Maximizing Search Results in Bio-pharmaceutical Space

Srinivasan Parthiban

Parthys
Reverse Informatics
The Art of Generating New Knowledge

ICIC 2011 23-26 October 2011, World Trade Center, Barcelona, Spain

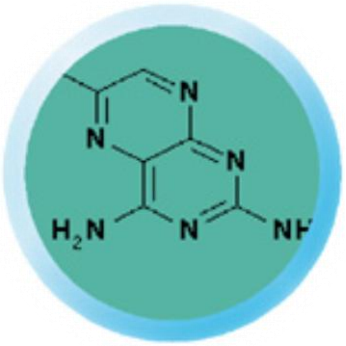
Parthys Reverse Informatics
IIT Madras Reserch Park – 8C
Kanagam Road, Taramani
Chennai 600 113, Tamil Nadu, India
T: +91-98409-75643, F: +91-44-4261-7070

parthi@reverseinformatics.com

26 October 2011

The Two Pillars in Drug Discovery

Bioassays



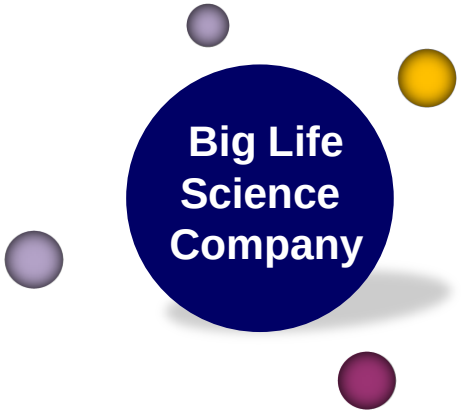
Biological Target

Chemical Compound

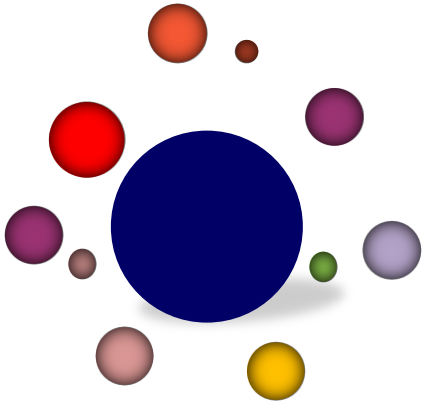
Parthys Reverse Informatics
The Art of Generating New Knowledge

The Changing Face of Life Science Industries

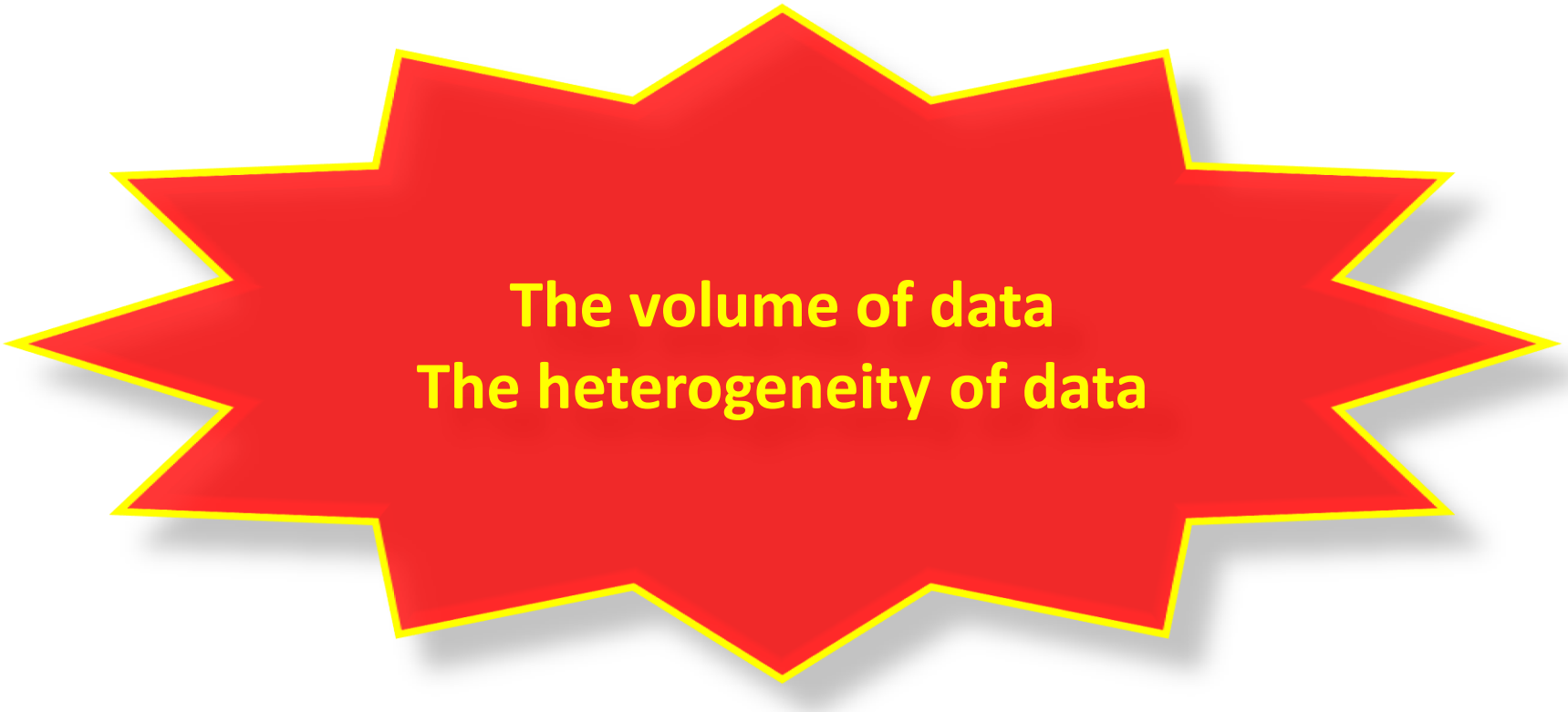
Yesterday



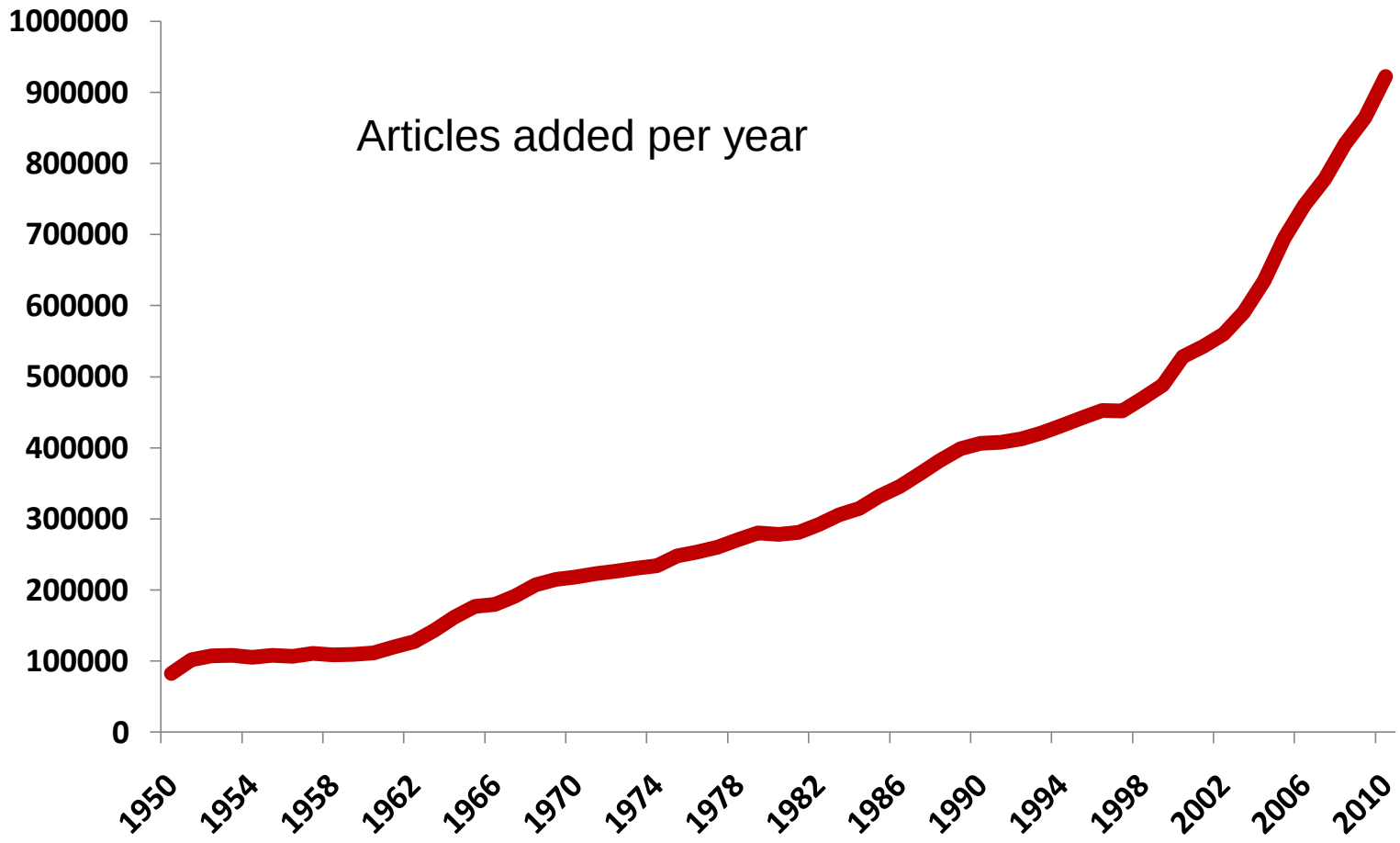
Today



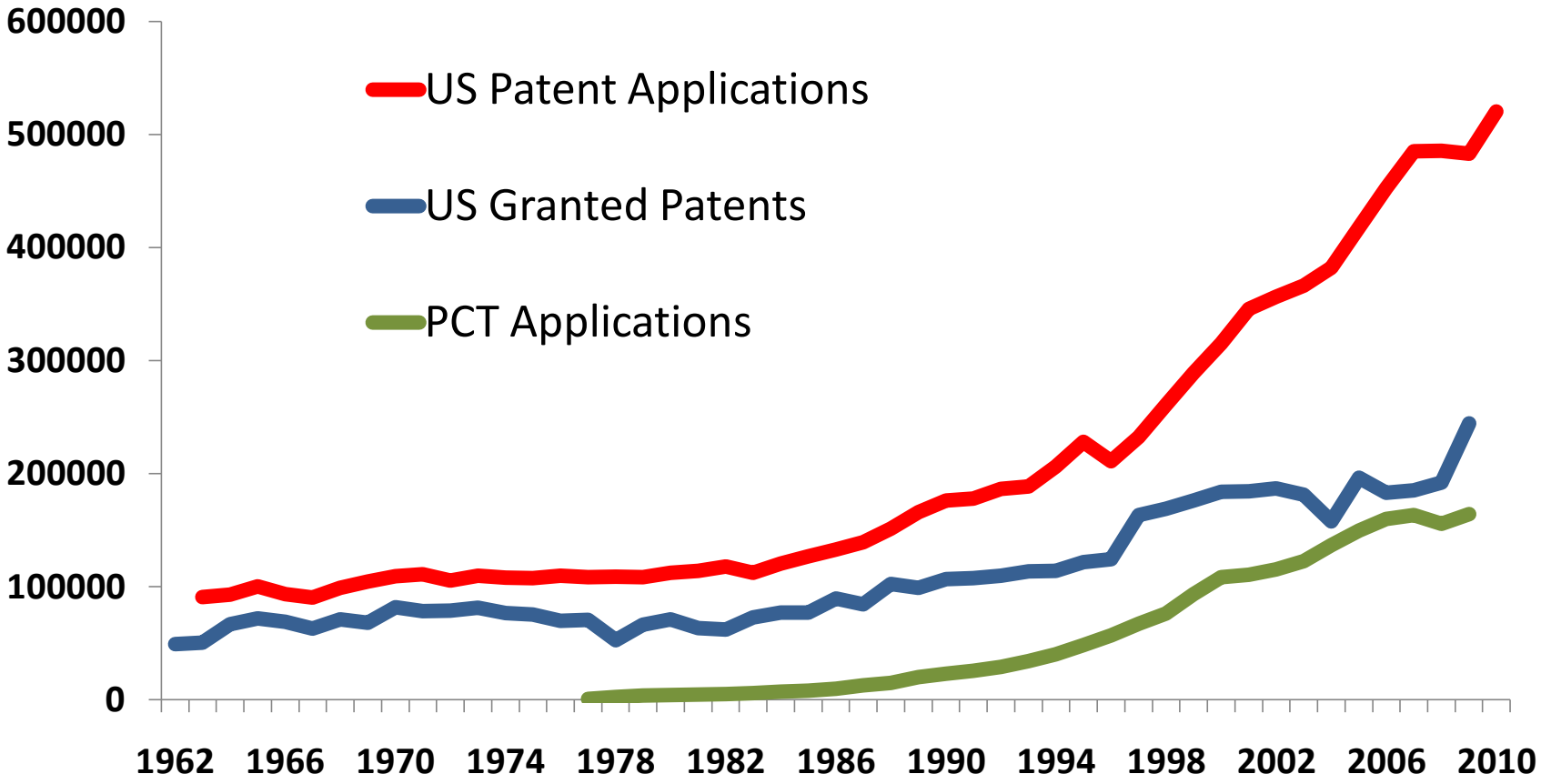
Why Industry Cares?



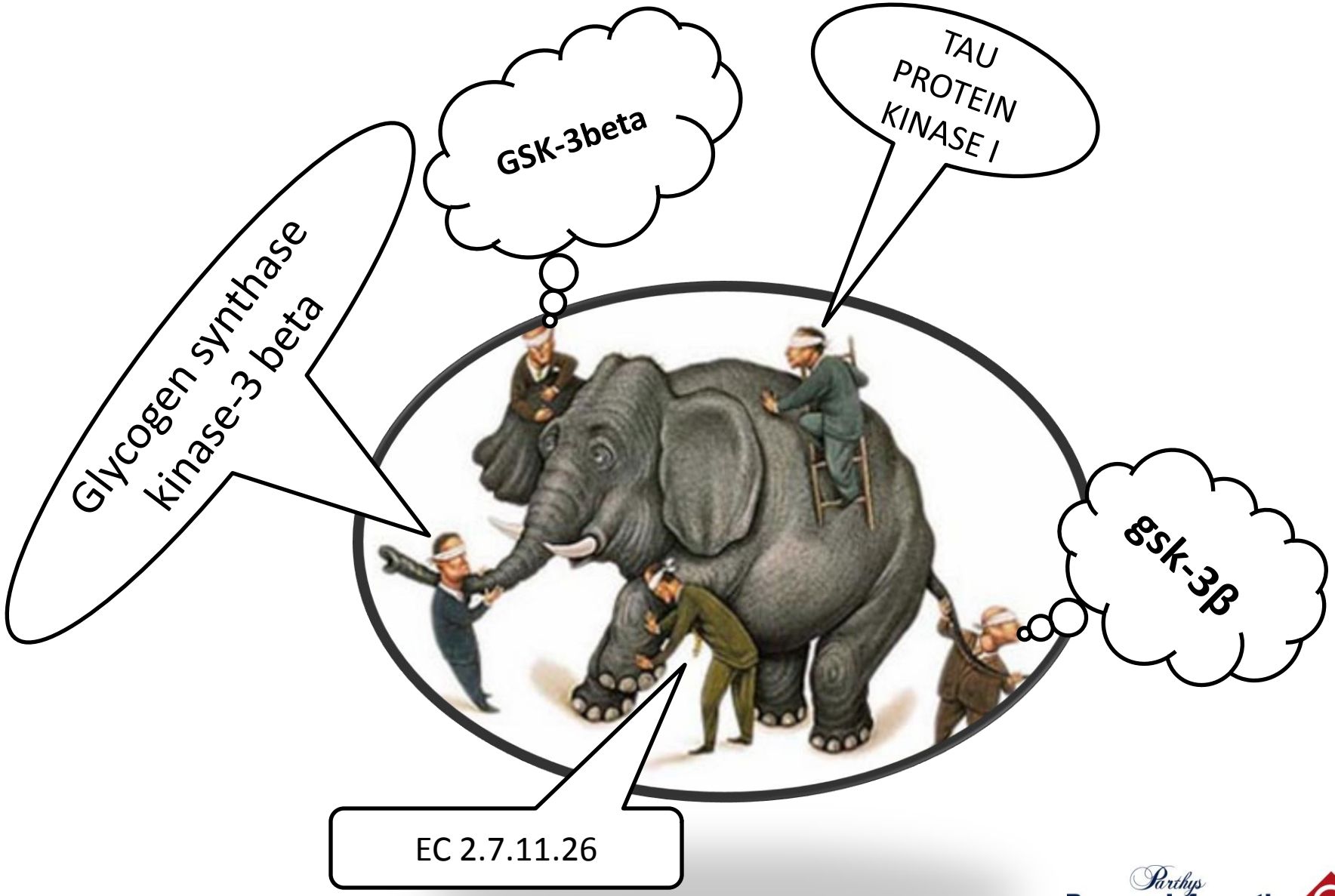
Growth of PubMed



Growth of Patents

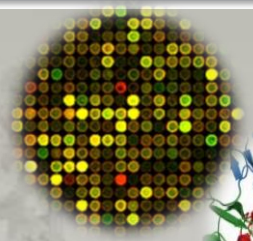


Many Biological Entities have Multiple Synonyms

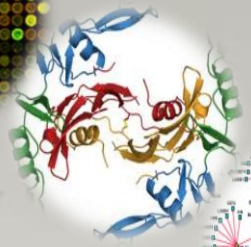


Parthys Reverse Informatics
The Art of Generating New Knowledge

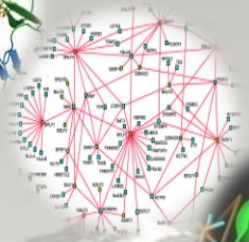
Understanding the Territory



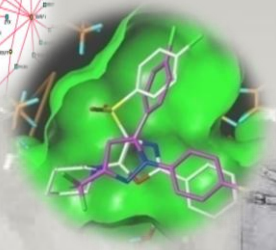
Gene Expression



Structural Biology



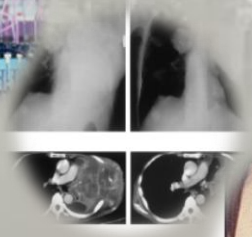
Protein Interaction Networks



Structure-Activity Relationships

Pharmacology

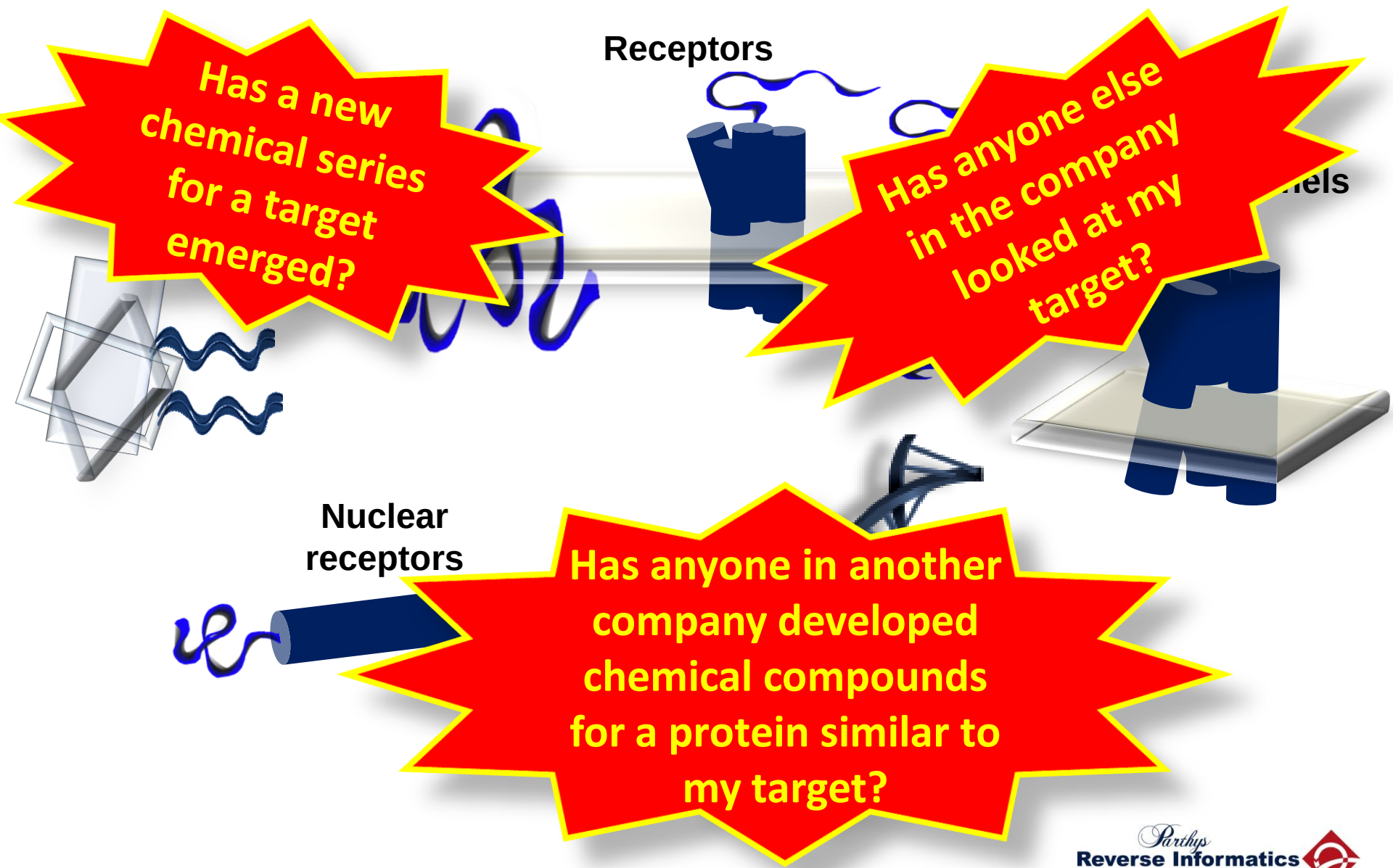
Toxicological Properties



Clinical Response

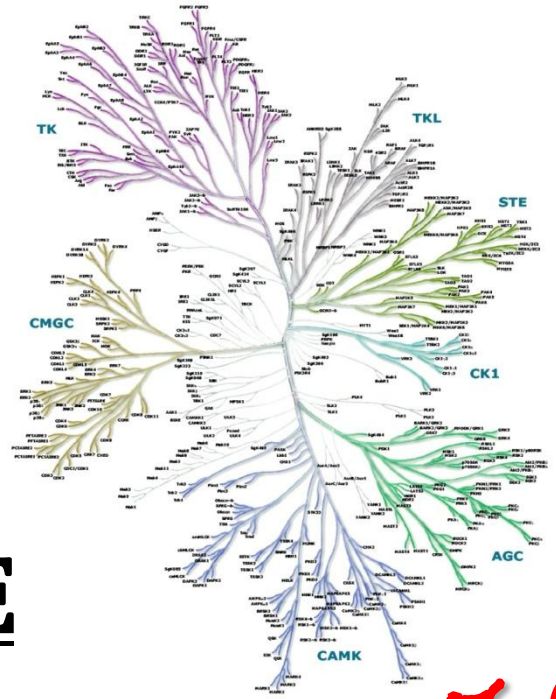
Genetic Variation

Drug Targets Literature Landscape



Parthys Reverse Informatics
The Art of Generating New Knowledge

Predominate Classes of Targets in Drug Discovery

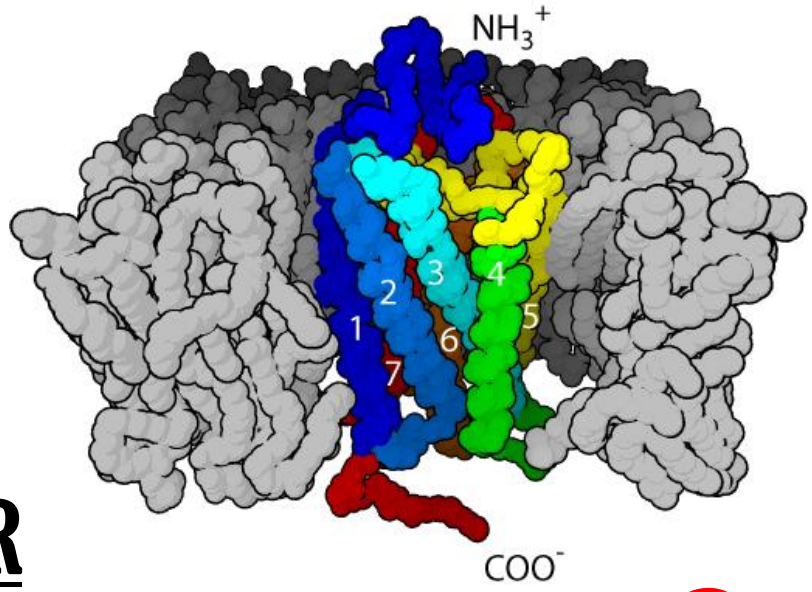


KINASE

>95% ☺

Single Word
NO textual variants

Synonyms: **Function**
Phosphotransferase



GPCR

<35% ☹

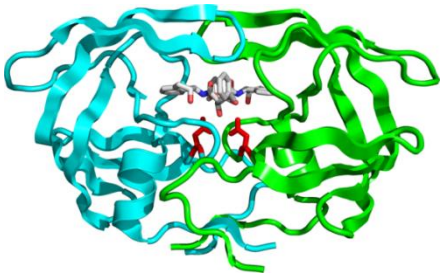
Abbreviation
Several Textual Variants

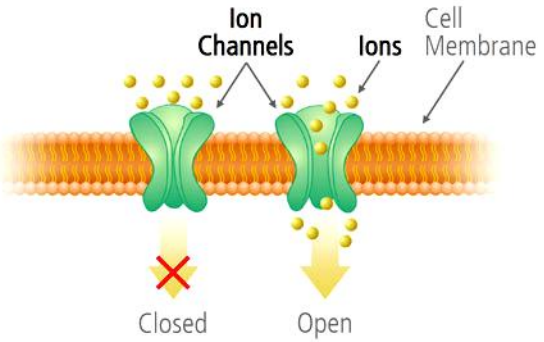
- *G protein coupled receptor*
- *G-protein-coupled receptor*

Synonyms: **Structure and Function**
Seven-transmembrane receptor (7TM)
Heptahelical membrane G-protein receptor

Parthys Reverse Informatics
The Art of Generating New Knowledge

Other Major Classes of Targets





> 80% ☺

> 70% 😐

> 65% 😐

Protease

Nuclear Receptor

Ion Channel

Single Word
NO textual variants

Two words
Some textual variants

Synonyms: **Function**

Peptidase/Proteinase

Intracellular membrane receptor

Ionotropic receptor

Parthys Reverse Informatics
The Art of Generating New Knowledge

Challenges in Patent Retrieval

Ambiguous Titles

Multiple Synonyms

Textual Variants

US7939554

FGR (e.g., FGF1R, FGF2R), FLT (e.g., FLT-1, FLT-2, FLT-3, FLT-4), FRK, FYN, GSK (e.g., GSK1, GSK2, GSK3-alpha, GSK3-beta, GSK4, GSK5), G-protein coupled receptor

US7923041

tau phosphorylation including cyclin-dependent kinase 5 glycogen synthase kinase-3b ("GSK-3b") and MAP kinases. Tau dephosphorylation appears to be primarily

US7897607

trimethyl-3,4-dihydro-2H-1,4-benzoxazin-8-yl)-6-propyl-2, [illegible] glycogen synthase kinase-3 beta (GSK-3β) [illegible] WO02/22074 and WO01/12607 disclose 3-aryl-4-quinolone

US7906281

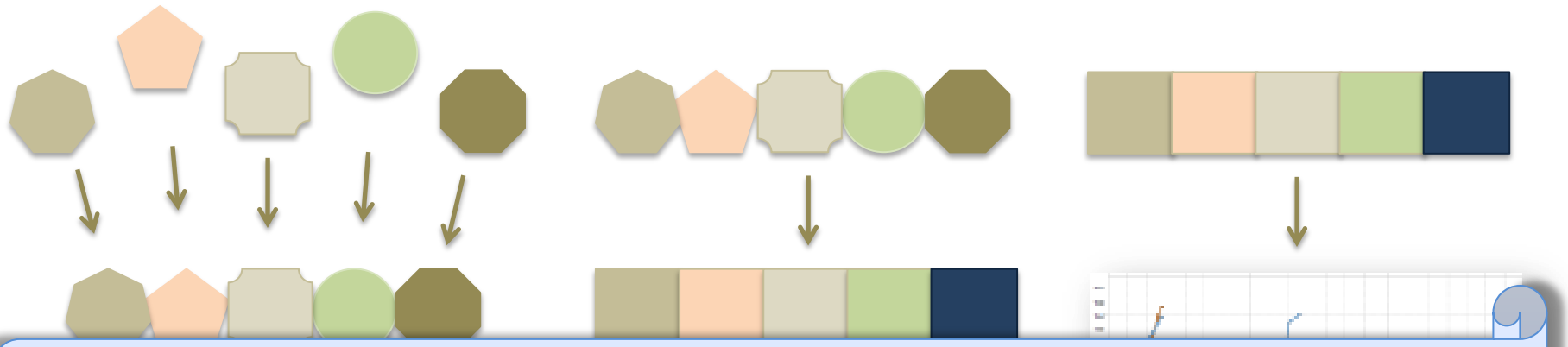
nucleotide positions in genes drawn from the group consisting essentially of ADRBK2, BNI, GSK3B, GRK3, IMPA1, IMPA2, INPP1, MARCKS, NTRK2 and/or NR1I2 within a

Way Forward

Identify Synonyms From various Sources

Standardize & Integrate the Data

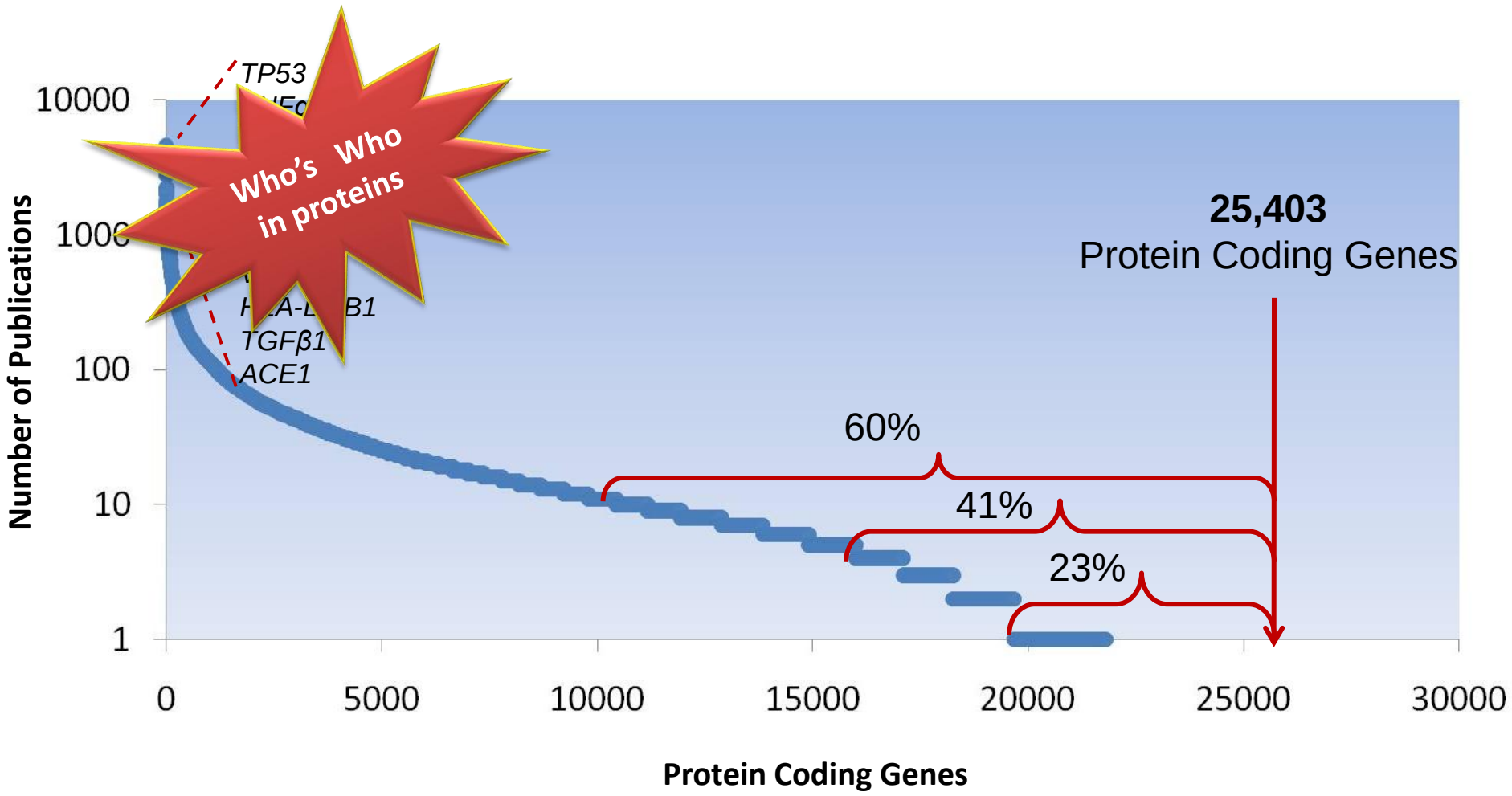
Search & Analyze



Our solution – Build a **unique synonym dictionary of bio-targets** and integrate it with a hierarchical classification to create Drug Target Ontology for patent retrieval

Parthys Reverse Informatics
The Art of Generating New Knowledge

Few Genes are well Studied



Parthys Reverse Informatics
The Art of Generating New Knowledge

Lost and Found!

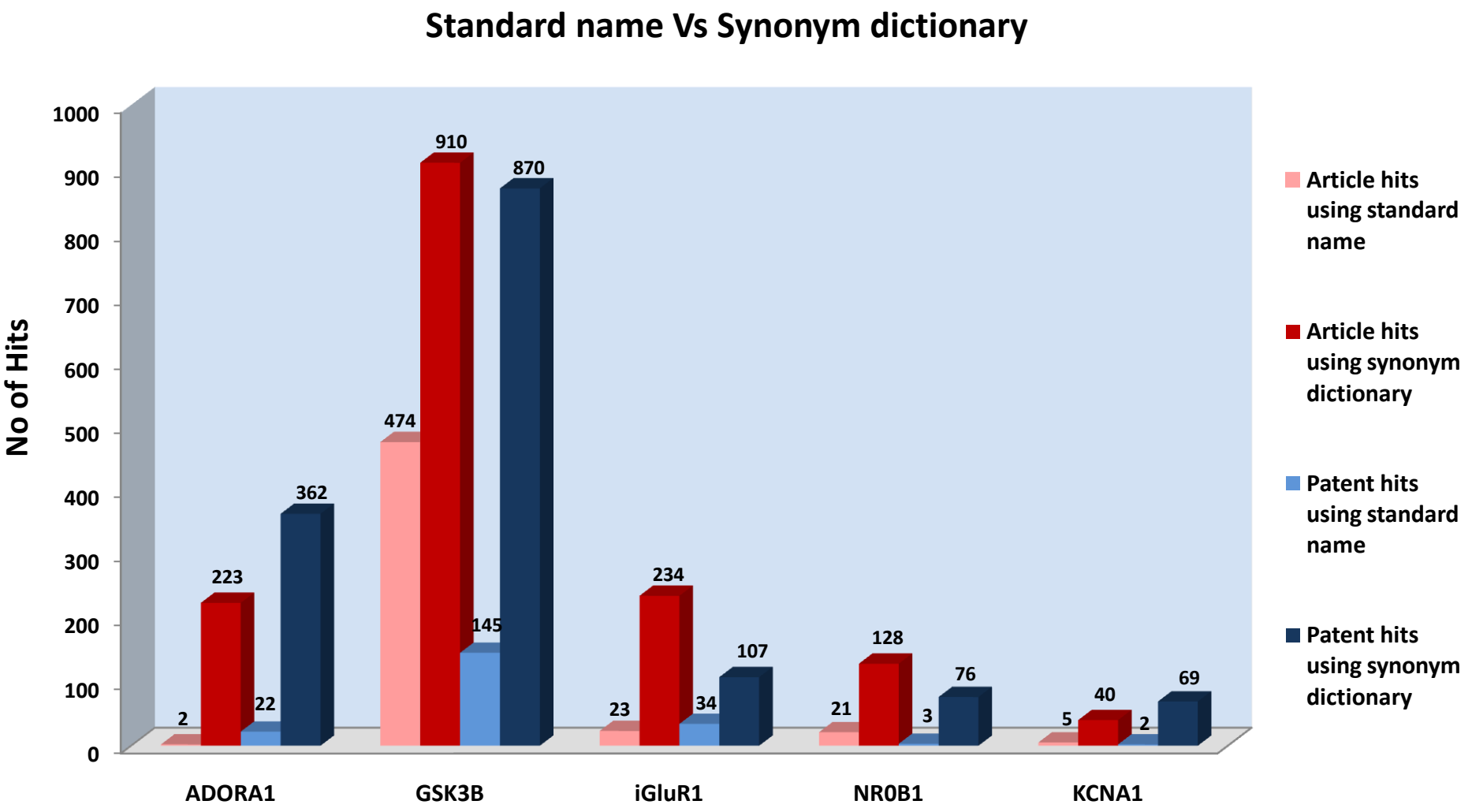
US7727235

TGF-beta 1

vast. Therapeutic agents which may be administered via the pharmaceutical compositions of the invention include growth factors, including bone morphogenic proteins (i.e. BMP's 1-7), bone morphogenic like proteins (i.e. GFD-5, GFD-7 and GFD-8), epidermal growth factor (EGF), fibroblast growth factor (i.e. FGF 1-9), platelet derived growth factor (PDGF), insulin like growth factor (IGF-I and IGF-II), transforming growth factors (i.e. TGF-beta I-III), vascular endothelial growth factor (VEGF); and other naturally derived or genetically engineered proteins, polysaccharides, glycoproteins, or lipoproteins.

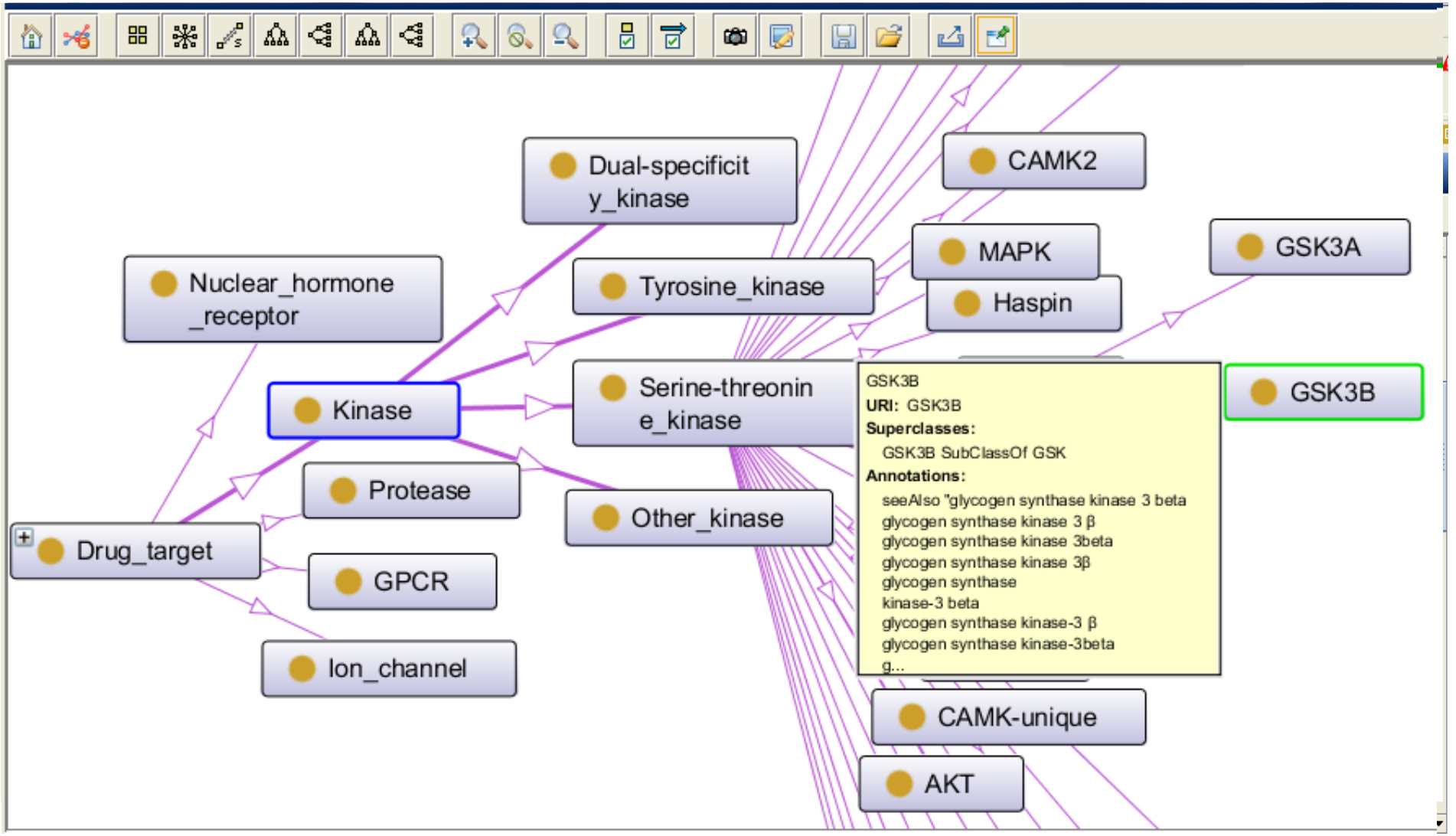
Assignee: Ethicon, Inc

Comparison of Drug Target Articles and Patents



Parthys Reverse Informatics
The Art of Generating New Knowledge

Screenshots of Drug Target Ontology



Parthys Reverse Informatics
The Art of Generating New Knowledge

Empowering the Research Community with Technology Alerts

End Use	Benefits of Technology Alert
Scientists & Researchers	Comprehensive list of patents in their area of interest
Patent Analysts	Exhaustive and accurate patent search results
CTO/HR	Head hunting the most active inventor in the field
CFO	Reduction in investment on redundant research
Competitive Intelligence	Research approach of competitors Most active competitor in the field of interest

Parthys Reverse Informatics
The Art of Generating New Knowledge

Better Results, Enable Success

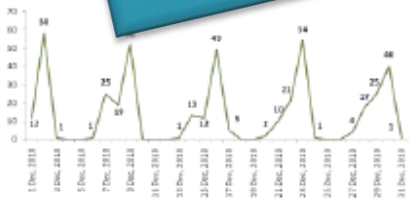
Automated Monthly Online Target Alert Report – December 2010

Geographical Distribution

The patent activity for the target EGFR in the month of December 2010 was analyzed. ***A total of 423 patents and applications were published in December*** While the patents and applications diverse countries, ***United*** with the

month o
AG. Othe
Mark Sli
significant

Publication

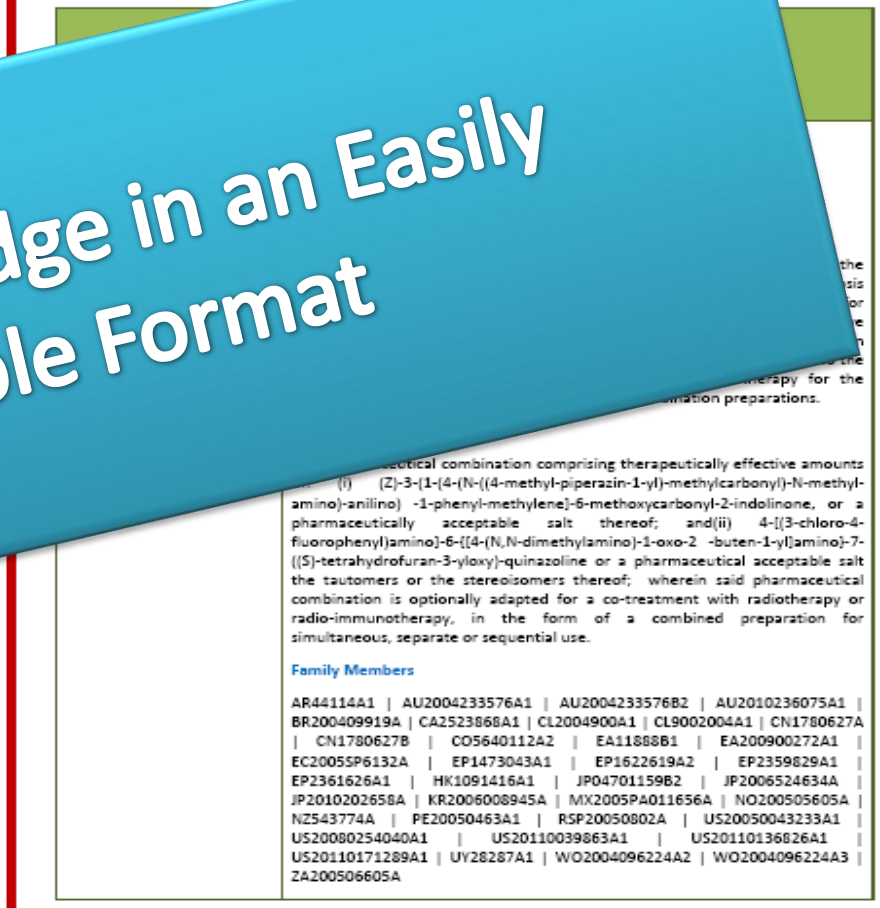


The publication of the issued patents and filed applications during the month December 2010 revealed an interesting fact that the publications peaked during the second half of every week i.e. Wednesday, Thursday and Friday. ***The most number of patents were published on 2nd December, 2010.***

Major Assignees

Genentech | NOVARTIS | AstraZeneca | UNIVERSITY of CALIFORNIA | gsk

Note: This is a sample report. The information contained in this report is only representative.



...therapy for the ...mation preparations.

...tical combination comprising therapeutically effective amounts ... (i) (Z)-3-(1-(4-(N-((4-methyl-piperazin-1-yl)-methylcarbonyl)-N-methyl-amino)-anilino) -1-phenyl-methylene]-6-methoxycarbonyl-2-indolinone, or a pharmaceutically acceptable salt thereof; and(ii) 4-[(3-chloro-4-fluorophenyl)amino]-6-{[4-(N,N-dimethylamino)-1-oxo-2 -buten-1-yl]amino}-7-((S)-tetrahydrofuran-3-yloxy)-quinazoline or a pharmaceutical acceptable salt the tautomers or the stereoisomers thereof; wherein said pharmaceutical combination is optionally adapted for a co-treatment with radiotherapy or radio-immunotherapy, in the form of a combined preparation for simultaneous, separate or sequential use.

Family Members

AR44114A1 | AU2004233576A1 | AU2004233576B2 | AU2010236075A1 | BR2004009919A | CA2523868A1 | CL2004900A1 | CL9002004A1 | CN1780627A | CN1780627B | CO5640112A2 | EA11888B1 | EA200900272A1 | EC2005SP6132A | EP1473043A1 | EP1622619A2 | EP2359829A1 | EP2361626A1 | HK1091416A1 | JP04701159B2 | JP2006524634A | JP2010202658A | KR2006008945A | MX2005PA011656A | NO2005056O5A | NZ543774A | PE20050463A1 | RSP20050802A | US20050043233A1 | US20080254040A1 | US20110039863A1 | US20110136826A1 | US2011017128 9A1 | UY28287A1 | WO2004096224A2 | WO2004096224A3 | ZA200506605A

Note: This is a sample report. The information contained in this report is only representative.

World's Knowledge in an Easily Computable Format

Parthys Reverse Informatics
The Art of Generating New Knowledge

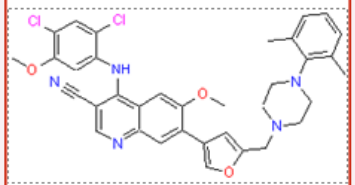
Kinase-Thematic Database

RIX Database

Inhibitor | Target | Scaffold | Assignee

Compound Information

Structure



Compound ID: XX 9999	Platform Name: KINASE
Target: src kinase	Compound Type: Inhibitor
Compound Name: 4-[(2,4-dichloro-5-methoxyphenyl)-amino]-7-(5-[[4-(2,6-dimethylphenyl) piperazin-1-yl]methyl]-3-furyl)-6-methoxyquinoline-3-carbonitrile	
Use: Anticancer	Disease: Chronic myeloid leukemia

Bibliographic Information

Journal/Patent: Patent Application	Reference ID: ZZ 000	Reference: US20060035930
Publication Date: 16-Feb-2006	First Author: Diane Harris Boschelli	Assignee: Wyeth, Patent Law Group

Title: 4-(2,4-dichloro-5-methoxyphenyl)amino-6-methoxy-7-[[5-substituted -amino)methyl]-3-furyl]-3-quinolinecarbonitriles as kinase inhibitors

Activity Information

Target Class	Target Name	Family	Subfamily
non-receptor Tyrkinase	src	src	DMPK

Bioassay ID	Bioassay	Assay Type	Test System	Species	Activity parameter	Unit	Prefix	Activity Value	Remarks
B 001	Src Kinase Assay	C	Recombinant human Src enzyme	Human	IC 50	nM			
B 005	Anchorage Independent Src-Transformed Fibroblast Proliferation Assay	C	Transformed Rat2 fibroblasts	Rat	IC 50	nM			
	HT29			Mice,					

Inhibitor | Target | Scaffold | Assignee

Summary | Phylogenetic Tree | Protein Interaction | Pathway | Orthologue Data

Kinase Information

Kinase ID: K 001	Kinase Name: Src(HS)	
Description: Proto-oncogene tyrosine-protein kinase Src	Synonyms:	
Organism: Homo sapiens	Uniprot Code: P12931	GENA ID: GHS018950

Classification

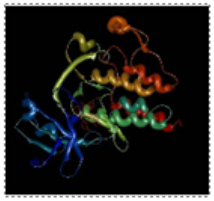
Group	Family	Sub Family
TK	src	DMPK

Family Description

The Src family, which is found in FYN (oncogene related to SRC), FGR, and YES), SRC (v-src avian sarcoma viral oncogene homologue), YES (v-yes-1 Yamaguchi sarcoma viral oncogene homologue), FGR (Gardner-Rasheed feline sarcoma viral oncogene homologue), LCK (lymphocyte-specific protein tyrosine kinase; required for antigen-activation of T-cells), HCK (hemopoietic cell kinase), and BLK (B lymphoid tyrosine kinase), has been implicated in many signal transduction pathways such as neuronal development and B-cell development. However, in C. elegans and Drosophila the Src family has only one sequence. Further, so far some reported pathways, including those for the Src family, have been different between Drosophila and mammals (Luttrell et al. 1997; Lu and Li 1999). The domain structure is conserved from worms to mammals, and the functional conservation, such as the structural plasticity of neurons, is expected.

Structure & Sequence Information

Structure



Sequence

GSNKSKPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPA SAUSHRGPSAAPAPAAAEPKLPGGHNSSLITVTSPGPKAGPLAG GVTTFVALYDYESRTETDLSFKKGERLQIVNNTEGDWWLAHSL STGQTGYIPSNYVAPSDSIQAEEWYFGKITRRESERLLLNAENP RGTFLVRESETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSGG FYITSRTQFNSLQQLVAYYSKHADGLCHRLTTVCPTSKPQTQG LAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTRVAIKT LKPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEPIYIVTE YMSKGSLLDFLKGETGKYLRLPQLVDMAAQIASGMAYVERMNY VHRDLRAANILVGENLVCKVADFGLARLIEDNEYTARQGAKFPI KWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNR EVLDQVERGYRMPCPPECPESLHDLMCQCWRKEPEERPTFE YLQAFLEDYFTSTEPQYQPGENL

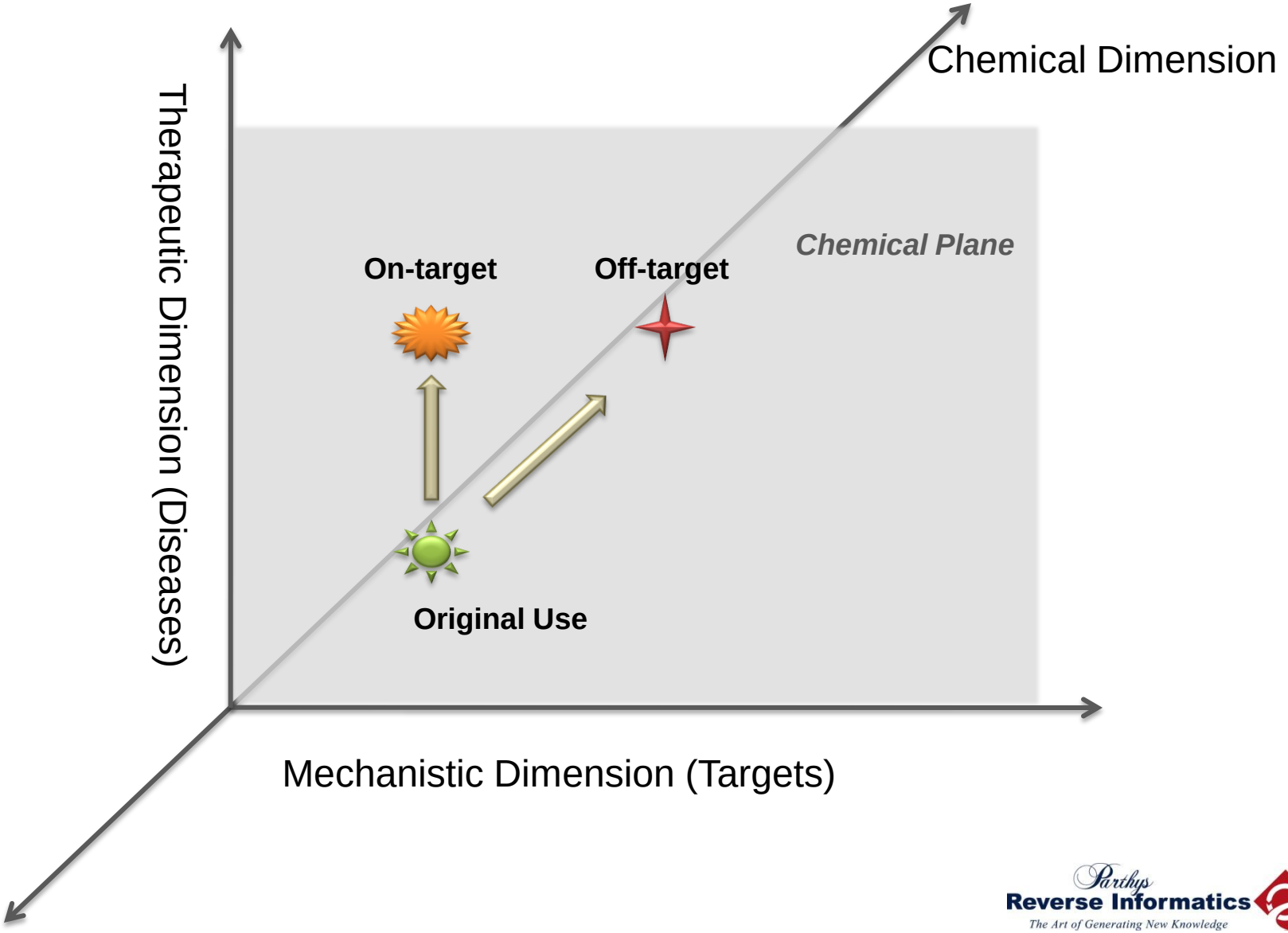
Compounds

S. No	Name	Type	Disease	References

Parthys Reverse Informatics

The Art of Generating New Knowledge

Repurposing Contextualized



Parthys Reverse Informatics
The Art of Generating New Knowledge

Summary

- **No two patent searches are ever the same**; many biological entities have multiple synonyms
- The information retrieval challenges in patents are discussed in terms of **ambiguous titles**, **multiple synonyms** and **textual variants**
- Our solution - Building a search platform to retrieve patents with **unique synonym dictionary** of bio-targets
- While **recall** improved with synonym dictionary, **precision** improved with relevant IPC code Filters
- Application of such comprehensive/complete retrieval of patents would enable us to provide **Technology Alert Services** to **empower the industry with better analytical capabilities for greater business impact**

Parthys Reverse Informatics
The Art of Generating New Knowledge

Thank you!

Parthys

Reverse Informatics

The Art of Generating New Knowledge