

ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича»
Лаборатория структурно-функционального конструирования лекарств

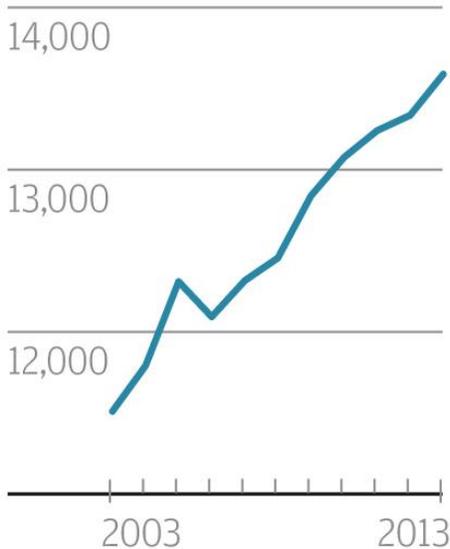
КОНОВА В.И.

**ВЕБ-СЕРВИС ДЛЯ ПРЕДСКАЗАНИЯ
ЦИТОТОКСИЧНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ
В ОТНОШЕНИИ ОПУХОЛЕВЫХ И НЕОПУХОЛЕВЫХ
КЛЕТОЧНЫХ ЛИНИЙ ЧЕЛОВЕКА**

Симпозиум «Биоинформатика и компьютерное конструирование лекарств»
Конгресс «Человек и лекарство»
Москва, 2016 г.

Rising menace

Total cancer deaths



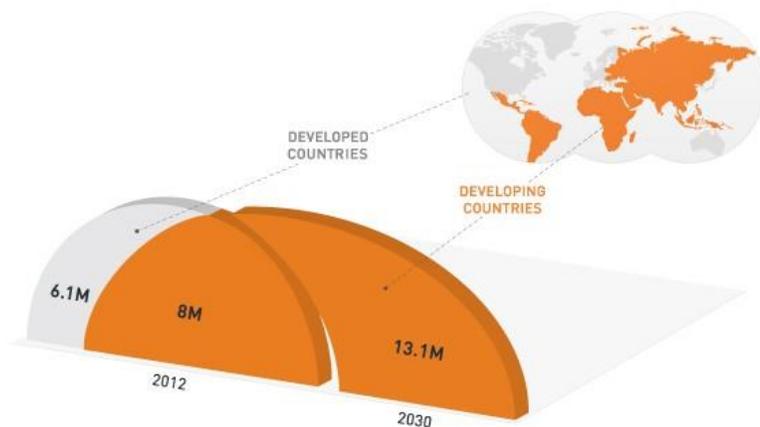
Sources: Department of Health, Hong Kong

Top 10 cancers for both sexes in 2012

| | Deaths | % |
|------------------------|--------|------|
| 1 Lung | 3,893 | 29.2 |
| 2 Colorectum | 1,903 | 14.3 |
| 3 Liver | 1,505 | 11.3 |
| 4 Stomach | 657 | 4.9 |
| 5 Breast | 604 | 4.5 |
| 6 Pancreas | 538 | 4 |
| 7 Prostate | 362 | 2.7 |
| 8 Non-Hodgkin lymphoma | 351 | 2.6 |

BURDEN ON THE DEVELOPING WORLD

NEW CANCER CASES



CANCER

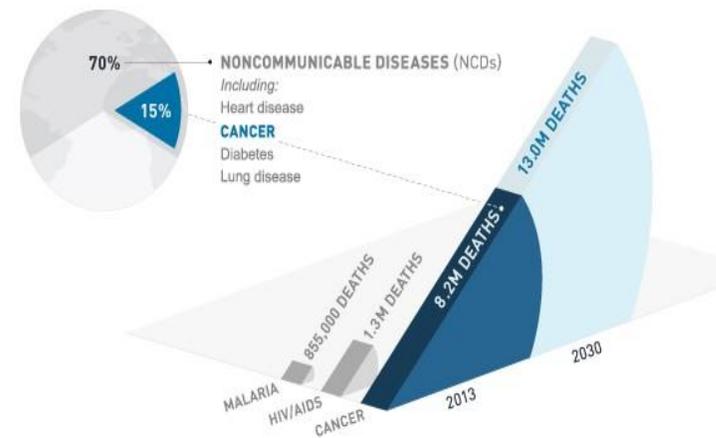


Leading types and their mortality rate each year

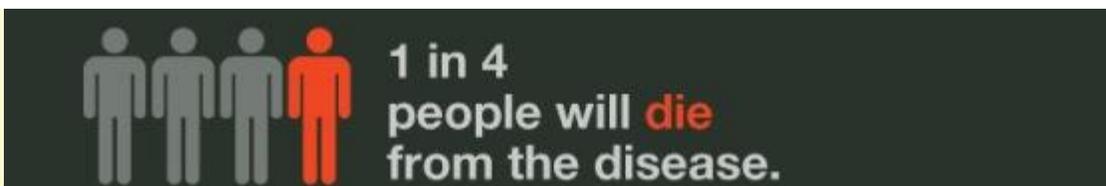
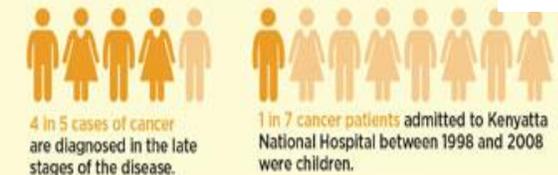
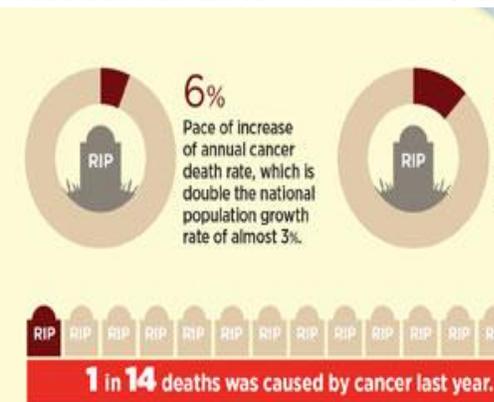
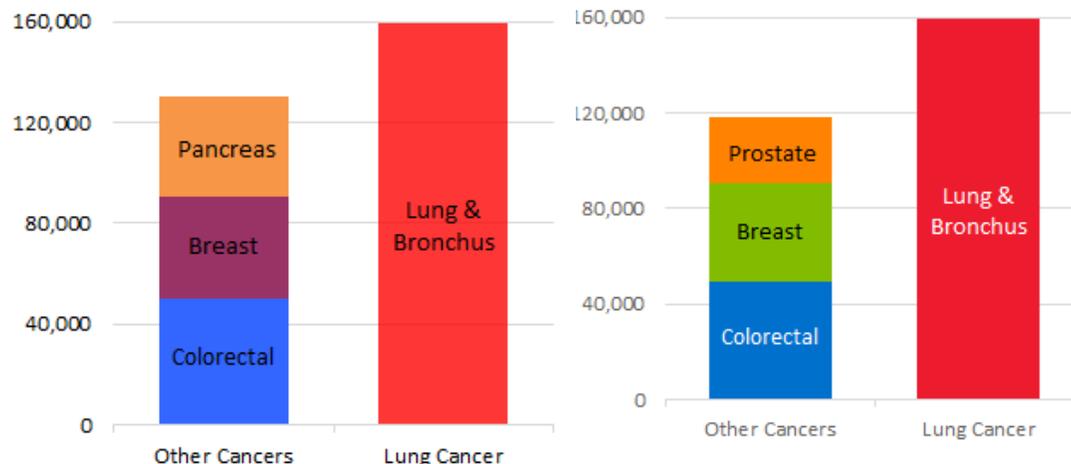


THE GLOBAL BURDEN

ANNUAL DEATHS



Estimated Cancer Deaths by Site, 2014² Estimated Cancer Deaths by Site, 2015²

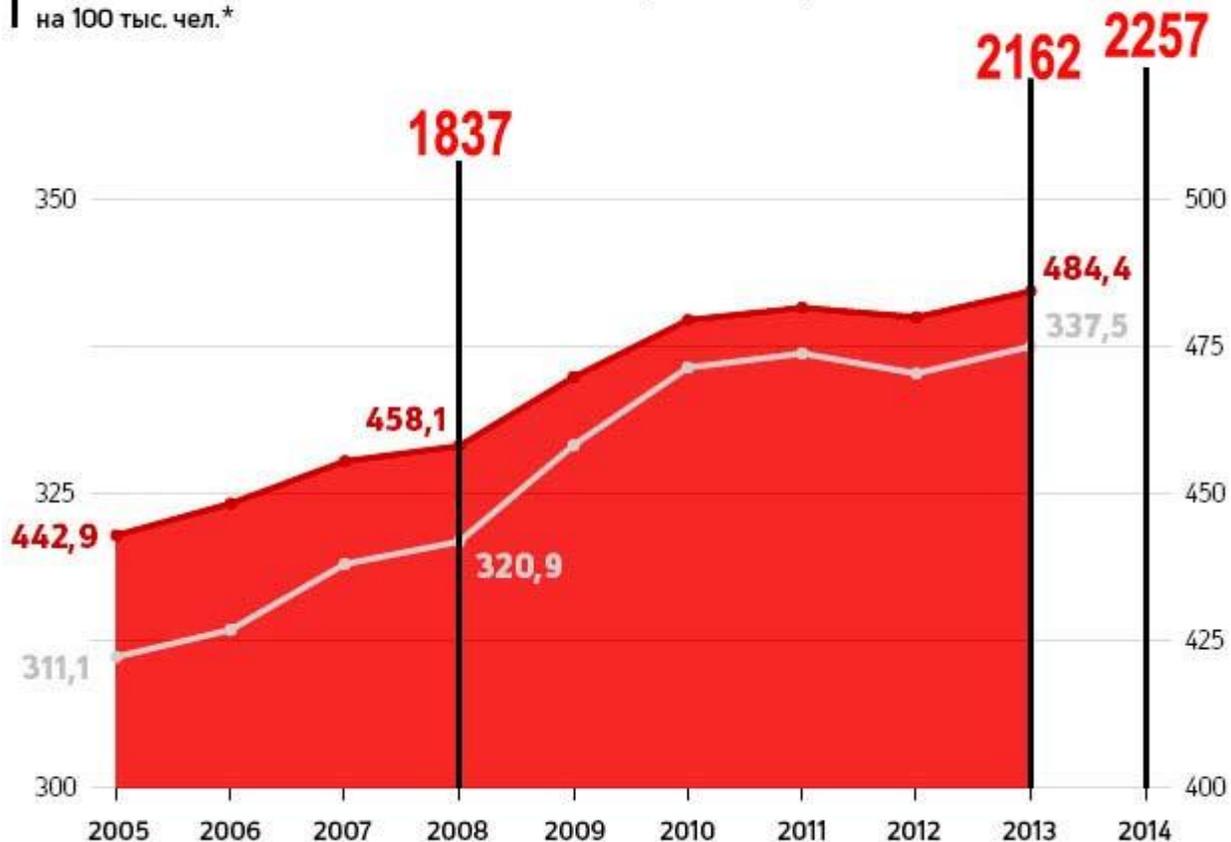


American Cancer Society, 2015
Siegel et al., Cancer statistics, 2015

Статистика заболеваемости раком в России

Взято на учет больных с впервые в жизни установленным диагнозом
 — на 100 тыс. чел. (левая шкала) ■ всего, тыс. чел. (правая шкала)

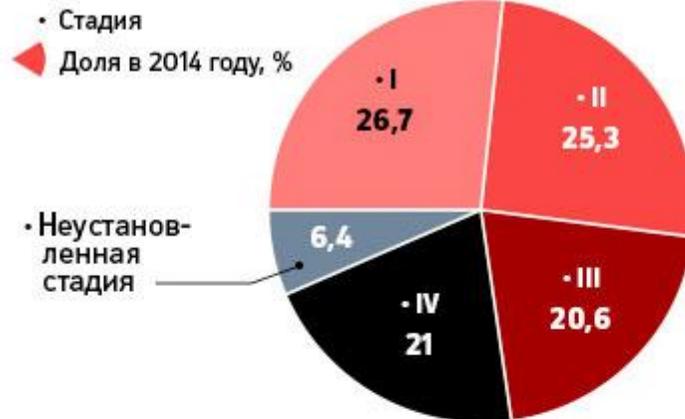
Общая заболеваемость злокачественными новообразованиями, на 100 тыс. чел.*



*Без учета Крыма и Севастополя

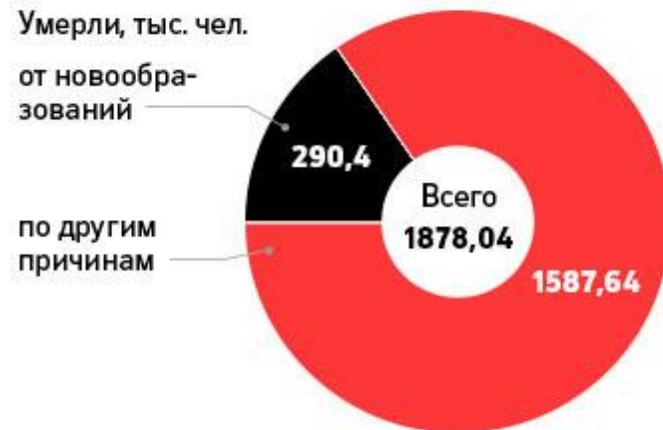
Источники: Минздрав, Росстат

Когда обнаруживают рак

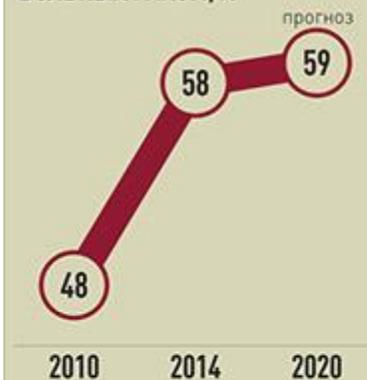


Сколько человек умерли от новообразований в 2014 году

По данным Росстата, новообразования занимали второе место среди причин смертности в России*

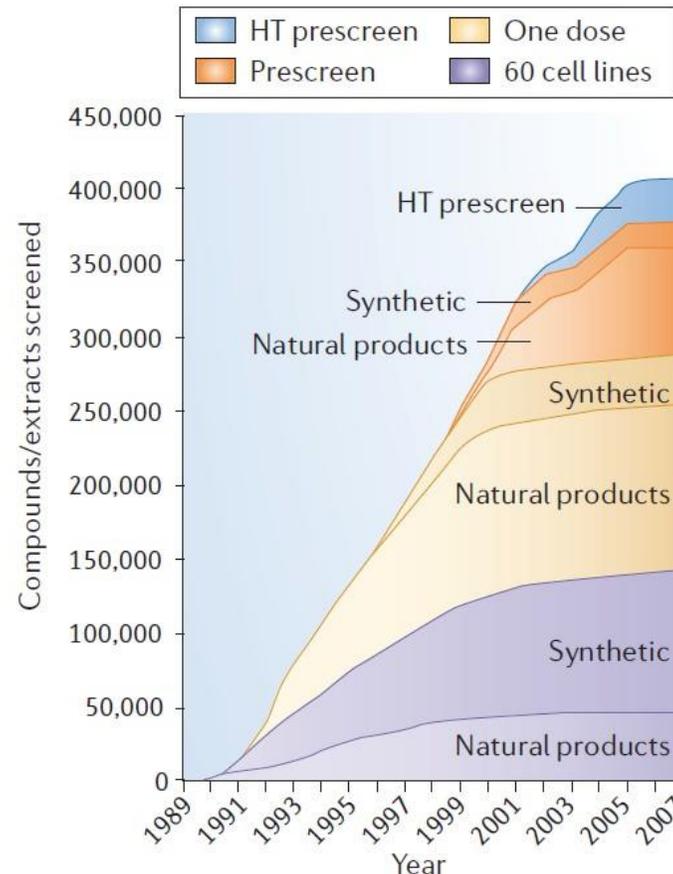
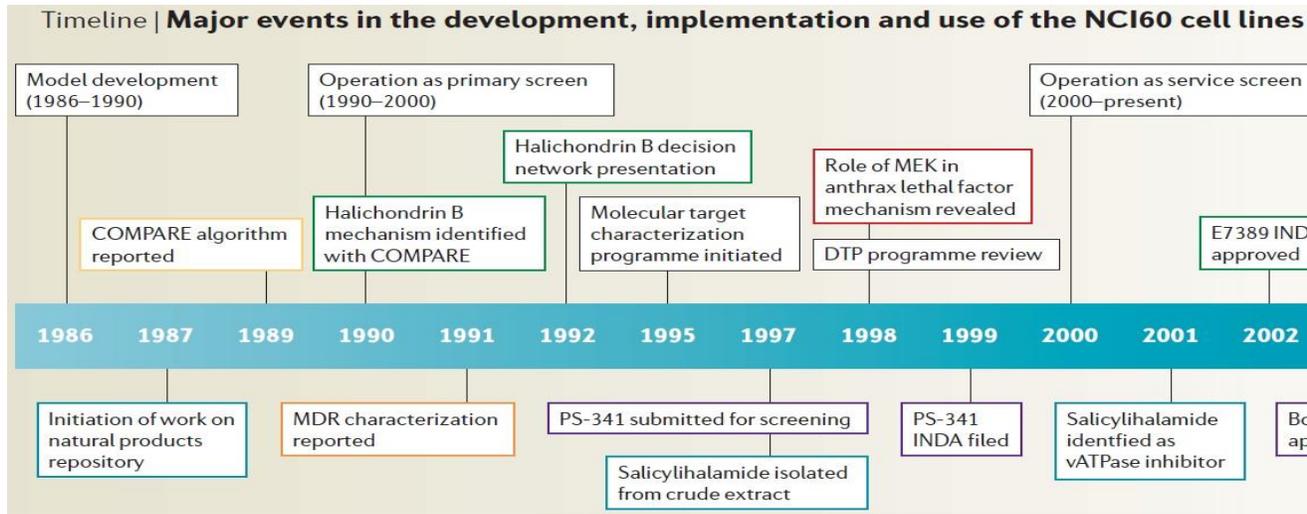
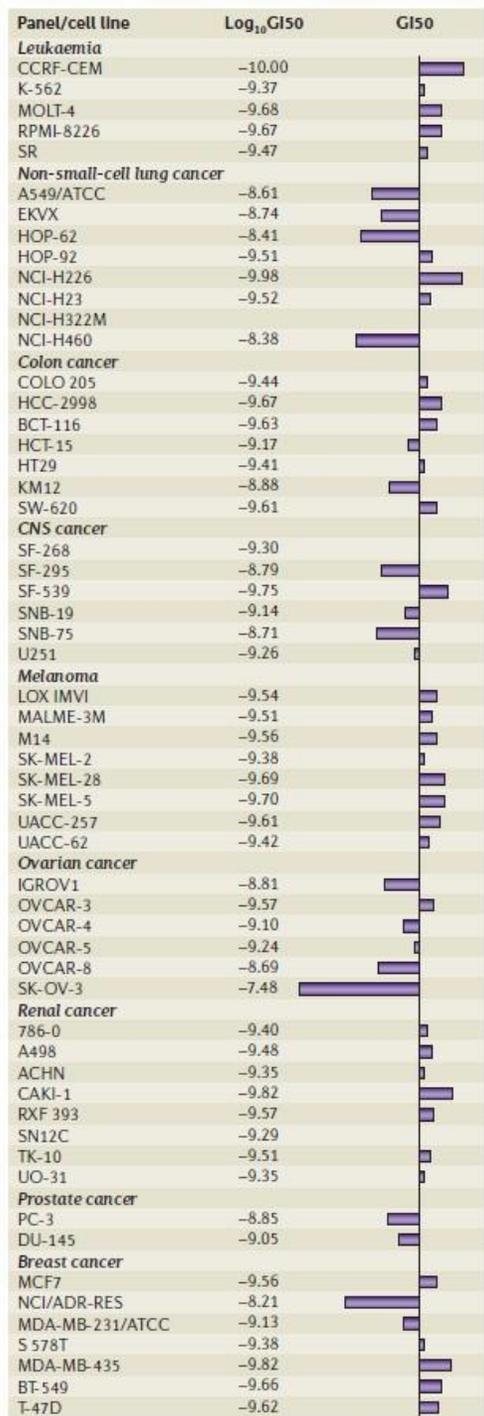


ПЯТИЛЕТНЯЯ ВЫЖИВАЕМОСТЬ БОЛЬНЫХ РАКОМ, %

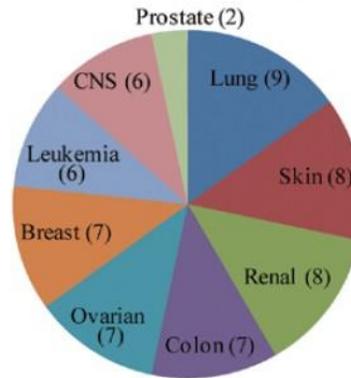


ВЫЯВЛЕНИЕ РАКА НА РАННИХ СТАДИЯХ

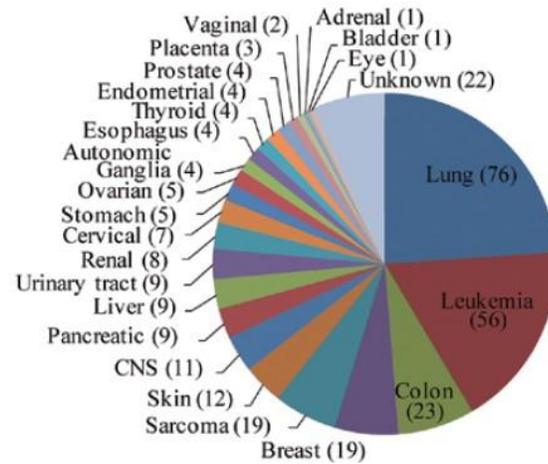




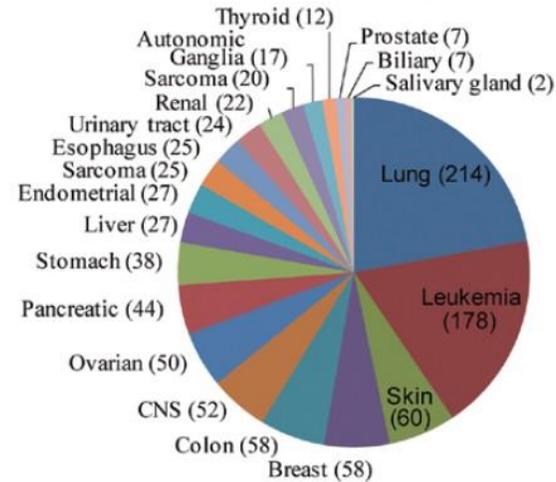
A
NCI60 cancer cell line panel



B
GSK cancer cell line panel



C
CCLE cancer cell line panel



NCI, National Cancer Institute – 60 cell lines
GSK, Glaxo Smith Kline – 318 cell lines
CCLE, Cancer Cell Line Encyclopedia – 967 cell lines

Cell Line Cytotoxicity Predictor (CLC-Pred)



Целью работы является создание веб-сервиса для компьютерного предсказания цитотоксичности химических соединений на основе их структурной формулы в отношении опухолевых и неопухолевых клеточных линий человека.

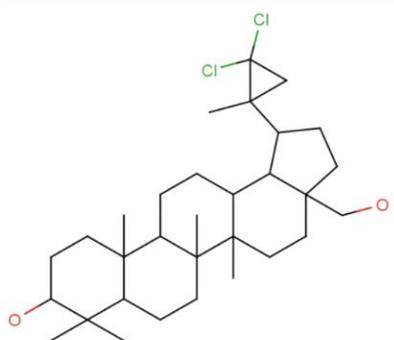
С использованием информации из базы данных **ChEMBLdb 19** о действии химических соединений на клеточные линии человека были созданы обучающие выборки для компьютерной программы **PASS** (Prediction of Activity Spectra for Substances) и построены предсказательные модели взаимосвязей «структура-цитотоксичность».



92% - средняя точность прогноза,
посчитанная по методу скользящего контроля с
исключением по одному, для цитотоксичности ХС.

- ✓ **Дополнительная оценка точности прогноза - тестовая выборка Genomics of Drug Sensitivity in Cancer project:**
135 соединений были протестированы на 714 опухолевых клеточных линиях.
Оценка: при $P_a > P_i$ IAP составляет 94,4%,
чувствительность 95,7%,
специфичность 71,3%,
точность 75,6%.

Оценка вероятности
наличия конкретного вида
биологической активности -
цитотоксичности



Pa>Pi ▾

Save *.sdf

Save *.csv

Cancer cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue | Tumor type |
|-------|-------|---------------------------|----------------------------------|---------|--------------|
| 0.875 | 0.003 | MSTO-211H | Pleural mesothelioma cells | Pleura | Mesothelioma |
| 0.672 | 0.020 | MCF7 | Breast carcinoma cells | Breast | Carcinoma |
| 0.601 | 0.014 | TE-671 | Human Rhabdomyosarcoma cell line | Muscle | Sarcoma |
| 0.525 | 0.017 | SK-OV-3 | Ovarian carcinoma cells | Ovarium | Carcinoma |
| 0.365 | 0.030 | T47D | Breast carcinoma cells | Breast | Carcinoma |

Non-tumor cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue |
|-------|-------|-----------------------------|---------------------------------|--------|
| 0.122 | 0.025 | Detroit 551 | Embryonic skin | Skin |
| 0.210 | 0.186 | IMR-90 | Embryonic lung fibroblast cells | Lung |

Предсказанный спектр биологической активности, где
Pa – вероятность наличия цитотоксичности к определенной клеточной линии,
Pi – вероятность отсутствия цитотоксичности

[» Home](#)[» Training Set](#)[» Products/Services](#)[» Interpretation](#)[» Contacts](#)[CLC data tumor](#)[CLC data non-tumor](#)

Traning set

Training sets on cytotoxicity of chemical compounds built on experimental data from ChEMBLdb were used to train PASS for "structure-cell line cytotoxicity" relationship prediction. The average prediction accuracy calculated by leave-one-out cross-validation procedure is approximately 92% for cytotoxicity prediction for cancer cell lines and non-tumor cell lines.

The outcome of the collection of experimental data was the training set of 53,774 structures of compounds, which reflects current knowledge about the cytotoxic substances in relation to 1,781 human cell lines

» Home | » Training Set | » Products/Services | » Interpretation | » Contacts

[CLC data tumor](#)

[CLC data non-tumor](#)

| Number of substance | Accuracy of the prediction* | Activity name |
|-------------------------------------|---|-------------------------------|
| 7182 | 0.87 | MCF7 |
| 6248 | 0.88 | HCT-116 |
| 5752 | 0.86 | A549 |
| 4193 | 0.89 | HT-29 |
| 3502 | 0.91 | HeLa |
| 3224 | 0.89 | K562 |
| 3222 | 0.89 | PC-3 |
| 2872 | 0.90 | HL-60 |
| 2706 | 0.89 | MDA-MB-231 |
| 2455 | 0.89 | DU-145 |
| 2360 | 0.91 | CCRF-CEM |
| 2299 | 0.90 | NCI-H460 |
| 1892 | 0.94 | A2780 |
| 1836 | 0.92 | HepG2 |
| 1727 | 0.94 | KB |
| 1467 | 0.90 | SK-OV-3 |
| 1217 | 0.88 | MDA-MB-435 |

* Leave-one-out cross-validation (LOO CV) procedure is performed using the whole PASS training set for validation of prediction quality. The prediction result is compared with known experimental data for the studied compound. The procedure is repeated for all compounds from the PASS training set; then the average Invariant Accuracy of Prediction (IAP=1-IEP) values are calculated for each biological activity and for all biological activities. IAP equals numerically to ROC AUC

» Home | » Training Set | » Products/Services | » Interpretation | » Contacts

[CLC data tumor](#)

[CLC data non-tumor](#)

| Number of substance | Accuracy of the prediction* | Activity name |
|-------------------------------------|---|-------------------------------|
| 565 | 0.95 | HUVEC |
| 363 | 0.92 | MRC5 |
| 250 | 0.92 | HEK293 |
| 208 | 0.97 | HaCaT |
| 195 | 0.95 | PBMC |
| 116 | 0.93 | WI-38 |
| 87 | 0.97 | HMEC |
| 62 | 0.99 | HFF |
| 61 | 0.96 | NHDF |
| 52 | 0.98 | NFF |
| 46 | 0.96 | WIL2-NS |
| 38 | 0.98 | HS27 |
| 37 | 0.97 | RPTEC |
| 37 | 0.96 | WRL68 |
| 29 | 0.86 | BJ |
| 29 | 0.96 | Detroit 551 |
| 25 | 1.00 | HTA975 |

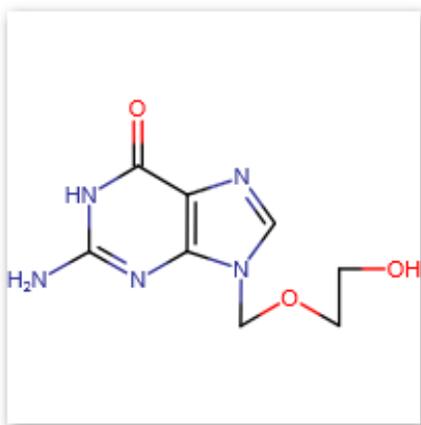
* Leave-one-out cross-validation (LOO CV) procedure is performed using the whole PASS training set for validation of prediction quality. The prediction result is compared with known experimental data for the studied compound. The procedure is repeated for all compounds from the PASS training set; then the average Invariant Accuracy of Prediction (IAP=1-IEP) values are calculated for each biological activity and for all biological activities. IAP equals numerically to ROC AUC

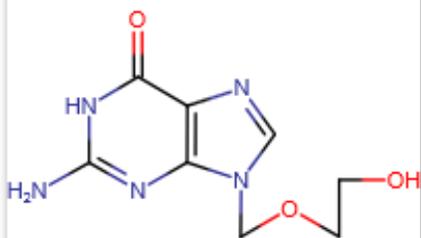
[» Home](#)[» Training Set](#)[» Products/Services](#)[» Interpretation](#)[» Contacts](#)[SMILES](#)[Use Files](#)[Marvin Molecular Editor](#)

CLC-Pred (Cell Line Cytotoxicity Predictor) is a web-service for *in silico* prediction of cytotoxic effect of chemical compounds in non-transformed and cancer cell lines based on structural formula. CLC-Pred provides a prediction of the cytotoxicity of a chemical compound to assess the relevance of the substance's inclusion in experimental screening.

Prediction is based on PASS (Prediction of Activity Spectra for Substances) technology (<http://www.way2drug.com/PASSonline>) and the training set created on the basis of data on cytotoxicity retrieved from ChEMBLdb (version 19) (<https://www.ebi.ac.uk/chembl/db/>).

Use SMILES, MOL file or Marvin Javascript applet for input data on a structure of test compound.

[Make prediction](#)

» [Home](#)» [Training Set](#)» [Products/Services](#)» [Interpretation](#)» [Contacts](#)[SMILES](#)[Use Files](#)[Marvin Molecular Editor](#)[Make prediction](#)CLC
effect
CLC
of th
Pred
(http
cyto
UsePa>Pi [Save *.sdf](#)[Save *.csv](#)

Cancer cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue | Tumor type |
|-------|-------|--------------------------|--|--------|----------------|
| 0.428 | 0.024 | CCRF-CEM | Childhood T acute lymphoblastic leukemia cells | Blood | Leukemia |
| 0.417 | 0.028 | IGROV-1 | Ovarian adenocarcinoma cells | Ovary | Adenocarcinoma |
| 0.409 | 0.044 | Hs-578T | Invasive ductal breast carcinoma cells | Breast | Carcinoma |
| 0.400 | 0.044 | MOLT-4 | Acute T-lymphoblastic leukemia cells | Blood | Leukemia |
| ... | ... | ... | Ovarian | ... | ... |

Non-tumor cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue |
|-------|-------|------------------------|--|----------|
| 0.243 | 0.142 | BJ | Foreskin fibroblast cells | Foreskin |
| 0.212 | 0.179 | IMR-90 | Embryonic lung fibroblast cells | Lung |
| 0.049 | 0.044 | RPTEC | Renal proximal tubule epithelial cells | Kidney |

toxic
mula.
ance
ology
a on
nd.

ChEMBL

- ChEMBL
- Downloads
- UniChem
- SureChEMBL
- Malaria Data
- ChEMBL-NTD
- ADME SARfari
- Web Services
- myChEMBL
- EBI RDF Platform
- FAQ
- Web status page New

EBI > Databases > Small Molecules > ChEMBL Database

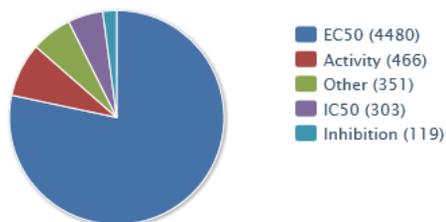
Target Report Card

Target Name and Classification

| | |
|-------------------------------|----------------|
| Target ID | CHEMBL614358 |
| Target Type | CELL-LINE |
| Preferred Name | BJ |
| Synonyms | |
| Organism | Homo sapiens |
| Species Group | No |
| Protein Target Classification | Not applicable |

Target Associated Bioactivities

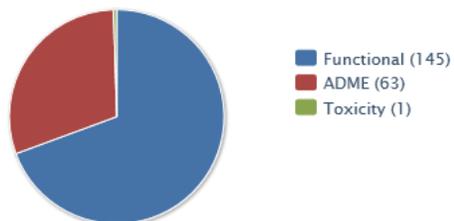
ChEMBL Activity Types for Target CHEMBL614358



Total: 5719

Target Associated Assays

ChEMBL Assays for Target CHEMBL614358



Total: 209

» Products/Services

» Interpretation

» Contacts

Pa>Pi

Save *.sdf

Save *.csv

Cancer cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue | Tumor type |
|-------|-------|--------------------------|--|--------|----------------|
| 0.428 | 0.024 | CCRF-CEM | Childhood T acute lymphoblastic leukemia cells | Blood | Leukemia |
| 0.417 | 0.028 | IGROV-1 | Ovarian adenocarcinoma cells | Ovary | Adenocarcinoma |
| 0.409 | 0.044 | Hs-578T | Invasive ductal breast carcinoma cells | Breast | Carcinoma |
| 0.400 | 0.044 | MOLT-4 | Acute T-lymphoblastic leukemia cells | Blood | Leukemia |
| 0.399 | 0.044 | MCF-7 | Ovarian | Breast | Carcinoma |

Non-tumor cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue |
|-------|-------|------------------------|--|----------|
| 0.243 | 0.142 | BJ | Foreskin fibroblast cells | Foreskin |
| 0.212 | 0.179 | IMR-90 | Embryonic lung fibroblast cells | Lung |
| 0.049 | 0.044 | RPTEC | Renal proximal tubule epithelial cells | Kidney |

toxic
mula.
anceology
a on

nd.

ChEMBL

- ChEMBL
- Downloads
- UniChem
- SureChEMBL
- Malaria Data
- ChEMBL-NTD
- ADME SARfari
- Web Services
- myChEMBL
- EBI RDF Platform
- FAQ
- Web status page New

EBI > Databases > Small Molecules > ChEMBL Database

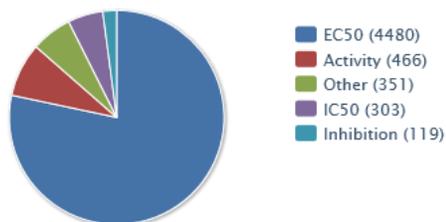
Target Report Card

Target Name and Classification

| | |
|-------------------------------|----------------|
| Target ID | CHEMBL614358 |
| Target Type | CELL-LINE |
| Preferred Name | BJ |
| Synonyms | |
| Organism | Homo sapiens |
| Species Group | No |
| Protein Target Classification | Not applicable |

Target Associated Bioactivities

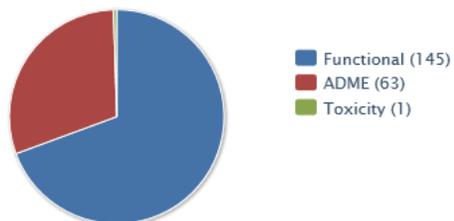
ChEMBL Activity Types for Target CHEMBL614358



Total: 5719

Target Associated Assays

ChEMBL Assays for Target CHEMBL614358



Total: 209

» Products/Service

Pa>Pi

✓ Результаты предсказания сохраняются в файле формата **SDF**, содержащем структуру химического соединения и прогнозируемых для него видах биологической активности в отношении различных клеточных линий человека с оценками вероятности проявления активности.

Cancer cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue | Tumor type |
|-------|-------|--------------------------|--|--------|----------------|
| 0.428 | 0.024 | CCRF-CEM | Childhood T acute lymphoblastic leukemia cells | Blood | Leukemia |
| 0.417 | 0.028 | IGROV-1 | Ovarian adenocarcinoma cells | Ovary | Adenocarcinoma |
| 0.409 | 0.044 | Hs-578T | Invasive ductal breast carcinoma cells | Breast | Carcinoma |
| 0.400 | 0.044 | MOLT-4 | Acute T-lymphoblastic leukemia cells | Blood | Leukemia |
| 0.399 | 0.044 | MCF-7 | Ovarian | | |

Non-tumor cell line prediction result

| Pa | Pi | Cell-line | Cell-line full name | Tissue |
|-------|-------|------------------------|--|----------|
| 0.243 | 0.142 | BJ | Foreskin fibroblast cells | Foreskin |
| 0.212 | 0.179 | IMR-90 | Embryonic lung fibroblast cells | Lung |
| 0.049 | 0.044 | RPTEC | Renal proximal tubule epithelial cells | Kidney |

SAR QSAR Environ Res. 2015;26(7-9):595-604. doi: 10.1080/1062936X.2015.1076516. Epub 2015 Sep 11.

Virtual screening of chemical compounds active against breast cancer cell lines based on cell cycle modelling, prediction of cytotoxicity and interaction with targets.

Konova V¹, Lagunin A^{1,2}, Pogodin P^{1,2}, Kolotova E³, Shtil A³, Poroykov V^{1,2}.

Author information

Abstract

Bio- and chemoinformatics methods are widely used for the detection of mechanisms of cancer, to search for potential drug targets and their ligands. Regulatory network analysis based on signalling pathways, and cell cycle regulation provides better understanding of diseases with multiple mechanisms of pathogenesis. We developed an approach for in silico prediction of the cytotoxic effect of chemical compounds in non-transformed and breast cancer cell lines. This approach combines the prediction of the interaction between chemical compounds and human proteins, cytotoxicity and regulatory network modelling taking into account gene expression. Application of our approach to virtual screening of libraries of commercially available compounds allowed selection of dozens of promising hits. These molecules are predicted to interact with the identified targets and exhibit cytotoxicity against breast cancer cell lines but not non-tumour human cell lines. Experimental testing of 49 selected compounds against MDA-MB-231 and MCF7 breast cancer cell lines confirmed the activity of eight compounds with IC50 values ranged from 0.8 to 50 μ M. Thus, the developed approach may be applied for virtual screening for cytotoxic compounds against tumour cell lines.

KEYWORDS: QSAR; breast cancer; cell cycle regulation; cytotoxicity prediction; drug target

- ✓ Разработанный подход был успешно использован для поиска новых фармакологически активных веществ, действующих на клетки рака молочной железы, что было подтверждено экспериментально на линиях MDA-MB-231 и MCF7.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2016610382

PASS CLC Pred

Правообладатели: **Конова Варвара Игоревна (RU), Погодин Павел Викторович (RU), Лагунин Алексей Александрович (RU), Филимонов Дмитрий Алексеевич (RU), Поройков Владимир Васильевич (RU)**

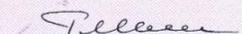
Авторы: **Конова Варвара Игоревна (RU), Погодин Павел Викторович (RU), Лагунин Алексей Александрович (RU), Филимонов Дмитрий Алексеевич (RU), Поройков Владимир Васильевич (RU)**

Заявка № 2015619364

Дата поступления 01 октября 2015 г.

Дата государственной регистрации
в Реестре программ для ЭВМ 11 января 2016 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ивлиев



Библиотеки коммерчески доступных образцов химических соединений

НЕТ

Взаимодействие с мишенью

ДА

НЕТ

Цитотоксичность к опухолевым клеточным линиям

ДА

ДА

Цитотоксичность к неопухолевым клеточным линиям

НЕТ

ДА

Острая токсичность на крысах (GUSAR)

НЕТ

НЕТ

Растворимость в воде (GUSAR)

ДА



Соединения-кандидаты для экспериментального тестирования

PASS Targets

Прогноз взаимодействия ХС с белками-мишенями

PASS CLC Pred

TRANSPATH
ArrayExpress

Моделирование регуляторных сетей клеточного цикла

Список перспективных фармакологических мишеней

ChEMBL

Прогноз цитотоксического действия фармакологически активных соединений на опухолевые и неопухолевые клеточные линии человека



ИБМХ Выборки структур образцов коммерчески доступных органических соединений

| Файл выборки | Структур ХС | Прогноз для клеток РМЖ | % активных соединений |
|--------------------------------|--------------------|-------------------------------|------------------------------|
| Asinex_Merged_Libraries-426400 | 435219 | 21537 | 4,9 |
| chemBlock-2012 | 125485 | 5643 | 4,5 |
| ChemBridge-DVS-2012 | 50000 | 1748 | 3,5 |
| ibs2012_dnc | 45895 | 3357 | 7,3 |
| ibs2013feb_sc1 | 230597 | 8956 | 3,9 |
| ibs2013feb_sc2 | 223165 | 12285 | 5,5 |
| Итого/в среднем | 1110361 | 53526 | 4,8 |

Результаты отбора структур образцов коммерчески доступных органических соединений по парам активностей для РМЖ

| База данных образцов коммерчески доступных ХС | Мишень | | | | | | | | |
|---|--------|------|------|------|-------|----------|------|------|------|
| | PI3K | CDK1 | CDK2 | CDK6 | АКТ-1 | P38ALPHA | НРК1 | МКК3 | МКК6 |
| Asinex_Merged_Libraries | | | 2 | 3 | 14 | | 2 | 23 | 5 |
| chemBlock-2012 | 1 | | | 2 | | | | 20 | 5 |
| ChemBridge-DVS-2012 | | | | | | | | 3 | |
| ibs2012_dnc | | | | | | | | | |
| ibs2013feb_sc1 | 2 | | 1 | 1 | | | | 7 | 5 |
| ibs2013feb_sc2 | 1 | 1 | 1 | 3 | | 1 | | | 4 |

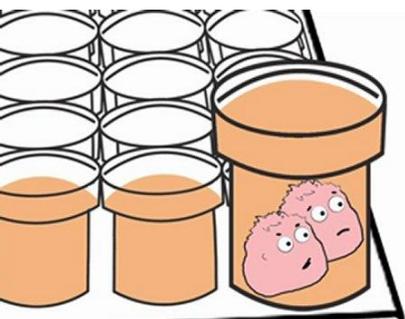
На основе скрининга библиотек коммерчески доступных образцов химических соединений InterBioScreen, Asinex, ChemBlock, ChemBridge были отобраны несколько десятков соединений, для которых с высокой вероятностью одновременно прогнозируется действие на выявленные мишени, цитотоксичность к клеточным линиям РМЖ и отсутствие цитотоксичности к 31 клеточной линии неопухолевых клеток человека.

При экспериментальном тестировании 49 соединений была подтверждена активность 8 соединений.

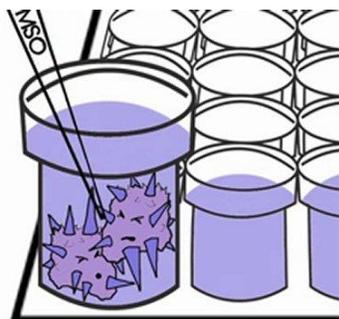
Примеры активных соединений из различных классов химических соединений, найденных с помощью *in silico* скрининга

- 8 активных соединений: $IC_{50} = 0,8 - 50 \mu M$
- Потенциальные кандидаты из различных химических классов:

- ✓ триазолкинолины
- ✓ азулены
- ✓ пиридиламины
- ✓ и др.

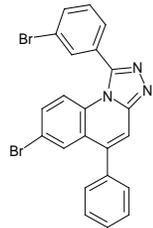
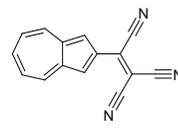
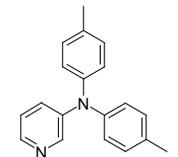
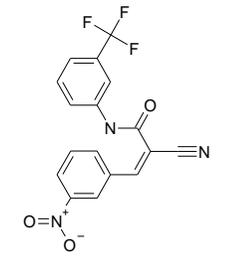


ССК-8



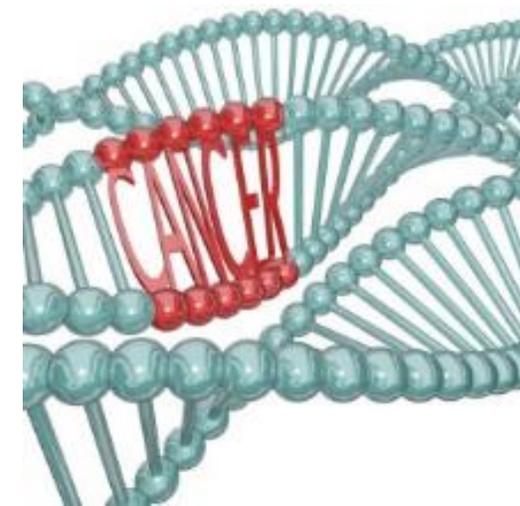
МТТ

(Konova et al., 2015)

| ID | Structural formula and name | IC_{50} against MDA-MB231, μM | IC_{50} against MCF7, μM |
|----|---|--------------------------------------|---------------------------------|
| 1 |  <p>7-Bromo-1-(3-bromo-phenyl)-5-phenyl-[1,2,4]triazolo[4,3-a]quinoline</p> | 12.5 | 14.0 |
| 2 |  <p>2-Azulen-2-yl-3-cyano-but-2-enedinitrile</p> | 22.5 | 40.0 |
| 3 |  <p>Pyridin-3-yl-di-p-tolyl-amine</p> | 22.5 | 18.0 |
| 4 |  <p>2-Cyano-3-(3-nitro-phenyl)-N-(3-trifluoromethyl-phenyl)-acrylamide</p> | 18.5 | 50.0 |

Результаты

- ✓ Создана обучающая выборка из 53774 структур ХС с данными о цитотоксичности химических соединений по отношению к 1781 клеточным линиям человека.
- ✓ С использованием программы PASS (<http://way2drug.com/passonline>) были построены зависимости «структура-цитотоксичность»,
- ✓ Со средней точностью 92%, посчитанной по методу скользящего контроля с исключением по одному, прогнозируется цитотоксичность в отношении 279 опухолевых и неопухолевых клеточных линий человека.
- ✓ С использованием технологии PASS и построенных зависимостей создан веб-сервис для прогноза цитотоксичности химических соединений в отношении опухолевых и неопухолевых клеточных линий человека на основе их структурной формулы CLC-Pred (<http://way2drug.com/cell-line>).



Благодарности



**ЛАГУНИНУ А.А.,
ПОГОДИНУ П.В.,
РУДИК А.В.,
ДРУЖИЛОВСКОМУ Д.С.,
ПОРОЙКОВУ В.В.**

Коллективу лаборатории Структурно-
функционального конструирования
лекарств ИБМХ

Работа выполнена в рамках Программы фундаментальных научных исследований государственных академий наук на 2013-2020 годы

Спасибо за внимание!