

システムがんNewsletter No. 7

2013年3月

システムの統合理解に基づくがんの先端的診断、治療、予防法の開発

文部科学省科学研究費補助金 新学術領域研究（複合領域：4201）

研究の紹介

治療によって修飾されるがんのゲノミクス解析と バイオマーカー探索

石川 俊平（研究代表者）

東京医科歯科大学難治疾患研究所ゲノム病理学分野 教授

新薬開発のためのバイオマーカーをなるべく 早い段階で発見し個別化医療に結び付ける

新薬を生み出すには多くの費用が掛かり、一品目につき100億円規模とも言われています。これだけ研究開発費を投じながら新薬が患者のベッドサイドに届く確率は消して高くありません。ある報告では臨床治験に入っただけのものうち8割近くは脱落すると言われ、その確率は科学技術の進歩によって年々上がるどころか低くなってきている傾向にあります。その原因としては新薬は常に既存の薬よりも高い効果のものを作らなければならない必要が出てきたからです。新薬は臨床治験でその効果の試験を行います。そこでは大半の医薬品の候補が効果が認められなかったり、副作用が出現したりという理由で脱落してしまいます。こうしたケースのほとんどにおいては一部の患者のみに効果が見られたり、副作用が出現します。ある新薬候補に効果のある患者、副作用のおこる患者を層別化するマーカーをバイオマーカーと呼びますがこれは近年ゲノム科学の発達によってたくさん報告されるようになってきました。例えば特定遺伝子の変異を有する癌症例だけが分子治療薬に奏功したり、特定のSNPを持つ患者だけに副作用が起こったりするという例はよく知られています。こうしたゲノムバイオマーカーを使うことによって効果的に医薬品が奏功する患者を選別したり



中央が石川

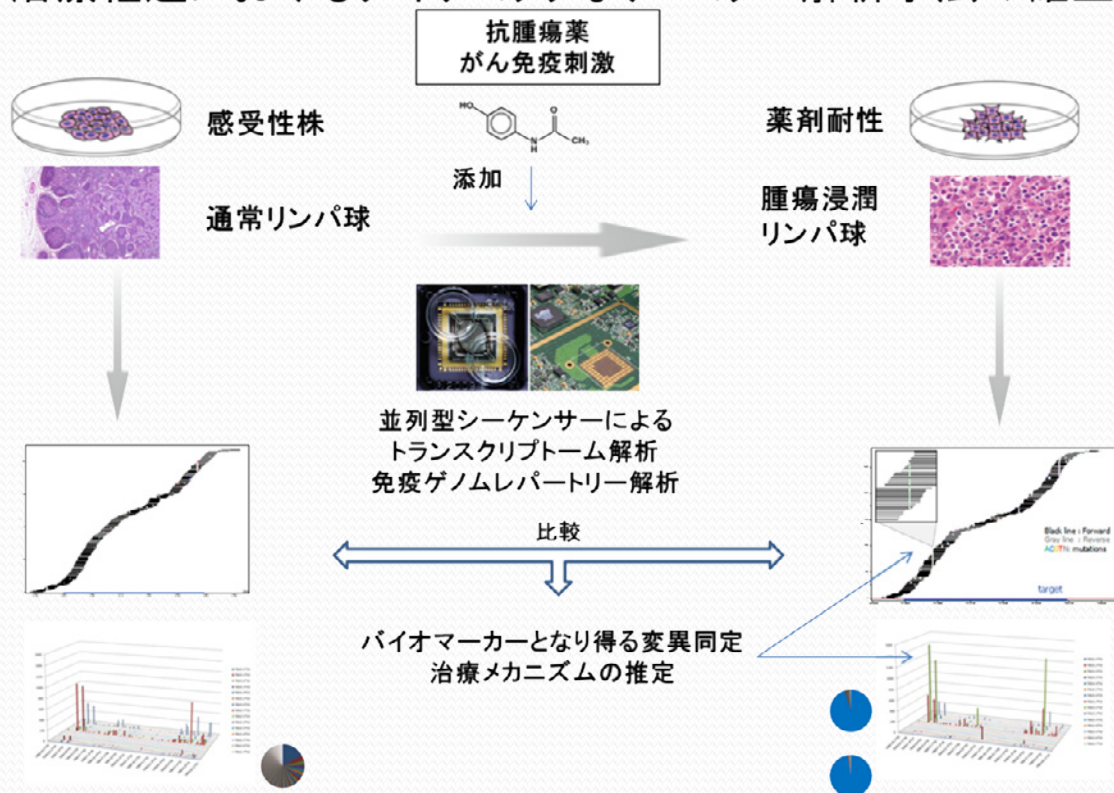
副作用の起こる患者への投与を避けることが可能です。非常に有用なゲノムバイオマーカーですが、現実にはこうしたバイオマーカーが判明するのはクスリが市販されてかなり経ってから、効果のある患者や副作用の患者さんのDNAを多数調べることによって判明するのが普通で実際に臨床治験に生かされることはほとんどありません。臨床治験は多くの期間と費用がかかり一度失敗して脱落した医薬品候補をもう一度治験をやり直すのはかなり難しい現実もあります。できれば臨床治験のかなり早期の段階まででこうした有効なバイオマーカーが判明すれば実際の臨床治験をそのバイオマーカーを使って効果的に進めることが出来ます。こうして前臨床治験までに多大な費用をかけた医薬品化合物がベッドサイドに届く確率をあげられれば医療資源の大きな有効活用になり医療費を抑制することが可能となります。

（次ページへ続く）

臨床治験の早期段階で医薬品候補の奏功性、副作用に関するバイオマーカーを特定することは容易ではありません。基本的には医薬品候補を投与した患者さんのDNA検体が十分手に入らない状態で探索することが必要となります。医薬品化合物の標的分子が開発段階からはっきりしている場合は、バイオマーカーの探索はそれほど難しいことはありません。しかし最近のマルチキナーゼ阻害剤や、HDAC阻害剤などのように正確にどのタンパクが最も重要なバイオマーカーとなり得るか簡単には想像できないものが増えてきています。標的分子以外のメカニズムで代謝、排泄が起こる場合も同様で多くの遺伝子を含むゲノム全体からバイオマーカーを探索する必要があります。実験室で細胞を使って探索できれば非常に有効ですが、従来の探索技術では多くの遺伝子を含むゲノム全体のなかから重要なこうしたバイオマーカーを選ぶことは難しいことでした。しかし近年のゲノム・ゲノミク

ス測定技術の発展は目覚ましいものがあり、特に並列型シーケンサーという技術の発達によって1塩基のレベルの変化を捉え、また極めて精度の高いデジタルカウントで核酸を定量できるようになりました。例えば細胞に薬剤を添加することによって変化するゲノム情報・ゲノミクス情報を高精度に解析することができれば医薬品候補のバイオマーカーを全ゲノムから探索することができる可能性があります。本研究では様々な医薬品候補化合物や治験中医薬品候補をがんの細胞株に加えその変化をゲノミクスの手法で解析します。それは短時間の遺伝子発現変化であることもあれば、長期の添加によって安定して獲得されるDNAレベルの変化であったりします。また機能的ゲノミクスと呼ばれる手法を用いて探索することもあります。本研究の目的は個別の医薬品化合物のバイオマーカーを探すことも重要ですがそれだけが目的ではありません。この研究で解析法が確立すれば大学や公的研究機関だ

治療経過におけるダイナミックなゲノミクス解析手法の確立



けでなく、実際に医薬品候補を作っている製薬企業もそれぞれの研究施設で同様の解析を定着させることによって、本当に早期からのバイオマーカーを生かした新薬の開発が可能となります。

また最近では分子標的治療薬とともに免疫治療ががんに対する新たな治療法として認知されてきます。日本でもペプチドワクチンを中心として多くの臨床研究が進んでいます。こうした免疫療法は一般的には治療効果がかなり時間がたってから判明するため、早期に効果を判定するバイオマーカーが求められている分野です。現在の良く使われている末梢血を用いた手法が十分ではなく、より高精度なバイオマーカーが求められています。免疫は非常に複雑なメカニズムで個々のリンパ球がその抗原受容体の遺伝子を再構成させて異なる配列を持っています。本研究では並列型シーケンサーを用いたこうしたリンパ球の個別の配列の解析も行い、免疫治療のバイオマーカーを探索することも目的としています。



スパコンを用いたがんの特異的な 変異を検出するソフトウェア "Genomon" の開発

千葉 健一（学術支援員）

東京大学医科学研究所ヒトゲノム解析センター 学術支援専門職員



近ごろテレビや新聞で遺伝子やゲノムを題材にした番組・記事を見かける機会が増えています。これには、遺伝子やゲノムの研究・解析の結果がもたらす新しい医療が現実味をおび、このような研究・解析に対しての世間からの興味が増えていることがあるのではないのでしょうか。遺伝子、さらにはゲノムのことが分かれば、「治せる病気がある」、「将来の病気への不安が軽減する」など期待があります。特にがんにおいては、個人の遺伝子やゲノムを解析することで早期発見や副作用の少ない的確な治療方法を見つけことが、そしてがんと診断されるより前からがんへの予防策を講じることができると期待されています。ところで、私たちの体にはなぜがん細胞というものができるのでしょうか。

ゲノムは生物の設計図であり細胞の振る舞いのプログラムです。体を作る細胞ひとつひとつのゲノムDNAには1日あたり5万～50万か所に傷がつくと見積もられていますが、この程度の傷であれば細胞のもつ自己修復能力の想定範囲であり、細胞は正常性を保ちます。異常な頻度のゲノムDNAの損傷や修復を担う遺伝子自身に傷を負うことがあります。この場合の細胞は、自らを破壊するアポトーシスの機能を働かせ、自律的に消滅することで自らの暴走を防ぎます。このような安全機構を幾重にも備える細胞ですが、しかし運が悪く深い傷が重なるなどの原因でこれらの機構を失い、傷を負った遺伝子が増殖し続け、秩序ある振る舞いができなくなった状態に陥ることがあります。これが「がん」細胞が生まれる瞬間です。

次世代DNAシーケンサーの登場によって、ゲノムDNAの塩基配列を読み取ることが比較的安価に短時間で可能になったことで、どこどここの遺伝子のDNAに運の悪い深い傷がつく（遺伝子変異が起こると言います）とがんになるのかを解明するための活

発な研究が世界的に始まっており、「システムがん」のチームは、その分野でトップレベルの技術を開発し、成果を上げてきました。本稿では、私たちの研究室で行なっているがんのゲノム、タンパク質の設計図を含む遺伝子領域（エクソン領域といいます）の全DNA塩基配列（約50万塩基対）から遺伝子変異を検出する方法とその解析を行うソフトウェア“Genomon”（<http://genomon.hgc.jp/>）を紹介します。

私たちは、ヒトゲノム解析センタースーパーコンピュータ（HGCスパコン）を使用してがん遺伝子の解析をおこなっています。HGCスパコンは、生物学系の利用者に使いやすいように設計・運用されており、ゲノム解析のための各国で開発されたソフトウェアを他の一般的なスパコンに比べ容易に利用することができます。

ゲノムDNA塩基配列の読み取りでは、まずとても長いDNAを超音波や薬品で数百塩基の長さのDNA断片に切断します。読み取り装置である次世代DNAシーケンサーはそのDNA断片を並列作業によってA、C、G、Tの塩基の文字として判別しながら読み取っていきます。あるDNA断片は下記のようなDNA塩基の配列断片（リードといいます）として出力されます。

```
GCAGAGGCCACCGAAAGTCAAAAGCGGGGCTTGG  
GAAGGCCCGCCGGGAGGCATGAGCTGGGCTGGGCC  
GAAAGAGGCCACTGGGAGGCAGGAGGAGCTGG
```

このリードは、このままではゲノムのどの部分の配列であるはわかりません。そこで、ヒトリファレンスゲノムと呼ばれる読み取りが完了している標準的なヒトゲノム配列（約30億塩基対）を利用します。ヒトリファレンスゲノムの中にリードに似た配列を探します。ジグソーパズルのピースの絵を完成図の中を探すようなものです。この作業をマッピングと呼びます。エクソ

ン領域の場合、DNA配列断片は数億リード出力されリードごとにリファレンスゲノム配列のどの部分に対応するかを探り出すことになります。これは単純作業でありコンピュータにとっては簡単なものと思われるかもしれませんが、しかしリードはシーケンサーの誤読や、がん細胞が元来持っている変異が含まれるため、マッピングではあいまいな一致を探すことになります。このような条件下での膨大なピース量のジグソーパズルを精度よく解くには多くの計算量が必要なことはお分かりいただけるでしょうか。

そこでGenomonではHGCスパコンを使用し、マッピングをおこなうソフトウェア"BWA"を並列実行することで、マッピング速度の大幅な改善を行いました。また並列実行には大量のメインメモリを要求しますが、HGCスパコンはこれに応え、滞りのない並列実行を実現できました。

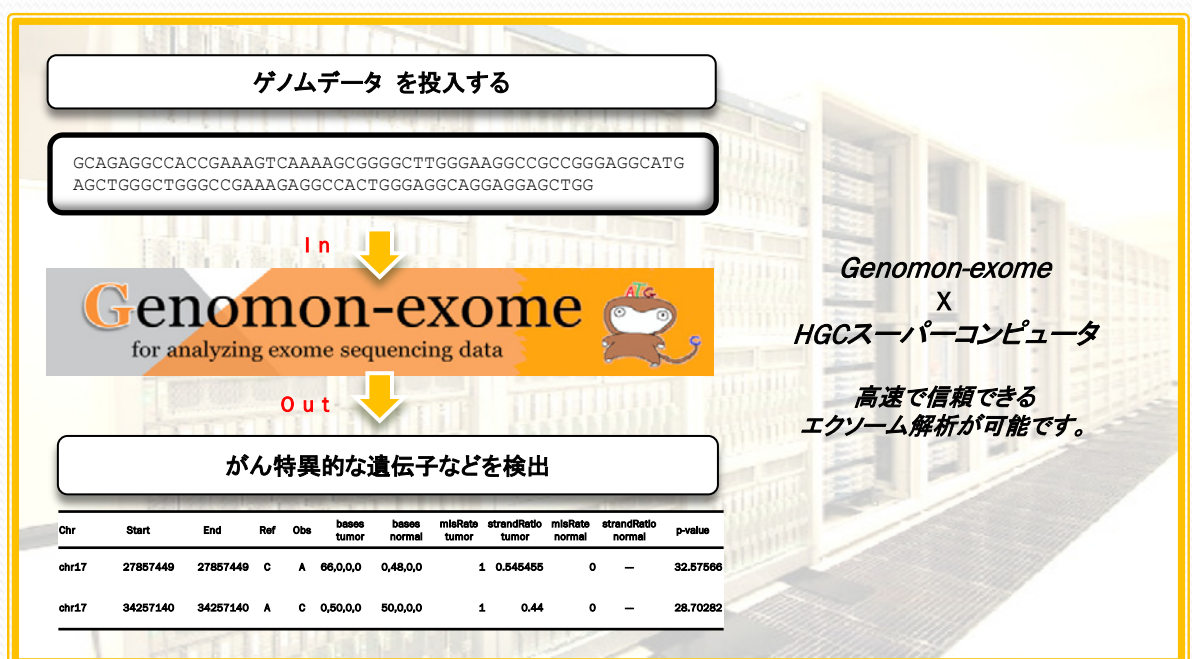
マッピング後、Genomonは解析を行います。マッピングの結果、DNAの配列断片がヒトの遺伝子のどの部分であるかがわかるようになるため、どの遺伝子のどの位置に変異があるかが検出できるようになります。あいまいさのある大量のデータに対しての解析は容易ではありませんが、統計的手法を用い、シーケンサーの読み取りエラーや、マッピングの間

違いに対処しつつ、がん特異的な変異を検出します。特に、「経験ベイズ法」を用いたがん特異的な変異の検出力[1] は世界トップレベルを誇り、ユーザーの方々からも極めて高い評価を得ています。これらの解析もHGCスパコンの並列処理能力を活かしたものです。

がん特異的な変異がある遺伝子がわかると、抗がん剤や分子標的治療薬の効果も予測することができるようになると期待されています。将来、予防・診断・治療のためにDNAを調べる人が増加し、何万人分ものデータを扱うようになると、より速く正確な結果を出力するために、スパコンと堅牢なデータ解析システムの構築が必要です。その上、情報保護の観点からは、セキュリティが万全で、地震や火事などの災害時にもデータが守られるかなど、不測の事態にも対応できるようなスパコンシステムの仕組みづくりも重要になります。

私たちは解析ソフトウェア“Genomon”を公開し、獲得したノウハウや開発した情報解析技術をバイオインフォマティクス分野で共有することによって、分野全体の効率化・レベルアップに貢献できればと考えています。

[1] <http://nar.oxfordjournals.org/content/early/2013/03/06/nar.gkt126.abstract>





文部科学省新学術領域研究

システムがん

システムの統合理解に基づくがん
の先端診断、治療、予防法の開発

オフィシャルWebサイト

<http://systemscancer.hgc.jp/>

システムがんの研究内容、構成研究者情報、発表論文、研究集会などの詳しい情報はこのWebサイトをご覧ください。



<http://twitter.com/SystemsCancer>

システムがんの研究成果、世界のがん研究、研究拠点、学会情報など最先端な話題をつぶやいています。

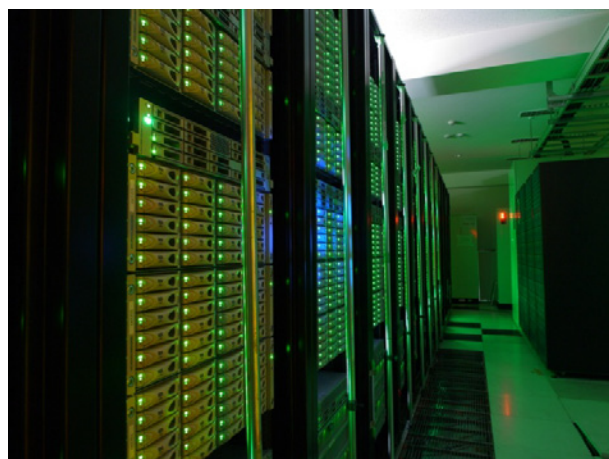


Human Genome Center

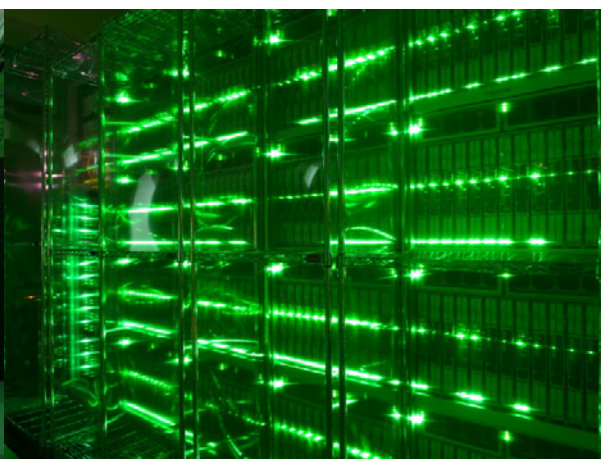
東京大学医科学研究所ヒトゲノム解析センター スーパーコンピュータ

<http://supcom.hgc.jp/>

システムがんで用いられるデータ解析ソフトウェアは、ヒトゲノム解析センターのスーパーコンピュータで動いています。スーパーコンピュータはどなたでもご利用になれます(有償)。



大容量のディスク装置



高密度の計算機

編集後記

さくら～さくら～♪

新学術領域研究「システムがん」ニュースレター No.6

発行日★平成25年3月31日(初版)、平成25年4月10日(二版)

発行★システムの統合理解に基づくがんの先端診断、治療、予防法の開発
領域代表者★宮野 悟

- 東京大学医科学研究所 ヒトゲノム解析センター DNA情報解析分野
- 〒108-8639 東京都港区白金台4-6-1
- TEL: 03-5449-5615 FAX: 03-5449-5442
- E-mail: miyanolab-jimu@edelweiss.hgc.jp

編集★サトウアユ