

The background of the slide features a complex digital and biological motif. It includes vertical columns of binary code (0s and 1s) in white and blue, overlaid on a dark, textured background. A faint, glowing blue DNA double helix is visible, winding through the lower portion of the image. The overall color palette is dominated by dark blues, blacks, and oranges.

# UNMATCHED

R&D Solutions

FOR PHARMA & LIFE SCIENCES

ELSEVIER

## Pathway Studio- Assisted Biological Research

Integration of Tools, Data and Expertize  
to solve a customer problem

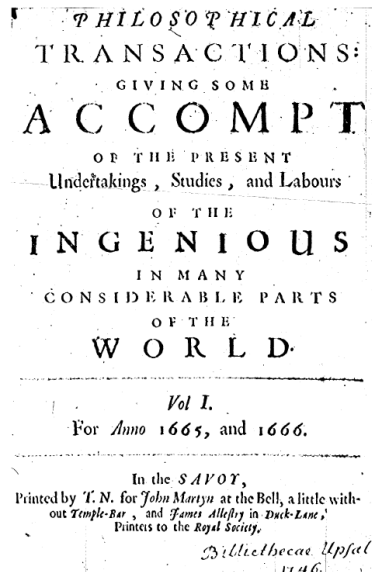
Stephen Sharp, Ph.D.

September 4, 2017



# Scientific information and data exchange in practice for over 400 years

- Most scientific data is still published in unstructured format
  - 17<sup>th</sup> Century
  - 21<sup>st</sup> Century



Royal Society of London  
Oldest learned society (1660)  
Oldest scientific journal (1665)



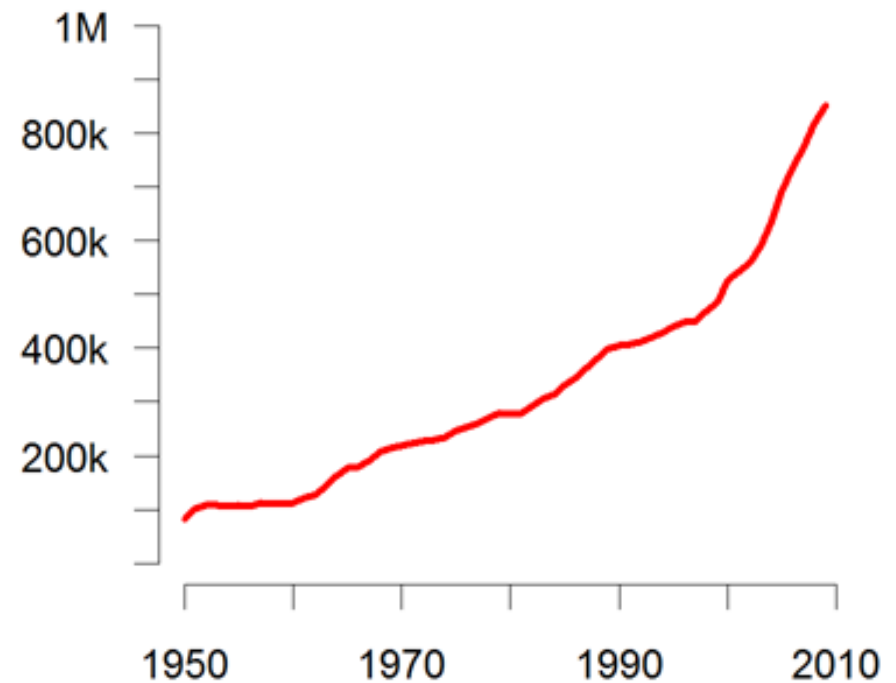


## Scientific literature is exploding

- More than 1M new citations/year in Medline – HOW TO KEEP UP?



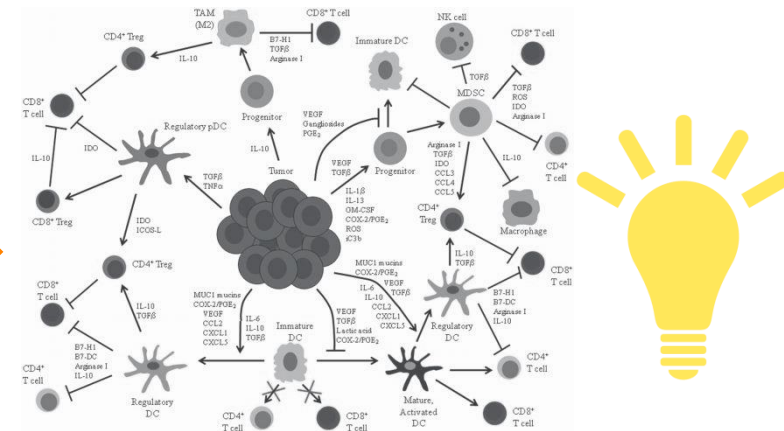
### MEDLINE-indexed articles published per year





# If only we knew what is known...

**Text mining:** analyzing text to extract information that is useful for particular purposes



## Amorphous information

- Hard to deal with
- Hard to deal with algorithmically
- Not scalable

## Structured information

- Search
- Visualize
- Network analysis
- Scalable
- Compressed



# History of MedScan Technology

- 2001- Ariadne Genomics developed **MedScan** – a tool to extract information for bio-molecular networks from literature

BIOINFORMATICS

Vol. 19 no. 13 2003, pages 1699–1706  
DOI: 10.1093/bioinformatics/btg207



***MedScan, a natural language processing engine for MEDLINE abstracts***

*Svetlana Novichkova, Sergei Egorov and Nikolai Daraselia\**

*Ariadne Genomics, Inc., 9100 Great Seneca HWY, Rockville, MD 20850, USA*

*Received on January 11, 2003; revised on February 11, 2003; accepted on March 17, 2003*

- 2012 - Elsevier acquired Ariadne Genomics and continued to develop **MedScan** and **PATHWAY STUDIO**



**PATHWAY STUDIO®**

**Human  
Mouse  
Rat**

- **2017- MedScan** engine can be used within **PATHWAY STUDIO** and also independently



## From text to fact

text



24 M abstracts  
3.5 M full texts

sentence



Our results suggest that a natural immune response mediated by  **$\gamma\delta$  T lymphocytes** may contribute to the immunosurveillance of melanoma



fact



standard  
name

standard  
link

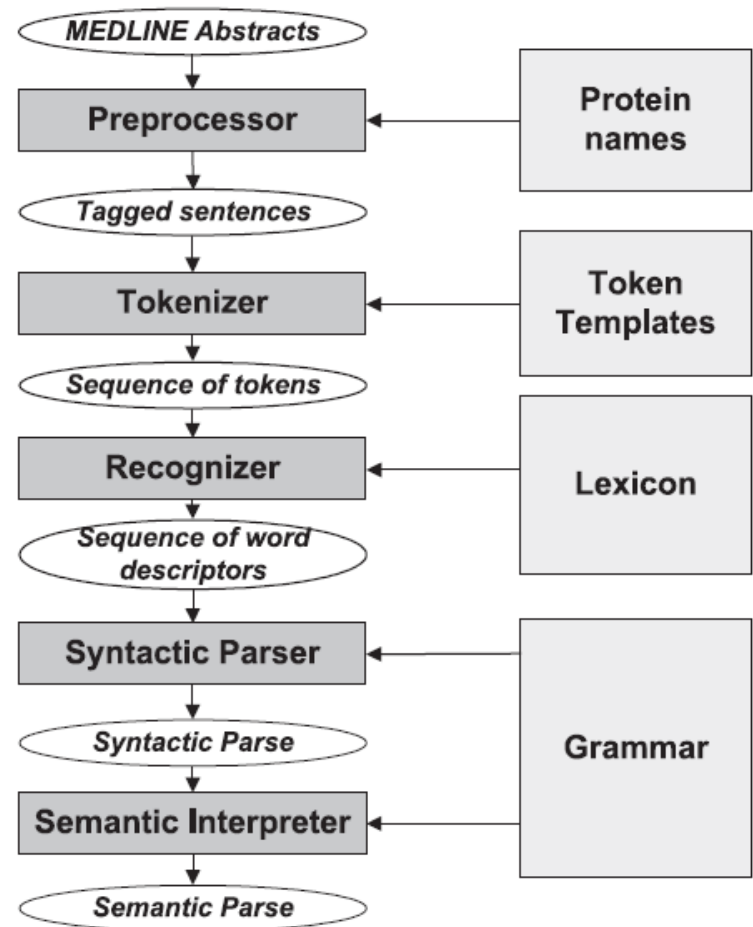
standard  
name

5.6 M facts



## Processing of Text using MedScan Technology

- Staged Processing:
  - Sentence tokenizing
  - Word recognition
  - Morphological analysis
  - Recognition of compound lexemes
  - Syntactic parsing
  - Semantic interpretation.





# Text Mining with MedScan Technology

## Natural Language Processing

The central idea of **MedScan's** NLP algorithm is decomposing natural language sentences into semantic relationships (which we will also call semantic triplets). Each triplet is designed to represent a single semantic relationship between two singular noun phrases (NPs).

11940574:7 Because **Axin2** has been shown to associate with and inhibit **beta-catenin** abundance and function, we hypothesized that **Axin2**, which is affecting proliferation of MEF cells can work in a negative feedback pathway, regulating **Wnt** signaling and thus controlling apoptotic process.

### Triplets identified:

- Axin2 associate beta-catenin abundance
- Axin2 inhibit beta-catenin function
- Axin2 associate beta-catenin abundance
- Axin2 inhibit beta-catenin function
- Axin2 affect MEF cell line proliferation
- Axin2 work negative feedback pathway
- Axin2 regulate Wnt signaling
- Axin2 control apoptotic process



## Data Extraction using Elsevier NLP

How MedScan Identifies “Entities” and “Relationships”

“Axin binds beta-catenin and inhibits GSK-3beta activity in hepatocytes.”



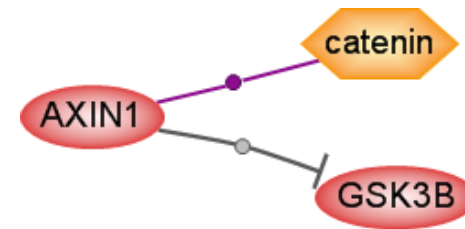
*Text Mining recognizes  
“Semantic Triplets”*

Entity	Relationship	Entity
<b>Axin</b>	<u>binds</u>	<b>beta-catenin</b>
<b>Axin</b>	<u>inhibits</u>	<b>GSK-3beta</b> activity

*Stores this  
information in  
the database...*



*Pathway Studio visualizes the  
information in graphical  
format*



Axin - beta-catenin, relation: Binding celltype: hepatocyte

Axin -> GSK-3beta, relation: Regulation, effect: Negative celltype: hepatocyte



# Solution overview

## Pathway Studio®

### Knowledgebase

Biological relations extracted from literature

Manually curated pathways

Ontologies Annotations

Structural similarity for chemicals

Variation annotation from public sources

### Tools

Search  
Summarization  
Navigation  
Visualization

Experiment analysis:  
• Gene expression  
• Proteomics  
• Metabolomics  
• NGS

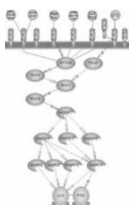
Text mining pipeline

Easy to use text mining interface

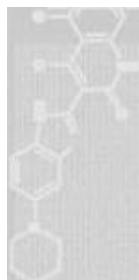
24M abstracts  
3.5M full texts



6.2M relations



>1800 pathways

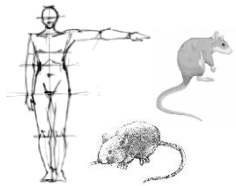


Gene	Accession	Length	Score	Rank
CA12	U08548	1089	100.0	1
TIMP1	U08549	1089	99.9	2
TIMP2	U08550	1089	99.8	3
MMP2	U08551	1089	99.7	4
MMP3	U08552	1089	99.6	5
CDK6	U08553	1089	99.5	6

Mining  
Models  
Retrieval  
Semantic  
Web  
Classification  
Linking  
Information  
Text



# Pathway Studio databases



Mammal

Protein  
centered

- Target interactions data
- Expression regulation
- Transport
- Synthesis
- Protein modification
- Functional characterization



Chem  
Effect

Chemical  
centered

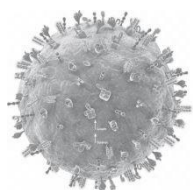
- Chemical-target interactions
- Adverse events
- Drug repurposing



Disease  
FX

Disease  
centered

- Biomarkers
- Drug repurposing
- Clinical trials



Cells

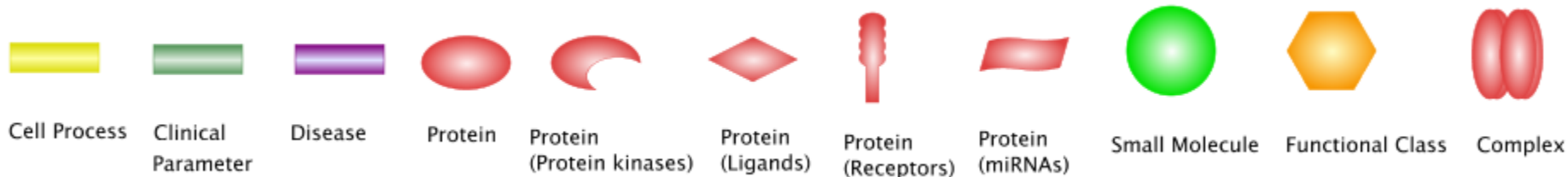
Cell  
centered

- Cell expression
- Cells role in the disease
- Cells as biomarkers



# Types of objects (Entities) in the database

Entity	Example
Protein, Complex, Functional class	PARP1, miR133b, IL23, cytokine
Drug, Metabolite, Ion, Chemical	glucose, aspirin, estradiol, cetuximab
Disease	malaria
Clinical Parameter	heart rate, lesion size, Gleason score
Cell Process	apoptosis
Treatment	heat shock
Cell	regulatory T cell





## Integrating text with other resources

All recognized terms have identifiers:

- Linking to other databases
- Mapping
- Integrating external data
- Use of ontologies

IUPAC Name: 4-[4-[[4-chloro-3-(trifluoromethyl)phenyl]amino]phenyl]pyridine  
 Molecular Formula: C<sub>21</sub>H<sub>16</sub>ClF<sub>3</sub>N<sub>4</sub>O<sub>3</sub>  
 PharmaPendium ID: Sorafenib Tosylate  
 Rotatable Bond Count: 5  
 XLogP-AA: 4.1  
 Molecular Weight: 464.82495

Notes: This gene encodes a chromatin-  
 proliferation, and tumor transfo  
 [provided by RefSeq, Jul 2008]  
 Alias: 5830444G22; 5830444G22Rik; A  
 polymerase); ADP-ribosyltransfe  
 ADPRT 1; ADPRT 1; ADPRT1; AI8  
 ADP-ribosyltransferase, nuclear;  
 nuclear NAD ADP-ribosyltransfe  
 polymerase; poly [ADP-ribose] p  
 ribosyl)transferase; poly[ADP-rib  
 Connectivity: 5614  
 Primary Cell Localization: [Nucleus](#)  
 Cell Localization: [Nucleus](#)  
 Organism: [Homo sapiens](#)  
 Human chromosome position: 1q41-q42  
 Rat chromosome position: 13q26  
 Mouse chromosome position: 1 84.44 cM  
 Owner: public  
 URN: urn:agi-llid:142  
 Date Created: 2015-12-18 06:22:20.026  
 Date Modified: 2015-12-18 06:22:20.027

Entrez GeneID: 11545; 142; 25591  
 Unigene ID: Hs.177766; Mm.277779; Rn.11327  
 Swiss-Prot Accession: A0A024R3T8; B1ANJ4; O35937; P09  
 Swiss-Prot ID: PARP1\_HUMAN; PARP1\_RAT  
 OMIM ID: 173870  
 MGI ID: MGI:1340806  
 Hugo ID: 270; HGNC:270  
 RGD ID: 2053  
 Ensembl ID: ENSG00000143799; ENSMUSG0000  
 GenBank ID: AAA51599; AAA51663; AAA60000;  
 ABM85752; AC121810; AC\_000023;  
 AK312339; AL359704; AL359742; A  
 CAA46478; CBX74363; CH466555; C  
 NM\_001618; NM\_007415; NM\_013  
 GO ID: 0000122; 0000723; 0003677; 00039  
 0023019; 0030225; 0032869; 00400  
 EC Number: 2.4.2.30  
 KEGG ID: hsa:142; mo:25591  
 Homologene ID: 1222  
 PIR ID: A29725; S21163; S26057  
 MedScan ID: 142



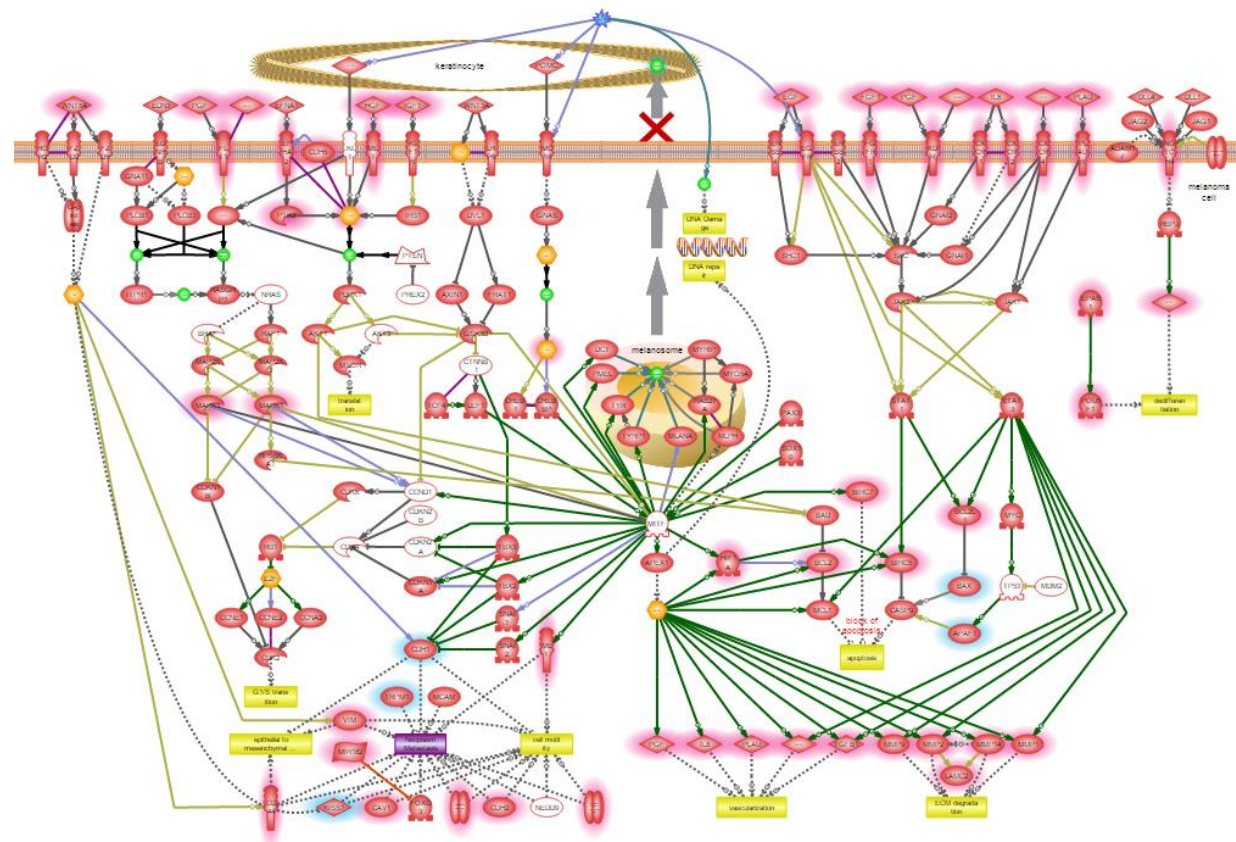
# Rich pathway collections for modeling and experiment analysis

Over 1,800 pathways manually built by PhD level scientists (ongoing project)

- Signaling
- Metabolic
- Cell processes
- Disease
- Immunological
- Expression targets
- Toxicity
- Nociception

## To use

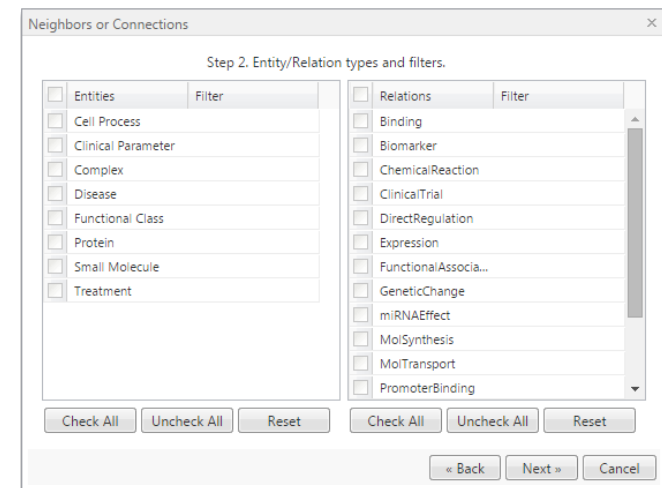
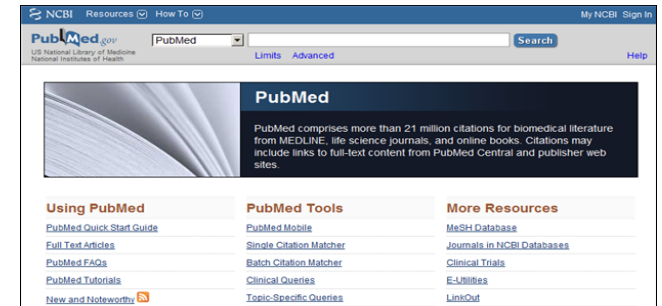
- As starting points to build pathways
- In experiment analysis
- In analysis of groups





# Interface and tools to find answers to complex biological questions

- Visualization
- Summarization
- Intersect, subtract, and union facts and lists
- Filter by
  - Identifiers
  - Bibliography
  - Tissues, organs, cells, cell cultures, organisms
  - Drugability









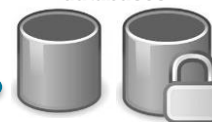
# Elsevier R&D Solutions: data, software, capabilities

## Content & Data

Elsevier and non-Elsevier  
textual information



Public and proprietary  
databases



## Capabilities

- Data extraction
- Data normalization
- Data integration

**Research &  
Discovery**

**Lead ID &  
Valid**

**Pre-clinical**

**Clinical**

**Post-launch**

- Disease modeling
- Target identification
- Biomarker discovery
- Drug repositioning

- Lead identification and characterization
- Synthesis optimization

- Lead prioritization for safety, delivery and efficacy
- Translational medicine

- Monitoring adverse events

## Databases Software

Pathway Studio

Reaxys<sup>®</sup>  
Reaxys Medicinal Chemistry

PharmaPendium<sup>®</sup>

Embase<sup>®</sup>

## Use-case centered integration and customization

Scopus

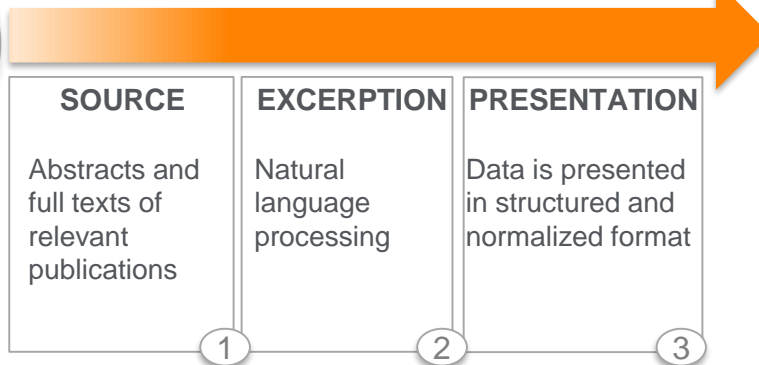
Professional Services Elsevier Text Mining



# Literature-extracted biology data: from disease mechanisms to targets

- What causes the disease of interest?
- What is the disease mechanism?
- What is it similar to?

Information  
published by  
researchers



Pathway Studio database contains:

**>285,000**  
Biological concepts

**>7**  
**MILLION**  
Relations between  
biological concepts

**>1,800**  
Manually built  
pathways, including  
disease models

The unique collection of data is sourced from:

**>10,000**  
Journals, including  
**>1,700**  
Full text journals

**>4**  
**MILLION**  
Full text articles

**>38**  
**MILLION**  
Supporting sentences

Pathway Studio®



# Rare diseases – when every piece matters



Nick Sireau at TEDx ImperialCollege  
<https://www.youtube.com/watch?v=B4UnVIU5hAY>

- Patients community
- Collaboration with medical researchers
- Drug repurposing candidate
- Fundraising
- Clinical Trial

findacure

ELSEVIER

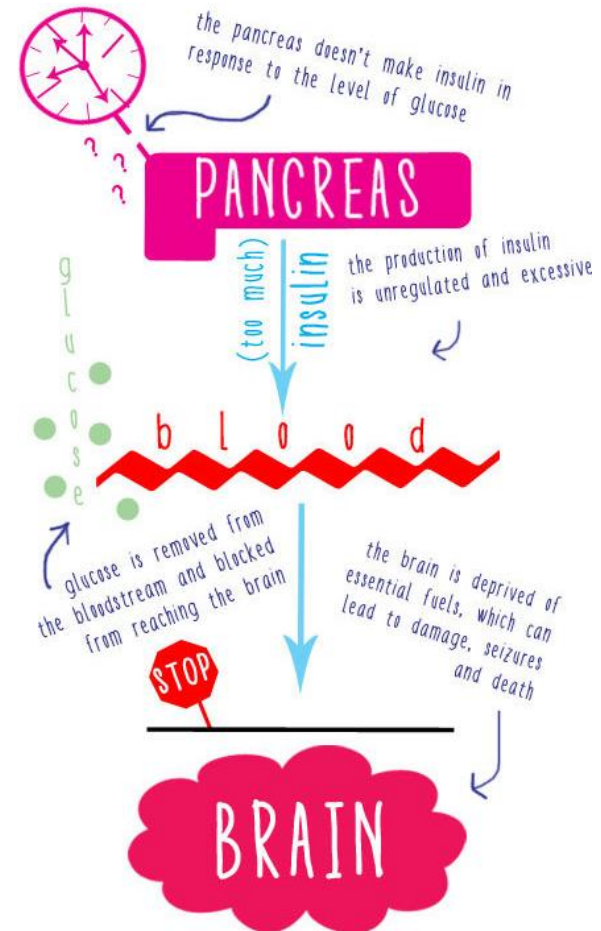
is a UK charity that is building the rare disease community to raise awareness, drive research and develop treatments.

is partnering with Findacure scientists to help identify and evaluate treatments for congenital hypersinsulinism



# Congenital hyperinsulinism (CHI)

- A rare genetic disease
- Permanently excessive level of insulin in the blood
- Develops within the first few days of life
- Can lead to brain injury or even death
- In the most severe cases the only viable treatment is the removal of the pancreas, consigning the patient to a lifetime of diabetes



[https://res.cloudinary.com/indiegogo-media-prod-cld/image/upload/c\\_limit,w\\_620/v1440424745/uzvnqzhvbpsrtthzxqpu.jpg](https://res.cloudinary.com/indiegogo-media-prod-cld/image/upload/c_limit,w_620/v1440424745/uzvnqzhvbpsrtthzxqpu.jpg)

How can we help?



# Congenital hyperinsulinism library

In support of Findacure's mission of education and knowledge sharing:

- Access to all Elsevier's ScienceDirect full-text publications covering CHI
- Collection of papers focused on different aspects of CHI
- Collection of papers focused on effects of sirolimus on CHI

ScienceDirect  
Mendeley

**Congenital Hyperinsulinism**

**By disease subtype**

- Diffuse CHI
- Focal CHI
- Persistent CHI
- Transient CHI

**By study type**

- Case reports
- Cohort studies
- Genetic studies
- Geographic location
- Reviews
- CHI-focused
- CHI-related
- Elsevier full-text publications

**Sirolimus**

- Sirolimus in CHI
- Sirolimus: insulin sensitivity and resistance
- Sirolimus: insulin synthesis and release

**CHI-focused in Congenital Hyperinsulinism**

Authors	Title	New	Published In	Added
Perina, Christine; Patel, Payal; Be...	Bornetars of insulin for the diagnosis of hyperinsulinism/hypoglycemia in Infants and Children	2015	Journal of Pediatrics	Feb 26
Zhang, Wen; Liu, Li; Wen, Zhe; Ch...	A compound heterozygous mutation of ABCC8 gene causing a diazoxide-unresponsive congenital hyperinsulinism with an atypical form: Not a focal l...	2015	Gene	Feb 26
Young, Li; Thornton, P	50% Long Acting Somatostatin Analogues: Early Experience in the Treatment of Patients with Congenital Hyperinsulinism	2015	Journal of pediatric nursing	Feb 26
Namukana, Okubako; Hashimoto, T...	Two patients with HNF4A-related congenital hyperinsulinism and renal tubular dysfunction: A clinical variant which includes transient hepatic dys...	2015	Diabetes Research and Clinical Practice	Feb 26
Jiao, Yuchan; Lundquist, Kimberley T...	Endothelial papillary network neovascularization in neonates with congenital hyperinsulinism and a de novo germline SHG gene mutation	2015	Pancreatology	Feb 26
Skolchekina, I O; Melnik, H A; Z...	Epilepsy and neurological manifestations in children with congenital hyperinsulinism	2015	European Journal of Pediatric Neurology	Feb 26
Bennett, James T; Velez, Valeria Z...	Molecular genetic testing of patients with monogenic diabetes and hyperinsulinism	2015	Molecular Genetics and Metabolism	Feb 26
Shi, Yanyan; Antipade, Hina B; Li, S...	Increased plasma insulin concentrations identifies a subset of patients with persistent congenital hyperinsulinism without KATP channel gene defects	2015	Journal of Pediatrics	Feb 26
Tornovsky-Gabay, Shanna; Dado...	Type 2 diabetes and congenital hyperinsulinism cause DNA double strand breaks and p53 activity in T cells	2014	Cell Metabolism	Feb 26
Norifu, Yohu	Congenital hyperinsulinism: current status and future perspectives	2014	Annals of Pediatric Endocrinology & Metab...	Feb 26
Jindal, Radhika; Ahmad, Anvisha S; S...	New mutation c.597_T88G in exon 5 of ABCC8 gene causing congenital hyperinsulinism	2014	Diabetes & Metabolic Syndrome	Feb 26
Kumar, G; Dhal, S S; Karanavathi...	ABCC8/ABCC9/PETCT revealing renal dominant scan: Lack of physiological uptake in the spleen of a neonate and the pituitary gland in congenital hyp...	2014	Revista Brasileira de Medicina Nuclear & Im...	Feb 26
Petratou, Ioannis; Bavaresco, Gerd...	Congenital hyperinsulinism: Clinical and molecular analysis of a large Italian cohort	2013	Gene	Feb 26
Palstra, Flavio; Afrascaniu, Emma; Giffert...	Hemolysis in a child with a paternally-inherited ABCC8 mutation and mosaic paternal uniparental disomy 12p causing focal congenital hyperinsulinism	2013	European Journal of Medical Genetics	Feb 26
Calton, Elizabeth A; Temple, L; Kar...	Pancreatic surgery in infants with Beckwith-Wiedemann Syndrome and hyperinsulinism	2013	Journal of Pediatric Surgery	Feb 26
Lipik, Pablo; Padalino, Andrew A; J...	Concomitance of two ABCC8 mutations causing an unresponsive congenital hyperinsulinism: Clinical and functional characterization of two novel ABCC...	2013	Gene	Feb 26
Palstra, Flavio; Sinder, Kara; Sheng...	Accuracy of PET/CT scan in the diagnosis of the focal form of congenital hyperinsulinism	2013	Journal of Pediatric Surgery	Feb 26
Lipik, Pablo; Stables, Lisa J; Zhu...	Congenital hyperinsulinism Associated With Beckwith-Wiedemann Syndrome	2012	Journal of Pediatric Nursing	Feb 26
Datta, Enay	Fetal adrenen among patients with persistent hyperinsulinemic hypoglycemia of infancy	2012	Journal of Pediatric Surgery	Feb 26
Al-Shanefy, Saad; Alkhatib, Huss...	Pancreatic head resection and Roux-Y pancreaticojejunostomy for the treatment of the focal form of congenital hyperinsulinism	2012	Journal of Pediatric Surgery	Feb 26
Lipik, Pablo; Stanley, Charles A; Pa...	Two genetic forms of hyperinsulinemic hypoglycemia caused by dysregulation of glutamate dehydrogenase	2011	Neurochemistry International	Feb 26
Stanley, Charles A	Pre-surgical application of preoperative genetic diagnosis and haplotyping for congenital hyperinsulinism	2011	Paediatric Endocrinology	Feb 26

**Details**

Type: Journal article

**A compound heterozygous mutation of ABCC8 gene causing a diazoxide-unresponsive congenital hyperinsulinism with an atypical form: Not a focal lesion in the pancreas reported by 18F-DOPA-PET/CT scan**

Authors: W. Zhang, L. Liu, Z. Wen et al.

Journal: Gene

Year: 2015

Volume: 572

Issue: 2

Pages: 222-226

**Abstract:**

Congenital hyperinsulinism (CHI) is a severe heterogeneous disorder due to dysregulation of insulin secretion from the pancreatic  $\beta$  cells leading to severe hypoglycemia in infancy. 18-Rules 1-3, 4-difluorophenylalanine positron emission tomography (18F-DOPA-PET/CT) is a useful tool in distinguishing between focal and diffuse disease preoperatively. But recent studies have suggested that the scanning may not be accurate as initially estimated. In the study we characterized a case of CHI with a compound heterozygous mutation of ABCC8 gene. The results of clinical investigation, gene mutation analysis, 18F-DOPA-PET/CT scan, and pathological examination showed some new characteristics that have never been reported. The patient was preoperative to medical therapy with diazoxide and received preoperative resection, which was confirmed a compound heterozygous mutation in ABCC8 genes. Imaging with 18F-DOPA-PET/CT indicated a focal lesion in the head of the pancreas. The pathological diagnosis was an atypical form of CHI. We suggest a comprehensive of atypical CHI unresponsive to diazoxide. It is considered that a relationship existed between the compound heterozygous mutation and the atypical form. 18F-DOPA-PET/CT is a useful tool in distinguishing between focal and diffuse forms preoperatively but the accuracy is not 100%. The scan result is best combined with genetic analysis and intra-operative biopsy to confirm the histological substrates. The combination will provide the optimal strategy for the surgical treatment of patients with CHI.

**Tags:**

**Author Keywords:**

18F-DOPA-PET/CT; ABCC8 gene; Congenital hyperinsulinism

**Type of Work:**

JOUR

**URL:**

ADD URL...

**Catalog IDs**

Arxiv ID:

DOI: 10.1016/j.gene.2015.07.012

ISBN: 03043828

PMID: 26362674



# Creating a comprehensive view of CHI with Elsevier R&D Solutions

- **CHI Library**
- **Disease, Target, Pathway, and Compound Analysis**
- **Research Landscape Analysis**

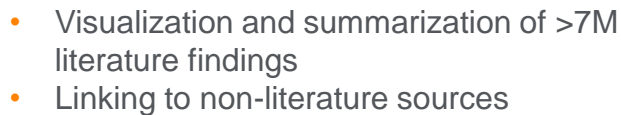
## Information Assets Applied

- **Content**  
Elsevier's vast set of literature and patent data
- **Data normalization**  
Taxonomies and dictionaries to normalize author names, institutions, drugs, targets, and other important terms
- **Information extraction**  
Finding semantic relationships, targets, pathways, drugs, and bioactivities





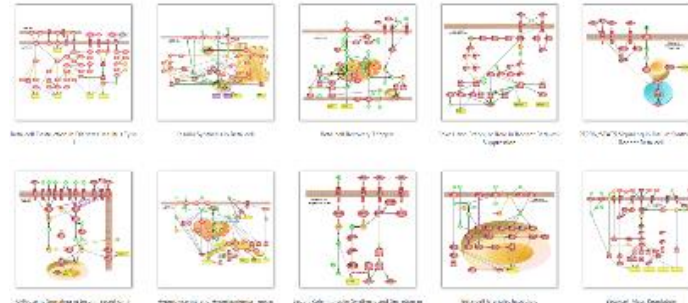
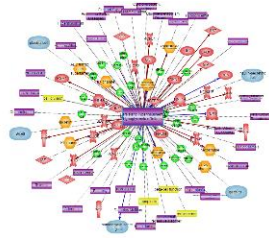
## Pathway Studio®





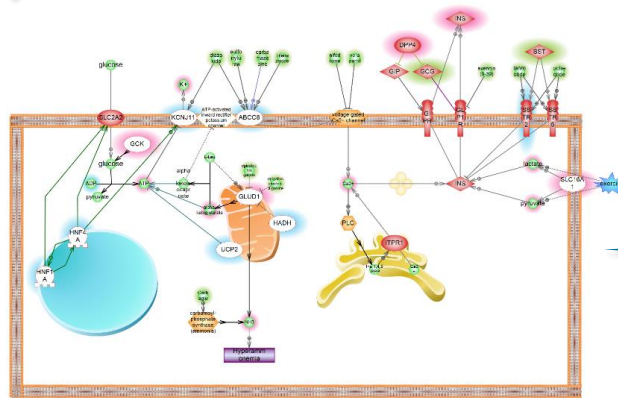
# Building and refining the disease model

Literature-extracted data



Relevant pathways  
(from a collection of 1800 models)

- Diseases
- Cell processes
- Signaling
- Metabolism
- Toxicity

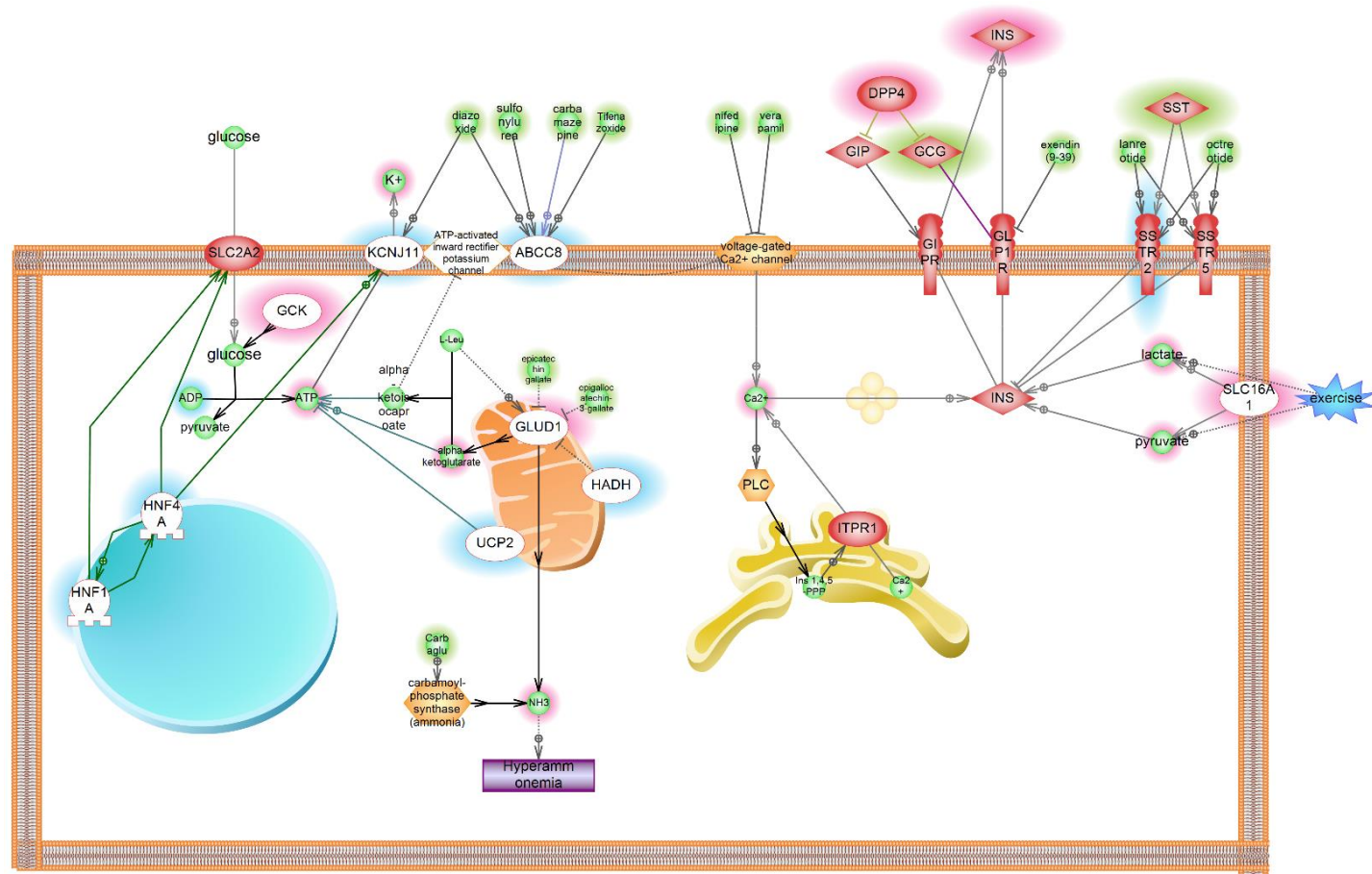


Disease  
overview



# CHI: Building and refining the disease model

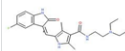

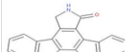
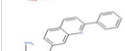

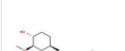
- Insulin secretion steps affected by CHI
- Role of mutated genes
- Drug targets
- Drugs





# Approved compounds that may treat CHI

- Each binds to one or more targets related to the disease
- Can easily be obtained and tested in preclinical studies
- List includes compounds known to treat hyperinsulinism

Molecule	N	NumberIn Group	Reaxys Registry Number (IDE.XRN)	Target	pX (DAT.PAURE US)_Median	Chemical Name (IDE.CN)
	125	5	9364276	AKT1 ROCK1 IGF1R SGK1 FAK	7.1300 6.3400 6.2550 6.4750 6.3600	N-[2-(diethylamino)je... [3H]-Sunitinib Sunitinib 5-(5-Fluoro-2-oxo-1,... sunitinib sunitinib SU 11248
	136	4	11751576	AKT1 JNK1 ROCK1 SGK1	7.1100 7.8300 7.3800 7.4200	[14C]-nilotinib Nilotinib molesanib N-(2,3-dihydro-3,3-d... N-(3,3-dimethyl-2,3-... N-(3,3-dimethylindol... AMC 706
	193	3	9305136	JNK1 ROCK1 IKKb	7.9600 7.2500 7.9200	Lestaurtinib (9S-(9a,10b,12a))-2,... CEP-701 CEP701 lestaurtinib
	231	3	18476426	INSR ROCK1 IGF1R	7.5100 6.2000 7.6200	Linsitinib OSI-906 cis-3-[8-amino-1-(2-... linsitinib cis-3-[8-amino-1-(2-... cis-3-[8-amino-1-(2-...
	366	2	11300181	Dipeptidyl peptidase 4 Dipeptidyl peptidase 4[Dipeptidyl peptidase 8[Dipeptidyl peptidase 2[Dipeptidyl peptidase 9	9 9	[14C]-Linagliptin [3H]-Linagliptin Linagliptin 1-[(4-methyl-quinazo... 8-[3(R)-aminopiperid... Trafalgar® Trafalgar®
	1133	2	5848501	TNF FRAP	7.2900 8.1550	[14C]-Rapamycin Rapamycin Sirolimus RAPAMTJNE® RAPAMUNE® temsirolimus NSC-226080

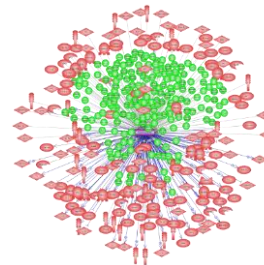


# From pathways to treatments

## Automated analysis to combine disease data with drug data

### Step 1

Find all targets that could be used to affect the disease state



- 88 targets related to hyperinsulinism with  $\geq 3$  literature references
- Full relationship information

Name	Effect	PMID	Selected_Sentences	Types_of_Memb ers	Number_of_ References
Regulation: ABCC8 ----> Hyperinsulinism		16075046 16075046 15356046 15356046 15292329 14715863 12414839	Mutations in ID{6833=SUR1 (ABCC8)} are the most common cause of ID{90000000,9010758=hyperinsulinism} of infancy, accounting for almost 50% of cases (6, 34). For ID{6833=SUR1}, impaired binding/hydrolysis and/or transduction of Mg nucleotides will decrease ID{12115272=KATP channel} activity, producing ID{90000000,9010758=hyperinsulinism} of infancy.	Protein -> Disease	47
Regulation: GCK ----> Hyperinsulinism		16075046 15919746 15356046 15356046 16574664 10488074 19717002	As described above, mutations in ID{0,2645,2746=GCK and GLUD1} cause ID{90000000,9010758=hyperinsulinism} of infancy, probably by enhancing metabolic ID{1190010=ATP} generation and decreasing ID{12115272=KATP channel} activity. CONTEXT{10003409} In addition, ID{2645=glucokinase} mutations that increase enzyme activity cause ID{90000000,9010389,9010758=hyperinsulinism}.	Protein -> Disease	46
Regulation: INSR ----> Hyperinsulinism		19770178 19752219 17018838 16644916 16214940 16214940 10194485	However, ID{3643=insulin receptor} inactivation in the ID{10000000,8803116=liver} also resulted in ID{90000000,9010758=hyperinsulinism}. ID{10000000,8803116=hepatic} ID{90000000,9012899=insulin resistance}, and peripheral ID{90000000,9012899=insulin resistance} (Michael et al. 2000, Baudry et al. 2002, Mauvais-Jarvis et al. 2002, Fisher & Kahn 2003).	Protein -> Disease	45



# From pathways to treatments

Automated analysis to combine disease data with drug data

## Step 1

Find all targets that could be used to affect the disease state

## Step 2

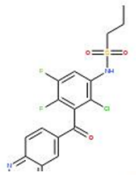
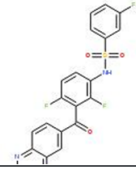
Query for each protein to find compounds that target it (>6 log units)

## Step 3

Collate data by compound to summarize the targets/activities related to disease that the compound hits  
Compute geometric mean of activities for ranking  
Rank by number of targets and geometric mean of activities against targets

Mean of activities among these targets

- All compounds that were observed to bind to targets in pathway
- Sorted by number of active targets.
- Too many targets may suggest lack of specificity.

Molecule	N	Number in Group	Reaxys Registry Number (DEVID)	Target	(DAT. Mean)	Geometric Mean	InChI Key (DE.INCHI)	H Bond Donors (CALC.HDONOR)	TPSA (CALC.TPSA)	H Bond Acceptors (CALC.HACCOR)	Rotatable Bonds (CALC.ROT BND)	Veber Number (CALC.VEBER)	Lipinski Number (CALC.LIPINSKI)	LogP (CALC.LOG P)
	1	5	23313886	JNK1a2 ROCK1 GDF1R FAK IKKb	8.3000 8.3000 8.3000 8.3000 8.3000	8.3	BILNBPXSG GOVVR- UHFFFAOY SA-N	3	129.66	8	9	2	4	3.081
	2	5	23313958	JNK1a2 ROCK1 GDF1R FAK IKKb	8.3000 8.3000 8.3000 8.3000 8.3000	8.3	SSMHMALO OCHCHK- UHFFFAOY SA-N	1	115.86	7	9	2	3	4.145

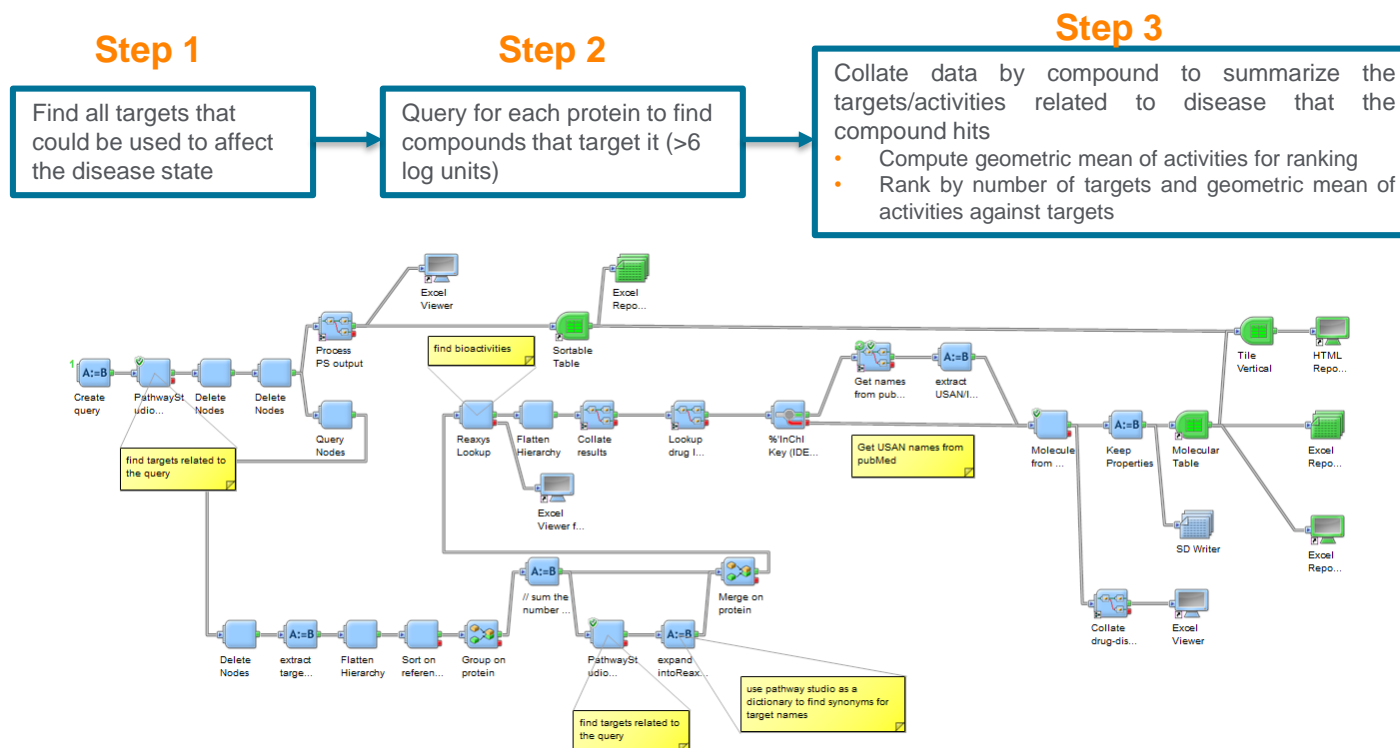
Drug-likeness metrics for sorting/classification

Targets and activities for each compound



## From pathways to treatments: PipelinePilot implementation combines data sources

## Automated analysis combines bioassay data with text-mined data





# From pathways to treatments: adjusting workflow

TARGETS

- Types of connections to a disease
- Place in a disease model
- Supporting evidence (good or bad)
- Adverse events
- Target class/localization
- Overall connectivity
- Biomarkers
- Signature-based
- Role in processes associated with a disease

DRUGS

- Drugs approved/passed safety
- Potential off-target activity
- Metabolism/transport
- Polypharmacology
- New drugs



## Summary

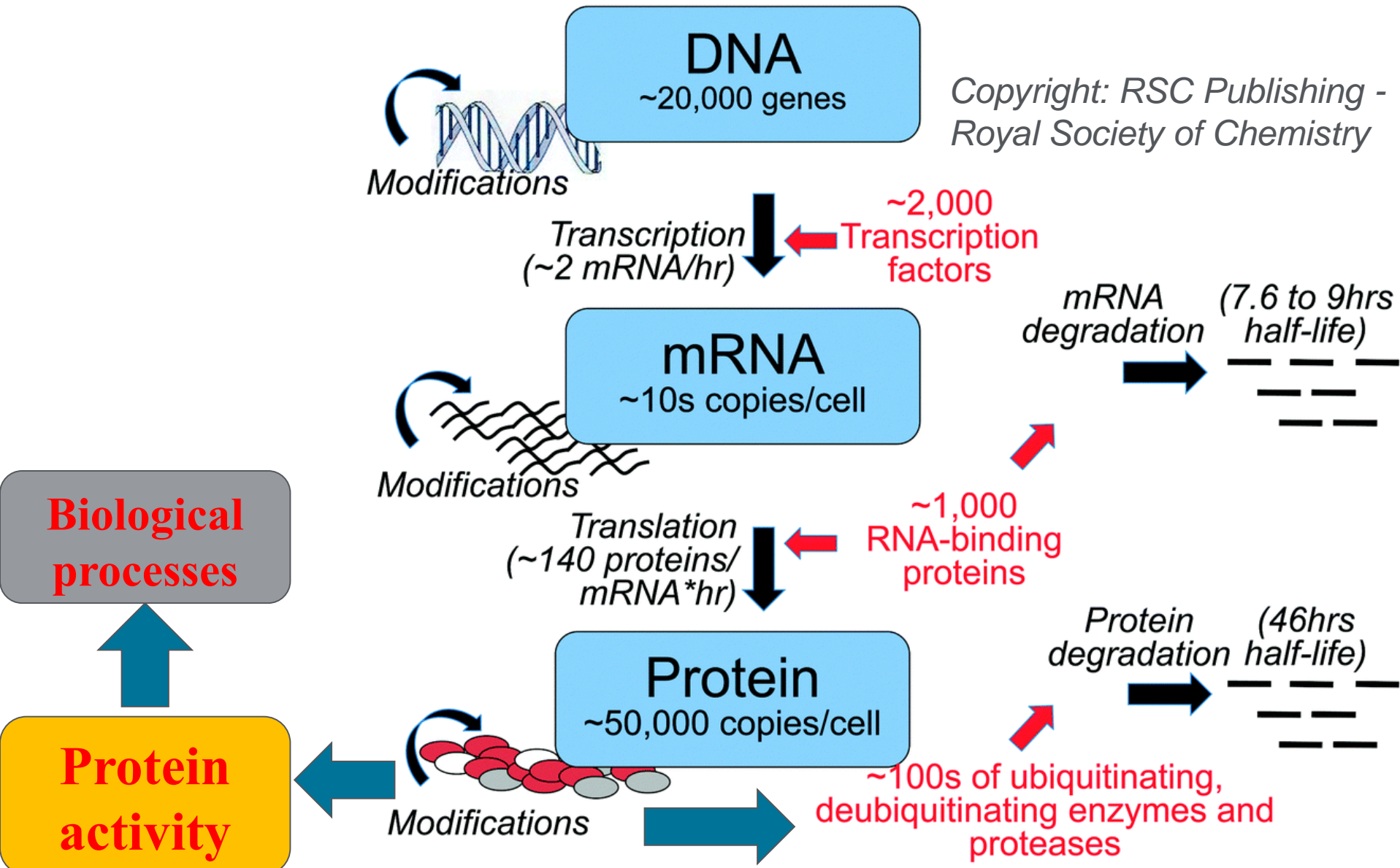
- Used extensive Elsevier's content, tools and capabilities to provide information about a rare disease:
  - Literature-extracted biology data to find targets and summarize what is known about the disease mechanism
  - Bioactivity data to find drugs that target those targets
  - Normalized names of authors and institution to find collaborators/research centers
- Once the output of interest is decided, **answer generation can be automated**:  
Provide a disease name and get:
  - List of targets with supporting information
  - Sorted list of approved drugs with supporting information
  - KOLs and institutes



# PATHWAY ANALYSIS FOR PERSONALIZED ONCOLOGY



# Central dogma 2017: How to measure protein activity?



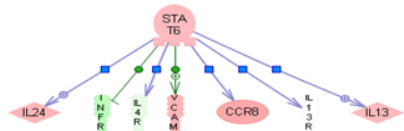
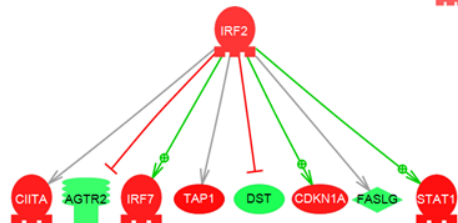
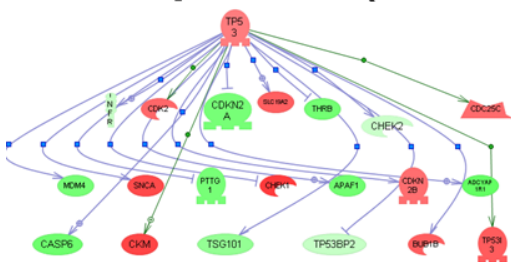


## SNEA: sub-network enrichment analysis

- Calculates protein activity from the observed changes of its downstream targets

SNEA	Reverse Causal Reasoning
Mann-Whitney enrichment test	Fisher's overlap test

Lower p-value (more significant)



- SNEA builds networks from all genes/proteins measured in the experiment using all relations in the database.
- SNEA can include indirect regulation i.e. expression regulatory cascades consisting of 2-3 steps
- Significant network centers may be found that are not measured in the primary dataset
- No prior curation of gene sets is required.
- Can work with partial information about TF targets. Does not require knowledge about all targets for TF
- P-value is sensitive to the size of the chip

Higher p-value (less significant)

Molecular networks in microarray analysis.

Sivachenko A, Yuryev A, Daraselia N, Mazo I. J Bioinform Comp. Biol.





[www.wakeforest-personalized-hemonc.com](http://www.wakeforest-personalized-hemonc.com)

11635 Northpark Drive, Suite 250, Wake Forest, NC 27587

*Gene expression profiling  
for targeted cancer treatment*

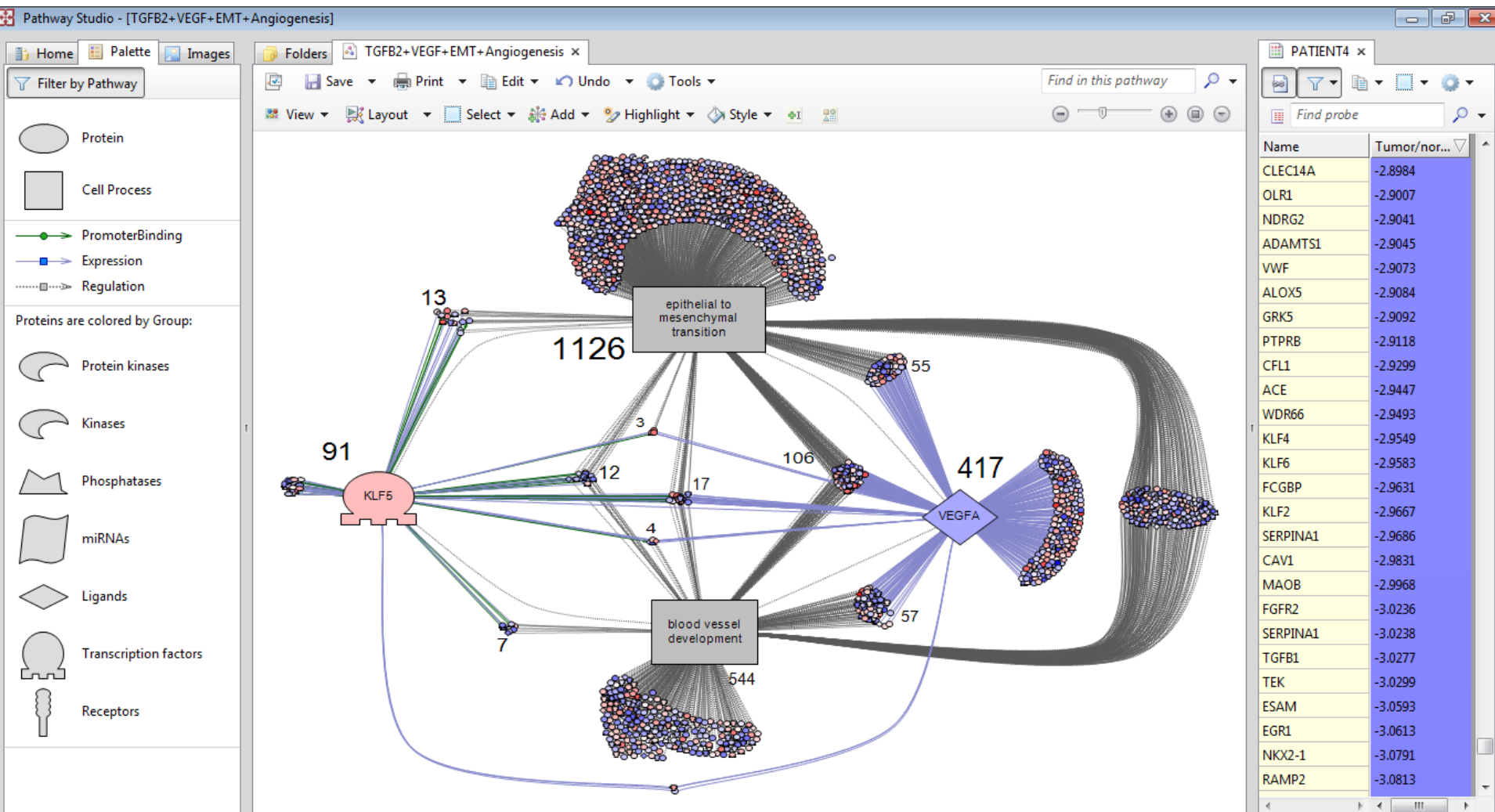
Luminita Castillos<sup>1</sup>, PhD, MBA, Francisco Castillos<sup>1</sup>, III, MD and Anton Yuryev<sup>2</sup>, PhD

<sup>1</sup>Personalized Hematology-Oncology of Wake Forest, PLLC, NC 27587, USA

<sup>2</sup>Elsevier, MD 20852, USA



# Example of expression regulators and Cell processes identified by SNEA in lung cancer patient

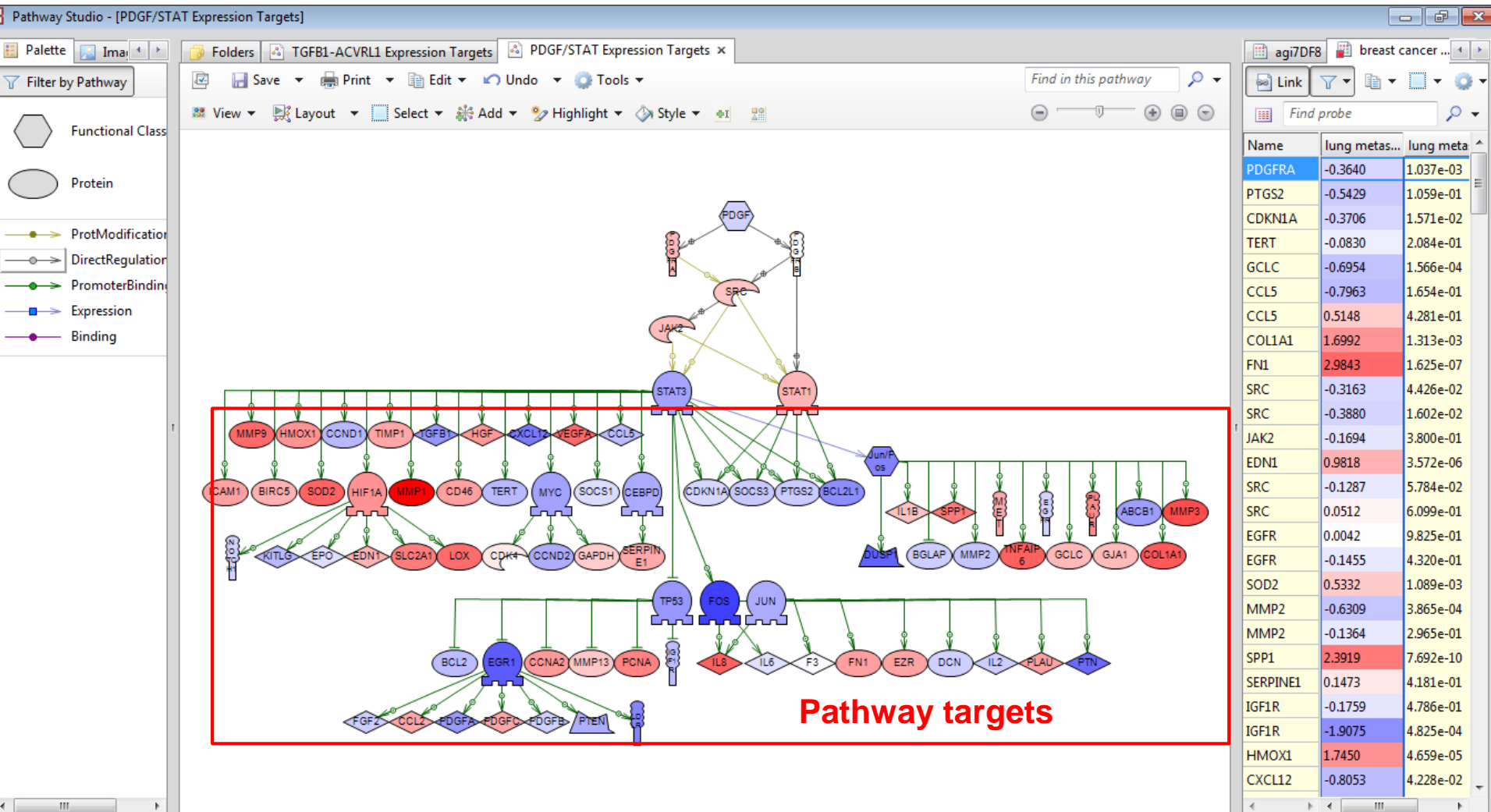




# Common misconception:

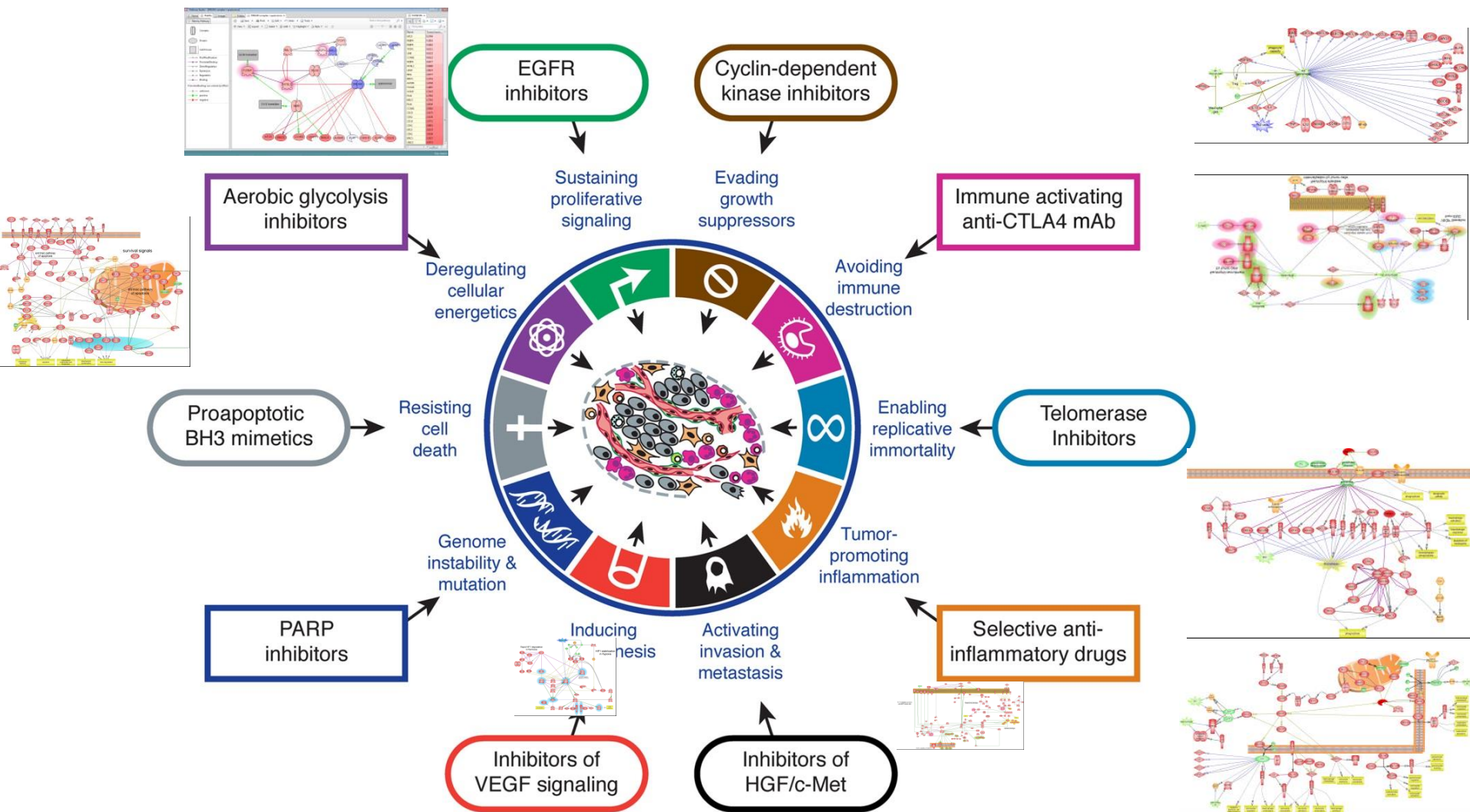
Pathway activity  $\neq$  Differential Expression of its components

Pathway activity  $\neq$  Differential Expression of its expression targets





# Pathway Activity signatures identify targets for anti-cancer drugs





## Major steps to calculate pathway activity signature

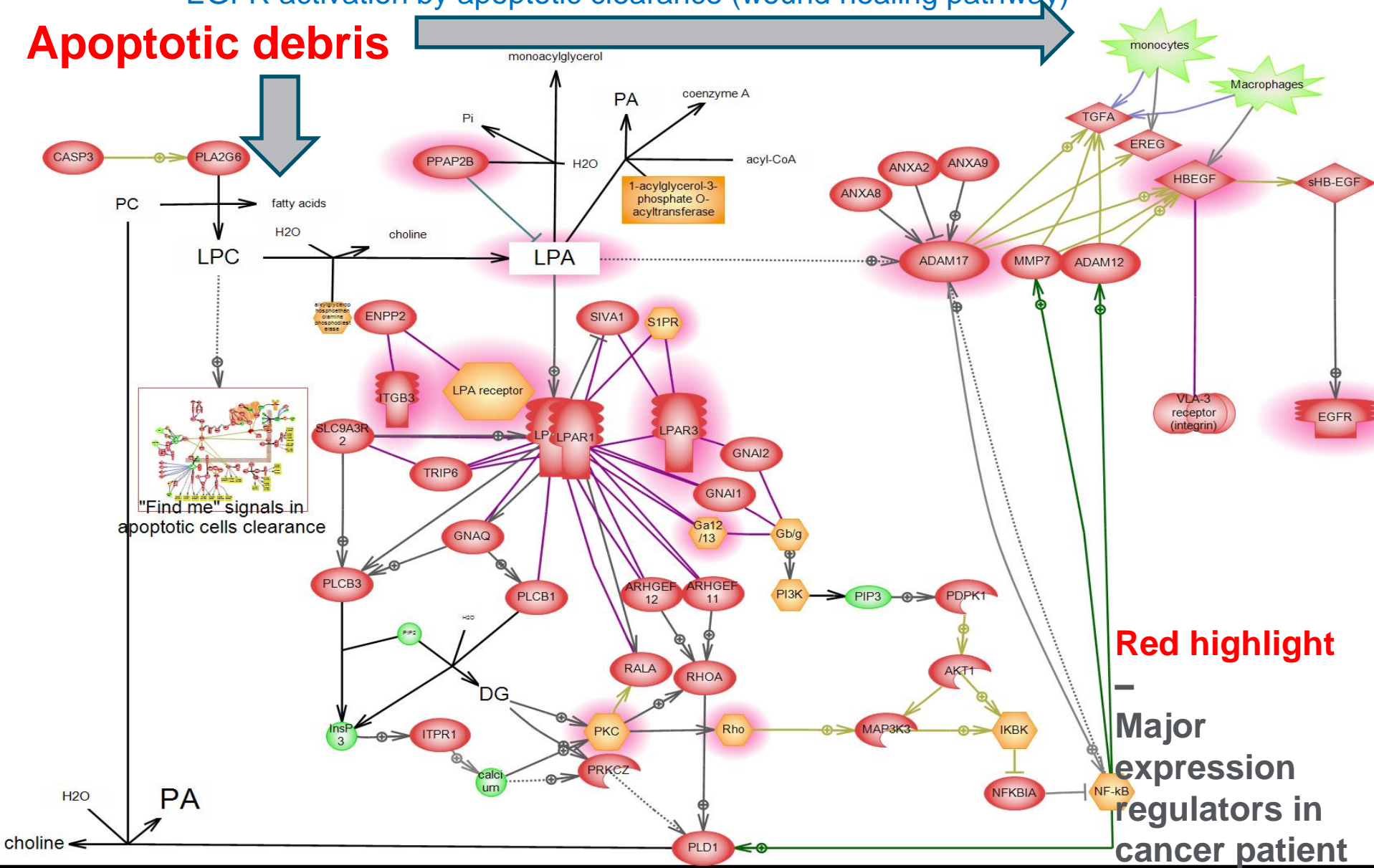
1. Calculates major expression regulators from the expression of their targets
2. Maps major expression regulators on cancer pathway collection
3. Calculate pathway activity signature
  1. Pathway activity signatures are short and therefore can classify patients better
  2. Pathway activity allow selection of drugs inhibiting the active pathway(s) instead of inhibiting single target



# Cancer pathways: Insights to cancer biology

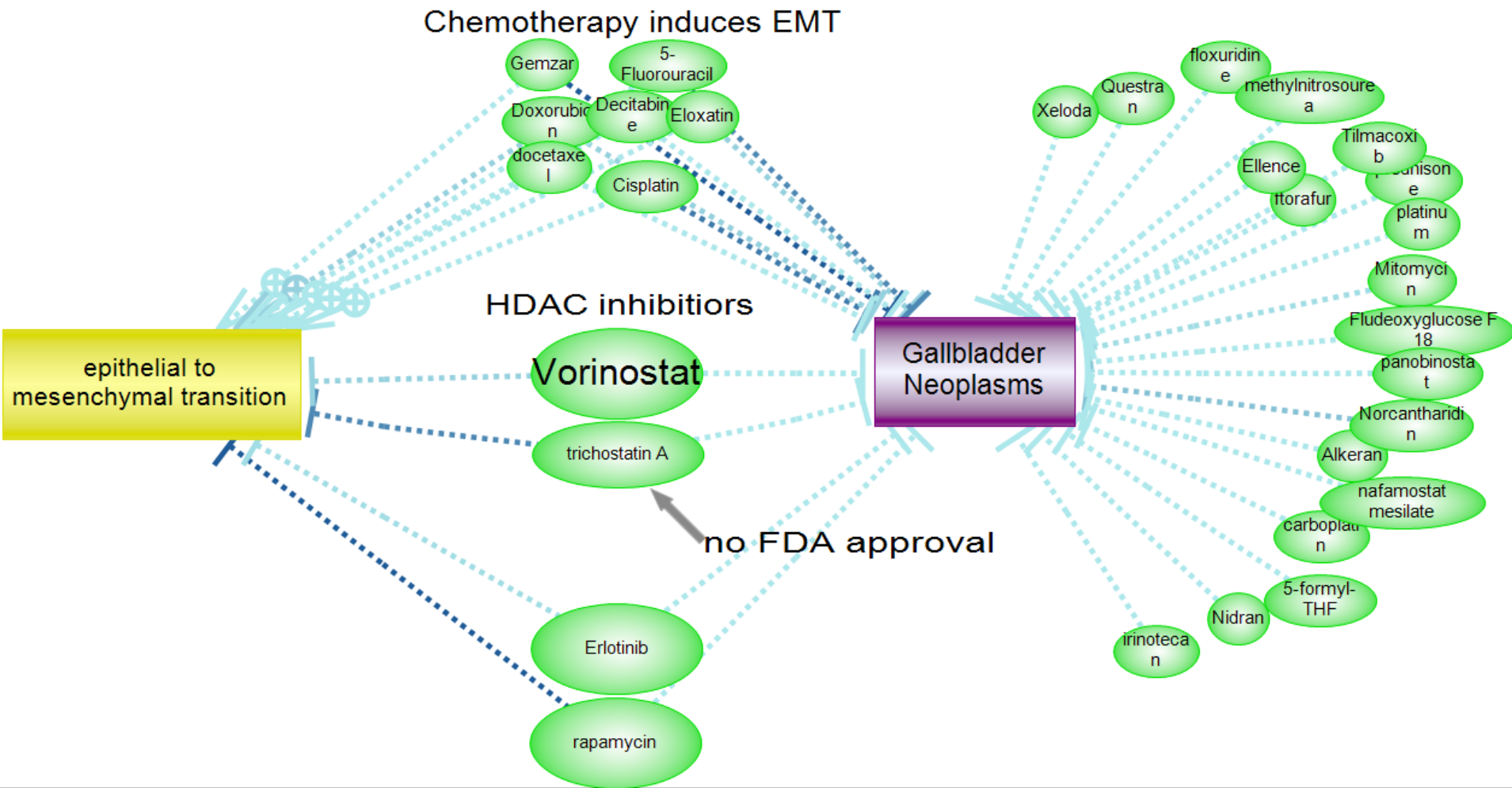
EGFR activation by apoptotic clearance (wound healing pathway)

**Apoptotic debris**





# How to select anti-cancer drugs in Pathway Studio



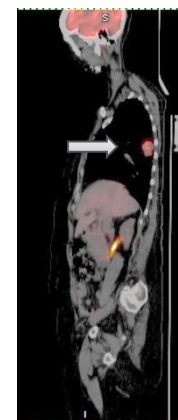


ELSEVIER

# Precision Oncology



Lung met before treatment      No lung met after treatment



ELSEVIER



Biopsy

Transcriptomics

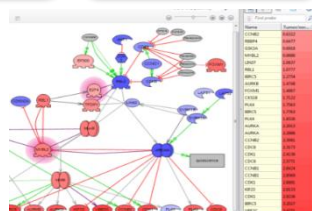
Data analysis

Targets

Drugs



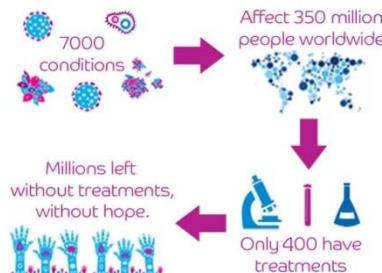
AATCCA  
TATGTC  
GGTATCA  
CAGG ...



# Rare diseases

ELSEVIER

findacure



Patient Empowerment

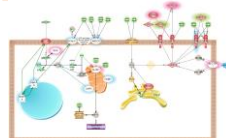
Disease mechanism

Targets

Drugs

Experts Network

Publications



findacure  
Congenital hyperinsulins



R&D Solutions

**Thank you**

