

P050 生物学情報への機械学習解析の応用 (Toxicogenomicsへの展開)

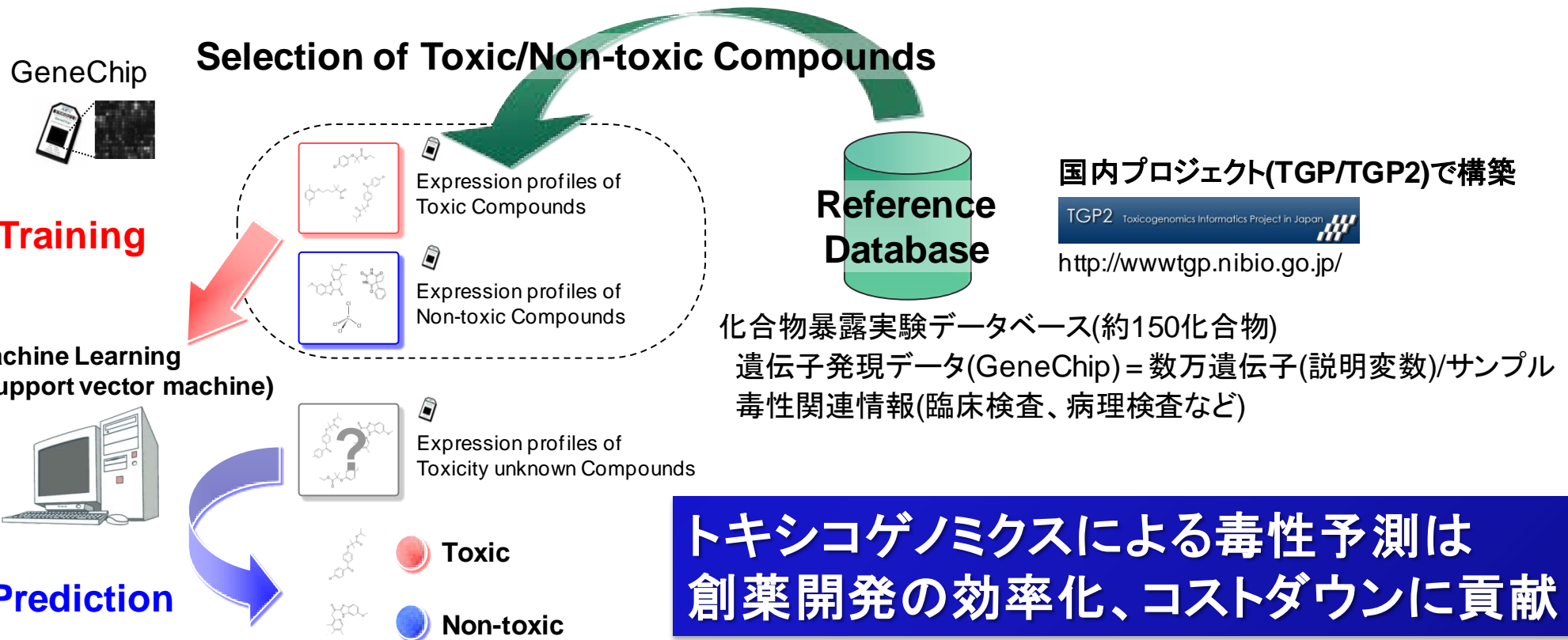
武藤裕紀*, 松下智哉, 芦原基起

*中外製薬株式会社 研究本部 創薬資源研究部

e-mail : mutohrn@chugai-pharm.co.jp

概要

機械学習をトキシコゲノミクスデータ解析へ応用し、データ前処理方法の選択による予測モデルへの影響検討および化合物の毒性予測を行った



結果

データ前処理方法として、①GeneChip数値化手法(Fig. 1)、②プローブセット選択手法の予測モデルへの影響を検討した

最適な手法の選択が予測精度の向上に重要であることが確認された

構築した予測モデルで、151化合物の毒性を予測した(Table 1)

機械学習を応用することで化合物の毒性を予測できる可能性が示唆された

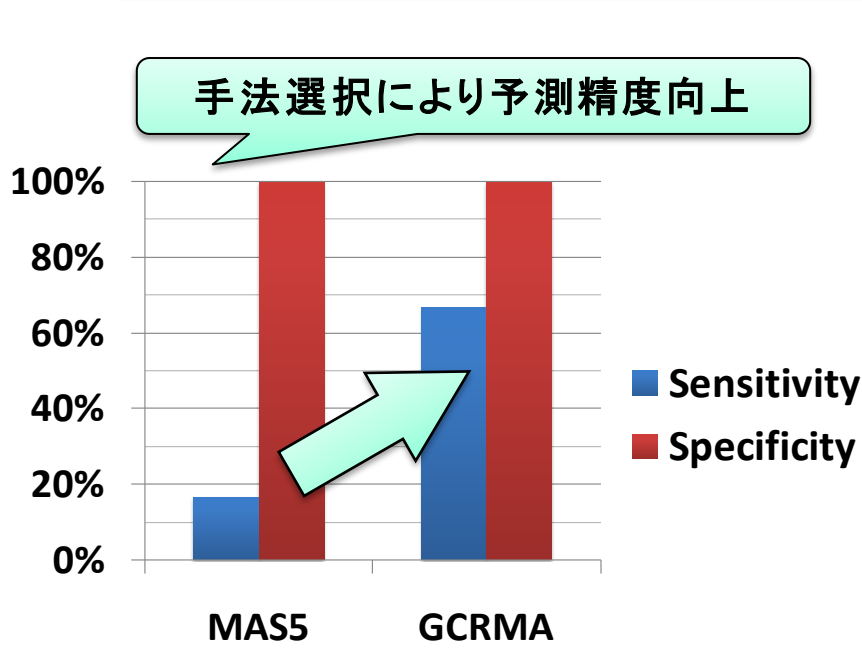


Fig. 1 GeneChip 数値化手法の影響検討

毒性有り化合物の多くを正しく分類

Table 1 毒性有りと予測された化合物

Compound // Time point & Dose	4 day			8 day			15 day			29 day			
	L	M	H	L	M	H	L	M	H	L	M	H	
acetamidofluorene						T		T	T	T	T	T	T*
allyl alcohol													T*
cisplatin													T
colchicine				T									
ethambutol						T				T			
lomustine										T			T*
lornoxicam				T			T						NA
meloxicam							T						NA
methapyrilene							T			T			T*
monocrotaline							T		T	T		T	NA
naphthyl isothiocyanate							T			T			T*
naproxen				T						T			NA
nitrosodiethylamine							T		T	T		T	NA
phalloidin	NA	NA	T	NA	NA	T	NA	NA	NA	NA	NA	NA	NA
thioacetamide				T						T		T	T*

L:Low dose, M:Middle dose, H:High dose, T:Toxic(predicted), NA:Not Available
 Red colored:Proliferation, bile duct was observed
 *Training sample