

ポスト「京」重点課題 2

個別化・予防医療を支援する統合計算生命科学

# NEWS LETTER

Vol. 10

## Contents

- Research Report

何してるの?なぜ?どうやって?ちょっと解りやすく  
教えて!に研究者が応える

スーパーコンピュータが照らし出す  
がんの多様性

- Information

お知らせやイベント情報

INTEGRATED  
COMPUTATIONAL  
LIFE  
SCIENCE  
TO SUPPORT  
PERSONALIZED AND  
PREVENTIVE MEDICINE



ポスト「京」  
重点課題 2  
個別化・予防医療を支援  
する統合計算生命科学

■ Research Report

Subtheme **A**

# スーパーコンピュータが照らし出すがんの多様性



サブ課題 A  
 東京大学医科学研究所  
 新井田 厚司

サブ課題 A「大量シーケンスによるがんの個性と時間的・空間的多様性・起源の解明」は、解析技術の進歩によって得られるようになったゲノムのビッグデータから、生命現象や病気の理解につながる情報を引き出すための研究を行っています。その一環として取り組んでいるのが、「がんの起源と集団内多様性を理解するためのシステムの開発」です。ポスト「京」をはじめとしたスーパーコンピュータを使ったシミュレーションを行うことにより、がんの一生をたどり、生命現象に潜む原理を探りだします。今回は、その内容や取り組みについて最新の治療戦略も交えながら報告します。

## がんは遺伝子の変異によって生じる病気

がんは多くの先進国において、患者数、死亡者数が増加し続けていて、さまざまな治療法が開発されているものの、発見が遅れると多くの場合いまだに不治の病となっています。がんは正常細胞の増殖能力の異常で起こる病気です。正常な細胞では、細胞分裂は厳密に制御されており、周りの細胞とコミュニケーションを取りながら、必要な時だけ分裂するようになっています。細胞は約 22000 種類のさまざまな働きをするタンパク質で構成されています。がん細胞は細胞分裂を制御するアクセルやブレーキの働きをするタンパク質が壊れているので、周りの細胞からの制止命令を受け取らず、自分勝手に分裂し最終的には体中に広がって患者を死に至らしめます。

なぜアクセルやブレーキに相当する細胞の部品が壊れてしまうかという、その設計図である遺伝子に変異と呼ばれるエラーが入るからです。遺伝子は、細胞の中心となる核と呼ばれる領域にある DNA という化学物質に書き込まれています。ヒトの DNA は、ATCG で表される 4 種類の塩基が 30 億個連なってでき、約 22000 個のタンパク質の設計図が暗号化され埋め込まれています。塩基の並びを DNA 配列と言い、この約 22000 個のタンパク質の遺伝子を含む 30 億塩基対の DNA 配列をゲノムと呼びます。ゲノムは生物種内ではほとんど一緒ですが、ヒトの場合 1000 塩基に一塩基程度、異なる部分があります。それが肌の色や背の高さ、特定の病気のなりやすさ等の個人の体質の多様性を決めています。

タバコや紫外線等の発がん物質は、

DNA に傷をつけ遺伝子に変異を生じさせます。一番よく起こる変異は、一つの塩基が他の塩基に置き換わる一塩基置換です。正常細胞は分裂する際に複製した DNA を二つの娘細胞に受け継ぎますが、その際変異が乗じる可能性があるので、歳を重ねるにつれて細胞のゲノムの

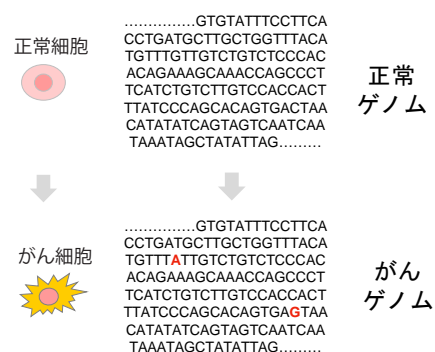


図1: がんの原因  
 正常細胞はゲノムに変異を蓄積することによってがん細胞に変化する。



## ■ Research Report

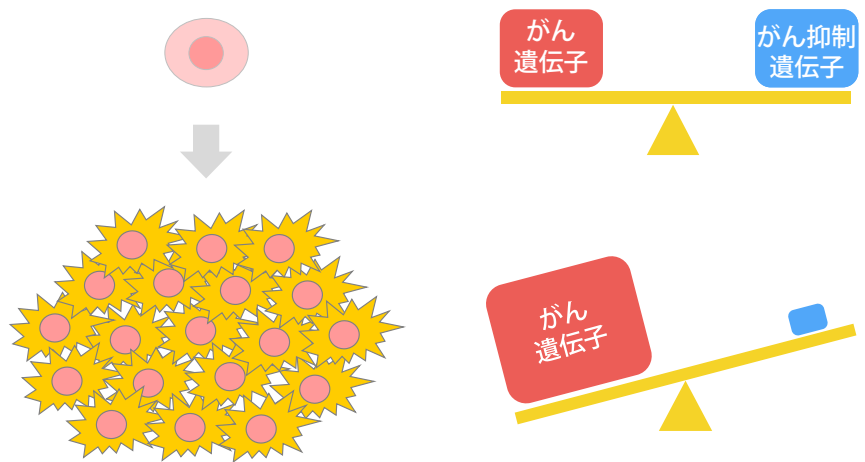
### Subtheme A

変異がさらに多くの変異を蓄積します(図1)。アクセルの働きをするタンパク質の元となる遺伝子のことをがん遺伝子、ブレーキではがん抑制遺伝子と呼びます。正常細胞ではがん遺伝子とがん抑制遺伝子の機能はうまくバランスが保たれていますが、それらにたまたま変異が入り、アクセルが過剰に働きブレーキが間かなくなることで細胞が異常増殖し始め、がん細胞に変化します(図2)。がん遺伝子の有名なものでは *KRAS*, *EGFR*, *MYC*、がん抑制遺伝子の有名なものでは *TP53*, *APC*, *RB1* が知られています。発がん物質や加齢ががんの原因となる機構は以上のように説明されます [1]。

### 情報解析によって進むがん患者間の多様性の理解

20世紀後半のがん研究は、一つ一つの遺伝子(およびその遺伝子に基づいて作られるタンパク質)の働きを細胞や動物を使った実験によって調べる分子生物学が中心で、それにより、さまざまながん遺伝子やがん抑制遺伝子の働きが明らかになりました。その結果にしたがって分子標的薬という新しいタイプの薬が作られています。従来の抗がん剤は増殖の早い細胞を無差別に狙う絨毯爆撃のような粗っぽい薬ですが、分子標的薬は変異を持った遺伝子から作られる異常なタンパク質を標的とするピンポイント爆撃のような切れ味の鋭い薬です。しかし、分子標的薬は劇的に効く場合がある一方で効かないことも多く、それは標的とする遺伝子に変異を持っている患者と持っていない患者の違いによることが分かってきました。つまり患者によって変異があるがん遺伝子、がん抑制遺伝子の種類が異なり、それによって適切な治療法が異なってきます。

このようながんの個性を理解するに



**がん遺伝子:** 細胞を増殖させるアクセル *KRAS*, *EGFR*, *MYC*

**がん抑制遺伝子:** 細胞増殖を停止させるブレーキ *TP53*, *APC*, *RB1*

図2: がん遺伝子とがん抑制遺伝子

変異によりがん遺伝子が活性化したりがん抑制遺伝子が不活性化した結果、細胞の異常増殖が起こる。

は、がん細胞の DNA 配列のどこに何の変異が入っているかを調べることで、言い換えればがん細胞のゲノム(以降簡単にがんゲノムと呼びます)を解読することが必要です。そのためにはまずシーケンサーという機器を使って DNA 配列を決定(シーケンス)します。近年、シーケンサーの性能が劇的に向上しています。ヒトゲノムプロジェクトがヒトゲノム解読終了を宣言したのが 2003 年ですが、その当時はヒトゲノム一人分をシーケンスするのに 13 年、約 3000 億円かかりました。その後の技術革新により、次世代シーケンサーという新しい機器が開発され、今日ではヒトゲノム一人分をシーケンスするのに 3 日、10 万円程度しかかかりません [2]。

安価で短時間に大量のゲノムデータの取得が可能になりましたが、次に問題になってくるのは情報解析のステップです。次世代シーケンサーの特徴として、大量の 100 文字の短い DNA 配列が出力されます。その短い DNA 配列からゲノムを解析するためには、元のヒトゲノム配列に

復元しなくてはなりません。それは例えるなら 30 億文字からなる書類をシュレッダーにかけて、出てきた 10 億個の紙くずを全て繋ぎ合わせて復元するようなものです。がんゲノムを解析する場合は、さらに正常細胞のゲノムと比較して、変異を同定し、さらにはがん遺伝子、がん抑制遺伝子(以降まとめてドライバー遺伝子と呼びます)を絞り込む細やかな計算が必要となります。このような膨大な計算には効率的なソフトウェアと大規模な計算資源が必要になります。がん研究分野ではこれまで主流だった分子生物学的実験に加えて情報解析が大きな柱となってきています。

次世代シーケンサーの登場以降、大量のがんゲノムデータが産出されるようになり、治療反応性の違いの裏に潜む、がんゲノムの多様性が明らかになってきました。例えば、国際がんゲノムコンソーシアムでは各国に大腸がん、乳がん、肺がん等、がんの種類を割り当てて、それぞれ数百人の患者から得たがん細胞の

■ Research Report

Subtheme **A**

DNA をシーケンスします。このような研究によってがんの種類間で異なるのはもちろん、同じがんの種類でも異なる患者間で著しいがんゲノムの多様性があることが明らかになってきました。この多様性は専門用語で腫瘍間不均一性と呼ばれています (図 3A)。情報解析を駆使することにより、これまでに見つかったいなかったドライバー遺伝子も多数同定され、発がんの機構についてもこれまでに無い深いレベルでの理解が可能となってきました [3]。

次世代シーケンサーによるがんゲノム解析は、基礎研究の枠を飛び越えて臨床に応用され始めています。病院では、患者のがん細胞から得た DNA をシーケンスし、解析したゲノム情報に基づいて治療方針を決める臨床シーケンスが導入され始めています。臨床シーケンスが本格的に行われるようになれば、これまで以上に膨大なデータが日々生み出され、また臨床応用の性格上、解析時間のさらなる短縮も求められます。今後、情報解析の重要性がますます高まるのは確実です。私たちはこのような将来を見据えて、大量のがんゲノムデータを迅速に処理す

るためのソフトウェアである Genomon パイプラインを開発し、スーパーコンピュータで運用を行なっています [4][5]。

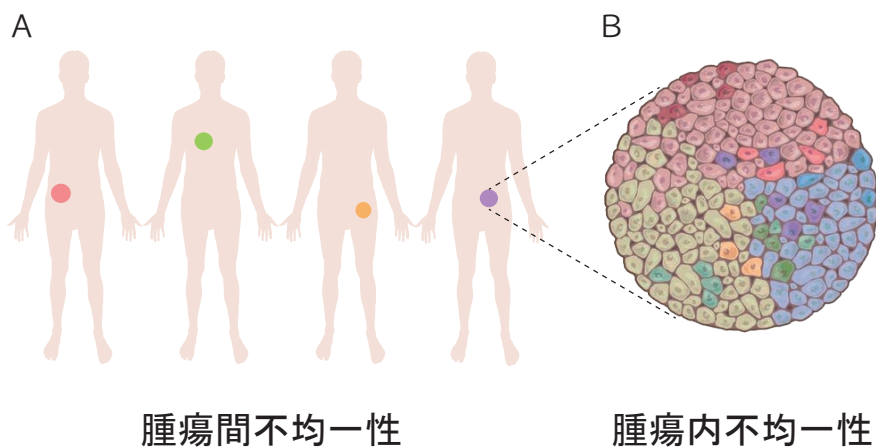
**さらなる多様性—腫瘍内不均一性—**

がんには腫瘍間不均一性以外にも、さらなるレベルの多様性が存在します。患者間だけでなく、一人の患者のがんの中にも異なるがんゲノムを持つ細胞集団が混在することが明らかになってきています (図 3B)。この多様性を腫瘍内不均一性と言い、腫瘍間不均一性と同様、非常に重要な問題です。分子標的薬がうまく効いて一時的にがんが小さくなくても、ほとんどの場合時間が経つと再びがんが増殖して再発に至ってしまう、治療抵抗性獲得の問題があります。この問題の要因が一人の患者の中のがんの多様性だと考えられています。例えば、大多数の細胞が、薬が効くタイプのがんゲノムを持っていたとしても、少数の細胞は薬が効かないタイプのゲノムだったとします。すると、初めは薬が効いたように見えても、少数の薬の効かない細胞が生き残り、それが

再増殖することで再発に至ります。これは雑草がいつぱいの庭に除草剤をまいて除草できたと思っても、しぶとく生き残る種類の雑草があると、それらが増えて、また庭が雑草でいつぱいになるのと似ています。

はじめの方で、がん細胞はヒトの正常細胞に変異が蓄積されてがん細胞に変化すると述べました。この発がんの過程は、詳しく見ると単一細胞からのダーウィン進化と捉えることができます。ダーウィン進化では環境に適した変異を持つ個体が自然による選択によって子孫を残す、自然選択によって進化が進みます。発がんの過程では、初めに単一の正常細胞のゲノムに変異がランダムに入ります。そのうちドライバー遺伝子への変異 (ドライバー変異) が入ると、その細胞は周りの細胞に比べて増殖速度が早くなり、自然選択を受け、同じゲノムを持った細胞集団であるクローンが形成されます。そのクローンに次のドライバー変異が入り同様の過程が繰り返されることで、単一の正常細胞が増殖能力の高いがん細胞集団へと直線的にダーウィン進化していきます (図 4A)。ダーウィン進化が腫瘍内不均一性を生み出す可能性もあります。一つのクローンに異なる複数のドライバー変異が入り、それぞれが自然選択を受けると、複数の新しいクローンが形成され腫瘍内不均一性が生まれるのです (図 4B)。

腫瘍内不均一性の研究は、腫瘍間不均一性の研究と同様に次世代シーケンサーの登場によって急速に進んでいます。通常のがんゲノムシーケンスにおいては、がんの一つの領域を切り出してそこから抽出した DNA をシーケンスしますが、腫瘍内不均一性に注目する際は、複数の領域を切り出し、それぞれから抽出した DNA をシーケンスする多領域シーケンスが用いられます (図 5)。多領域シーケン



**図3:** がんの二種類の多様性  
異なる患者間のがんの多様性を腫瘍間不均一性、一人の患者のがんの中の多様性を腫瘍内不均一性という。(16)のFigure 1を改変)

Research Report

Subtheme **A**

スで得られた多領域変異プロファイルには通常、すべての領域に共通して存在するファウンダー変異と多様性を形成しているプログレッサー変異の二種類が観察されます。プログレッサー変異は進化の後半に複数のクローンに別れる際に獲得された変異だと考えられます。これまでにこの多領域シーケンスによって複数のがんの種類についての腫瘍内不均一性の解析が報告されています [7]。

**がんの進化シミュレーションによる腫瘍内不均一性の理解**

大腸がんは、発がん過程で獲得されるドライバー変異の種類や順番等に関しては長い間研究され、直線的なダーウィン

進化によって生じてくると考えられてきましたが、腫瘍内不均一性の詳細な解析はなされていませんでした。そこで、私たちは大腸がんの腫瘍内不均一性を明らかにすべく多領域シーケンスを行いました。その結果、予想以上に強い腫瘍内不均一性が大腸がん中存在することがわかりました (図 6AB)。ダーウィン進化で腫瘍内不均一性が形成されていれば、プログレッサー変異にドライバー変異が観察されるはずですが、これまでに知られているドライバー変異は主にファウンダー変異に存在し、プログレッサー変異には観察されませんでした。そこで私たちは、発がんにおける進化の過程をシミュレーションすることにより、その強い腫瘍内不均一性を生み出す原理を探る

ことにしました。  
シミュレーション自体はとても簡単なものです。はじめに細胞をコンピュータの中で分裂させながらゲノムに変異をランダムに入れて行きます。ゲノムの中にドライバー変異が一つ入るごとに分裂速度が一定の割合で増加します。最後にシミュレーションで生じたがんから、実際の大腸がんからの実験データと同様の多領域シーケンスをコンピュータの中で行い、多領域変異プロファイルを取得します。このステップによって、変異確率、ドライバー変異の数、強さ等の条件を変えながら、スーパーコンピュータ「京」を使い数十万回のシミュレーションを行い、実験データと同様の強い腫瘍内不均一性ができる条件を探りました。

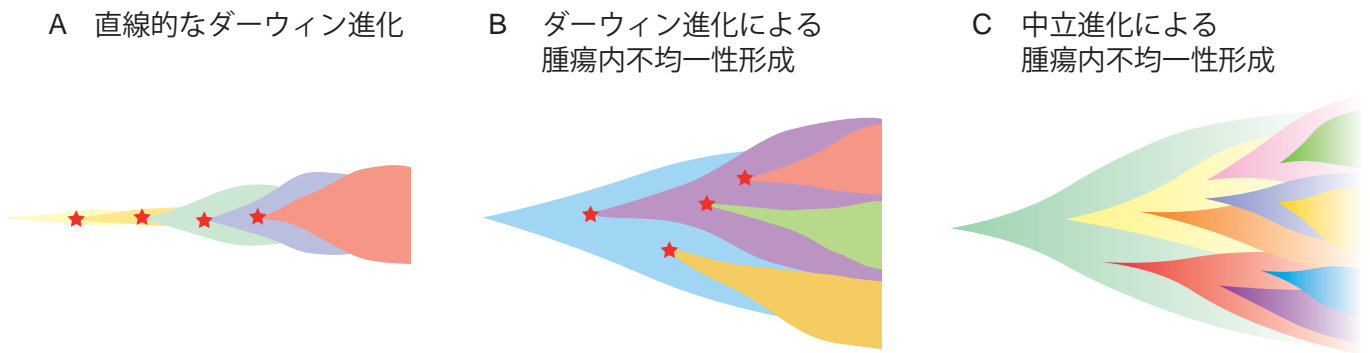


図4: 三つのがんの進化様式 赤い星印はドライバー変異を示す。

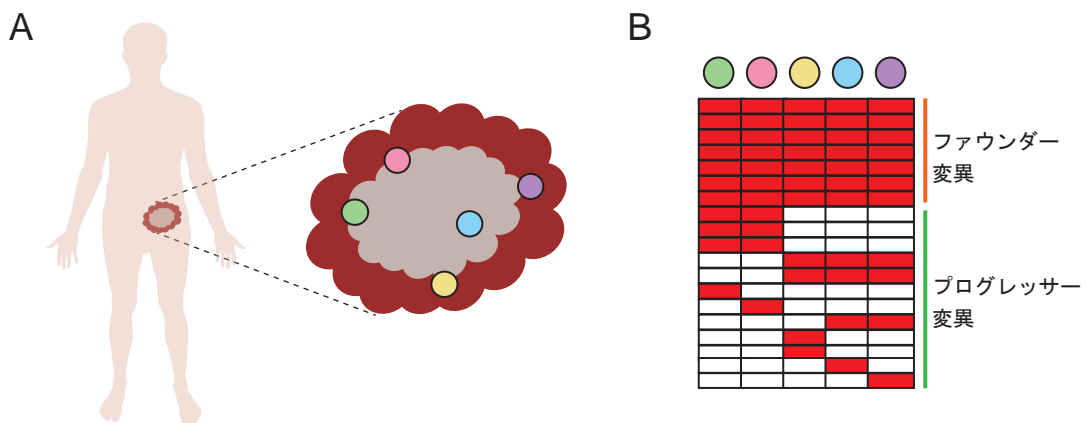


図5: 多領域シーケンス  
一つのがんの複数領域 (A) からDNAを抽出し、多領域変異プロファイル (B) を得る。

■ Research Report

Subtheme **A**

その結果、変異率が高い場合に実際の大腸がんの多領域変異プロファイルと同様の高い腫瘍内不均一性が再現されました(図6CD)。また、実験データと同様に、シミュレーションデータではドライバー変異はファウンダー変異に存在し、プログレッサー変異には存在していませんでした。これは腫瘍内不均一性が中立進化によって生み出されていることを示しています(図4C)。中立進化とは自然選択によるダーウィン進化に対して、増殖に有利でも不利でもない、中立な変異の蓄積の結果起こる進化の形式です。中立進化説は1968年に木村資生により提唱され、発表された当時は古典的なダーウィン進化支持者による激しい抵抗にあいましたが、現在では生物種間の遺伝的な多様性はおもに中立進化により生じるということは広く受け入れられています。また更に、シミュレーションによってできたがんを詳細に調べることによって、

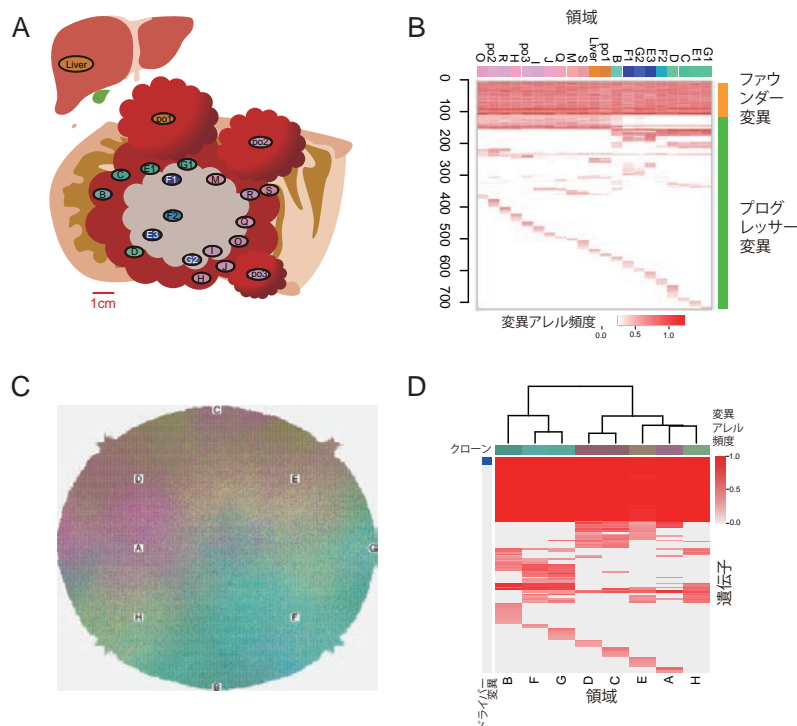
多領域シーケンスでは捉えられない無数の異なる中立変異を有するクローンが存在することも明らかになりました。

これらは大腸がんの進化のはじめに変異率が上がり、それが中立進化の引き金になっていることを示唆します。DNAの変異を修復しているがん抑制遺伝子が、変異して壊れることにより変異率自体が上がることは、これまでも報告されており、もっともらしい仮説です。また、ドライバー変異を蓄積したのちに中立進化により生み出される無数の異なる中立変異を有するクローンが、分子標的薬を使った治療時の再発にも関わっていると考えられます。中立な変異と言ってもそれは環境に依存するので、分子標的薬治療によって環境が変化すると、治療抵抗性の原因となる治療抵抗性変異に変わる可能性があります。つまり無数の異なる中立変異を有するクローンが存在すると、その中から治療抵抗性を持ったクローン

が必然的に出現してしまい、再発に至ると考えられます(図7)。

**治療抵抗性獲得の克服に向けた研究**

ではこのような腫瘍内不均一性に起因する治療抵抗性獲得の問題を克服するにはどうすれば良いでしょうか。最近の研究からその戦略が見えつつあるので最後に紹介したいと思います。通常、分子標的薬は、副作用の許容される最大容量で使用され、短期的にできるだけがんを縮小させるスケジュールで行われます。しかしシミュレーションによって容量やスケジュールを調整することにより、薬が効く細胞を完全に殺さず、あえてある程度残して薬が効かない細胞と競合させることで、短期的にはがんの縮小があまり見られなくても、長期的にがんの増殖を抑えられる可能性が示されています(図8)。



**図6:** 大腸がん多領域シーケンスとシミュレーション  
大腸がんの多領域シーケンスの結果 (A, B) および高い変異率を仮定したシミュレーション結果 (C, D)。



■ Research Report

Subtheme **A**

除草剤が効く草は効かない草と養分をとりあっているため、完全に除草剤が効く草を除いてしまうと、効かない草が自由に育ってしまいます。あえて、除草剤を調整し、効く草を残しておくことで、効かない草の生育を妨げるのです。庭の場合は多少見栄えが悪いかもかもしれませんが、がん患者にとっては長く生きるに越したことはありません [8]。

以上、スーパーコンピュータを用いたがん研究の実際を簡単に紹介させていただきました。次世代シーケンサーの登場によって、分子生物学的実験が中心であったがん研究の現場は状況が一変しました。今後、膨大なデータからがんの多様性を照らし出すスーパーコンピュータを利用した情報解析およびシミュレーションはますます重要性を増していくと思われます。少しでもスーパーコンピュータを用いたがん研究の面白さが伝わることを願って結語とします。

<参考文献>

[1] Weinberg, R. (2013). *The biology of cancer*. Garland science.  
 [2] Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333.  
 [3] Garraway, L. A., & Lander, E. S. (2013). Lessons from the cancer genome. *Cell*, 153(1), 17-37.  
 [4] <http://genomon.readthedocs.io/ja/latest/>  
 [5] 伊東聡, 矢留雅亮. がんの遺伝子解析とスーパーコンピューティング. ポスト「京」重点課題2 個別・予防医療を支援する統合計算生命科学 NEWS LETTER. 2017, vol.6, p. 2-5.  
 [6] Edward J.Fox & Lawrence A.Loeb. One cell at a time. *Nature* 512, 143-144(2014).  
 [7] Niida, A., Nagayama, S., Miyano, S., & Mimori, K. (2018). Understanding intratumor heterogeneity by combining genome analysis and mathematical modeling. *Cancer science*, 109(4), 884-892.  
 [8] <https://www.nature.com/news/cancer-therapy-an-evolved-approach-1.19746>

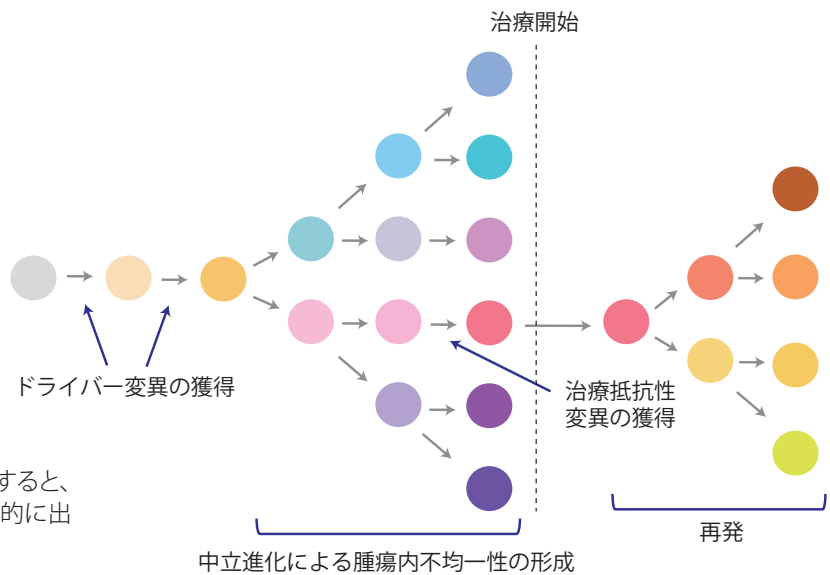


図7: 治療抵抗性獲得のしくみ

無数の異なる中立変異を有するクローンが存在すると、その中から治療抵抗性を持ったクローンが必然的に出現して再発に至る。

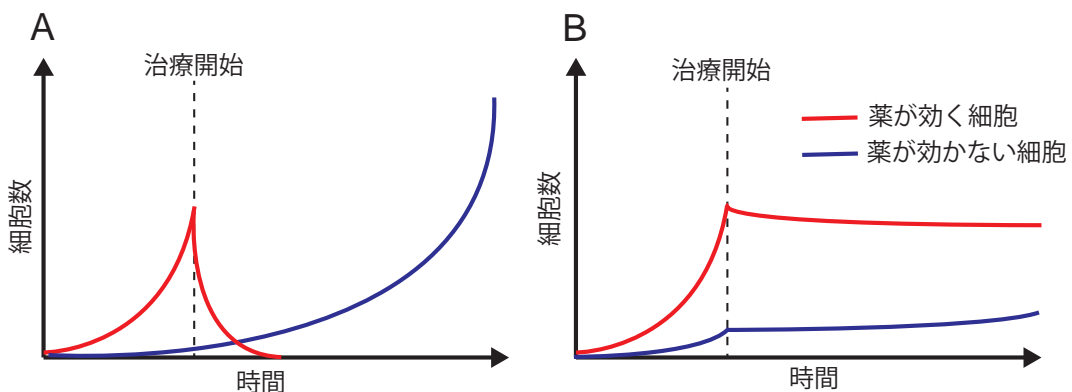


図8: 治療抵抗性獲得に対する戦略

薬が効く細胞を全て殺せば薬が効かない細胞が自由に増えてしまうので (A)、ある程度残して薬が効かない細胞と競合させることで、再発を遅らすことができる可能性がある (B)。

## Information

News & Events

### SS研HPCフォーラム2018 次世代HPCの息吹

- 日程：8月28日(火)  
フォーラム10:30～17:05 懇親会17:25～18:45
- 場所：汐留シティセンター 24階 大会議室  
富士通株式会社プレゼンテーションルーム (東京都港区)
- 参加費：無料 (懇親会500円)
- 参加申し込み：8月上旬受付開始予定
- Web：  
<http://www.sskn.gr.jp/MAINSITE/event/2018/20180828-hpcf/index.html>

#### ○ 講演

- 演者：宮野 悟 (東京大学医科学研究所)
- タイトル：スパコンで暴き出すがんの複雑さ

### 17th European Conference on Computational Biology (ECCB2018)

- Date：September 8 (Sat) - 12 (Wed)
- Venue：Athens, Greece
- Web：<http://eccb18.org/>

#### ○ Poster

- Author：Satoshi Ito, Masaaki Yadome, and Satoru Miyano  
(The University of Tokyo)
- Title：Virtual Grid Engine: A pseudo grid engine on MPI parallel environments

### 2018年度統計関連学会連合大会

- 日程：9月9日(日)～13日(木)
- 場所：中央大学・後楽園キャンパス (東京都文京区)
- 主催：応用統計学会、日本計算機統計学会、日本軽量生物学会、日本行動計量学会、日本統計学会、日本分類学会

#### ○ 企画セッション

- 日程：9月11日(火) 15:30～17:30
- 場所：B会場 (5533教室)
- タイトル：  
(22)データ科学から迫る生命医科学研究のフロンティア

- オーガナイザー：島村 徹平 (名古屋大学医学部)、新井田 厚司 (東京大学医科学研究所)、白石 友一 (国立がん研究センター研究所)
- Web：[http://www.jfssa.jp/taikai/2018/program\\_3.html](http://www.jfssa.jp/taikai/2018/program_3.html)

### 生命医薬情報学連合大会 (IIBMP2018)

- 日程：9月19日(水)～21日(金)
- 場所：荘銀タクト鶴岡、鶴岡アートフォーラム (山形県鶴岡市)
- Web：<http://iibmp2018.org/wp/>
- BoFセッション
- 日程：9月20日(木) 15:00～16:30
- 場所：大ホール
- タイトル：医学・情報学融合研究が牽引するメディカルイノベーション
- 座長：島村 徹平 (名古屋大学医学部)、新井田 厚司 (東京大学医科学研究所)

### 平成30年度一般向けスパコンセミナー もっと知りたくなるスパコンの世界 ～シミュレーションが創り、支える私たちの健康・医療～

- 日程：9月30日(日)  
見学会12:30～14:00 講演会14:15～16:15
- 場所：神戸大学 先端融合研究環 統合研究拠点コンベンションホール (兵庫県神戸市)
- 参加費：無料
- 主催：兵庫県、神戸市、計算科学振興財団
- 定員：350名 (先着順) ※要事前申し込み
- 参加申し込み：  
[https://www.fbri-kobe.org/kbic/event/detail.php?event\\_id=135](https://www.fbri-kobe.org/kbic/event/detail.php?event_id=135)
- 詳細は、参加申し込みホームページをご覧ください。

#### ○ 講演

- 演者：久田 俊明 (株式会社UT-Heart研究所)
- タイトル：  
スパコンの中で拍動するあなたの心臓～心臓シミュレータUT-Heartによるテーラーメイド医療～



#### 文部科学省 ポスト「京」開発事業

重点的に取り組むべき社会的・科学的課題に関するアプリケーション開発・研究開発  
重点課題2 個別化・予防医療を支援する統合計算生命科学

Integrated Computational Life Science to Support Personalized and Preventive Medicine

#### ■ 問い合わせ先

国立大学法人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  
ポスト「京」重点課題2 個別化・予防医療を支援する統合計算生命科学 事務局

〒108-8639 東京都港区白金台 4-6-1 TEL: 03-5449-5615 FAX: 03-5449-5442  
E-mail: [icls-office@hgc.jp](mailto:icls-office@hgc.jp) URL: <http://postk.hgc.jp/>



ポスト「京」重点課題は、国家基盤技術としてスーパーコンピュータ「京」の後継機となるポスト「京」を活用し、国家的に解決を目指す社会的・科学的課題に戦略的に取り組み、世界を先導する成果の創出を目指す文部科学省の事業です。重点課題2「個別化・予防医療を支援する統合計算生命科学」は、東京大学医科学研究所を代表機関として、ポスト「京」によって初めて実現できる「情報の技術」、「物理の原理の応用」、そして「ビッグデータの活用」により、病態の理解と効果的な治療の探索法の研究を行い、その成果を個別化・予防医療へ返す支援基盤となる統合計算生命科学を確立することを目的としています。