

医療健康分野のビッグデータ活用研究会
報告書 vol.1

2016年7月

日本製薬工業協会
医薬産業政策研究所

無断転載を禁ずる

医療健康分野のビッグデータ活用研究会 報告書 vol.1

森田 正実

(医薬産業政策研究所 統括研究員)

鈴木 雅

(医薬産業政策研究所 主任研究員)

本報告書は研究上の討論のために配布するものであり、著者の承諾なしに転載、複写・複製することを禁ずる。

本報告書に記された意見や考えは著者の個人的なものであり、日本製薬工業協会および医薬産業政策研究所の公式な見解ではない。

内容照会先：

日本製薬工業協会 医薬産業政策研究所

〒103-0023 東京都中央区日本橋本町 2-3-11 日本橋ライフサイエンスビルディング 7F

TEL: 03-5200-2681; FAX: 03-5200-2684

E-mail: opir-sp@jpma.or.jp

URL: <http://www.jpma.or.jp/opir/>

謝辞

本研究を実施するにあたり、エキスパート勉強会講師として、ご指導・ご協力をいただいた田中博先生（東京医科歯科大学名誉教授、東北メディカルメガバンク機構 機構長特別補佐、東北大学医療情報 ICT 部門特任教授）、奥野恭史先生（京都大学大学院医学研究科臨床システム腫瘍学教授、理化学研究所 客員主管研究員・副グループディレクター（併任）、先端医療振興財団 客員グループリーダー・部長（併任）、菅野純夫先生（東京大学大学院新領域創成科学研究科メディカルゲノム専攻ゲノム制御医科学分野教授）、岡田随象先生（大阪大学大学院医学系研究科遺伝統計学教授）に謝意を表します。

目次

ページ

メンバーリスト	2
エキスパート勉強会講師	3
スケジュール	4
はじめに	5
米国の Precision Medicine の進展と日本の状況	8
バイオバンク／コホート研究	16
ビッグデータ解析のための人工知能（AI）	20
ゲノム／オミックスビッグデータを中心とした創薬	27
まとめと今後の課題	34

メンバーリスト (2015年7月～2016年6月時点)

研究員・事務局

○医薬産業政策研究所

森田正実 統括研究員

鈴木 雅 主任研究員

協力研究員

○日本製薬工業協会・研究開発委員会

蓮岡 淳 武田薬品工業(株) 医薬研究本部本部長室主席部員

藤田和也 アステラス製薬(株) 研究本部研究統括部企画管理グループ課長代理

赤塚浩之 田辺三菱製薬(株) 渉外部担当部長

○日本製薬工業協会・医薬品評価委員会

・臨床評価部会

近藤充弘 大塚製薬(株) メディカルアフェアーズ部オペレーション担当部長

・PMS 部会

宮崎 真 MSD(株) グローバル研究開発本部ファーマコビジランス領域
安全対策&薬剤疫学課

○日本製薬工業協会・産業政策委員会

・イノベーション推進部会

中平博之 大日本住友製薬(株) 研究本部研究本部長付

オブザーバー (製薬協委員会委員以外の定例会議メンバー)

坂田恒昭 塩野義製薬(株) グローバル医薬開発本部オフィスシニアフェロー

本間光貴 理研横浜 ライフサイエンス技術基盤研究センター

望月洋明 三井情報(株) ソリューションセンター バイオメディカル室

小仲一朗 NEC(株) プロセス業ソリューション事業部

関西第二医薬インテグレーション部 部長

南 英夫 富士通(株) 製薬ビジネス戦略推進室 室長

五十嵐夕子 (株)シードプランニング 執行役員 リサーチ&コンサルティング部

メディカルバイオチーム 主任研究員

*NEC 福間衡治氏 (医薬インテグレーション部 部長)、宮澤和秀氏 (医薬インテグレーション部 マネージャー) も適宜参加。

エキスパート勉強会講師

第1回 田中博先生

東京医科歯科大学名誉教授
東北メディカルメガバンク機構機構長特別補佐
東北大学医療情報 ICT 部門特任教授
日本オミックス医療学会理事長
地域医療福祉情報連携協議会会長
次世代生命医学研究所理事長

第2回 奥野恭史先生

京都大学大学院医学研究科臨床システム腫瘍学教授
理化学研究所 客員主管研究員・副グループディレクター（併任）
先端医療振興財団 客員グループリーダー・部長（併任）

第3回 菅野純夫先生

東京大学大学院新領域創成科学研究科メディカルゲノム専攻
ゲノム制御医科学分野教授

第4回 岡田随象先生

大阪大学大学院医学系研究科遺伝統計学教授

上記の先生方には、本報告書内に用いた図表についても提供いただいた。

スケジュール (1年目： 2015年7月～2016年6月)

7月 8日 (水) &10日 (金) キックオフ

7月 31日 (金) 14:30～17:00 第一回勉強会 (東京医歯大／東北大・田中先生)

8月 4日 (火) 10:00～12:00 研究会 (勉強会振り返りなど)

9月 8日 (火) 10:00～15:00 研究会 (レポートチェックなど)

10月 5日 (月) 13:30～16:00 第二回勉強会 (京都大・奥野先生)

10月 23日 (金) 13:00～15:00 研究会 (勉強会振り返りなど)

11月 10日 (火) 13:00～17:00 研究会 (レポートチェックなど)

1月 15日 (金) 13:00～17:00 第三回勉強会 (東京大・菅野先生)

2月 4日 (木) 15:00～17:00 研究会 (勉強会振り返りなど)

3月 16日 (水) 13:00～17:00 研究会 (レポートチェックなど)

3月 18日 (金) 13:00～17:00 サテライト勉強会 (新潟大・鈴木先生)

4月 15日 (金) 14:00～16:30 第四回勉強会 (大阪大・岡田先生)

4月 26日 (火) 15:00～17:00 研究会 (勉強会振り返りなど)

5月 18日 (水) 13:00～17:00 研究会 (レポートチェックなど)

6月 8日 (水) 13:00～17:00 中間まとめ

はじめに

ヒトゲノム計画では13年という年月、3500億円の費用をかけて、ヒトの全ゲノムの解読を行ったが、その後10年間で、次世代シーケンサーやコンピュータの性能が飛躍的に上がり、およそ1日、1000ドルの経費で個人の全ゲノムの解析ができるようになった。それに加えて、最近では生体センサー、ウェアラブルデバイスなどの機能が上がり、医療健康に関する多様な指標が継続的に計測できるようになった。このような大量の医療健康データ収集技術の発展の中で、医療健康情報を取り巻く環境は大きく変わりつつある。製薬企業においても、医療健康分野のビッグデータの活用により、近未来の医療や創薬における今後のパラダイムシフトや方向性が注目されている。

このような状況の中、医療・健康分野のビッグデータの活用により、将来の医療や医学研究のあり方がどう変わるのかを推測し、その変化によって疾患の概念や診断・治療、そして製薬企業の創薬等の活動が具体的にどのように変わるのかをまとめることを目的とし、医薬産業政策研究所（政策研）では、医療・健康分野のビッグデータの活用方法や課題および解決策を探るための調査研究を行うこととした。

この政策研究を行うにあたっては、カバーする範囲が広範にわたり製薬企業の内外でタイムリーに動いている内容であることから、政策研に「医療健康分野のビッグデータ活用研究会」を立ち上げた。研究会では、政策研を研究主体に、製薬協内の研究開発委員会、医薬品評価委員会、産業政策委員会から協力研究員が参画し、外部からもIT関連企業を中心にオブザーバーを迎えて、医療健康分野のビッグデータの最新の状況を踏まえた調査研究を進めた。また、その調査研究の補完や理解を進める目的で、外部からエキスパートの先生を招いて、定期的な勉強会も開催してきた（初年度は4回実施）。

研究を進めるにあたり、医療健康分野のビッグデータについて全体をみると、この分野のビッグデータには実に多くのデータが含まれることがわかる（図1）。電子カルテやレセプト、ゲノムやオミックス、さらには、いわゆるウェアラブルデバイスで測定できるデータなど、「医療健康分野のビッグデータ」という言葉からイメージするものは、人によって多様である。実際、この医療健康分野のビッグデータについて、各所でセミナー／講演会が開催され、報道がなされているが、これらが混在しているために、論点が定まっていないことも否めない。異なる地域や機関に存在している医療健康に関する様々なデータが、ICT技術の進展によって、名寄せ、追跡、集積できることでビッグデータ化することが目指されている。一方、それによって期待されている成果は多岐に及んでいる。

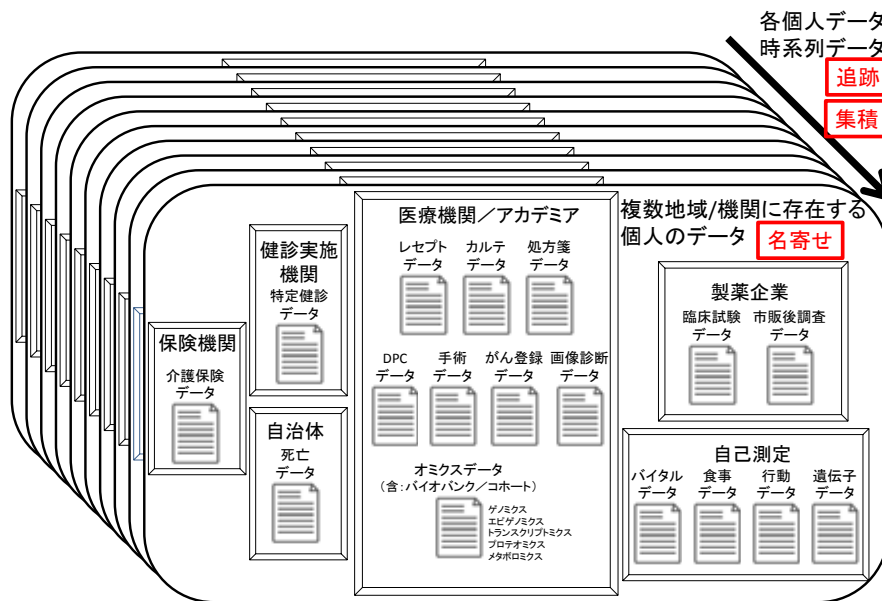


図1 医療・健康分野のビッグデータとは？

医療健康分野のビッグデータの活用に対する期待は立場により多様で、オールジャパンで医療ビッグデータの活用をしようという中であっても、例えば、医療機関では医療連携や医療サービスの向上を、国や保険機関は、医療費適正化などといった具合に、各ステークホルダーの思惑が異なることも、このビッグデータの活用を考える際の議論が薄まっている一因である。

製薬企業に関連の深い項目を中心に見ても、多くが含まれていることがわかる(図2)。

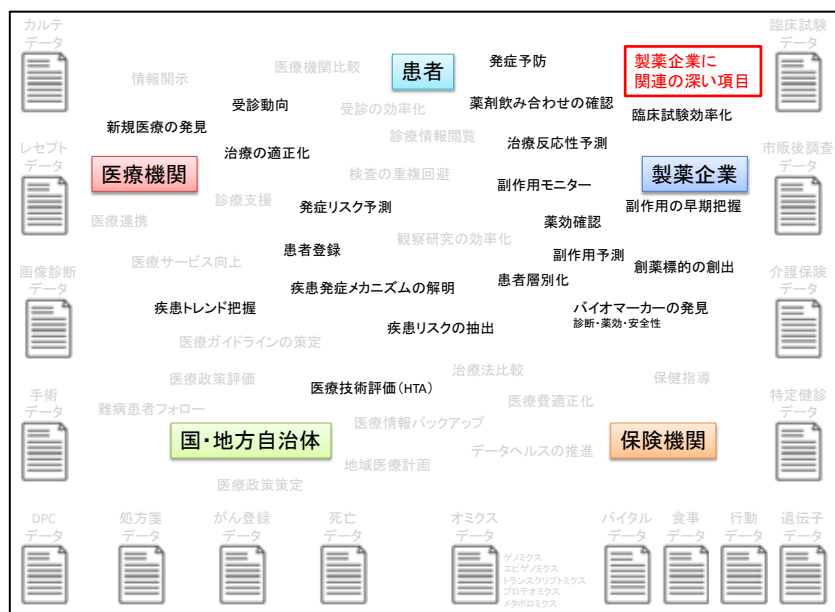


図2 医療・健康分野のビッグデータ活用への期待

さらに、製薬企業のバリューチェーンのどこに影響するかということを考えても多様で、上市品の販売戦略といった育薬から、開発の臨床試験効率化、研究の創薬の標的の創出、さらに他の産業への展開なども期待されている（図3）。

そのために、業界内で医療健康分野のビッグデータの活用を検討しようとしても、その重要視するポイントの違いによって検討が希薄化してしまう危険性がある。

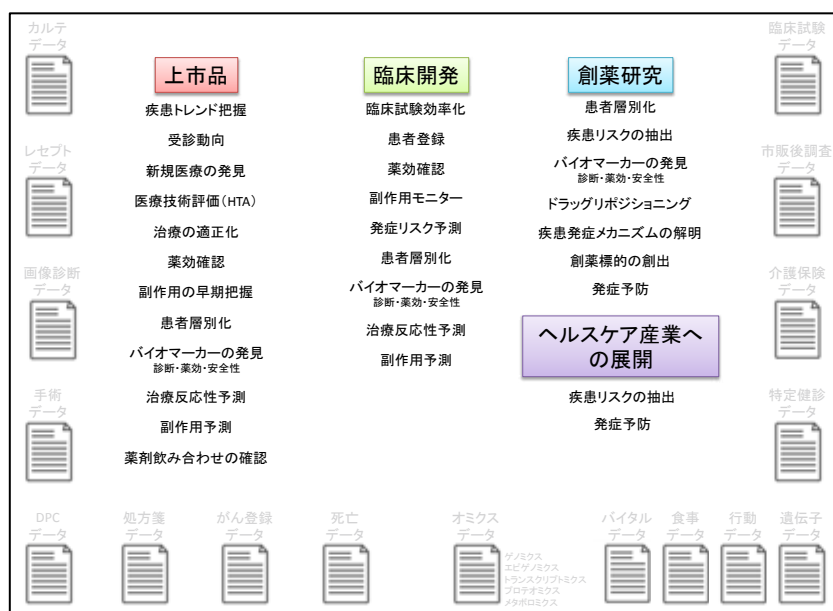


図3 医療健康分野のビッグデータの製薬企業への影響

そこで、本研究会での検討を進めるにあたっては、「医療健康分野のビッグデータ」をある程度ポイントを絞って検討を行うこととした。NDB（レセプト情報・特定健診等情報データベース）の活用、MID-NET（医療情報データベース基盤整備事業）による的確な安全性情報の迅速な抽出などについては、製薬協内の各関連委員会で実務的な対応が行われているので、その対応に任せることとし、本研究会の中心テーマ（初年度）としては、レセプトや電子カルテといった臨床情報とゲノムを中心とするオミクスデータを連結させていく診療や研究のためのビッグデータを対象とすることとした。

米国等、医療現場でゲノムビッグデータの取り込みを行っている海外の先進状況と日本の動きを比較しつつ、現時点ではまだ一部でしか実践されていないが、近い将来訪れるであろう臨床ゲノム／オミクスデータの実臨床での取得、活用により、Precision Medicine や疾患分類の細分化、人工知能（AI）の活用、創薬手法の変化がどう進むのかなどについて検討を行ってきた。

本報告書では、それらについて報告するとともに、日本における基盤整備上の課題や製薬企業に対する期待について後述する。

米国の Precision Medicine の進展と日本の状況

2015 年年頭のオバマ大統領の一般教書演説で述べられた Precision Medicine Initiative¹⁾では、100 万人のコホートによって遺伝的素因と環境素因との相互作用を徹底的に調べることが発表された。総額 2 億 1500 万ドルの大統領予算を付けて、1.3 億ドルで National Institute of Health (NIH) が 100 万人コホート調査情報を収集し、7 千万ドルで National Cancer Institute (NCI) ががんのドライバー遺伝子の発見を実施し、1 千万ドルで FDA (Food and Drug Administration) がコホート調査のデータベース開発を行い、500 万ドルで国家医療 IT 調整官室 (ONC : Office of the National Coordinator for Health Information Technology) が、それに関する標準規格・情報プライバシー・セキュリティを研究する。

Precision Medicine という言葉は 2011 年に全米研究評議会 (NRC : National Research Council) が、これまで表現型や臓器別に作られていた病気の大系は、疾病の本質とは異なるため、これからはゲノム/オミックス機序から作った方がいいという勧告を出した際に提案されている。内在的な疾病発症機序の詳細な差異に対して診断・治療を行う観点から、一定のレベルで層別化されたグループでの「個別化」が適切であるという考えである。

Personalized Medicine として遺伝子の多型や変異だけが個別化の指標だった時代から、最近では、環境・生活習慣要因の方が遺伝的素因よりも重要、あるいは遺伝的素因と環境素因との相互作用が重要だと認識されつつある。Personalized Medicine にこの環境・生活習慣要因がプラスされているのが Precision Medicine の概念であるとも言える。

この Precision Medicine に重要なのは、単純にゲノムの多型性の違いで層別化し、個人化するのではなく、それぞれの多型がどう臨床的転機等に影響したのか、あるいはするかという表現型とセットになっていることである。これで初めて、意味のある治療のための単位として、疾患を層別化することができる。

Precision Medicine を実現するためには、層別化のエビデンスの構築に、疾病発症や臨床過程に関する基盤情報が必要であり、患者毎に継続的なデータ収集が可能なモバイルヘルスや追跡型のバイオバンク、ゲノムコホートの取組みが重要となってくる。その意味では Precision Medicine Initiative が大統領教書演説として発表されたことは、アメリカではそれほど積極的に進められてなかったバイオバンクに、今後は手を付けていくという宣言がされたということでもある。

医療においては Population Medicine (疾患ごとの標準的治療を基礎とする治療) による包括的な医療はもはや成り立たず、個別化医療をどのように体系化するかということが課題である。Precision Medicine は個別化パターンを網羅的に調べるといふ新しいビッグデータの活用の観点から、患者の臨床情報 (フェノタイプ) や環境、生活習慣などの要因の解析を含めたゲノム/オミックス医療を目指した動きとなる。次世代シーク

1) 医薬産業政策研究所「米国の Precision Medicine Initiative」 政策研ニュース No.45 (2015 年 7 月) 参照

エンサーの進展などにより、ゲノム情報を中心とした多種多様の情報が収集できるようになってきた状況もあり、個人の網羅的な遺伝子診断と他の情報との紐付けが Precision Medicine ではさらに重要になってくる。

またオバマ大統領が唱えた Precision Medicine Initiative は、医療に効率化をもたらす方策でもある。世界的に医療費や薬剤費の高騰、高額化が課題となっており、がん治療においては薬の 75%は効いていないというようなことも言われているため、ゲノム／オミックス情報をベースにして治療毎に有効な患者層を突き止めるということは、一つの重要な目的でもある。

米国では 2010 年に、主として単一遺伝子病で、未診断疾患（診断することが非常に難しい疾患）の原因遺伝子の同定というスタイルで、臨床現場でのゲノム／オミックス医療が始まった。これは 2005・2006 年から次々と発表された次世代シーケンサーが、2007・2008 年から急速なシーケンス能力の拡大とコストの低減を実現し、「シーケンス革命」と呼ばれる事態を引き起こしたインパクトにより実現できた動きである。

当初は生命科学研究が対象であったが、2010 年頃から研究だけではなく診療に次世代シーケンサーを使用する事例が出現し始めた。シーケンス革命からわずか 3 年で、臨床実装が開始されたのである。

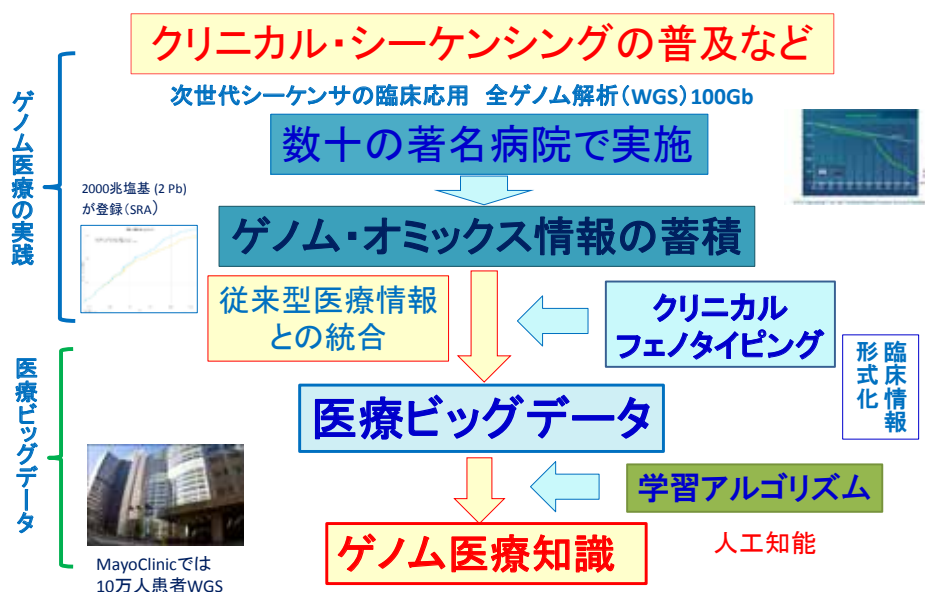


図 4 米国における Precision Medicine の進展

Institution	Major Projects	Institution	Major Projects
MC Wisconsin	Using whole genome sequencing to establish diagnosis in patients with currently undiagnosed genetic disorders	Alabama	Planning stages for projects in risk assessment, <u>pharmacogenetic</u> analysis, identification of families for further research
Mount Sinai	<ul style="list-style-type: none"> CYP2C19 testing for antiplatelet <u>rx</u> post percutaneous coronary intervention Personalized decision support for CVD risk management incorporating genetic risk info 	Baylor	Whole <u>exome</u> and whole genome sequencing in <u>Mendelian</u> disorders to improve <u>diagnosis</u> .
Northwestern	Using pharmacogenomics evidence (from GWA genotyping) to guide prescriptions in primary care and assess risk for other conditions such as HFE/hemochromatosis	Geisinger	<ul style="list-style-type: none"> Selection for gastric bypass surgery <u>vs</u> other <u>wt</u> loss means based on genetic variants predictive of long-term benefit from surgery IL28B variants and response to hepatitis C treatment KRAS and BRAF mutational analysis in thyroid cancer patients
Cleveland Clinic	Tumor-based screening for Lynch syndrome, endometrial cancer	Ohio State	<ul style="list-style-type: none"> Personalized genomic med study of CHF and HTN <u>pts</u> randomized to genetic counseling <u>vs</u> usual care CYP2C19 testing in interventional cardiovascular procedures for <u>clopidogrel</u>.
UCSD	<ul style="list-style-type: none"> Screening for actionable mutations in malignant <u>gliomas</u> and <u>glioblastomas</u> for biomarker based RCTs Targeted <u>rx</u> (such as RET inhibitor) of metastatic solid tumors based on tumor mutation status 	Harvard	Whole genome sequencing with integration in EMR and CDS; pilot of 3 patients to start
Morehouse	<ul style="list-style-type: none"> <u>Exome</u> sequencing of 1200 early onset severe African American hypertension cases and 1200 controls 	U Penn	Genotyping for assessment of MI risk in Preventive Cardiology program
Duke	<ul style="list-style-type: none"> Computer-based family <u>hx</u> collection and CDS tool with 1-yr follow-up for perceptions, attitudes, behaviors related to thrombosis and breast, ovarian, and colon cancer SLCO1B1*5 genotyping and statin adherence Effect of genetic risk info on anxiety and adherence in T2DM 	St. Jude's	Pre-emptive <u>PGx</u> genotyping in children
		Vanderbilt	Pre-emptive <u>PGx</u> genotyping for <u>clopidogrel</u> , warfarin, or high-dose simvastatin
		U Maryland	Develop and apply evidence-based gene/drug guidelines that allow clinicians to translate genetic test results into actionable medication prescribing decisions
		Mayo	<ul style="list-style-type: none"> <u>PGx</u> driven selection/dosing of antidepressants CYP2C19 genotyping for antiplatelet <u>rx</u> post PCI
		Inter-Mountain	Tumor-based screening for Lynch syndrome

表1 National Human Genome Research Institute (NHGRI) Working Group のリストから見るゲノム／オミックス医療の実践

表1にリスト化している米国で実施されてきたゲノム／オミックス医療の多くは、生得的ゲノム（生殖細胞系列ゲノム）情報の変異・多様性に基づいて、疾患原因遺伝子や疾患感受性遺伝子の同定が行われる第一世代のゲノム／オミックス医療である。病因未知の疾患原因遺伝子を全ゲノムシーケンス (WGS) や全エキソームシーケンス (WES) で見つけることで有名になった Wisconsin 医科大学や Baylor 医科大学病院、また薬物代謝酵素の多型性判定と電子カルテへの DNA 記載の実装を行っている Vanderbilt 大学病院²⁾などを含めて、ほとんどのものが第一世代であることがわかる。

一方、WGS／WESで難治性のがんドライバー遺伝子変異の同定を行っている Mayo Clinic や MD Anderson がんセンターでは、後天的な体細胞変異を見ており、厳密に言えば、(体細胞) オミックス医療の世代に属するが、第一世代と同様に次世代シーケンサーのインパクトが影響した臨床実装である。

この米国のゲノム／オミックス医療の臨床実装の取組みを見てみると、2010年に Wisconsin 医科大学の小児病院での3歳の小児の原因不明の腸疾患へのゲノム医療の取組みと成功で火がついて、2013年には NIH の BD2K (Big Data to Knowledge) Initiative という国家プロジェクトが発足し、2015年にオバマ大統領の年頭教書で Precision Medicine Initiative が発表されており、当初先端医療施設の個別の取り組みから始まった医療が、5年という短期間のうちに国家的取り組みに強化され、既に多くの有名病院でクリニカルシーケンスが日常診療で実践され、ゲノム／オミックス医療の時代が始まっていることがわかる (図5)。

2) 医薬産業政策研究所「Learning Healthcare System-実臨床データによる医療の検証・改善-」政策研ニュース No.46 (2015年11月) 参照

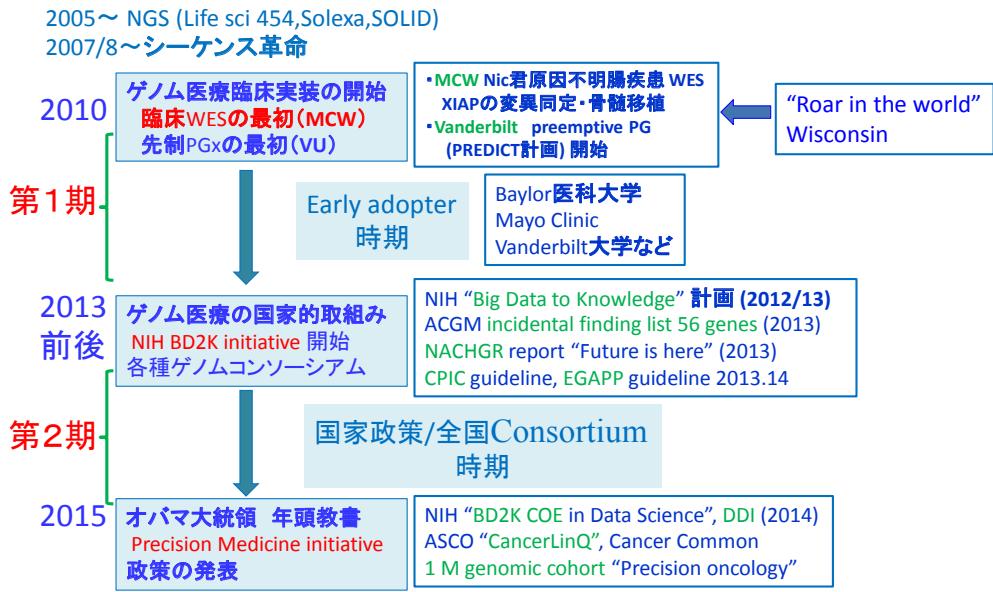


図5 ゲノム/オミックス医療の米国での経緯

このゲノム/オミックス医療を進めることに寄与している関連プロジェクトとして eMERGE (Electronic Medical Records and Genomics) プロジェクト¹⁾ (2007年～phase I、2011年～Phase II、2015年～Phase III) がある。

Phase Iにおいては、電子カルテ (EMR : Electronic Medical Record) を通じて臨床フェノタイピングを行ったり、臨床フェノタイピングとバイオレポジトリに基づく GWAS (Genome-Wide Association Study) が可能かについて検討され、また利用に関する ELSI (Ethical, Legal and Social Issues) 側面を検討するなどの取組みが5施設で実施された。

さらに Phase II では、電子カルテと遺伝情報の統合の実装を目指し、更に4施設が加わり、電子カルテへのゲノム情報の統合を行い、PGx (Pharmacogenomics) の臨床応用に関する試行プロジェクトや診療における結果回付に関する検討 (遺伝カウンセリング等を含む) が行われている。また、CSER (Clinical Sequencing Exploratory Research) コンソーシアムと協同して電子カルテ等の臨床情報に遺伝子情報を統合した、実装での臨床活用についても検討されており、臨床情報と一体になって臨床的転帰を改善するための WGS/WES 解析が具体的に進められている (図6)。



図6 eMERGE 実施拠点

このように米国では、多くの先端医療病院でクリニカルシーケンス（日常臨床における遺伝子解析）が行われ、ゲノム／オミックスの臨床実装と医療研究活用が始まっている。この臨床実装では日本は米国に大きく差をつけられている。

また、米国では既に数多くの医療情報データベースが存在し、活用されている³⁾（図7）。電子カルテと臨床ゲノム情報がリンクしたデータベースについては、NextBioClinical（Illumina 社）など既に商用で売られ始めているものもある。このような臨床実装や医療情報データベースにより、臨床ゲノムに個々の患者の臨床診断などの詳細情報や、分子レベルの解析情報を統合的に検討することで、新しいバイオマーカーや疾患原因探索、創薬標的の創出などを進めつつある。現在は医療情報データベースの活用により、比較有効性研究や保険医療政策研究、直観的医療の裏付け、治療法の妥当性の確認などが中心である。

しかし、米国においても広く一般の医療機関においては、ゲノム／オミックス情報を含むデータ連携はまだ十分進んでおらず、民間における利用にも制限がある。その意味では創薬標的の創出や新規ヘルスケアサービスの提供といった本格的なビッグデータの2次活用までは必ずしも結びついていない。

今後は Precision Medicine Initiative の進展によりゲノム／オミックス情報の電子カルテへの組み込みが急速に進んでいくことが推測される。その進展に伴い、このフェノタイプとゲノム／オミックスの連結されたデータ解析により、画期的な治療法や創薬標的が見出されることが期待されている。

3) 医薬産業政策研究所 「米国における医療情報データベース」政策研ニュース No.46（2015年11月）参照

DB/DB運用組織	規模	含まれるデータ	二次利用	活用例
NIS	700万人 (各年)	診断および処置情報、患者情報、施設情報 予想される支払元、総費用 退院時の状態、滞在の長さ、 重症度と併存疾患対策等	制限なし (350ドル/1年分)	・研究者、政策立案者による各項目の全米推計 (ヘルスケアの利用、アクセス、料金、品質、 成果等) * マーケティング目的での利用は禁止
CMS(メディケア・ メディケイド サービスセンター)	メディケア (5,200万人超) メディケイド (6,400万人超)	診断および処置情報、患者情報、 施設情報 サービス提供者および支払い金額、 外来やリハビリ、 薬局に対する請求データ (メディケアクレームファイルの場合)	集計データは制限なし 個票データは 公的利用のみ (ResDACを通じて提供)	・研究者、政策立案者による各データ項目の解析 * 個票データは申請・審査を経て研究・政策 利用での使用が許されている
Kaiser Permanente	900万人	診断および処置情報、患者情報、 疾患レジストリ、健康調査、 コホート研究等	外部への データ提供サービス 無し	・FDAとの連携によるVioxx (NSAID) の市販後安全性 (心疾患リスク) ⇒Vioxxの市場からの撤退
Optum	7,400万人	請求データとリンクしたEHR	Humedica NorthStar、 Clinformatics 等	・製薬企業による市場セグメンテーションや比較有効性 分析による製品発売計画段階から製品成長のための 戦略立案 ・疾患に関する市場情報・患者情報の解析によるマーケ ティングへの提供 ・臨床試験の条件に一致する患者を診断している医師 情報の提供
Truven Health Analytics	2億人以上	請求データとリンクしたEHR、 健康リスク評価、臨床検査値、 病院データ等	MarketScan Databases 等	・比較有効性研究 ・製品差異化ポイントの解析による製品価値の実証
Practice Fusion	500万人	EHR	Insight 等	・特定の疾患の罹患の傾向把握 ・処方薬の市場シェアの毎週変化の把握
Heritage Provider Network	70万人以上	EHR、画像制御システム、 業務管理の統合データ	外部への データ提供サービス 無し	・患者の病歴と処方履歴の検証による薬物相互作用の チェックによる投薬管理
IMS Institute	30億件以上の 処方箋データ 毎年15億件以 上の医薬品の 取引データ	処方箋 取引価格	Xponent, National Prescription Audit, National Sales Perspective, National Disease and Therapeutic Index 等	・新規処方箋数の追跡 ・医師の処方箋の他剤へのスイッチ分析 ・医薬品の分配に関する各種パターン(地理、年齢、性 別、支払方法等による階層化)の分析 ・医薬品の実取引価格の調査 ・疾患の治療パターンの解析

図7 米国における主な医療情報データベース

一方、日本でのゲノム/オミックス医療の実装は、いくつかの医療機関で試行されているが、まだ研究の色合いが強い。例えば、研究費を用いた試行的ゲノム医療としては、国立がん研究センター東病院や静岡県立がんセンター⁴⁾で、遺伝子検査によるがんのドライバー遺伝子の診断、研究があり、東大病院ゲノム医学センターでは、先天的神経筋疾患の診断、研究が行われている。他に、がんの患者から取った組織をシーケンズして、適した治験薬を使う産学連携の取り組み「SCRUM-JAPAN」が、国立がん研究センターを中心に進められている。

また、2015年4月から、京大病院がんセンターと三井情報の共同プロジェクトで、原発不明がん、希少がん、標準治療不応がんなどを対象に「OncoPrime」というがん関連遺伝子解析検査を用いた抗がん剤の選択が実施されている。臨床現場でのゲノム解析であるため、精度、品質管理がされたクリニカルクオリティでの解析が非常に重要であるが、日本の場合はクリニカルクオリティの基準が整備されていないため、現在は米国のCLIA (Clinical Laboratory Improvement Amendments) 基準を満たした米国のラボで解析を行っている。この検査は、現在自由診療で行っているため、京大病院の場合は、90万円弱の患者負担がかかる。2015年4月にスタートして、半年で60例を越える検査の実績があるが、例数が少なく、未だ十分な解析には至っていない(2015年10月現在)。今後は、岡山大、北大、千葉大で同じスキームを広げ、ゲノム情報と薬剤反応性などのデータを広く収集分析して、薬剤反応性予測のモデルを作ることが考えられ

4) プロジェクト HOPE (http://www.scchr.jp/press_releases/20140818/20140818.html)

ている。

ゲノム医療の推進については、内閣官房の健康・医療戦略本部に設置された「ゲノム医療実現推進協議会」や厚労省の中に設置された「ゲノム医療実現本部」などにより、日本でも国策としても進めようとしているが、実用化については欧米の取り組みに対して遅れている。保健医療での取り扱いや個人情報保護等の倫理の問題など、制度的にも課題があり、また、治療法とアウトカムのデータベースがなく、ゲノム診断の結果から適した治療法の選択が難しいということや、臨床情報と遺伝子情報の統合やデータベース化が困難であること、クリニカルシーケンスを行う体制ができていないことなど、ゲノム医療の実装自体に関わる課題も解決されていない。

2016年度の日本医療研究開発機構（AMED）の概算要求予算として、文科省 20 億円、厚労省 30 億円の新規のゲノム解析予算が確保され（図 8）、文科省は東北メガバンク（東北 MMB）や理化学研究所の BioBankJapan（理研 BBJ）といったバイオバンクでのゲノム解析プロジェクト、厚労省は疾患ゲノム情報総合データベース整備事業を中心に医療の ICT 化の施策を進めようとしている。更にはナショナルセンターを中心にがんや遺伝病で、実際の治療をするときにどれぐらいのゲノム情報が使えるかについて、実証実験のようなことをやるという計画も入っており、このような取組みは Precision Medicine を目指した施策としてマッチングする。

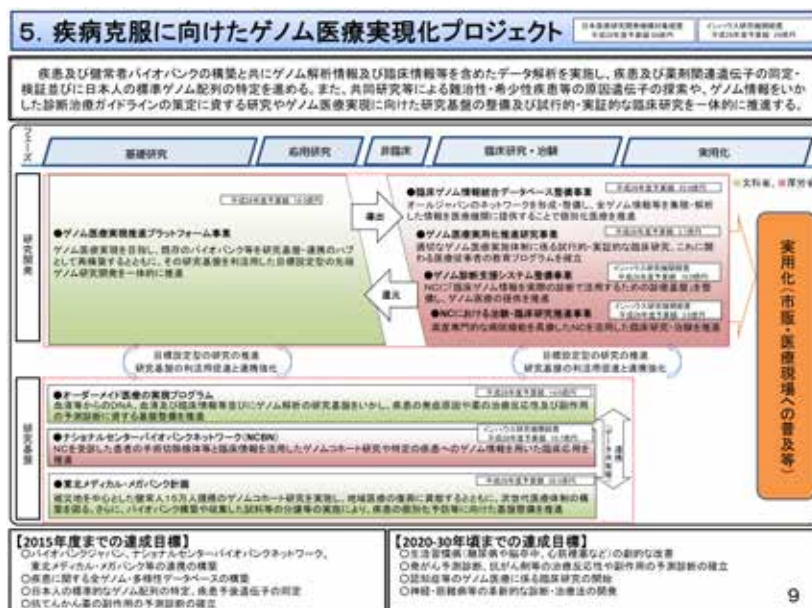


図 8 疾患克服に向けたゲノム医療実現化プロジェクト

データベースの基盤整備などを行う「医療分野の ICT 化の推進」には 16 億円の予算がついている。「クリニカルイノベーションネットワーク（CIN）の構築」については、31 億円の予算となっている。文科省の取組みでは理研 BBJ や東北 MMB といった既に

稼働しているゲノム研究プラットフォームを活用しようとしている。がんや脳など疾患プロジェクト等のデータベースの連携や、目標設定型の先端ゲノム研究などもある。

東北 MMB の事業はロードマップによれば 28 年度で 15 万人リクルートを達成する計画である。1000 人全ゲノム解析のデータも、いわゆるコントロールアクセスという形で公開された。また、理研 BBJ のオーダーメイド実現化プログラムもゲノムプラットフォーム事業の基盤を支える事業である。これまでにたくさんの論文を排出し、アカデミアが申し込んで審査に受ければ、サンプルやデータの使用もできるが、必ずしもデータは公開されてはおらず、自由なアクセスは制限されている。しかし現在第一次プロジェクトの対象 20 万人の臨床情報などのまとめ作業が進められており、収集してきた生体試料や紐付け情報の活用に焦点が移っている。

さらに AMED は ICT をはじめとした医療研究を対象として、2016 年度に 175 億円程度の調整費の予算を組みこもうとしている。疾患ゲノム情報総合データベースなど Precision Medicine 実現化を後押しする施策の推進は、産業界への影響も大きい。

米国の Precision Medicine Initiative により、今まで研究レベルであったゲノムが実臨床で使われそうであるということが日本においてもかなり大きなインセンティブになっており、今後実用化のための取り組みが進むことが期待される。

バイオバンク／コホート研究

バイオバンクの目的や機能は、従来、生体標本（生体試料）の収集や臨床研究の試料保存、非臨床試験等の試料供給といった役割が主体であったが、近年はゲノム／オミックス医療を実現するためのヒト健康・疾患基盤情報の収集が重要な役割となっている。つまり、バイオバンク／コホート・疾患レジストリなどで診療データ等の患者情報が紐付けされたゲノム／オミックス情報の集積や生体試料の活用等により、新しいEBMの構築が期待されており、この取り組みをこれから急速に進展させることで、ゲノム／オミックスの医療研究や臨床活用の範囲、意義が著しく高まると考えられる。

バイオバンクの機能を大きく二つに分けると、疾患型バイオバンクと集団型バイオバンクがある。疾患型バイオバンクでは疾患罹患患者の網羅的分子情報とそれに紐付けされる臨床表現型情報や患者環境情報などを収集し、疾患ゲノム／コホートや疾患レジストリなどと連携して疾患の原因探索、個別化医療や創薬のエビデンスの供給源となる。

疾患レジストリを活用したスウェーデンで臨床試験では、試験の質を担保しながらコストを大幅に抑えたと報告されている。臨床表現型（フェノタイプ）だけではなく、遺伝子型を対象患者層別化に加えることで、特定の薬剤の反応性に合わせた患者のリクルートと割り付けが可能となる。疾患レジストリに参加する多施設の医療機関と情報を共有、プロトコルを統一化することで、Registry Based Randomized Clinical Trialが可能となる。その疾患の治療やフォローをそれぞれの施設で行うことで、リアルワールドデータ（real world data）の活用や臨床試験の費用削減につながることを期待できる。また、被験者のフォローも一般の診療の中で長く観察が可能であり、血液や組織の試料採取や保管の方法を規定し、バイオバンクとの連携を高めることにより、後追いでゲノム／オミックス研究の機会も担保できる。

一方、集団型バイオバンクでは健常者前向きコホート研究等と連携し、健常時点での網羅的分子情報と臨床環境情報を集めて、生涯を追跡することにより、予防医療や先制医療などのエビデンスの供給源となる。

東北MMBのコホート研究とバイオバンクは2013年から活動を開始している。宮城県、岩手県で住民コホート8万人、後述するオランダのコホートに倣った3世代コホートが7万人の計15万人の健常人コホートを目指している。既に初期の成果として日本人の健常人1000人の全ゲノム配列情報を公開し、その過程で欧米では未発見の1200万のSNPを日本人の解析で見つけている。日本人のデータベースの構築は、これからの疾患ゲノムの研究のプラットフォーム事業として高い評価をされている。この日本人全ゲノム配列情報はNBDCがリポジトリとして公開されており、今後日本のデータベースの充実を図っていこうとするものである。

15万人コホートの全被験者のゲノム・オミックス情報の解析を実施することは物理的にも費用的にも困難が伴うため、適切な生体試料のサンプリングの重要性は高い。被験者の疾患発症などのフェノタイプ変化に際して、過去の試料のゲノム／オミックス解

析を実施できるように準備しておく必要がある。特に予防・先制医療の研究を検討するにあたって、試料保管の意義が高い。

更に、日本の今後のゲノム情報の充実を考えて、東北 MMB では日本人に最適化された SNP アレイである「ジャポニカアレイ」の開発を行った。未だに次世代シーケンサーによる全ゲノム解析の費用が 1 人 2~30 万円かかる現状で、このジャポニカアレイでは 2 万円の費用で SNP 解析ができ、被験者の疾患関連遺伝子の解析やリスク検診に活用できる。

2000 年ぐらいから多くの先進諸国で、大規模なバイオバンクの構築が行われているが、一番有名なのは英国の UK Biobank で、50 万人の健常人データを収集し、健診データ等と紐付けしている。英国 Genomics England プロジェクトは希少疾病、がん患者等を中心に 2013 年に疾患ゲノムの収集を開始して、2017 年までに 10 万人のゲノム配列を集めようとしている。

英国ではまず英国人の標準 SNP の分布を調べることを目的に UK10K (1 万人) が行われ、引き続いて 100K (10 万人) で全ゲノムシーケンスを行い、順次公開している。基本的に対象としているのは遺伝子疾患であり、その原因遺伝子探索を重要テーマとしている。データベースは単にシーケンスデータがあるだけではなく、クリニカルの特徴データ等も付随する被験者データとなっている。日本のプロジェクトも東北 MMB を始め、公開を心がけてはいるが、データ公開や 2 次活用において、UK100K をお手本としたデータセットが必要である。

また欧州全体としては、BBMRI (Biobank/Biomolecular Research Infrastructure) という欧州各国の 250 以上のバイオバンクを統合した組織を持っている。欧州では小さなバイオバンクが多い中、オランダは Lifeline という 16 万人を超える規模のバイオバンクを 2006 年から開始しており、このバイオバンクでは、おそらく世界で初めての試みとして、東北メディカルメガバンクに先駆けて 3 世代コホートで情報を集めている。

通常、個々人のゲノムの違いを任意の 2 人で比べた場合は 0.1% から 0. 数% 以内ぐらい、つまり 300~1000 万個ぐらいの違いがあるが、父親、母親とその子のゲノムの違いは 50 個程度である。すなわち、親から遺伝情報を受け継ぐときに生じる遺伝子の変異 (新生突然変異) は非常に少ないため、配列測定誤差を除去するのも容易で、3 世代コホート研究は多くの意味のある研究をもたらすと考えられている。

米国では Precision Medicine Initiative で 100 万人の疾患ゲノムコホートからデータを取り、新しく創設するバイオバンクでゲノム/オミックス情報の統合データ化を目指しているが、既存のデータを含めた情報の統合化という動きが主体である。米国のこの動きは 2015 年に始まったばかりであり、日本もこのバイオバンク/コホートの取組みについては欧米の活動から大きく遅れてはいない⁵⁾⁶⁾ (図 9)。

その意味で、ゲノム/オミックス医療の医療現場での実装が遅れている日本では当面はこのバイオバンク/コホート・疾患レジストリにおいて、フェノタイプ情報と統合化

5) ヒューマンサイエンス振興財団 創薬資源調査報告書「次世代医療に向けたコホート研究の動向 ―創薬並びに個別化医療や予防医療への貢献の道を探る―」 http://www.jhsf.or.jp/paper/report/report_201404.pdf

6) ヒューマンサイエンス振興財団 研究資源委員会調査報告書「バイオバンク・ネットワーク ―個別化医療及び創薬の基盤整備―」 http://www.jhsf.or.jp/paper/report/report_no81.pdf

されたゲノム／オミックス情報の活用により、臨床エビデンスを蓄積していくことが必要である。

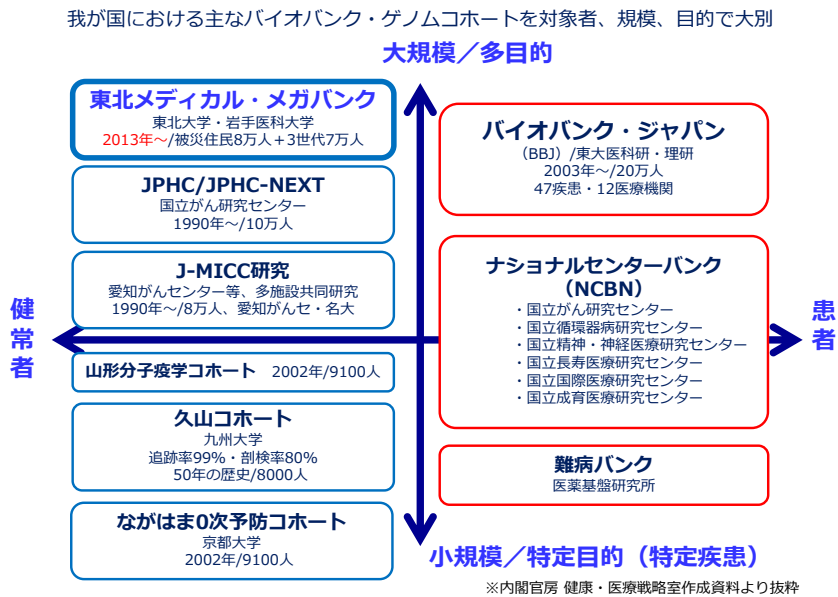


図9 日本国内の主なバイオバンク／ゲノムコホートの状況

バイオバンクに対する企業側の要望として、提供されている試料や情報の質の担保の問題がある。バイオバンクの設備、試料の採取、保管管理、取り扱い(SOP、規格)、情報等の面での標準化や質の確保が望まれるが、海外のバイオバンクとの連携の問題も絡んで、ISO化への動きが出てきている。東北 MMB では ISO9001 (品質管理) と ISO27001 (情報・セキュリティ) を取得している。

また、ゲノム／オミックス情報の重要性が増す中、ゲノム解析センターを併設したバイオバンクの形態が望まれている。東北 MMB は 2013 年の発足当時から解析センターを併設した『複合バイオバンク』として創設されている。試料提供と一体化したゲノム／オミックス情報の提供が可能である。将来的には 15 万人の全てのホールゲノムの解析情報を提供する予定である。

バイオバンク／コホートを日本の強みとして活用しようと考えた場合、そのサンプリングや測定についてはまだ課題や、改善点も指摘されている。例えば、以下のような点である。①ゲノムはどこで測っても同じだが、エピゲノムはどの細胞で測るかがポイントになり、メタゲノム、メタボロームについても同様で、サンプリングの部位、量、タイミングには制限が多い(脳の細胞をどう取るのかなど、採取部位によってはサンプリング技術の課題もある)。②感度レベルとしても、プロテオームはまだ、数千も読めず、相互に比較できるようなデータになっていない。そのため、現時点では核酸系(ゲノム、トランスクリプトーム、エピゲノム)のデータがどれだけ取れるかがポイントになって

いる（メタボローム、プロテオームは、個別にいくつかのサンプルを確認する際に使う技術であり、ビッグデータ解析ができるレベルにはない）。③バイオリソースのクオリティとして、施設によって、保存の仕方や採取の仕方が違い、データの比較ができない。各ナショナルセンターレベルでも、プロトコールの違いがある（例えば保存液が何かについても統一されていない）。

ビッグデータ解析のための人工知能（AI）

ゲノム／オミックスが臨床実装された臨床現場では個々人のゲノムや診療情報からどのような診断、治療、予後予測といった医学判断をするかという活用が問われることになるが、そのためにもゲノム医療知識の集積や選択が必要となる。このゲノム医療知識は文献的な情報の収集を行うことはもちろんであるが、医療研究や実診療でのゲノム／オミックス情報の蓄積と臨床表現型解析とのデータ統合によって、今後急速に深まっていくと考えられる。またこの過程で、膨大な情報の解析と活用のために、スーパーコンピュータによるシミュレーションやAIの活用が重要である（図10）。

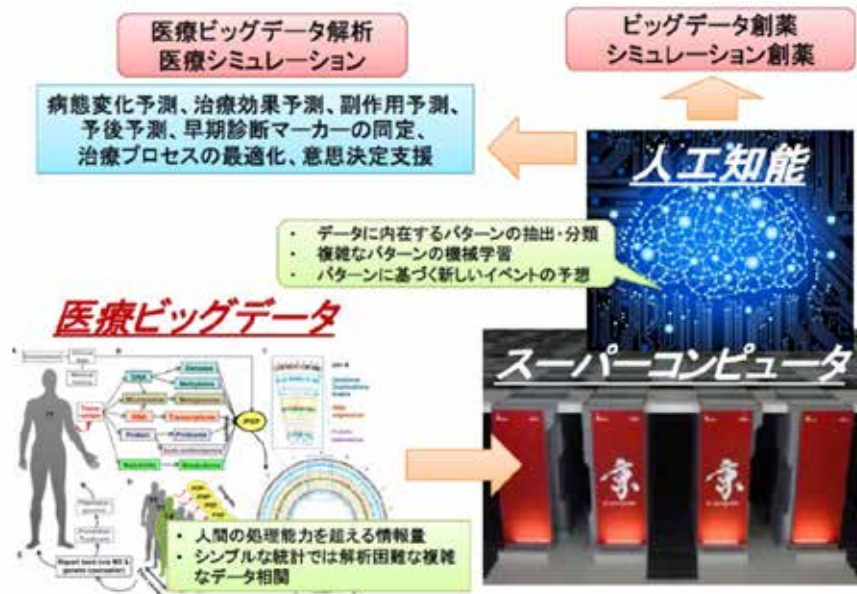
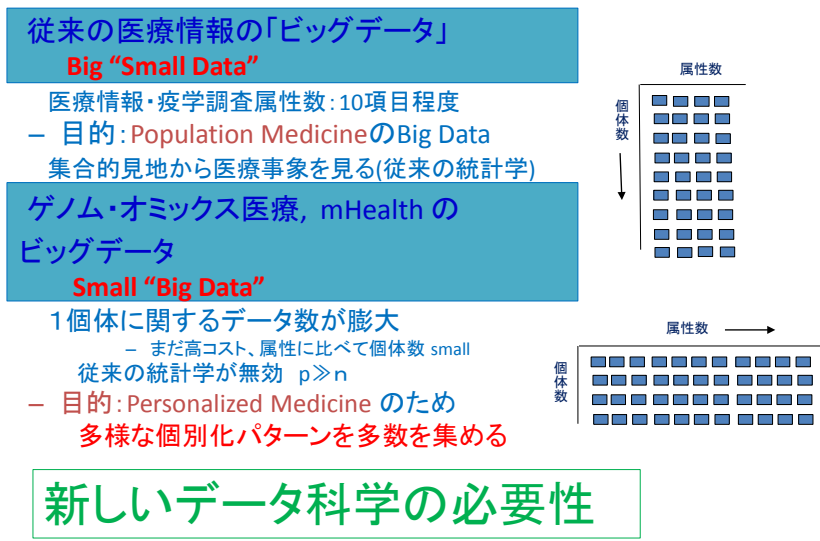


図10 医療ビッグデータの解析とシミュレーション

医療ビッグデータは単にデータ量が多くなったということだけではなく、医療や創薬等のプラットフォームを大きく変革していく力となる。その意味ではここでいう医療ビッグデータとは、診療情報だけではなく、ゲノム情報、オミックス情報（網羅的タンパク発現、遺伝子発現、代謝物、エピゲノム情報など）、環境・生活情報などの多くの情報（時系列変化）を含み、それらが紐づけされた情報であることが重要である。

多人数の個人の生得的ゲノム情報（生殖細胞のゲノム多型性など）、後天的オミックス情報（体細胞の網羅的分子プロファイルなど）、診療情報、環境・生活情報などの多くの情報をデータベース化し、そのビッグデータを分析活用するというスタンスが必要であり、この際『多様な個別化パターンを多数集める』という新しいデータサイエンスの手法が必要となってくる（図11）。

<目的もデータ形式も従来型と違う>



新しいデータ科学の必要性

図 1 1 医療のビッグデータ革命～何が新しいのか～

医療ビッグデータの解析の構成要素としては網羅的分子情報（ゲノム／オミックス）と臨床環境表現型の大量データと、知識発見システム（データサイエンス）が必須となるが、米国では既にいくつかの動きがある（図 1 2）。

- 網羅的分子情報 (Genome/Omics)
 - ゲノム・オミックス情報、multi-omics
- 臨床環境表現型 (Phenotyping)
 - eMERGE-I, PheKB、EWAS
- 知識発見システム (DataScience)
 - Data-mining, Knowledge Discovery, 人工知能

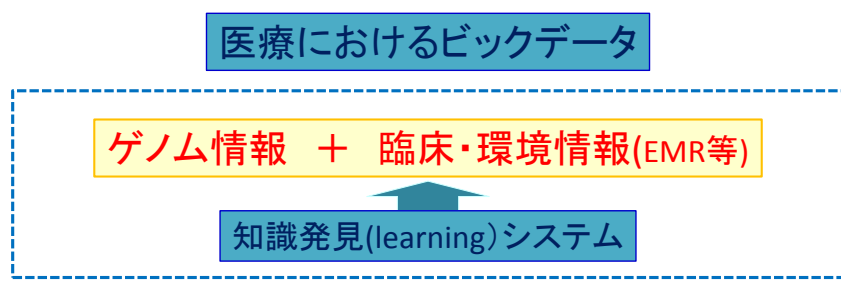


図 1 2 バイオ・医療における Big Data の構成要素

例えば、米国 NIH が推し進める BD2K 施策に関連して、米国臨床がん学会 (ASCO) では CancerLinQ Initiative が設定されている。この CancerLinQ は、診療現場から電

子カルテで大量の診療データを集め、データがバラバラの状態を改善、集合化しデータベースを構築して統合的な分析を行うものである。17 万人規模のがんの症例データベースを作って、各がん種で 1~2 万人ぐらいの症例を集めている。

さらに、学会が主導的に、統計学的学習あるいはニューロネットを駆使して学習（ディープラーニング）していくシステムを構築して診療を行っていくスタンスを取っている。機械学習（ディープラーニング）を用いた‘Rapid Learning’の概念をがん治療において実現しようとしている一例である。既に、CancerLinQ のプロトタイプは完成しており、最も進んでいる乳がんに関しては 10 万人以上の症例が収集されていると報告されている。

また IBM の Watson についてはクイズ番組（Jeopardy）でクイズ王に勝利したことで注目を浴び、AI としての能力を評価されているが、基礎は自然言語理解に応用される統計学的学習理論での機械学習であり、大量の情報により機械学習を繰り返すことによって AI としての機能や活用を拡大している。

Watson には以下の特徴／システムが備わっている。①自然言語処理や大量データベース探索、確信度付回答を行う「Deep QA システム（Jeopardy で活用）」、②質問をシンプルな質問に分解して解答する、MIT（マサチューセッツ工科大学）の START と呼ばれる「オンライン自然言語 QA システム」、③CMU（カーネギーメロン大学）の Open Advancement of Question-Answering Initiative（OAQA）を骨格とするシステム、④質問解答に最も適切なテキスト資料を特定する「知識源拡張アルゴリズム」によるテキストからの知識の自動的な抽出、⑤大量の情報資料の大規模情報抽出、構文解析、知識推論による一般知識情報源への変換。

Watson はいくつかの研究機関と学習プログラムを行っている。以下の研究機関との共同研究が、医療関係の初期の主な研究である。①Memorial Sloan-Kettering Cancer Center (MSKCC) による『患者の最善の治療方針を決定する学習プログラム』、②New York Genome Center による『がん専門医がより良い個別患者ケアを選択する支援ツールシステム』③クリーブランドクリニックによる『Watson を対話的に利用して症例を分析する、問題解決型学習プログラム』④MD Anderson による『治験に適切な患者を診療情報から選別するプログラム』などである。

MSKCC では The Oncology Expert Adviser Software (OEA) を作成し、Watson の計算能力および自然言語処理技術と臨床治験データ（分子・ゲノムデータ、がん病歴の膨大なリポジトリなどを含む）を組み合わせ、個々の患者にとって最高の治療方針を決定するための診断情報や治療の選択肢を見出すことを目標としている。また、New York Genome Center ではがん専門医ががん患者に対してより良い個別ケアを提供できるように支援するためのツールとして、ゲノム研究専用デザインされた試作システムを構築している。最初の対象としては脳腫瘍を選択し、ゲノム配列と医療情報、医療文献から患者毎の治療の提案が行われている。

Watson はデータベースと自然言語処理の機械学習からなる簡潔なシステムである。Watson では、コンピュータで処理しにくい形のデータ（非構造化データ）を数値化するなどコンピュータで処理しやすい形（構造化）にしてデータベースを作るという処理がまず行われる。この非構造化データ（文献データや画像データ）のデータベースを作成できることが Watson の強みである。

自然言語（論文等）処理については、Google なども Word2Vec⁷⁾ というような手法を使って、検索エンジンを使うとユーザーが知りたいことが最初に出てくるという仕組みを持っているが、この処理はコンピュータにとっては難しいため、形態素解析（単語の切り出し）、構文解析（文法）、語義の曖昧性解消（日本語の「青い」など）、照応解析（「人」などの語が指すのがある対照群か一般か等）などの自然言語処理を行っている。

もう一つの要素は機械学習である。機械学習には、教師あり学習と教師なし学習があり、決定木、ニューラルネットワーク、遺伝的プログラム、帰納論理プログラミング、サポートベクターマシン、クラスタリング、ベイジアンネットワークなどが関連する。今はニューラルネットワークのディープラーニングが大流行りであるが、現状の機械学習は全てベースに統計があるため、第 5 世代コンピュータの数学的な知識がコアとなる。機械学習の手法において Watson が特別に卓越しているわけではなく、日本にも機械学習に非常に強い研究者群がいるが、その研究者群の多くは生命科学のデータを扱っていない。

他に、IT の重要な課題として、ビッグデータのストレージ、分散した状況でのクラウド上のデータマッチング（クラウド・ネット）、十分短い時間での結果の返却（実時間処理）、さらには結果の可視化、などがある。これら AI の実現に必要な情報インフラを単品として持っている IT 企業はあるが、システムインテグレーションができている企業は少なく、Watson が他に対して優れているところはこのシステムインテグレーションとデータアクセスの面である。

Watson Oncology Advisor は MSKCC (Memorial Sloan Kettering Cancer Center) の電子カルテ、診療情報について自然言語処理を使って構造化したデータベースであり、機械学習による検索ができる。同様に Watson Discovery Advisor は、Medline や特許のデータを自然言語処理し、構造化してデータベース化し、機械学習によって検索に使えるようにしたものである。Watson Clinical Trial Matching は、Mayo クリニックの持っている Phase1 臨床試験のデータを検索できるようにしてある。Watson Medical Record Advisor は、現状は電子カルテの読み込みを全自動ではできないため、クリーブランドクリニックの研究者と共同研究を行い、クリーブランドクリニックの全ての電子カルテ情報を読み込んで使える形にしたものを販売している。さらに Watson Genomics Advisor は Watson Discovery Advisor に COSMIC、ClinVar、NCI Pathways 等のいくつかのデータベースを相互に関連づけられるようにし、ユーザーの解析対象の遺伝子変異情報を合わせたもので、ユーザーの所有している変異情報によって非常に大

7) 単語をベクトル化して表現する定量化手法。日本人が日常的に使う語彙数は数万から数十万といわれるが、Word2Vec では各単語を 200 次元くらいの空間内におけるベクトルとして表現する。単語同士の類似度や単語の「意味」を捉えられるようになった。

きなシステムができるという発想を持っている。

このように上記のサービスは全て Watson による自然言語処理で、カルテ情報等を読み込んで構造化するプロセスを経て、実用化している。

ベイラー医科大学と IBM による探索研究への Watson の活用事例では、がん関連タンパク質である p53 の活性化と不活性化を導くタンパク質を予測するために、Watson に科学論文を学習させ、p53 に関する 7 万の科学論文の自動分析を実施した。Watson は p53 を修飾する新たな関連タンパク質を 6 つ、数週間で特定することができた。従来のベイラー大学の探索研究では、新たな疾患関連タンパク質の発見は年に 1 つ程度であったことから、AI 創薬による発見のスピードと質の高さに期待が膨らんでいる。

また、薬物や治療法の開発と評価に使用された臨床試験の結果を詳述する科学論文を Watson に学習させ、さらに毒性情報を Watson に学習、解釈、抽出、理解させて、既存薬の別の適応症の候補を絞る際の意思決定支援に用いるといった製薬各社による Watson の活用も発表されており、創薬の現場での活用も開始されている。

その他の AI 活用の動きとしては以下の事例がある。①Cancer Commons Initiative という民間会社ががん患者のコンサルを行う際に、患者データの提供を受ける仕組みに対して'Rapid Learning'を活用するインフラ整備を行っている、②Craig Ventor 氏が立ち上げた Human Longevity Inc. という企業が 7000 億円の初期投資で年間 4 万ゲノムを目標にゲノム情報を集め、健康寿命延長させるためのデータベースを構築しようとしており、これらの取組みに機械学習の専門家を加え、システムを構築している、③Google もプロジェクトを起こして、Conrad 氏を中心に Duke 大学や Stanford 大学の協力を得て、健康に関する尺度発見を行おうとしており、現在 AI を活用して先制医療的なバイオマーカーの探索を進めている。

これらの取組みにおいて、知識発見システムの構築がその研究の中心となっている。AI を用いることで、相互の関係が十分に知られていない生命現象の分子の動きとフェノタイプの発現の関係の解明等につながる可能性が期待されている⁸⁾。

この AI の学習を進めるアルゴリズムがディープラーニングであり、ディープラーニングのゲノム/オミックス医療への応用が研究の焦点となっている。昔は 3 層ぐらいしかなかったニューロネット(図 1 3)を何層もつないで機械学習をさせるという手法で、このような研究を進めるためにはかなり大きな研究費が必要であるが、画像理解などを始め多くの分野で他のアルゴリズムを引き離して優秀な結果が出ている。

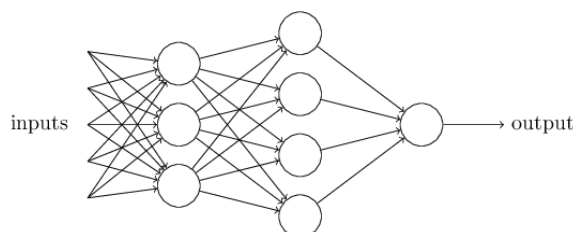


図 1 3 ニューラルネットワークの概念と多層パーセプトロン

8) 医薬産業政策研究所「AI 創薬への動き」政策研ニュース No.47 (2016 年 3 月) 参照

日本では、京大奥野教授を中心に製薬企業 20 数社がコンソーシアムを作り、ディープラーニングによるドラッグデザインが実施されている。コンソーシアムでは、スーパーコンピュータによる計算速度の向上と機械学習の手法を取り入れて、正確性が著しく向上し、かつ短時間判定ができるバーチャルスクリーニング手法である **Chemical Genomics-Based Virtual Screening 法 (CGBVS 法)** を開発している (図 1 4)。更に、機械学習による最適化をコンピュータ自らがデザインする「De Novo ドラッグデザインシステム」により、標的タンパク質名を指定しただけで、活性化化合物の化学構造を自動的に生成する AI の開発も行っている。

他にも、ディープラーニングの手法を用いて、ゲノムシーケンスからエピゲノムを予測するシステムなどの応用例が報告されており、近い将来にディープラーニング技術が創薬標的の特定に活用されることが期待されている。

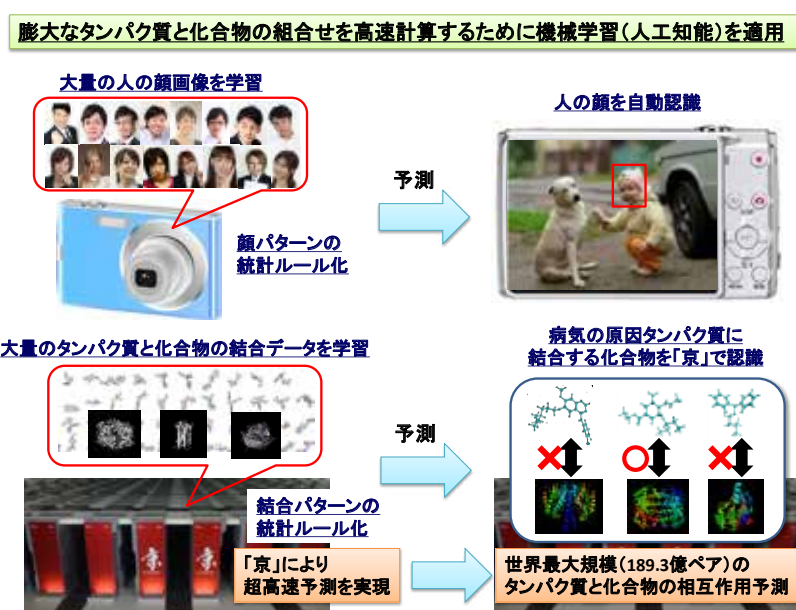


図 1 4 AI 創薬の一例：「CGBVS 法」

AI を活用して、診断 (予防・先制医療)、治療、予後予測などの診療支援を行っていくことが現実的になってきている。例えば、がん患者を例にとると、遺伝的背景や環境背景はそれぞれ人によって違い、がん発症の原因となった分子異常もしばしば異なっている。Precision Medicine の概念で患者の層別化や最適治療の選択する際の判断に AI が活用されるのはもちろんであるが、それぞれ体の状態が変化し、ある人はある時点で、死の階段を登り始め、一方、他のある人はもう少し長い間生きるという臨床上の差が非常に問題になる。そういった状況では、それぞれの患者の変化に関する各イベントの境界点を見極めることが重要である。体の状態の変化を表す検査値のパターンを使って、いろんなシミュレーションを実施することができる。現在の定期的に計測する検査値の

みでは患者の状態変化を的確に予測することはまだ不十分であるが、新たなパラメータを元に患者の状態をシミュレーションするアルゴリズムの開発も行われており、近い将来に患者の予後予測の精度も格段に上がってくる。

健康人の病気発症予測の場合も、がん患者と同じようなロジックで考えられる。定期検診の結果は、長期的な体の状態変化を捉えることに役立つ可能性がある。さらに、生活習慣のパターンに加え、遺伝的背景を考慮することで、ある人はあるときに病気になり、ある人はずっと健康でいる、さらには、この健康と病気の状態の違いが何に起因するのかを突き止めることができる。

また、ヒトの一生を通じたデータ（ライフコースデータ）として、この健康と病気の状態をビッグデータ解析の立場から AI を活用して把握していくことも考えられる。

一生を通じて、健康と病気や死につながる転換期を把握できれば、予防医療や先制医療の効果的な活用が可能となる。これらを実現するためには健康状態である時からの継続的データの取り方を工夫する必要がある。ウェアラブルデバイス等の普及が一助となることは間違いないが、ウェアラブルデバイスを着けてデータ計測することにインセンティブを付与するなどの工夫を講じる必要がある。また体の状態が変化する境界点の前後における生検サンプルを取得し、分子レベルの情報を入手できるようにしておくことも必要である。このようなデータは継続的なデータが必要となるため、膨大なデータ量となり、スーパーコンピュータや AI の活用が必須となる。

ビッグデータの解析や判断における AI の役割は、大きく二つに分かれる。一つは人間の判断を支援する AI であり、もう一つは人間を超えてより高度な判断を行う AI である。人間の判断を支援する AI は Google などキーワード検索し、何かのヒットを見るのと類似した位置付けであり、ヒットで知識を得たあと、それを頭の中で解釈し、使う人間が結果に基づく予想をする必要がある。つまり、そのタイプの AI というのは予測ではなく、関係性のある答えを優先順位をつけて提案しれくれるものであり、このビッグデータの解析を予測につなげるためには、ヒットの結果から何らかの予測の方法論を作っていくか、シミュレーションを用いた強化学習などにより人工知能の解析能力を上げて、確度の高い予測ができるようにしていかなければならない。

碁の世界で人間のプロ棋士を負かして、人間を超える高度な判断を行える事例となった **alfaGo** はディープラーニングとその後の強化学習（AI 同志の対戦と学習）で、瞬間に人間を超える能力（棋力）を持った。このタイプの AI は、多くの情報、知識を習得し、さらに新たな情報を継続的に習得し続けることにより、高い確度の予測・判断が可能となる。これからの医療にパラダイムシフトを起こす可能性を示唆しており、医学研究や創薬などの研究にも大きな革新をもたらすことが期待される。

ゲノム／オミックスビッグデータを中心とした創薬

ゲノムワイド関連解析 (GWAS) を代表とする遺伝統計において、重要なのは、単に有意差検定を行い、遺伝子の関与の有無を見るのではなく、そのゲノム解析の成果を活用して疾患病態の解明や新規創薬につなげることであるが、膨大な解析データがある中で、残念ながら最終の成果につながる研究は多くはなかった。

しかし、バイオリソースの各種データの蓄積は、ここ数年指数的な伸びを示しており、生物学や医学や創薬に関する多彩なデータベースが存在する。GWAS 等による大規模ヒトゲノム解析のため、ゲノムを中心として、SNP や遺伝子のデータベースを中心として、エピゲノム修飾、RNA やタンパクの発現量、あるいはネットワーク解析、ノックアウトモデルや創薬にかかわるデータなど多くのバイオリソースデータを駆使すれば、データ間の関係を見出すことができ、疾患病態解明や創薬へと結び付けることができる可能性が高い⁹⁾ (図 15)。

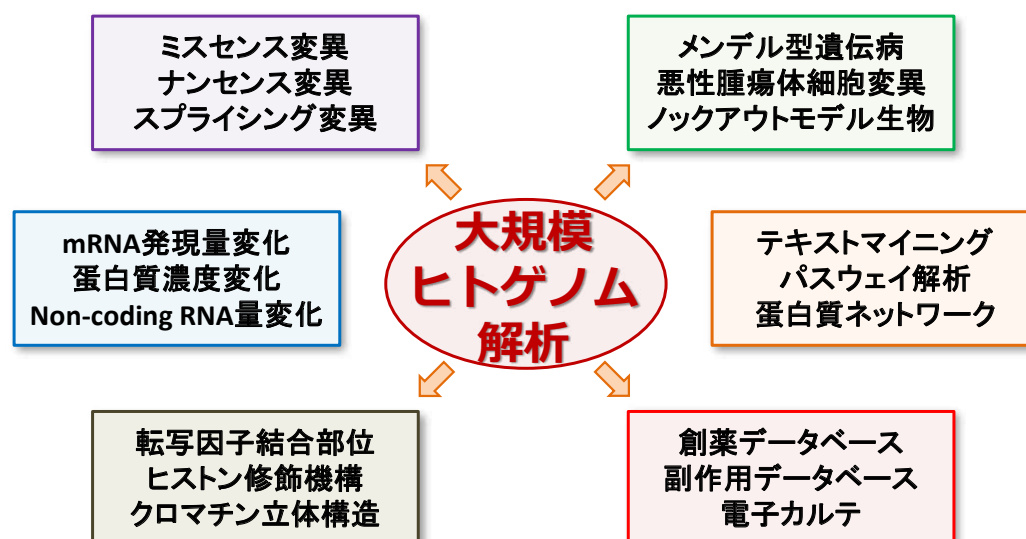


図 15 ヒトゲノム解析から疾患病態解明・新規創薬へ

例えば、これまでに欧米を中心に、肥満関連の遺伝子が多く見出されている。この肥満患者の大規模ゲノム解析データと組織特異的エピゲノム解析 (ENCODE プロジェクト) データを統合させて解析することにより、肥満を決める遺伝子変異は中枢神経組織の遺伝子発現領域に含まれていることが報告されている。つまり、摂食行動をきちんと制御できないために肥満となる「脳疾患としての肥満」が明らかになっている¹⁰⁾。

また、関節リウマチについては、およそ 100 個の関連遺伝子を分類すると、6 割ぐらいはミスセンス変異 (アミノ酸配列が変わる) や免疫細胞特異的な cis-eQTL (mRNA の量が変わる) が起こっており、特定の細胞でタンパク質の発現量が変わることが疾患発症の一因だということが報告されている¹¹⁾。併せて制御性 T 細胞の関与が重要で、

9) 医薬産業政策研究所「ビッグデータ時代のゲノム解析を中心とした病態解明・新規創薬— 遺伝統計学への期待 —」政策研ニュース No.48 (2016 年 7 月) 参照

10) Locke AE, et al. Genetic studies of body mass index yield new insights for obesity biology Nature 518, 197–206, 2015, Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics 47, 1228–35, 2015

11) Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–81, 2014

そのヒストン修飾機構の異常も疾患の発症リスクであることも確認されている。さらに、がんの原因遺伝子のデータベースを用いて、関連性を調査した結果としては、固形がんのリスク遺伝子には関節リウマチと共通するものがないものの、血液がんでは共通しているものがあることが示されている。以前から、関節リウマチ患者は非患者に比較してリンパ腫になりやすく、関節リウマチとリンパ腫に同じ薬が効くということも認知されていたが、原因遺伝子レベルで遺伝子ネットワークを共有していることが要因であったことが改めて確認されたことになる。

このように疫学についてもゲノム解析データが研究のヒントを与えてくれることがある。同様に関節リウマチの事例として、数十年前から疫学的に認識されていた関節リウマチ患者は統合失調症になりやすく、逆に統合失調症患者は関節リウマチになりやすいことをゲノムレベルで裏付けようと、それぞれ数万人レベルの統合失調症と関節リウマチの GWAS の比較が行われた。その結果、それぞれの疾患で、相手の疾患への成り難さをゲノムレベルで持っていることが確認されている¹²⁾。

疾患へのなりにくさはゲノムパターンだけですべてを説明できるわけではないが、今まで疫学で得られていた知見に対して、ゲノムである程度の説明ができることが示された。さらに、このような疾患間の遺伝子ネットワークの共有がわかったことにより、統合失調症の発症原因の探索も進められた。つまり、複合した発症原因の要因の一つとして、免疫反応の関与を推察することができた。このようにゲノムと疫学を組み合わせることによっても、新しい発見ができる。

上述したようにヒトゲノム解析のデータを中心に、臨床データや医学、生物学のデータ、医薬品関連のデータなどとの統合的な研究を進めると、疾患病態の解明や創薬につながる可能性がある。従来、創薬ではマウスなどの実験動物のデータから、ヒトへの外挿性を予測したものの、治験で効果がなかったというような失敗を数多く経験しているが、これからはヒトのゲノム解析を活用した効果の予測も考えられる。

数千人と単位でヒト集団のゲノムの変異を捉えると、その遺伝子の機能を劇的に落とすような変異を持っている人が見つかる。これは遺伝学で理論的に担保されており、このような状態を LOF (Lose of function variant : 機能喪失型遺伝子多型) という。これをノックアウトマウスに代わるノックアウト人間と考えることができる。

つまり、そうした人を探してきて、何が起きているかを調べれば、遺伝子機能の変化が起きた時にヒトで何が起きるかということを推察できる。このように、倫理的に不可能なヒトへの介入実験を、バーチャルに観測ができるということが示され、「ヒト集団こそがベストのモデル生物」という考えが提唱されている。

今までの創薬はまず疾患を理解し、そのメカニズムを解明して、それに効きそうな化合物を目利きで探していく形で行われてきた。このアプローチは、かつてはうまく新薬発見に繋がっていたが、近年では難しくなっている。

それを打破する目的で始まった従来のゲノム創薬は、ゲノムから対象とする疾患の発

12) Lee SH, et al. New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis. *Int J Epidemiol* 44, 1706-1721, 2015

症メカニズムなどを理解した上でその疾患の治療薬のターゲットを発掘し、ターゲットに作用する薬剤シーズを見つけ、その最適化を行うというアプローチであった。この場合、研究の途中で疾患の発症原因をどう考え、標的として何を選択するか、さらに標的に対して化合物を最適化できるか、のそれぞれにハードルがあり、最終の製品にたどり着かないということが大半であった。つまり、疾患を理解する道具としてゲノムを使おうというアプローチが創薬までのハードルを上げていたと考えられる。

これからのゲノム創薬では、疾患はあくまでもサンプルとして、そのサンプルから因果関係が強いゲノムの情報を入手し、そのゲノム情報に基づいた標的を選択して、薬の探索を行うことを手始めに行い、結果的に疾患の治療薬になるかどうかをバーチャルに確認する、という姿を目指すべきかもしれない（図16）。つまり、疾患の完全なメカニズムの理解を目指すのではなく、疾患特有のゲノム情報からシンプルにその遺伝子変異に対する薬を見つけていくことが、これからのゲノム創薬となる可能性が期待できる。

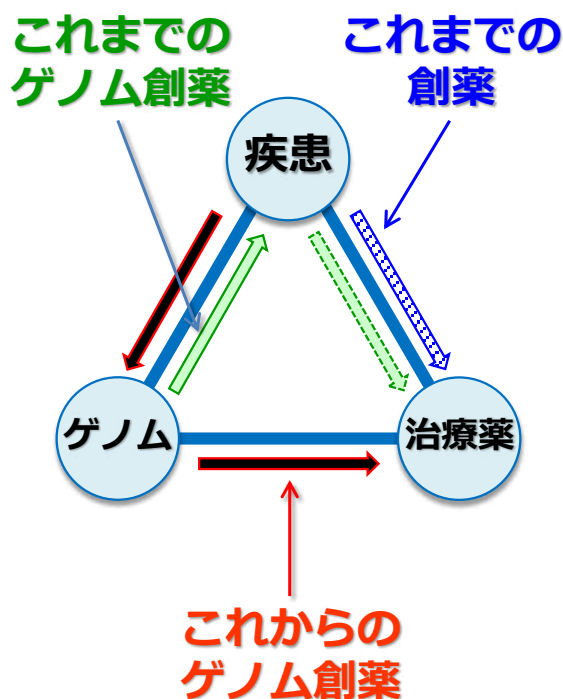


図16 ゲノム創薬の新しいモデル

例えば、家族性 LDL 血症という生まれながらに LDL が高い家系のゲノム解析の結果から、PCSK9 の遺伝子の機能亢進変異によって LDL が高値になるということが知られていた。また、PCSK9 の機能低下変異を持っている人では、LDL が低くなり、かつ機能低下変異により大きな別の障害が発症しないことが確認されたことから、この PCSK9 阻害薬が抗コレステロールの薬として開発され、現在大型製品になりつつある。同様にキサンチン尿症を発症しやすい家系から、XOH という感受性遺伝子が発見され

た。これも XOH の機能低下変異を持つ人が別の大きな障害を発症していないことが確認され、その阻害薬からフェブキソスタットという抗痛風治療薬が開発され大型製品となっている。

これらは、遺伝子の機能を抑えた時に何が起きるかということそのゲノムの変異を持つヒトを使って予測した事例で非常にうまくいった例である。しかしどちらもメンデル型希少疾患であった。一般に、一つの遺伝子がいろいろな機能をネットワークとしてカバーしており、阻害剤を作っても、他の疾患のネットワークに影響を与え、副作用が出て、薬にならないという例も多い。先の 2 例の成功例は希少疾患で、他の疾患ネットワークとの関与がなかったということが幸運であったが、この希少疾患で成功したゲノム創薬の新しいモデルをコモンディジーズ（一般的な疾患）の領域にまで広げていくということが、これからの方向性としては重要である。

コモンディジーズとしての関節リウマチを対象として、このゲノム創薬の新しいモデルが適応できるかについて、既存の創薬のデータベースなどを用いて検討が行われた。その結果、関節リウマチのゲノム解析結果により知られていた感受性遺伝子と現在用いられている関節リウマチ薬の標的遺伝子と比較することによって、全ての薬の標的遺伝子は感受性遺伝子と繋がっているということが確認されている（図 17）。

このような疾患ごとのゲノム解析結果と治療標的遺伝子のネットワークの作成により、ドラッグリポジショニングが進むことも期待できる。他にも、抗悪性腫瘍薬剤である GSK3B 阻害剤が 2 型糖尿病に対しても治療薬としての可能性があることが報告されている¹³⁾。

これらの事例によりゲノム創薬の新しいモデルがコモンディジーズへも適応できることや、ドラッグリポジショニングに寄与することが示唆されている。

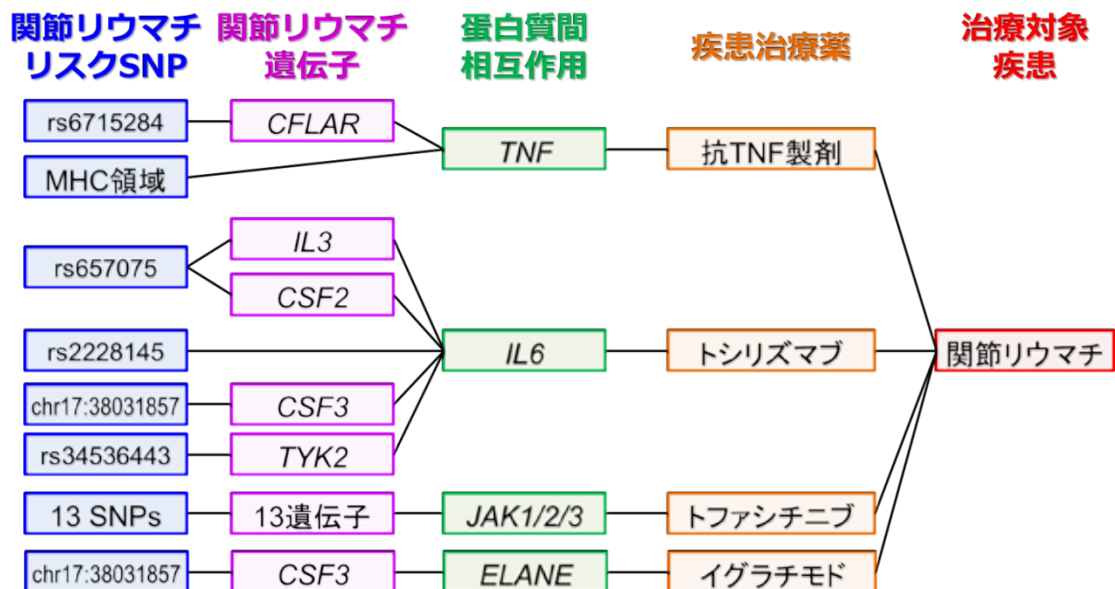


図 17 関節リウマチ感受性遺伝子と治療薬標的遺伝子

13) Imamura M, Okada Y, et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes Nature Communications 7, 10531, 2016

薬剤投与時の遺伝子の発現を見ることによって創薬への活用ができる。よく知られている活用手法は遺伝子発現プロファイルによる有効性予測、毒性予測である。これには、薬剤投与による遺伝子プロファイルの変化を見た Massachusetts Institute of Technology (MIT) の Broad Institute の Connectivity Map (CMAP) や、疾患特異的遺伝子発現プロファイルの変化を集めた Gene Expression Omnibus (GEO) を用いる遺伝子発現シグネチャ逆位法がある。

この遺伝子発現シグネチャ逆位法は、遺伝子発現プロファイルの変化を分子ネットワークの変化ととらえて、疾患特異的な遺伝子発現シグネチャと薬剤特異的な遺伝子発現シグネチャが負に相関した際に疾患の治療薬になる、という仮説を生成する考え方である。逆に、正に相関する場合には、薬剤を投与した際に起こる副作用をその疾患と類似の症状として予測できる可能性がある。このように、CMAP と GEO を用いて解析することによって、どの薬がどの疾患に効くかが予測できる。これはドラッグリポジショニング (DR) にも応用できる (図 1 8)。

• 遺伝子発現シグネチャ逆位法 (signature reversion)

- 薬剤特異的遺伝子発現シグネチャ
- 疾患特異的遺伝子発現シグネチャ
- **有効性予測**: 両者が負に相関する
- Non-parametric な相関尺度で評価
 - Gene Set Enrichment Analysis (GSEA): ES score
 - 対照と比較して順位づけられた遺伝子リストの上位に密集しているかの尺度
- 例: 炎症性腸疾患IBDに抗痙攣剤(topiramate), 骨格筋委縮にウルソール酸

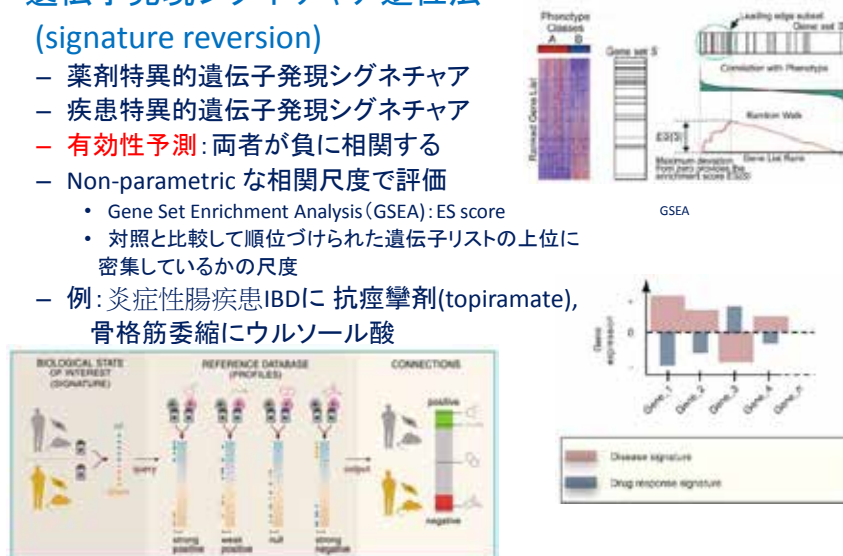


図 1 8 遺伝子発現プロファイルによる有効性予測

一方、ランダムにどの薬がどの病気に効くかどうかを予測するには、Diseasome¹⁴⁾ という考え方がある。これもドラッグリポジショニング (DR) に応用されている。

Online Mendelian Inheritance in Man (OMIM)¹⁵⁾ から疾患とその原因遺伝子を抽出し、A 疾患の原因遺伝子と B 疾患の原因遺伝子の共有遺伝子からネットワーク (疾患ネットワーク) を作る。この疾患ネットワークの近くの疾患に対しては、同じ薬が効く可能性が高いのではないかと予測するのが第 1 世代型の Diseasome からの DR である (図 1 9)。

14) genome+ome で Genome, protein+ome で Proteome であるのと同様に disease+ome で表される網羅的な疾患ネットワーク

15) 遺伝子変異による疾患に関する情報データベース

- OMIMから 1,284 疾患と 1,777 疾患遺伝子を抽出
- ヒト疾患ネットワーク(HDN)
 - 867疾患は他疾患へリンクを持つ 細胞型や器官に非依存
 - 516疾患が巨大クラスターを形成
 - 大腸がん、乳がんがハブ形成
 - がんはP53 やPTENなどにより最結合疾患 がんなどは後天的変異
 - 疾患を網羅的に見る見方:臓器や病理形態学に非依存
 - リンネ(12疾患群分類)以来300年続いた分類学を越える
- 疾患遺伝子ネットワーク(DGN)
 - 1377遺伝子は他遺伝子へ結合
 - 903遺伝子が巨大クラスター
 - P53がハブ
- ランダム化した疾患/遺伝子ネットワークに比べ
 - 巨大クラスターのサイズが有意に小さい
- 疾患遺伝子は機能的なモジュール構造
 - 同じモジュールに属する遺伝子は相互作用し
 - 同一の組織で共発現し、同じGOを持つ

図 1 9 第 1 世代型 Diseasesome

これに対して、疾患特異的な遺伝子発現パターンを並べて、臓器別、あるいは器官別ではなくて、発現パターンの類似性から疾患体系をくくり直す考え方が第 2 世代型の Diseasesome である。ゲノム/オミックスから、疾患の同一性を作り直すと、例えば、心筋梗塞とデュシェンヌ型ジストロフィーが慢性増殖性炎症の遺伝子の発現パターンとして近い疾患ということになる。このように疾患の近似性から分類を行った上で、第 1 世代と同様に、Diseasome 上で近い疾患の薬が効くかもしれないという仮説から既存薬剤の適用拡大を検討するのが第 2 世代の DR である (図 2 0)。

Gene-Expression Nosology of Medicine

- 疾患を平均遺伝子発現パターンよりクラスター分類
 - 臓器別疾患分類では予想できない疾患間の親近性
 - 分類項目はサイトカインの遺伝子発現と相関
 - 疾患の再体系化に基づいた医薬の repositioning
- さらに656種類の臨床検査を結合した分析
- 心筋梗塞・デュシェンヌ型筋ジストロフィーが近い

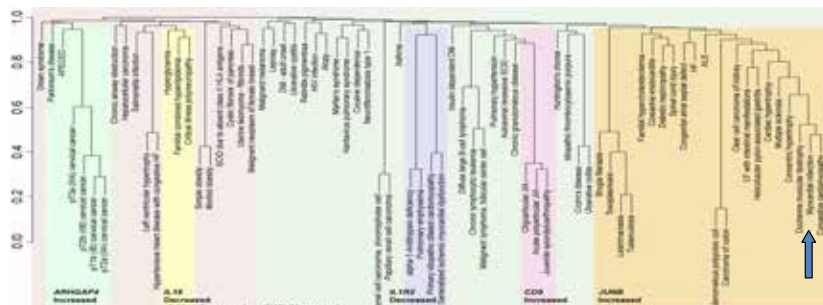


図 2 0 第 2 世代型 Diseasesome

さらに、第3世代は、Protein-Protein Interaction (PPI) のネットワークを 4620 の Module に分けて、病気でどの疾患ネットワークが発現しているかを見ることで、PPI の類似性で疾患ネットワークを作成する考え方である。これにより、マラリアの薬とクローン病の薬が実は同じ病気に対して効くということが見つけられた実績もある (図 21)。

Transcriptomeの変化をPPIに投影した 疾患ネットワーク(Butte)

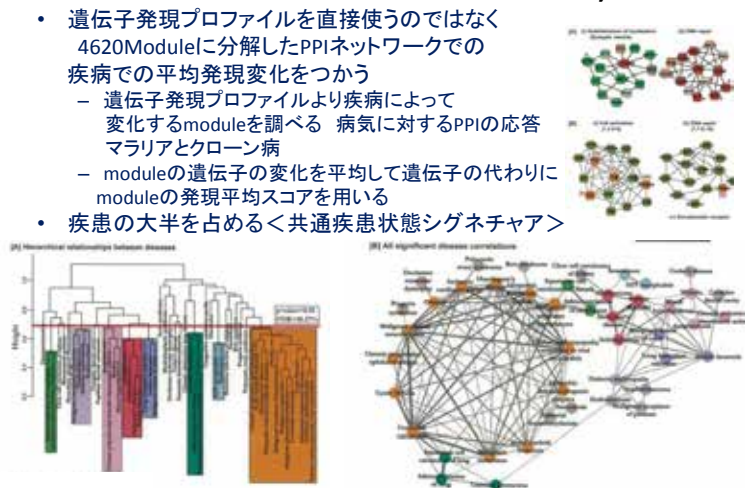


図 2 1 第3世代型 Diseaseome

欧米では、大規模ヒトゲノム解析を中心として、オミックスやエピゲノム、エクソソーム、バイオームなどの生命分子情報と疾患との関係を探る研究が急速に進展している。優秀なゲノム研究者は既に実用化の時代に入ったと、アカデミアの研究施設から製薬企業にスピンアウトとするものも多い。米国などでは、そうやってゲノム創薬を経験した研究者が再びアカデミアに戻るなど、産学の距離が非常に近いことが強みである。また、製薬企業が行う臨床試験についても、ゲノムを一通り集め、並行してゲノム解析を行ってデータベース化している。さらに、国際的なゲノム研究プロジェクトに、資金だけでなく自分たちが保持するゲノムデータも出して参画している。これにより、数年前からプロジェクトの解析結果を知り、また最新トレンドの解析方法についての情報収集を行うなどのメリットを使って、少しでも早く新しい薬を開発しようとする姿勢がうかがえる。

このように、ゲノム創薬を何となく絵空事で話していた時代から、既に実際にできるのではないかという時代に移りつつある。

まとめと今後の課題

米国では、クリニカルシーケンスが普及し、実臨床でのゲノム／オミックス情報の蓄積、活用、医療情報データベースの整備、2次利用が始まっている。Precision Medicine Initiative においては、がん領域のゲノムドライバ特定への取組みを短期目標としつつも、多くの多因子疾患に対して患者のサブグループを探索する中で疾患と生体分子等の関与因子との関係を明確にしていくという目標を掲げている。また、eMERGE プロジェクトや BD2K Initiative を通じて、ゲノムなど生体分子情報を電子カルテに取り込み迅速な医療実践につなげていくための方策や、そこで得られるアウトカムデータを用いた医療の改善も、同時に進める体制を構築しつつある。そういう基盤整備の進展の中で創薬に向けたビッグデータの活用という点でも、産官学連携が進んでいる¹⁶⁾¹⁷⁾。

一方、日本でのゲノム／オミックス医療の実装は、いくつかの医療機関で試行されているが、まだ研究の色合いが強い。政府における検討も、ゲノム医療推進協議会や次世代医療 ICT 基盤協議会などで進められているが、「ゲノム／オミックス」と「レセプト／電子カルテ」の活用が個別に議論されており、ICT を活用し、ゲノム情報のみならず、より詳細なオミックス情報を電子カルテへ取り込み、ゲノム／オミックス医療の早期の臨床実装を目指す動きが求められる。

バイオバンク／コホートに関して、米国では、Precision Medicine Initiative で 100 万人の疾患ゲノムコホートからデータを取り、新しく創設するバイオバンクでゲノム／オミックス情報の統合データ化を目指している。欧州では英国を始め、バイオバンク／コホート研究に力を入れている国が多く、BBMRI (Biobank/Biomolecular Research Infrastructure) という欧州各国の 250 以上のバイオバンクを統合した組織を持っている。欧米各国がこれからのゲノム／オミックスの医療活用の際に重要な基盤整備として積極的な取組みを行う中、日本もヒト組織の利用が進まないなどの課題はあるもののこのバイオバンク／コホートの取組みという面においては欧米の活動から大きく遅れてはいない。

近年、医療に関連するデータ量が指数的に増大しており、リアルワールドデータからエビデンスを見出していく期待が膨らむ中、データから意味のある知見を見出す作業は、困難を増している。そのような中で、ディープラーニング技術による AI の発展が、患者の様々な種類の膨大なデータを短時間に分析して、診断や治療、予後判断を行う際、さらには新たな治療法（疾患メカニズムや創薬ターゲットを含む）の発見に対して、大きな支え手となってきている。実臨床での生体分子情報や診療情報の解析において、人間の処理能力を超えるデータ解析を短時間で処理し、今後の Precision Medicine を進展させるためには、AI 活用は必須でありキーとなる。

Precision Medicine の実践やそのためのデータベース整備、AI の普及などによって医療はパラダイムシフトが起こるとともに、臨床の広範なデータの活用で創薬研究のス

16) 医薬産業政策研究所「米国の Accelerating Medicines Partnership」政策研ニュース No.46 (2015 年 11 月) 参照
17) 医薬産業政策研究所「欧州の Big Data for Better Outcomes」政策研ニュース No.47 (2016 年 3 月) 参照

マイルも大きく革新することが想定される。多様なヒトのゲノム情報やフェノタイプとの詳細な解析により、ヒト独自の創薬標的の発見が進むことが期待される。

これからの創薬は、ヒトゲノム等の生体分子の解析のデータを中心に、臨床データや医学、生物学のデータ、医薬品関連のデータなどとの統合的な解析を進めることにより、疾患病態解明や標的分子発見等がさらに進展することが期待される。

実臨床やコホート研究、疾患レジストリなどのビッグデータ解析の意義はこれから益々高くなる。リアルワールドデータに基づくエビデンス創出が望まれ、既に BDBM (Big Data Based Medicine) という新しい概念が提唱されている。国民の医療健康、そして関連する産業の技術革新を進めていくためにも、我が国においてもこの BDBM ができるインフラ整備を進めていく必要がある。

【日本の課題①】 クリニカルシーケンスの臨床実装のスピードアップ

今後、日本においても、米国で提唱された Precision Medicine を進展させていくことが必要であり、この医療を実践していくためにはクリニカルシーケンスの臨床実装が必須である。米国と比較してクリニカルシーケンスの体制は非常に遅れており、この基盤整備のスピードアップが必要である。現在、内閣官房のゲノム医療実現推進協議会やタスクフォースでの検討なども活発化しているが、クリニカルシーケンスを始め、ゲノム/オミックス研究の実臨床での実装の促進についてはまだまだ十分ではない。

網羅的なクリニカルシーケンスは、日本の医療保険制度の中では馴染めず、研究費もしくは患者の自由診療の範囲で行われている現状がある。特区や特定のプロジェクト等を利用するなど、ゲノム/オミックスの実臨床における実用化を具体的に検討する必要がある。また、医療保険制度の中でのクリニカルシーケンスの適用の検討についても早急に開始すべきである。

【日本の課題②】 ゲノム/オミックス情報を含めた医療関連情報の標準化と統合

米国では昨年年初のオバマ大統領の一般教書演説で Precision Medicine Initiative が提唱される 8 年前より eMERGE プロジェクトが開始され、電子カルテと遺伝子情報の統合活用のための検討が進められていた。このことが、米国におけるゲノム/オミックス医療の実装の推進に対して一つのアクセラレーターとなっていた。

日本における現状のカルテの ICT 化や標準電子カルテの議論の中には、ゲノム情報の付加についての検討はほとんど入っていない。「日本版 eMERGE」を開始するなど、ゲノム/オミックス情報の電子カルテへの取り込みも踏まえた臨床表現型情報の標準化と統合にも注力すべきである。

【日本の課題③】 バイオバンク／コホート、疾患レジストリの更なる活用

ゲノム／オミックス医療の実装が遅れている日本では、当面はバイオバンク／コホートや疾患レジストリの医療情報の ICT 化を進め、フェノタイプ情報と統合化されたゲノム／オミックスデータの活用により、臨床エビデンスを蓄積していくことが望まれる。このコホート研究や疾患レジストリの取り組みについては、患者の診療アウトカム情報や多様な環境情報、継時的な検査指標といった Precision Medicine の実践に資する情報等を更に取得できる体制も望まれる。

また、例えば Clinical Innovation Network (CIN) 構想における疾患レジストリの対象患者はゲノム／オミックス情報を取得するという運用があれば、患者の層別化を更に的確にして効率的な治験被検者の抽出が行え、また治験後のフォローを行うことにより、有益な臨床エビデンスが入手できることが期待できる。

このようなコホート・疾患レジストリの一定の範囲に、試験的なゲノム医療の運用といったインフラを国が整備することで、臨床エビデンスの蓄積を進め、また治験においては“Registry-Based Randomized Clinical Trial¹⁸⁾”により、リアルワールドデータを反映した評価を行うことなども可能となる。

【日本の課題④】 医療・創薬分野における AI 開発

患者の診断、治療、予後予測の支援のための AI について、医療健康分野での開発や普及を図り、臨床実装のデータをフィードバックすることにより、AI 機能・性能の改善を進める必要がある。日本における AI 研究については、産総研に人工知能研究センターが、理研に革新知能統合研究センターが設立されるなど、ここへ来て活性化しているが、医療・創薬分野における AI 研究をさらに発展させるためには、ハード、ソフト（アプリケーション）、データのそれぞれの面で AI 適用を促進する取り組みが必要である。その上で、産学連携、他業種連携により、インプットするデータを効率よく収集し、アウトプットされた結果を評価・フィードバックすることで、AI アルゴリズムを洗練していくサイクルを回すことが望まれる。

その過程において、Google, Amazon, Facebook, IBM などと 1 社で太刀打ちするのは難しく、オールジャパンで、IT 関連業種と医療・製薬関連業種、アカデミアが連携し、コンソーシアムを形成するなど工夫して AI 開発を進め、同時にデータサイエンティスト人材の育成も積極的に進めるべきである。

【日本の課題⑤】 医療ビッグデータを活用する研究プラットフォームの構築

現在、政府を中心に進められてはいるが、医療・ゲノム情報の 2 次利用や産官学共同研究といった研究プラットフォームの構築と推進について、標準化の推進や基盤データの蓄積・データベース化をさらに強力に進めていくべきである。例えば、難病・がん等のレジストリにおけるゲノム／オミックス情報の蓄積とデータベース化、また AMED

18) Michael S., et al. The Randomized Registry Trial — The Next Disruptive Technology in Clinical Research? N Engl J Med 369:1579-1581, 2013

を中心としたバイオバンク事業でのゲノム／オミックス研究の推進と基盤データベースの構築は必須である。

このような標準化の推進や基盤データの整備の上で、医療 ID の付与、代理機関の創設等による二次利用の促進が期待されている。そのために電子カルテとゲノム情報の統合化等の推進を行い、個人情報保護法などを適切に設定すること等により、必要な法・制度の整備をすすめ、有意義な医学研究、ゲノム研究が促進される環境構築がなされることが期待される。

これらの課題の解決により、ゲノム／オミックスを含む医療ビッグデータの蓄積と活用が臨床実装され、医療のパラダイムシフトが進むとともに、疾患原因を始めとするエビデンスの構築や創薬等の研究の 2 次活用も進むことを期待したい。

【製薬企業への期待①】創薬研究における医療データ活用の活発化

医療健康分野のビッグデータを活用した創薬に向けて、製薬企業は、今ある臨床、創薬データの活用を活発化するべきである。ビッグデータ活用の波は既にくており、一般に公開されているデータベースも国内外に広く存在する。医療健康分野のビッグデータがいつ使えるようになるのかといった待ちの姿勢ではなく、他分野で進むビッグデータに関する解析ノウハウの蓄積に遅れることなく、製薬企業においても創薬に向けた医療データの活用をさらに活発化すべきである。

製薬企業では、特に創薬研究分野において、医療データに対する信用は実験データよりも低いという価値観があるが、このマインドセットを変えて、企業が蓄積している臨床データや医療データベースなどの探索研究への積極的な活用を検討していく必要がある。

【製薬企業への期待②】医学薬学の着想を入れたビッグデータ解析の推進

ビッグデータ解析に医学薬学の着想を入れて製薬企業の強みを生かす取り組みを進めていくべきである。そのためには、統計的なアプローチだけでなく、疾患と分子のネットワークやパスウェイ、代謝システム等々の分子生物学や医学、薬学などの知識に基づいた解析の着想が、他者との差異化ポイントとなる。

上述した AI 研究においても、アルゴリズム開発同様に、インプットデータとアウトプットされる結果の評価が重要であり、どのようなデータを使って結果を出し、それをどう評価するか、製薬企業ならではの強みを加えたビッグデータの活用があるはずである。そのためにも、まず、ビッグデータ解析を実践してみることが必要である。

【製薬企業への期待③】オリジナルデータを用いたデータ解析の推進

NDB 等の国のビッグデータや医療機関、バイオバンクといった外部のビッグデータを用いた連携がまだ難しい中、企業内に蓄積しているデータの活用を目指すことを推奨したい。

ビッグデータ活用の強みとなる重要なポイントは、オリジナルデータを用いて独自のアウトプットを出すことである。各社が持つ既存あるいは今後取得する臨床研究、臨床試験、市販後調査データの解析も一つのパイロット研究である。その中で、独自データを生かす、もしくは生かせる仕組み、例えば同意の取得の仕方の整備などをしていく必要がある。

以上の取り組みを進めることで、製薬企業内でのビッグデータの活用によるノウハウの蓄積と人材育成が進み、ひいては研究環境の進展により、創薬標的をはじめとする新たな知見の発掘と、日本における創薬の活発化につながることを期待したい。