

平成29年度厚生労働科学研究費補助金（化学物質リスク研究事業）
分担研究報告書

化学物質の有害性評価手法の迅速化、高度化に関する研究
— 新型反復暴露実験と単回暴露実験の網羅的定量的遺伝子発現情報の対比による毒性予測
の精緻化と実用版毒性予測評価システムの構築 —

分担研究課題：「システムトキシコロジー解析基盤の研究開発」

研究分担者： 北野 宏明
特定非営利活動法人システム・バイオロジー研究機構 会長

研究要旨：

毒性機序の複雑性に対応すべく、大規模データ解析技術の開発として ensemble learning system に加え、deep learning system の開発を進め、実際のゲノミクスデータを用いた検証を進めた。ゲノム解析とその関連データベースの整備としては先行研究で開発したソフトウェアの一層の強化(軽量化、高速化)と Garuda プラットフォーム上への実装、一般公開を進めた。

A. 研究目的

本研究は、これまでの蓄積を基盤に、システム毒性学のさらなる発展と実用レベルに達する成熟化を目指し、その中で本プロジェクトの成果を国際的に大きく普及させることを目指していく。特に、本研究フェーズでは、技術実証のフェーズから、より現実的な応用を意識し、広範な化学物質の低濃度長期反復曝露時の生体毒性の計算機予測を可能とし、OECD AOP (Adverse Outcome Pathway) との連動性を高めた解析技術を開発する。

ここまでの研究では、解析の基本技術の開発を中心に行ってきたが、生体内分子相互作用ネットワークは、大規模データからの推定を行う方式であった。しかし、広範な毒性予測を行うには、事前に高精度な計算モデルを構築し、そのモデルと大規模データから推定された分子間相互作用の変化を比較参照しながらより精度の高い毒性予測を行う必要がある。さらに、マウスでの実験結果と人間での違いや、マウスを利用した毒性試験ができない場合に、どのように人間での毒性リスクを予測するかの技術を開発することがより重要性を増している。

そこで本研究計画では、主要な毒性に関係するアウトカムに関連するモデルを構築し、その挙動を予測するための技術と情報基盤群の開発を行うこととする。これは、

これまでのプロジェクトが、最初に特定の化学物質を取り上げ、その影響を順次解析するという流れで研究を進めたアプローチとは、逆のアプローチであり、毒性アウトカム(Adverse Outcome)という生理学的な結果から、それを引き起こす原因にさかのぼるという方法論である(図1)。すでに、今までの研究で、化学物質からその影響を予測する方法は、一定の開発が進められており、その方法のさらなる洗練化は継続する必要があるものの、この段階で、逆側からのアプローチを新たに導入することで、全体の解析系が確実に連動し、実用的な結果をもたらすことをより確実にすると考える。

また、複数の化学物質の複合的效果を考える際に、Forward Effect Analysis のみでは、検討すべきターゲットの範囲が広範になると同時に、実際に Adverse Outcome として関連するターゲットはそのごく一部であるという極めて非効率的な解析を行わざるを得ない。これが、今まで複合的な曝露に対して計算論的アプローチが十分な有効性を持ちにくかった理由の一つである。しかし、実際には知りたい Adverse Outcome は明確に定義することができ、そのような結果をもたらすであろう遺伝子や分子の変動は、一定の組み合わせに限定することが可能である。その組み合わせが、ある程度広範囲にわたる場合においても、その範囲が定義されることは非常に重

要である。これは、例えば、代謝工学で使われる Metabolic Flux Analysis(FBA)などの制約駆動型システム状態分析手法を、逆方向に利用するというイメージである。この考えをもとに、AOP の中間段階の分子の変動や遺伝子変動の注意すべきパターンを類型化し、そこから、複数の候補化合物の組み合わせを使った Forward Effect Analysis を実行することで、より効率的かつ精度の高い複合曝露の計算予測ができること期待される。

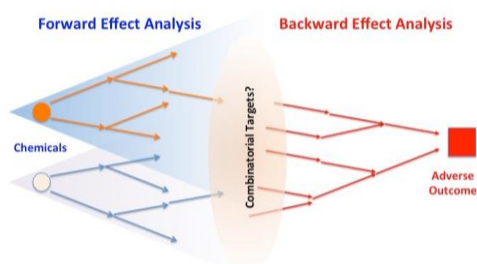


図 1 : Forward Effect Analysis と Backward Effect Analysis の融合

この一連の処理を行うには、複数のソフトウェアの連動が必須であり、本研究では、これに関わる一連のソフトウェアの開発を行う。

(A) AOP 連携

Adverse Outcome Pathway(AOP) は、Pathway の名称を使っているものの、分子間相互作用ネットワークのレベルでの Pathway ではなく、より抽象的な、生理学的変化がどのように引き起こされるかを概念化したネットワーク化知識体系である。そこで、一連の毒性に関連する AOP を Garuda Platform 上におけるツールである PhysioDesigner で定義し、分子間相互作用ネットワークモデル構築ツールである CellDesigner によるモデルと連動させる。ここで重要なのが、分子レベルや遺伝子レベルでのどのような変動が、AOP レベルでの変動に関連するののかの同定と検証である。AOP レベルの記述が、生理学的なレベルであるので、最終的に分子レベルや遺伝子レベルのモデルでの予測を AOP レベルに連動させる必要がある。逆に、これが実現できれば、マウス実験からのヒトでの影響推定やマウスを使わない場合でのヒトでの影響推定を可能とする技術に一步踏み出すことができる。

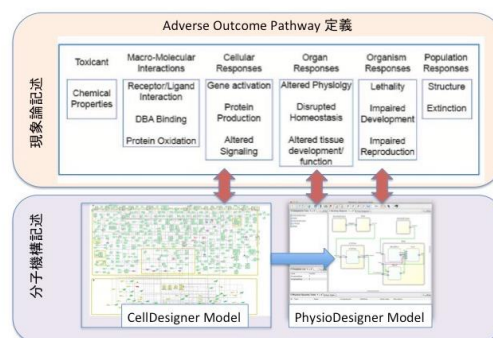


図 2 : 現象論的記述が主体の AOP 定義と分子機構記述モデルとの連携

(B) Garuda Systems Toxicology Platform の開発

これらの研究と今までの研究成果を統合的に運用し、さらに実用的な毒性予測研究に資することができるプラットフォームを開発する。基盤は、これまでの研究で開発をしている Garuda Platform として、そこに毒性研究に特化したツール群、データベースなどを装備し、標準運用手順(Standard Operating Procedure)を定義する。これによって、標準的な毒性の計算機予測の手順が定義される。これらの機能を実現したシステムは、Garuda Systems Toxicology Terminal として実用化される。(図 3)



図 3 : Garuda Systems Toxicology Terminal

B. 研究方法

本課題で成し遂げようとしている目標は広範であり、多元的アプローチとそれらを統合することが必要となる。そのため、
 (1) 大規模データ解析技術の開発と
 (2) ゲノム解析とその関連データベースの整備を行った。

(1) 大規模データ解析技術の開発
 近年、大規模かつ多次元的な生物学データ

が蓄積されつつある。特に、化合物が毒性を引き起こすメカニズムは非常に複雑であり、Percellome等の化合物の毒性に関する発現データベースは、この複雑性を内包している。したがって、Percellomeデータベースから、有用な情報を抽出するためには、この様なデータの複雑性に対処できる解析法が必要不可欠である。しかしながら、この様な複雑性の高いデータから有用な情報を引き出すことは、従来の研究で用いられてきた統計解析法などでは不可能である。

前年度の Ensemble Learning 手法に引き続き、今年度は、Deep Learning を利用して、Percellomeデータベースから、一定の特徴量を認識して分類する問題を設置して、その実現可能性を検討した。

Percellomeデータベースでは、化学物質投与下の遺伝子の発現情報を画像で表す。これまで、Percellomeデータベースの解析では、化学物質投与下における有意な遺伝子を同定するために、長年経験とトレーニングを積んできた研究者が、それぞれの遺伝子の発現パターンを画像を目で見て、判別を行ってきた (図4)。

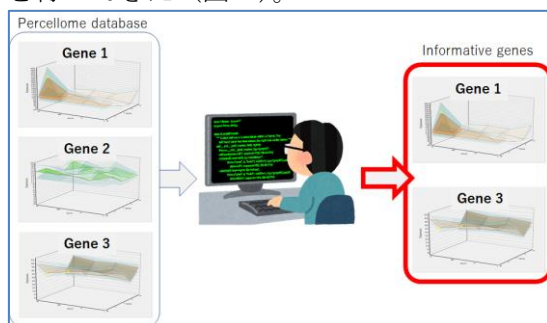


図4：Visual inspectionによる、有意な遺伝子の同定

しかしながら、Percellomeデータベースは、100以上の化学物質それぞれに対し、40,000以上のプローブの発現データが含まれており、400万以上の発現データを記述した画像で構成されている。この様な、大量の画像の全てを目で見て、有意な遺伝子を判別するためには、多大な労力と時間が必要となる。

ところで、近年、機械学習の手法の発展は目覚ましく、機械学習を用いた画像解析の手法も数多く開発されつつある。特に、deep learningの手法の発展はめざましく、分野によっては、人間の能力を超

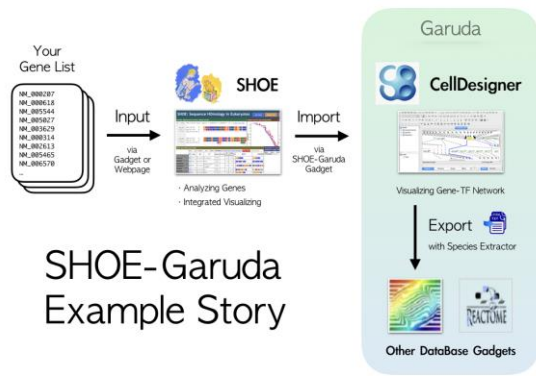
えつつある (David Silver et al (2006) Mastering the game of Go with deep neural networks and tree search. Nature 529:484 – 489.)。例えば、deep learningの技術を使用した画像解析については、非常に高精度に皮膚がんの分類を行うことのできる deep learning モデルが、発表されている (Andre Esteva et al. (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 52:115-118)。

そこで、今年度、我々は、percellomeデータベースの画像データから、効率良く有意な遺伝子を判別する手法の構築を目的として、最新の多数の層で構成される deep learning algorithm を用いた、画像解析システムの実装を行った。実装した deep learning に基づく画像解析システムの予測精度の検証を、熟練した研究者が目で見えて作成した画像データセットを用いて行った。その結果、今回実装した画像解析システムにより、percellomeデータベースの画像データから、有意な遺伝子の判別を、高い精度で出来ることを確認した。

(2) ゲノム解析とその関連データベースの整備

同時に、ゲノム関連のデータベースとの連携をより緊密にし、例えば、上記 Deep Learning で、同定された遺伝子リストなどからネットワークや転写因子群 (TFs) を同定し、そこから毒性の推定を行うなどの一連の処理が必要となる。

この領域では、従来から開発を行っていたSHOEのGaruda Platform上への実装を引き続き行い、さらに機能強化と連携の拡大を行い、この一連の処理に関わる連携を実現した (図5)。



SHOE-Garuda Example Story

図5：SHOE-GARUDA 連携の強化

C. 研究結果

(1) 大規模データ解析技術の開発
今年度、最新の deep learning algorithm (resnet アルゴリズム, Kaiming He (2015) Deep Residual Learning for Image Recognition. arXiv:1512.03385) を用いて、遺伝子の発現パターンを記述した画像データから自動で特徴量を抽出し、それぞれの遺伝子が有意かどうかを判別する、画像解析システムの構築を行った (図6)。このシステムは、画像データ (トレーニングデータとテストデータ) の前処理を行うコンポーネント、deep learning に基づく画像データから有意な遺伝子を予測する分類器を構築するコンポーネント、そして、構築した分類器の予測精度の検証を行うコンポーネント、で構成される。

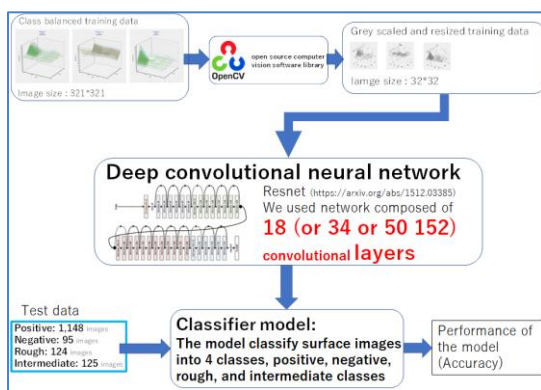


図6：deep learning algorithm を用いた、画像解析システム

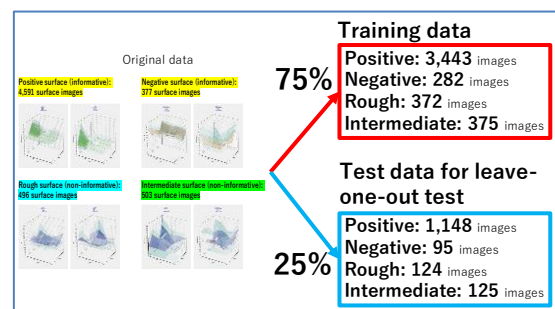
画像データの前処理を行うコンポーネント：このコンポーネントでは、オリジナルの、遺伝子発現を記述した RGB 画像を、グレースケール画像へ変換する。その後、

グレースケール画像のサイズの縮小を行う。

deep learning に基づく画像データから有意な遺伝子を予測する分類器を構築するコンポーネント：このコンポーネントは、多数の層 (畳み込み層やプーリング層等) で構成される。層の数としては、18 層、または、34 層、または、50 層、または、152 層とした。これらの層を用いて、画像データから、特徴量の自動抽出を行う。その後、最後の層において、抽出した特徴量に基づき、それぞれの遺伝子を、4つのクラスに分類する。4つのクラスは、positive class (化学物質により発現が up regulate されるクラス)、negative class (化学物質により発現が抑制されるクラス)、rough class (化学物質により up regulate も抑制もされないクラス)、intermediate class (positive または negative class と rough class の間のクラス)、である (図7)。Positive と negative class に分類された遺伝子が、有意な遺伝子と予測されるものである。

今回、熟練した研究者が、実際に目で見てクラス分けをした画像データを用いた。全データの内、75 パーセントの画像を、分類器を作成するためのトレーニングデータとして、残りの 25 パーセントを、作成した分類器の予測精度を評価するための、テストデータとして用いた (図7)。

図7：テストデータとトレーニングデータ



その結果、18 層で構成される deep learning アルゴリズムを用いた場合、予測の正解率は 83 パーセントと、高い正答率を示した (図8)。この結果は、我々の deep learning に基づく画像解析システム

により、有意な遺伝子を高い精度で推測できることを示唆する。しかしながら、34層、50層、152層の場合の正答率は、0.76、0.7、0.72と、より層の浅い18層の場合よりも低かった。この結果は、層を厚くしすぎると、トレーニングデータに対して過学習を起こしてしまう可能性を示唆している。

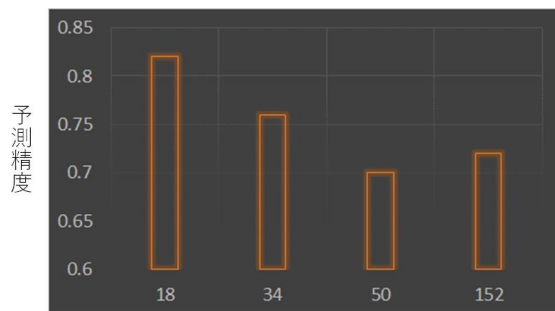


図8：18層、34層、50層、152層の deep learning model を用いた場合の予測精度

また、18層の場合について、4クラスそれぞれに対する、予測精度の検証を行った(図9)。その結果、positive class、negative class、intermediate class のエラーの割合は、12～16パーセントであり、クラス間で大きな差はなかった。しかしながら、rough class のエラーの割合は、28パーセントであり、他のクラスよりも予測精度が低かった。このことから、rough class の判別は、他の class よりも困難であることが示唆される。この原因としては、rough class (有意でない遺伝子) に分類される遺伝子が大部分であり、このクラスには、多様な発現変動パターンを示す多様な遺伝子が含まれるためである、と考えられる。

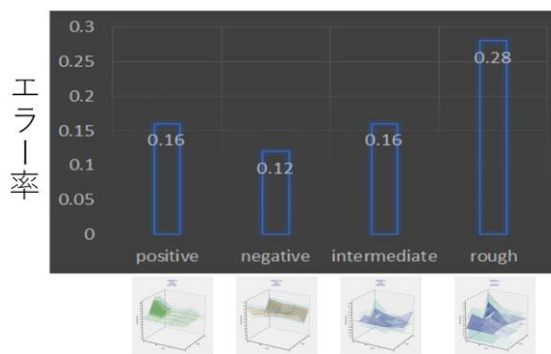


図9：18の deep learning model を用いた場合の、4つのクラスそれぞれの予測精度

今後は、構築した画像解析システムに改良を加え、より高精度な分類器の構築を進める。例えば、現在、deep learning モデルを作る際には、ひとつの角度からみた画像データしか用いていないが、研究者は、角度を変えながら見て判別しているの、より高精度なモデルの構築のために、多様な角度の画像データを用いて予測を行うシステムの構築を進める。

(2) ゲノム解析とその関連データベースの整備

今年度はSHOEを起点とした一連の処理を実現した。まず、SHOE 自体は、Garuda Platform の Gadget のみならず、Cloud Service としても提供可能となった(図10)。どちらも同様の検索結果が表示される(図11)。

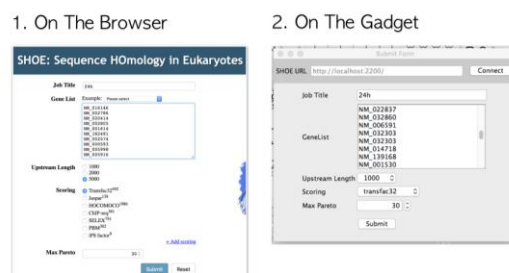


図10：SHOE の Cloud Version (左)と Garuda Gadget Version(右) (遺伝子リスト指定画面)

SHOE calculates Transcription Factors, Scores, etc ...

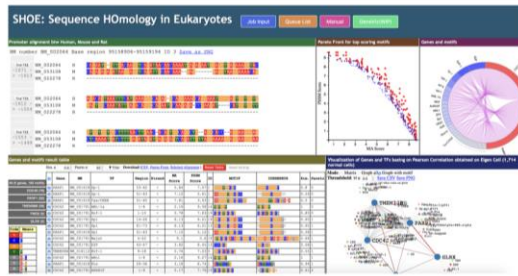


図 1 1 : SHOE 検索結果表示画面

さらに SHOE の結果から、CellDesigner に連動させ、遺伝子-転写因子マップの表示が可能となった(図 1 2)。

Import Results to CellDesigner in Garuda

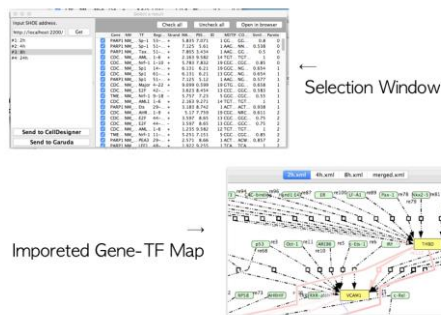


図 1 2 : SHOE-CellDesigner 連携: SHOE の結果から遺伝子・転写因子マップが検索され表示される

さらに複数の解析結果をマージし、より全体を俯瞰できる遺伝子-転写因子マップを構築することも可能となった(図 1 3)。

Working with CellDesigner Gadgets

Merged multiple maps with Merge Model gadget

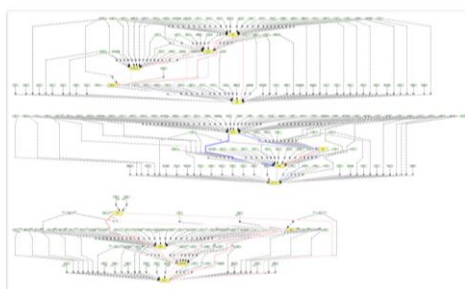


図 1 3 : 複数の解析結果をマージし、俯瞰的な遺伝子・転写因子マップを構成する

またこの拡大されたマップから、遺伝子と転写因子を抽出する(図 1 4)。

Species Extractor extracts Genes & TFs from the map

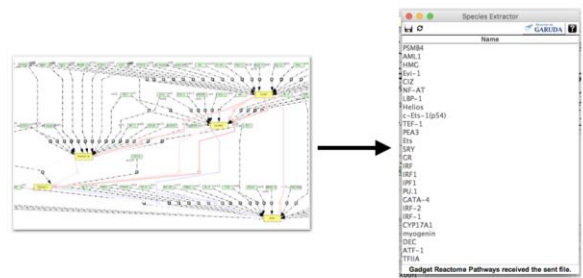


図 1 4 : 俯瞰的マップから、遺伝子・転写因子のリストを抽出する

抽出されたリストから Percellome DB や Reactome DB へとアクセスし、関連するデータなどにアクセスする(図 1 5)。

Extracted Genes & TFs can be also imported other gadgets

Example:

Percellome: searches elements from Percellome Project DataBase

REACTOME Pathways: searches Reactome Pathway for genes and proteins

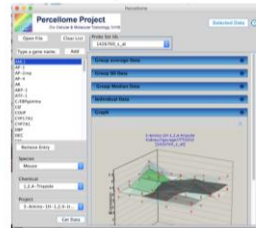


図 1 5 : Percellome や Reactome へと連動する。

これらの一連の解析を続けると極めて大きなネットワークが形成される。そこで、我々は、これに対応できる可視化ツールを開発した(図 1 6)。

[New Feature] GeneViz

Gene-TF Map Visualizer on the SHOE Web

Features

- Visualizing Multiple Maps
- Importing results from SHOE
- Finding common elements through maps
- Search elements
- Search by Attributes (receptor ...)

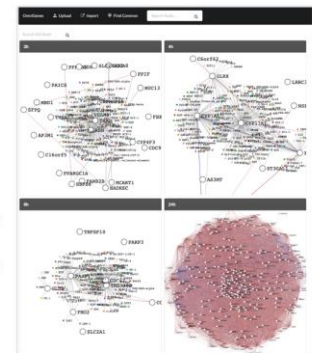


図 1 6 : 複雑なネットワークに対応する可視化ツール GeneViz

D. 考察

複雑な毒性機序の解析には、大規模データベースを柔軟に検索し、さらに高度な遺伝子動態から注目すべき遺伝子群を自動的に同定する技術が必要である。今回、我々は、Deep Learning を用いた手法でこれを実現し、その精度を確認した。また従来から開発を行っていたSHOEに関しては、その機能強化(軽量化、高速化)をさらに進めると同時に、Garuda Platform 準拠による他ソフトウェアとの連動性強化をさらに推し進めた。これらの結果、解析パイプラインの人工知能駆動化の可能性が見えてきた。

E. 結論

システムトキシコロジー解析技術の基盤整備及び応用開発についても予定通り推移しており、一般向けのリリース版も完成した。これらを基盤にプロジェクトの最終目標の達成、即ち毒性解析パイプラインの構築を進める。

F. 研究発表

1. 論文発表

- ① Nozomu Yachie; Robotic Biology Consortium; Tohru Natsume. Robotic crowd biology with Maholo LabDroids. *Nature biotechnology*. 35, 4, 310-312, April 2017.
- ② 北野宏明. ブロックチェーンの活路は人工知能との連携にあり. *ハーバード・ビジネス・レビュー*, 2017年8月号, 24-37, Aug. 1, 2017.
- ③ 北野宏明. ノーベル・チューリング・チャレンジ: 人工知能による科学的発見の再定義と生命科学の加速をめざして. *実験医学別冊「あなたのラボに AI (人工知能) × ロボットがやってくる」*, 22-31, Dec. 15, 2017.
- ④ 北野宏明. システム論とデータ駆動分析から可能となる老化研究. *実験医学増刊号「総力戦で挑む老化・寿命研究」*, 35, 20, 115-123, Dec. 15, 2017.

2. 学会発表

- ① Kitano, H. Nobel Turing Challenge. International workshop "From Genetic Networks to a Cellular Wiring Diagram", Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Apr. 26, 2017. (invited)
- ② Kitano, H. Nobel Turing Challenge: The Day AI win The Nobel Prize and Future of Civilization. LCSB seminar, Luxembourg Centre for Systems Biomedicine, Luxembourg, May 11, 2017. (invited)
- ③ 北野宏明. AI 駆動生命科学イノベーションとグランドチャレンジ. AI × Life Science シンポジウム, 日本橋三井ホール, 東京, May 17, 2017. (Plenary talk)
- ④ Kitano, H. Impacts of Artificial Intelligence for Pharmaceutical Industry: Disruptive Innovations in Scientific Discovery and Biomedical Sciences. The FIP 6th Pharmaceutical Sciences World Congress 2017(PSWC 2017), Stockholm Masaan, Stockholm, Sweden, May 21, 2017. (Plenary talk)
- ⑤ Kitano, H. Nobel Turing Challenge: Grand Challenge of AI, Robotics, and Systems Biology. ICRA 2017, Sands Expo and Convention Centre, Singapore, May 31, 2017. (Plenary talk)
- ⑥ 北野宏明. 人工知能による医療革命. 山口大学シンポジウム「人工知能・システム医学による難治性疾患への新たな挑戦」, ANA クラウンプラザホテル宇部, 山口, June 24, 2017. (invited)
- ⑦ 北野宏明. 人工知能とグランドチャレンジ: RoboCup から Nobel Turing Challenge まで. 第1回 AI・人工知能 EXPO, 東京ビッグサイト, 東京, June 28, 2017. (Plenary talk)
- ⑧ 北野宏明. 人工知能によるシステム・トキシコロジーの加速. 第44回日本毒性学会学術年会, パシフィコ横浜, 神奈川, July 11, 2017. (invited)
- ⑨ 北野宏明. Nobel Turing Challenge. THE NEW TEXT CONFERENCE, 虎ノ門ヒルズ, 東京, July 25, 2017. (invited)
- ⑩ Kitano, H. Systems Biology: Tools and Integrated Platforms for the 3Rs. 10th World Congress on Alternatives and

Animals in the Life Sciences, Washington State Convention Center, Seattle, USA, Aug. 23, 2017. (Plenary talk)

- ⑪ 北野宏明. 人工知能が切り拓く新たながん研究. 第76回日本癌学会学術総会, パシフィコ横浜, 神奈川, Sep. 29, 2017. (invited)
- ⑫ Kitano, H. Nobel Turing Challenge: Objectives and Implications. Symposium on AI and Society, Toranomon Hills, Tokyo, Oct. 10, 2017. (Keynote)
- ⑬ 北野宏明. System Biology, Data and AI. 第4回日本橋ライフサイエンスシンポジウム, 日本橋ライフサイエンスハブ, 東京, Oct. 14, 2017. (invited)
- ⑭ Kitano, H. Nobel Turing Challenge: Grand Challenge of AI, Robotics, and Systems Biology. DSAA2017 - The 4th IEEE International Conference on Data Science and Advanced Analytics, Shinagawa Prince Hotel, Tokyo, Oct. 20, 2017. (Keynote)
- ⑮ Kitano, H. Noble Turing Challenge. JSPS-IVA Seminar “The Nobel Turing Challenge - When will the first Nobel Prize be awarded for AI?”, IVA's Conference Centre, Stockholm, Sweden, Oct. 25, 2017. (invited)
- ⑯ Kitano, H. Nobel Turing Challenge. OECD conference ‘AI: Intelligent Machines: Smart Policies’, OECD, Paris, France, Oct. 26, 2017. (invited)
- ⑰ 北野宏明. Nobel Turing Challenge: Grand Challenge of AI and Systems Biology. 平成29年度日本バイオインフォマティクス学会・生命システム理論研究会「AIを用いた細胞システムの解明に向けて」, 京都大学 iPS 細胞研究所, 京都, Dec. 15, 2017. (invited)

G. 知的財産権の取得状況
なし