

化学構造によるグルーピング及びリードアクロスによる神経毒性の *in silico* 予測

研究分担者：吉成浩一 静岡県立大学 教授

## 研究要旨

近年、インビトロ及びインシリコ手法を活用した発達神経毒性（DNT）評価手法の開発が用いられている。私達の研究室では、リードアクロスや階層的クラスタリングなどの手法を活用し、化学構造からインシリコ手法により DNT を予測する手法の開発に取り組んできたが、DNT 陰性物質の情報が限られており、データの少なさ、不確かさが予測手法の精度向上を阻んでいる可能性が考えられた。そこで本研究では、DNT 陽性物質の化学構造情報のみを利用して未試験物質の DNT リスクを予測する手法の開発を試みた。まず、論文情報から DNT 陽性物質と DNT 不明物質を定義し、それらの分子記述子を計算した。この分子記述子情報を利用して DNT 陽性物質間距離を計算し、陽性物質の近さを閾値として定義した。次に、各 DNT 不明物質と全陽性物質との距離を計算し、①閾値内に含まれる DNT 陽性物質数、②最も類似性が高い DNT 陽性物質との距離、を基準として、それぞれ各 DNT 不明物質の DNT リスクを予測した。その結果、両指標では異なる物質群が高リスクと予測され、それらにはカルバメート系農薬や有機リン系農薬、抗てんかん薬、枯葉剤の成分、オピオイド、レチノイン酸異性体、既知の神経毒性物質などが含まれていた。さらに、2つの指標によるリスクのランクの積を用いた統合評価を行い、単独の予測では検出されなかった物質の検出もでき、それらは神経毒性が疑われている物質であった。以上、本研究により、未試験物質の DNT リスクを化学構造情報からインシリコ手法で簡便に予測する手法を開発できた。今後、クラス分け等による精度の向上、インビトロ試験による検証等を行っていきたい。

## A. 研究目的

化学物質の発達神経毒性（DNT）を動物実験で評価するには多大なコストと時間を要する。また、近年動物福祉の観点から、実験動物を用いた安全性試験の代替法の開発が求められており、インビトロ及びインシリコ手法を活用した評価手法の研究開発が進められている。しかし、DNTに関しては、公的に利用可能な毒性試験データベースがないことなどの理由から研究が遅れている。

そこで私達は、本研究班において、DNT評価におけるインシリコ手法の有用性を提案するために、既存情報からDNT予測のためのデータセットを構築し、DNTを示す物質のグルーピング解析や階層的クラスタ解析並びにリードアクロス手法によるDNT評価手法の開発に向けた基礎検討を行ってきた。これまでに、化学構造情報を数値化した分子記述子を利用して化学物質の類似性を評価し、その結果に基づいたクラスタリングやリードアクロスを行い、使用する分子記述子の種類、毒性予測に利用する類似物質の数などの検討を行った。その結果、化学構造に基づく手法によりある程度の精度でDNTの予測が可能であることが明らかになった一方で、使用したデータセットの不確かさが結果に大きな影響を与えていること、特にDNT陰性物質の正確性に疑問が感じられた。実際、インシリコ手法で得られた予測結果を共同研究者の協力のもとインビトロ試験等で検証すると、データベース上の陰性物質がインビトロ試験では陽性であったケースも認められた。これらの結果から、インシリコ手法による予測を単独で用いるのではなく、インビトロ試験と組み合わせた予測が有用であると考えられた。

そこで本年度は、これまでに収集したDNT陽性物質の情報を活用して、DNTリスクが高い物質を化学構造情報からインシリコ手法で簡便にスクリーニングする方法

の開発を試みた。

## B. 研究方法

### 1. データセット

これまでの研究で利用してきた文献 (*Neurotoxicol Teratol*, 52:25-35, 2015) から、DNTを示す物質及びDNT情報が無い物質を選択し、前者を「DNT陽性物質」（164物質）、後者を「DNT不明物質」（197物質）として使用した。これら化学物質の化学構造情報はPubChemから収集した。

### 2. 化学物質の類似性評価

alvaDescソフトウェア (Alvascience, ver 2.0) を利用して、対象物質の分子記述子を計算した。解析には、搭載されている33ブロックに分類された5,666記述子のうち、化学構造との関連を理解しやすい6ブロック (Constitutional indices, Ring descriptors, P\_V SA-like descriptors, Functional group counts, Atom-centred fragments, Molecular properties) に含まれ、被験物質で計算可能であった376記述子を利用した。

各分子記述子について、最大値が1、最小値が0となるように正規化したのち、物質間のユークリッド距離をRを利用して計算した。物質間距離は理論上の最大距離（使用した分子記述子数の平方根）で除して相対距離とした。この相対距離を物質間の類似性の指標とした。

## C. 研究結果

### 1. 類似物質数による予測

164種のDNT陽性物質について、それぞれの最近傍物質との相対距離セットのうち、90%が含まれる値を閾値として設定した。次に、197種のDNT不明物質それぞれについて、全陽性物質（164種）との相対距離を算出し、

横軸にDNT不明物質、縦軸に陽性物質との相対距離をプロットした(図1)。この図を元に、各DNT不明物質について、閾値内に含まれた物質数をカウントした。結果をヒストグラムとして図2Aに示した。全197物質の最低値は0、最大値は56、平均値は16.1、中央値14であった。約30%の物質は、0~3物質しか含まない一方で、12物質は、陽性物質の25%(41物質)以上を含んでおり、後者の物質は、非常に多くのDNT陽性物質と化学構造が類似しており、DNTを示す可能性があるのではないかと考えられた。

そこで、これら12物質について確認したところ、カルバメート系農薬が2物質、他に抗てんかん薬(paramethadione)、オピオイド(tramadol)が含まれ(図3A)、これらはDNTを示す可能性が高いと考えられた。また、*N,N*-diethyl-3-methylbenzamide及びmethyl methacrylateでは神経毒性の懸念が報告されていた。

## 2. 物質間距離による予測

上記の解析では、近傍物質数でDNTリスクを予測していることから、用いたDNT陽性物質セットの構造的な偏りに左右される可能性が高い。すなわち、あるDNT陽性物質と非常に類似しているにもかかわらず、類似する化学物質の数が少ない場合には、DNTリスクは小さく(ランクが低く)判定される。

そこで次に、各DNT不明物質について、最も構造が類似しているDNT陽性物質との相対距離(最近傍相対距離)に着目し、その最近傍相対距離の短さによりDNTリスクをランク付けした。

最近傍相対距離の分布を図2Bに示した。また、図3Bには、距離が短い15物質とそれぞれの最近傍のDNT陽性物質を示した。15物質の化学構造を見ると、pentane、propyleneoxide、isopropyl alcoholのような比較的分子量から、isotretinoinや $\alpha$ -noracetylmethadolのような比較的大きな分子までが含まれていた。また、いずれも最近傍のDNT陽性物質とはメチル基等の官能基1つ程度の違いしかなく、化学構造が非常に類似していたことから、DNTを示す可能性が非常に高いと考えられた。実際、神経毒性を示すと考えられる有機リン系農薬が2物質、枯葉剤の成分である2,4,5-trichlorophenoxyacetic acid(別名2,4,5-T)、抗てんかん薬paramethadione、レチノイン酸異性体のisotretinoin、中枢神経毒性が知られているメトキシ酢酸の関連物質2-methoxyethanolなどが含まれていた。

前述の閾値内のDNT陽性物質数でのリスク判定で挙げられた12物質と共通していたのは、paramethadioneの1物質のみであった。

## 3. 物質数と物質間距離による統合的予測

閾値内のDNT陽性物質数によるリスク予測並びにDNT陽性物質の最近傍相対距離によるリスク予測において、重複する化学物質は1物質しか認められなかったことから、両手法による判定は異なる物質群を高DNTリスクと予測していると考えられた。そこで、これらの解析を組み合わせた予測を行うために、両手法でのDNTリスクを低い順にランクをつけ、2つのランクの値の積を「DNTリスクスコア」として算出した。

図3Cに上位15物質を示した。このうち、図3A、図3Bに含まれた物質数はそれぞれ6物質、7物質であり、抗てんかん薬のparamethadioneが共通していた。その共通性から示唆されるように、paramethadioneが本評価において最も高リスクと予測された。2-Ethoxyethyl acetate、styrene、nitrobenzeneの3物質は本予測においてのみ認められたが、いずれも神経系の毒性を示

すことが知られている。

DNTリスクスコアによる予測結果と個別予測結果の関連性を確認するために、全DNT不明物質について、物質数によるDNTリスク予測ランクと最近傍相対距離によるDNTリスク予測ランクをそれぞれ縦軸、横軸として散布図を作成した(図4)。その結果、両者のランクには有意な正の相関が認められた(Spearmanの $\rho$ 値:0.794、 $p < 0.001$ )。また、リスクの上位15物質もほぼ右上の領域に含まれていた。

## D. 考察

私達の研究室では、化学構造からDNTの有無を予測するためのインシリコ手法の開発を進めてきたが、利用可能な物質セットが限られていること、特にDNT陰性物質の情報がほとんどないことから、参考文献の情報を参考に、文献調査によりDNT情報がない物質をDNT陰性物質として解析を行ってきた。リードアクロスや階層的クラスタリングにより6割程度の物質の判別が可能であったが、様々な条件検討を行っても、それ以上の精度向上は認められなかった。その理由として、データセットには、複数のDNT陽性物質と化学構造が類似したDNT陰性物質が多く含まれており、それらが「偽陽性」として判定されること、また分子量が100未満程度の小さな化学物質や500を超える大きな化学物質についてはデータセットに類似物質がほとんど存在しないことなどが、精度向上が認められない主要な原因であると考えられた。

実際、前者の「偽陽性」と判定された物質が「真陽性」である可能性については、共同研究者のインビトロ試験結果から、真陽性である可能性が示された物質も存在した。そのため、これまでで使用してきたDNT陰性物質は、実際にはDNT陽性物質を多く含む可能性が考えられた。また、この一連の過程により、インシリコ手法によるDNT懸念物質のスクリーニングとインビトロ試験による検証という組み合わせは、DNT評価の動物実験代替法として有用であることが示唆された。そこで本研究では、これまでの「DNT陰性物質群」を「DNT不明物質」として再定義し、既知のDNT陽性物質の情報から「DNT不明物質」のDNTリスクの予測を試みた。

本研究では、164種のDNT陽性物質を活用して、それらの類似性からDNT不明物質のDNTリスクを3つのアプローチにより予測した。1つは、DNT陽性物質間での類似性(物質間相対距離)から、類似物質を定義する物質間相対距離の閾値を定義し、各DNT不明物質についてその閾値内に含まれるDNT陽性物質数による予測、2つ目は、各DNT不明物質と最も類似している(相対距離が近い)DNT陽性物質との相対距離による予測、3つ目はこれら2つの手法を掛け合わせた予測である。

1つ目の手法によりDNTリスクが高いと予測された12物質について確認したところ、カルバメート系農薬や抗てんかん薬、オピオイドなど、神経毒性を示すと考えられる物質が多く含まれていた。また、2つ目の手法においてDNTリスクが高いと予測された上位15物質についてその化学構造を確認したところ、既知の陽性物質との類似性が非常に高い物質が多数見出された。例えば、高用量曝露により催奇形性を示すと考えられるall-trans-レチノイン酸(tretinoin)の異性体であるisotretinoin、有機リン系農薬、枯葉剤成分、神経毒性を示すメトキシ酢酸の関連物質などが含まれていた。これらの結果から、2つの基準によるDNTリスク予測は共にある程度の精度でDNTリスクの高い化学物質を同定できていると考えられた。

一方で、構造を確認した化学物質数は限られている

ものの、両手法で共通して同定された物質は、抗てんかん薬のparamethadioneのみであった。DNT陽性物質であるtretinoinの異性体であるisotretinoinについては、最近傍相対距離による予測では検出されたが、類似物質数による予測では検出されなかった。これは、DNT陽性物質群中に、tretinoinやisotretinoinと類似した物質がほとんど含まれないことを示唆しており、実際、isotretinoinは閾値内にDNT陽性物質を2物質しか含まなかった。逆に、類似物質数による予測でDNTリスクが高いと予測されたカルバメート系農薬や枯葉剤成分の2,4,5-trichlorophenoxyacetic acidは、最近傍相対距離による予測では検出されなかった。これらのDNT不明物質は、その化学構造や既知情報からDNT陽性である可能性が非常に高いと考えられることから、2つの基準を用いた予測は共に有用であるものの、いずれの手法もDNTリスクが高い物質を見落とす可能性があることが示唆された。

そこで、両基準を加味した予測を行うために、両予測で算出された値をDNTリスクとしてランク化し、両ランクの値の積を「DNTリスクスコア」として定義し、予測に用いた。リスクが高かった上位15物質を確認したところ、1位～11位及び14位の12物質は、単独の基準による予測のいずれかで同定された物質であり、総合評価による予測で同定された物質は3物質であった。これらの結果から、統合予測は有用であると考えられた。

しかし、isotretinoinは197物質中102位となり、比較的lowリスクであると予測された。今回の解析では、ランク値の積を用いたことから、一方の指標が著しく低い場合には、総合評価では上位に入る可能性は低くなる。これらの問題点を克服するために、今後、2つ指標（図3の縦軸及び横軸の値）を複数のクラスに分類してマトリクス化することで、全DNT不明物質やデータセット外の物質のDNTリスクを複数のクラスに分類できるのではないかと考えている。さらには、DNTリスクが高いと予測された物質について、共同研究者の協力を得ながらDNT関連インビトロ試験等で検証することで、本予測手法の妥当性を検証したいと考えている。

## E. 結論

本研究では、DNT陽性物質の化学構造情報を活用した新たなインシリコDNTリスク予測手法を開発した。本手法は、①類似物質数、②類似性の高さ、並びに③その組み合わせにより、被験物質のDNTリスクをランク付けることが可能であり、実際に神経毒性が知られている有機リン系農薬やカルバメート系農薬、抗てんかん薬、中枢神経毒性物質などを検出することができた。本手法は、簡便に既知のDNT陽性物質の類似物質を検出することが可能であり、DNTリスクのスクリーニング手法としての活用が期待される。

## F. 研究発表

### 1. 論文発表

なし

### 2. 学会発表

高村真弥、大村奈央、保坂卓臣、志津怜太、竹下潤一、吉成浩一：分子記述子を用いたリードアクロスによる発達神経毒性予測手法の開発、フォーラム2023 衛生薬学・環境トキシコロジー、2023年9月13日、広島市

## G. 知的所有権の取得状況

### 1. 特許取得

なし

### 2. 実用新案登録

なし

### 3. その他

なし

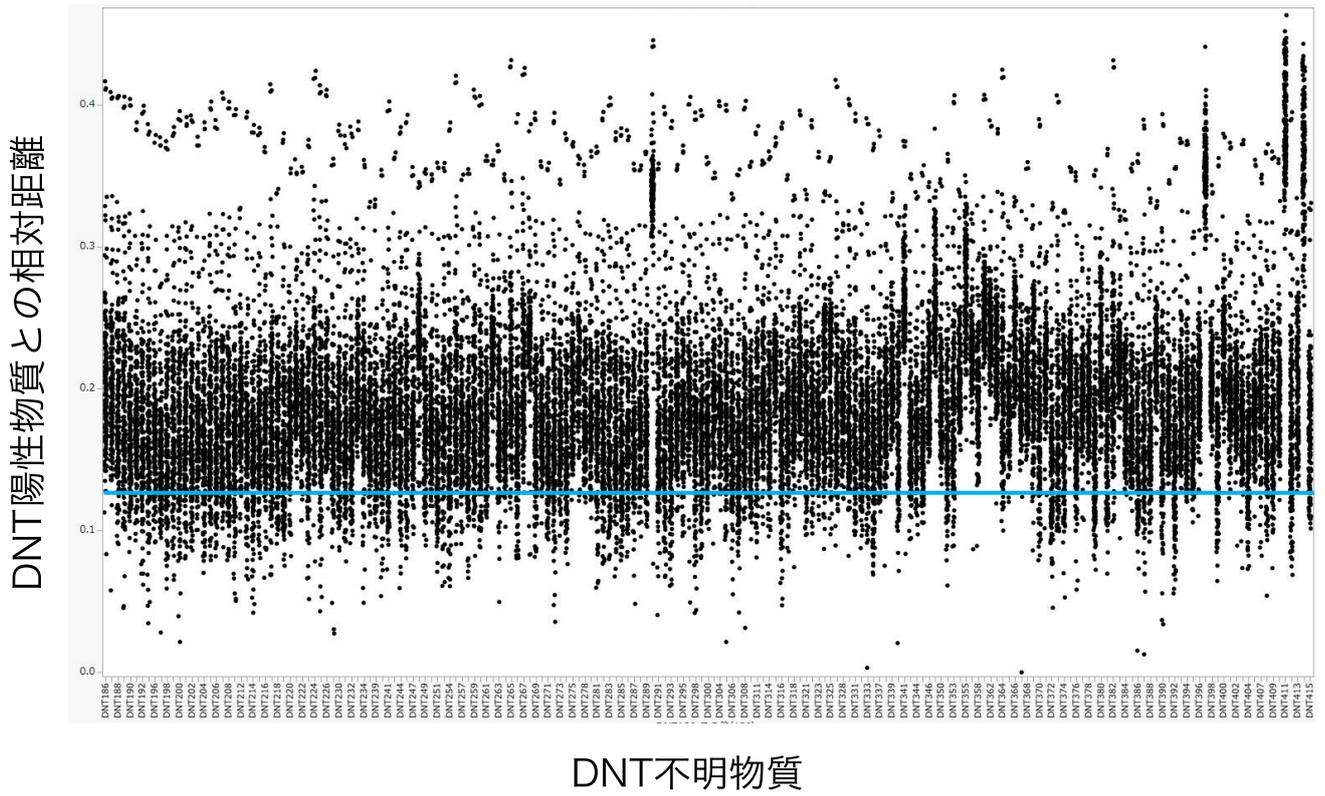
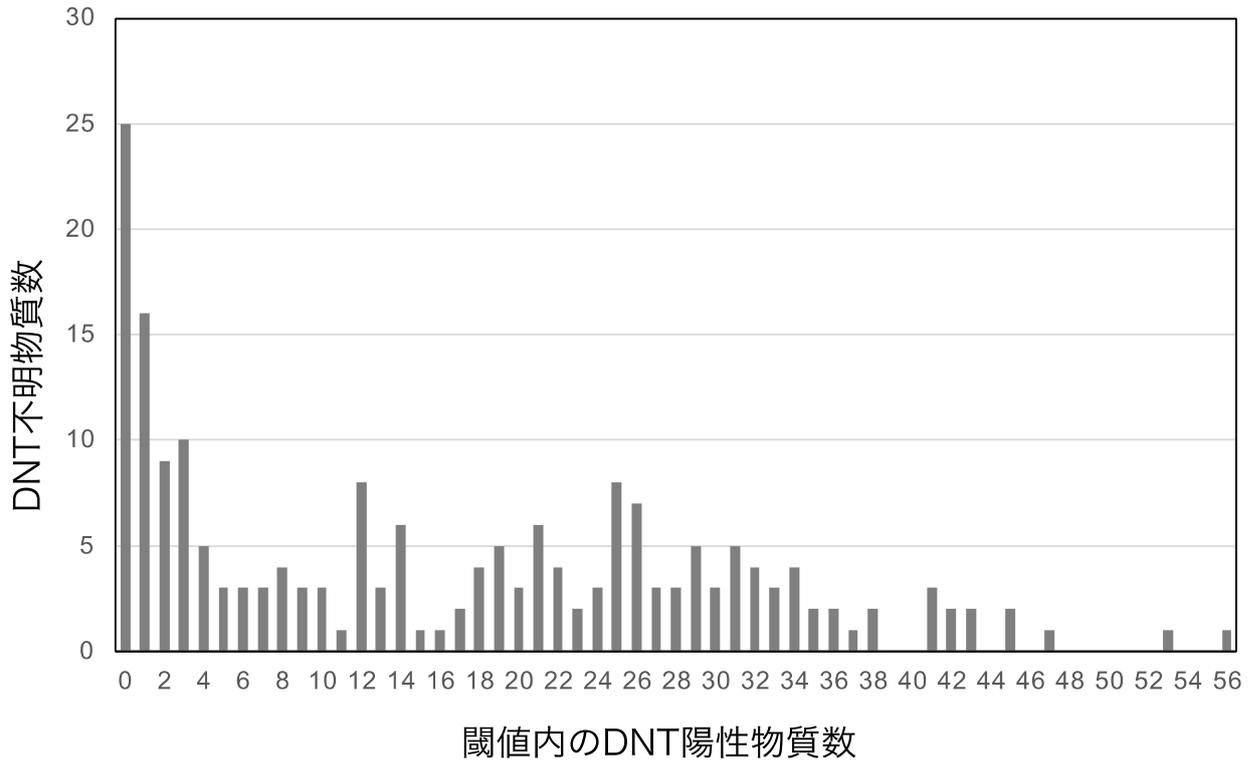
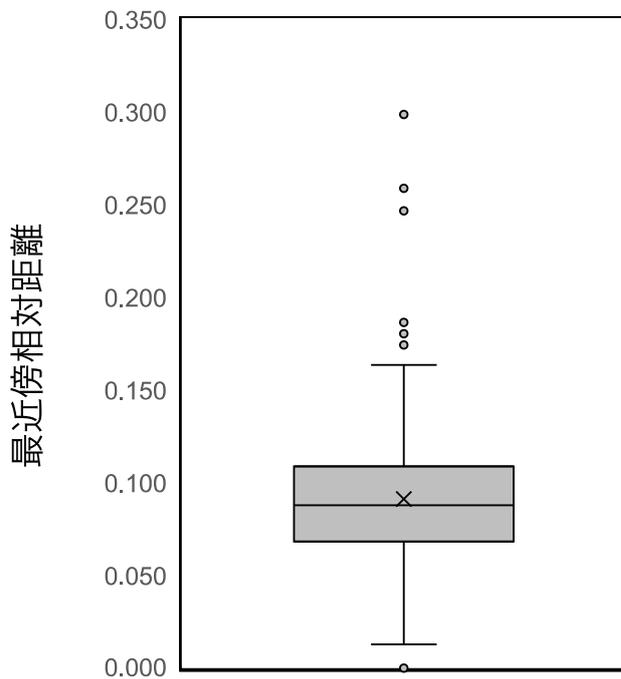


図1 各DNT不明物質とDNT陽性との物質間相対距離  
 197種のDNT不明物質（横軸）について、それぞれ164種のDNT陽性物質との相対距離を算出し、プロットした。青線は、DNT陽性物質で決定したDNTリスク判定閾値（0.128）を示している。

**A**



**B**



最小値	0.2987
最大値	0.0000
平均値	0.0911
10パーセンタイル	0.0466
第1四分位	0.0683
中央値	0.0881
第3四分位	0.1078
90パーセンタイル	0.1329

**図2 DNTリスク指標の分布**

A. 類似物質数による評価：図1において、197種のDNT不明物質について、閾値内に含まれるDNT陽性物質数を算出し、ヒストグラムとして示した。B. 各DNT不明物質と最も距離が近いDNT陽性物質との相対距離（197個）の分布を箱ひげ図で示し、統計量を表に示した。

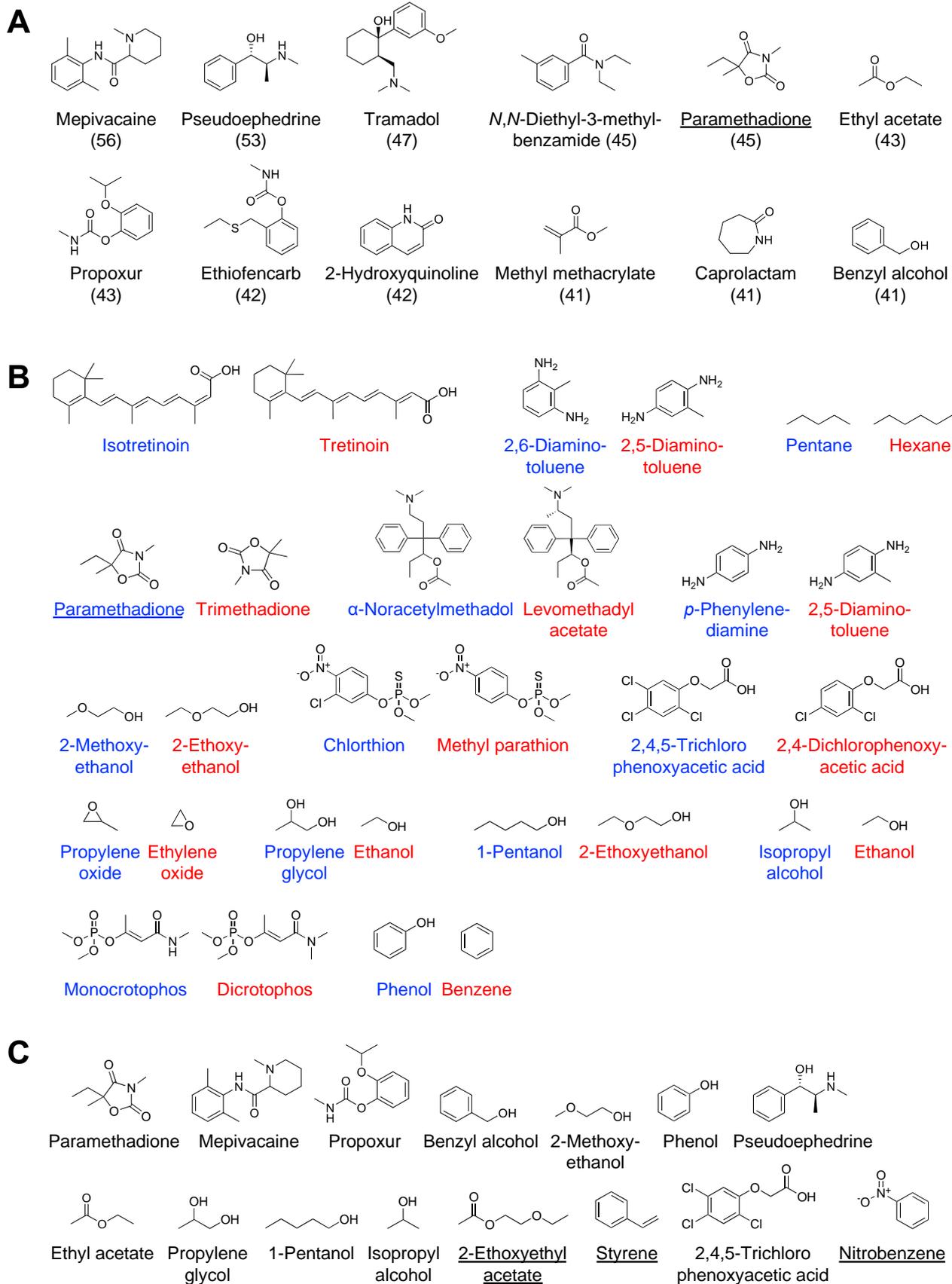


図3 DNTリスクが高いと予測されたDNT不明物質のリスト

A. 類似物質数が41以上であった12物質の化学構造を示した。B. 最近傍相対距離が小さかった上位15物質（青字）とその最近傍のDNT陽性物質（赤字）の化学構造を示した。下線はAにおいても認められた物質を示す。C. 類似物質数と最近傍相対距離から算出されたDNTリスクスコアの上位15物質の化学構造を示した。下線はCにおいてのみ認められた物質を示す。

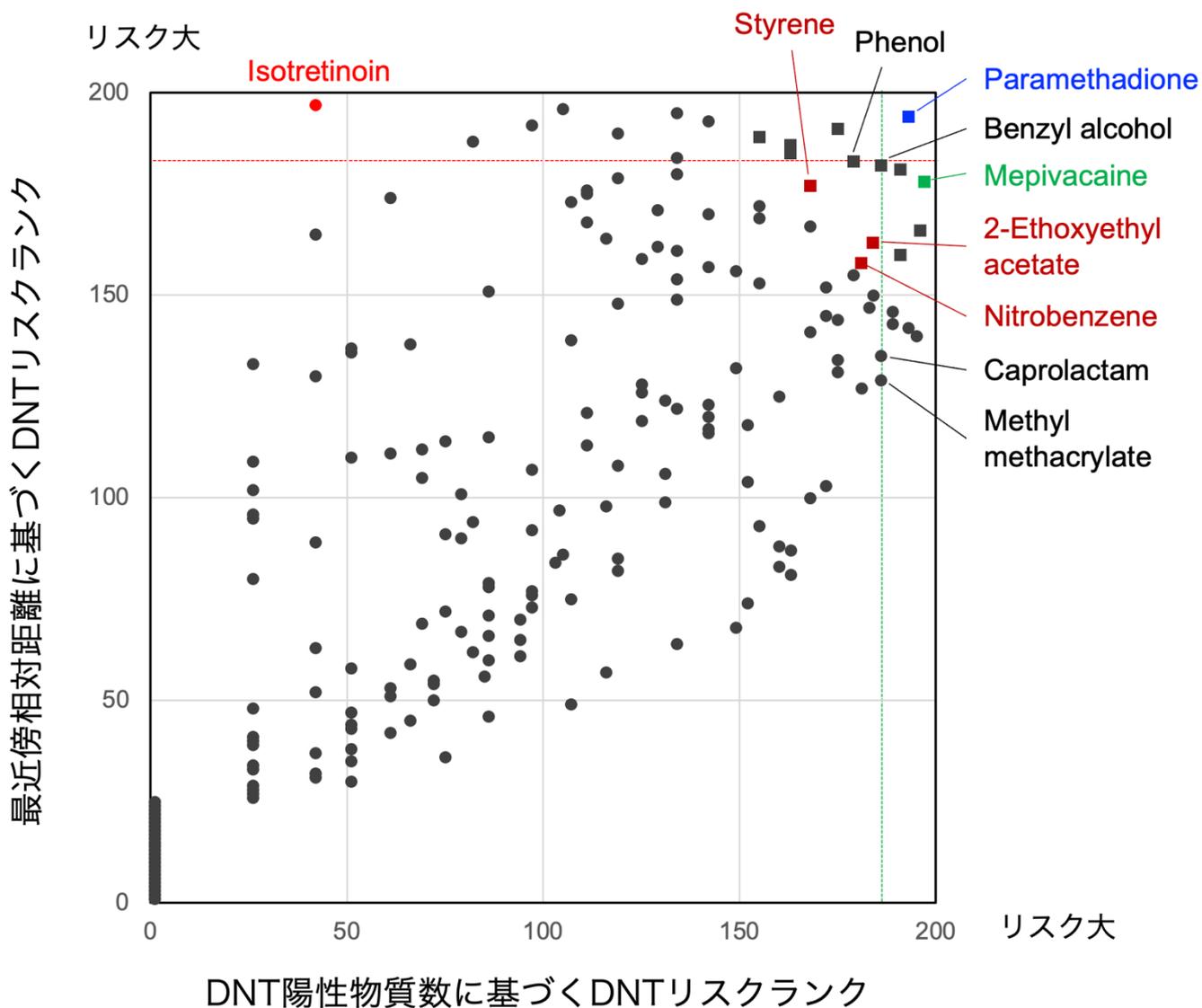


図4 2種類のDNTリスク予測ランクにおけるDNT不明物質の分布

196種のDNT不明物質について、類似物質数によるDNTリスク予測ランク（横軸）と最近傍相対距離によるDNTリスク予測ランク（縦軸）をプロットした。縦軸、横軸ともにランクが大きいほどDNT陽性物質との類似性が高く、DNTリスクは大きい。図3Cに示したDNTリスクスコアの上位15物質を■で示した。また、本文中に記載のある代表的な物質の名称を記した。