

東京大学医科学研究所 これからのがん治療にはAIと スパコンが欠かせない

東京大学医科学研究所（略称東大医科研）は、日本では唯一の研究附属病院を有する国内最大規模の医科学の研究所であり、がん、感染症やその他の難治疾患を対象にした最先端の研究と医療が進められている。今般、同研究所ヒトゲノム解析センター長である宮野悟教授に、人工知能（AI）を活用した最先端の個別化ゲノム医療や、米国での医療関係者・政府・民間企業の連携による医療などについて、インタビューをさせていただく機会を得た（取材・文責：JOI総務部 佐々木永市、内野美佳）。

スパコンによりがんの黒幕を暴き出し、AIにより有効な治療方法や薬剤を探し当てる

東大医科研では、2015年7月1日より、IBM社の人工知能 Watson Genomic Analytics（現在、Watson for Genomicsに改称）（以下、ワトソン）を導入し、がんの個別化ゲノム医療を開始している。導入当時は、北米を除いて世界で初めてであった。抗がん剤治療により病状の改善がみられなかった白血病患者のゲノム情報をワトソンで分析したところ、わずか10分の解析で特殊な白血病であることがわかり、治療法を変えた結果、60代女性患者の命が救われたことは、「人工知能が人の命を救った」としてNHKなどのメディアでも多く取り上げられた。

がんは、ゲノム異常の病気であり^注、ゲノム変異が蓄積することで、悪性の細胞（がん細胞）になってしまう病気である。しかしながら、すべての遺伝子異常（遺伝子変異）が、がん発症の原因になるわけではない。がんになるために重要な役割を果たしている遺伝子変異（原因遺伝子）があり、この原因遺伝子は、がんの種類やタイプにより異なる。がんの個別化ゲノム医療では、ゲノム解析等により、がんの原因となっている変異を特定し、個々の患者にあった治療法を見出すことが求められている。

注：ゲノム・DNA・遺伝子の意味

ヒトの遺伝情報はDNA（デオキシリボ核酸）と呼ばれる物質に記されている。DNAは、アデニン（A）、グアニン（G）、シトシン（C）、チミン（T）の4種類の塩基という構成要素（文字）が2つずつの対でつながってできた鎖であり、人の細胞にはこの塩基の対が約30億個結合したDNAが2セットある。DNAの1セット分（30億個分）をゲノム、DNAの中の主にタンパク質の合成にかかわる情報部分を遺伝子（約23000～24000種類）と呼ぶ。ゲノムの中で、遺伝子として働いている部分は全体の約1.5%である（古川洋一著『変わる遺伝子医療』（ポプラ社）より）。

個別化ゲノム医療のプロセス

なぜ個別化ゲノム医療では、スパコンや人工知能（AI）が必要となるのであろうか。

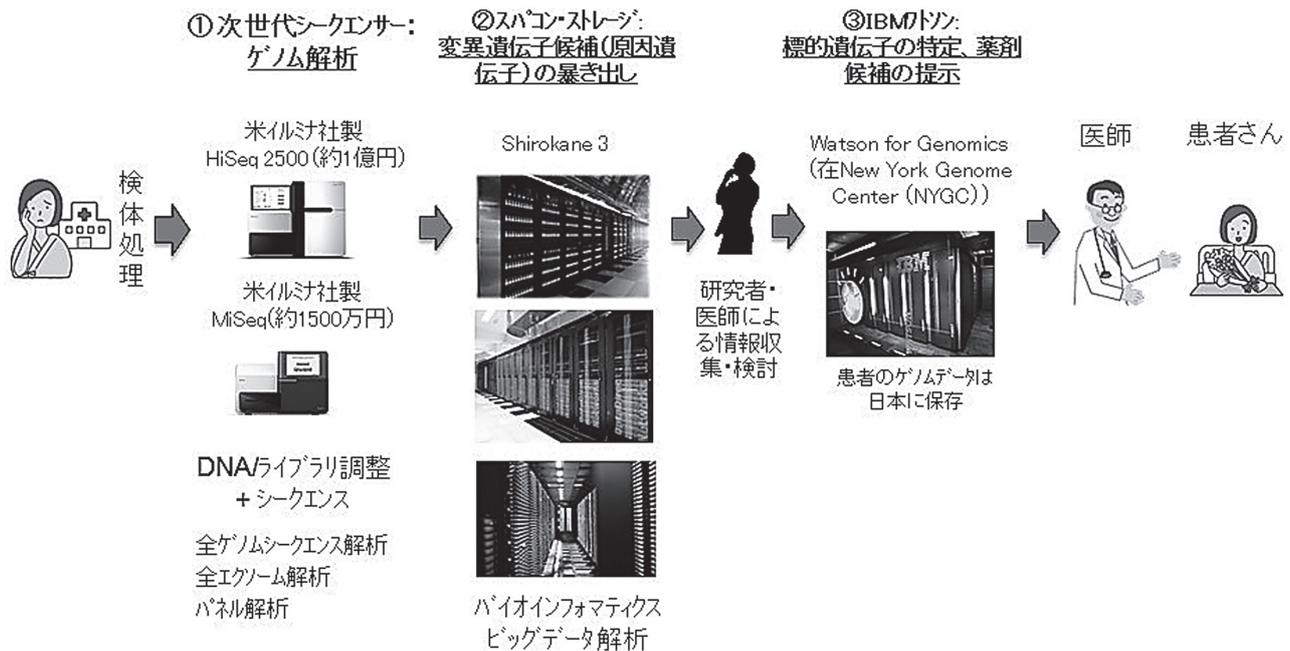
東大医科研の個別化ゲノム医療では、①次世代シーケンサーと呼ばれるゲノム解析装置にてゲノム解析（DNA情報の読み取り）を行い、②読み取られた膨大なDNA情報を、スパコンで解析し、がん細胞変異遺伝子候補を特定（がんの原因遺伝子の暴き出し）、さらに、③人工知能ワトソン、ならびに専門家によりがんの原因遺伝子と膨大な研究論文・薬特許情報・パスウェイ情報等を統合し、標的遺伝子（がん細胞を抑えるための標的対象となる遺伝子）とその薬剤候補を特定するというプロセスが取られている。順を追って見ていきたい（図1参照）。

①次世代シーケンサーによるゲノム解析

次世代シーケンサーは、A、T、C、Gの文字でつづられるDNA情報をコンピューターで読めるように取り出すゲノム解析装置である。東大医科研で利用している次世代シーケンサーは米イルミナ社製 HiSeq 2500（装置価格約1億円、1回で1テラベース（文字）を出力する）、MiSeq、HiSeq 500やサーモフィッシャーサイエンティフィック社製のIon Proton、Ion PGMなどであり、目的に応じて使い分けている。

ヒトゲノムのシーケンスのコストは、以前は大変高価だったが、現在では一人当たり約1000ドルまで価格が下がっており、将来は所要時間1時間以内、1万円以下の簡易シーケンスも可能なシーケンサーの登場が期待されている。これに伴い、解析・蓄積されるヒトゲノムのデータは指数的に増加しており、米国カリフォルニア大学サンタクルーズ校（UCSC）には、

図表1 東大医科研：がんの個別化ゲノム医療のプロセス



出所：東大医科学研究所

2015年当時で、毎月1ペタバイト (PB) (1PB=1000テラバイト) ペースでシーケンスデータが蓄積されるなど、2014~18年の間に、世界で次世代シーケンサーにより解析・蓄積されるヒトゲノムのデータは、2エクサバイト (1エクサバイト=1000ペタバイト=100万テラバイト) を超えるといわれている。

次世代シーケンサーから出てくる生のデータは、正常ゲノムの場合、ヒトゲノムの30億文字が印刷された書類を30部コピーしたもの (900億文字) をシュレッダーにかけて100文字ごとの断片にした、いわばジグソーパズルのピースの山のようなもの。がんのゲノムの場合には、40コピー (30億文字×40コピー=1200億文字、12億ピース) が必要となり、合わせると、実に1検体当たり合計21億ピースのパズルが出てくる。この膨大な量のパズルを解くために必要なプロセスが、次のステップのスパコンによる解析である。

②スパコン (Shirokane3) と巨大ストレージを用いたバイオメディカルビッグデータ解析による変異遺伝子候補 (原因遺伝子) の暴き出し

東大医科研ヒトゲノム解析センターでは、2015年4月よりスーパーコンピューターシステムShirokane3の一般運用を開始した。スパコンでは、次世代シーケンサーで取得した膨大なゲノム情報 (1検体当たり21

億個のジグソーパズルピース) の文字の並びをつなぎ合わせ、どの配列がゲノムのどの部分に相当するかを調べ、標準配列との違いや、変異遺伝子かどうかなどを解析する。親から受け継いだゲノム、がんを発症する鍵遺伝子の変化、環境因子による変化を受けたゲノム、それぞれの特性を解析することで、がんを発症させたシステム異常の原因遺伝子を暴き出していく。

なお、余談となるが、東大医科研ヒトゲノム解析センターでのスパコンの導入、利用に当たっては省エネが大きな課題となった。東京大学は、東京電力の発電量全体の1%の電力を使用していた。スパコンによる電力使用量は大きく、ヒトゲノム解析センターでは、間接蒸発式冷却器の導入やスパコン・ストレージ設置室の室内温度を、トラブルが出現しないギリギリの27~28℃とすることにより、電気使用量の大幅な削減を実現し、電力消費の削減に貢献してきた。

ゲノムシーケンスおよびスパコンによる解析により見えてくる変異候補は、数百から数百万に及ぶ。従来は、この変異候補を人力・人智で解析・解釈し、研究論文などと突合することで、がんを発症させている遺伝子とその変異を絞り込み、また、利用可能な薬剤情報をもとに、どの遺伝子を標的にどの薬剤を使って治療を行うかを模索してきた。これには膨大な時間がかかり、また時間をかけたとしても、原因や有効な治

療法を見出せないこともある。

米保健福祉省下の国立衛生研究所 (National Institutes of Health, NIH) のデータベース MEDLINE上には、2016年現在、2600万件の研究論文要旨が登録されており、世界中の誰もが無償で利用可能である。これは論文本体の印刷物の厚さにすると富士山の高さ (高度 4 km) を超える。論文数は指数関数的に増えており、2050年には大気圏外 (高度 100km) に達する高さになるであろう。このうち、がんおよびゲノム関連の論文は約6分の1を占めており、がんだけで年間20万件以上増加しており、人の手で新しい論文を調べることでさえも、到底無理である。また、この中には間違った論文も含まれていること、がんは、その複雑性から、がんだけの知識では原因が分からないことにも留意すべきである。そこで、人工知能の出番となる。

③ワトソンによる、変異遺伝子候補と研究論文・薬剤情報等との突合による、標的遺伝子、治療法および薬剤候補の特定

ワトソンは学習する辞書である。金融など医療以外のビジネスでも幅広く活用されているが、ゲノム医療での活用は、2014年にNew York Genome Center (NYGC) とIBMによるNYGC Watson initiativeにより開始された。研究目的で2015年7月1日に東大医科

研がワトソンを導入した際には、2000万件超の研究論文要旨 (MEDLINEデータベース) と1500万件超の薬の特許データ、がんおよびゲノム関係のがん・ゲノム関連データベース (英COSMIC (Catalogue of Somatic Mutations in Cancer)、米NIHのClinVar、National Cancer Institute Pathwayなど) をすでに学習していた。その後もNYGCや、東大医科研を含むワトソンを導入した数十の医療研究機関・医療機関等により情報を拡充、学習を進めており、ゲノム医療の基盤となっている。

東大医科研では、スパコンにより解析された変異遺伝子候補をワトソンのシステムにアップロードする。すると、ワトソンは、膨大なデータから学習した知識をもとに、がんの原因遺伝子、がんを抑えるためにターゲットとすべき標的遺伝子とその根拠の強さ、遺伝子群でのパスウェイ (遺伝子と遺伝子の関係メカニズム)、標的遺伝子に有効な薬剤候補を提示する (図2参照)。たとえば、大腸がん細胞株RKOの変異候補4237個をワトソンにアップロードすると、12個の候補遺伝子を見出し、その根拠の強さのレベル順に、有効な薬剤候補が提示された。薬剤候補は、米国で承認済の薬剤のみならず、現在臨床治験中の薬剤、さらには、他の類似疾患で承認された薬剤まで提示される。研究者・医師は、ワトソンから提示された内容を確認し、合理性をチェックして、薬剤候補の検討を行うこととなる。

図表2 Watsonが提示する3つの薬剤カテゴリー

Relevance	Alteration	Approved for Colorectal Carcinoma	Investigational for Colorectal Carcinoma	Off Label
■■■■	BRAF V600E	Regorafenib	MLN2480 Vemurafenib Cobimetinib(GDC-0973 XL518) PD0325901 Binimetinib(MEK162) Trametinib	Dabrafenib
■■■■	PIK3CA H1047R	-	Buparlisib(BKM150) Apatinib(BYL719) MLN1117	-
■■■■	BRCA1 O435Y	-	Talazoparib(BMN 673) Olaparib(AZO-2281)	-
■■■■	ATM O2913G	-	Talazoparib(BMN 673) Olaparib(AZO-2281) AZD6738	-
■■■■	RPTOR A245T	-	AZD2014 INK128 Sunitinib Everolimus Temozolomide	-
■■■■	IGF1R M122I	-	Ganitumab(AMG 479)	-
■■■■	DDR1 A59G	-	Imatinib Dasatinib	Nilotinib
■■■■	AURKA L222P	-	Afisertib(MLN8237)	-

変異の入っている標的遺伝子(左)と変異体の名称(右)

大腸がんに対して米国で承認されている薬剤 (注: 日本では未承認)

大腸がんに対して米国で現在臨床治験中の薬剤

オフ・ラベル: 他のがんについて米国で承認されている薬剤

出所: 東大医科学研究所



インタビューにご対応いただいた東京大学医科学研究所
ヒトゲノム解析センター長 宮野教授

個別化ゲノム医療の最先端を行く米国

30億文字の情報で構成されるヒトゲノムの解析は、1990年から米国を中心とした国際的な協力関係のもと「ヒトゲノム解析計画」が進められ、2003年に完了した。

米国政府はヒトゲノム解析完了がゴールではなく、ヒトゲノムを医療に応用するための研究が必要として、その後も膨大な予算を投入している。2013年1月、米NIHは、大量のバイオメディカルデータの解析がボトルネックとなるとの認識のもと、Associate Director for Data Scienceを設置、スパコンや大規模ストレージなどのリソースの整備とデータ解析の人材育成を開始した。Big Data to Knowledge (BD2K) Initiativeとして、2015年までに全米12箇所の大学等にBD2Kセンターを設置、ゲノム医療に応えられる人材の育成および研究を行う体制・環境を充実させている。2015年1月20日、オバマ前大統領は一般教書演説の中でPrecision Medicine Initiativeを発表、同月末には100万人以上の遺伝子情報および健康医療情報を集め、その解析により病気の原因を解明する計画を発表している。さらに、直近2016年12月、The 21st Century Cures Act (Cures) が米上院を圧倒的賛成多数で通過し、オバマ前大統領が署名した。この法律は、ワトソンをはじめとしたClinical Decision Support (CDS) を含む健康関連ソフトウェアを活用して個別化ゲノム医療を進めていくうえでの画期的な一歩となった。CDSでは、その機能が、医療機器・IVDからの画像データ・信号の入手、加工、分析を行うものではない範囲において、(i) 患者情報や文献・ガイダンスの情報を表示、

分析、印刷するもので、(ii) 医療従事者に予防・診断・治療に関するリコメンデーションを提示するもので、かつ、(iii) 上記のリコメンデーションの根拠を医療従事者が独自にレビューできるもの、となっている。つまり、医療従事者が医療上の判断を下す際に、最初から当該推奨を当てにすることのないようにとの制限がある。

これは東大医科研でワトソンを研究利用している要件に合致したものとなっていた。米国ではこのような法律を迅速に整備してICTの参入が促進されている。

米国では、IBMをはじめとして多くの民間企業が個別化ゲノム医療をビジネス機会ととらえ投資を進めており（IBMはワトソン関連で約4000億円を投資）、新たなゲノム医療関連企業が続々と登場するなど、政府・民間・医師会の協働で個別化ゲノム医療の実施体制が急速に整備されている。

日本のゲノム医療：失われた10年

一方、日本では、2003年のヒトゲノム解析計画完了をもってゲノム研究は終わりという認識があり、ゲノム関係の研究予算は削減された。ここ数年になってようやく、ゲノム医療への注目が上がりつつあるが、失われた10年は大きい。

ゲノム情報の蓄積は、日本最大のバイオバンクジャパンで20万人分であり、また、東大医科研では100万検体分の保存ストレージを整備している。これに対し、英国オックスフォード大学が参画する、中国・北京に設立されたChina Oxford Centre for International Health Researchでは国内約1万箇所の病院から検体と臨床情報を収集するシステムを稼働、2016年7月現在で600万検体を収集済、2017年には900万検体を収集予定であり、1800万検体分のストレージ施設を有し、かつ、ここでの研究成果は英国のトップジャーナルを通じ全世界で紹介される枠組みも用意されるなど、日本とは桁の違う検体とデータの集積体制の構築、それによる研究開発が行われている。

また、米国のCures Actによる規制緩和に対し、日本の改定薬事法ではそのようになっているようには見えない。

上述のとおり、アメリカでは個別化ゲノム医療をビジネス機会ととらえ、官民学が連携して取り組みを進めており、日本のはるか先を進んでいる。今後、日本においても、同様の取り組みが進むことを期待したい。

