立腺癌, 子宮内膜癌
RB1	P06400	網膜芽細胞腫遺伝子	なし	家族性網膜芽細胞腫	網膜芽細胞腫, 肉腫, 乳癌, 小細胞肺癌
SDHD	O14521	コハク酸脱水素酵素シトクロムB小サブユニット遺伝子	なし	家族性傍神経節腫	傍神経節腫
MADH4	NP_005350	ショウジョウバエの <i>mothers against decapentaplegic 4</i> 遺伝子の相同遺伝子	なし	若年性ポリポシス	胃腸ポリープ, 大腸癌, 膵臓癌
SMARCB1	Q12824	Swi/Snf5 アクチン依存性クロマチンマトリックス付着調節因子遺伝子	なし	桿状素因症候群	悪性桿状腫瘍
TP53	P04637	腫瘍抑制p53遺伝子	p73遺伝子, p63遺伝子	リー・フラウメニ症候群	肉腫, 副腎皮質癌, 神経膠腫, その他
TSC1	Q92574	結節性硬化症1遺伝子	なし	結節性硬化症1	過誤腫, 腎細胞癌
TSC2	P49815	結節性硬化症2遺伝子	なし	結節性硬化症2	過誤腫, 腎細胞癌
VHL	P40337	フォン・ヒッペル・リンドウ症候群遺伝子	なし	フォン・ヒッペル・リンドウ症候群	腎細胞癌, 血管腫, 褐色細胞腫
WT1	P19544	ウィルムス腫瘍1遺伝子	Znフィンガー遺伝子群	家族性ウィルムス腫瘍	ウィルムス腫瘍

\*50個以上のアミノ酸配列の50%以上が同一のもの

これらの塩基配列の大部分は、そのクローンに挿入されたDNA断片のどちらか一方の末端配列である。次に、ゲノム塩基配列の中でこれらの2つのcDNA塩基配列に適合するものを探し、染色体上の位置を決定した(補完資料参照)。

同じcDNAクローン由来の2つの塩基配列の大部分は、これらが1つの正常転写産物由来なら予想されるとおりに、ゲノム地図上の同じ位置に決定された。ところが、1つのcDNAクローン由来の2つの塩基配列がゲノム地図上の別々の位置に定まる場合がいくつかあった。これらは、染色体転座によって生じたキメラ遺伝子の転写産物に由来するのだろうか。その可能性もあるが、この戦略には限界がある。この概念上は単純な実験にさえも、低頻度の反復配列や多数の同一配列をもつ遺伝子のコピーのようにゲノムに本来備わる複雑性がつきまわった。そのうえ我々の解析からは、キメラと思われる転写産物が正常組織からも癌組織からも同じ比率で(どちらの場合も出発時のクローンの3%)得られた。これは、偽陽性が有意に検出される可能性を示している。これらの偽陽性は、cDNAライブラリ作成時の人為的に生じたキメラクローン、塩基配列の誤読、データベースの注釈と管理上の誤りあるいは概要配列の誤整列から生じたのだろう。

当然のことだが、この実験ではこのような研究用には計画されていなかった材料を使ったので、それほど成果が得られなかったとしても驚くにあたらない。偽陽性のほかにも問題がある。塩基配列が決定されたのはライブラリの中の比較的少数のクローンであり、ライブラリの多くは正規化されていない(あまり多くない転写産物の量が低めに見込まれる)。しかも、調べたcDNAの多くは完全長ではない。このことは、既知のキメラ遺伝子をもつ5種類の癌から作成したライブラリについてCGAP注釈を探索したときに、5種のキメラ転写産物に關与する10個の遺伝子のうち1個しか見つからなかったことからわかった。ここで強調しておくが、ゲノムレベルで癌の複雑性を説明するには、癌ゲノム由来の塩基配列のデータがもっとたくさん必要である。そして、その塩基配列のデータは、この課

題に使えるように適切に配置される必要があるだろう。

したがって、癌細胞ゲノムの異常の本質が今回の概要配列からすぐに明らかになることはないだろう。これらの解析を促進するには、完成配列が必要である。完成配列は、癌細胞ゲノムと正常ゲノムを大規模に比較するような新世代の研究を形成する構造的枠組みとなるだろう。最終的には、完成配列が当初のゲノム上の位置決定に依存する戦略とは別の戦略への脱皮を促し、体系的なゲノム全体にわたる変異の探索が可能になるだろう。そのためには新しい技術が必要である。癌細胞に存在するあらゆる様式の異常(大きな欠失、再編成、塩基置換、小さな挿入と欠失、増幅、およびメチル化などの後成的変化)を一挙に検出する技術は、今のところ存在しない。癌ゲノムから作成したゲノムライブラリの塩基配列の解読は、この目標への一番の近道なのだろうが、癌の多様性と、十分な被覆度をもったヒトゲノム配列を得るのに必要な労力と費用を考えると、これは大いなる挑戦である。

#### 引用文献

1. Higginson, J., Muir, C. & Munoz, N. *Human Cancer: Epidemiology and Environmental Causes* (Cambridge Monographs on Cancer Research, Cambridge, UK, 1992).
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001).
3. Birney, E. & Durbin, R. Using GeneWise in the Drosophila annotation experiment. *Genome Res.* 10, 547-548 (2000).
4. Kaelin, W. G. Jr The p53 gene family. *Oncogene* 18, 7701-7705 (1999).
5. Rowley, J. D. The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.* 32, 495-519 (1998).
6. Bell, R. S., Wunder, J. & Andrulis, I. Molecular alterations in bone and soft-tissue sarcoma. *Can. J. Surg.* 42, 259-266 (1999).

補完資料は(<http://www.nature.com>)で閲覧でき、印刷版はネイチャーのロンドン編集部から入手可能。

[原文]

Futreal, P.A. *et al.* *Cancer and genomics.* *Nature* Vol.409 No.6822, pp. 850-852 (15 February 2001)