Institute of Biomedical Chemistry (IBMC)
Moscow, Russia

# **DIGEP-PRED 2.0:** A WEB-SERVICE FOR PREDICTING DRUG-INDUCED CELL SIGNALING AND GENE EXPRESSION CHANGES

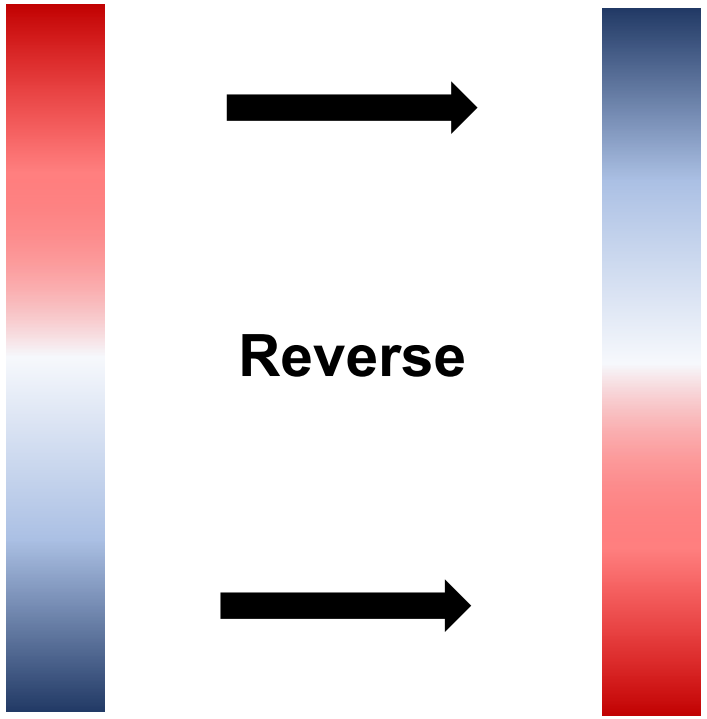
S.M. Ivanov, A.V. Rudik, A.A. Lagunin, D.A. Filimonov, V.V. Poroikov

# Main applications of drug-induced gene expression data

## Connectivity Map

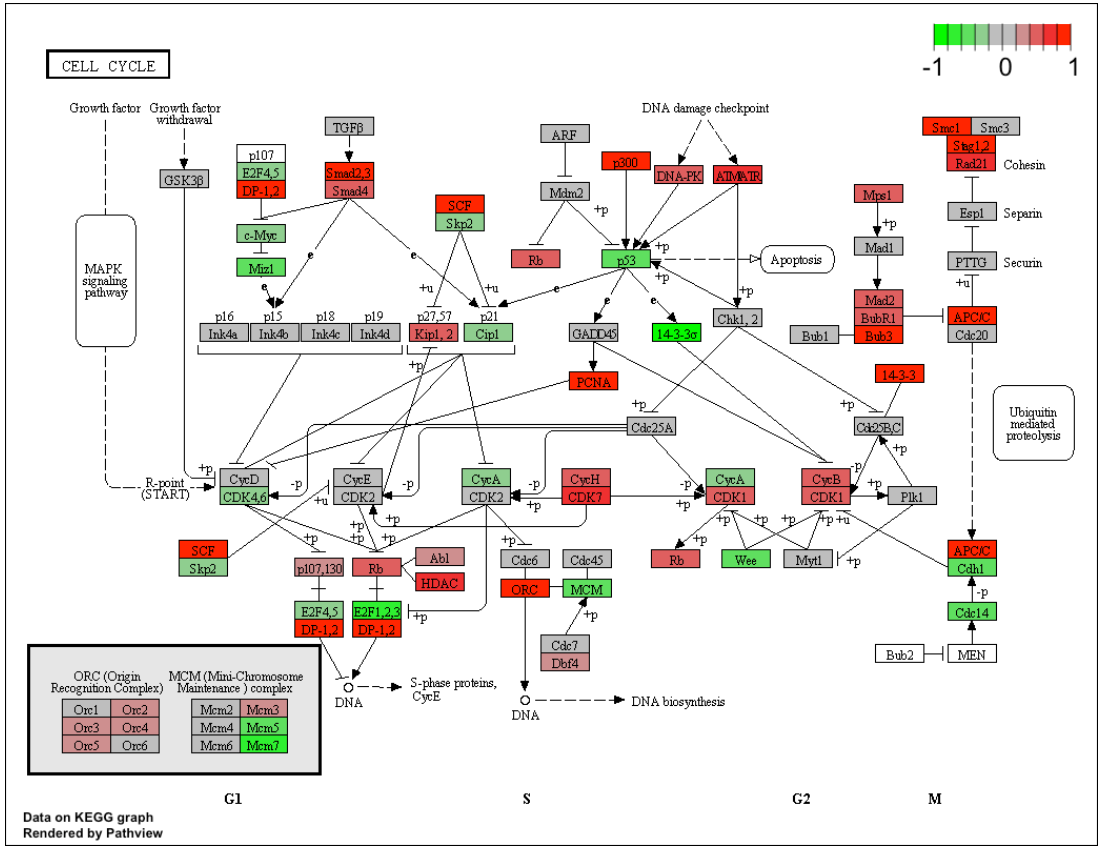
**Drug signature**

**Disease signature**



**Reverse**

## Pathway analysis

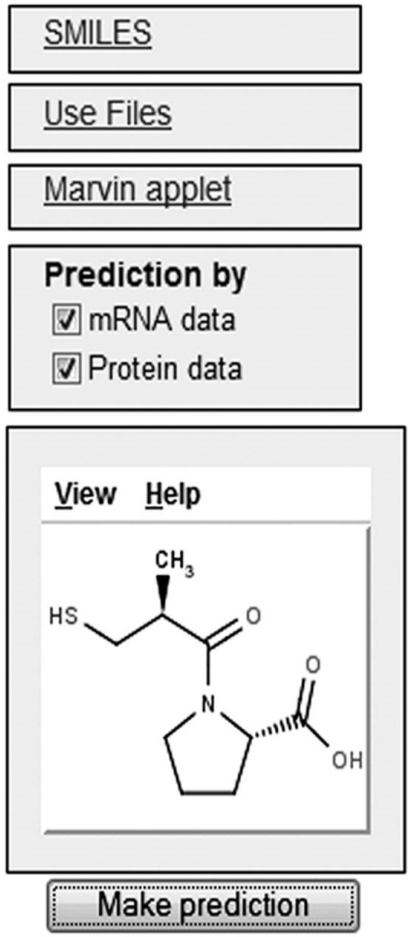


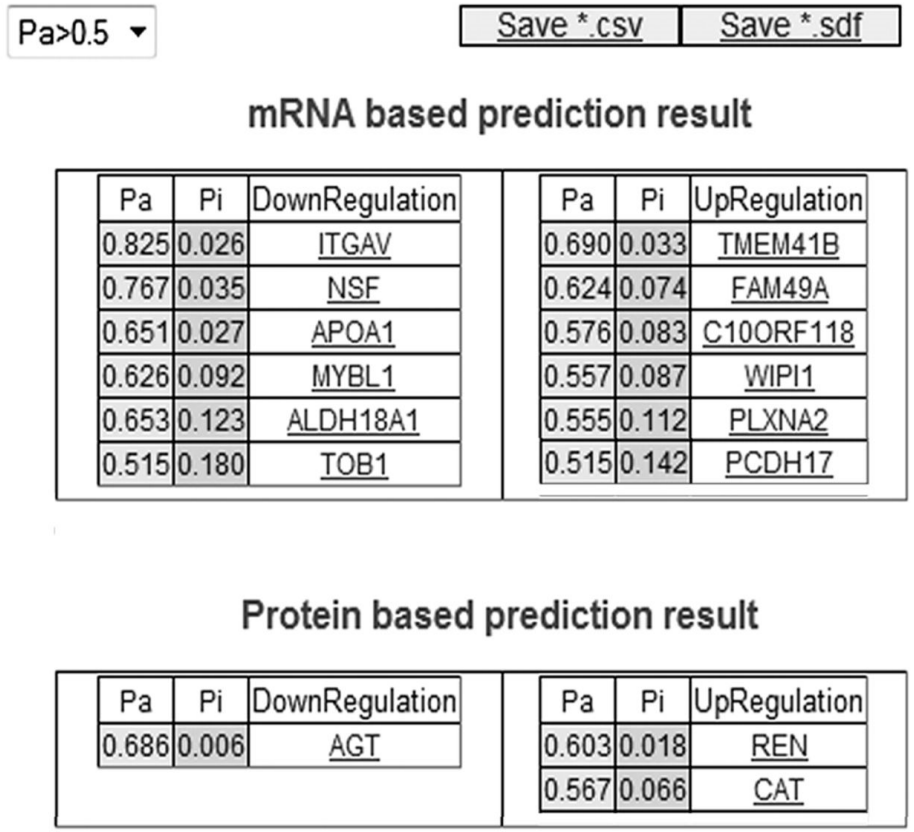

Musa A. et al. Brief Bioinform. 2018; 19(3): 506-523.

# DIGEP-Pred: web service for in silico prediction of drug-induced gene expression profiles based on structural formula

Alexey Lagunin, Sergey Ivanov, Anastasia Rudik, Dmitry Filimonov, Vladimir Poroikov

Bioinformatics. 2013; 29(16): 2062-3.

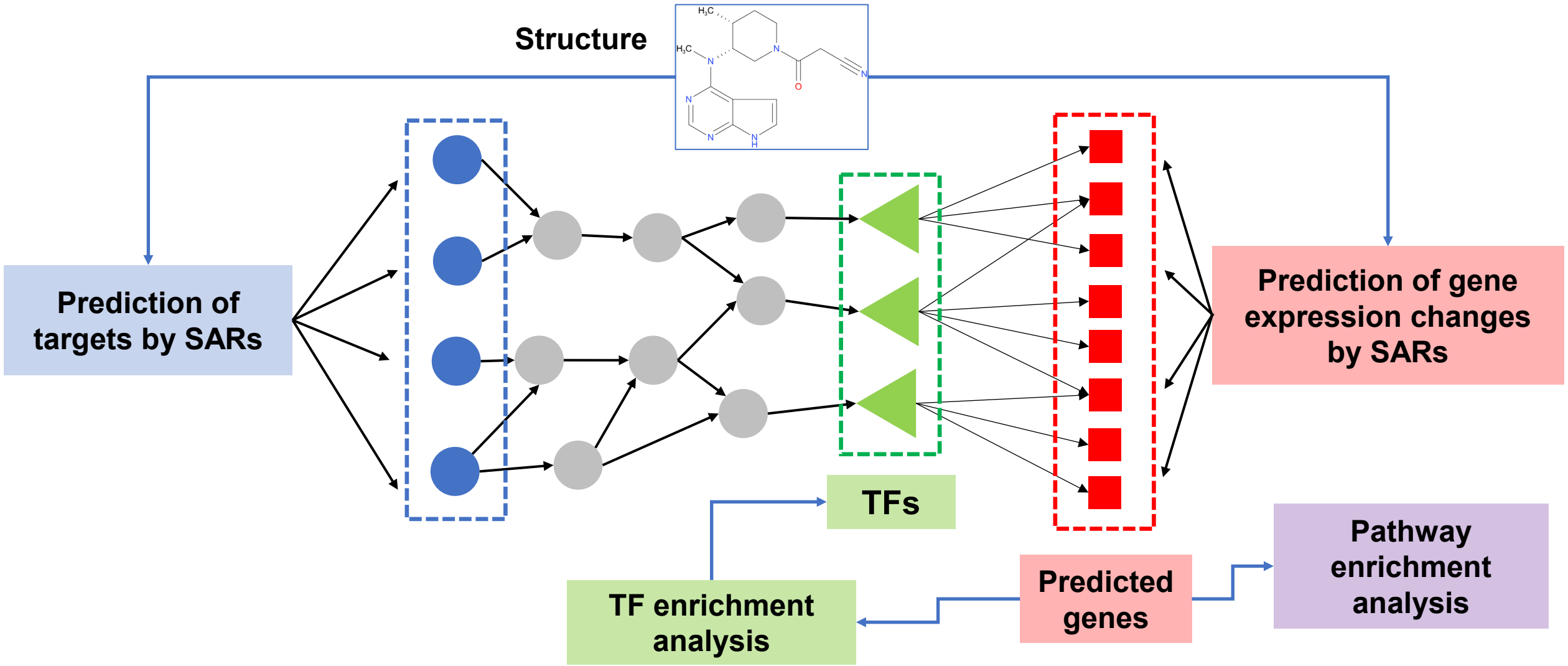



*https://ctdbase.org/*

*http://www.way2drug.com/PASSOnline*

# DIGEP-Pred 2.0 web application: main ideas



# Prediction of drug-induced gene expression changes

ConnectivityMap

| | | | logFC | |
|---|---|---|---|---|
| MCF7 | | | 0.5 | |
| HL60 | X | | 0.7 | |
| PC3 | | | 1 | |
| | | | 1.5 | |
| | | | 2 | |

**15 training sets**

**Comparative Toxicogenomics Database**

mRNA level

Protein level

**2 training sets**

***Examples of biological activity:***

"ADGRB3 UpRegulation", "BCL2 DownRegulation"

*http://www.way2drug.com/PASSOnline*

MNA descriptors

Modified naïve Bayes approach

**Average accuracy = 87.5 %**

# Prediction of drug-induced gene expression changes

| Dataset | \| logFC \| | N of compounds | Selected activities | N of unique genes | Average accuracy, % |
|---|---|---|---|---|---|
| cMAP MCF7 | 0.5 | 856 | 3414 | 3027 | 84.6 |
| | 0.7 | | 2404 | 2179 | 86.0 |
| | 1 | | 1378 | 1283 | 86.9 |
| | 1.5 | | 495 | 482 | 88.1 |
| | 2 | | 218 | 216 | 88.9 |
| cMAP PC3 | 0.5 | 830 | 2408 | 2227 | 83.9 |
| | 0.7 | | 1842 | 1709 | 85.2 |
| | 1 | | 1063 | 1008 | 85.9 |
| | 1.5 | | 442 | 433 | 85.8 |
| | 2 | | 200 | 196 | 86.2 |
| cMAP HL60 | 0.5 | 538 | 3589 | 3334 | 85.6 |
| | 0.7 | | 2722 | 2568 | 87.1 |
| | 1 | | 1564 | 1508 | 89.4 |
| | 1.5 | | 526 | 522 | 91.2 |
| | 2 | | 200 | 200 | 91.1 |
| CTD mRNA | - | 2620 | 18425 | 13377 | 86.5 |
| CTD protein | - | 2671 | 3695 | 2932 | 94.8 |

# Pathway enrichment analysis

| | |
|---|---|
| **Signaling and metabolic pathways** | **KEGG pathways** (https://www.genome.jp/kegg/pathway.html) |
| | **Reactome pathways** (https://reactome.org/) |
| **Cellular processes** | **Gene Ontology biological processes** (https://geneontology.org/) |
| **Diseases** | **DisGeNET** (https://www.disgenet.org/) |
| **Transcription factors** | **CollecTRI** (https://github.com/saezlab/CollecTRI)<br>(Müller-Dott S. et al. Nucleic Acids Res. 2023;51(20):10934-10949) |

Khatri P. et al. PLoS Comput Biol. 2012; 8(2): e1002375.

# Prediction of protein targets (molecular mechanisms of action)

ChEMBL

PubChem

Ki, IC50 < 10 μM

Percent of inhibition > 50%

**656011 compounds for training**


*http://www.way2drug.com/PASSOnline*

**1940 individual proteins**

**2170 molecular mechanisms of action**

***Examples:***
Adenosine receptor A2a agonist,
Amphiregulin inhibitor

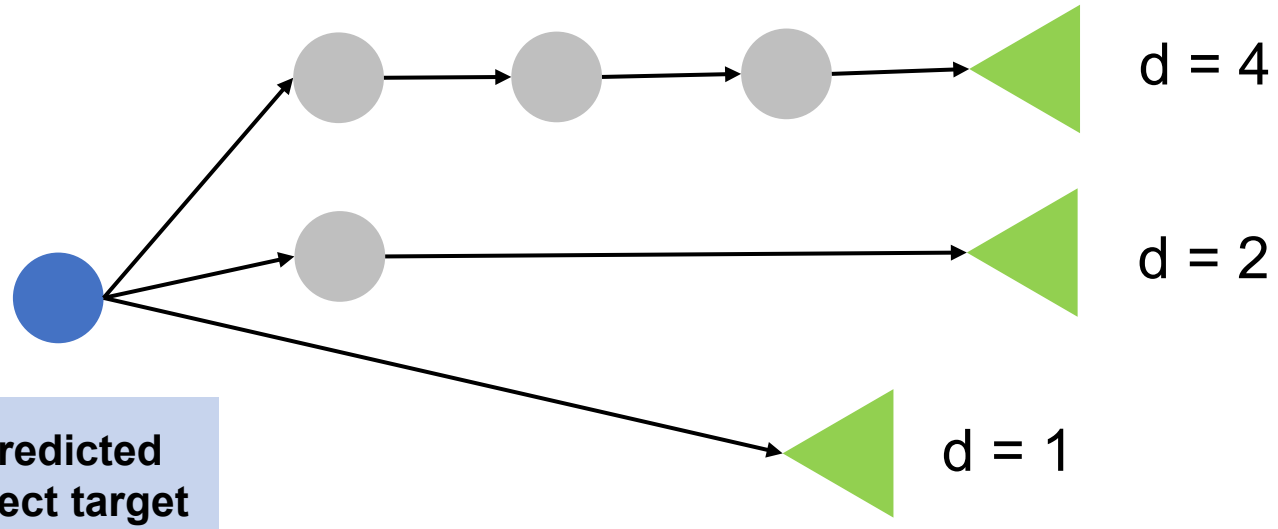
**Average accuracy = 98 %**

Lagunin A.A. et al. Int J Mol Sci. 2023; 24(2): 1689.

# “Upstream” analysis

- ✓ Transcription factors (TFs) from enrichment analysis results
- ✓ Signaling network from OmniPath database (https://omnipathdb.org/) (Türei D. et al. Nat Methods. 2016;13(12): 966-967)

**Shortest path lengths d**




**Predicted direct target**

**Predicted transcription factors**

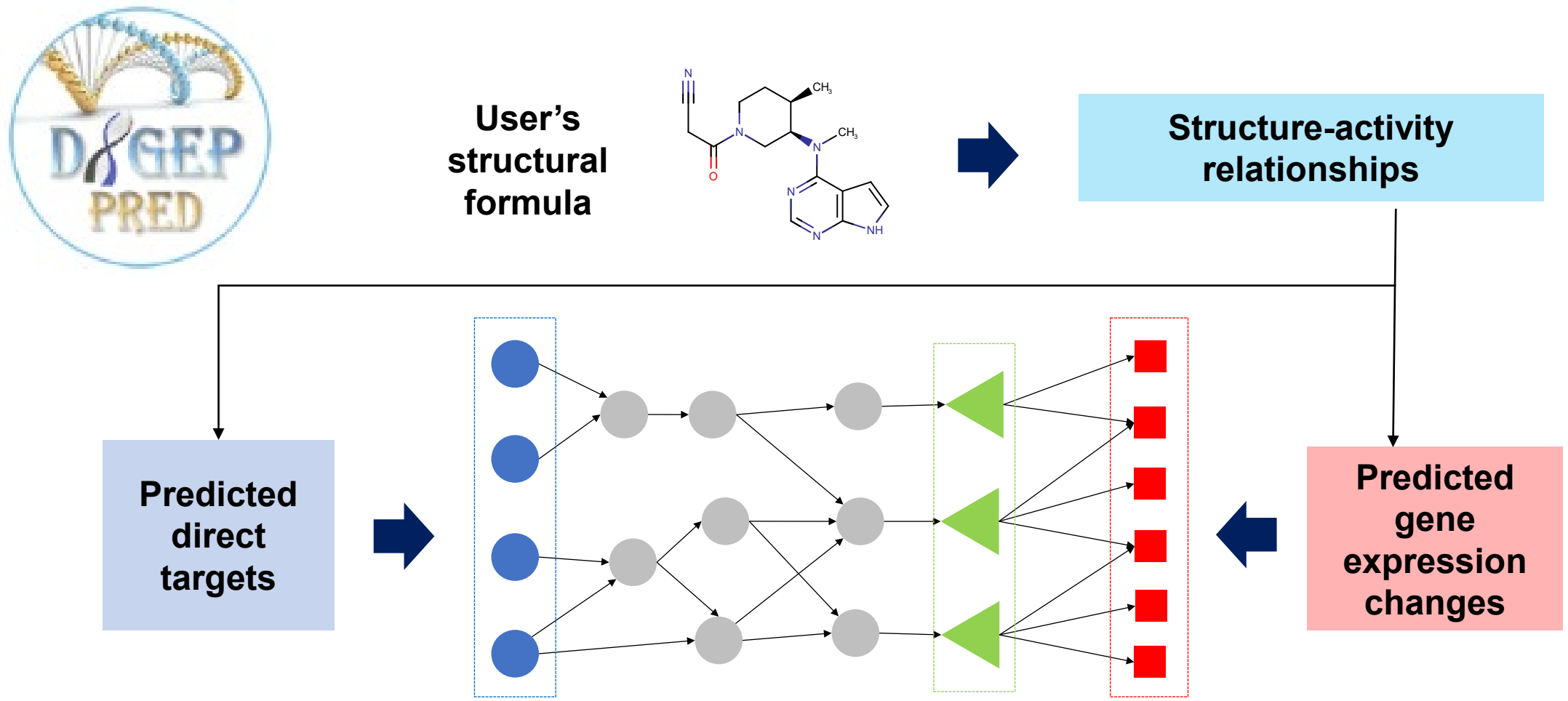
$$Score = \sum_{i=1}^{n} \frac{1}{1 + d_i}$$

Yu H. et al. BMC Syst Biol. 2016;10 Suppl 1(Suppl 1):2.

Lee T., Yoon Y. BMC Bioinformatics. 2018; 19(1): 446.

# DIGEP-Pred 2.0 free available web application

https://www.way2drug.com/digep-pred/



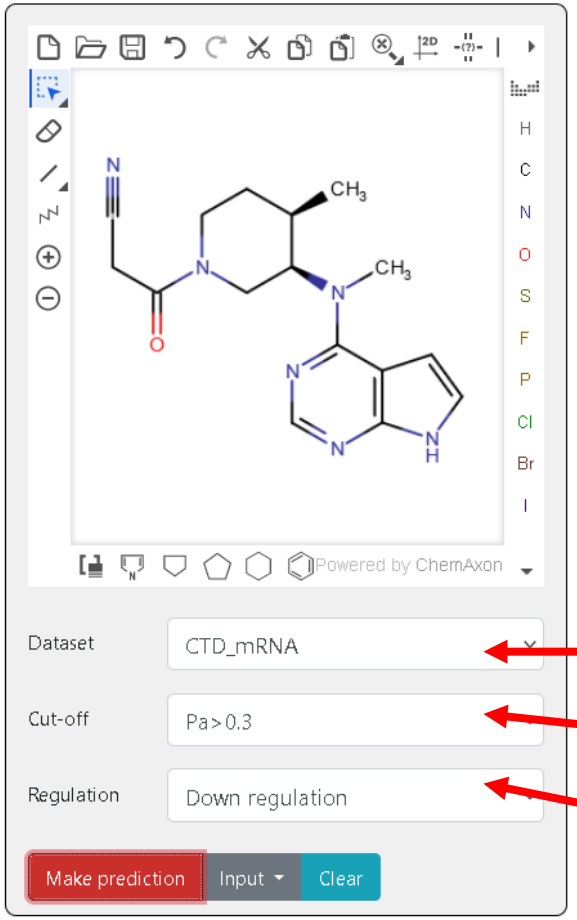
# Predicted gene expression changes for tofacitinib

Way2Drug PREDICTIVE SERVICES
Understanding Chemical-Biological Interactions

*DIGEP-Pred 2.0: Prediction of drug-induced changes of gene expression profile*

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

**Prediction results analysis**




PDF | Copy | CSV | Excel | Enrichment analysis | Potential target search | Show 10 entries

| Pa | Pi | Genes | IAP, LOO CV |
|---|---|---|---|
| 0.995 | 0.000 | CXCL11 | 0.755 |
| 0.980 | 0.001 | IFI35 | 0.773 |
| 0.961 | 0.001 | USP18 | 0.751 |
| 0.949 | 0.000 | IRF7 | 0.819 |
| 0.920 | 0.001 | RSAD2 | 0.785 |
| 0.910 | 0.002 | IFIT3 | 0.779 |
| 0.909 | 0.001 | IFI27 | 0.803 |
| Pa | Pi | Search Genes | IAP, LOO CV |

Showing 1 to 10 of 59 entries

Previous 1 2 3 4 5 6 Next

**Choose model to perform prediction**

**Choose probability threshold**

**Choose direction of gene expression changes**

# Disease enrichment results

Disgenet diseases | Gene Ontology processes | KEGG pathways | Reactome pathways

Copy CSV Excel Show 10 entries Search:

| Name | ID disease | n | N | odds_ratio | -lg10(p.value) | -lg10(adj.p) |
|---|---|---|---|---|---|---|
| Myocardial Ischemia | C0151744 | 8 | 218 | 16.625 | 7 | 5 |
| Celiac Disease | C0007570 | 5 | 48 | 44.838 | 7 | 5 |
| Arthritis, Adjuvant-Induced | C0003865 | 4 | 43 | 39.345 | 5 | 4 |
| Arthritis, Collagen-Induced | C0971858 | 4 | 43 | 39.345 | 5 | 4 |
| Arthritis, Experimental | C0993582 | 4 | 43 | 39.345 | 5 | 4 |
| Rheumatoid Arthritis | C0003873 | 6 | 190 | 13.789 | 5 | 4 |

# Potential anti-COVID-19 mechanisms of tofacitinib

Disgenet diseases | Gene Ontology processes | KEGG pathways | Reactome pathways

Copy | CSV | Excel | Show 10 entries | Search:

| Name | ID pathway | n | N | odds_ratio | -lg10(p.value) | -lg10(adj.p) |
|---|---|---|---|---|---|---|
| Coronavirus disease - COVID-19 | hsa05171 | 8 | 232 | 15.614 | 7 | 6 |

**Category**: 6. Human Diseases

**Subcategory:** 6.3 Infectious disease: viral

**Gene:** CCL2; CXCL10; IL6; MMP1; MMP3; MX1; OAS1; STAT1

**p.value:** 1.54e-7

**adj.p:** 3.06e-6

# Transcription factor enrichment results

Potential targets | Transcription factors

Copy | CSV | Excel | Show 10 entries | Search:

| Transcription factor | n | N | odds_ratio | -lg10(p.value) | -lg10(adj.p) |
|---|---|---|---|---|---|
| STAT1 | 14 | 303 | 23.646 | 13 | 11 |
| IRF9 | 7 | 27 | 115.664 | 12 | 10 |
| IRF1 | 10 | 168 | 28.126 | 11 | 9 |
| IRF2 | 7 | 61 | 51.240 | 9 | 8 |

**Genes** CXCL10; CXCL11; IFI27; IFI35; IL6; TNFSF10; USP18

**p.value:** 3.86e-10

**adj.p:** 1.14e-8

# Tofacitinib protein targets – master regulators

Potential targets | Transcription factors

Copy | CSV | Excel | Show 10 entries | Search:

| Mechanism of action | Pa | Pi | Gene symbol | Uniprot ID | ChEMBL family level 1 | ChEMBL family level 2 | Network score | Rank |
|---|---|---|---|---|---|---|---|---|
| Tyrosine-protein kinase JAK2 inhibitor | 0.606 | 0.003 | JAK2 | O60674 | Enzyme | Kinase | 25.0 | 1 |
| Tyrosine-protein kinase Lck inhibitor | 0.286 | 0.009 | LCK | P06239 | Enzyme | Kinase | 27.1 | 2 |
| Mitogen-activated protein kinase kinase kinase kinase 1 inhibitor | 0.393 | 0.045 | MAP4K1 | Q92918 | Enzyme | Kinase | 25.8 | 3 |

# Conclusions

A new version of the DIGEP-Pred 2.0 web application has been created. It can be used for:

1. Prediction of gene expression changes induced by drug-like compounds in various conditions
2. Estimation of pathways, cellular processes, transcription factors and diseases associated with studied compound
3. Identification of the most probable protein targets of compound which are potentially responsible for induced gene expression changes

To obtain all this information, the user has to provide only the structural formula of a compound

APPLICATION NOTE

molecular informatics

# DIGEP-Pred 2.0: A web application for predicting drug-induced cell signaling and gene expression changes

Sergey M. Ivanov[1,2] | Anastasia V. Rudik[1] | Alexey A. Lagunin[1,2] | Dmitry A. Filimonov[1] | Vladimir V. Poroikov[1]

[1]Department of Bioinformatics, Institute of Biomedical Chemistry, Moscow, Russia

[2]Department of Bioinformatics, Pirogov Russian National Research Medical University, Moscow, Russia

**Correspondence**
Sergey M. Ivanov, Department of Bioinformatics, Institute of Biomedical Chemistry, Pogodinskaya Street, 10 bldg. 8, Moscow 119121, Russia.
Email: smivanov7@gmail.com

**Funding information**
Program for Basic Research in the Russian Federation, Grant/Award Number: No. 122030100170–5

**Abstract**

The analysis of drug-induced gene expression profiles (DIGEP) is widely used to estimate the potential therapeutic and adverse drug effects as well as the molecular mechanisms of drug action. However, the corresponding experimental data is absent for many existing drugs and drug-like compounds. To solve this problem, we created the DIGEP-Pred 2.0 web application, which allows predicting DIGEP and potential drug targets by structural formula of drug-like compounds. It is based on the combined use of structure-activity relationships (SARs) and network analysis. SAR models were created using PASS (Prediction of Activity Spectra for Substances) technology for data from the Comparative Toxicogenomics Database (CTD), the Connectivity Map (CMap) for the prediction of DIGEP, and PubChem and ChEMBL for the prediction of molecular mechanisms of action (MoA). Using only the structural formula of a compound, the user can obtain information on potential gene

Ivanov S.M., Rudik A.V., Lagunin A.A., Filimonov D.A., Poroikov V.V. DIGEP-Pred 2.0: A web application for predicting drug-induced cell signaling and gene expression changes. Mol Inform. 2024:e202400032. doi: 10.1002/minf.202400032.

# Acknowledgement

Anastasia Rudik, Ph.D

Alexey Lagunin, Dr.Sci., Prof.

Dmitry Filimonov, Ph.D

Vladimir Poroikov, Dr.Sci., Prof.,
Corresponding Member of RAS



This work has been supported within the framework of the Program for Basic Research in the Russian Federation for a long-term period (2021–2030) (No. 122030100170-5).

## If you have any questions, please feel free to contact us:

Sergey Ivanov, Ph. D.

Laboratory for Structure-Function Based Drug Design

Institute of Biomedical Chemistry, Moscow, Russia

Tel: (7-499) 246-09-20

E-mail: smivanov7@gmail.com

# Thank you for your attention!