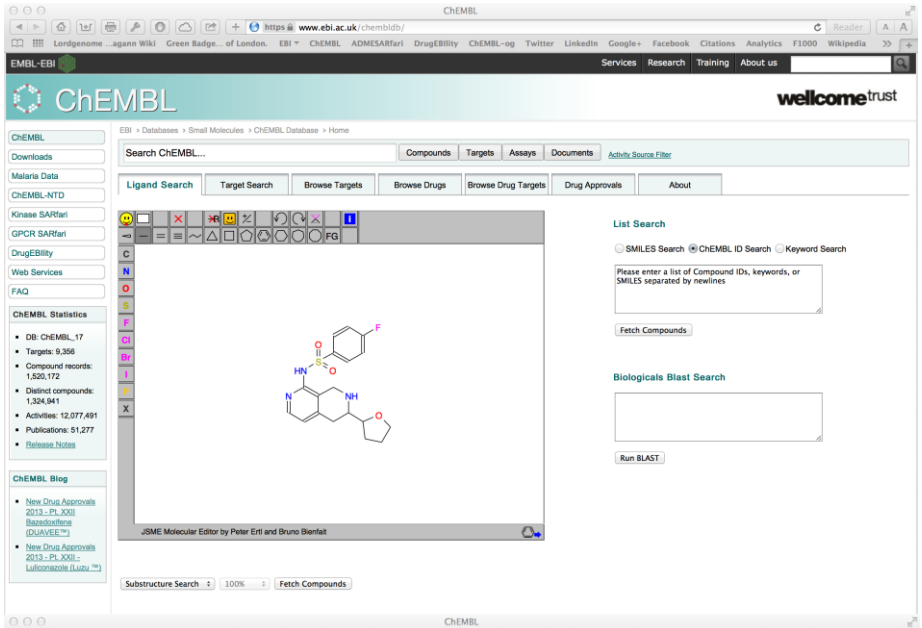
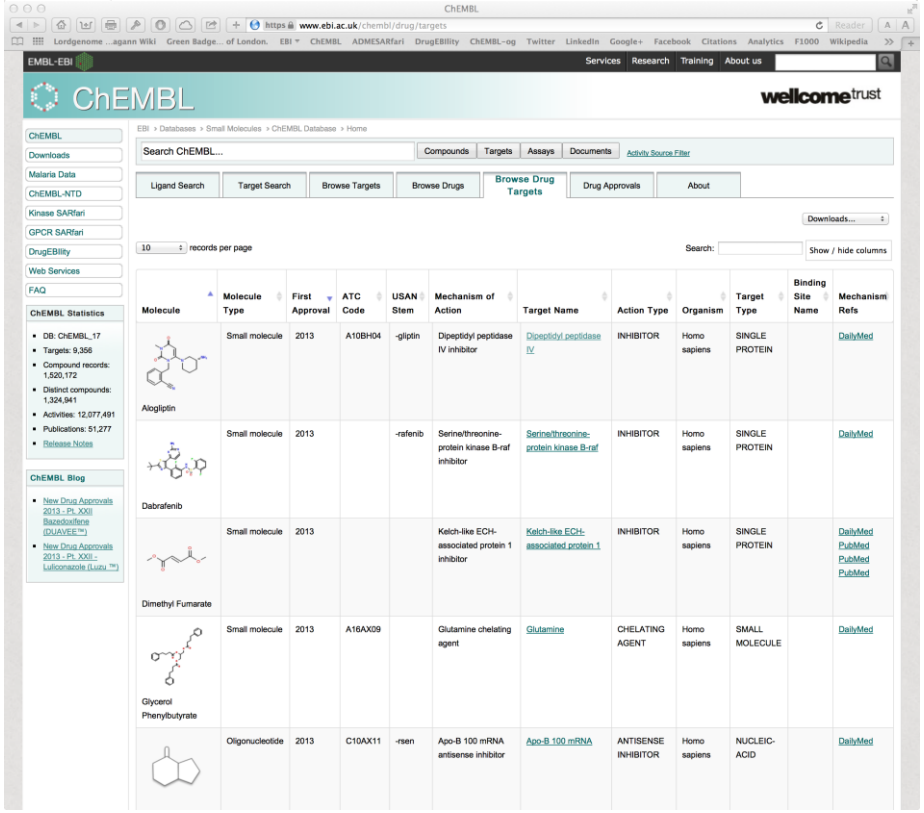
ChEMBL an Open Data Resource of Medicinal Chemistry and Patent Data

John P. Overington

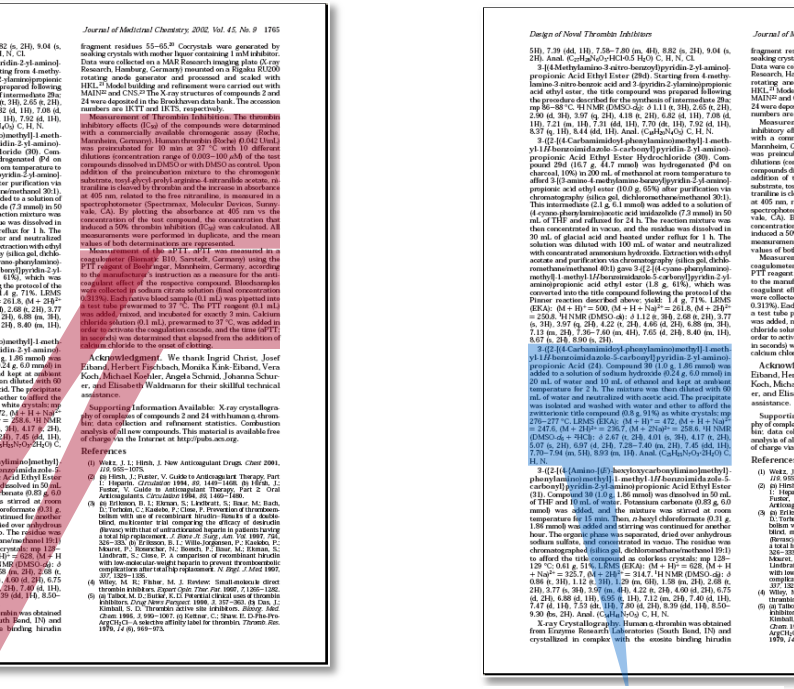
European Molecular Biology Laboratory -
European Bioinformatics Institute





ChEMBL

- The world's largest primary public database of medicinal chemistry data
  - ~1.4 million compounds, ~9,000 targets, ~12 million bioactivities
- Truly Open Data - CC-BY-SA license
- Many download/access formats
  - Semantic Web
    - RDF download, SPARQL endpoint at http://rdf.ebi.ac.uk/chembl
  - ChEMBL Applicances
    - myChEMBL – linux VM
    - ChEMpi – raspberry pi
- ChEMBL 18 released next week



>Thrombin

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRANTFLEEVRKGNLERECVEETCSY
EEAFEALESSTATDVFWAKYTACETARTPRDKLAACLEGNCAEGLGTNYRGHVNITRSGIECQLWRS
RYPHKPEINSTTHPGADLQENFCRNPDSSTTGPWCYTTDPTVRRQECSIPVCGQDQVTVAMTPRSEG
SSVNLSPPLEQCVPDRGQQYQGRLAVTTHGLPCLAWASAQAKALSKHQDFNSAVQLVENFCRNPDGD
EEGVWCYVAGKPGDFGYCDLNYCEEAVEEETGDGLDEDSDRAIEGRTATSEYQTFFNPRTFGSGEAD
CGLRPLFEKKSLEDKTERELLESYIDGRIVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRWVL
TAAHCLLYPPWDKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDIALMKLK
KPVAFSDYIHPVCLPDRETAASLLQAGYKGRVTGWGNLKETWTANVGKGQPSVLQVVNLPIVERPVC
KDSTRIRITDNMFCAGYKPDEGKRGDACEGDSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFY
THVFRLKKWIQKVIDQFGE

Inhibition of human Thrombin

PTT (partial thromboplastin time)

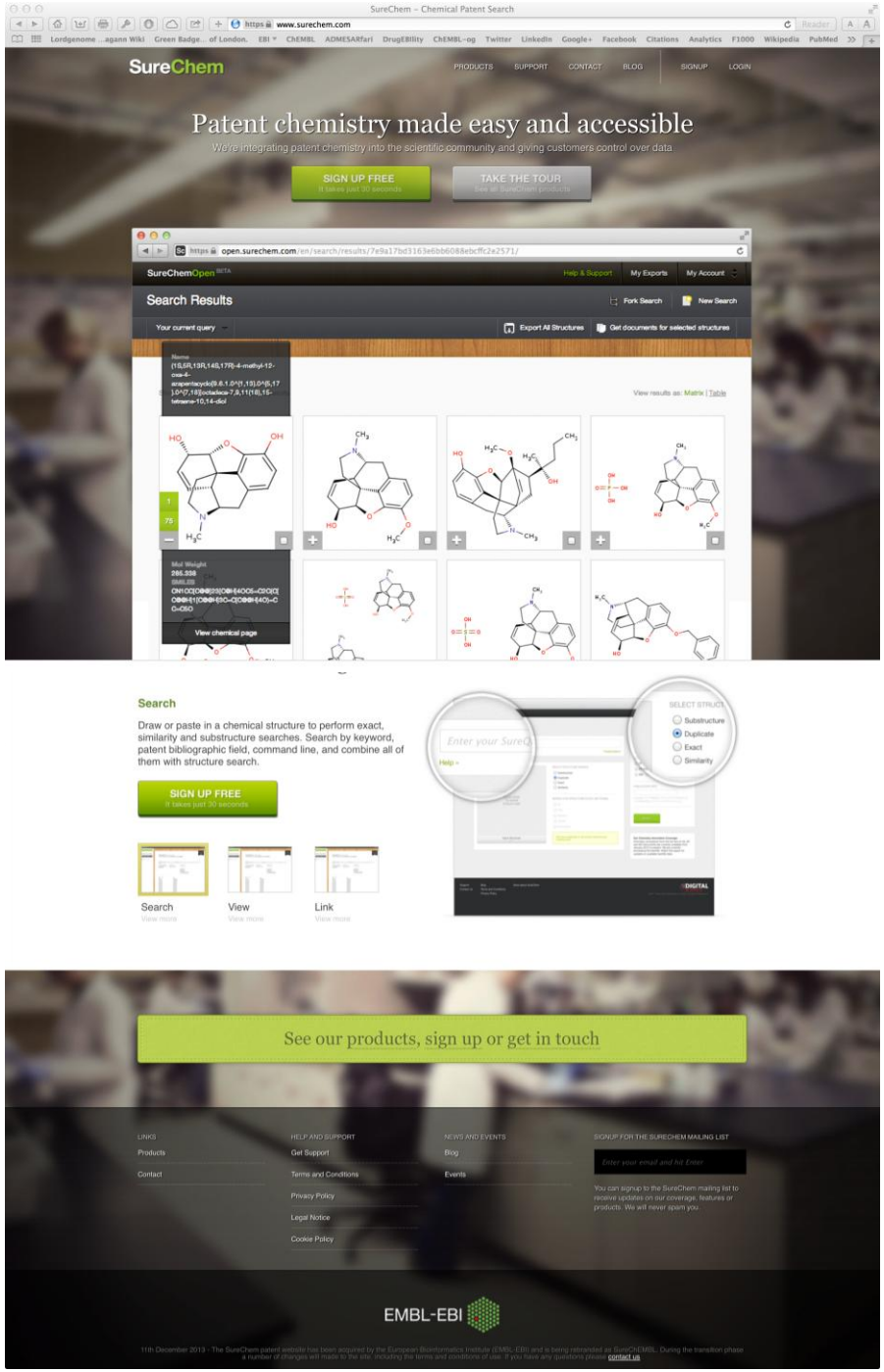
Assay

Compound

SAR Data

K_i=4.5 nM

ED_2=230 nM



SureChEMBL

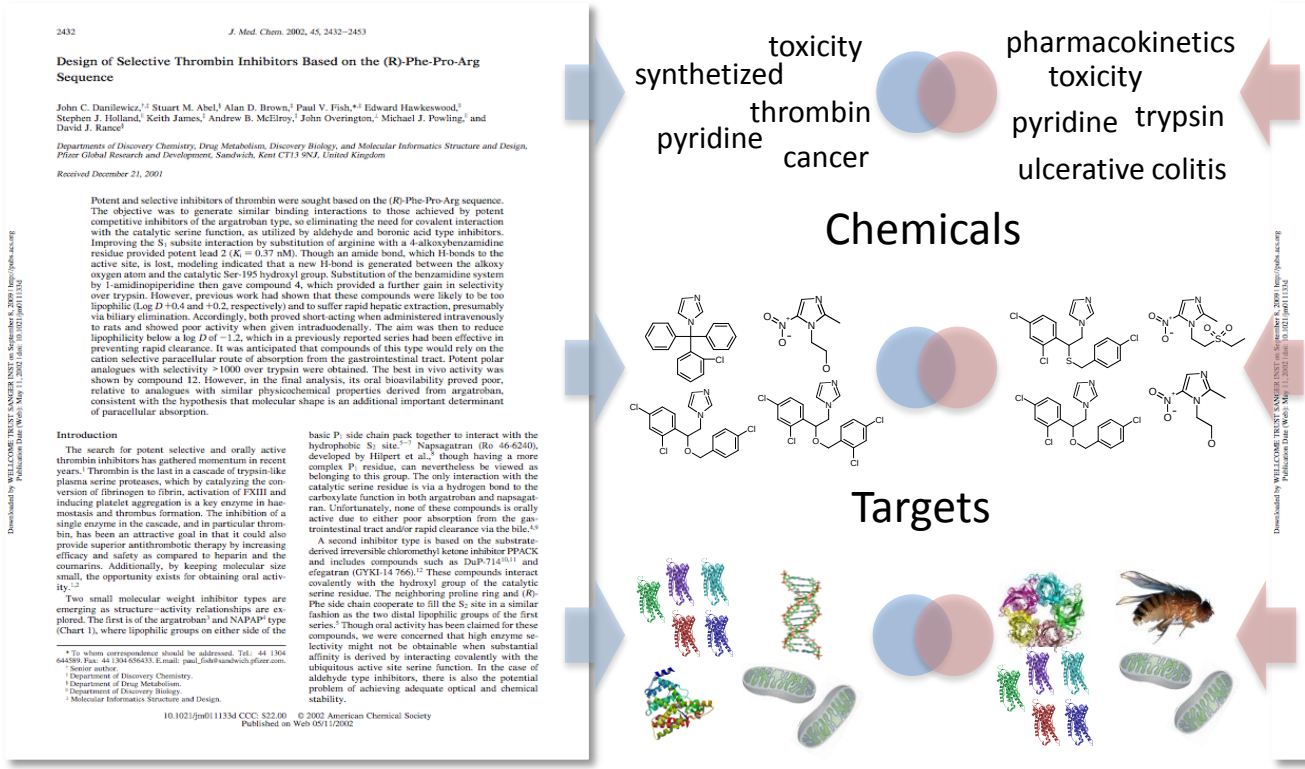
- EMBL-EBI have acquired the SureChem product from Digital Science
  - >15 million chemical structures
  - Automatically extracted chemical structures from full-text patent
- EMBL-EBI will Provide an ongoing free, Open resource to entire community
  - Add target, sequence, disease, animal model, cell-line indexing

Document Similarity

Document 1

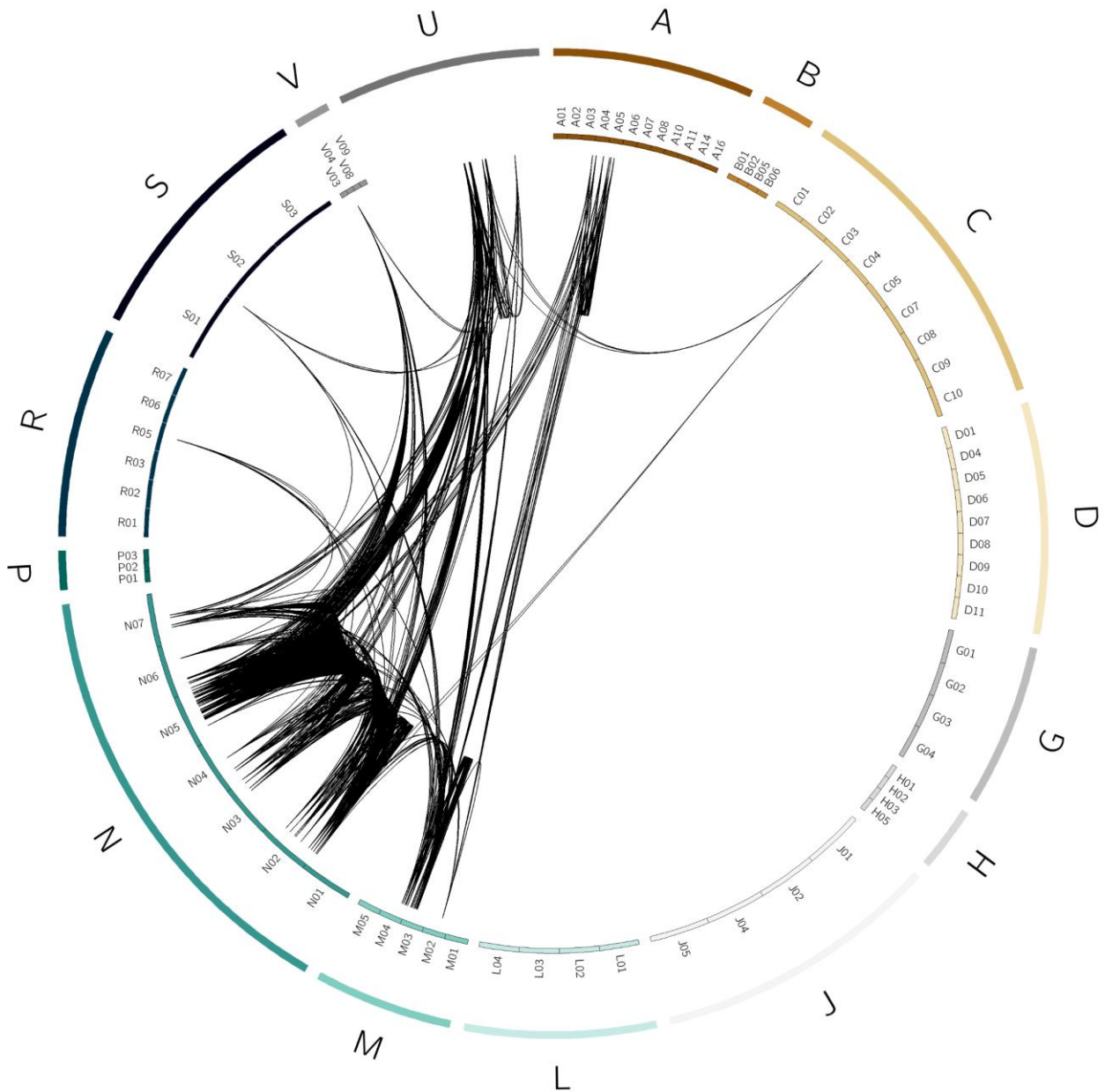
Words, n-grams, …

Document 2



- Document similarity methods currently rely on word/concept co-occurrence
- Possible to extend to include overlap/similarity of shared molecular objects
  - Sequences and Ligands
  - Greater richness possible in similarity measures and searches
    - Sequences – Sequence similarity, domains, structures,…
    - Ligands – Tanimoto similarity, scaffolds,….

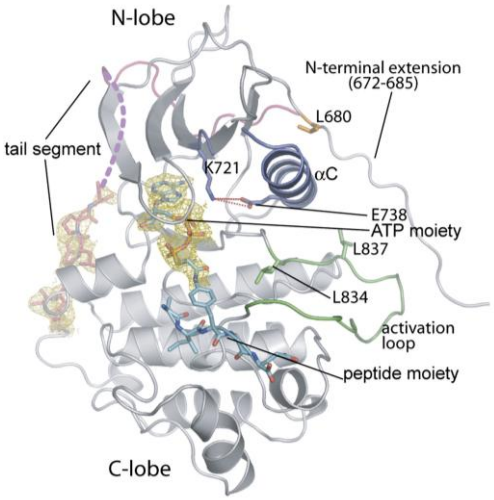
Polypharmacology *via* Binding Sites

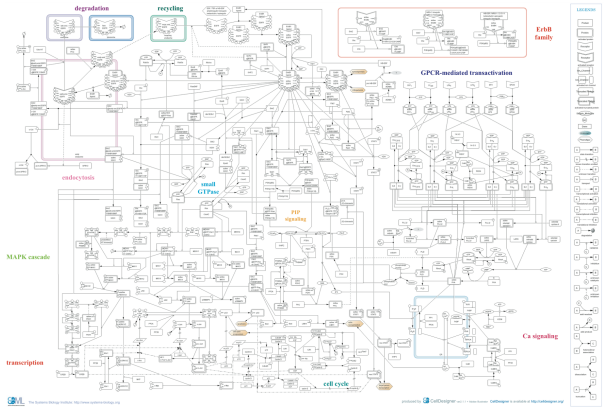


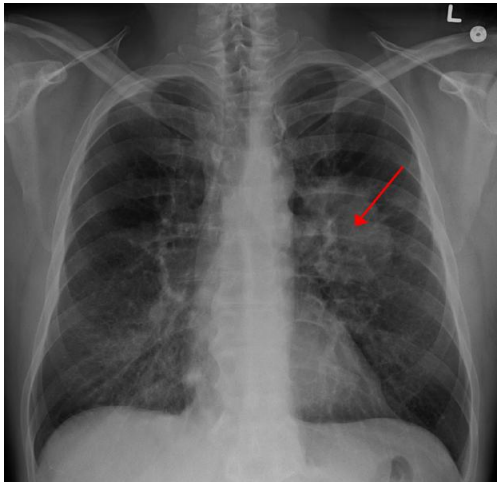
ATC Disease indications linked by shared similarity of target site, for Ligand-gated ion channels (Pfam PF02932)

Santos *et al*, submitted

Target – Pathway - Disease







Assays in Drug Discovery

Human
clinical trial

- Traditional medicines
  - Aspirin, Artemesinin, Arsenic trioxide….
  - Very slow and error prone
  - Not hypothesis led, *ad hoc* discovery

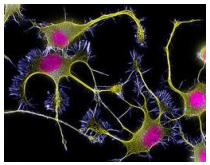
Assays in Drug Discovery

Animal
disease model

Human
clinical trial

- +ve
  - Higher Throughput
  - Greater Safety
  - Faster, cheaper, smaller scale
- -ve
  - Less predictive

Assays in Drug Discovery



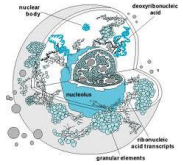
Functional assay

Animal disease model

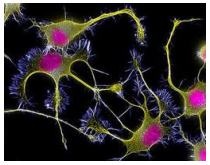
Human clinical trial

- +ve
  - Higher Throughput
  - Mechanistic insights and use of advances in basic science
  - Faster, cheaper, smaller scale
- -ve
  - Less predictive

Assays in Drug Discovery



Cell-based screen



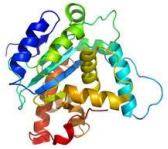
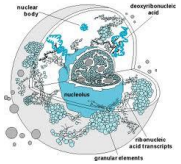
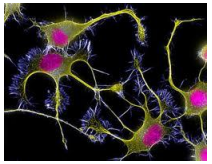
Functional assay

Animal disease model

Human clinical trial

- +ve
  - Higher Throughput
  - Mechanistic insights and use of advances in science
  - Faster, cheaper, smaller scale
- -ve
  - Less predictive

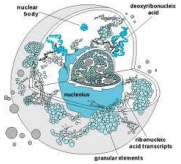
Assays in Drug Discovery

1980s	1960s	1950s	1920s	Ancient
				
Biochemical assay	Cell-based screen	Functional assay	Animal disease model	Human clinical trial

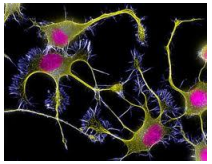
- +ve
  - Higher Throughput
  - Mechanistic insights and use of advances in science
  - Recombinant DNA technology and Genomics
  - Faster, cheaper, smaller scale
- -ve
  - Less predictive

Drug Discovery Assay "Cascade"

Biochemical assay



Cell-based screen



Functional assay

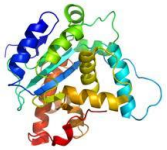
Animal disease model

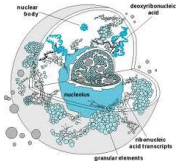
Human clinical trial

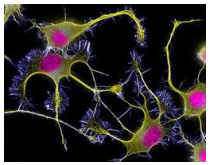
- Move from *'quick, low-cost, less predictive'* assays to '*slow, high-cost, more predictive'* assays
- Make selection of which compounds to progress to later assays on basis of activity in earlier screens
  - Early, cheap assays are used a lot of times; later, expensive assays rarely
  - Attrition – failure of compounds in that screening pipeline
  - Clinical trials configured in staged Phase 1, 2, 3 mode

Assay Costs

Costs are estimates!!

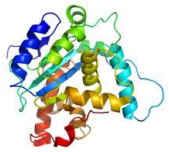




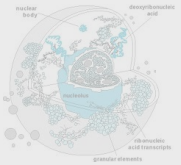


	In silico	Biochemical assay	Cell-based screen	Functional assay	Animal disease model	Human clinical trial
Cost (€)	0.0001	10	100	1,000	10,000	100,000,000

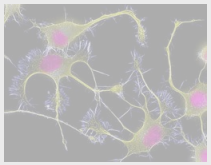
Targets & Diseases Connected *via* Drugs



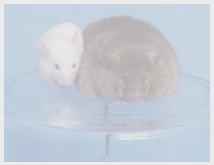
Biochemical assay



Cell-based screen



Functional assay



Animal disease model

Human clinical trial

PPARγ

PPARα

SUR1

K(ATP) channels

DPP-IV

GLP1R

Thrombin

Factor Xa

Target

.....

Type 2 diabetes

.....

.....

.....

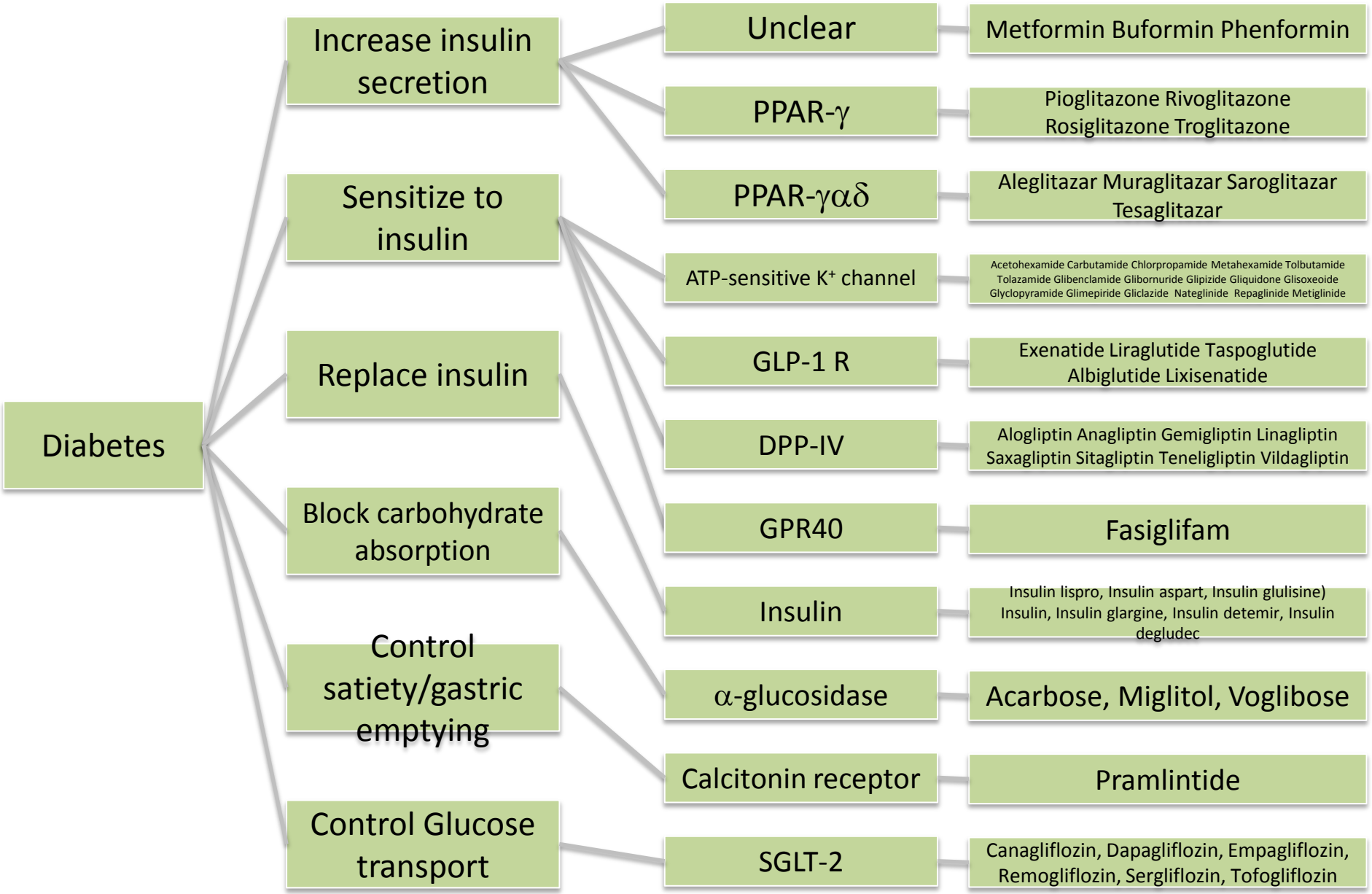
.....

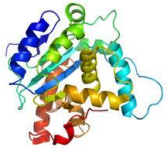
Deep vein thrombosis

.....

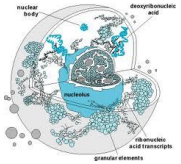
Disease

Ontology of Diabetes Drugs

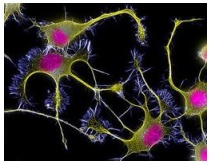




Biochemical assay



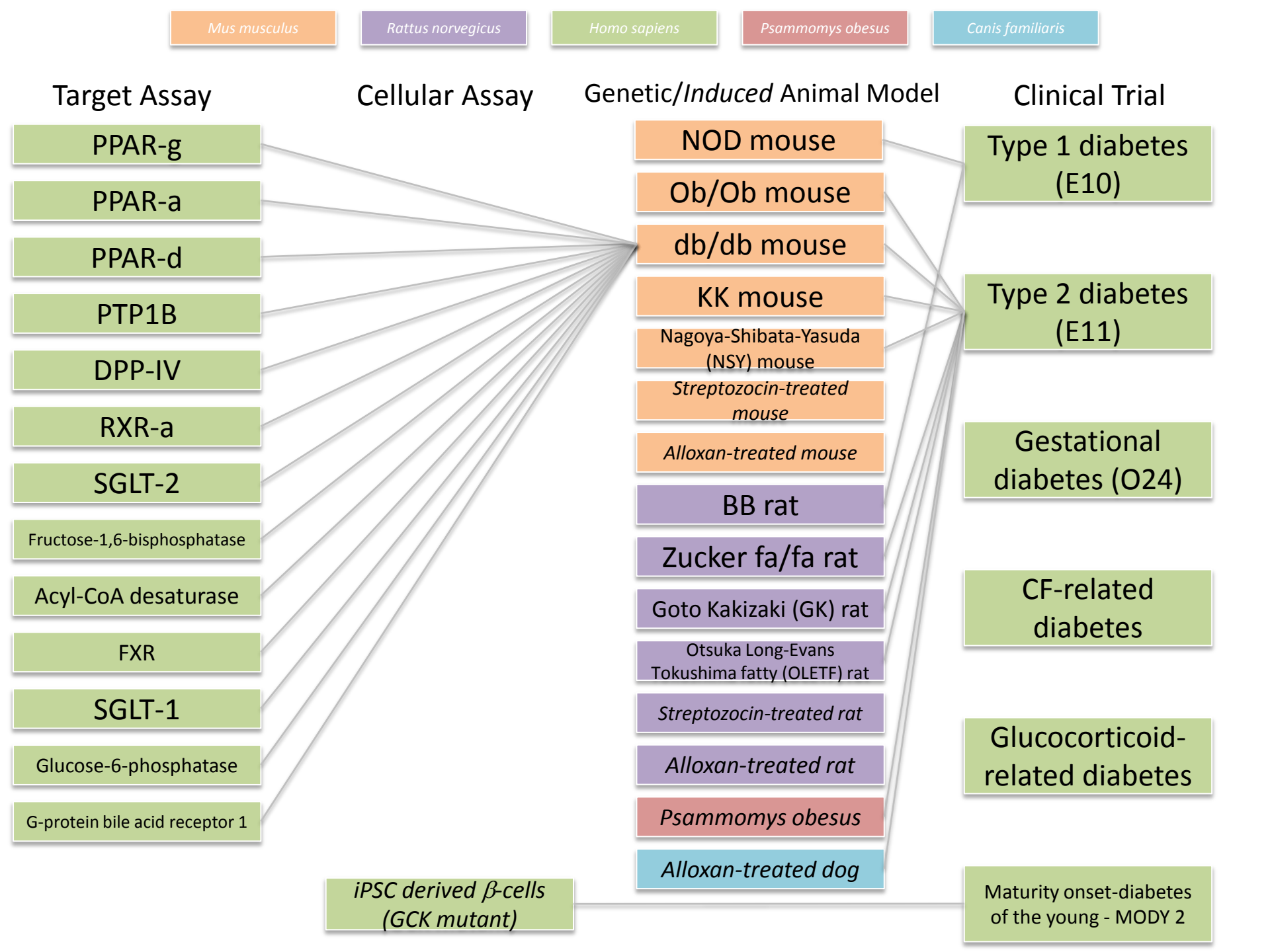
Cell-based screen



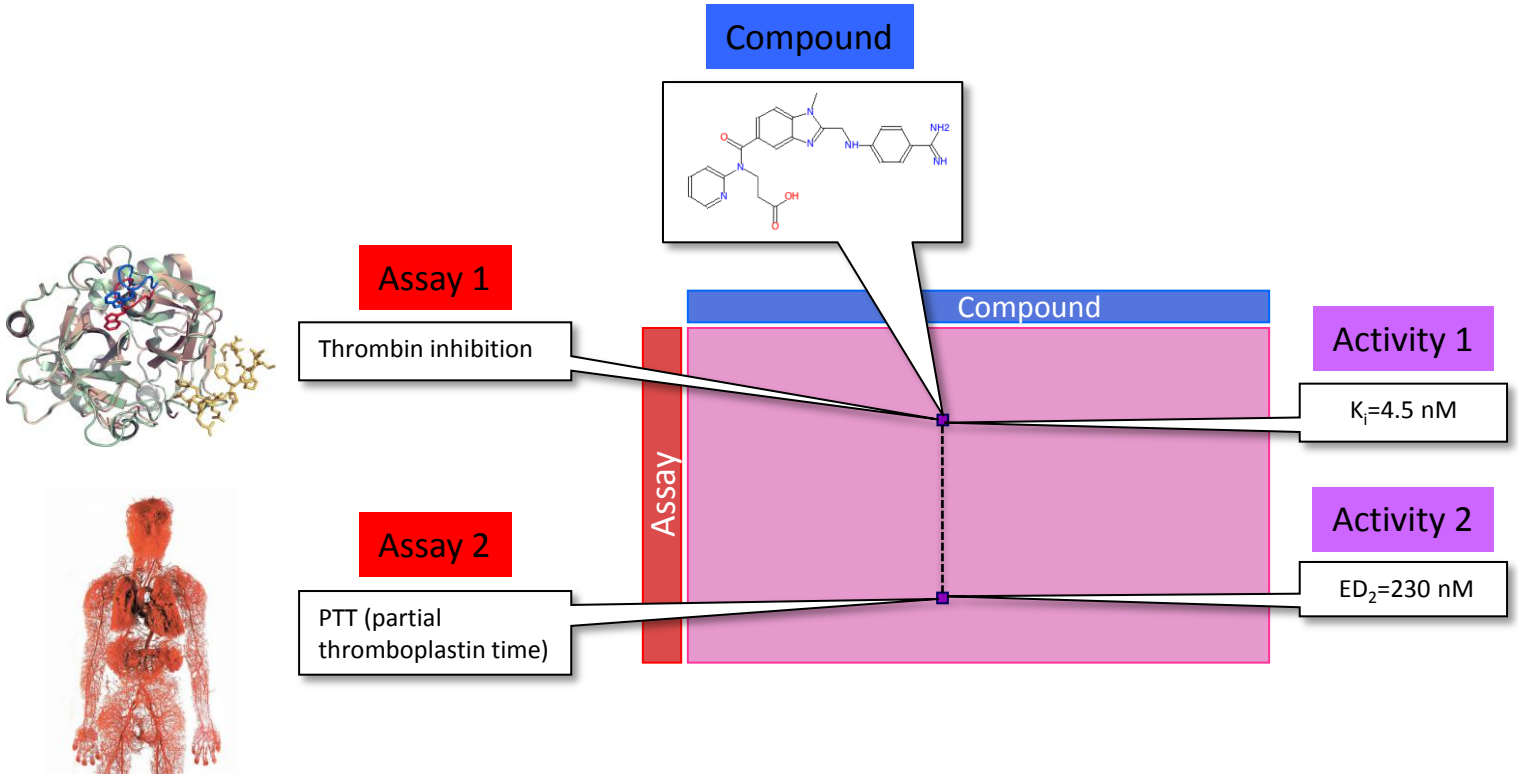
Functional assay

Animal disease model

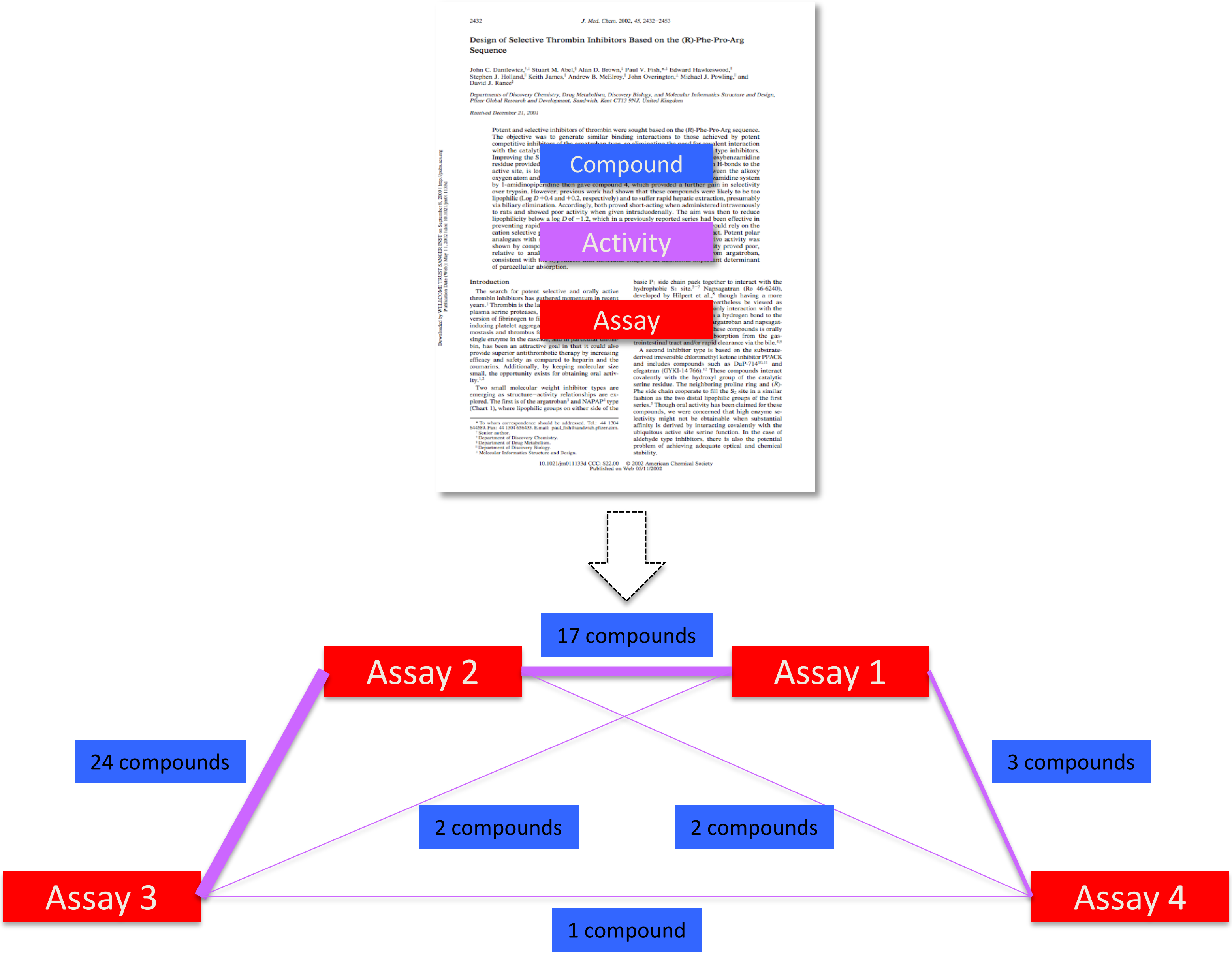
Human clinical trial


Mus musculus
Rattus norvegicus
Homo sapiens
Psammomys obesus
Canis familiaris
Target Assay
Cellular Assay
Genetic/Induced Animal Model
Clinical Trial
PPAR-g
PPAR-a
PPAR-d
PTP1B
DPP-IV
RXR-a
SGLT-2
Fructose-1,6-bisphosphatase
Acyl-CoA desaturase
FXR
SGLT-1
Glucose-6-phosphatase
G-protein bile acid receptor 1
NOD mouse
Ob/Ob mouse
db/db mouse
KK mouse
Nagoya-Shibata-Yasuda (NSY) mouse
Streptozocin-treated mouse
Alloxan-treated mouse
BB rat
Zucker fa/fa rat
Goto Kakizaki (GK) rat
Otsuka Long-Evans Tokushima fatty (OLETF) rat
Streptozocin-treated rat
Alloxan-treated rat
Psammomys obesus
Alloxan-treated dog
iPSC derived β-cells (GCK mutant)
Type 1 diabetes (E10)
Type 2 diabetes (E11)
Gestational diabetes (O24)
CF-related diabetes
Glucocorticoid-related diabetes
Maturity onset-diabetes of the young - MODY 2

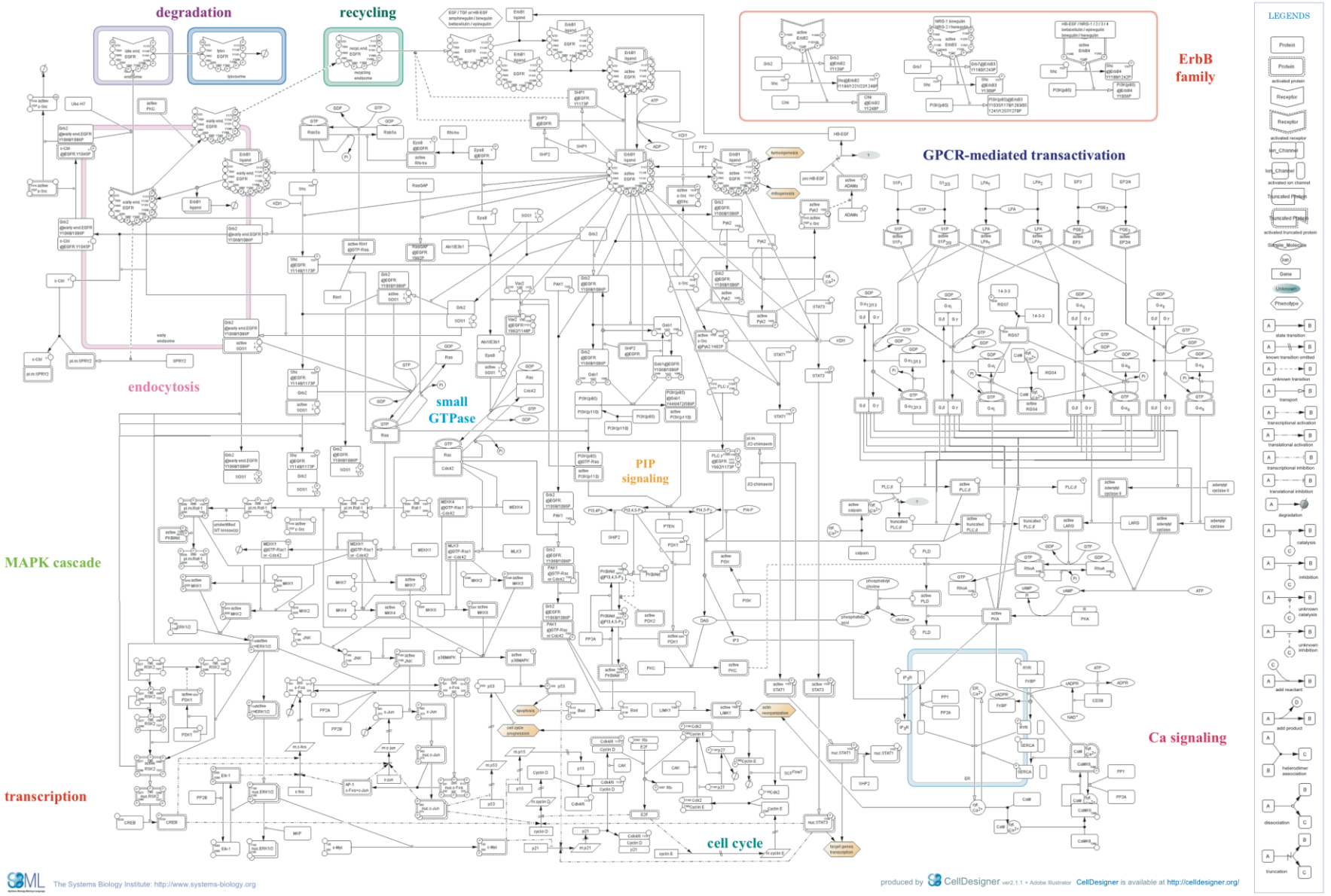
Compounds & Bioassays in ChEMBL



ChEMBL Assays as a Graph



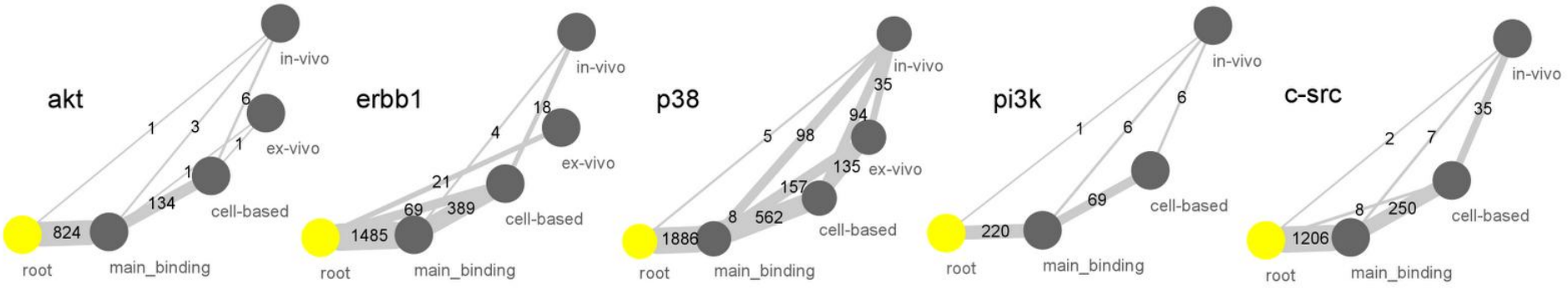
EGFR Signaling Pathway



K. Oda, *et al. Mol. Syst. Biol.* **1** DOI:10.1038/msb4100014

EGFR Assay Cascades From ChEMBL

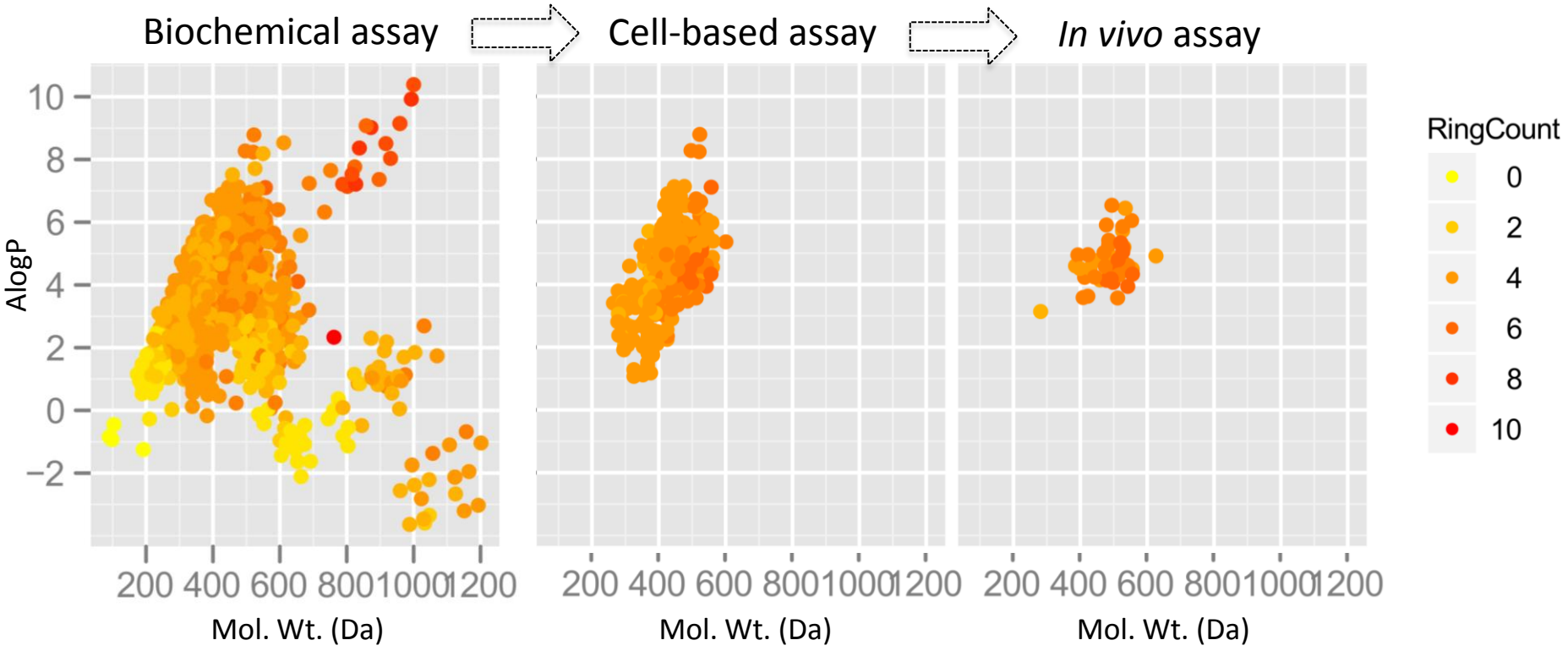
Assay Network for EGFR pathway inhibitors



F. Krueger (unpublished)

EGFR Assay Cascades from ChEMBL

Physicochemical properties for cSrc EGFR pathway inhibitors



F. Krueger (unpublished)

Extraction & Curation of PK Data

Single Imatinib 400 mg dose

Bioequivalence & Bioavailability

Research Article

Bioavailability of a New Generic Formulat
Tablets Versus Glivec in Healthy Male Ad

Dalia Jawhari, Mahmoud AlSwisi and Mahmoud Ghannam*

Abstract

Imatinib is a highly selective inhibitor of tyrosine kinase use
the cost of the drug is prohibitive especially in the developing
pharmacokinetics profile of a new Imatinib generic formulation (
Hikma Pharmaceuticals PLC) with those of Glivec, (batch number:
in healthy male volunteers /fed state. The study was single cen
2-period, 2-sequence, crossover design. The study was perfon
Canada) in accordance with Good Clinical Practices and the app
non- or ex-smokers, of at least 18 years of age but not older than 5
or equal to 18.5 and below 30 kg/m2 were selected according to
period, a single 400 mg dose of Imatinib was orally administered
a 10-hour overnight fast, thirty (30) minutes after the start of a h
seated for at least the first 4 hours following each drug admini
samples were collected by venipuncture in pre-cooled Vacutaine
4 mL) was collected prior to drug administration while the others
2.67, 3, 3.33, 3.67, 4, 4.5, 5, 6, 8, 10, 14, 18, 24, 48 and 72 hours
were separated by at least 14 calendar days. Urine drug and eth
period of the study. Hematology and biochemistry tests were repeated after the collection of the last blood sample
of the study. Safety was evaluated through the assessment of adverse events, and laboratory tests. Imatinib plasma
samples were analyzed employing a validated HPLC method using MS/MS detection. For a 400 mg dose of Imatinib,
the analytical range was approximately 10 ng/mL to 4000 ng/mL. Descriptive statistics were used to summarize
adverse events, safety results and demographic variables (age, height, weight and BMI). The main pharmacokinetic
parameters of interest for this study were C_{max}, AUC_{0-t} and $AUC_{0-\infty}$. Other parameters such as T_{max}, AUC_{0-t}, K_e
and $T_{1/2}$ were provided for information purposes only. The natural logarithmic transformation of C_{max}, AUC_{0-t} and
$AUC_{0-\infty}$ was used for all statistical inference. The mean (C.V. %) of C_{max}, T_{max}, AUC_{0-t} and $AUC_{0-\infty}$ for Imatinib were
1760.5 ng/ml (26.6%), 3.67 hrs (26.4%), 30946.5 ng.h/ml (28.0%) and 31912.5 ng.h/ml (28.2%) versus 1779.4 ng/
ml (25.8%), 3.67 hrs (39.0%), 31073.6 ng.h/ml (25.7%) and 32270.9 ng.h/ml (26.4%) for Glivec. The 90% confidence
intervals of C_{max}, AUC_{0-t} and $AUC_{0-\infty}$ for Imatinib 400 mg were (92.00%-105.52%), (95.69%-102.31%) and (95.23%-
101.55%) respectively. The ratio of the geometric LS means for the test to reference C_{max}, AUC_{0-t} and $AUC_{0-\infty}$ for
Imatinib 400 mg were 98%, 99% and 99% respectively with low ISCV 12.9 % for C_{max} and 8.3 % for AUC_{0-t} and 6.0%
for $AUC_{0-\infty}$. The results indicated that the products are equivalent and switchable according to FDA rulings.

Keywords: Imatinib, Bioequivalence, Glivec, Tyrosine kinase inhibitor

Introduction

Imatinib is the 4-[(4-methyl-1-piperazinyl) methyl]-N-[4-methyl-3-[(4-(3-pyridinyl)-2-pyrimidinyl] amino) phenyl] benzamide, which is represented by the following structure (Figure 1).

Protein tyrosine kinases participate in signal transduction pathways which regulate diverse cellular processes such as growth, metabolism, differentiation, adhesion and apoptosis [1,2]. Deregulation of protein tyrosine kinase activity has been associated with the pathogenesis of various cancers, including chronic myeloid leukemia (CML) and gastrointestinal stromal tumour (GIST), as well as other proliferative diseases [3].

CML is a hematologic stem-cell disorder associated with a specific chromosomal translocation known as the Philadelphia Chromosome (Ph). Ph is detected in 95% of patients with CML from which 20% of them have acute lymphocytic leukemia [4]. The molecular consequences of the Ph translocation is the fusion of the Abl (Abelson leukemia virus) proto-oncogen on chromosome 9 to the Bcr (breakpoint cluster

Figure 1:

*Corresponding author: Mahmoud Ghannam, Jordan, E-mail: mghannam@apm.com.jo

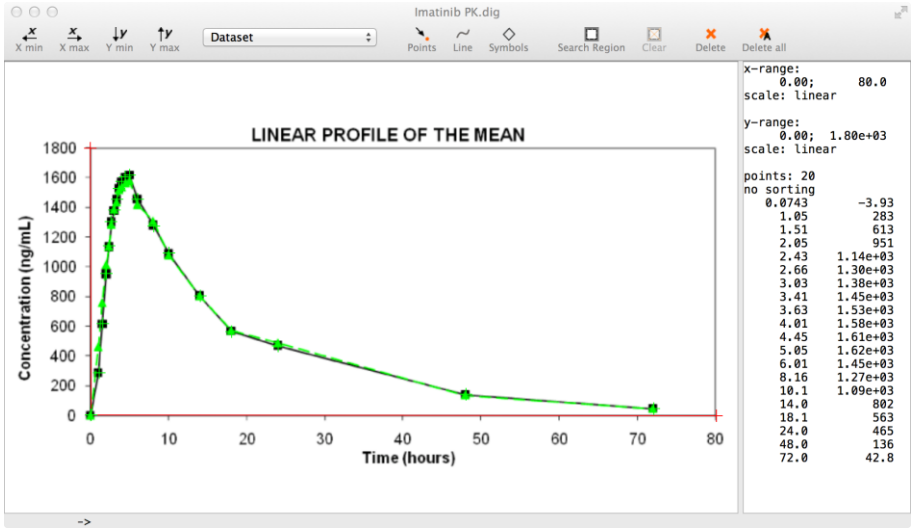
Received May 17, 2011; Accepted July 29, 2011; Published August 01, 2011

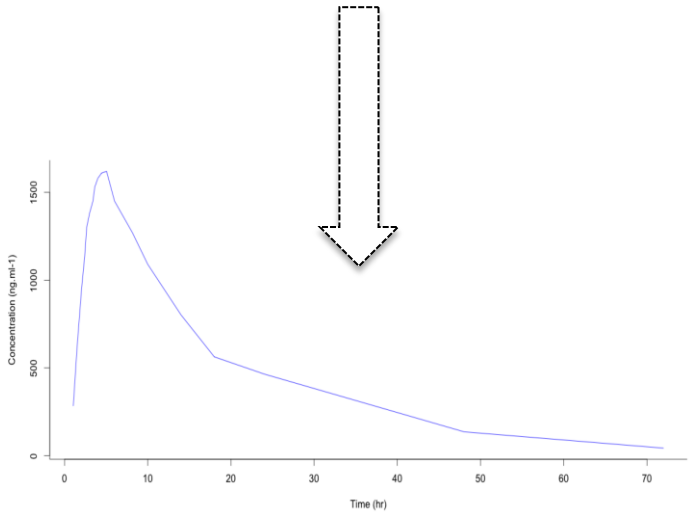
Citation: Jawhari D, AlSwisi M, Ghannam M (2011) Bioavailability of a New Generic Formulation of Imatinib Mesylate-400mg Tablets Versus Glivec in Healthy Male Adult Volunteers. J Bioequiv Availab 3: 161-164. doi:10.4172/jbb.1000077

Copyright: © 2011 Jawhari D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

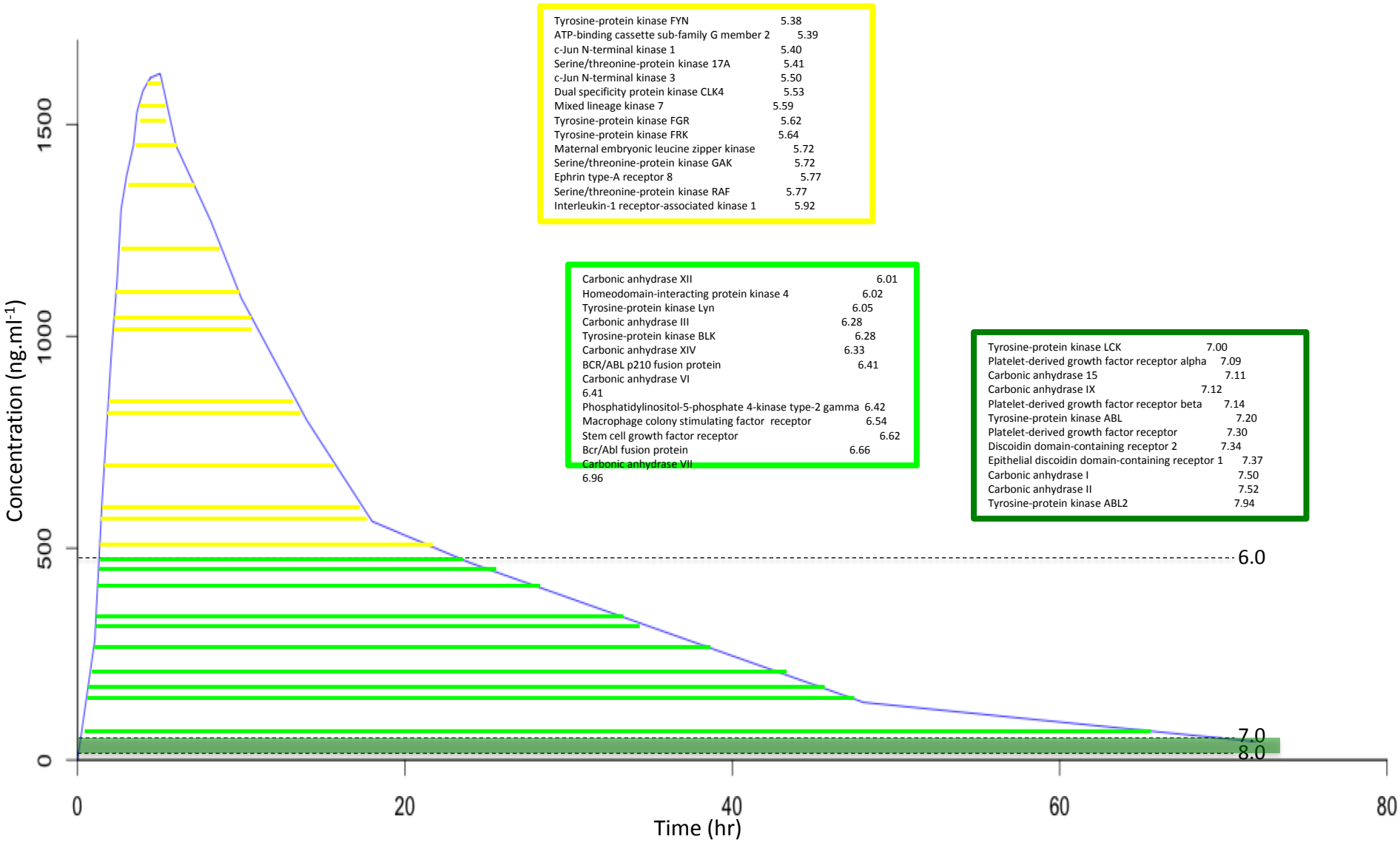
J Bioequiv Availab
ISSN:0975-0851 JBB, an open access journal

Volume 3(7): 161-164 (2011) - 161



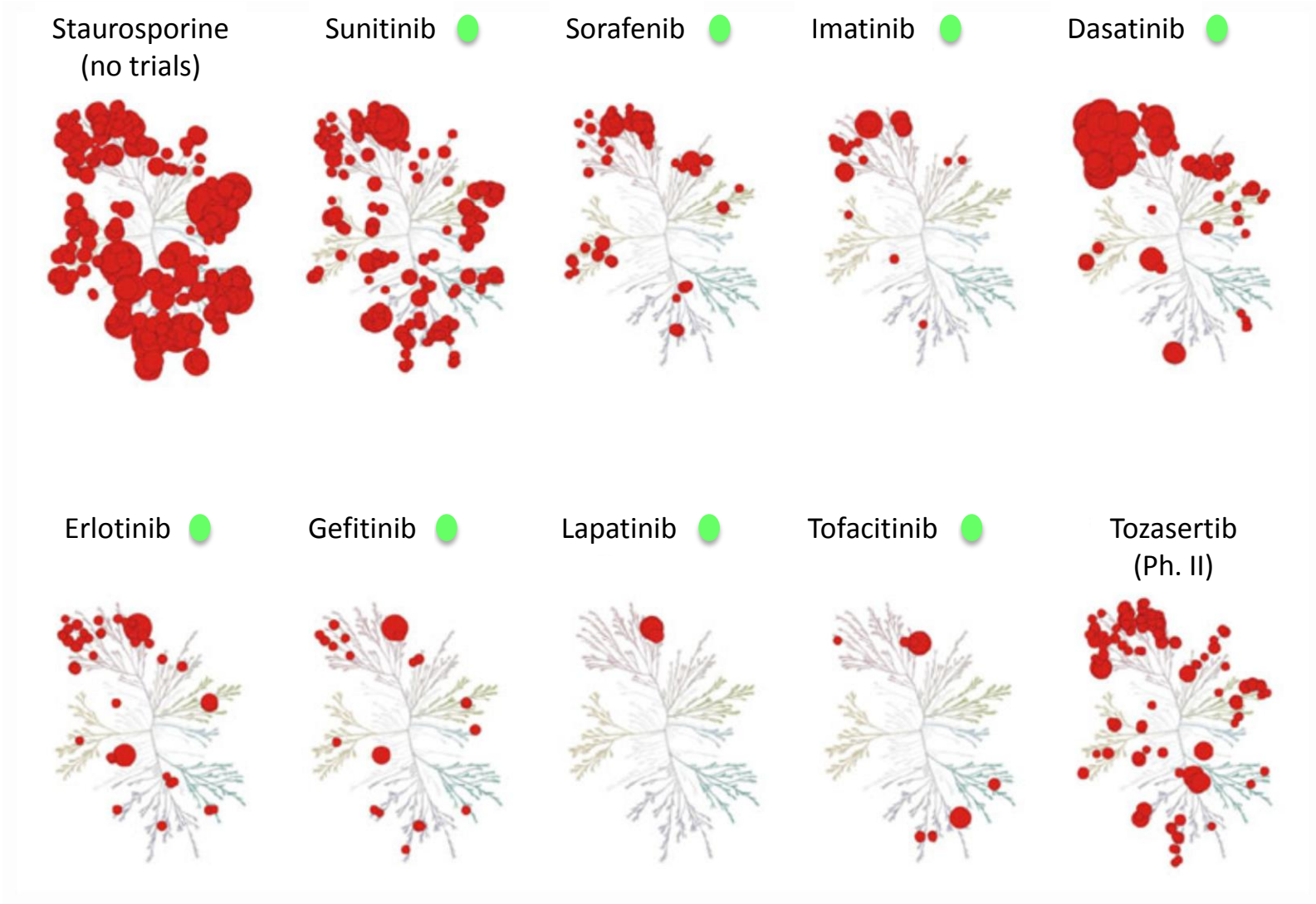


Imatinib Polypharmacology Spectra



Imatinib 400 mg single dose from Jawhari *et al* (2011) J Bioequiv Availab 3: 161-164; Data is median pChEMBL for human targets from ChEMBL 16

Kinase Inhibitor Polypharmacology

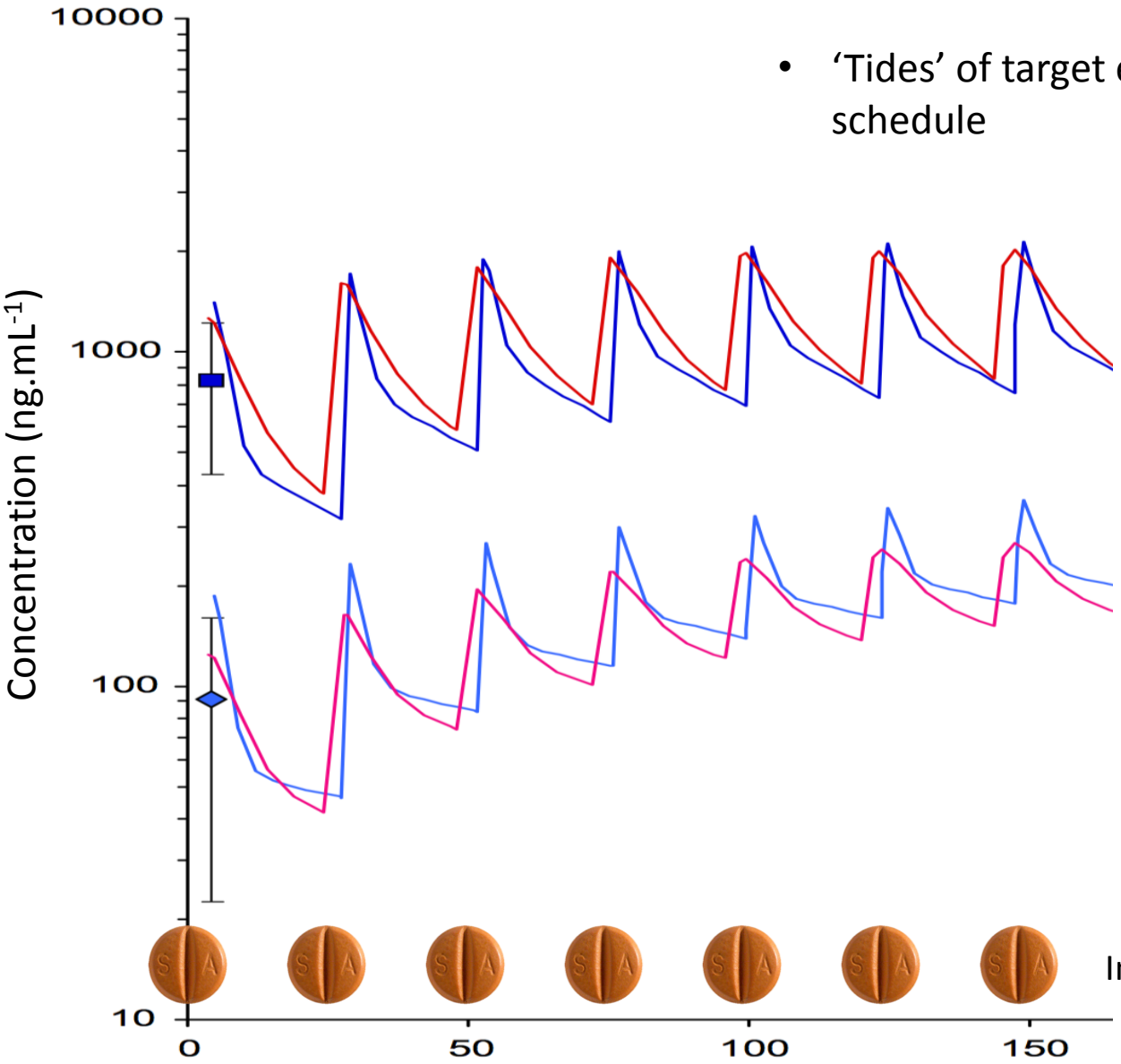


US launched

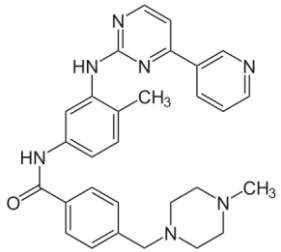
Adapted from Ghoreschi *et al*, *Nature Immunology* **10**, 356 - 360 (2009)

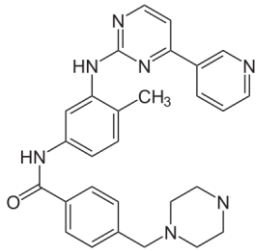
Imatinib Pharmacokinetics

- 'Tides' of target exposure during dosing schedule



imatinib





N-desmethyl imatinib

Imatinib, 400 mg uid

Note log scale for concentration axis

Acknowledgements

- ChEMBL
  - Anne Hersey
  - Anna Gaulton
  - Louisa Bellis
  - George Papadatos
  - Mark Davies
  - Michal Nowotka
- UniChem
  - Jon Chambers
- eTox
  - Ruth Akhtar
  - Francis Atkinson
  - Patricia Bento
- Research
  - Felix Kruger
  - Ben Stauch
  - Rita Santos
  - Joey Bach Hardie
  - Grace Mugumbate
  - Gerard Van Westen