

AI×バイオ

研究開発の俯瞰と最前線

2020年9月28日～11月30日

ライフサイエンス・臨床医学ユニット 島津博基



国立研究開発法人科学技術振興機構 研究開発戦略センター
Center for Research and Development Strategy Japan Science and Technology Agency

なぜバイオにAIとデータを用いることが今後の主流になるか

- 次世代シーケンサーの進展によるゲノム解析の高速化・低コスト化
- 技術の進展により計測機器やセンサなどからデータがハイスループットに排出される時代に
- AI・機械学習の進化により、科学的発見（従来の方法論、人間の認知能力では見つけれなかった発見）の可能性が拡大

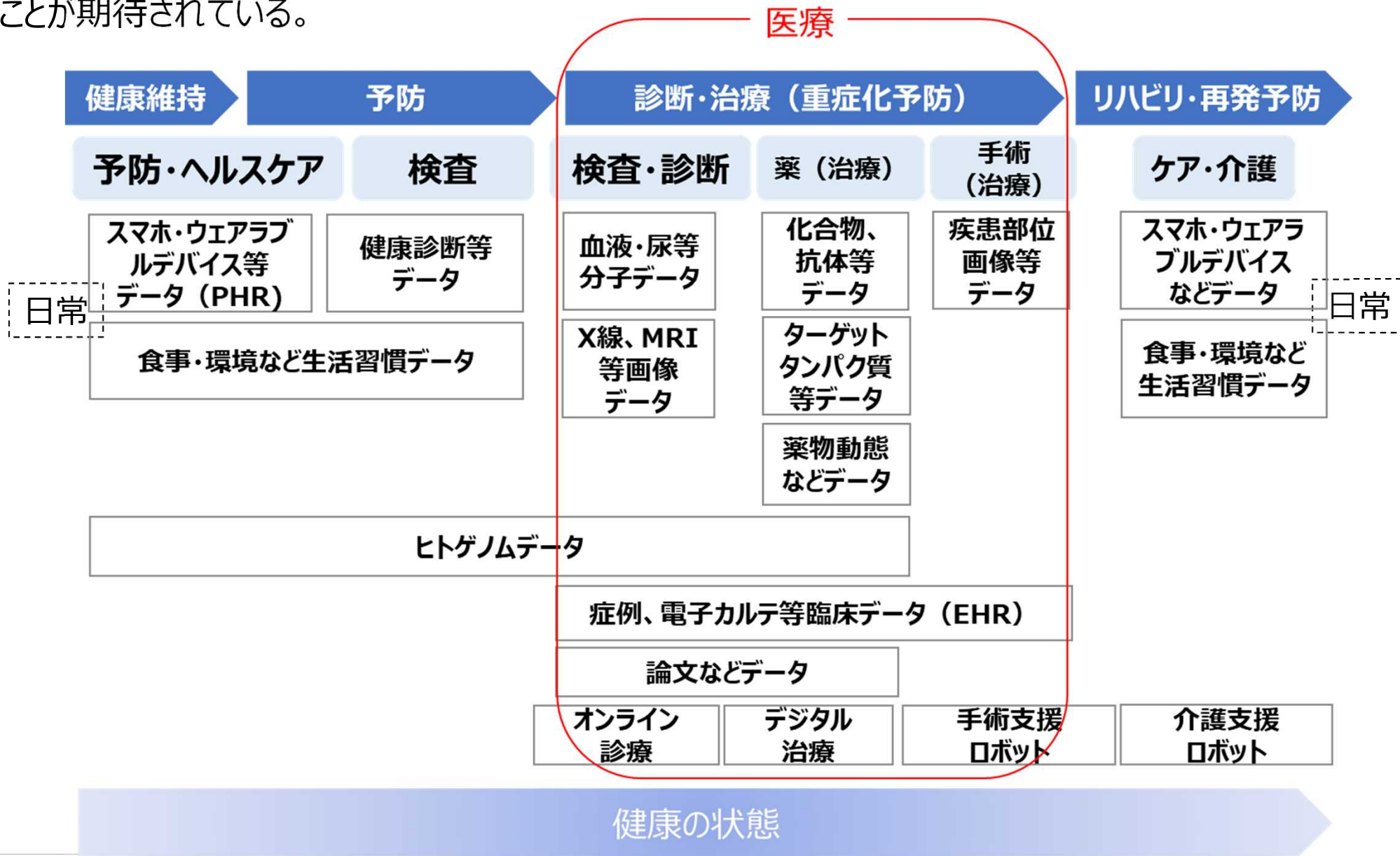
技術の早い進展が社会・経済や科学の要請と合致

- ヘルスケアは画一から個別へ、治療から予防へ
- 創薬のコストと時間の短縮（新薬創出確率の上昇）へ
- 生命は個別因子から複雑系システムの理解へ

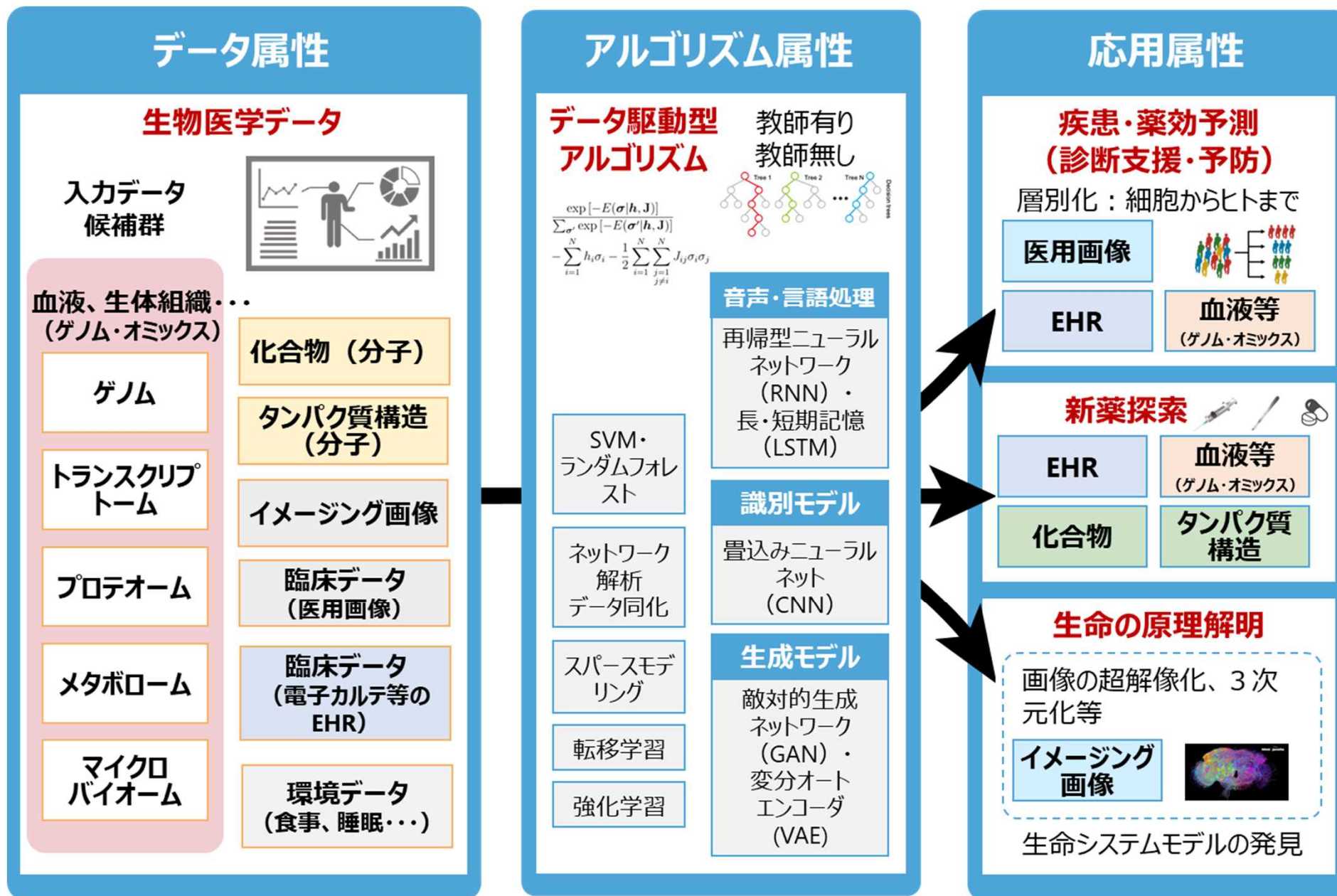
生化学・分子生物学のような決定論的な要素還元的手法と
量子化学、AI・機械学習のような確率論的な複雑系を対象とした手法の
両輪で生命科学・医科学関連研究は進展していく時代に。

ヘルスケアデータの活用

日常生活から研究機関、医療機関にわたる様々なデータが疾病の予防や早期診断等に用いられることが期待されている。



AIを用いた研究の全体像



技術的な背景・経緯

識別モデル

- **2012年**、「畳み込みニューラルネットワーク（CNN）」が画像認識で従来手法と比較して圧倒的精度を発揮（ILSVRC）。
- 2015年、人間による画像認識のエラー率を下回る（ILSVRC）。
- 2016年、17年に網膜や皮膚がんの画像にディープラーニングを用いた論文が出て高被引用に。
- **2018年**、米国FDAは糖尿病網膜症を検出するAIを用いたデバイスを医療機器として初めて承認。

生成モデル

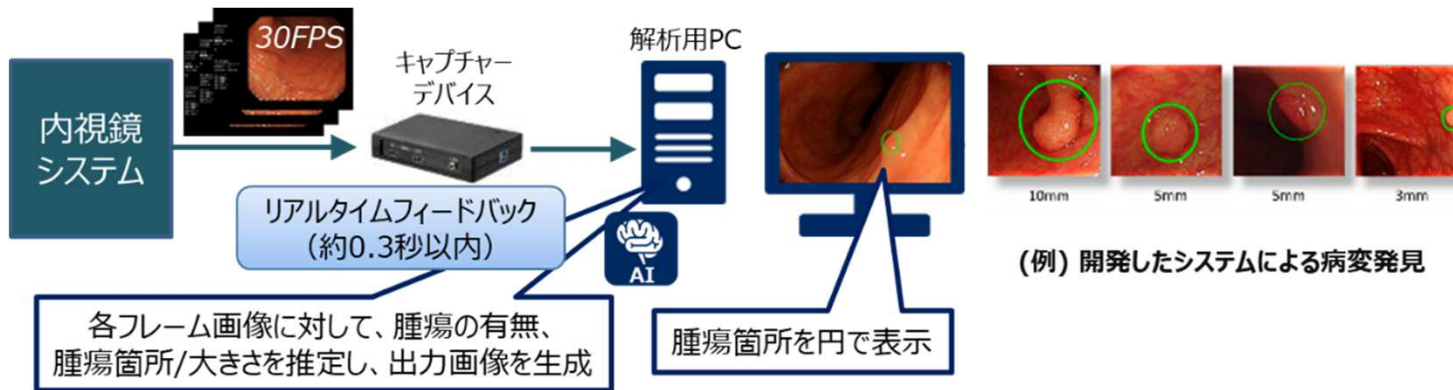
- **2014年**、画像生成のためのアルゴリズム「敵対的生成ネットワーク（GAN）」発表。
- 2016年頃から、世界中でAI（創薬）ベンチャーが続々と大手製薬会社と共同研究を開始。
- 2019年9月、米ベンチャーInsilico Medicine社は、分子化合物の特定から前臨床試験まで2～3年かかっていたところを、AIを用いて、わずか21日で完了、という論文を発表。
- **2020年1月**、大日本住友製薬と英ベンチャーExscientiaはAIを活用して創製した新薬候補化合物のフェーズ1試験を開始と発表。業界平均で探索研究に4年半かかるところ、12カ月未満で完了したとしている。同時期にTransgene社とNEC、AIによって開発されたがんワクチンの治験を開始と発表。

⇒
CRDSのHPから「報告書」をDL可能



早期疾患リスク予測（医用画像データ）

内視鏡画像から腫瘍の存在診断、治療効果の判定



現在の精度

✓評価用データ*
病変**画像：705枚
非病変画像：4135枚

(*学習用に用いていない画像、答えを教えず解析)
(**病変、前癌病変としてのポリープと早期がん)

	感度 (検出した病変数/全病変数)	特異度(誤検出したbox数/処理画像枚数)	
		病変なし	病変あり
全病変 (751病変)	97.3% (731病変/751病変)	99.0% (43個/4135枚)	90.3% (68個/700枚)
隆起病変のみ (641病変)	98.0% (628病変/641病変)	99.0% (43個/4135枚)	89.0% (65個/592枚)
平坦病変のみ (110病変)	93.6% (103病変/110病変)	99.0% (43個/4135枚)	96.3% (4個/109枚)

隆起型の病変：感度98%、特異度99%、正診率98.8%

(検出だけでなく、病変の位置まで正しく示した場合に正解とした)

Scientific Reports volume 9, Article number: 14465 (2019)

出典：浜本隆二（国がん）

- 内視鏡に限らず、様々な医用画像に関する**CNN**を用いた成果が産出。
- 2019年、オリンパス社、AIを搭載した内視鏡画像診断支援ソフトウェア「EndoBRAIN」を発売。
- FDAは、2018年以降、AI医療機器を60種類以上承認。
- 2020年3月現在FDAに承認されているAI医療機器については、パラメータとアルゴリズムを固定（追加の学習は認めない）して承認を受けており、AI機器が確定診断をできるわけではなく、あくまで診断補助（医師が判断の責任ももつ）として用いられている。

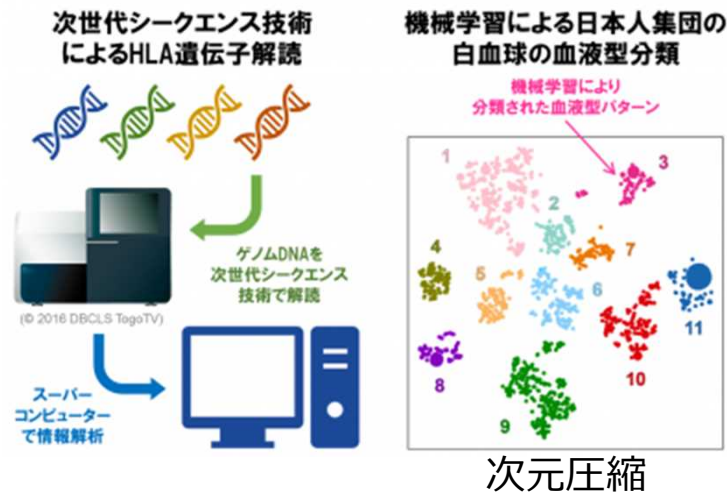
今後は、1) 質的診断、2) 転移予測、3) 予後予測等への発展が見込まれる。

- 課題
- ①画像のアノテーション（ラベリング）のコストが高い
 - ②付与した情報の範囲を超えた分類はできない
 - ③根拠・説明が欲しい

ヒト・患者の層別化と疾患リスク予測（血液・尿等分子データ）

バイオマーカーは、従来の一遺伝子（分子）・一表現型から、生体分子群の特徴を機械学習を用いてパターンで識別する方向に

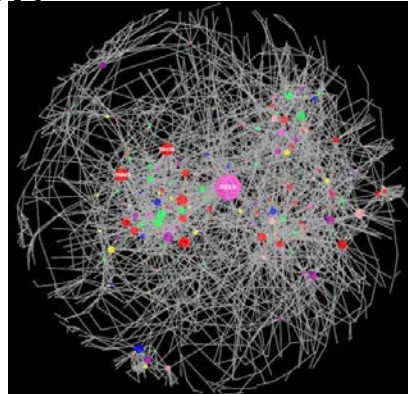
日本人集団のHLA（白血球型）を解明



Nature Genetics volume 51, pages470–480(2019)

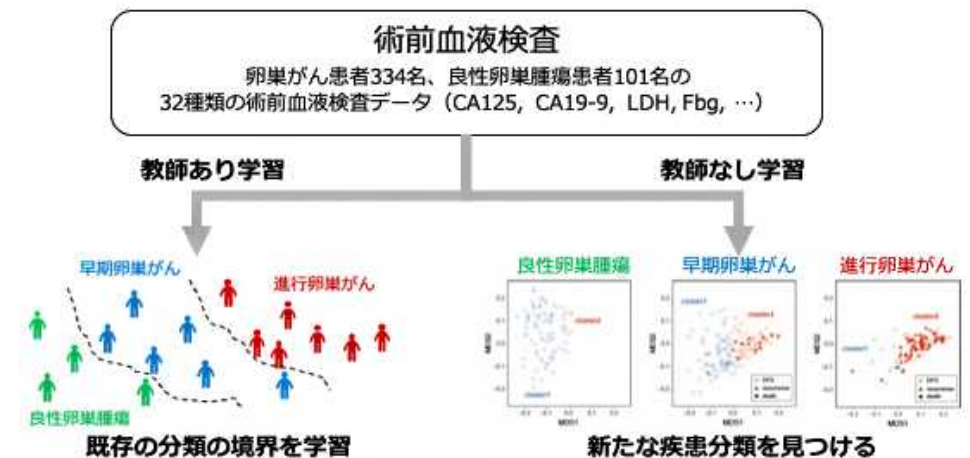
血中マイクロRNAを用いた認知症発症リスク予測モデルの構築 東レ

予測モデルで用いられたmiRNAのターゲット遺伝子の関係ハブ遺伝子として EXOC5, DDX3X, YTHDF3を抽出
(EXOC5はADと、DDX3XとYTHDF3は脳腫瘍との関連が知られている)



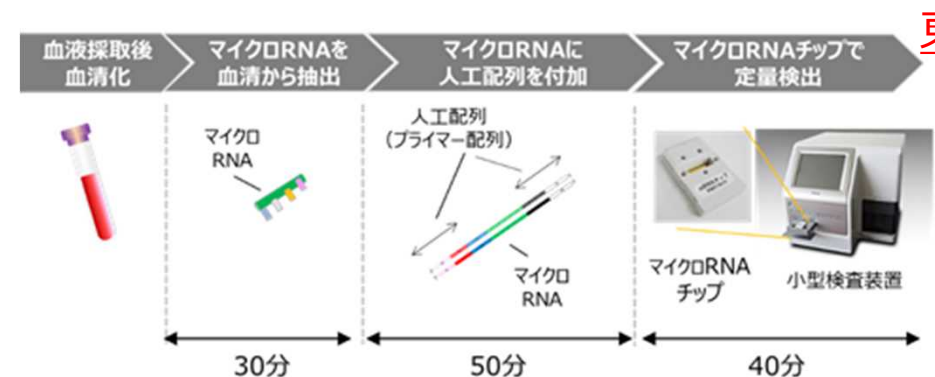
Communications Biology volume 2, Article number: 77 (2019)

卵巣腫瘍の特性を術前予測



Clin Cancer Res April 11 2019

マイクロRNA検出技術を用いて、13種類のがんを高精度で2時間以内に網羅的に検出 東芝



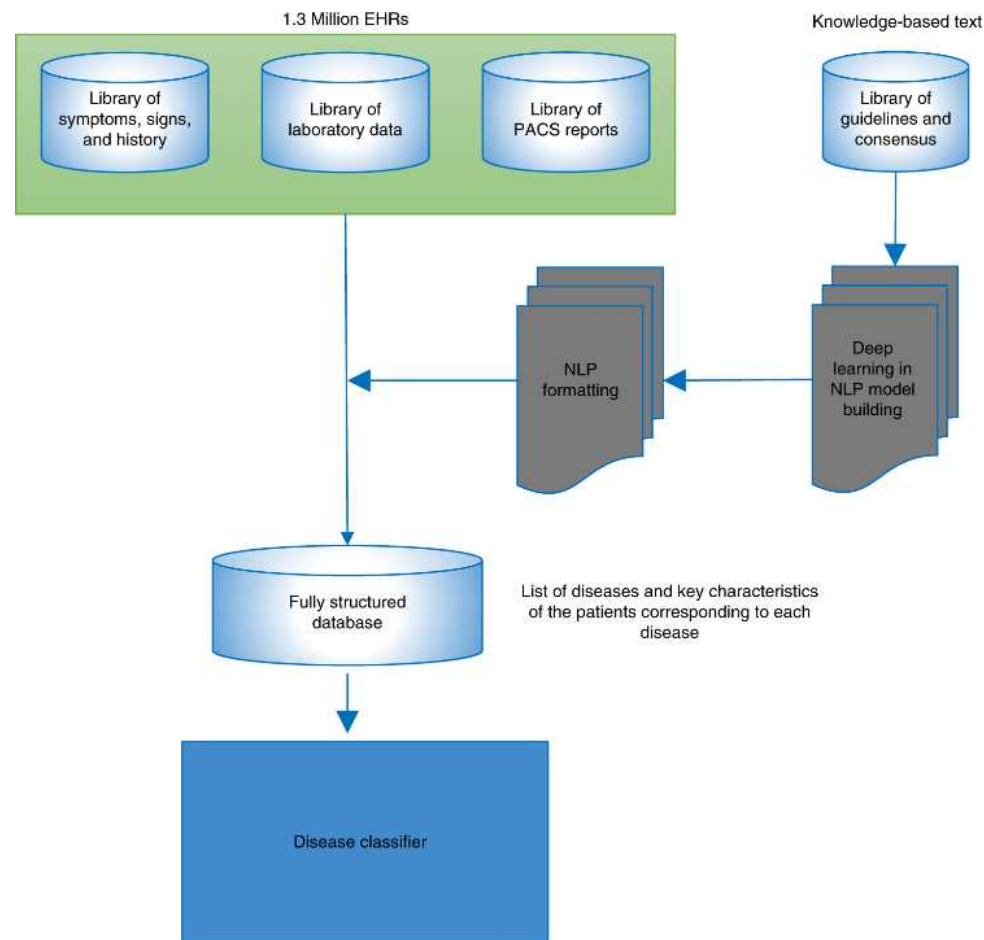
出典：AMEDプレス2019年12月

診断支援（AI医師in中国）（電子カルテ等データ）

- 自動化された**自然言語処理**システムを適用して、電子健康記録（EHR）から臨床的に関連する情報を抽出
- 1,362,559人の小児患者の訪問からの計1,016万のデータポイントを分析して、トレーニングおよび検証
- 一般的疾患（インフルエンザ、手足口病など）と、危険もしくは生命を脅かすような疾患（急性の喘息発作、髄膜炎など）の診断など、複数の臓器システムにわたって、55種の小児疾患の診断において経験豊富な小児科医に匹敵する高い診断精度を実証
- このシステムで3万276回の診断を行い、診断と臨床の一致度は87.4%にのぼる。

広州医科大学広州女性小児医療センター
カリフォルニア大学サンディエゴ校

Nature Medicine volume 25, pages433–438(2019)



- 上海交通大学医学院附属上海兒童医学センターの外来で、小児科の「AI医師」が勤務を開始。スマホアプリを起動し、音声か文字で症状を入力すると、AI医師が診療案内や受診手続きをしてくれ、患者に最もふさわしい医師を推薦してくれる。さらに「診療前AI検査サービスを受ける」を選ぶと、AI医師が症状を詳しく聞き、必要な検査を推薦してくれる。
- 患者は受診時間を1時間以上短縮できる。

診断支援（AI医師in中国）（電子カルテ等データ）

世界最大の保険会社「平安保険グループ」

- コアとなる医療サービス。オンライン問診や遠隔診療により診察の効率化。
- 平安医好生（Ping An Good Doctor）では、2018年半ばまでに、登録ユーザーは2億2,800万人、1日平均531,000人の相談がある。
- 診察データは累計で3億件以上が蓄積されており、これをもとにAI診断を行い、人間の医師をサポート。これにより、従来の5倍近い効率で診察が行われている。
- 外部のドクターは4万に達する。さらに、3,100の病院、1,100の診療所、500のデンタルクリニック、1万の薬局と提携をしており、プラットフォームの体をなしている。

**O2O（Online To Offline）から
OMO（Online Merges with Offline）へ**



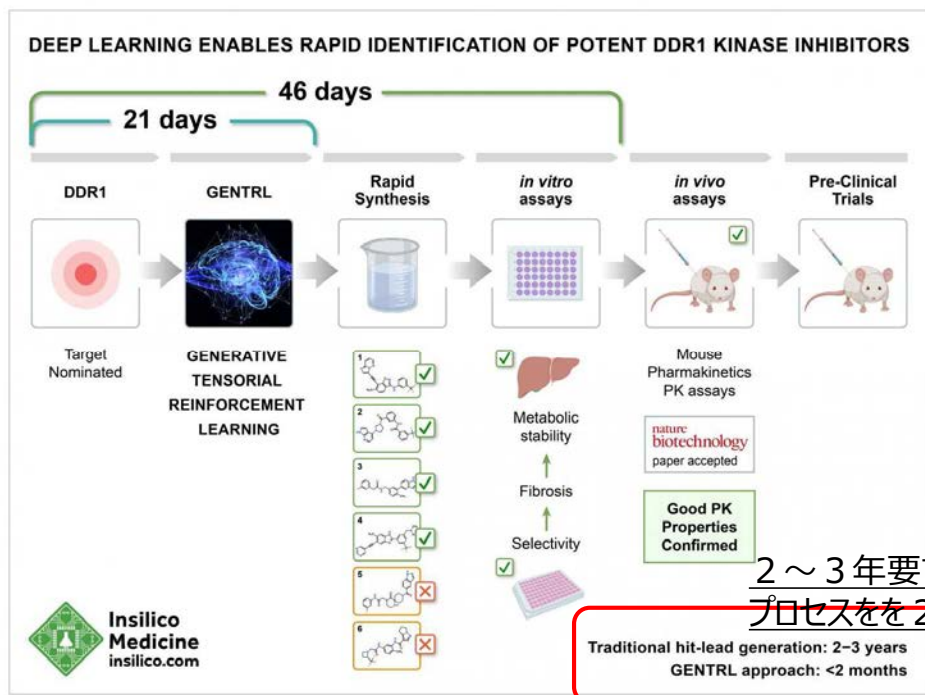
- **テンセント（騰訊控股）**傘下の医療企業WeDoctor（微医）では、自治体との契約によって、農村部に定期的に移動診療を実施し、健康診断を行い、ここで集められたデータを、AI訓練用として同社のデータベースに蓄積
- データベースには2,000種の疾患と5,000以上の症状が記録されており、AIによる診断の精度は90パーセントを超える
- 自治体の医療従事者はこのシステムを利用することができる

人口が多く、医師不足であり、かつ個人情報の活用の障壁が高くない中国だからこそいち早く実装

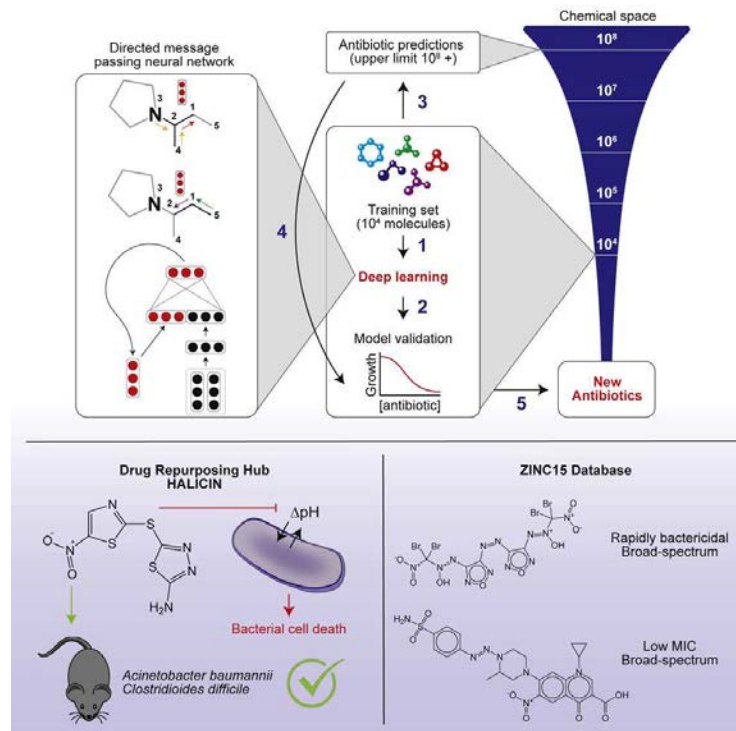
創薬への応用（化合物・分子データ）

- AI創薬ベンチャーのInsilico Medicine社（米国）では、新規の低分子設計のために、生成テンソル強化学習（GENTRL）を開発。合成の実現可能性、新規性、および生物活性を最適化することができる。
- これを使用して、線維化やその他の疾患に関するキナーゼターゲットであるディスコイジンドメイン受容体1（DDR1）の強力な阻害剤を20日間程で発見（2019年9月）。

- MITでは、化合物のグラフ表現と抗菌活性を紐付けるニューラルネットワーク（Message Passing Neural Network）を用いて、大腸菌を殺すのに効果的であることが知られている約2500の化学構造で最初に訓練され、約6000の化合物ライブラリを用いてスクリーニングし、他の既知の抗生物質とは異なる構造を持つ潜在的なリード化合物を特定。
- 「Halicin」と名付けられた抗生物質はこれまでのものとは異なる作用メカニズムを持っており、臨床試験では数々の薬剤耐性菌を殺すことができた。



Nature Biotechnology volume 37, pages1038-1040(2019)



Cell Volume 180, Issue 4, 20 February 2020, Pages 688-702.e13

製薬業界でも 成果が見える形に

2020年1月30日

各位

大日本住友製薬株式会社
Exscientia Ltd.

大日本住友製薬と Exscientia Ltd.の共同研究 人工知能(AI)を活用して創製された新薬候補化合物のフェーズ1試験を開始

大日本住友製薬株式会社(本社:大阪市、代表取締役社長:野村 博、以下「大日本住友製薬」と Exscientia Ltd.(本社:英国 オックスフォード、CEO: Andrew Hopkins、以下「Exscientia」)は、このほど、両社の共同研究(以下、「本研究」)を通じて人工知能(AI)を活用し創製された DSP-1181(以下、「本剤」)のフェーズ1試験を日本で開始しましたので、お知らせします。今後、大日本住友製薬は強迫性障害を予定適応症として本剤の開発を進める予定です。

本剤は、大日本住友製薬のモノアミン GPCR 創薬における創薬経験・知識と、Exscientia 独自の AI 創薬プラットフォームである Centaur Chemist™ との融合による共同研究を通じて見い出されました。その高い相乗効果により、業界平均※で4年半を要するとされる探索研究を12ヶ月未満で完了しました。

大日本住友製薬HP : https://www.ds-pharma.co.jp/ir/news/pdf/ne20200130_2.pdf

Transgene社とNEC、AIによって開発されたがんワクチンTG4050の2つの治験を卵巣がんと頭頸部がんで開始

News Room



経営戦略／業績／人事



サステナビリティ



研究開発／新技術



2020年1月7日

Transgene SA

日本電気株式会社

- TG4050は、NECの最先端AI技術とTransgene社のウイルスベクター技術を組み合わせることにより、治療用ワクチンをデジタル世代のものにします
- 新しい免疫治療は、患者さん毎の腫瘍の変異を標的にする免疫反応を引き出すように患者さん一人一人に対して十分に調整されます
- メイヨークリニックとツールーズオンコポールにて2つの第Ⅰ相臨床試験への患者さんの登録が行われました。

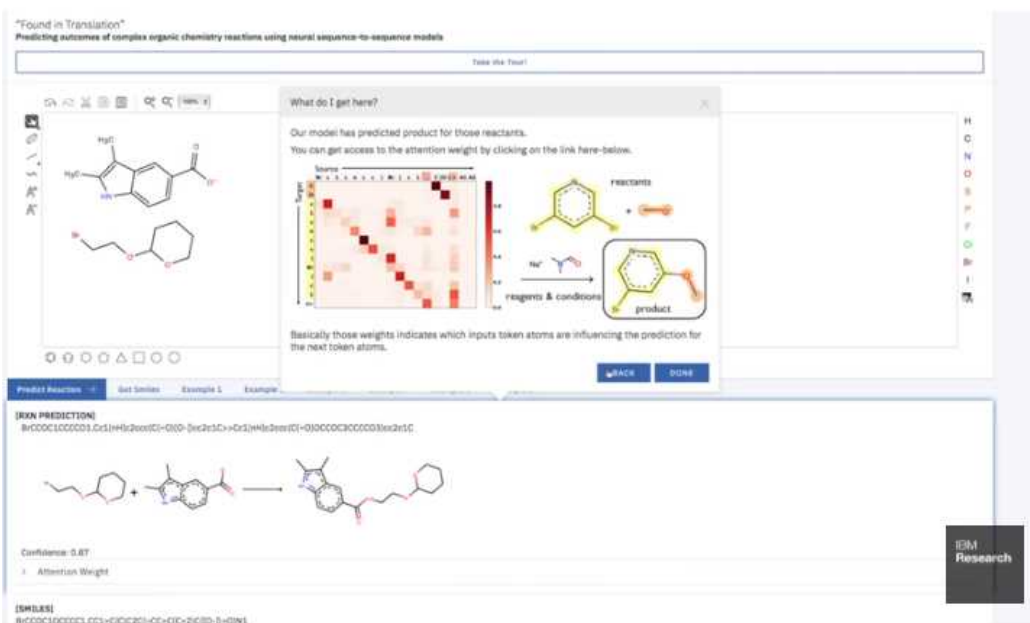
化学反応・創薬等への応用

反応物から生成物を予測 (化学反応予測)

Chem. Sci., 2018,9, 6091-6098

- **RNN**を用いて、39万5496件の化学反応を含んだデータセットから未知の化学反応を予測
- 予測の正解率は、80%。

反応予測アプリケーション IBM RXN for chemistry



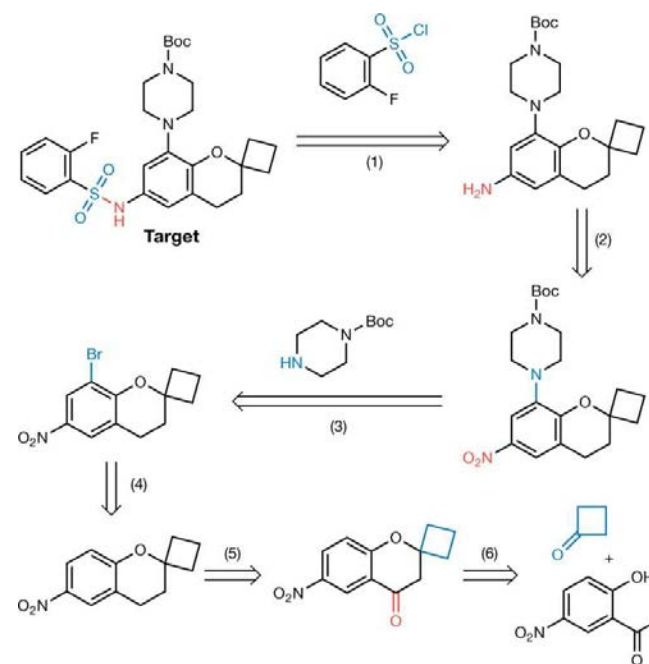
特許と書籍から集めた200万の反応を利用

生成物から反応物を予測する試み

Nature 555, 604–610 (2018)

- AIによって提示された候補分子群をどうやって作るか研究者が考えるプロセスが律速に。
- これまでに知られている一段階有機化学反応のほとんど全て約1240万をAIに学習させた。
- アルファ碁と同じアルゴリズムを利用。

2015年に報告された薬物候補合成の中間体の6段階のルートの例は、アルゴリズムによって5.4秒で見出された。

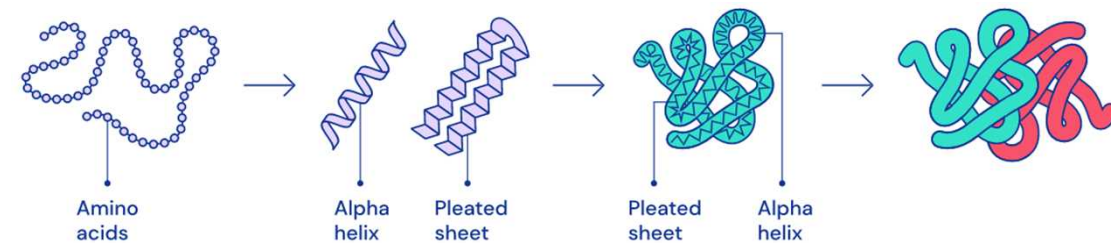


その他・創薬等への応用

タンパク質の立体構造予測

- アミノ酸配列からタンパク質の3次元構造（フォールディング）を解析
- アミノ酸同士の距離、化学結合の間の角度、エネルギー効率の良い配置

AlphaFold By DeepMind(Google)



この時点では、6 オングストロームの誤差がある
(クライオ電顕の解像度は 1 ~ 2 オングストローム)。

クライオ電顕等で結晶構造解析をしなくともタンパク質の構造をある程度正確に予測できる時代が来るかも？

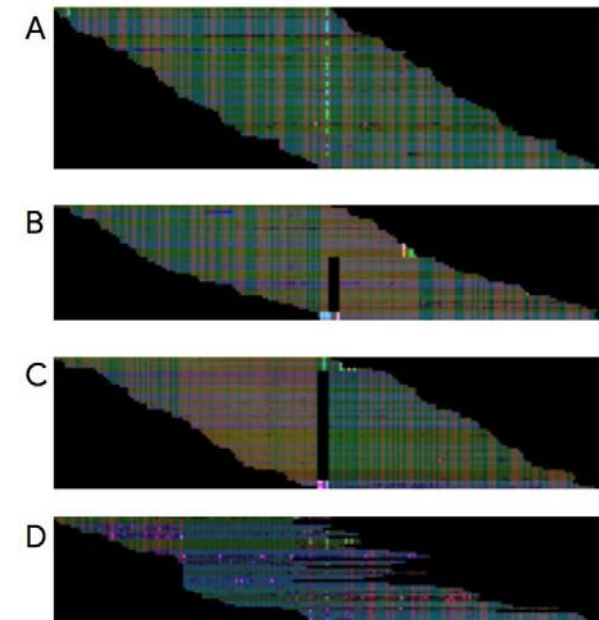
Nature volume 577, pages706–710(2020)

ゲノム変異（遺伝的バリエーション：一塩基SNPと挿入/削除）解析

- 配列決定技術の急速な進歩にもかかわらず、個々のゲノムに存在する遺伝的変異体を数十億の短いエラーのある配列読み取りから正確に読むことは依然として困難
- DNA鎖をコンピューターが認識できる画像に変換

DeepVariant By Verily LifeSciences(Google)

image classification

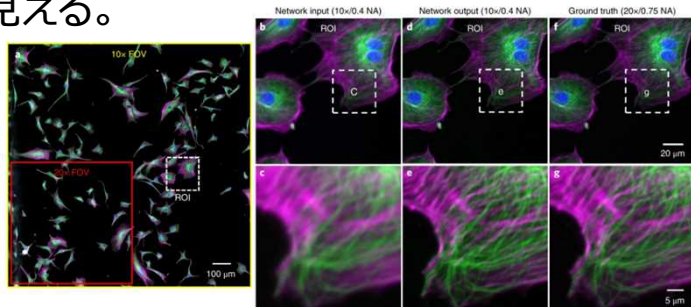


Nature Biotechnology volume 36, pages983 – 987 (2018)

画像の超解像化、3次元化等

GANによる光学顕微鏡画像の超解像化

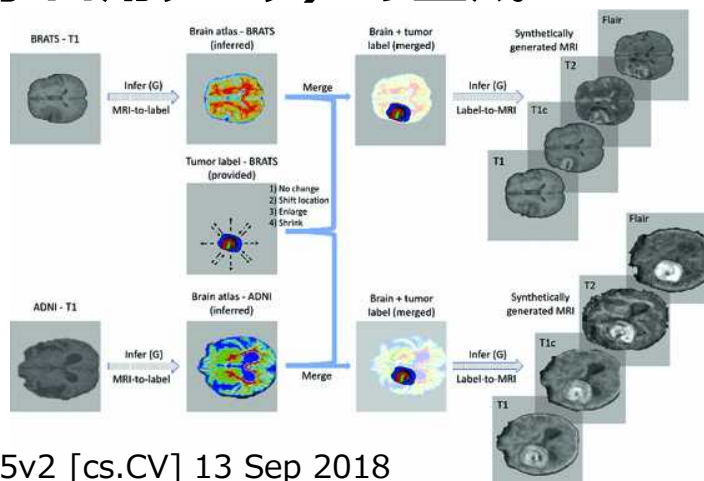
- 光学顕微鏡で高い分解能でサイズの大きいものを見ようとすると分解能と撮像に要する時間のトレードオフ関係がある。
- これまで1枚の撮像に分単位を要したが、秒単位を可能にし、ダイナミクスが30ナノメートル以下の分解能で見える。



Nature Methods volume 16, pages103–110(2019)

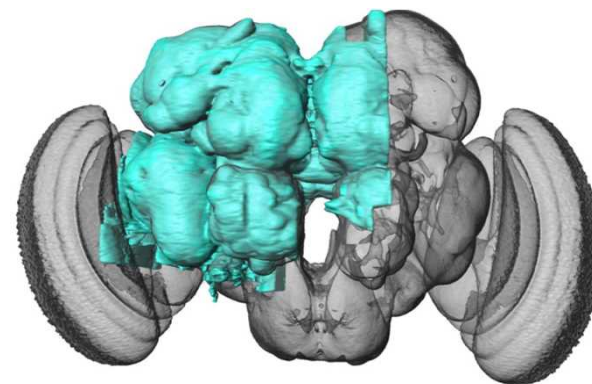
GANによる医用画像（学習用データ）の生成

- GANで生成された合成画像が学習用データに使われ、機械学習のアウトプット（識別器としての効率）が挙がる事例が現実に出始めている。



電子顕微鏡（SEM連続断面観察法）を用いた細胞や臓器の3次元再構成

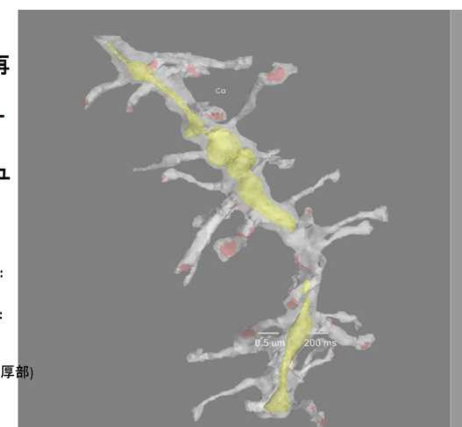
ショウジョウバエの脳にある2万5千個のニューロンとその間の2,000万のつながりを示す配線図が示された。



A Connectome of the Adult Drosophila Central Brain
bioRxiv January 21, 2020.

微小領域の再構成により、細胞内シグナル伝達分子(Ca²⁺)のシミュレーションが可能

White colored area : (細胞膜)
Yellow colored area : (ミトコンドリア)
Red colored patch : (PSD, 後シナプス肥厚部)
Blue points : Ca²⁺ ions



出典：石井信（京都大学）
Scientific Reports volume 9,
Article number: 19413 (2019)

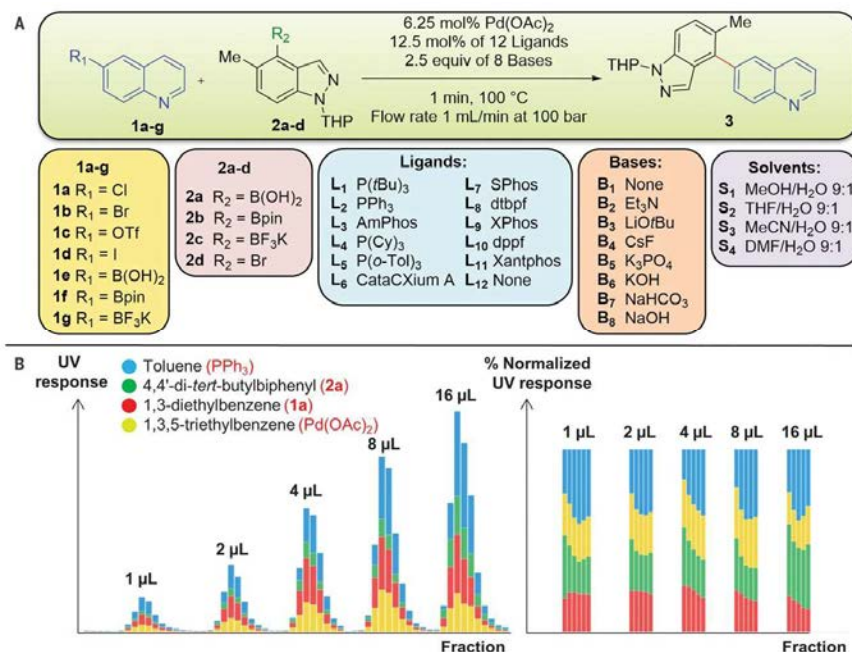
データとアルゴリズムの課題の克服に向けて

課題	対応策
データ品質と数量 ●新NP問題（サンプル数 $\ll$ パラメータ数） ●大量のラベルなしデータと少量のラベルありデータ ●事後アノテーションのコスト高	■ データプラットフォーム＋分散統合学習 ■ 研究環境の自動化（ロボットなどの活用） ■ 自然言語処理（NLP） ■ 生成モデル（GAN、VAE）
ブラックボックス	■ 次元削減技術や重み付け可視化技術
学習範囲（内挿外挿）	■ 次世代AI ■ 生成モデル（GAN、VAE） ■ データ同化（実験とのサイクル）／研究環境の自動化（ロボットなどの活用）
複雑系生命理解（生命システムのモデルや時空間問題）	■ 計測技術＋AIアルゴリズム
個人情報保護	■ フェデレーションラーニング

化学研究の自動化

有機合成のデジタル化

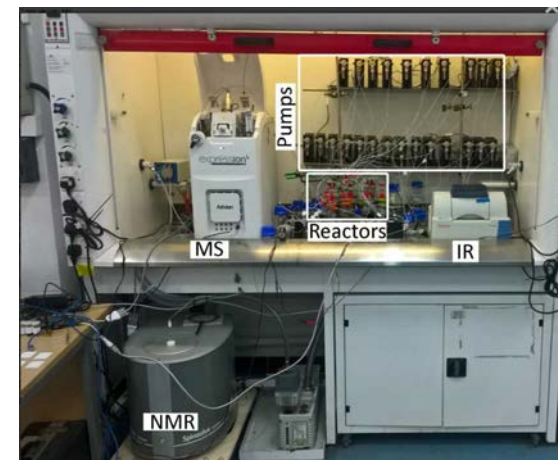
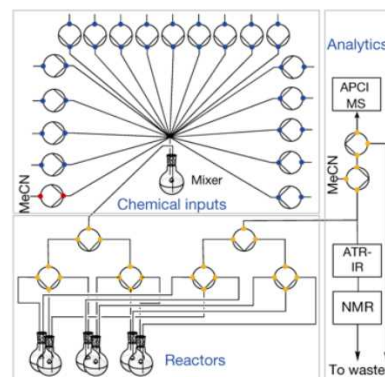
- 複数の離散変数（触媒、リガンド、塩基）および連続変数（温度、滞留時間、圧力）（24時間あたり1500反応を超える速度で合計5,760反応ポイントでのLC-MSデータを生成）を使用した鈴木-宮浦カップリングのハイスループット反応スクリーニングを実証。
- フロー装置が薬物発見プロセスの早い段階で反応の最適化を加速できること、および他の研究者が機械学習アルゴリズムを構築できるようにする信頼できるデータを提供できることを示した。



Science 26 Jan 2018:Vol. 359, Issue 6374, pp. 429-434

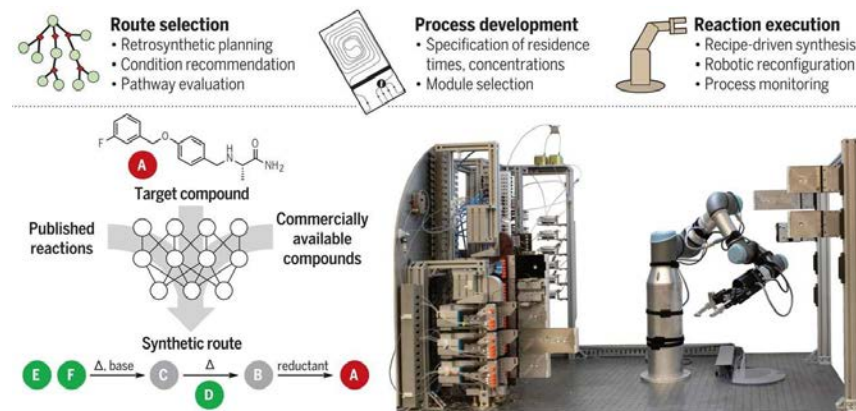
ロボットケミスト

約1000個の反応組合せのわずか10%を実現することで、実験を実施することなく、残りの90%の結果を80%以上の精度で予測



Nature volume 559, pages377–381(2018)

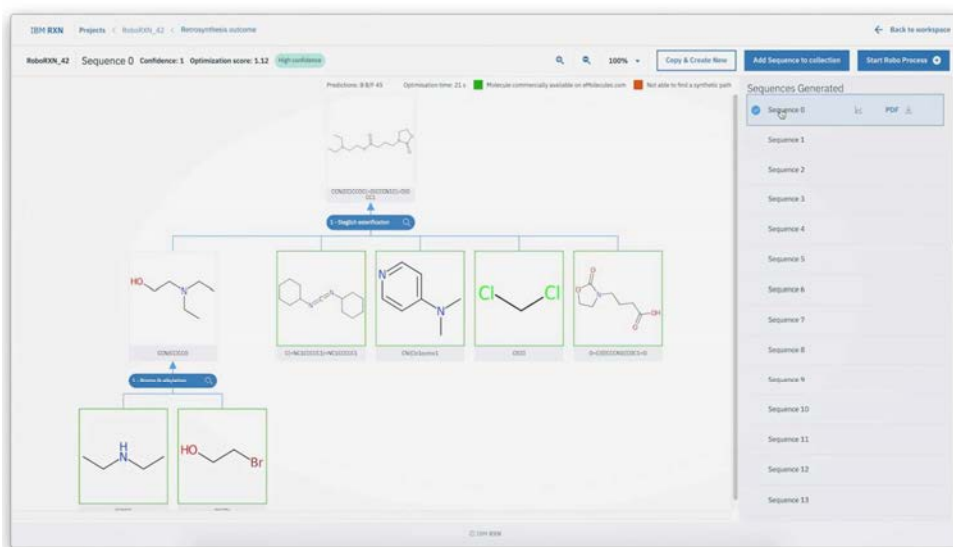
最初に人工知能に導かれたソフトウェアが分子を合成するためのルートを提案。次に専門の化学者がこのルートを確認して化学的な「レシピ」に仕上げ。最後にハードウェアを自動的に組み立てて反応を実行するロボットプラットフォームにレシピを送信



Science 09 Aug 2019:Vol. 365, Issue 6453, eaax1566

AIが有機合成を立案し、ロボットが合成

- IBMは、人工知能モデルとクラウド・コンピューティングのプラットフォーム、ロボットを組み合わせ、
「ロボRXN (RoboRXN)」と呼ぶ化学研究所をクラウド上に構築した。研究者が在宅で新たな分子を設計し、合成することを支援する。
- 研究者はWebブラウザ経由で、作成したい有機化合物の骨格構造をロボRXN上の白紙のキャンバスに描くと、ロボRXNが機械学習を用いて必要な材料と調合の手順を予測、遠隔地の研究所のロボットに指示を送ってその合成を自動実行させる。実験が完了したら、ロボRXNが結果と共に報告書を利用研究者に送信する仕組み。
- 現時点では医薬品関連化合物が主に想定されている。



利用者は目的の有機化合物の骨格構造を、ロボRXNのキャンバスに描く

<https://rxn.res.ibm.com/>

バイオ研究の自動化

Robot scientist Eve

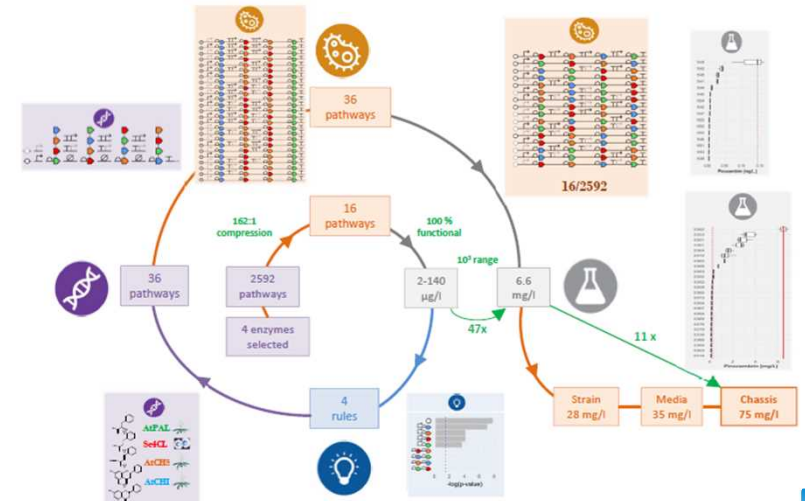
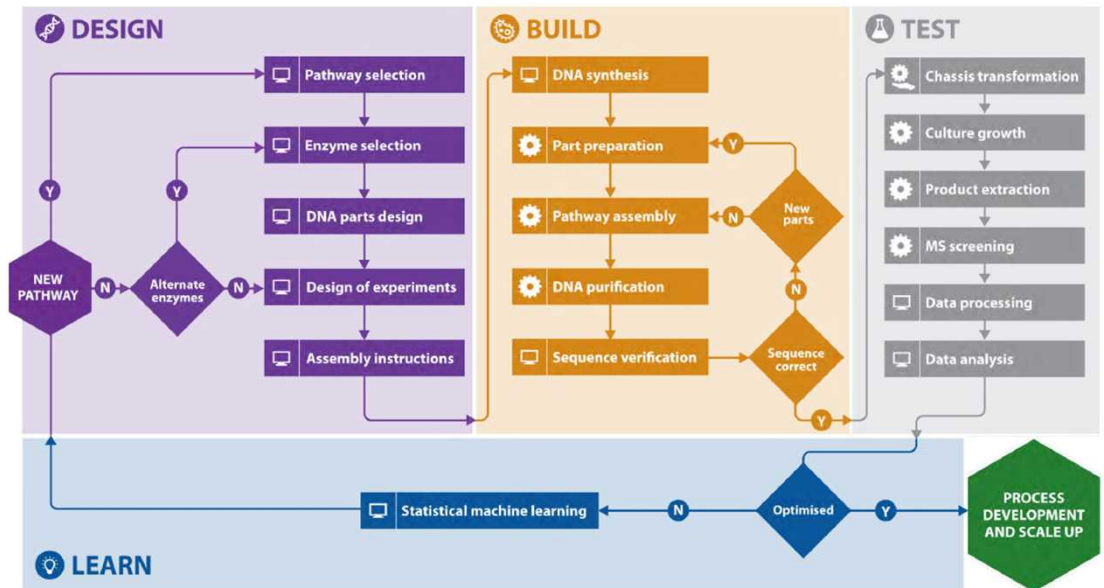
- ロボットサイエンティスト「イブ」をハイスループットスクリーニングで採用し、多くの練り歯磨きに含まれる成分であるトリクロサンがマラリア原虫の、従来知られていたENRだけでなく、DHFRと呼ばれる酵素を特異的に阻害することにより、寄生虫の成長に影響することを発見。
- マンチェスター大学、ケンブリッジ大学等のチームによって開発された。
- 自動的に仮説をたて、テストして観察結果を解釈し、実験を実行し、結果、仮説を修正するサイクルを繰り返して、ハイスループット仮説主導の研究を自動化。



ケンブリッジ大学HP : <https://www.cam.ac.uk/research/news/artificially-intelligent-robot-scientist-eve-could-boost-search-for-new-drugs>

DBTLサイクルと実現パイプライン

- 大腸菌によるフラボノイド (2 S) -ピノセンブリンの生産に適用。
- 2回のサイクルの適用により、最大88 mg/L -1に到達。収量が500倍改善された生産経路が確立。



サマリー

2016年以降、計測（画像）×識別モデル、スクリーニング×生成モデルが躍進。
今後、自然言語処理、AI×ロボティクスの活用が進むことが見込まれる。

現在

- 医用画像による疾患の発見や予測 by CNN（識別モデル）
- 診断（意思決定）支援 by RNN（自然言語処理）
- 創薬プロセスにおける活用（時間とコストの大幅削減） by GAN/VAE等（生成モデル）
- イメージングの超解像化、ハイスループット化、3次元化 by CNN/GAN（識別モデル／生成モデル）
- 細胞分取、細胞培養 by CNN/スパースモデリング等

5年後

- 微量血液等からほとんどの疾患のリスク予測
- 生命システム原理（数理物理モデル）の部分的解明

10年後

- 薬の臨床試験においてデジタルモデルが人の代替（デジタルツイン）
- 脳波測定ウェアラブルデバイスを用いた応用
- AI×ロボティクスにより自動化から自律化（仮説生成）へ

21世紀中 バイオ革命

AIによる超複雑系の解明①

■ ゲノム、エピゲノムとあらゆる表現型との因果解明→ゲノム・分子治療→生命の設計、人間拡張

AIによる超複雑系の解明②（AIとバイオと量子の接点）

■ 精神・神経疾患の治療→意識・思考・意志のメカニズム解明→人工意識、人間拡張

今後の方向性

1. 戦略的なデータプラットフォーム～オープンサイエンス～の構築

- ゲノム、がん、脳神経、免疫、マイクロバイオーーム等で先行

2. 研究の自動化・自律化 ～オンデマンド化～

- AIとロボティクスの相性の良さから化学、バイオ、材料でホットトピックス。
2018年からScience、Nature本誌を賑わす。

3. 電子カルテ等（EHR）等と自然言語処理（NLP）

- 中国でAI医師が通常になりつつある。AmazonやApple等が追いかける展開。

4. 生命システム（複雑系・ダイナミクス）原理の理解

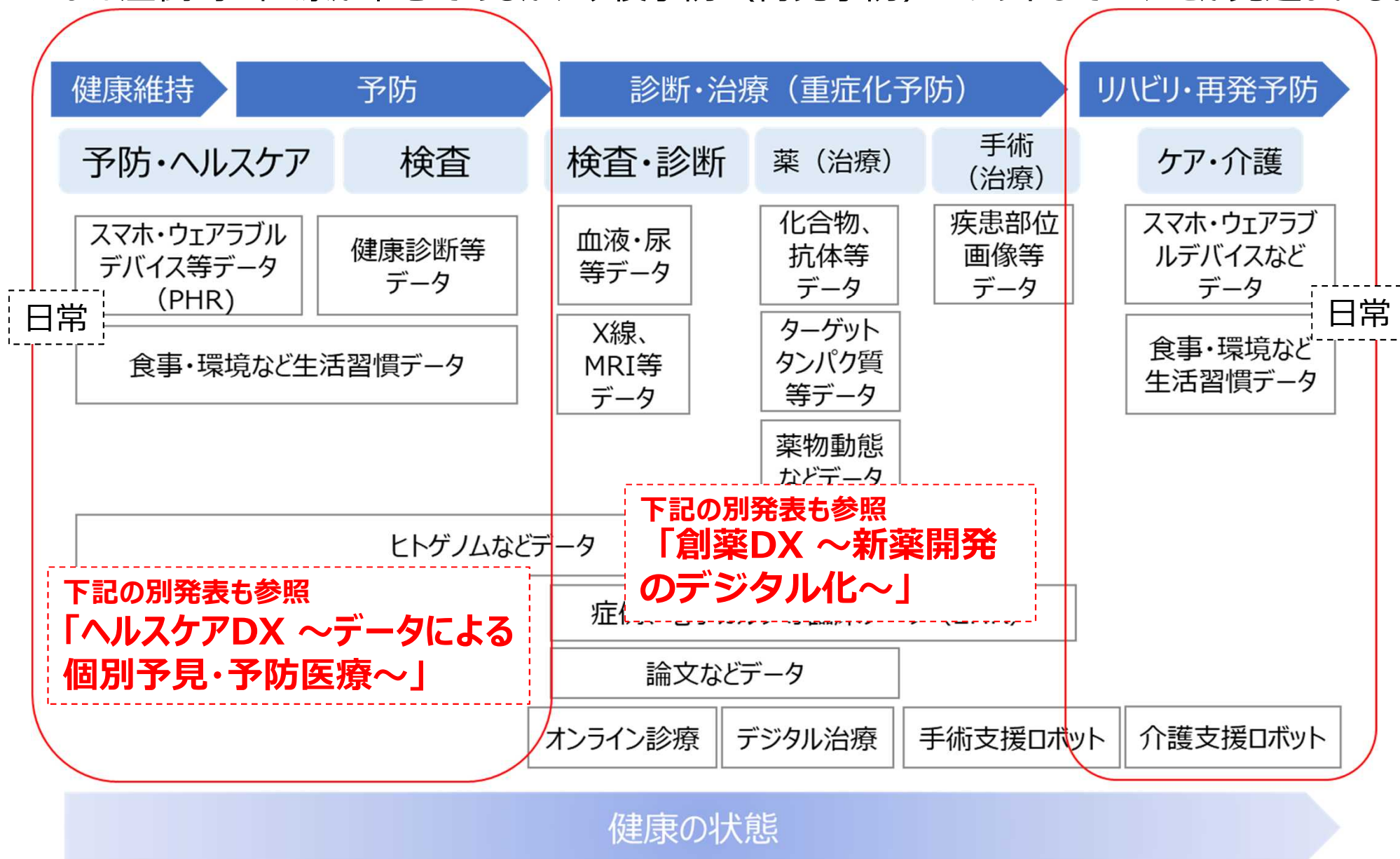
- 今後システム理解（脱要素還元、複雑なものを複雑なまま理解）にAIの活用が期待。

5. 脳神経科学と将来AIアルゴリズム

- 次の脳神経系インスパイアードアルゴリズムはまだまだこれから

今後の方向性

いまは圧倒的に医療が中心であるが、今後予防（再発予防）へシフトしていくことが見込まれる。



最後までご覧頂きありがとうございました。

■ CRDSの活動・成果・コラムはウェブサイトでご覧いただけます



<https://www.jst.go.jp/crds/>



各種報告書・コラム無料公開中

■ 最新情報はメールマガジン、Facebookでも配信中



<https://www.jst.go.jp/melmaga.html>



<https://fb.com/170314426446196/>

■ その他 お問い合わせはこちらまで



crds@jst.go.jp