



# Predictive Toxicology using (Q)SAR: Effective Use of Various Toxicity Endpoint Models

Предсказательная токсикология с использованием (Q) SAR: эффективное использование различных моделей для разнообразных токсических эффектов

Dr Roustem D Saiakhov

MultiCASE Inc, Beachwood, OH USA

[www.multicase.com](http://www.multicase.com)

saiakhov@multicase.com

# (Q)SAR Моделирование: Что это?



- (Q)SAR = (Quantitative) Structure-Activity Relationship
  - Моделирование устанавливает связи между фрагментами химических структур и биологической активностью ( токсичностью)
  - Общее предположение: похожие молекулы имеют похожие химические и биологические
  - Биологическая активность может быть объяснена химической структурой
- Модель обучается на результатах реальных лабораторных экспериментов
  - Компьютер исследует “кусочки” химических структур чтобы найти фрагменты связанные с биологической активностью → structural alerts
  - Факторы, модулирующие активность также могут быть определены
- Модель может быть использована для предсказания токсичности нового соединения исходя только из его структуры

QSAR – количественные – модели основанные на статистике

SAR – качественные – модели основанные на экспертных правилах

} (Q)SAR

- **Модели основанные на статистических алгоритмах**
  - E.g., linear regression, naïve Bayes, partial least squares regression analysis (PLS), support vector machines (SVM), discriminant analysis, k-nearest neighbors (kNN)
  - Используют классический набор соединений для обучения
  - Быстро генерируются
  - Различаются в легкости интерпретации результатов
- **Модели основанные на экспертных правилах**
  - Создаются на основе знаний реальных экспертов
  - Часто сопровождаются информацией о химических и биохимических механизмах и ссылками на публикации
  - Легко понять результаты
  - Очень нелегко построить такие модели
- **Подходы основанные на молекулярном подобию**
  - Используют классический набор соединений для обучения
  - Быстро генерируются
  - Основаны на оценке молекулярного подобию искомого соединения и химических соединений из базы данных модели
  - Различаются в легкости интерпретации результатов

# (Q)SAR за и против



- За
  - Быстрые и менее дорогие по сравнению с лабораторными тестами
  - Тестовая субстанция не нужна, только химическая структура
  - Часто дает представление о предполагаемом механизме токсичности
  - Не требуются лабораторные животные
- Против
  - Может быть “черным ящиком”
  - Не всегда надежны
- Слабости
  - Зависит от доступности данных
  - Очень зависит от качества данных. Data curation.
  - Встречаются противоречащие данные.

# Можем ли мы использовать QSAR модели чтобы предсказать что угодно?



- Можно предсказать:
  - Небольшие и среднего размера органические соединения, встречающиеся в природе или синтезированные
- Можно предсказать но требуются специальные процедуры:
  - Полисахариды, полиэтиленгликоли итд., (используем короткие фрагменты)
  - Органометаллические соединения (требуется изменение SMILES)
- Нельзя предсказать:
  - Полимеры (~>1000 Da)
  - Неорганика
    - Простые неорганические соли
  - Органические соединения которые:
    - Смеси
    - Уникальные по своей структуре соединения
- Не различаются: Stereoisomers

# Какие QSAR модели и платформы использовать?



- Коммерческие платформы
  - CASE Ultra, DEREK/SARAH, QSAR Applier. (Все используются US FDA)
- Бесплатные платформы
  - OECD Toolbox, Toxtree, EPI Suite
- Наиболее полезные модели

- **Bacterial Mutagenicity**

- Genotoxicity

- Carcinogenicity

- Developmental toxicity

- Hepatotoxicity

- Renal Toxicity

- Cardiotoxicity

- **Skin** and Eye toxicity

- Ecotoxicity

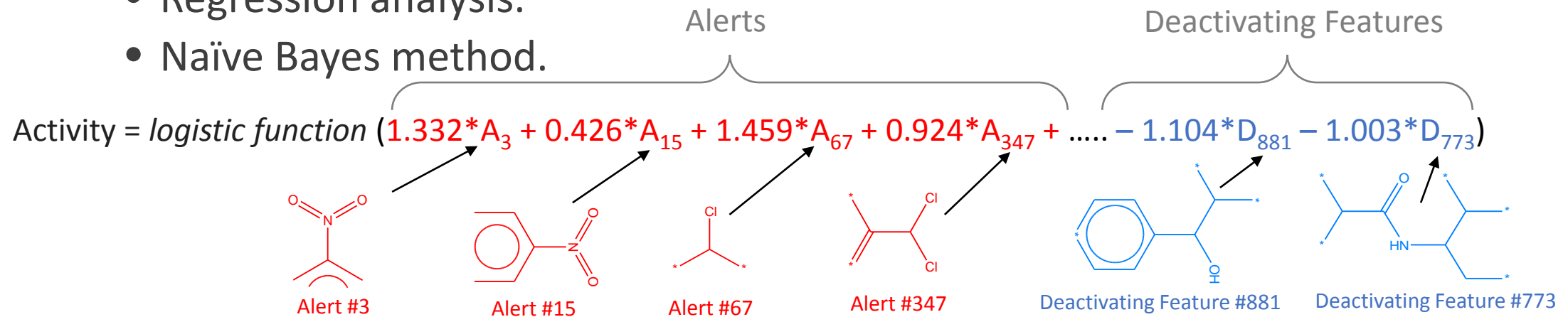
- Acute Toxicity

- Inhalation toxicity

- ADME

# Что такое QSAR модель?

- Обычно модели это сочетание набора структурных атрибутов биологической активности (alerts) и база данных поддержки.
- Обычно вычисляются коэффициенты влияния каждого структурного атрибута (structural alerts).
- Могут использоваться разнообразные статистические алгоритмы:
  - Regression analysis.
  - Naïve Bayes method.



# Заключение эксперта чрезвычайно важно для надежного предсказания токсичности



**Результаты полученные от модели** “... могут быть рассмотрены с использованием экспертов чтобы получить дополнительные доказательства актуальности любого положительного, отрицательного, противоречивого или неубедительного результата и предоставить обоснование для поддержки окончательного заключения.”

Грамотно выполненная экспертиза может помочь в следующих ситуациях:

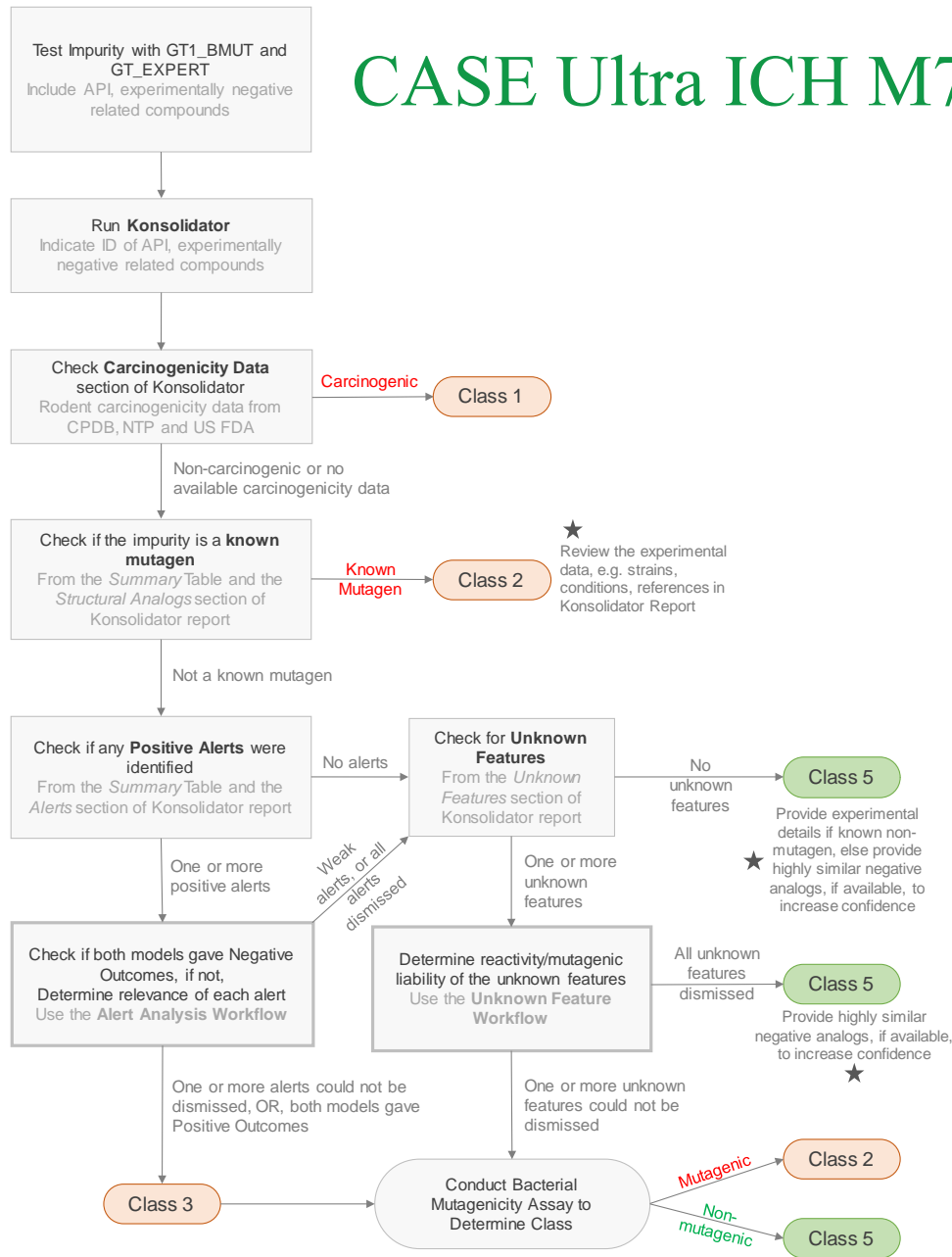
- Противоречивые или неопределенные предсказания.
- Определяемая структура за пределами химического пространства модели.
- Для определения достоверности идентифицированных атрибутов токсичности.
- Механистическая интерпретация атрибутов токсичности.
- Чтобы включить данные от аналогов не включенных в базу поддержки модели.
- Чтобы повысить уверенность в предсказаниях токсичности.



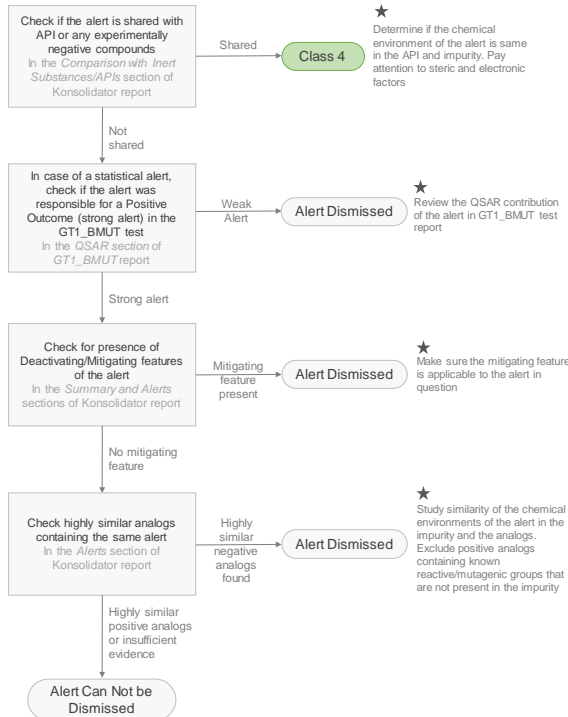
# CASE Ultra ICH M7 процесс предсказания



Выглядит сложно? Используйте Konsolidator Tool



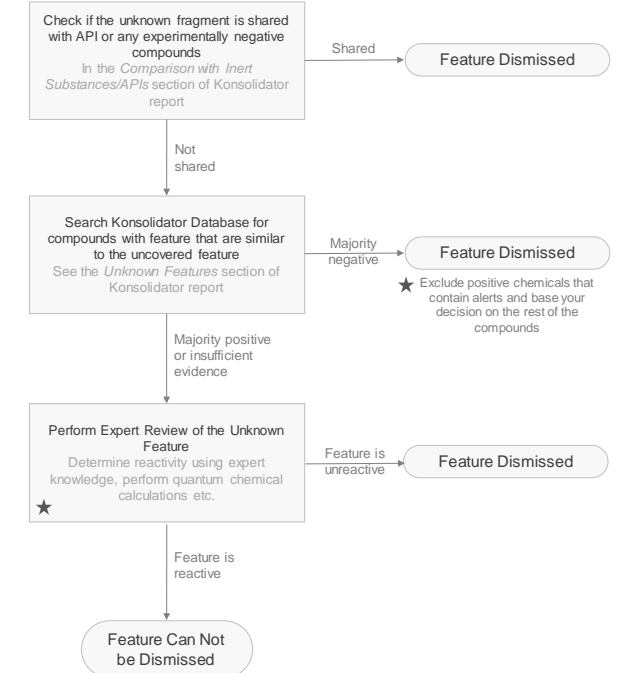
## Alert Analysis Workflow



★ Consideration for Expert Analysis

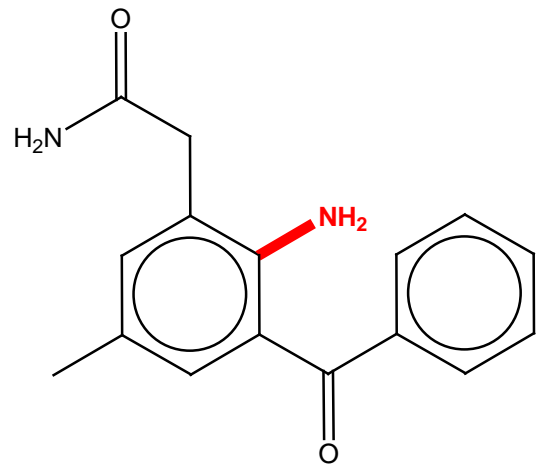
Copyright © 2018 MultiCASE Inc. All rights reserved.

## Unknown/Uncovered Feature Analysis Workflow

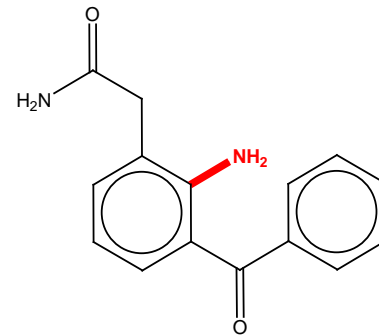


★ Consideration for Expert Analysis

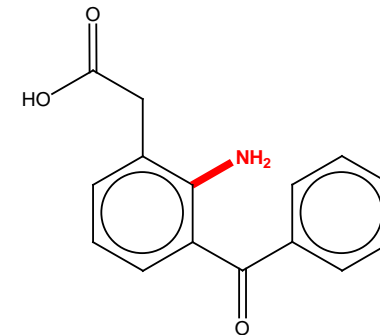
# Использование аналогов для подтверждения результатов в CASE Ultra Konsolidator



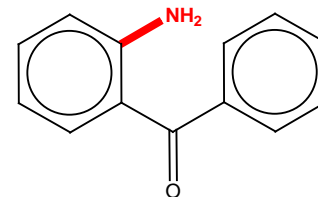
Тестируемая структура



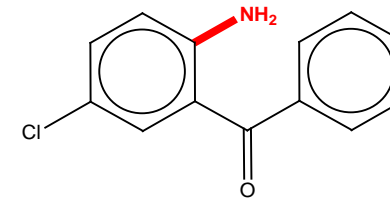
AES = 0.933  
Negative



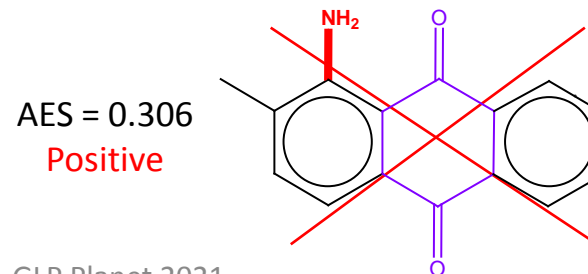
AES = 0.673  
Negative



AES = 0.363  
Negative



AES = 0.322  
Negative



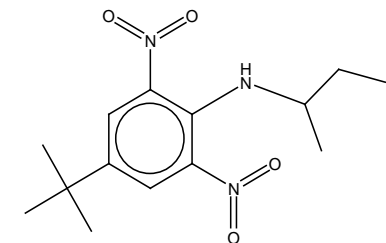
AES = 0.306  
Positive

**AES: Alert Environment Similarity** Подobie структурного окружения

# Пример #1: Противоречивые результаты

	<b>Модель</b>	<b>Предсказание</b>	
	Экспертная	Negative	
	Статистическая	Positive	
	<b>Заключение эксперта</b>	<b>Negative</b>	<b>▶ Class 5</b>

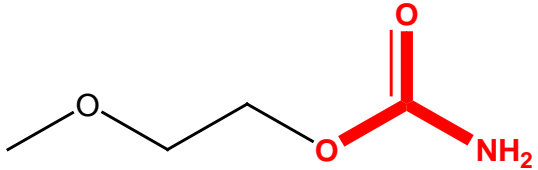
- Было обнаружено несколько похожих неактивных аналогов, один очень близок к тестируемой структуре.
- Было обнаружено экспертное правило объясняющее снижение активности (Nitro-aromatics, ortho-groups with steric or electronic deactivation)



AES = 0.727

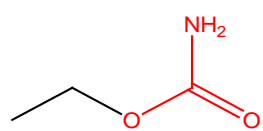
Negative

# Пример #2: Противоречивые результаты

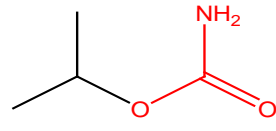
	Модель	Предсказание
	Экспертная	Negative
	Статистическая	Positive
	<b>Заключение эксперта</b>	<b>Positive</b>

► **Class 3**

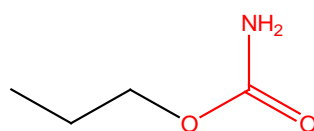
- O-Carbamate фрагмент был идентифицирован обеими моделями, и было обнаружено несколько активных аналогов.



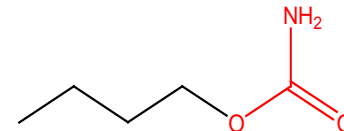
AES = 0.600  
Positive



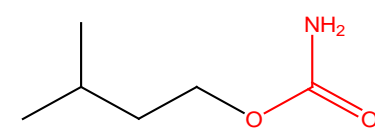
AES = 0.577  
Positive



AES = 0.500  
Positive



AES = 0.429  
Positive



AES = 0.429  
Positive

# Прогнозирование сенсibilизации кожи у людей

Model Name	Version	System Level	Modeled Event	Model Description	Training Set Size
LLNA_CAT2_STRONG	1.8.0.2	ORGAN	Activation of Immune Response	Strong sensitizers (LLNA EC3<1%)	383
LLNA_CAT3_MODERATE				Moderate and strong sensitizers (LLNA EC3<10%)	792
LLNA_CAT4_WEAK				Weak, Moderate and Strong sensitizers (LLNA EC3<100%)	891

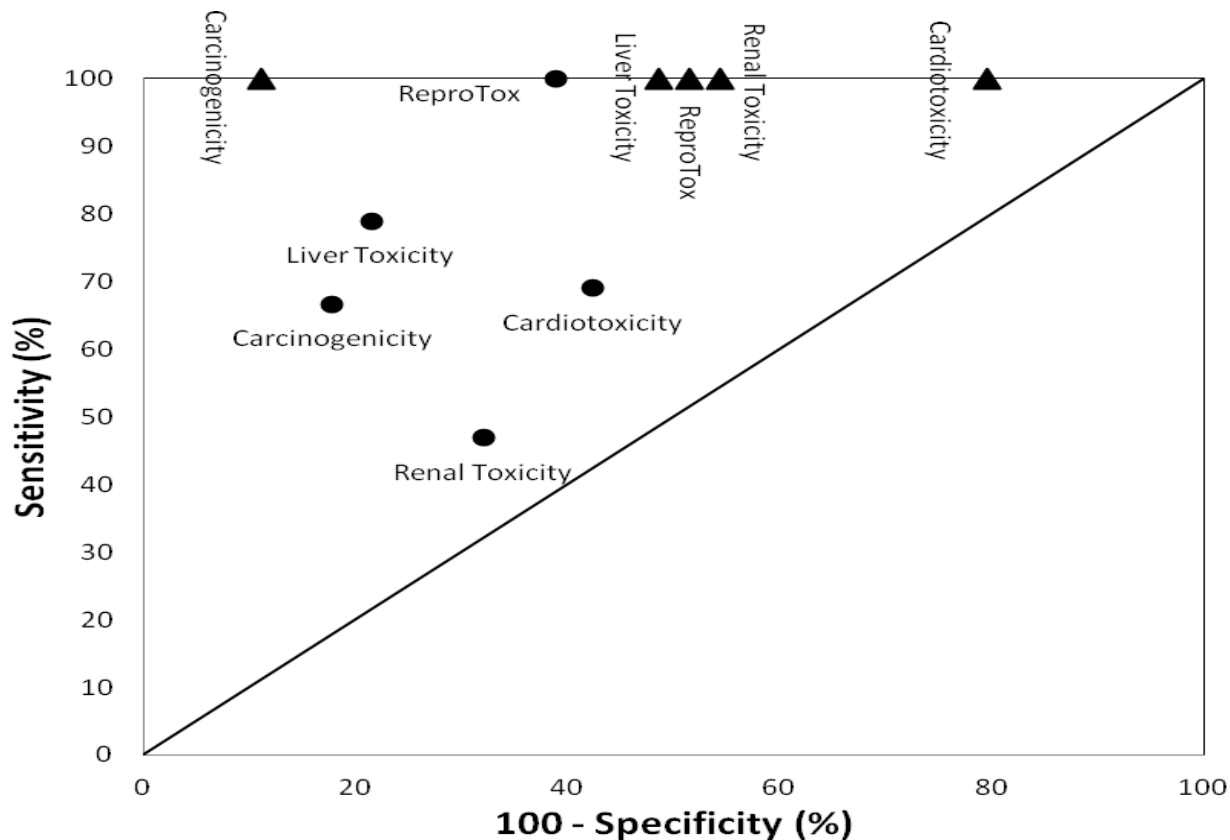
- Связано с недавно принятым OECD 497 Guideline on Defined Approaches for Skin Sensitization
- Обучающий набор получен из MultiCASE базы данных с 1821 соединениями с имеющимися ECETOC категориями и EC3 значениями
- Тестовый набор сенсibilизаторов человека состоял из 19 соединений, 14 активных and 5 неактивных.
- Coverage 94.74%; Sensitivity 92.31%; Specificity 80.00%; Concordance 88.89%  
Pos accuracy 92.31%; Neg accuracy 80.00%

# Решение сложных проблем: побочные эффекты у людей



- Целью данного пилотного исследования была проверка возможностей основанной на фрагментах химических структур системы QSAR CASE Ultra для прогнозирования побочных эффектов лекарственных препаратов.
- Был протестирован набор из 870 лекарств из набора данных о побочных эффектах SIDER, и их профиль побочного действия был предсказан с точки зрения использования CASE Ultra в отношении канцерогенности, генетической токсичности, токсичности для печени, кардиотоксичности, токсичности для почек и репродуктивной токсичности.
- Набор из 47 лекарств, которые были отозваны или сняты с рынка с 1950-х годов, был составлен и подвергал их токсичности, также был оценен на предмет потенциальных рисков с использованием CASE Ultra. Цель состояла в том, чтобы предсказать потенциальные риски и сравнить их с фактическими причинами отзыва лекарств.

# Решение сложных проблем: побочные эффекты у людей



Относительное расположение результатов прогнозирования различных побочных эффектов для 870 тестируемых химических веществ из базы данных SIDER на графике частоты ложных положительных результатов (100 - specificity) и истинных положительных результатов (sensitivity). Черные кружки представляют результаты для всех протестированных соединений, черные треугольники представляют результаты только для тех лекарств, которые не были использованы при разработке QSAR моделей. Совершенно случайное предположение дало бы точку на диагональной линии.

Для коллекции веществ из базы данных SIDER чувствительность и специфичность прогнозов канцерогенности составляют 66,67% и 82,17% соответственно; на печеночную токсичность - 78,95%, 78,50%; кардиотоксичность - 69,07%, 57,57%; почечная токсичность - 46,88%, 67,90%; репродуктивная токсичность - 100,00%, 61,10%

## US FDA список изъятых лекарств

- CASE Ultra правильно распознала соответствующие токсические эффекты у 43 из 47 отмененных препаратов.
- Для 9 препаратов, которые не входят в обучающие наборы большинства моделей (alpidem, buformin, lumiracoxib, methaqualone, rimonabant, tegaserod, terodiline, ximelagatran and zimeldine), CASE Ultra смогла правильно идентифицировать все (100%) как небезопасные.
- Из этих 9 препаратов 3 были сняты с продажи из-за побочных эффектов, которые не рассматривались в данном исследовании
- Для пяти из оставшихся 6 (83,3%) (alpidem, lumiracoxib, tegaserod, terodiline and ximelagatran) были предсказаны те же побочные эффекты, из-за которых они были сняты с рынка.

Effectiveness of CASE Ultra Expert System in Evaluating Adverse Effects of Drugs; Saiakhov, R.D., Chakravarti, S.K. and Klopman, G.; Molecular Informatics, 2012, 32, 87-97. DOI: 10.1002/minf.201200081



# Заключение



Правильно разработанные и поддерживаемые в актуальном состоянии платформы in-silico токсикологии являются мощными, надежными и подходящими инструментами для подготовки нормативных документов, а также для исследовательской деятельности.

To get the answers to your questions please contact us at:

+1-216-831-3740

[saiakhov@multicase.com](mailto:saiakhov@multicase.com) or [info@multicase.com](mailto:info@multicase.com)

23811 Chagrin Blvd Ste 305, Beachwood, Ohio, 44122, USA

# Acknowledgments

- This presentation would be impossible without great help of Dr Suman Chakravarti (VP and CSO), Gianna Cioffi (Customer care), Mounika Girireddy and Radha Alla (software and database developers).
- Several slides from Dr Chakravarti's and Dr Kruhlak's presentations were used in this presentation directly or as an inspiration.