

令和7年度 厚生労働行政推進調査事業費補助金（化学物質リスク研究事業）
分担研究報告書

分担研究課題名： Ames/QSAR の活用に関する研究

研究分担者	増村健一	国立医薬品食品衛生研究所	安全性予測評価部・部長
研究分担者	山田隆志	国立医薬品食品衛生研究所	毒性部・部長
研究協力者	城島光司	国立医薬品食品衛生研究所	安全性予測評価部第三室
研究協力者	川村智子	国立医薬品食品衛生研究所	安全性予測評価部第三室
研究協力者	藤井真人	国立医薬品食品衛生研究所	安全性予測評価部第三室

研究要旨

化審法の既存化学物質スクリーニング評価の有害性情報を用いて、Ames/QSARの予測精度の検討を行った。昨年度の試行結果を踏まえて対象物質の再分類とQSAR予測を行った。既存化学物質481物質のうち、スクリーニング評価におけるAmes変異原性の判定は、436物質（91%）が陰性、1物質が疑陽性、40物質（8%）が陽性、4物質（1%）が強い陽性であった。Derek Nexus（知識ベース）とSarah Nexus（統計ベース）、CASE UltraにおいてはGT_EXPERT（知識ベース）とGT1_BMUT（統計ベース）を用いたQSAR予測の結果、4つのモデルはいずれも高い正確度（accuracy）を示した。2つのモデルを組み合わせた場合の予測性能は単独モデルとほぼ同等であった。陰性物質へのデータの偏りが大きいことや、各モデルが既存化学物質の既知見情報を学習済みであることが予測性に影響したと考えられる。一方で、モデルに既知物質として含まれていない陽性物質に対する検出感度は低い可能性が示唆された。今後は正しく予測されなかった物質について分析し予測の信頼性を検討する。

A. 研究目的

生活環境中の化学物質の多くについて十分な毒性情報がないためリスク評価が進展していない現状は化学物質管理における国際的な課題となっている。化学物質の人健康リスク評価を効率的に進めるためには、*in silico* 評価手法の開発と行政利用への取り組みが重要である。化学物質の構造情報を用いた定量的構造活性相関（QSAR）、類似物質の情報からの類推（リードアクロス）、グループ化による包括的な評価等の *in silico* 手法を用いた毒性予測及び評価法の開発と基盤整備を行うことが重要である。また、世界的な動物実験削減と 3Rs の流れに対応し、*in vivo* 試験にかわる New Approach Methodology (NAM) の活用が求められる。有害性発現経路（Adverse Outcome Pathway: AOP）に基づいた統合的アプローチ（Integrated Approaches to Testing and Assessment: IATA）により、*in silico*, *in vitro* のデータを活用した人健康影響評価の科学的な基盤を構築することが必要である。本研究では化学物質管理の現場の課題解決に資する有用な *in silico* 評価手法の開発・提案を行う。「化学物質の審査及び製造等の規制に関する法律」（以下、「化審法」という。）の毒性エ

ンドポイントに着目し、人健康影響評価への利用を目指す。規制での利用が進む Ames/QSAR の予測性・信頼性向上を目指すとともに、より複雑な毒性についても統合的評価に取り組み、専門家判断を支援する統合的リスク評価系の構築を目指す。本分担研究「Ames/QSAR の活用に関する研究」においては、Ames/QSAR の行政利用場面の拡大を目的として、化審法における既存化学物質情報を用いた Ames/QSAR の予測精度検討を実施する。予測性と適用範囲の分析を行い、一般化学物質の評価における Ames/QSAR 活用場面の考察と提案を行う。

B. 研究方法

化審法における既存化学物質の情報を用いて Ames/QSAR を実施し、QSAR 予測結果と Ames 試験情報との一致率等から予測精度と適用範囲を検討し、Ames/QSAR の利用場面を提案する。R7 年度は、化審法の既存化学物質スクリーニング評価に用いられる Ames 試験情報と判定結果に対して、QSAR モデルによる Ames 変異原性の予測精度評価を行った。化審法における既存化学物質の評価においては、厚生労働省が担当する人健康影響評価に関して、国立医薬品食品

衛生研究所安全性予測評価部が評価書等の原案作成や行政支援を行っている。本研究では、既存化学物質スクリーニング評価において安全性情報の収集と整理を行っている安全性予測評価部第1室（井上薫室長）の協力により、平成27年度以降の既存化学物質スクリーニング評価を対象とした。Ames試験の判定情報として、既存化学物質スクリーニング評価の有害性クラス判定に用いられる有害性情報のうち、「Ames試験（復帰突然変異試験）」の情報（「陽性」「陰性」等）を用いた。この判定結果は、国内外の評価機関による評価書・データベース等の情報に基づいて評価を行い、化学物質安全性評価委員の専門家が確認し、化学物質調査会資料として提出、公開されたものである。

国立医薬品食品衛生研究所安全性予測評価部が2015年から2024年までに評価した既存化学物質情報をもとにAmes/QSAR検討用のデータセットを作成した。R6年度に行った試行結果を踏まえて、R7年度は曖昧さを除くために再分類を行った。具体的には、既存化学物質スクリーニング評価において審議会に提出された評価対象物質を対象として、リードアクロスのために検討された類似物質は除いた。また、化学構造情報（SMILES記法）の取得方法を一元化し、対象物質のCAS RN®を用いてSciFinder®で構造情報を検索しSMILESを取得した。

スクリーニング評価を実施した619物質について、高分子化合物、無機化合物、構造未特定化合物等を除いた481物質を本研究の対象としてAmes/QSARを実施した。R6年度にAmes/QSARを試行した結果を踏まえて、再度予測を行った。2種類のQSARソフト（Lhasa社のNexus 2.9.1およびMultiCASE社のCASE Ultra 1.9.2.7）を用いた。Nexusにおいては、Derek Nexus 6.5.1（知識ベース）とSarah Nexus 5.1.0（統計ベース）で予測を行った。また、CASE Ultraにおいては、GT_EXPERT（知識ベース）とGT1_BMUT（統計ベース）で予測を行った。QSARソフトはそれぞれ解析時点での最新バージョンを用いた。

予測精度の検討においては、各ソフトの独自判定項目を「陽性」「陰性」の2値に集計した。具体的には、Derek Nexusではequivocal, plausible, probableを「陽性」、inactiveを「陰性」として集計した。Unclassified or misclassified features判定の物質は集計から除いた。Sarah Nexusではequivocal, positiveを「陽性」、negativeを「陰性」とした。Outside Domainの物質は集計から除いた。GT_EXPERTおよびGT1_BMUTではinconclusive, posit

ive, known positiveを「陽性」、negative, known negativeを「陰性」とした。Out of Domainの物質は集計から除いた。Ames判定結果とAmes/QSAR予測結果の2x2行列を用いて、各QSARモデルの感度（sensitivity）、特異度（specificity）、正確度（accuracy）等を算出した。また、単独モデル及び2つのモデルを組み合わせた場合の予測結果の比較を行った。さらに、CASE Ultraを用いて、モデルに「既知の陽性/陰性」として含まれていないAmes陽性物質に対する検出感度を検討した。

（倫理面の配慮）

本研究では動物を用いた実験は実施しない。

C. 研究結果

R6年度に行った試行の結果を踏まえて、R7年度は対象物質の再分類を行い、あらためて化学構造（SMILES）の取得およびAmes/QSAR予測を行った（図1）。

既存化学物質スクリーニング評価サマリー表の対象物質（1135物質）のうち、審議会番号の有無に基づいて619物質を抽出した。次に、対象物質のCAS RN®を用いてSciFinder®で構造情報を検索し、SMILESを取得した。619物質のうち、対象物質のAmes試験判定結果があるものが571物質、ないものが48物質であった。この48物質には類似物質の情報を用いたリードアクロスによる評価を行った物質が含まれるが、今回の解析対象からは除いた。さらに、以下に該当するものはAmes/QSAR予測の対象から除いた。1. 高分子化合物（28物質）、2. 無機物（8物質）、3. 構造不定物質（53物質）、4. その他（1物質）。その結果、481物質をAmes/QSAR予測の対象とした。

Ames試験に基づくスクリーニング評価結果は、対象とした481物質中の436物質（91%）が陰性、1物質が疑陽性（equivocal）、40物質（8%）が陽性、4物質（1%）が強い陽性であった。2種のQSARソフト（NexusおよびCASE Ultra）を用いてAmes変異原性を予測し、各ソフトの独自判定項目を「陽性」と「陰性」に分類して、参照データ（Ames判定結果）に対する予測結果を比較した（図2）。

4つのモデル（Derek（知識ベース）、Sarah（統計ベース）、GT_EXPERT（知識ベース）、GT1_BMUT（統計ベース））における感度（sensitivity）はそれぞれ0.452, 0.674, 0.651, 0.651であり、Sarahが最も高値、Derekが最も低値であった。特異度

(specificity) はそれぞれ0.972, 0.938, 0.981, 0.970であり、GT_EXPERTが最も高値、Sarahが最も低値であった。正確度 (accuracy) はすべてのモデルで0.9以上 (0.914~0.952) であった。

また、2つのモデルを組み合わせた場合に予測性が向上するかどうかの検討を行った。具体的には、2つのモデルの両方が陰性の場合に陰性とし、1つ以上のモデルで陽性の場合には陽性とする手法を用いた。その結果、感度 (sensitivity) はSarah Nexus + GT_EXPERTおよびSarah Nexus + GT1-BMUTの組み合わせが最も高値 (0.814) であり、特異度 (specificity) はGT_EXPERT + GT1-BMUTの組み合わせが最も高値 (0.967) であった。2モデルを組み合わせた場合の正確度 (accuracy) はすべての組み合わせで0.9以上 (0.911~0.939) であり、単独モデルの場合と同等であった。

さらに、未知物質に対する予測性を検討するため、CASE Ultraについて、既知見情報を意味するknown positive及びknown negativeと判定された物質を除いて集計を行った (図3)。対象とした481物質中、GT_EXPERTでは447物質 (93%) がknown positive or negative判定、6物質 (1%) がOut of Domain判定であった。残りの28物質中、Ames陽性判定物質は1物質であり、QSAR陰性であった。Ames陰性判定物質は27物質であり、うち26物質がQSAR陰性、1物質がQSAR陽性 (inconclusive判定) であった。GT1-BMUTでは481物質中379物質 (79%) がknown positive or negative判定、8物質 (2%) がOut of Domain判定であった。残りの94物質中、Ames陽性判定物質は1物質あったが、QSAR陰性であった。Ames陰性判定物質は93物質であり、うち86物質がQSAR陰性、7物質がQSAR陽性 (positive判定5物質 + inconclusive判定2物質) であった。両モデルが陰性判定したAmes陽性判定1物質は同一物質であった。限られた物質数での条件ではあるが、モデルに「既知の陽性/陰性」として含まれていないAmes陽性物質に対する感度は低い結果となった (0/1)。

D. 考察

化審法における既存化学物質のスクリーニング評価のために国立医薬品食品衛生研究所安全性予測評価部が作成している安全性情報サマリー表に基づいて、Ames変異原性予測に用いるための対象物質の抽出と整理を行った。R6年度の試行では、解析対象物質数を増やすために幅広く対象物質を抽出してQSAR解析を行った。その結果を踏まえて、R7年

度はより曖昧さを除いた物質選択、構造情報収集を行い、QSAR解析時点で最新バージョンのQSARソフトを用いて再解析を行った。CAS番号を用いてAmes/QSAR予測に必要な化学構造情報 (SMILES) を取得する際、複数の情報源から収集するとCAS番号と化学構造が適切に対応しているかどうか確認が難しくなる場合があるため、SciFinderを用いて統一的にSMILESを取得した。一部の物質では類似物質の情報を用いたリードアクロスによって評価が行われているが、本解析では評価対象物質のAmes試験判定情報がない物質は除外した。また、Ames試験判定情報がある571物質のうち、無機物、高分子化合物、構造不定物質等のAmes/QSAR解析に適さない対象外物質が90物質あった。このことは、既存化学物質スクリーニング評価の対象物質のうちAmes/QSARが適用できない物質群が約16% (90/571) であることを示唆している。これは昨年度の試行結果の値 (12%) とほぼ同等であり、評価が必要な既存化学物質の8~9割はAmes/QSARが利用できる可能性があることを示唆していると考えられた。今回の解析対象からは除いているが、構造類似物質の試験情報を用いて毒性の類推を行うリードアクロスについても、妥当性判断にAmes/QSARを利用できる可能性があると考えられ、構造類似物質のAmes/QSAR予測性の解析は今後の課題と考える。

CAS番号、SMILES情報、Ames試験判定結果が収集できた既存化学物質481物質についてAmes/QSAR予測を行った。Ames試験判定結果は、436物質 (91%) が陰性、1物質が疑陽性 (equivocal)、40物質 (8%) が陽性、4物質 (1%) が強い陽性であった。陰性に大きく偏ったデータセットとなっており、このような偏りは統計解析に影響することから、予測性能を検討する際には注意が必要である。一方で、同様の特徴は先行研究であるAmes/QSAR国際チャレンジプロジェクトでもみられており、一般化学物質の約1割がAmes陽性物質であることが今回の既存化学物質データでも確認されたものと考ええる。Ames/QSAR予測について、ICH M7 (医薬品中DNA反応性 (変異原性) 不純物の評価及び管理) ガイドラインでは、相補的な2つのQSAR法 (専門的経験に基づく知識ベースの方法及び統計ベースの方法) を用いて実施することとされている。今回はLhasa社のNexusおよびMultiCASE社のCASE Ultraを用いた。これらは化審法の新規化学物質審査においてAmes/QSAR参考資料の作成に使用されているソフトである。Nexusにおいては、Derek Nexus (知識ベース) とSarah Nexus

(統計ベース)を用いて予測を行った。CASE Ultraにおいては、GT-EXPERT(知識ベース)とGT1-BMUT(統計ベース)を用いて予測を行った。陽性陰性の予測結果の定義がソフトによって異なるため、各ソフトの独自判定項目を「陽性」「陰性」の2値に集計した。その際、inactiveまたはnegativeと判定されたものを「陰性」とし、それ以外を「陽性」とする保守的な分類を行った。QSARモデルの学習データに存在しない構造の物質については予測の信頼性が低いと考えられるため、out of domainと判定された場合は集計から除いた。

4つのQSARモデルについて、単独モデルを用いた予測では、正確度(accuracy)が全て0.9以上であった。この高い数値は、母集団の9割が陰性物質であるというデータの偏りを反映していると考えられる。感度(sensitivity)についてはDerek Nexusが、特異度(specificity)についてはSarah Nexusが他のモデルと比べてやや低値を示したが、この2つはそれぞれ知識ベースと統計ベースのモデルであり、両者を組み合わせて相補的に用いることで予測性を高められる可能性が考えられた。

今回用いた4つのQSARモデルについて、任意の2つのモデルを組み合わせた場合の予測精度について検討した。この場合、2つのモデルでいずれも陰性予測となったものを「陰性」、それ以外を「陽性」とする保守的な手法を用いており、単独モデルでの予測よりも感度(sensitivity)は高く、特異度(specificity)は低くなると予想された。6通りの組み合わせについて集計した結果、正確度(accuracy)は全て0.9以上であり、単一モデルとほぼ同等であった。この結果から、今回用いた既存化学物質のAmes/QSAR予測においては、どのモデルも高い予測性能を持っており、2つのモデルを組み合わせた場合の予測性能の向上は限定的であることが示唆された。知識ベースと統計ベースの組み合わせによる予測性能の向上は、今回の条件では明らかでなかった。原因としては、データが陰性物質に偏っているため高い陰性予測率が統計に影響したこと、また、多数の既存化学物質に関する既知の試験結果がQSARモデルの学習に組み込まれており、今回予測対象とした物質の多くについてQSARモデルが既知見情報として予測結果を出力していることが挙げられる。今回用いた商用Ames/QSARモデルは既存化学物質の既知見情報を学習しており、化審法スクリーニング評価における有害性情報の判定と高い確率で一致した(正確度(accuracy)0.9以上)。

このことは、Ames/QSARソフトは既知の有害性情報を含めた情報検索と有害性予測に有用であることを示唆している。

さらに、モデルに既知陽性/陰性として含まれていない既存物質に対する検出感度を検討するため、CASE Ultraを用いて、モデルが既知陽性/陰性として予測した物質を除いた集計を行った。GT-EXPERTでは、Ames試験陽性判定1物質中の0物質を陽性と予測し、Ames試験陰性判定27物質中の26物質を陰性と予測した。GT1-BMUTでは、Ames試験陽性判定1物質中の0物質を陽性と予測し、Ames試験陰性判定93物質中の86物質を陰性と予測した。陰性物質の予測性能は概ね高いと考えられた。一方で、1物質のみではあるが、モデルに既知陽性/陰性として含まれていない陽性物質に対する検出感度は0/1であり、十分とは評価できない結果であった。この結果は、学習データにない新規化学物質や化学構造が大きく異なる物質に対する陽性予測の限界を示唆している可能性がある。Ames/QSARはハザード同定のためのスクリーニングとして用いられる場面が多いと考えられるため、対象物質群の中に少数存在する陽性物質を見落とさずに検出できる性能は重要である。利用場面としては、Ames/QSARで陽性と予測された物質についてAmes実試験でフォローアップを行うケースが考えられる。このような利用法について妥当性を検証するためには、既知見に含まれない陽性物質を用いた予測精度の検討が重要である。

また、各予測値と参照データとの差異については、陽性陰性の予測結果の定義がソフトによって異なる点にも留意し、個別の分析が必要である。特に既知見の利用とデータ解釈の違いは、いわゆる専門家判断に関わる部分であり、判定結果に大きく影響する可能性がある。

現在の化審法の新規化学物質の評価において複数のAmes/QSAR予測結果が参考資料として利用されているが、各QSARモデルの予測精度、または知識ベースと統計ベースの精度比較や相補性の検討は十分でない。個々のQSAR予測結果をどのように総合判断に用いるかは今後の課題である。

E. 結論

化審法の既存化学物質スクリーニング評価の有害性情報を用いて、Ames/QSARの予測精度の検討を行った。昨年度の試行結果を踏まえて、スクリーニング評価におけるAmes変異原性の判定情報とAmes/QSAR予測の比較を行った。4つのQSARモデルの予測

結果はいずれも高い正確度を示した。2つのモデルを組み合わせた場合の予測性能は単独モデルとほぼ同等であった。陰性物質へのデータの偏りが大きいこと、各モデルが既存化学物質の既知見情報を学習済みであることが予測性に影響したと考えられた。一方で、モデルに既知の陽性として含まれていない陽性物質に対する検出感度は低い可能性が示唆された。今後は正しく予測されなかった物質について分析し予測の信頼性を検討する。

F. 健康危険情報

該当なし。

G. 研究発表

1. 論文発表

(発表誌名・頁・発行年等も記入)

1. Murata Y, Akagi J, Doi Y, Iso T, Umamo T, Masumura K, Matsumoto M, Toyoda T, Ogawa K: Evaluation of 13-week repeated-dose oral toxicity of zirconium(IV) butoxide in Crl:CD(SD) rats. Regul Toxicol Pharmacol. 2026;164:105968.
2. Wada NA, Takeiri A, Motoyama S, Matsuzaki K, Tanaka K, Matsuo S, Fujii-Takeiri E, Tateishi H, Matsumoto K, Niimi N, Sassa A, Grúz P, Masumura K, Mishima M, Jishige KI, Sugiyama KI, Nohmi T: Mitomycin C-induced DNA double-strand breaks are enhanced by catalytical inactivation of DNA polymerase κ in mice. Genes Environ. 2025;47(1):22.
3. Matsumura S, Hosoi S, Hirose T, Otsubo Y, Saito K, Miyazawa M, Kawade A, Hakura A, Kakiuchi D, Asakura S, Koyama N, Okada Y, Chikura S, Kimoto T, Masumura K, Suzuki T, Sugiyama KI. Whole genome mutagenicity evaluation using Hawk-Seq™ demonstrates high inter-laboratory reproducibility and concordance with the transgenic rodent gene mutation assay. Genes Environ. 2025;47:13.
4. Matsumoto M, Natsume M, Iso T, Umamo T, Murata Y, Masumura K, Horibata K, Sugiyama KI: Increased mutagenicity in the liver of MutaMouse females following oral treatment with commercial-grade toluene diisocyanate. Genes Environ. 2025;47:12.
5. Iso T, Umamo T, Sugiyama KI, Masumura K, Matsumoto M: Verification of health-based guidance values for carbendazim. Regul Toxicol Pharmacol. 2025;162:105839.
6. Yamazoe Y, Ambe K, Tohkin M, Yamada T, Masumura K. CYP1B1- and CYP1A1-Template systems and their application to metabolism and inhibition. Genes Environ. 48;1;2026.

2. 学会発表

1. 増村健一, 安東朋子, 堀端克良, 石井雄二, 杉山圭一: マウスを用いたアクリルアミドのin vivo 変異原性と次世代に誘発される生殖細胞系列突然変異の解析. 第52回日本毒性学会学術年会(2025. 07)
2. 磯貴子, 馬野高昭, 村田康允, 広瀬望, 若山美智子, 杉山圭一, 増村健一, 松本真理子: カルベンダジムの小核形成の用量反応評価からの指標値の導出. 第52回日本毒性学会学術年会(2025. 07)
3. 城島光司, 山田隆志, 藤井真人, 増村健一: ラット肝毒性予測モデルの構築と予測根拠の可視化による解釈性の向上～化学物質安全性評価への機械学習の活用を目指して～. 第52回日本毒性学会学術年会(2025. 07)
4. Yamada T, Yamamoto S, Yoshida K, Matsumoto M, Masumura K: An open-source database for inhaled industrial chemicals in PBK modeling: Towards high-accuracy parameter prediction. The American Society for Cellular and Computational Toxicology (ASCCT) 14th Annual Meeting (2025. 10)
5. Jojima K, Yamada T, Fujii M, Masumura K: Evaluating pre-trained transformer models for toxicity prediction task: Effectiveness and performance for hepatotoxicity. CBI学会2025年大会(2025. 10)
6. 増村健一, 城島光司, 川村智子, 藤井真人, 牛田和夫, 甲斐薫, 井上薫, 山田隆志: 化審法の既存化学物質情報を用いたAmesQSAR予測性の検討. 日本環境変異原ゲノム学会第54回大会(2025. 11)
7. 松本真理子, 磯貴子, 馬野高昭, 大道友香, 広瀬望, 若山美智子, 増村健一, 堀端克良, 杉山

主一：グリシジルメタクリレート経口投与による雄MutaMouse におけるin vivo 変異原性の検討. 日本環境変異原ゲノム学会第54回大会(2025.11)

H. 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

評価対象(及び対象外)とした既存化学物質の内訳(R7年度再解析版)

スクリーニング評価実施
(619物質)

評価対象物質の Ames試験結果		解析対象外の内訳				計
		高分子	無機物	構造不確定	その他	
あり	571	28	8	53	1 (Carbon black)	481 ← 今回の解析対象

図2

Ames/QSAR予測性の検討(単体モデル及び2モデルの組み合わせによる評価)

Reference		QSAR						
			Posi	Nega				
		Posi	tp	fn				
		Nega	fp	tn				
		tp/(tp+fn)	tn/(tn+fp)	(sensitivity+specificity) / 2	tp/(tp+fp)	tn/(tn+fn)	(tp+tn)/all	2*(precision*sensitivity) / (precision+sensitivity)
QSAR model		sensitivity	specificity	balanced accuracy	true positive rate(precision)	true negative rate	accuracy	f1
Derek Nexus		0.452	0.972	0.712	0.613	0.948	0.926	0.521
Sarah Nexus		0.674	0.938	0.806	0.518	0.967	0.914	0.586
GT_EXPERT		0.651	0.981	0.816	0.778	0.966	0.952	0.709
GT1-BMUT		0.651	0.970	0.810	0.683	0.965	0.941	0.667
Derek Nexus+GT_EXPERT		0.698	0.960	0.829	0.638	0.969	0.936	0.667
Derek Nexus+GT1-BMUT		0.698	0.953	0.825	0.600	0.969	0.930	0.645
Derek Nexus+Sarah Nexus		0.721	0.921	0.821	0.477	0.971	0.903	0.574
Sarah Nexus+GT_EXPERT		0.814	0.933	0.873	0.547	0.980	0.922	0.654
Sarah Nexus+GT1-BMUT		0.814	0.921	0.867	0.507	0.980	0.911	0.625
GT_EXPERT+GT1-BMUT		0.651	0.967	0.809	0.667	0.965	0.939	0.659

*太字は最高値

既知陽性/陰性として予測された物質を除いた分割表 (CASE Ultra)

Reference \ GT-EXPERT	Posi	Nega
	Posi	Nega
Posi	0	1
Nega	1	26

Out of domain およびKnown Positive とKnown Negative を除いた場合
(total: 28/481)

Reference \ GT1-BMUT	Posi	Nega
	Posi	Nega
Posi	0	1
Nega	7	86

Out of domain およびKnown Positive とKnown Negative を除いた場合
(total: 94/481)