

生体影響予測を基盤としたナノマテリアルの統合的健康影響評価方法の提案  
分担研究課題名：(1)ナノマテリアルによる細胞内網羅的遺伝子発現データベースの構築、  
(2)機械学習による生体影響予測の試み

研究分担者：花方信孝 物質・材料研究機構 技術開発・共用部門 部門長

**研究要旨：**3年間の研究期間において、まずはナノマテリアルによる細胞内網羅的遺伝子発現データベースの構築のため、既存データベースを調査して、それらからデータを抽出し適切なラベリングやデータ処理を行なって基礎的なデータベースを構築した。続いて、このデータベースに対して、機械学習のアルゴリズムやパラメータの検討を行ない、生体影響予測を試みた。特に機械学習のラベリング方法および入力特徴量の調整を検討や、マイクロアレイ解析の一色法と二色法のデータ変換法の開発を進めた。また、実測遺伝子発現データを得るために、標準ナノマテリアルとして酸化チタンを用いた曝露毒性試験の実施および遺伝子発現マイクロアレイ解析を行なった。

## A. 研究目的

現在一般社会で開発・使用されているナノマテリアルが社会的に受容されるためには、そのリスクについて十分な安全評価手法が必要である。しかしながら、その手法として代表的な動物実験は費用的にも時間的にも高コストであることに加えて、近年の動物愛護原則から敬遠され、代替手法が求められている。そこでin vitro評価法の一環として、ナノマテリアルによる細胞内網羅的遺伝子発現データベースの構築方法および機械学習による生体影響予測モデルを検討することを研究目的とした。

具体的には、データベースを構築する上で元データとして使用する既存のデータベースを選定するために生命科学系データベースの調査、データベースの構築法および機械学習における学習用サンプル情報のラベリング方法および入力特徴量の調整など予測モデルの検討、マイクロアレイ解析の一色法と二色法のデータ変換法の開発、標準ナノマテリアルとして酸化チタン曝露時の毒性試験の実施および遺伝子発現マイクロアレイ解析を目指した。

## B. 研究方法

（既存データベースの調査）

生命科学系データベースのカタログであるIntegbioデータベースカタログ（科学技術振興機構、<https://integbio.jp/dbcatalog/>）に登録された1,684件（調査時）のデータベースを中心にインターネット上で公開されているデータベースの中から遺伝子発現や化学物質の暴露などに関するものなど本研究目的に適したものを検索して調査した。

（機械学習の実施環境）

使用するコンピュータのOSはMS WindowsとCent OS (Linux)の両方を用意した。ディープラーニング等の機械学習のライブラリとしてGoogleのKeras / Tensor Flowを採用した。プログラムを組むための言語はPythonを使用することとした。PythonのディストリビューションとしてAnacondaをインストールした。

（データベース構築および機械学習の調整と検討）

マイクロアレイ解析の既存の遺伝子発現データベースを利用するには一色法と二色法のデータをコンバートする必要があるその手法を開発する。機械学習においてはサンプル情報をラベリングしなければならないが、その方法を検討する。また、機械学習における入力特徴量について調整する。

#### (酸化チタンの毒性評価)

標準ナノマテリアルとして酸化チタンのナノ粒子を使用することとし、まずは毒性評価を行なうためWST-8アッセイを実施した。評価を行なった酸化チタンは国立医薬品食品衛生研究所より分与された以下の7サンプル。

- MT-150A (Lot#651105)
- MT-500B (Lot#1880902)
- AMT-100 (Lot#181102)
- TKP-102 (Lot#4190101)
- AMT-600 (Lot#6553)
- TiO<sub>2</sub> (Lot#1001)
- TiO<sub>2</sub> (Lot#1005)

WST-8アッセイには株式会社同仁化学研究所のCell Counting Kit-8 (CCK-8)を使用した。細胞はTHP-1細胞とRAW264細胞を用いて、それぞれ500,000cells/mL, 70,000cells/mLに調製した。96ウェルプレートにそれぞれ50 $\mu$ L/well, 100 $\mu$ L/wellで播種した。THP-1細胞には酸化チタンを50 $\mu$ L/well添加し、37 $^{\circ}$ Cで24時間後にCCK-8を10 $\mu$ L/well添加し4時間後に波長450nmで吸光度を測定した。RAW264細胞は播種から24時間後に培地を除去してから酸化チタンを100 $\mu$ L/well添加し、さらに24時間後にCCK-8を同様に添加して測定した。

なお、酸化チタンのナノ粒子の大きさは動的光散乱光度計 (DLS) により測定した。

#### (ZnO曝露細胞の一色法マイクロアレイ解析)

- 細胞： THP-1細胞およびA549細胞
- 曝露ナノマテリアル： 酸化亜鉛 (ZnO)
- 曝露濃度： 300 $\mu$ g/mL (THP-1細胞) または60 $\mu$ g/mL (A549細胞)
- 曝露時間： 6時間または24時間
- マイクロアレイ： Agilent SurePrint G3 Human GE Microarray 8x60K Ver. 3.0 (G4851C; GEO platform:GPL21185) 1枚
- ハイブリダイゼーション： 一色法
- マイクロアレイの割り当て: not shown
- マイクロアレイスキャナー： Agilent Surescan G2600D
- 画像数値化処理ソフトウェア： Agilent Feature Extraction v11.5

### C. 研究結果

(既存データベースの調査)

Integbioデータベースカタログを中心に調査を行なったところ、本研究の目的に適するいくつかのデータベースが見出された(表1)。そのうち最大のもはよく知られているNCBI運営のGene Expression Omnibus (GEO)であり、DataSetsで4,348件、Seriesで105,964件、Samplesで2,783,483件の遺伝子発現データが登録されている(2018/12/14調査時点)。今後はこのGEOのデータを主に利用することとした。また、Integbioデータベースカタログには収録されていないが、化学物質の曝露と遺伝子や病気との関連をまとめたデータベースとしてノースカロライナ州立大学運営のComparative Toxicogenomics Database (CTD)が存在し、補足的に利用可能と思われる。なお、公開されている既存データベースの中でも、全データを一括ダウンロードして再利用可能なデータベースと、検索はできるものの一部データしかダウンロードできないデータベースが存在した。後者のタイプのデータベースからウェブスクレイピングにより全データをダウンロードすることは可能かもしれないが、当該データベースのライセンス条件的に問題がないかどうか慎重な検討が必要と思われる。

#### (機械学習の実施環境)

機械学習ライブラリKeras / Tensor Flowをインストールしてプログラムが実行可能であることを確認した。マイクロアレイ発現データを用いた機械学習の前に、GEOデータベース上のデータと実測したマイクロアレイ発現データを用いたデータマイニングを試みた。

#### (データベース構築および機械学習の調整と検討)

同一RNAサンプルに対する一色法と二色法のマイクロアレイ解析データからコンバート手法の開発を進めた。遺伝子発現データベースは主にGEOデータベースを利用した。機械学習におけるサンプル情報のラベリングを手動で行うのは大量のデータについては困難なので、サンプルのディスクリプション情報から自動でラベリングしてみたが、精度に問題がある。入力特徴量については計算機の能力からある程度削減する必要があり、変動の小さい遺伝子を除く方向で検討した。

#### (酸化チタンの毒性評価)

標準ナノマテリアルとして酸化チタンのナノ粒子の使用することとした。これにより各分担研究

間の比較も可能となる。まずはMMT-8アッセイによる毒性評価を行なった(図1、図2)。浮遊細胞のTHP-1細胞に対しては7種類とも顕著な影響は認められなかった。

一方で付着細胞のRAW264細胞においては、MT-150A, MT-500B, TiO<sub>2</sub>-1005の3種類では影響がみられなかったものの、AMT-100, AMT-600, TKP-102, TiO<sub>2</sub>-1001の4種類では毒性が認められた。THP-1細胞とRAW細胞で実験の日時に差があり、ナノ粒子径はその都度DLSにより測定したが多少のばらつきがある。しかしながら、RAW細胞に毒性を示した酸化チタンとナノ粒子径との間に相関は認められない。

#### (ZnO曝露細胞のマイクロアレイ解析)

発現強度(ノーマライズ)：ノーマライズ方法は第3四分位値を基準とする75パーセンタイル法を使用した。発現強度の分布は概ね均等であった。

発現強度(プローブ)：今回使用したマイクロアレイスライドはスポット数としては1アレイあたり60,901個あるが、1つのプローブが複数のスポットにある場合があるので、プローブ単位で発現強度をまとめた。なお、プローブの総数は58,201個である。

階層的クラスタリング：全8アレイのデータのうち8アレイとも発現比が求まったプローブ22,003個の発現強度データで階層的クラスタリングを行なった(図3)。まず、THP-1細胞とA549細胞でクラスタが大きく分かれている。そしてそれぞれの細胞においてZnOに暴露したかしていないかでクラスタが分かれた。6時間後と24時間後の違いは比較的小さいことが分かった。また、A549細胞よりもTHP-1細胞の方がZnOの影響が大きいことが明らかになった。

#### D. 考察

生命科学系データベースとしてマイクロアレイ解析の生データが登録されている GEO データベースを選択したが、サンプル条件などのラベル付けについてはなお課題がのこった。手動でのラベル付けだけでなく、データのメタ情報からうまくラベルを生成する方法について検討の余地がある。Gene Ontology エンリッチメント解析のような結果をラベルとして利用することも検討したが、有意な結果がでないケースがあるので、利用は限定

的となる。

機械学習を実行する上で、ヒトの遺伝子の数はRefSeq データベースに限っても 26,000 個ほどもあるため、入力特徴量としては扱いにくい。そのため顕著の遺伝子に限って1/10程度の量に減らしてモデルを作成し機械学習を実行させている。この部分の効率的なデータの選択が本研究において重要である。

実測データとしてZnOを曝露した細胞の系を扱っていたが、他の分担研究と協調するためにも共通となる標準ナノマテリアルを使うことが適切として酸化チタンのナノ粒子を用いることとした。今回は7種類に酸化チタンについて毒性評価を行なったが、THP-1細胞とRAW264細胞で異なる結果を示すナノ粒子もあった。これらについてマイクロアレイ解析も実測し、機械学習のモデルを検証するデータとして利用できる。

#### E. 結論

細胞内網羅的遺伝子発現データベースの構築法および機械学習の予測モデルを開発することを目的として、既存データベースを調査して、それらからデータを抽出し適切なラベリングやデータ処理を行なって基礎的なデータベースを構築した。機械学習における学習用サンプル情報のラベリング方法および入力特徴量の削減方法について検討した。分担研究間で共通して使用する標準ナノマテリアルとして酸化チタンのナノ粒子を採用し曝露時の毒性試験を実施した。

#### F. 研究発表

##### 論文発表

1. Chinnathambi S, Hanagata N, Yamazaki T, Shirahata N. Nano-Bio Interaction between Blood Plasma Proteins and Water-Soluble Silicon Quantum Dots with Enabled Cellular Uptake and Minimal Cytotoxicity. *Nanomaterials*. 10(11), 2250, 2020.

#### G. 知的所有権の取得状況

本年度はなし。

表1.本研究目的に適すると思われる既存データベース

データベース名	URL	説明
Gene Expression Omnibus (GEO)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	NCBIが運営するマイクロアレイ解析等の遺伝子発現データのレポジトリ。
GEO Data Sets	<a href="https://www.ncbi.nlm.nih.gov/gds/">https://www.ncbi.nlm.nih.gov/gds/</a>	GEOのデータセット部分。実験条件等の情報が整備。
ArrayExpress	<a href="https://www.ebi.ac.uk/arrayexpress/">https://www.ebi.ac.uk/arrayexpress/</a>	EBが運営する遺伝子発現データアーカイブ。
CBEX	<a href="https://cibex.nig.ac.jp/data/">https://cibex.nig.ac.jp/data/</a>	DBJが運営する遺伝子発現データレポジトリ。
CellMontage	<a href="http://cellmontage.cbr.c.nrc.ca/">http://cellmontage.cbr.c.nrc.ca/</a>	マイクロアレイデータ検索・解析システム。
Open TG-GATEs	<a href="https://toxico.nibiohn.go.jp/">https://toxico.nibiohn.go.jp/</a>	医薬品等のin vivo/in vitro曝露データ集積データベース。
抗がん剤遺伝子発現データベース	<a href="http://scads.jfcr.or.jp/db/cs/">http://scads.jfcr.or.jp/db/cs/</a>	化合物の制がん作用と関連する遺伝子発現情報を提供するデータベース。
Comparative Toxicogenomics Database (CTD)	<a href="https://ctdbase.org/">https://ctdbase.org/</a>	環境曝露と人間の健康に関するデータベース。
Online Mendelian Inheritance in Man (OMIM)	<a href="https://www.ncbi.nlm.nih.gov/omim/">https://www.ncbi.nlm.nih.gov/omim/</a>	ヒトの遺伝子変異と遺伝病のデータベース。
Gene Signature Database (GeneSigDB)	<a href="https://genesigdb.org/genesigdb/">https://genesigdb.org/genesigdb/</a>	遺伝子発現解析結果からの変動遺伝子群のデータベース。
Integrative Disease Omics Database (IDoDB)	<a href="https://gemdb.ncc.go.jp/omics/">https://gemdb.ncc.go.jp/omics/</a>	多層的な疾患オミクス解析の統合データベース。
Functional Annotation of the Human Genome (FANTOM)	<a href="http://fantom.gsc.riken.jp/jp/">http://fantom.gsc.riken.jp/jp/</a>	ゲノムDNAから転写されているRNAの機能をカタログ化したデータベース。
COXPRESdb	<a href="https://coxpresdb.jp/">https://coxpresdb.jp/</a>	ヒトマウスラット他4種の共発現遺伝子データベース。
Human Protein Atlas	<a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a>	ヒトの器官、組織、細胞におけるタンパク質の発現および局在に関するデータベース。
H-ANGEL	<a href="http://www.h-invitational.jp/hinv/h-angel/">http://www.h-invitational.jp/hinv/h-angel/</a>	H-invitationalプロジェクトの構築した転写産物データに対する遺伝子発現データ。
BiGPS	<a href="http://biogps.org/">http://biogps.org/</a>	ヒトおよびマウスの遺伝子発現情報データベース。

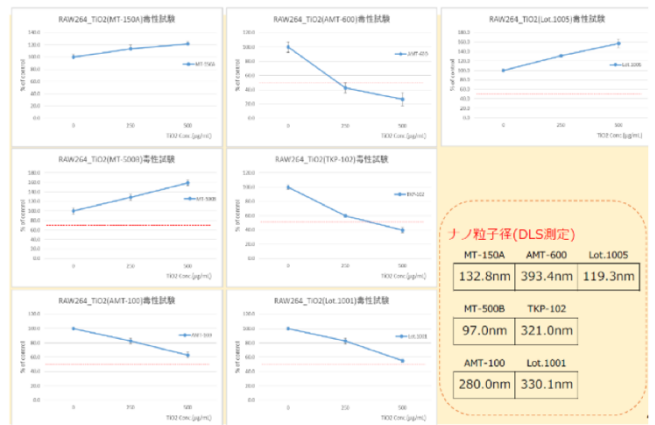


図2. RAW264細胞に対する酸化チタンの毒性評価

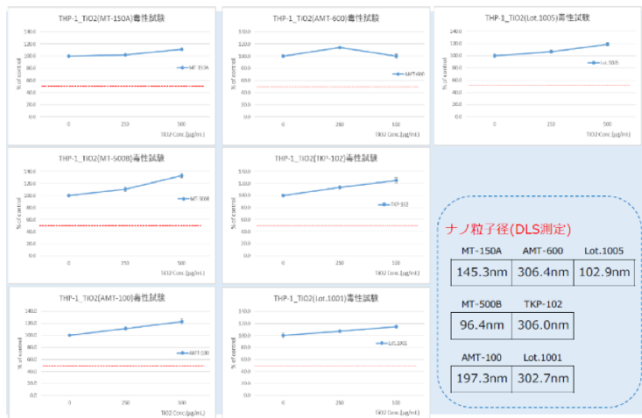


図1. THP-1細胞に対する酸化チタンの毒性評価

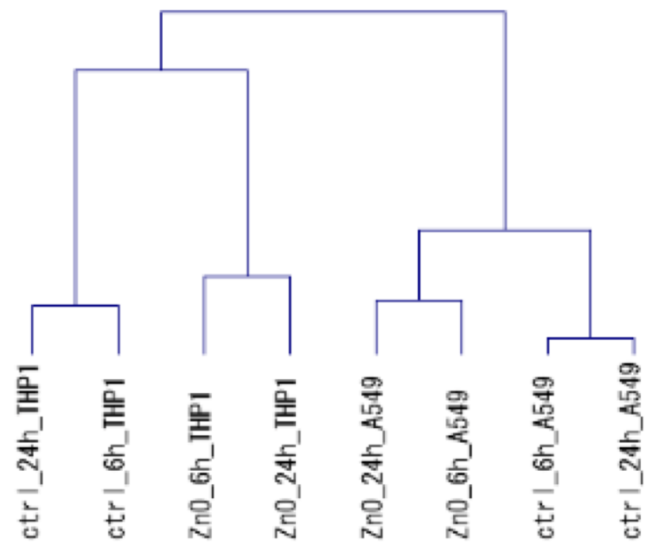


図3.階層的クラスタリング