

ポスト「京」(スーパーコンピュータ「富岳」) 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

Vol. 17

NEWS LETTER

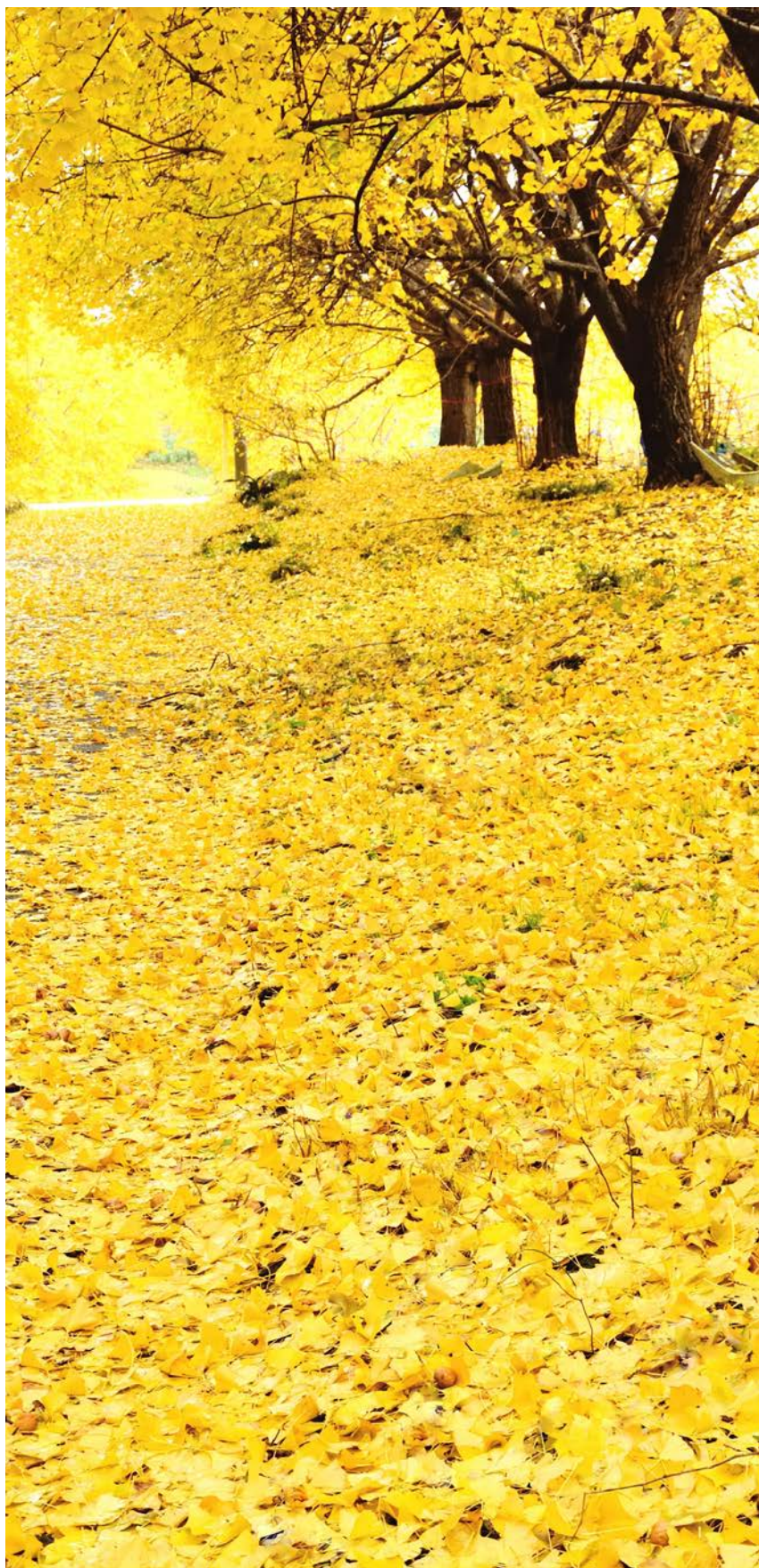
Contents

- Special Topic
重点課題2サブ課題Aの軌跡、5年間を振り返る
- 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



■ Special Topic

Subtheme **A**

重点課題 2 サブ課題 A の軌跡、5 年間を振り返る

サブ課題 A

京都大学大学院医学研究科
東京大学医科学研究所

小川 誠司
宮野 悟
(左から)



スーパーコンピュータ「富岳」を用いて重点的に取り組むべき社会的・科学的課題に向けたアプリケーション開発を目的として、2015年からスタートしたポスト「京」（「富岳」）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」（以下、重点課題 2）が2020年3月に終了します。

重点課題2のサブ課題 A は、スーパーコンピュータとビッグデータを駆使して一人ひとりに的確ながんの個別化予防・治療戦略を創出することを目指しています。がんは、多くの先進諸国で死因の1位を占めており、日本では一生において二人に一人が直面する疾患です。

進化する複雑なゲノム異常データとの勝負であるがんに、サブ課題 A の宮野悟課題責任者（東京大学医科学研究所）と小川誠司分担責任者（京都大学大学院医学研究科）が一丸となって研究を実施するなかで、どのような成果が得られているのでしょうか。

今回のニュースレター 特集「Special Topic」では、サブ課題 A の主な成果を通して、5年の軌跡を振り返ります。

より大規模に鮮明にがんのシステム異常を検出する

重点課題 2 サブ課題 A が開発しているアプリケーションは Genomon^{*1} です。これはスーパーコンピュータを用いて次世代シーケンサーの出力する大量の遺伝子データを処理（マッピング）し、統計解析などを駆使してがん化に強い関連性のある遺伝子変異を高精度・高感度に検出することが可能なアプリケーションです。小川教授のイニシアティブのもと、スーパーコンピュータ「京」上で Genomon を動かし世界に類のない大規模なスケールでゲノムデータを解析することにより、世界トップ研究の成果を少しずつですが創出することができました。こ

の Newsletter Vol.17「重点課題 2 サブ課題 A の軌跡、5 年間を振り返る」では、それらの成果をご紹介します。

免疫系を回避するがん細胞のゲノム異常の解明

免疫チェックポイント^{*2}分子の一つである PD-L1 タンパク質を作る情報を持った（タンパク質をコードする）PD-L1 遺伝子の発現調整で、3' 非翻訳領域（遺伝子のタンパク質に翻訳されない領域の後半部分）が重要な役割を担っていることを明らかにしました。また、正常な 3' 非翻訳領域が失われた結果 PD-L1 の遺伝子発現が上昇したがん細胞は、免疫による監視を回避して増殖することを、世界で

初めて明らかにした研究です。

この成果に先立ち、「京」による Genomon での大規模データ解析で成人 T 細胞白血病リンパ腫 49 例の全ゲノム、43 例の RNA シーケンズデータからそのゲノム変異の全貌を明らかにしてしました（Nature Genetics 2015）。これまでは中程度サイズ（10～1,000 塩基）の挿入や欠失の構造変異を検出するソフトウェアが確立されておらず、次世代シーケンサーを用いたゲノムデータ解析の盲点となっていました。しかし Genomon はその盲点である数十～数百塩基の構造異常の高精度検出を数理的方法で可能にしており、その結果 PD-L1 遺伝子の 3' 非翻訳領域の構造異常が見つかりました。同時にこの構造異常をもつ検体の PD-L1 遺

■ Special Topic

Subtheme A

伝子の発現が高いことも観察されました。そこで33種類のがん種10,210検体のデータをGenomonで解析すると、複数のがん種において同じ構造異常が見いだされ(図1)、マウスを使ったゲノム編集でその機能解析を行い、がんにおける新たな免疫回避のメカニズムを解明しました。この成果は2016年5月のNature誌に掲載されました。この成果についてはNewsletter Vol.2「大規模データ解析で生命システムを明らかに」で詳しく取りあげています。

急性白血病を起こす2ステップの遺伝子異常のパターンを発見

世界で初めて2,250例の骨髄異形成症候群と二次性急性白血病患者さんに対して、次世代シーケンサーや「京」を用いて大規模な遺伝子解析を行い、慢性の骨髄異形成症候群から急性白血病を起こす遺伝子異常を詳細に調べました。これまでになく多数例において網羅的な遺伝子解析を時系列で複数回行った結果、

骨髄異形成症候群から二次性急性白血病に進行するときには、2つのタイプの遺伝子異常パターンがあることを発見しました。

骨髄異形成症候群から二次性急性白血病へ進行する際に、2つのタイプのゲノム異常が細胞に蓄積し異常クローン^{※4,5}が増大すること、そして、低リスク(白血病リスクの低い状態)の骨髄異形成症候群に「タイプ2ゲノム異常^{※6}」が起こると高リスクに移行し、「タイプ1ゲノム異常^{※7}」が起こると急性白血病が起こるこ

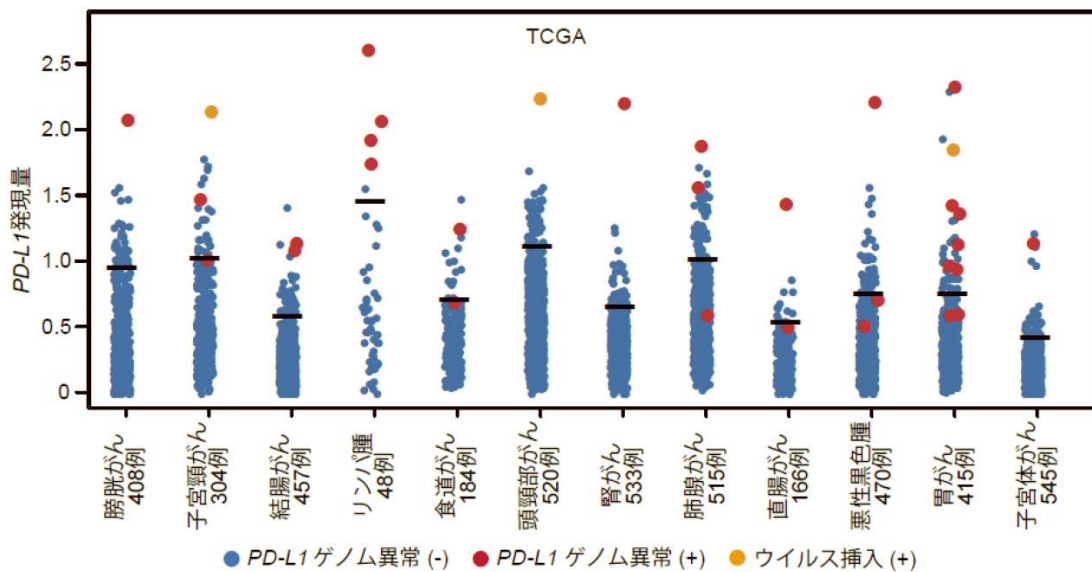


図1: 33のがん種10,210検体(がんゲノムアトラス^{※3})における遺伝子変異解析
2種類のがん種を含む、32症例でPD-L1遺伝子の3'非翻訳領域の異常(赤色と黄色の●印)が生じていた。とくにB細胞悪性リンパ腫(図ではリンパ腫と表記)と胃がんで高頻度にみられる。

用語解説

※1 Genomon: がんの変異を高精度・網羅的に暴き出すソフトウェア。機械学習などのAI技術が投入されている重点課題2のターゲットアプリ。

※2 免疫チェックポイント: 免疫の恒常性を保つために、自己に対する免疫応答を抑制するとともに、過剰な免疫反応を抑制する分子群。PD-L1は、がん細胞に発現する「免疫チェックポイント」を構成する一群の分子の一つで、がん細胞を攻撃する細胞障害性T細胞が発現するPD-1分子に作用して、細胞障害性T細胞を抑制する。

※3 がんゲノムアトラス (The Cancer Genome Atlas: TCGA):

米国立がん研究所 (NCI) と米国立ヒトゲノム研究所 (NHGRI) が共同で支援している2005年から始まった米国のがんゲノム解読プロジェクト。さまざまな組織におけるがんのゲノム、エピゲノム、トランスクリプトームなどのデータを集約し公開している。

■ Special Topic

Subtheme **A**

とが分かりました。また *SF3B1* 変異陽性症例は、タイプ1とタイプ2異常を持たない場合、ほとんど白血病を発症しないことも観察されました。

これらの遺伝子異常は、白血病の進行を予測するマーカーとしてスクリーニングに用いられることが期待され、発見や治療をより早く始められ、高い治療効果が得られる可能性があります。この成果は2016年12月のNature Genetics誌に掲載されました。

加齢に伴う正常組織の遺伝子異常とがん化のメカニズムの解明

実体顕微鏡を組み入れた0.5mm微小サンプル解析システムを開発し、食道がんと食道正常上皮組織を1mmメッシュ間隔で数百箇所の切片を採取して、ゲノムシーケンスを行いました(図3)。その結果は、食道がんに頻繁に見られる遺伝子変異が乳児期から生じ、加齢とともに増加して70歳を超える高齢者では全食

道面積の40~80%が、がん遺伝子の変異をもった細胞に置き換わるという衝撃的なものでした。とくに、飲酒や喫煙をしている人は遺伝子異常が高頻度に認められました。これは、がんがなぜ高齢者に生じ、飲酒や喫煙によって促進されるのかを解明した発見となりました。がんが生じる初期のメカニズムを解明する突破口を開く意義のある成果です。この成果は2019年1月のNatureに掲載され、また研究成果の意義について解説記事も同時に出ました。

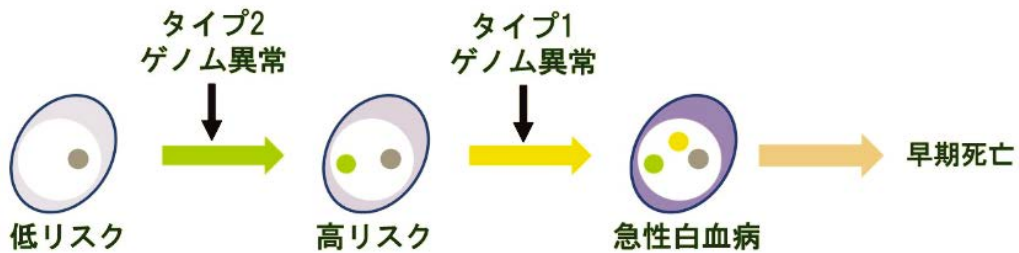


図2: 骨髄異形成症候群から急性白血病進行への2ステップ

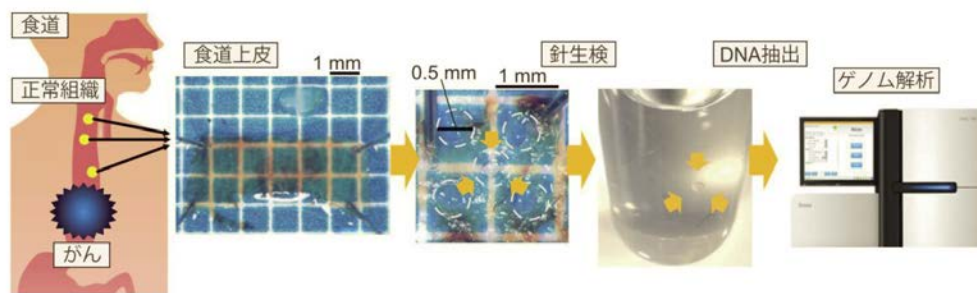


図3: 正常組織での遺伝子変異の解析手順

正常組織での発がん前に生じるクローン拡大を解析するため、食道上皮の生検検体から、直径0.5mmの非常に小さい生検針を用いて食道上皮組織を採取しゲノム解析を行った。

用語解説

- ※4 クローン: 一つの細胞由来の均一な遺伝情報をもつ細胞集団。
- ※5 がんのクローン進化: がんの起源となる細胞系列に生ずる遺伝子変異によって、さまざまに変化する環境に適応した、あるいは環境下で生存に有利な細胞群とそれらの子孫の細胞群を含む、極めて多様な細胞集団に進化すること
- ※6 タイプ2ゲノム異常: *TP53, GATA2, KRAS, RUNX1, STAG2, ASXL1, ZRSR2, TET2* 遺伝子変異
- ※7 タイプ1ゲノム異常: *FLT3, PTPN11, WT1, IDH1, NPM1, IDH2, NRAS* 遺伝子変異

■ Special Topic

Subtheme **A**

慢性活動性EBウイルス感染症の原因解明

原因不明の難病である慢性活動型EBウイルス^{※8}感染症の遺伝子解析を行い、感染症の原因を解明しました。またこの感染症が血液がんの一種であることを解明しました。

EBウイルスは、感染しても多くの場合は命にかかわる症状は出ませんが、ごく一部の人に血液がんやその他の難病を引き起こすことが知られています。しかしなぜ重い病気を発症するのかは明らかになっていませんでした。

この感染症の患者さん 80 人の遺伝子解析から、ウイルスに感染した血液細胞の遺伝子に、がんに似た遺伝子変異が生じていることが分かりました。さらに病気が進行し血液がんのような状態になった一人の患者さんについて遺伝子解析をした結果、血液がんで見られる *DDX3X* 遺伝子変異が新たに生じていました。前述の *PD-L1* 遺伝子の変異も見られました。病気の進行に伴い、新たな遺伝子変異が生じる過程はがんの特徴的な性質の一つであり、慢性活動型EBウイルス感染症は、がんの振る舞いをす

た(図4)。

続いて、ヒトの遺伝子(約 20,000 遺伝子)とEBウイルスの遺伝子(約 80 遺伝子)の全解析により、慢性活動型EBウイルス感染症に関わるEBウイルスは、いくつかの遺伝子を失っており(図5)、その結果、EBウイルスが異常に活性化して、がんを発症させることがわかりました。遺伝子が欠損したEBウイルスは他の血液がんでも見つかり、同じ仕組みでさまざまな血液がんを発症させることも分かりました。この成果は2019年1月の *Nature Microbiology* 誌に掲載されました。

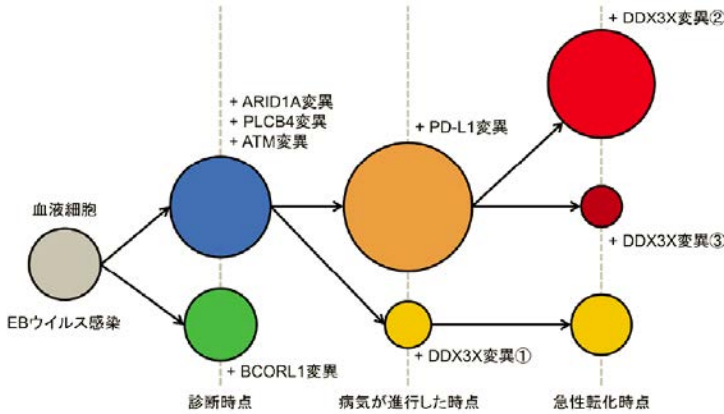


図4: EBウイルス感染細胞の遺伝子変異(一例)
病気が進行するにつれ、EBウイルス感染した血液細胞はさまざまな遺伝子変異を生じていた。

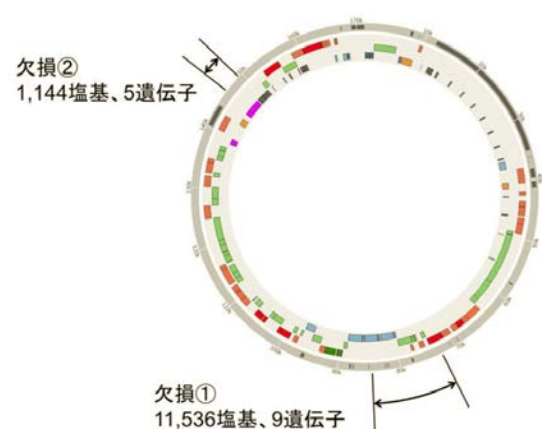


図5: EBウイルスのゲノムで見つかった大きな欠損(一例)
EBウイルスのゲノムは環状で、約80の遺伝子が存在する。この一例では、欠損①の領域では9つの遺伝子が欠損しており、その中にウイルス粒子を作るために必要な遺伝子が含まれていた。欠損②の領域では、ウイルスが潜伏感染するために必要な遺伝子が含まれていた。

用語解説

※8 EBウイルス (Epstein-Barr ウィルス): ヒトに感染するウィルス。世界中に蔓延しており、95%の人が一度は感染する。通常は大きな害のないウィルスだが、一部の人には血液がんやその他難病を発症させる。

■ Special Topic

Subtheme **A**

肺がんの増殖に関わる新規長鎖ノンコーディングRNAの発見

約 8,000 検体の超大規模遺伝子発現データを、「京」でネットワーク解析を行いました。その結果、2 万以上あるノンコーディング RNA から MYC を操っている長鎖ノンコーディング RNA を世界で初めて突き止めました。

MYC は代表的ながん遺伝子で、肺がん細胞の増殖に必須な転写因子です。「京」を用いて肺がんのがん組織での遺伝子の発現情報を解析し、MYC の発現制御を通じて肺がん細胞の増殖を維持する長鎖ノンコーディング RNA を発見しました。

MYMLR (MYC-modulating lncRNA) と命名したこの長鎖ノンコーディング RNA を抑制することにより、肺がん細胞の増殖を抑制することができます。

この研究では、MYC を発現制御する MYMLR の肺がん細胞の増殖における重要性とその機能の詳細を明らかにしました。この成果により、MYMLR を標的とする革新的な治療法の開発が進むと期待されています。2019 年 7 月の The EMBO Journal 誌に掲載されました。この MYMLR の発見は、サブ課題 A の協力研究者である愛知県がんセンター高橋隆総長の牽引により、HPCI 戦略プログラム 分野 1 (2011 年 4 月～2016 年 3 月) の時代から協働で進めてきた研究の成果です。

「富岳」を活用して、データ科学を躍進させる

以上にあげた 5 つの成果から見えてくるのは、がんゲノムの解読が、本質的にゲノムの疾患であるがんの解明にとっても有効な手段であること、そして医療や健康に有用な情報を抽出するためには、大規模データ解析を可能にするスパコンとアプリケーションの活用が不可欠ということです。

一般的なスパコンでは、がんゲノムのシーケンスデータを解析するには 1 検体あたり 24 時間程度と見積もられますが、「京」をフル活用した場合には同じ時間で 100 検体のデータ解析が可能になりました。さらに「富岳」では 1,000 検体を

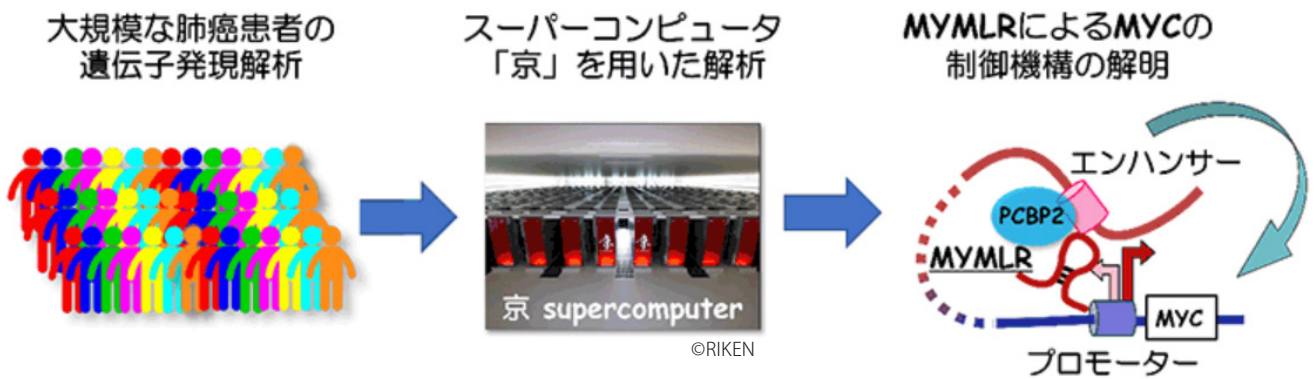


図6: 「京」によるMYMLRの発見
 (右図) MYMLRは、協調して働くタンパク質PCBP2と結合してMYC遺伝子近傍のゲノムDNAを湾曲させ、遺伝子の発現制御を担うエンハンサー^{※9}をプロモーター^{※10}に近接させて、MYCの転写を制御している。

用語解説

- ※9 **エンハンサー**: 転写調節因子 (活性化因子と抑制因子) が結合するDNAの調節配列。プロモーターに働きかけて遺伝子の発現を調節するゲノム領域。
- ※10 **プロモーター**: 遺伝子の転写が始まるDNAの塩基配列。遺伝子上流に位置し、遺伝子発現の開始や発現量の制御にかかわるゲノム領域。

■ Special Topic

Subtheme A

大幅に超えるデータ解析が可能になります。スパコンに求められるのは速さだけではなく、どれ程精緻に解析できるか質も大切です。一人ひとりのがんを捉えるためには、タンパク質をコードしていない部分を含めた全ゲノム解析に基づいて、1%以下の頻度の変異を網羅的に見出すことが必要です。1%以下の頻度の変異を網羅的に見つけるためには約2,000検体を解析することが必要で[1]、例えば50種のがんでは、「京」で5,000日を要します。「富岳」では50日程度で可能になり、今まで見逃されていたがんに関係する遺伝子変異も見つけられるようになります。

がん免疫やがん発生のメカニズムに関して、網羅的にがんのゲノム異常の探索

を行うことにより、これまで未踏だったがんの分子病態の理解を飛躍的に加速しました。世界でのゲノム解析競争が激化する中、ほんの僅かの差で成果を上げることができたのは、スパコンパワーだけでなく大規模データ解析を考慮したスパコンの運用方法にもありました。「京」での運用経験を活かして、「富岳」ではスパコンのシステムをアプリケーションと協調的に開発する「コデザイン」体制が取り入れられています。さらなる社会的・科学的成果を創出に向けて、運用面でも前進が期待されます。プロジェクト終了まであと残り約5カ月ですが、「富岳」による飛躍的なデータ科学へと繋ぐべく、将来を見据え引き続き粛々と研究を進めていきます。

<参考文献>

[1] Lawrence MS, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 505(7484):495-501, 2014.

<引用文献>

Newsletter Vol.2「大規模データ解析で生命システムを明らかに」
http://postk.hgc.jp/_media/library/newsletter2.pdf

Newsletter Vol.6「がんの遺伝子解析とスーパーコンピューティング」
http://postk.hgc.jp/_media/library/newsletter6.pdf

Peter Walter他, Essential 細胞生物学, 中村桂子・松原謙一(監訳)他, 南江堂, 2016.

【参考資料】今回登場した文献

プレスリリース	論文 / doi	掲載
がん細胞が免疫から逃れるメカニズムの解明 —免疫チェックポイント阻害剤の効果予測への応用に期待—	Aberrant <i>PD-L1</i> expression through 3'-UTR disruption in multiple cancers. doi: 10.1038/nature18294.	Nature 2016.5
骨髄異形成症候群におけるクローン進化の解明 —急性白血病を起こす2ステップの遺伝子異常のパターンを発見—	Dynamics of clonal evolution in myelodysplastic syndromes doi: 10.1038/ng.3742.	Nature Genetics 2016.12
加齢に伴う正常組織の遺伝子異常とがん化のメカニズムを解明 —食道上皮は加齢に伴いがん遺伝子の変異を獲得した細胞で再構築される—	Age-related remodelling of oesophageal epithelia by mutated cancer drivers doi: 10.1038/s41586-018-0811-x.	Nature 2019.1
慢性活動性 EB ウイルス感染症の原因と、身近なウイルスががんを引き起こす仕組みを解明	Defective Epstein Barr-virus in chronic active infection and related hematological malignancy doi: 10.1038/s41564-018-0334-0	Nature Microbiology 2019.1
スーパーコンピュータ「京」から迫る肺がんの分子機構 —肺がんの増殖に関わる新規長鎖ノンコーディング RNA の発見—	Divergent lncRNA MYMLR regulates MYC by eliciting DNA looping and promoter-enhancer interaction doi: 10.15252/embj.201798441	The EMBO Journal 2019.7

2019年度 理化学研究所 計算科学研究センター 一般公開

- 日時: 11月9日(土) 10:00 ~ 16:30 (入場は16:00まで)
- 場所: 計算科学研究センター (兵庫県神戸市)
- 入場料: 無料
- 問い合わせ先:

理化学研究所 計算科学研究推進室 広報グループ
TEL:078-940-5555 (平日10:00 ~ 17:00)

- WEB:
<https://www.r-ccs.riken.jp/jp/outreach/library/event/openhouse2019.html>

○ ポスター展示

- 場所: 計算科学研究センター 1F
- 時間: 10:00 ~ 16:30
- 展示ブース名: スパコン「富岳」でできること、わかること

スパコン「富岳」を活用して社会や私たちの生活にどのように役立てられるのかについて、重点課題の研究機関や大学が紹介します。

- ・「富岳」が可能にする創薬・医療の未来 (重点課題1)
- ・私だけの医療で健やかに生きる社会を (重点課題2)
- ・スパコンのなかの仮想地球で探る 将来の気象や地球環境 (重点課題4)
- ・働く電子・分子を、CGとシミュレーションで見てみよう! 一次世代太陽電池、人工光合成、次世代リチウム電池、CO₂分離・回収 (重点課題5)
- ・革新的クリーンエネルギーシステムの実用化 (重点課題6)
- ・つよい磁石をつくる (重点課題7)
- ・「富岳」で変わるものづくり (重点課題8)

一般公開では、ほかにもスパコン「富岳」を設置するための工事の様子を見学者ホールから見るすることができます。また、「富岳」の試作機 (CMU・CPUパッケージ) の展示や、ミニ講演会なども開催予定です。どうぞお立ち寄りください。

第73回 国立病院総合医学会 令和における国立医療の挑戦~明日は変えられる~

- 日時: 11月8日(金) ~ 9日(土)
- 場所: 名古屋国際会議場 (愛知県名古屋市)
- 運営事務局: (株)コングレ中部支社内 73nms@congre.co.jp
- WEB: <http://www.congre.co.jp/73nms/index.html>

○ 特別講演1

- 時間: 11月8日(金) 10:00 ~ 11:00
- 場所: 第1会場
(名古屋国際会議場 1号館2階センチュリーホール)
- 座長: 長谷川好規 (NHO名古屋医療センター 院長)
- 講演者: 宮野 悟
- 講演タイトル: 人工知能とスーパーコンピュータが支えるがんゲノム研究と医療

化学コミュニケーションのフロンティア 第6回公開シンポジウム

- 日時: 12月9日(月) ~ 10日(火)
- 場所: 慶応義塾大学日吉キャンパス 協生館
藤原洋記念ホール (神奈川県横浜市)
- 入場料: 無料
- 問い合わせ先: 京都大学大学院薬学研究科 システムケモセラピー (制御分子学)分野内 新学術領域研究「化学コミュニケーションのフロンティア」総括班事務局
fr_chemcomm@pharm.kyoto-u.ac.jp
- WEB: http://www.pharm.kyoto-u.ac.jp/fr_chemcomm/event/event06.html

○ 特別講演

- 時間: 12月9日(月) 17:05 ~ 17:50
- 講演者: 宮野 悟
- 講演タイトル: 人工知能に支援されたSociety 5.0時代のがんゲノム医療と病院システム



文部科学省 ポスト「京」(スーパーコンピュータ「富岳」) 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 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」の後継機となるスーパーコンピュータ「富岳」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、「富岳」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。