ChEMBL:
The Organization of Drug Discovery Data

4th Summer School in Chemoinformatics
Strasbourg

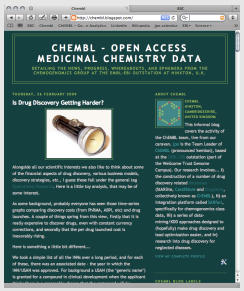
John P. Overington
25th June 2014

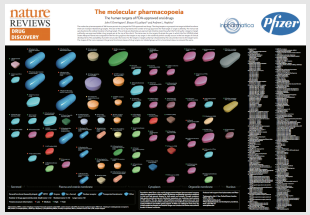

DE QUOI T'OCCUPES TU EXACTEMENT ?
DE LA RÉIFICATION
NON, JE ME PROMÈNE. PRINCIPALEMENT JE ME PROMÈNE.
JE VOIS, C'EST UN TRAVAIL TRÈS SÉRIEUX, AVEC DE GROS LIVRES ET BEAUCOUP DE PAPIERS SUR UNE GRANDE TABLE.

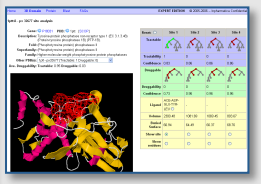
ChEMBL – Data for Drug Discovery

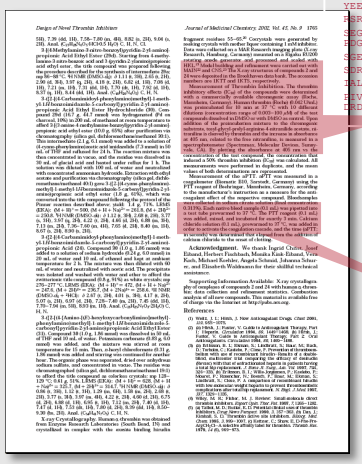
1. Scientific facts

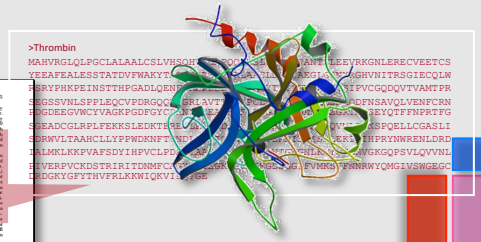
3. Insight, tools and resources for translational drug discovery

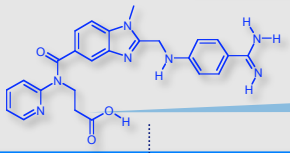


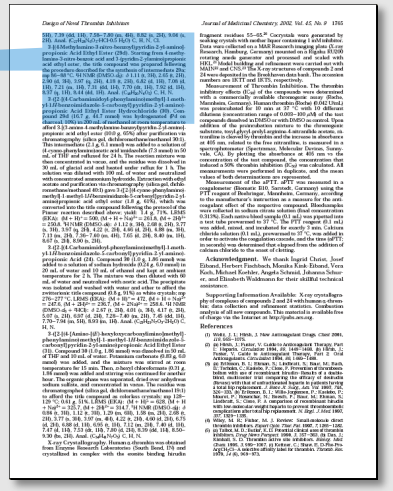












Compound

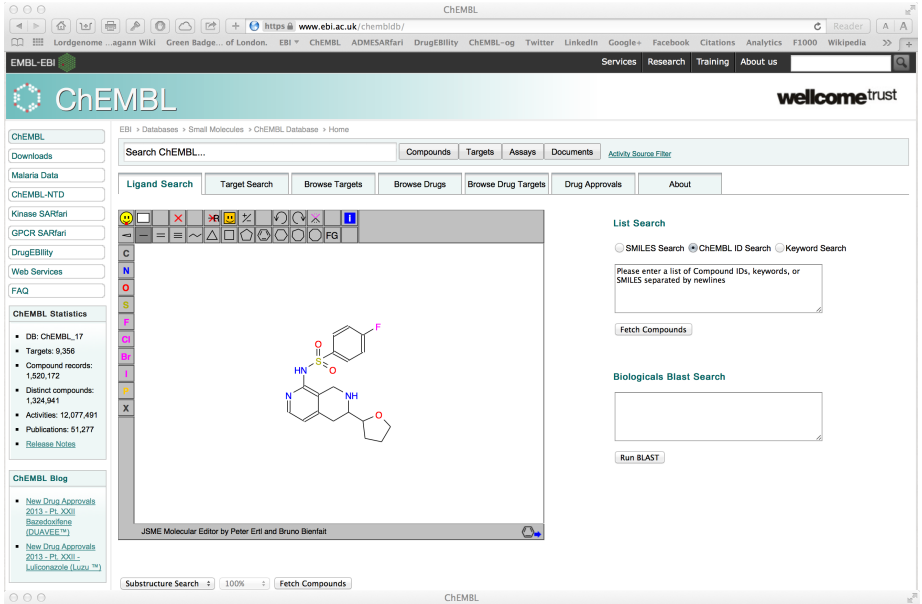
K_i = 4.5nM

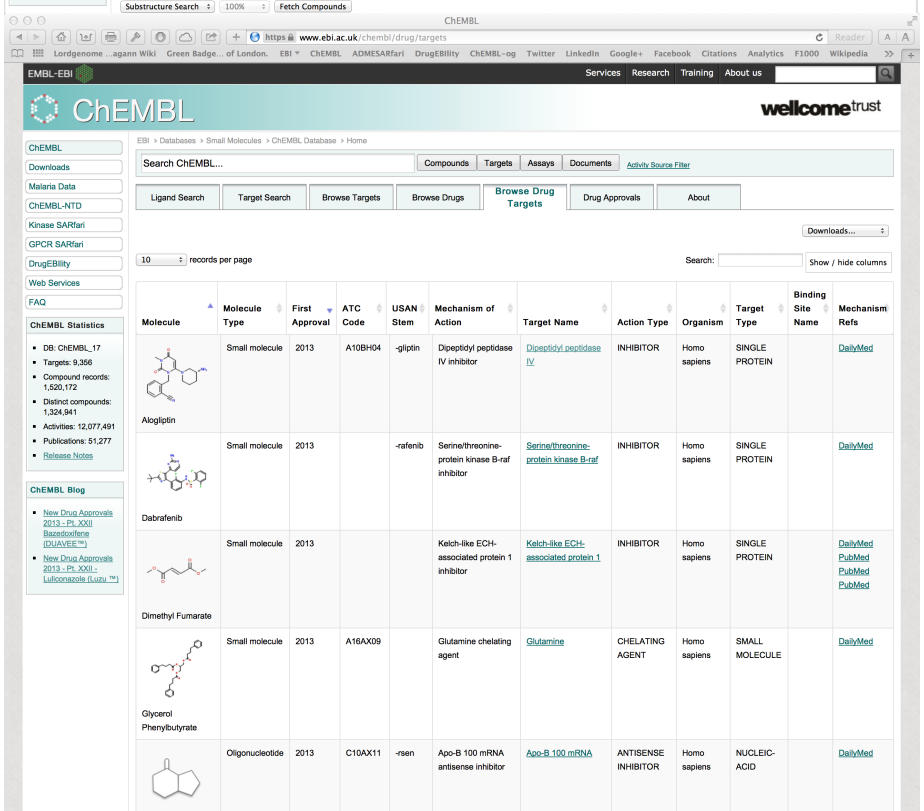
Assay/Target

Bioactivity data

APTT = 11 min.

2. Organization, curation and standardization of pharmacology data





ChEMBL

https://www.ebi.ac.uk/chembl

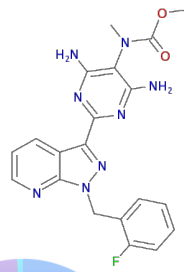
- The world's largest primary public database of medicinal chemistry data
  - ~1.4 million compounds, ~9,000 targets, ~12 million bioactivities
- Truly Open Data - CC-BY-SA license
- ChEMBL data also loaded into BindingDB, PubChem BioAssay and BARD

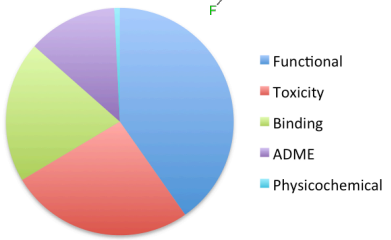
A. Gaulton et al (2012) *Nucleic Acids Research Database Issue.* **40** D1100-1107

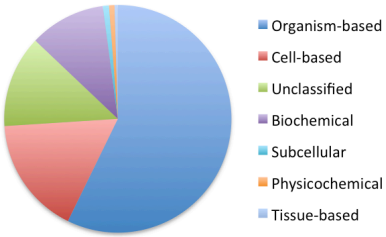
MyChEMBL - 8GB FOSS!

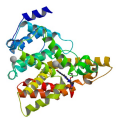
ChEMBL - Added Value in Curation

- Annotation of approved drug and USAN information
  - Small molecules and biotherapeutics
  - Human and pathogen targets
  - Direct molecular targets
  - Mode of action assigned
- Standardization and correction of chemical structures
- Classification of assays
  - assay type, format, organisms, cell-lines
- Standardization and validation of assay endpoints
  - unit conversion, identification of errors, duplicates
- Detailed representation of molecular targets
  - single proteins *vs.* protein complexes, protein-protein interactions *etc.* and protein family classification



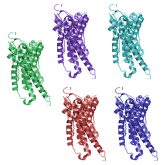






PDE5

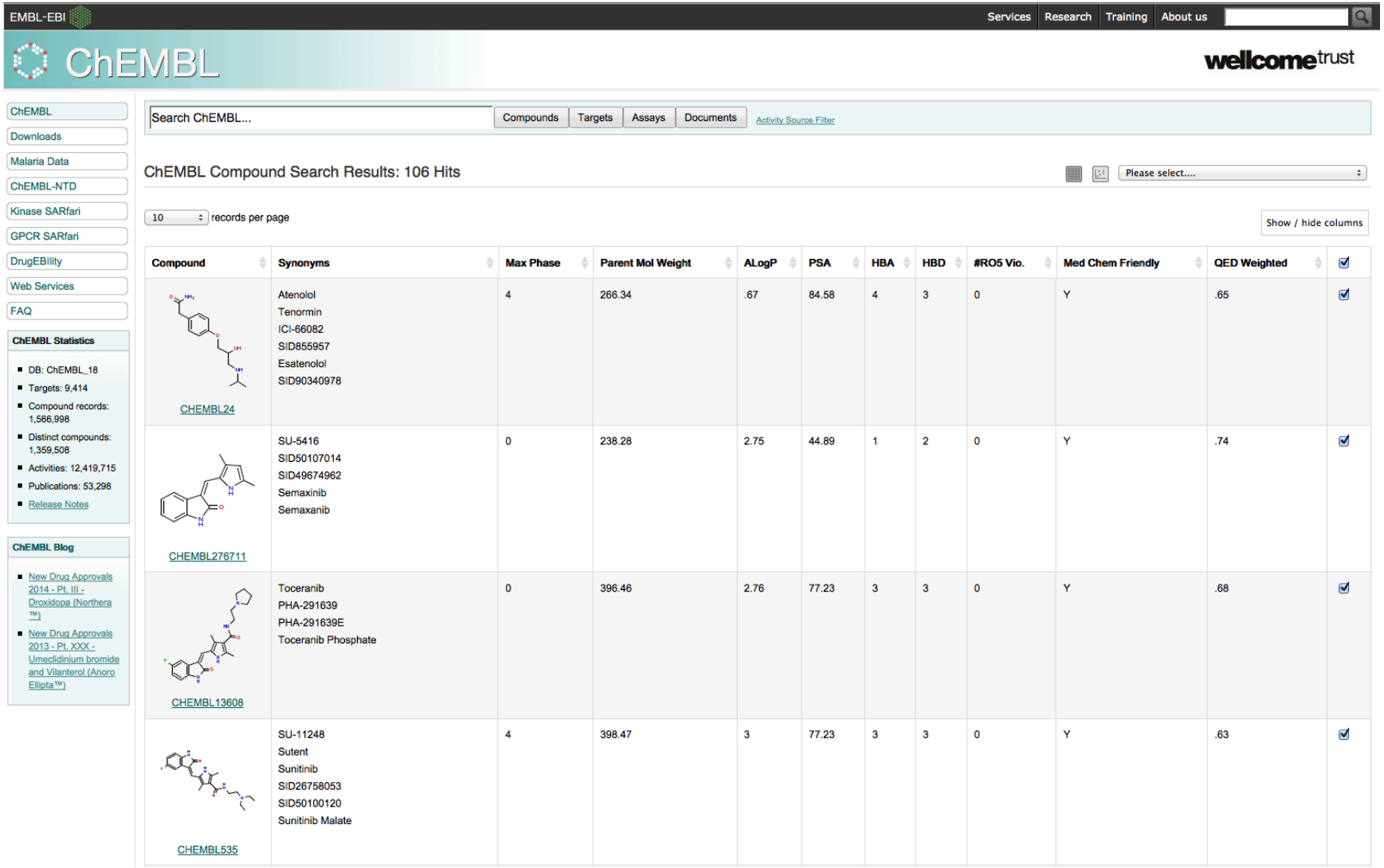
GABA-A receptor



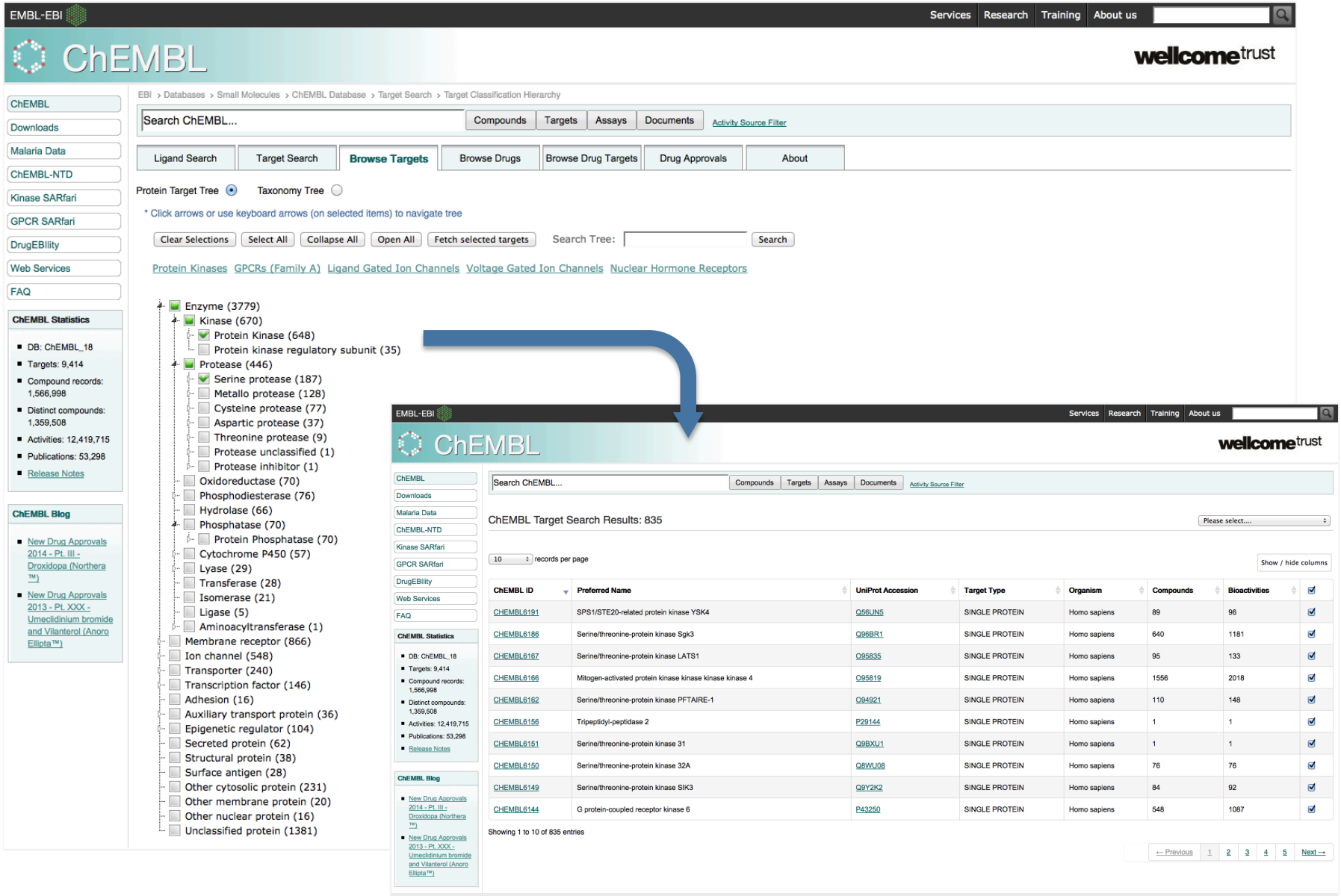
Muscarinic receptors

E = enzyme
MR = membrane receptor
IC = ion channel
TF = transcription factor
T = transporter
PO = protein, other

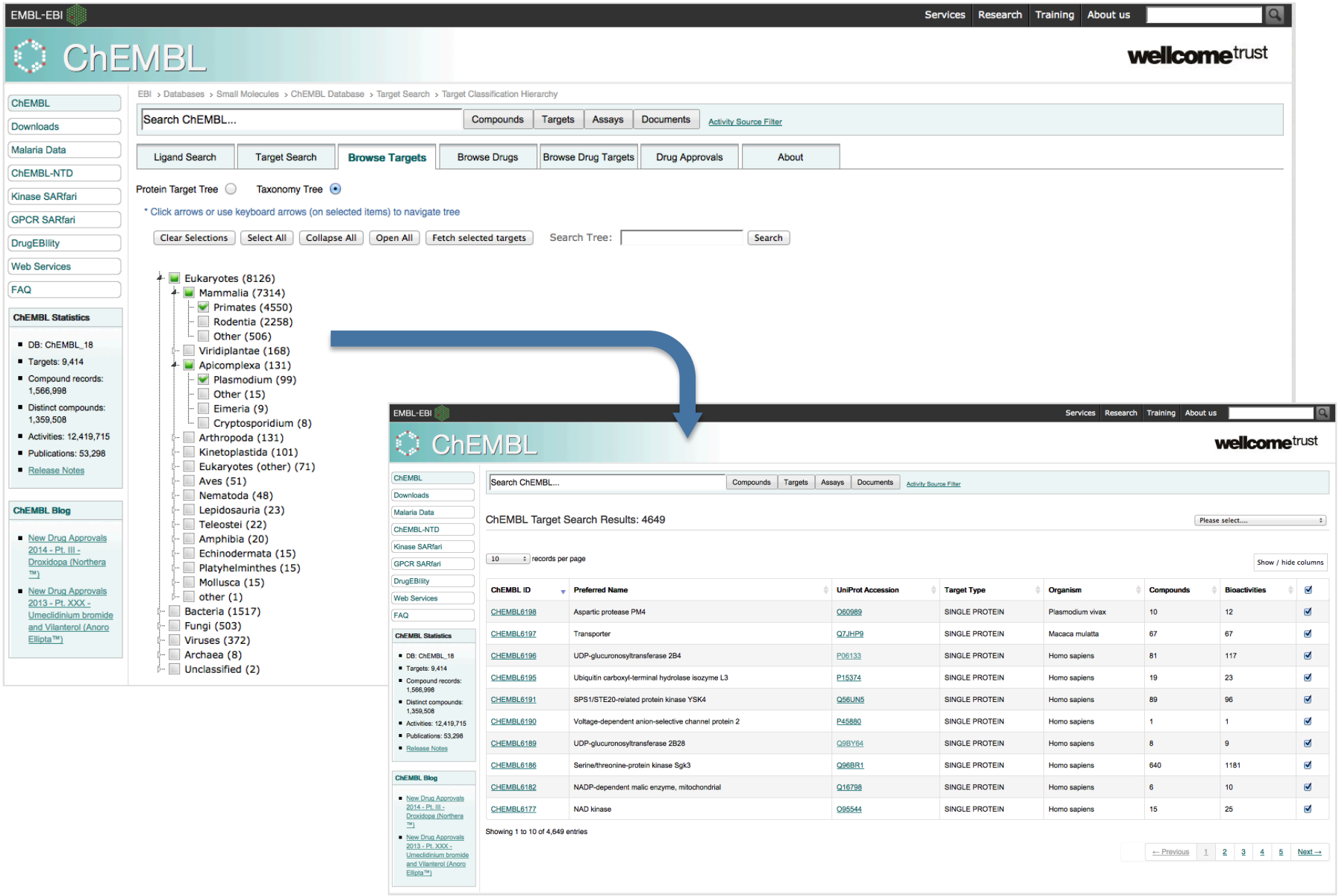
Spreadsheet Views



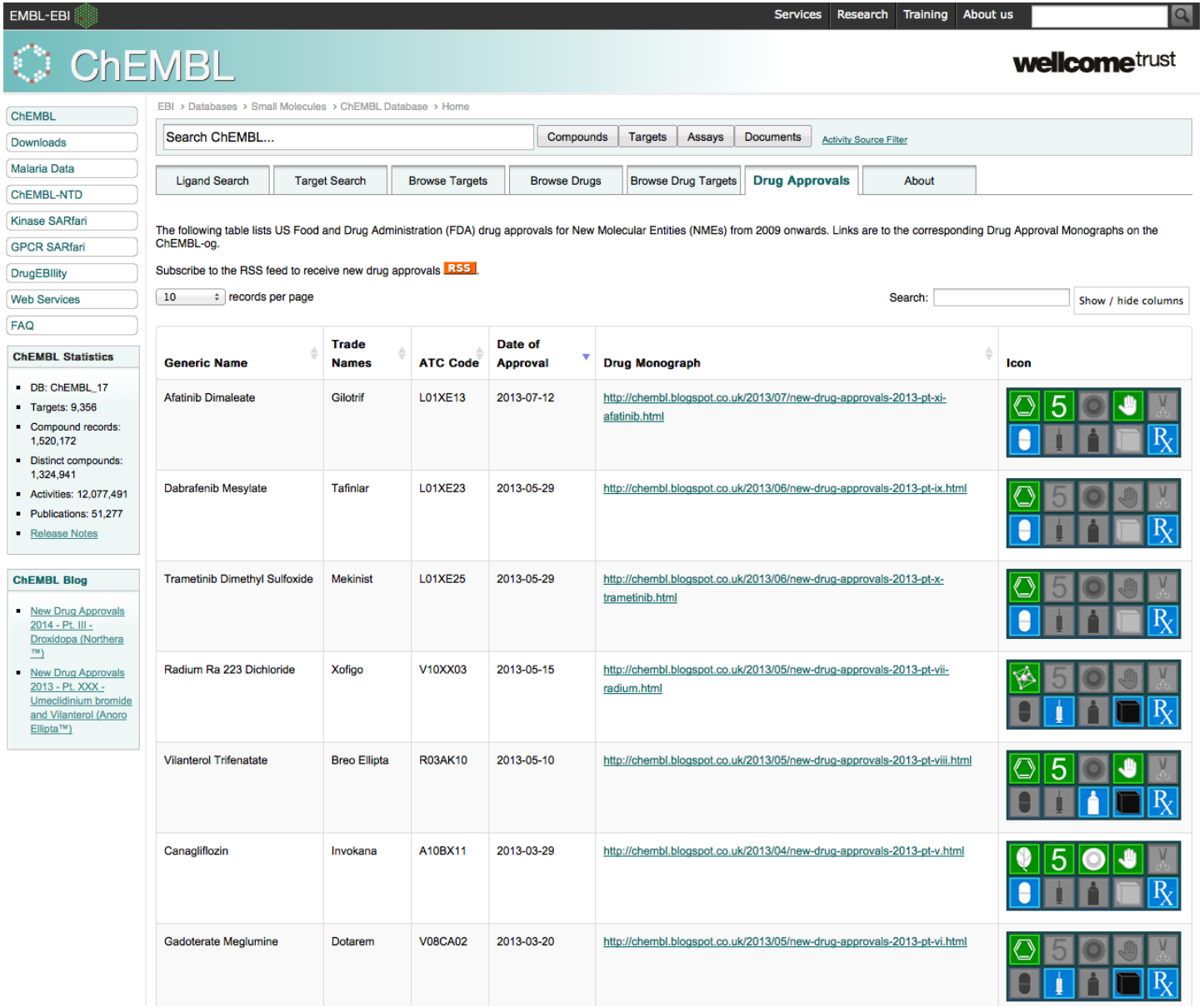
Target Class Data



Assay Organism Data



Drug Approvals



EMBL-EBI | Services | Research | Training | About us

ChEMBL

wellcome trust

EBI › Databases › Small Molecules › ChEMBL Database › Home

Search ChEMBL... | Compounds | Targets | Assays | Documents | Activity Source Filter

Ligand Search | Target Search | Browse Targets | Browse Drugs | Browse Drug Targets | **Drug Approvals** | About

ChEMBL
Downloads
Malaria Data
ChEMBL-NTD
Kinase SARfari
GPCR SARfari
DrugEBIlity
Web Services
FAQ

ChEMBL Statistics

- DB: ChEMBL_17
- Targets: 9,356
- Compound records: 1,520,172
- Distinct compounds: 1,324,941
- Activities: 12,077,491
- Publications: 51,277
- Release Notes

ChEMBL Blog

- New Drug Approvals 2014 - Pt. III - Droxidopa (Northera ™)
- New Drug Approvals 2013 - Pt. XXX - Umeclidinium bromide and Vilanterol (Anoro Ellipta™)

The following table lists US Food and Drug Administration (FDA) drug approvals for New Molecular Entities (NMEs) from 2009 onwards. Links are to the corresponding Drug Approval Monographs on the ChEMBL-og.

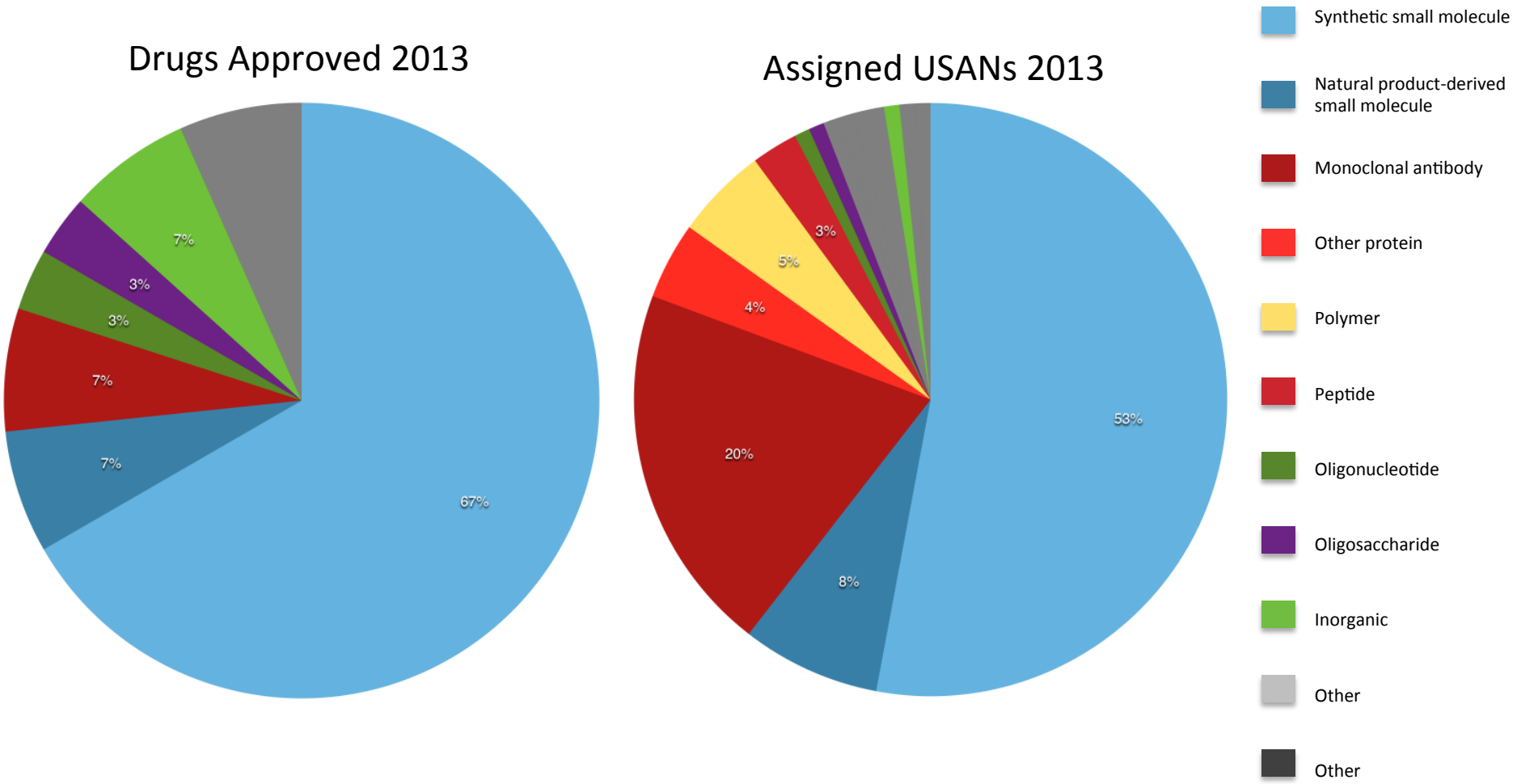
Subscribe to the RSS feed to receive new drug approvals RSS.

10 records per page

Search: | Show / hide columns

Generic Name	Trade Names	ATC Code	Date of Approval	Drug Monograph	Icon
Afatinib Dimaleate	Gilotrif	L01XE13	2013-07-12	http://chembl.blogspot.co.uk/2013/07/new-drug-approvals-2013-pt-xi-afatinib.html	
Dabrafenib Mesylate	Tafinlar	L01XE23	2013-05-29	http://chembl.blogspot.co.uk/2013/06/new-drug-approvals-2013-pt-ix.html	
Trametinib Dimethyl Sulfoxide	Mekinist	L01XE25	2013-05-29	http://chembl.blogspot.co.uk/2013/06/new-drug-approvals-2013-pt-x-trametinib.html	
Radium Ra 223 Dichloride	Xofigo	V10XX03	2013-05-15	http://chembl.blogspot.co.uk/2013/05/new-drug-approvals-2013-pt-vii-radium.html	
Vilanterol Trifenatate	Breo Ellipta	R03AK10	2013-05-10	http://chembl.blogspot.co.uk/2013/05/new-drug-approvals-2013-pt-viii.html	
Canagliflozin	Invokana	A10BX11	2013-03-29	http://chembl.blogspot.co.uk/2013/04/new-drug-approvals-2013-pt-v.html	
Gadoterate Meglumine	Dotarem	V08CA02	2013-03-20	http://chembl.blogspot.co.uk/2013/05/new-drug-approvals-2013-pt-vi.html	

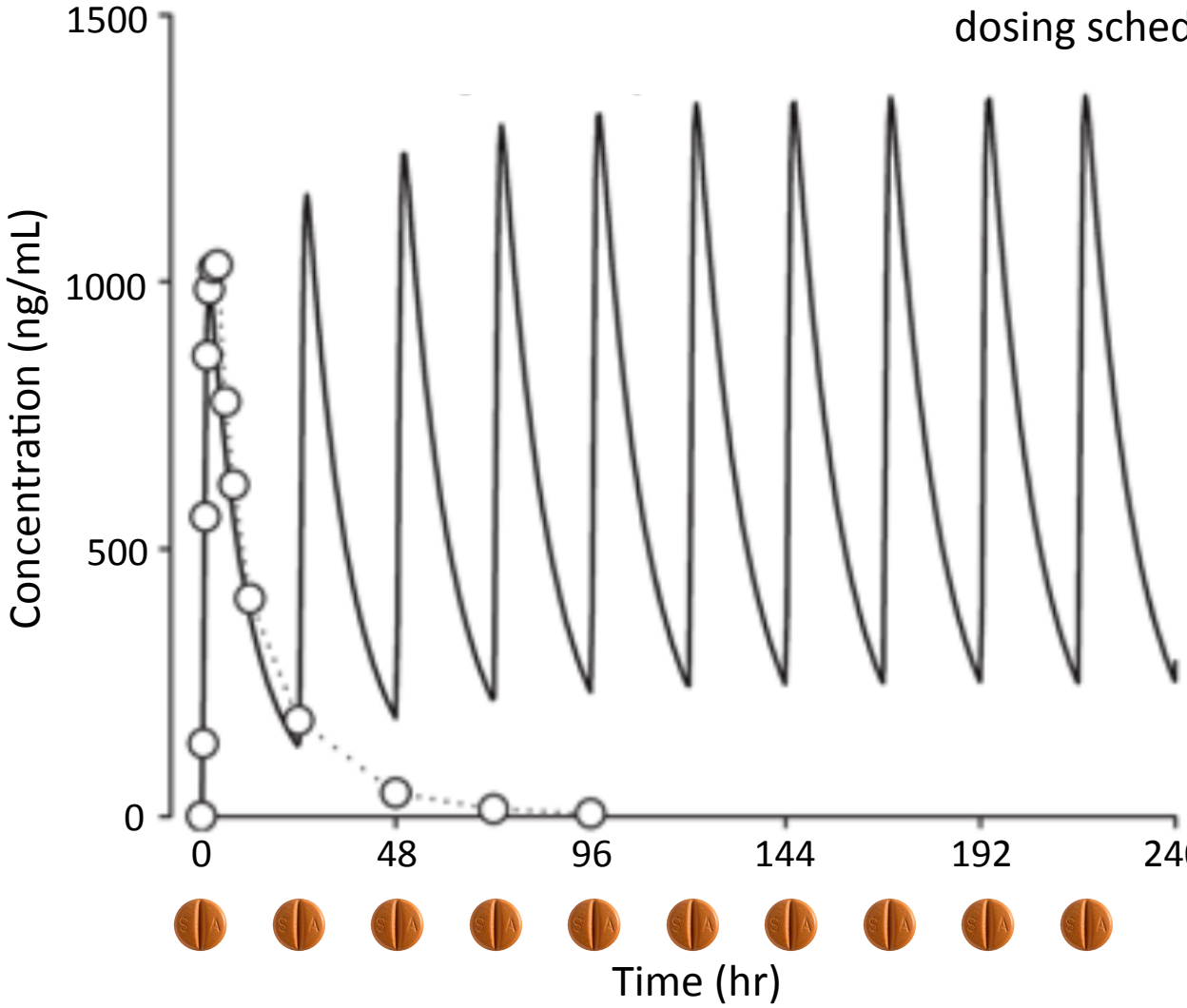
Different Types of Drugs

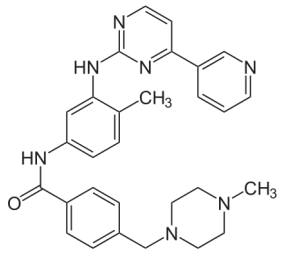


Santos *et al*, *unpublished*

Filling The Gaps – Exposure Data

- Large 'tides' of target exposure during dosing schedule

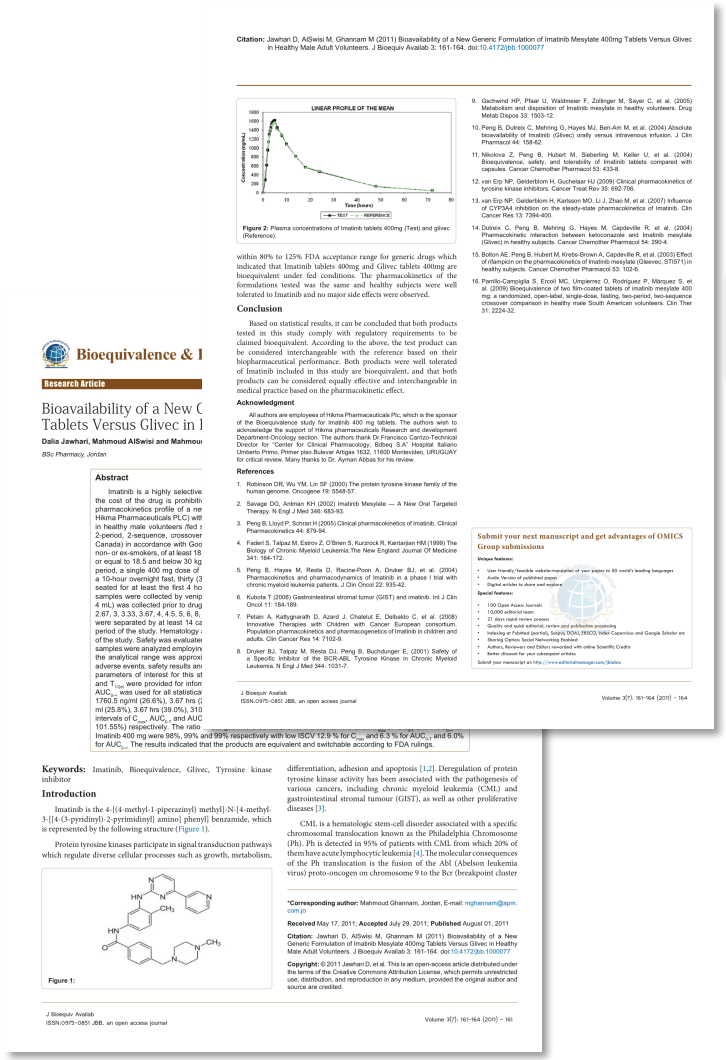




Imatinib, 200 mg uid

Filppula *et al, Clin. Pharmacol & Therap* (2013) **94** 383-389

Filling the Gaps – Exposure Data



Bioequivalence & I

Research Article

Bioavailability of a New (
Tablets Versus Glivec in I

Dalia Jawhari, Mahmoud AlSwisi and Mahmou

BSc Pharmacy, Jordan

Abstract

Imatinib is a highly selective
the cost of the drug is prohibiti
pharmacokinetics profile of a ne
Hikma Pharmaceuticals PLC) wit
in healthy male volunteers (fed s
2-period, 2-sequence, crossover
Canada) in accordance with Goo
non- or ex-smokers, of at least 18
or equal to 18.5 and below 30 kg
period, a single 400 mg dose of
a 10-hour overnight fast, thirty (3
seated for at least the first 4 ho
samples were collected by venip
4 mL) was collected prior to drug
2.67, 3, 3.33, 3.67, 4, 4.5, 5, 6, 8,
were separated by at least 14 da
period of the study. Hematology
of the study. Safety was evaluate
samples were analyzed employin
the analytical range was approxi
adverse events, safety results an
parameters of interest for this st
and T_{12h} were provided for infor
AUC_{∞} was used for all statistica
1760.5 ng/ml (26.6%), 3.67 hrs (
ml (25.8%), 3.67 hrs (39.0%), 310
intervals of C_{max}, AUC_{∞} and AUC
101.55%) respectively. The ratio
Imatinib 400 mg were 98%, 99% and 99% respectively with low ISCV 12.9 % for C_{max} and 6.3 % for AUC_{∞} and 6.0% for AUC_{0-t}. 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

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

LINEAR PROFILE OF THE MEAN

Time (hours)

Figure 2: Plasma concentrations of imatinib tablets 400mg (Test) and glivec (Reference).

within 80% to 125% FDA acceptance range for generic drugs which indicated that Imatinib tablets 400mg and Glivec tablets 400mg are bioequivalent under fed conditions. The pharmacokinetics of the formulations tested was the same and healthy subjects were well tolerated to Imatinib and no major side effects were observed.

Conclusion

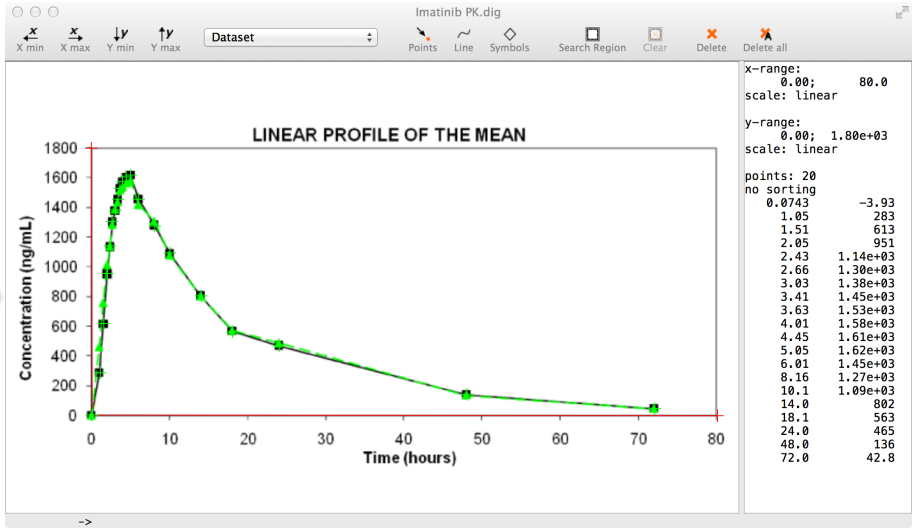
Based on statistical results, it can be concluded that both products tested in this study comply with regulatory requirements to be claimed bioequivalent. According to the above, the test product can be considered interchangeable with the reference based on their biopharmaceutical performance. Both products were well tolerated of Imatinib included in this study are bioequivalent, and that both products can be considered equally effective and interchangeable in medical practice based on the pharmacokinetic effect.

Acknowledgment

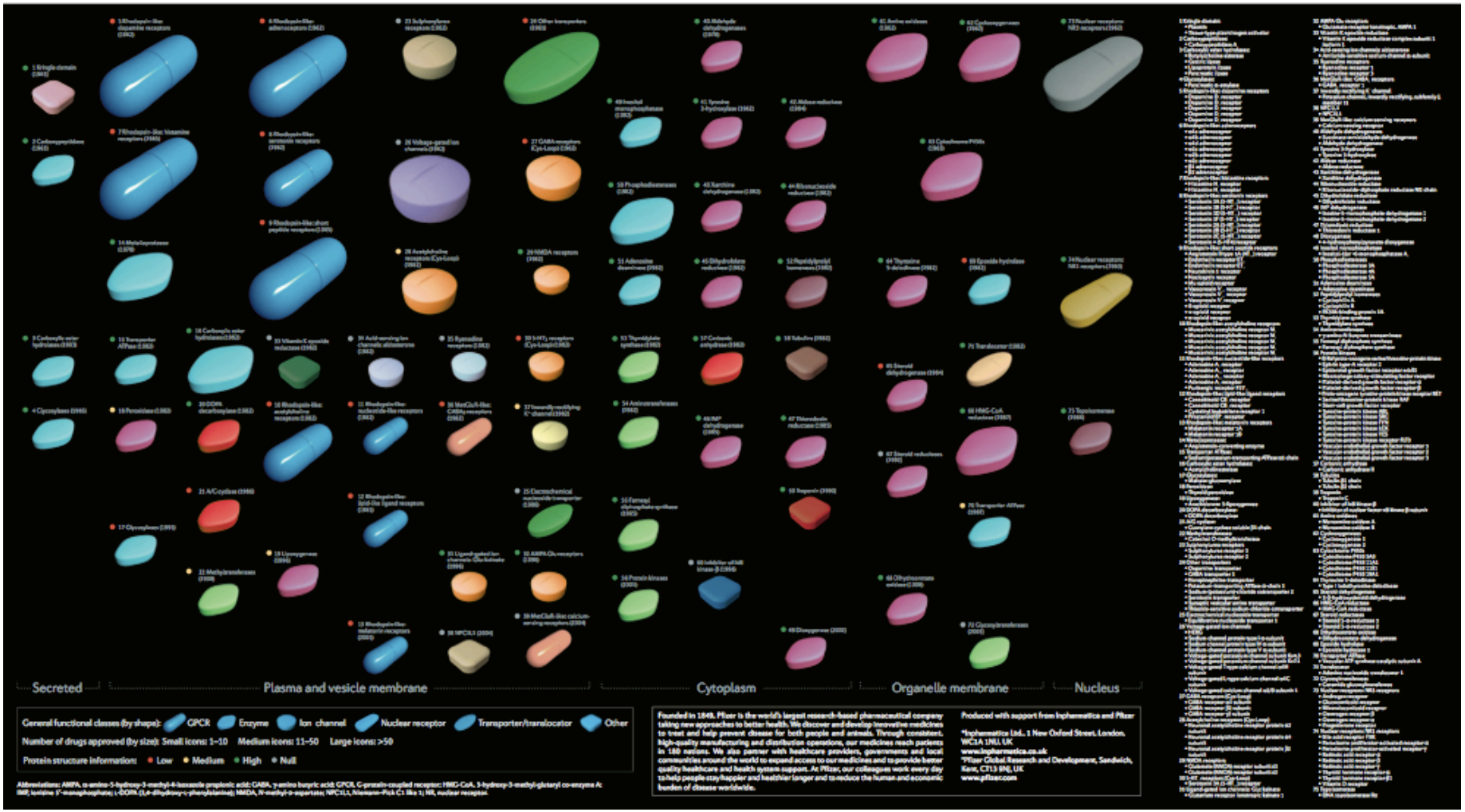
All authors are employees of Hikma Pharmaceuticals Plc, which is the sponsor of the Bioequivalence study for Imatinib 400 mg tablets. The authors wish to acknowledge the support of Hikma pharmaceuticals Research and development Department-Oncology section. The authors thank Dr.Francisco Carrizo-Technical Director for "Center for Clinical Pharmacology , Bidneg S.A" Hospital Italiano Umberto Primo. Primer piso.Bulevar Artigas 1632. 11600 Montevideo, URUGUAY for critical review. Many thanks to Dr. Ayman Abbas for his review.

References

1. Robinson DR, Wu YM, Lin SF (2000) The protein tyrosine kinase family of the human genome. Oncogene 19: 5548-57.
2. Savage DG, Antman KH (2002) Imatinib Mesylate — A New Oral Targeted Therapy. N Engl J Med 346: 683-93.
3. Peng B, Lloyd P, Schran H (2005) Clinical pharmacokinetics of imatinib. Clinical Pharmacokinetics 44: 879-94.
4. Fadeel B, Tabaz M, Estrov Z, O'Brien S, Kurzrock R, Kantarjian HM (1999) The Biology of Chronic Myeloid Leukemia.The New England Journal Of Medicine 341: 164-172.
5. Peng B, Hayes M, Resta D, Racine-Poon A, Druker BJ, et al. (2004) Pharmacokinetics and pharmacodynamics of imatinib in a phase I trial with chronic myeloid leukemia patients. J Clin Oncol 22: 935-42.
6. Kubota T (2006) Gastrointestinal stromal tumor (GIST) and imatinib. Int J Clin Oncol 11: 184-189.
7. Petain A, Kattygnarath D, Azard J, Chatelut E, Delbaldo C, et al. (2008) Innovative Therapies with Children with Cancer European consortium. Population pharmacokinetics and pharmacogenetics of imatinib in children and adults. Clin Cancer Res 14: 7102-9.
8. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, (2001) Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. N Engl J Med 344: 1031-7.
9. Gschwind HP, Pfaar U, Waldmeier F, Zollinger M, Sayer C, et al. (2005) Metabolism and disposition of Imatinib mesylate in healthy volunteers. Drug Metab Dispos 33: 1503-12.
10. Peng B, Dutreix C, Mehring G, Hayes MJ, Ben-Am M, et al. (2004) Absolute bioavailability of Imatinib (Glivec) orally versus intravenous infusion. J Clin Pharmacol 44: 158-62.
11. Nikolova Z, Peng B, Hubert M, Sieberling M, Keller U, et al. (2004) Bioequivalence, safety, and tolerability of Imatinib tablets compared with capsules. Cancer Chemother Pharmacol 53: 433-8.
12. van Erp NP, Gelderblom H, Guchelaar HJ (2009) Clinical pharmacokinetics of tyrosine kinase inhibitors. Cancer Treat Rev 35: 692-706.
13. van Erp NP, Gelderblom H, Karlsson MO, Li J, Zhao M, et al. (2007) Influence of CYP3A4 inhibition on the steady-state pharmacokinetics of Imatinib. Clin Cancer Res 13: 7394-400.
14. Dutreix C, Peng B, Mehring G, Hayes M, Capdeville R, et al. (2004) Pharmacokinetic interaction between ketoconazole and Imatinib mesylate (Glivec) in healthy subjects. Cancer Chemother Pharmacol 54: 290-4.
15. Bolton AE, Peng B, Hubert M, Krebs-Brown A, Capdeville R, et al. (2003) Effect of rifampicin on the pharmacokinetics of Imatinib mesylate (Gleevec, STI571) in healthy subjects. Cancer Chemother Pharmacol 53: 102-6.
16. Parrilla-Campiglia S, Ercoli MC, Umpierrez O, Rodriguez P, Márquez S, et al. (2009) Bioequivalence of two film-coated tablets of imatinib mesylate 400 mg: a randomized, open-label, single-dose, fasting, two-period, two-sequence crossover comparison in healthy male South American volunteers. Clin Ther 31: 2224-32.



Only ~1% of Genome is a Drug Target

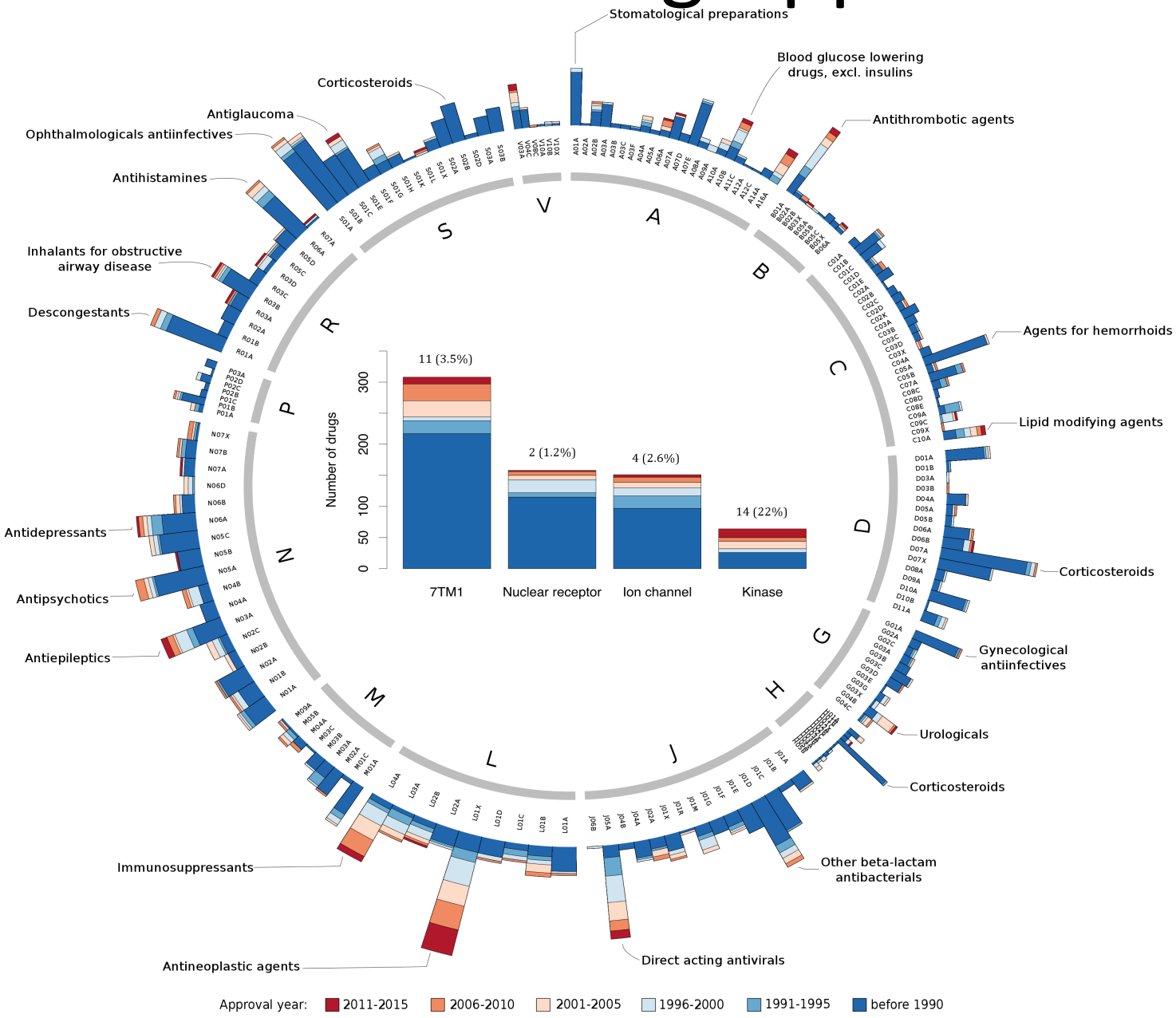


Drug Targets and Drugs

	Targets			Drugs		
Drug target Class	**Total targets**	**Small-molecule drug targets**	**Biotherapeutic drug target**	**Total drugs**	**Small molecules**	**Biotherapeutics**
Human Protein	315	243	86	1133	951	182
Pathogen Protein	52	49	4	205	200	5
Other human biomolecules	15	3	13	75	50	25
Other pathogen biomolecules	8	7	2	102	99	3

Santos *et al*, *unpublished*

Innovation in Drug Approvals

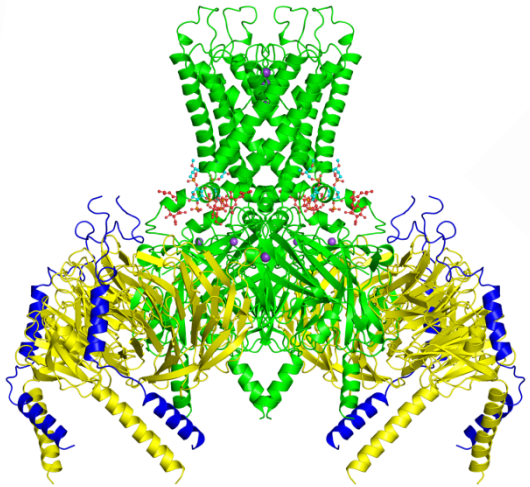


Privileged Target Families

Rhodopsin-like GPCR
PDBe: 3sn6

22% of drug targets
33% of small mol drugs

Ion channels
PDBe: 4kfm

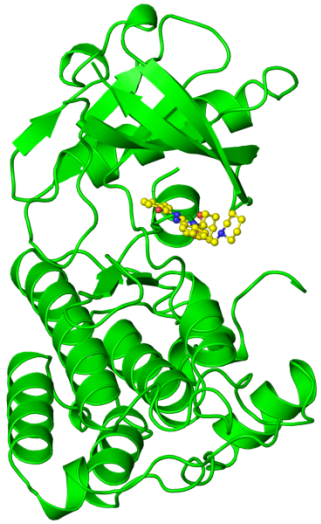


12% of drug targets
18% of small mol drugs

Nuclear receptors
PDBe: 3e00

6% of drug targets
17% of small mol drugs

Protein kinases
PDBe: 4foc

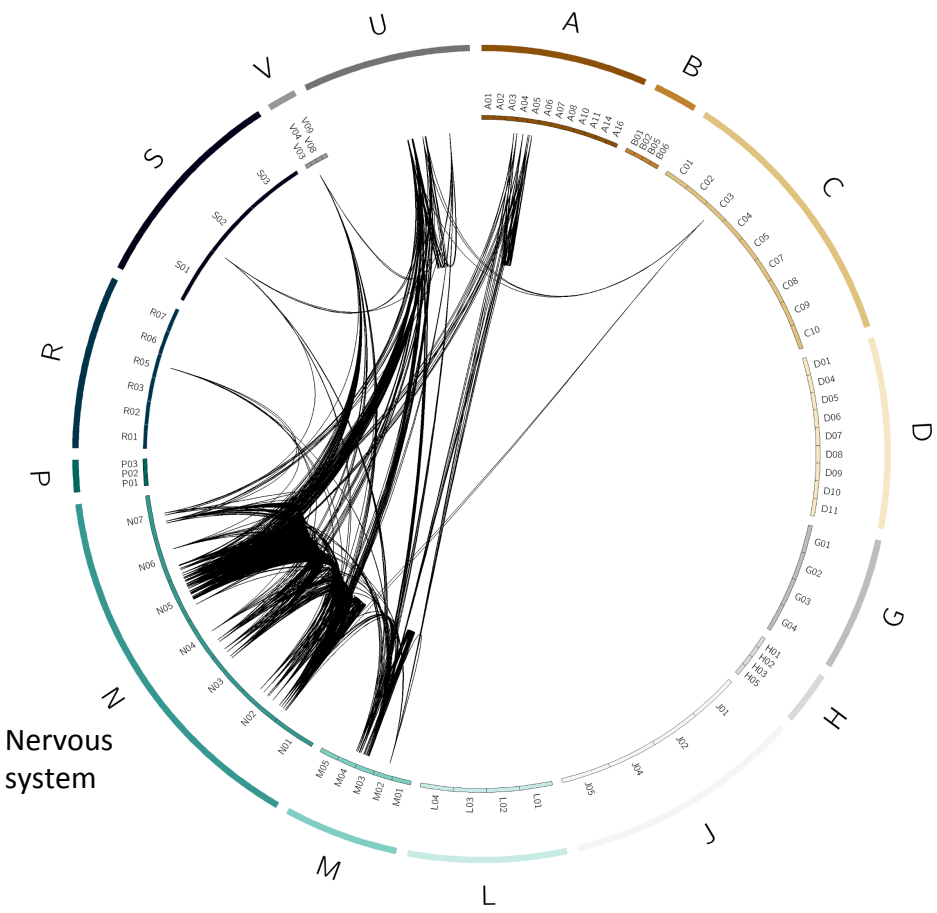


13% of drug targets
2.4% of small mol drugs

Over 53% of all targets and 70% of drugs modulate these four target classes

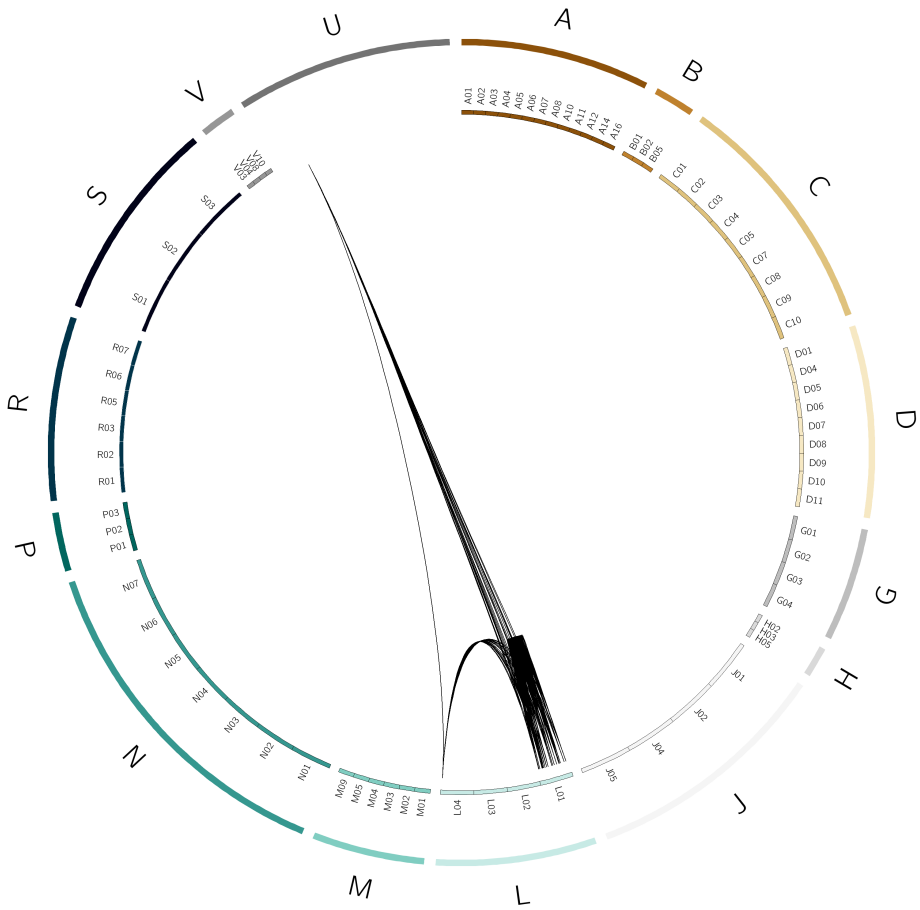
Footprint of Target Classes Across Disease

Ligand-gated ion channels



Nervous system

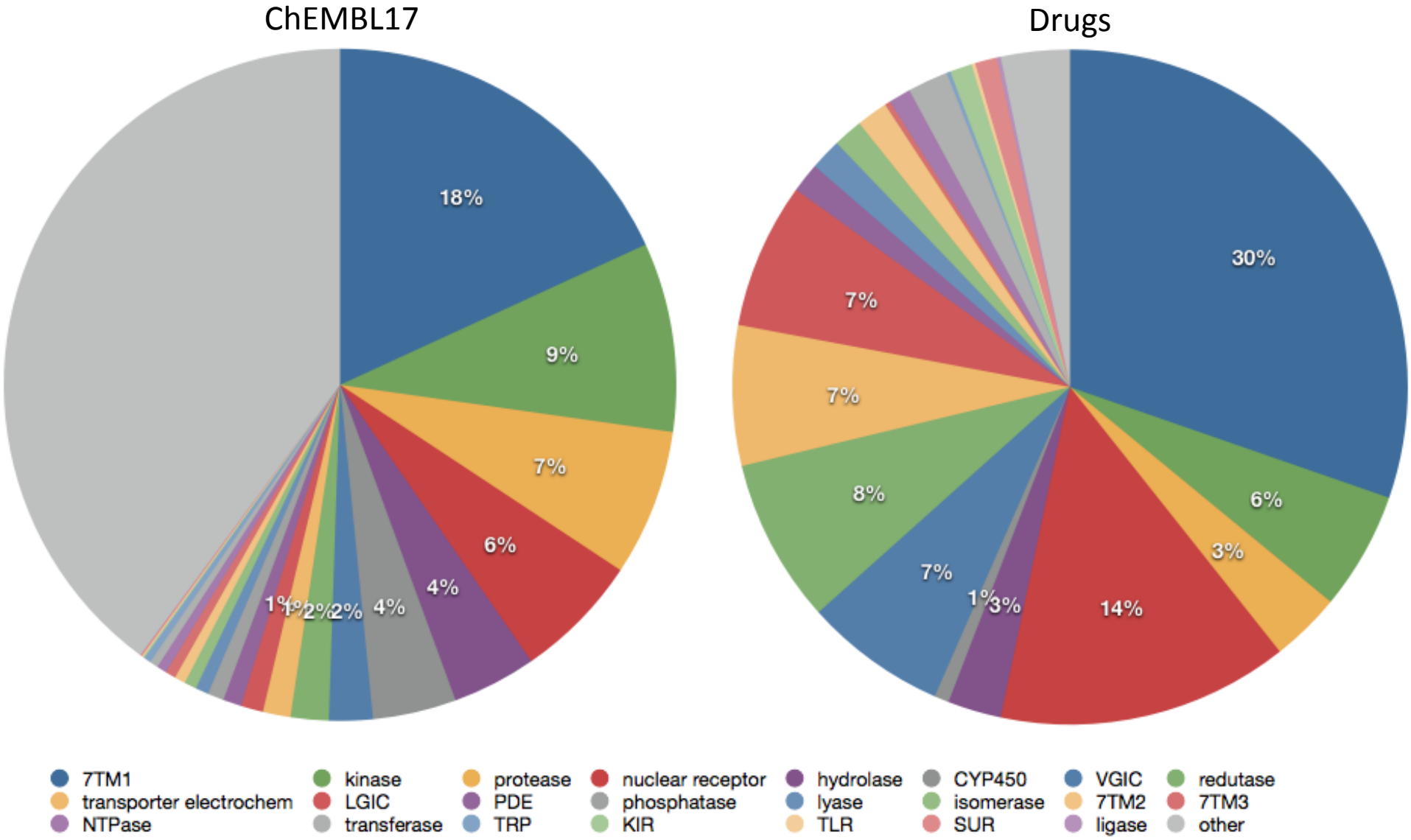
Protein kinases



Cancer and inflammation

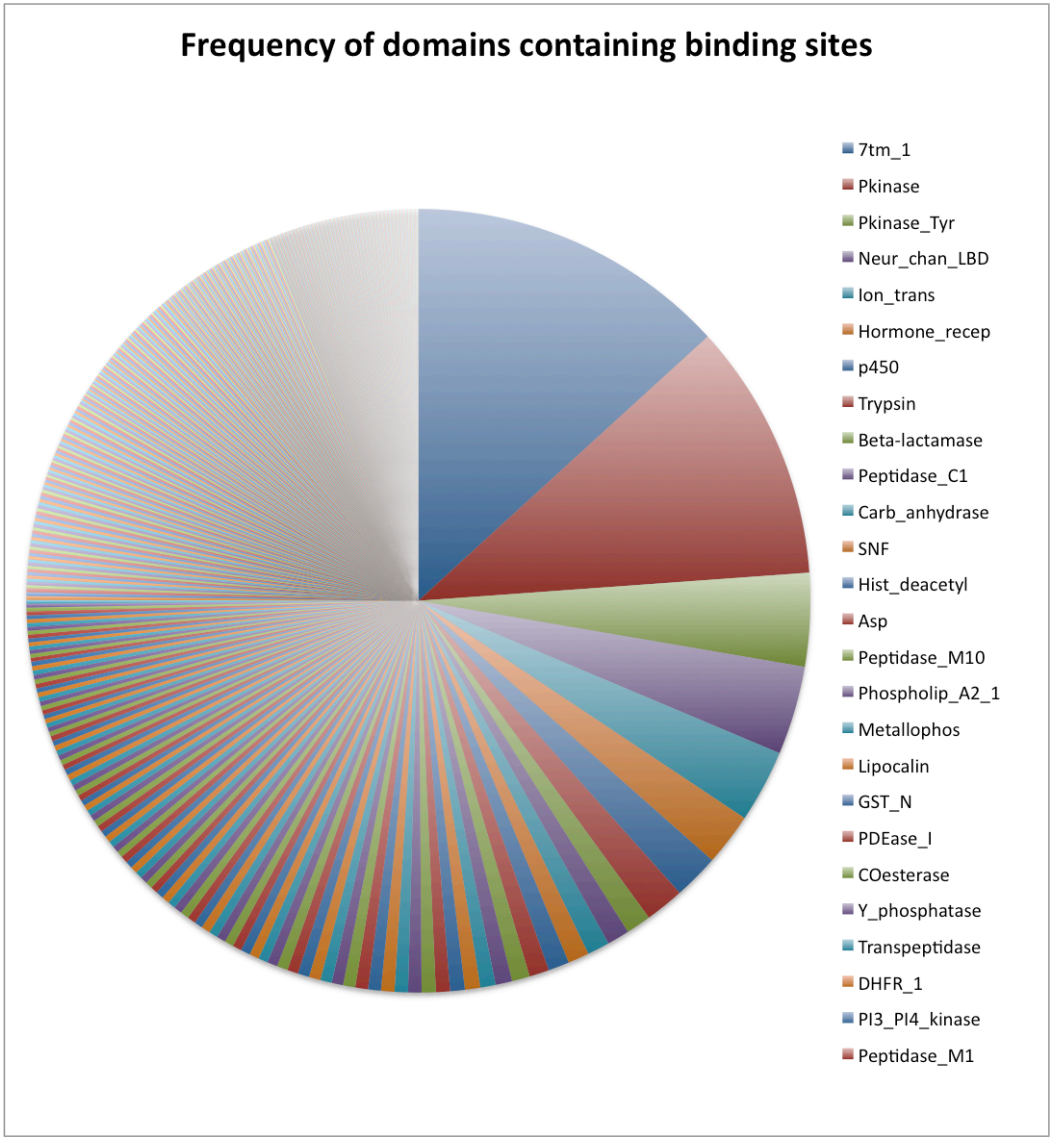
Santos *et al*, unpublished

Privileged Target Families



Santos, *unpublished*

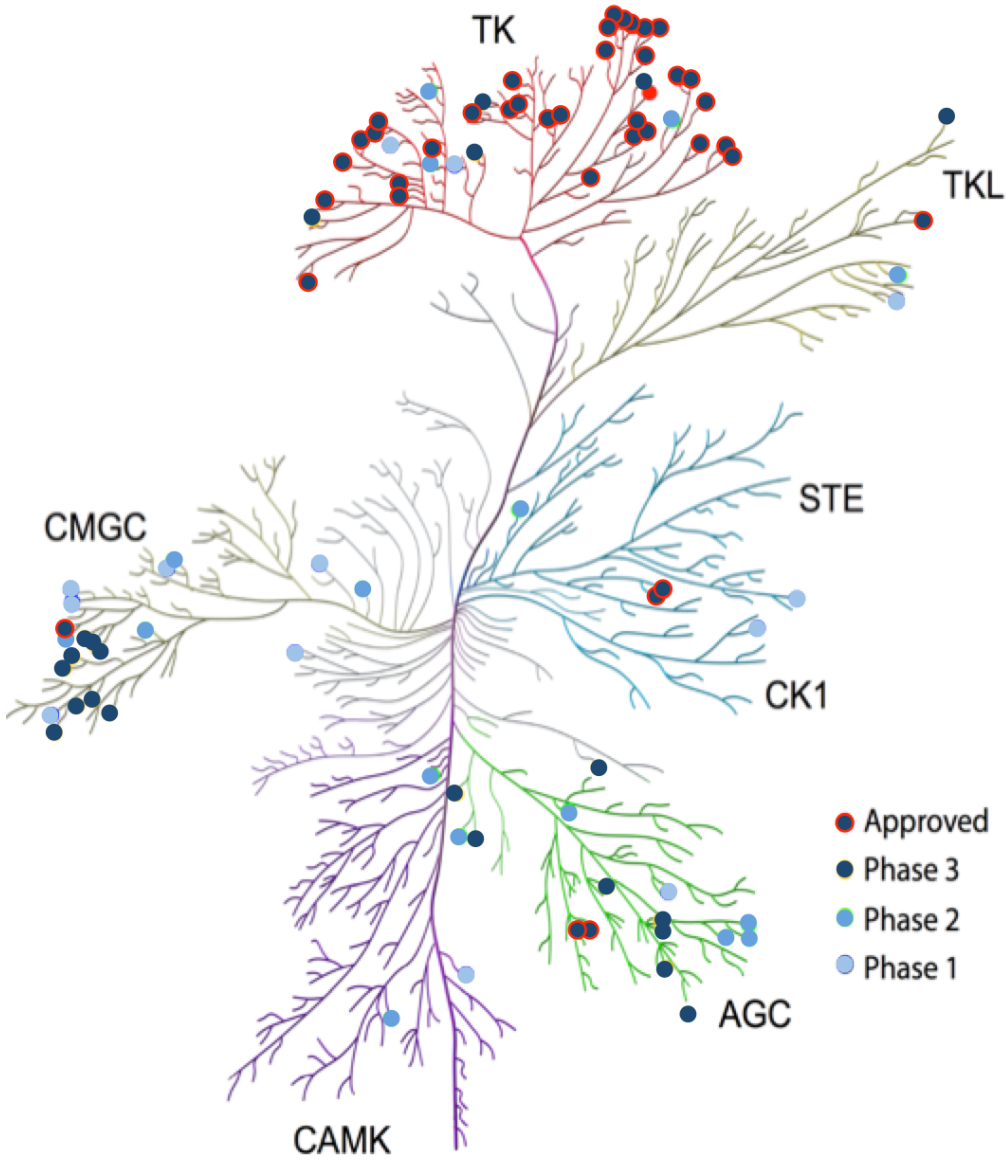
Domain-level Binding Site Annotation



Depleted and Enriched Pfam Domains

Neur_chan_memb	-1.63
zf-C4	-0.94
ANF_receptor	-0.88
SH2	-0.83
Pkinase_C	-0.70
fn3	-0.53
SH3_1	-0.51
Lig_chan	-0.50
C2	-0.50
C1_1	-0.50
Guanylate_cyc	-0.46
HATPase_c	-0.46
I-set	-0.44
adh_short	-0.39
PH	-0.39
Ank	-0.39
…..	
Metallophos	0.35
Phospholip_A2_1	0.38
Peptidase_M10	0.41
Asp	0.45
SNF	0.48
Hist_deacetyl	0.48
Carb_anhydrase	0.50
Peptidase_C1	0.51
Trypsin	0.51
Beta-lactamase	0.57
p450	1.00
Hormone_recep	1.19
Ion_trans	1.66
Neur_chan_LBD	2.02
Pkinase_Tyr	2.12
Pkinase	5.87
7tm_1	7.30

Clinical Kinome

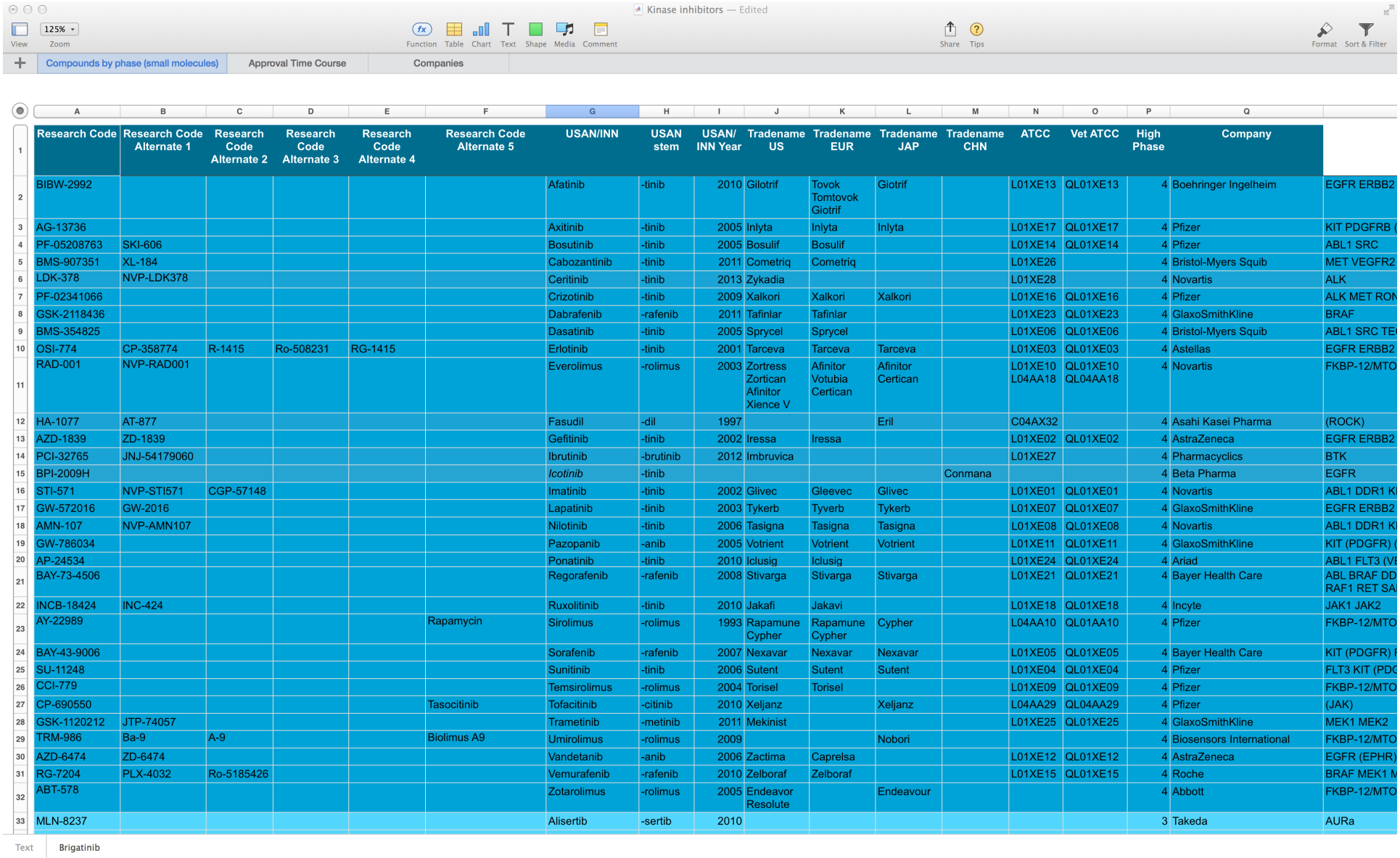


Overington, Al-Lazikani & Wennerberg, *unpublished*

Clinical Kinome

- 399 Clinical stage human kinase inhibitors
  - 31 Approved small molecule kinase inhibitors
    - 17 -tinib – tyrosine kinase inhibitors
    - *5 -rolimus – mTor inhibitors*
    - 4 -rafenib – Raf inhibitors
    - 2 -anib – angiogenesis inhibitors
    - 1 -metinib – met inhibitor
    - 1 brutinib – Bruton tyrosine kinase inhbitors
    - 1 -dil – Rho kinase inhibitor (Japan only)
  - 38 Phase 3
  - 143 Phase 2
  - 189 Phase 1
    - Phase 1:2 ratio is atypical due to many kinase inhibitor trials being phase 1/2 oncology trials

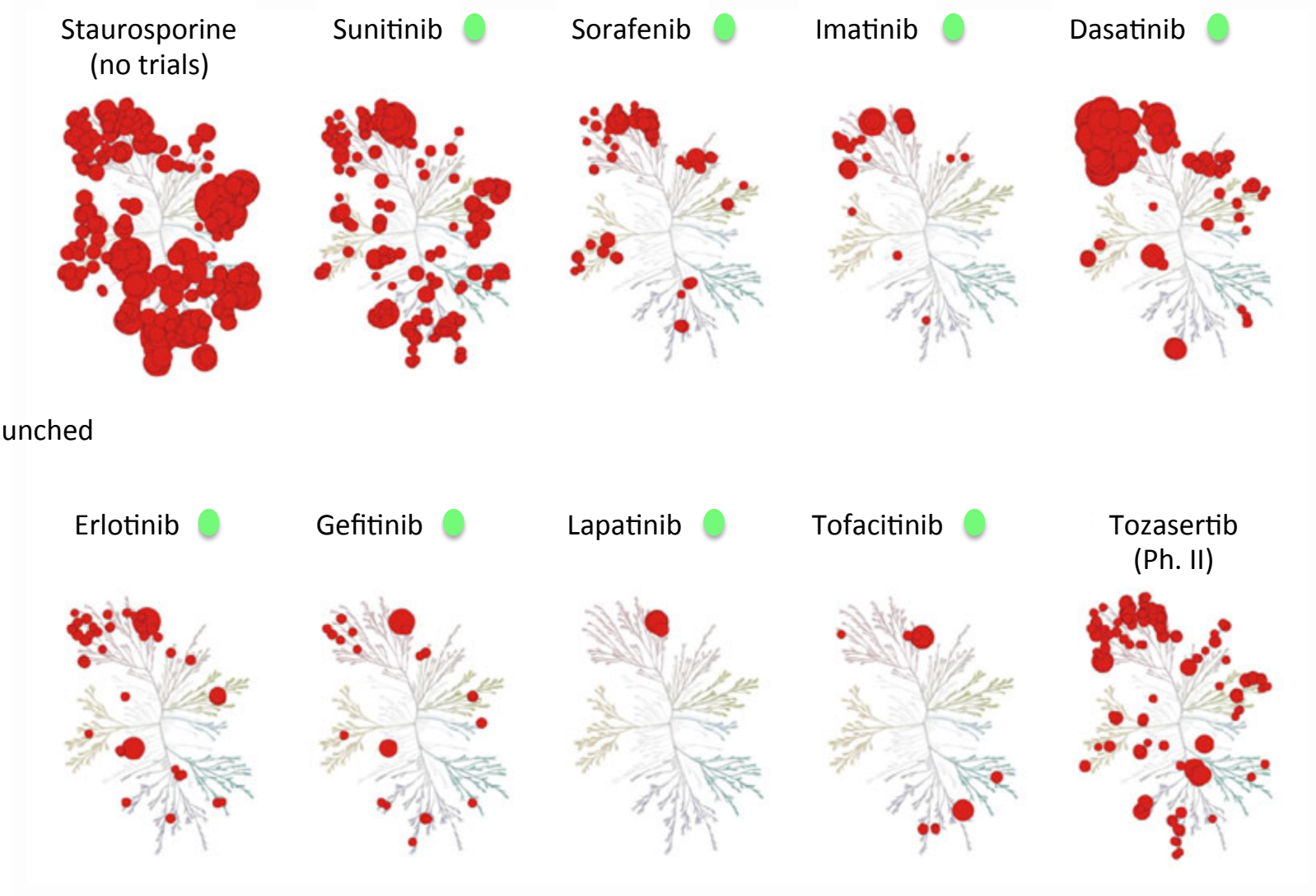
Kinase Inhibitors in Clinical Development



Research Code	Research Code Alternate 1	Research Code Alternate 2	Research Code Alternate 3	Research Code Alternate 4	Research Code Alternate 5	USAN/INN	USAN stem	USAN/ INN Year	Tradename US	Tradename EUR	Tradename JAP	Tradename CHN	ATCC	Vet ATCC	High Phase	Company	
BIBW-2992						Afatinib	-tinib	2010	Gilotrif	Tovok Tomtovok Giotrif	Giotrif		L01XE13	QL01XE13	4	Boehringer Ingelheim	EGFR ERBB2
AG-13736						Axitinib	-tinib	2005	Inlyta	Inlyta	Inlyta		L01XE17	QL01XE17	4	Pfizer	KIT PDGFRB (
PF-05208763	SKI-606					Bosutinib	-tinib	2005	Bosulif	Bosulif			L01XE14	QL01XE14	4	Pfizer	ABL1 SRC
BMS-907351	XL-184					Cabozantinib	-tinib	2011	Cometriq	Cometriq			L01XE26		4	Bristol-Myers Squib	MET VEGFR2
LDK-378	NVP-LDK378					Ceritinib	-tinib	2013	Zykadia				L01XE28		4	Novartis	ALK
PF-02341066						Crizotinib	-tinib	2009	Xalkori	Xalkori	Xalkori		L01XE16	QL01XE16	4	Pfizer	ALK MET RON
GSK-2118436						Dabrafenib	-rafenib	2011	Tafinlar	Tafinlar			L01XE23	QL01XE23	4	GlaxoSmithKline	BRAF
BMS-354825						Dasatinib	-tinib	2005	Sprycel	Sprycel			L01XE06	QL01XE06	4	Bristol-Myers Squib	ABL1 SRC TE
OSI-774	CP-358774	R-1415	Ro-508231	RG-1415		Erlotinib	-tinib	2001	Tarceva	Tarceva	Tarceva		L01XE03	QL01XE03	4	Astellas	EGFR ERBB2
RAD-001	NVP-RAD001					Everolimus	-rolimus	2003	Zortress Zortican Afinitor Xience V	Afinitor Votubia Certican	Afinitor Certican		L01XE10 L04AA18	QL01XE10 QL04AA18	4	Novartis	FKBP-12/MTO
HA-1077	AT-877					Fasudil	-dil	1997			Eril		C04AX32		4	Asahi Kasei Pharma	(ROCK)
AZD-1839	ZD-1839					Gefitinib	-tinib	2002	Iressa	Iressa			L01XE02	QL01XE02	4	AstraZeneca	EGFR ERBB2
PCI-32765	JNJ-54179060					Ibrutinib	-brutinib	2012	Imbruvica				L01XE27		4	Pharmacyclics	BTK
BPI-2009H						*Icotinib*	-tinib					Conmana			4	Beta Pharma	EGFR
STI-571	NVP-STI571	CGP-57148				Imatinib	-tinib	2002	Glivec	Gleevec	Glivec		L01XE01	QL01XE01	4	Novartis	ABL1 DDR1 K
GW-572016	GW-2016					Lapatinib	-tinib	2003	Tykerb	Tyverb	Tykerb		L01XE07	QL01XE07	4	GlaxoSmithKline	EGFR ERBB2
AMN-107	NVP-AMN107					Nilotinib	-tinib	2006	Tasigna	Tasigna	Tasigna		L01XE08	QL01XE08	4	Novartis	ABL1 DDR1 K
GW-786034						Pazopanib	-anib	2005	Votrient	Votrient	Votrient		L01XE11	QL01XE11	4	GlaxoSmithKline	KIT (PDGFR) (
AP-24534						Ponatinib	-tinib	2010	Iclusig	Iclusig			L01XE24	QL01XE24	4	Ariad	ABL1 FLT3 (VE
BAY-73-4506						Regorafenib	-rafenib	2008	Stivarga	Stivarga	Stivarga		L01XE21	QL01XE21	4	Bayer Health Care	ABL BRAF DD RAF1 RET SA
INCB-18424	INC-424					Ruxolitinib	-tinib	2010	Jakafi	Jakavi			L01XE18	QL01XE18	4	Incyte	JAK1 JAK2
AY-22989					Rapamycin	Sirolimus	-rolimus	1993	Rapamune Cypher	Rapamune Cypher	Cypher		L04AA10	QL01AA10	4	Pfizer	FKBP-12/MTO
BAY-43-9006						Sorafenib	-rafenib	2007	Nexavar	Nexavar	Nexavar		L01XE05	QL01XE05	4	Bayer Health Care	KIT (PDGFR) F
SU-11248						Sunitinib	-tinib	2006	Sutent	Sutent	Sutent		L01XE04	QL01XE04	4	Pfizer	FLT3 KIT (PDG
CCI-779						Temsirolimus	-rolimus	2004	Torisel	Torisel			L01XE09	QL01XE09	4	Pfizer	FKBP-12/MTO
CP-690550					Tasocitinib	Tofacitinib	-citinib	2010	Xeljanz		Xeljanz		L04AA29	QL04AA29	4	Pfizer	(JAK)
GSK-1120212	JTP-74057					Trametinib	-metinib	2011	Mekinist				L01XE25	QL01XE25	4	GlaxoSmithKline	MEK1 MEK2
TRM-986	Ba-9	A-9			Biolimus A9	Umirolimus	-rolimus	2009			Nobori				4	Biosensors International	FKBP-12/MTO
AZD-6474	ZD-6474					Vandetanib	-anib	2006	Zactima	Caprelsa			L01XE12	QL01XE12	4	AstraZeneca	EGFR (EPHR)
RG-7204	PLX-4032	Ro-5185426				Vemurafenib	-rafenib	2010	Zelboraf	Zelboraf			L01XE15	QL01XE15	4	Roche	BRAF MEK1 M
ABT-578						Zotarolimus	-rolimus	2005	Endeavor Resolute		Endeavour				4	Abbott	FKBP-12/MTO
MLN-8237						Alisertib	-sertib	2010							3	Takeda	AURa

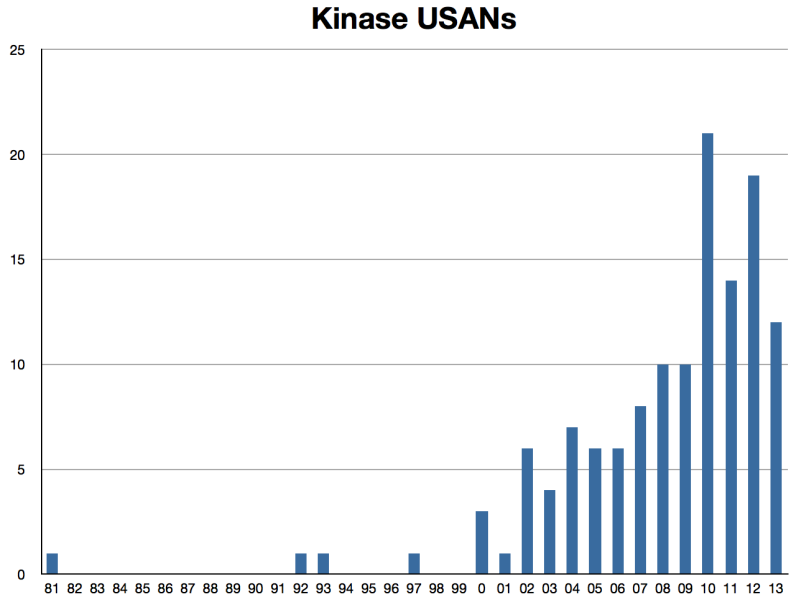
Overington, Bellis, Al-Lazikani & Wennerberg, *unpublished*

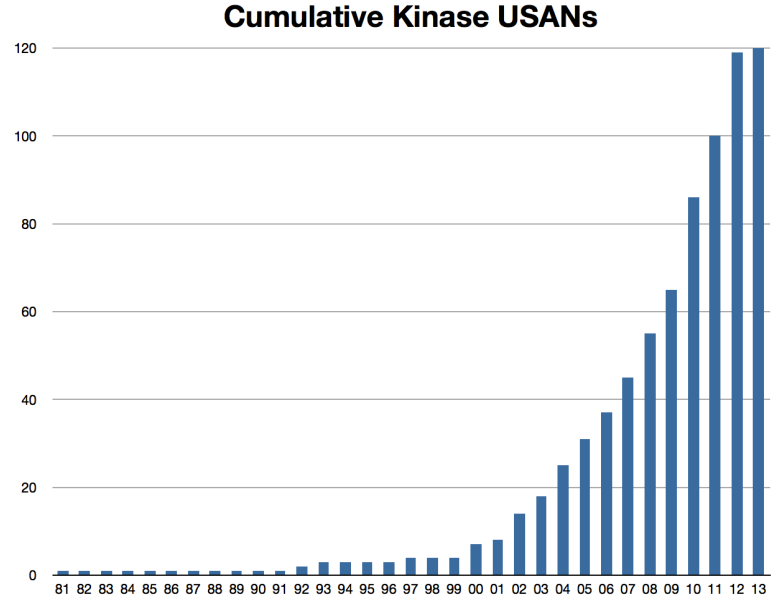
Kinase Inhibitor Polypharmacology

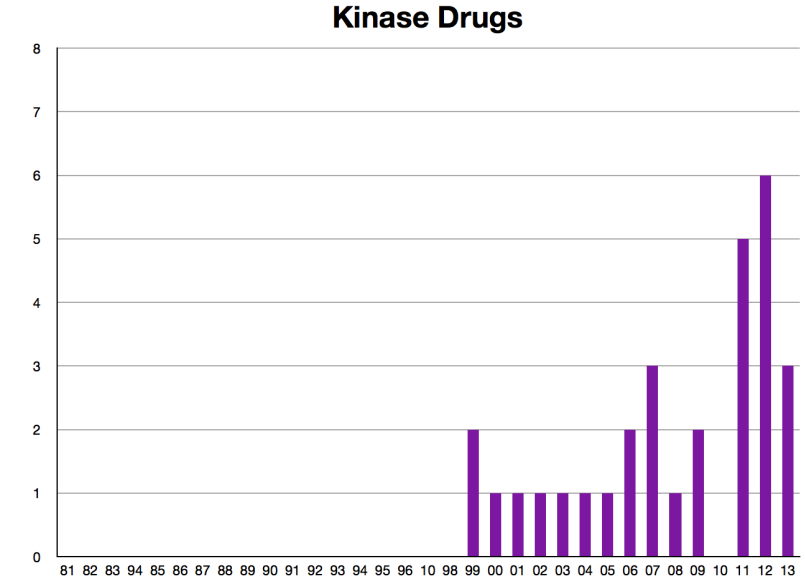


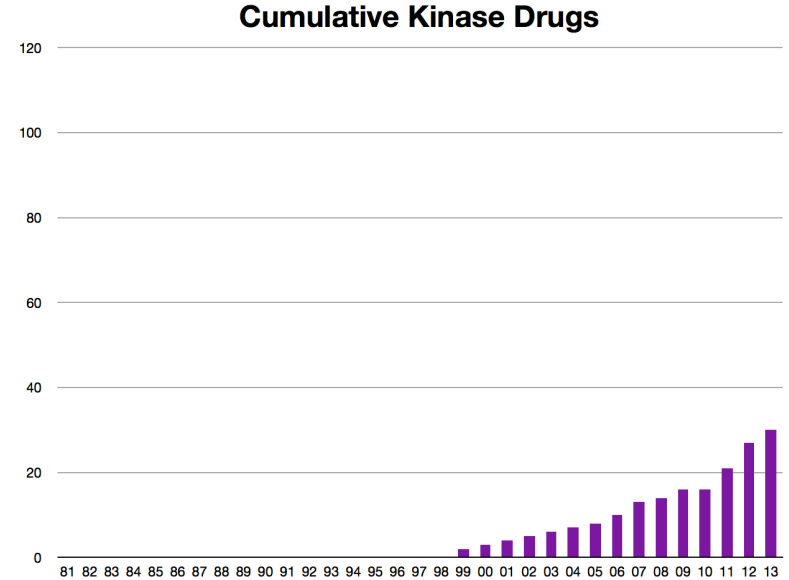
Adapted from Ghoreschi *et al*, *Nature Immunology* **10**, 356 - 360 (2009)

Kinase Inhibitor Attrition





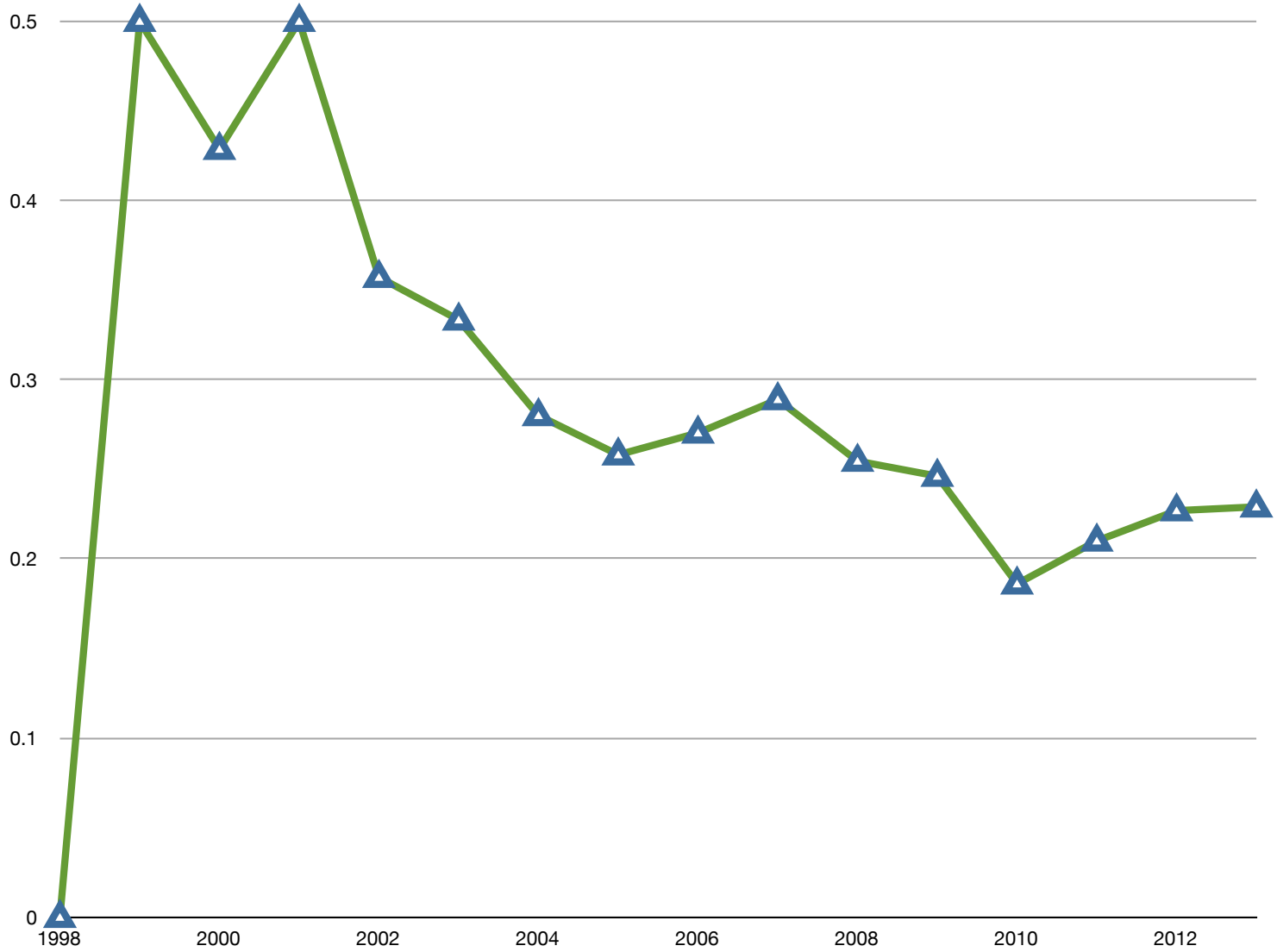




Overington, *unpublished*

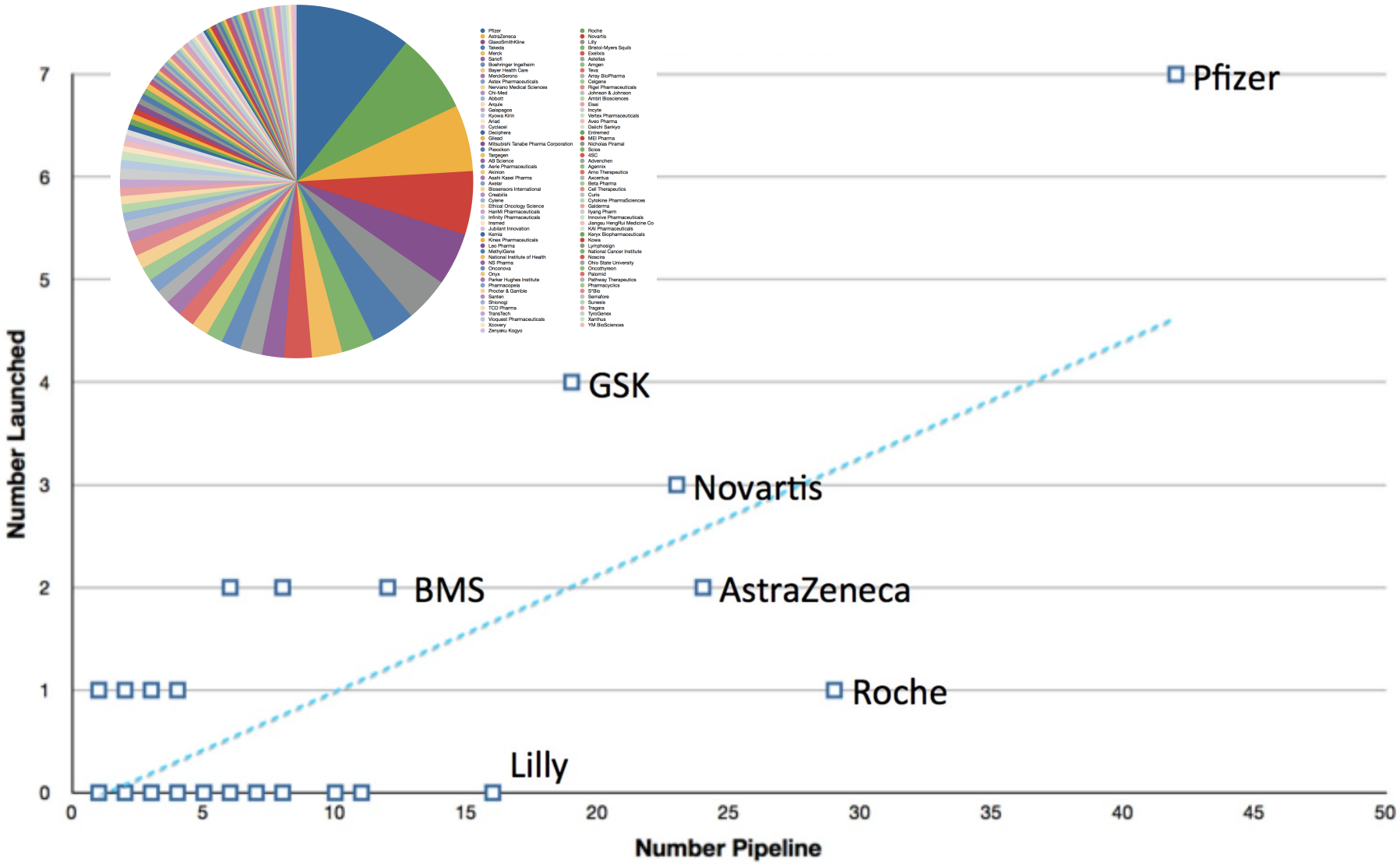
Kinase Inhibitor Attrition

USAN to approved fraction! – ~0.2 is long term mean for all drugs across all classes



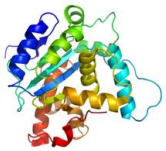
Overington, *unpublished*

Kinase Inhibitor Productivity

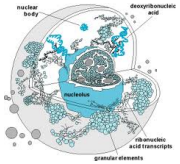


Overington, *unpublished*

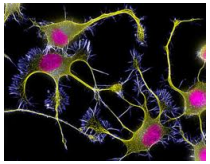
Modern Drug Discovery Assay Cascade



Biochemical assay



Cell-based screen



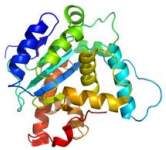
Functional assay

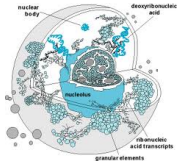
Animal disease model

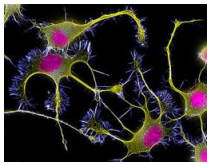
Human clinical trial

- Move from *'quick, low-cost, less predictive'* assays to '*slow, high-cost, more predictive'* assays
- Make selections on which compounds to progress to later assays on basis of positive activity in earlier screens
  - Early, cheap assays are used a lot of times, later expensive assays rarely
  - Attrition – failure of compounds in that pipeline

Efficacy Assay Cascade







Biochemical assay

Cell-based screen

Functional assay

Animal disease model

Human clinical trial

PPARγ

SGLT2

Calcitonin receptor

GLP-1

DPP-IV

...

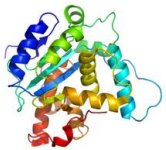
...

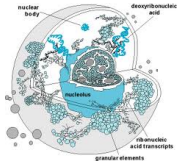
← Mechanism →

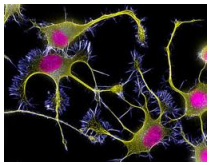
Target

← Disease →

Efficacy Assay Cascade







Biochemical assay

Cell-based screen

Functional assay

Animal disease model

Human clinical trial

PPARγ

SGLT2

Calcitonin receptor

GLP-1

DPP-IV

...

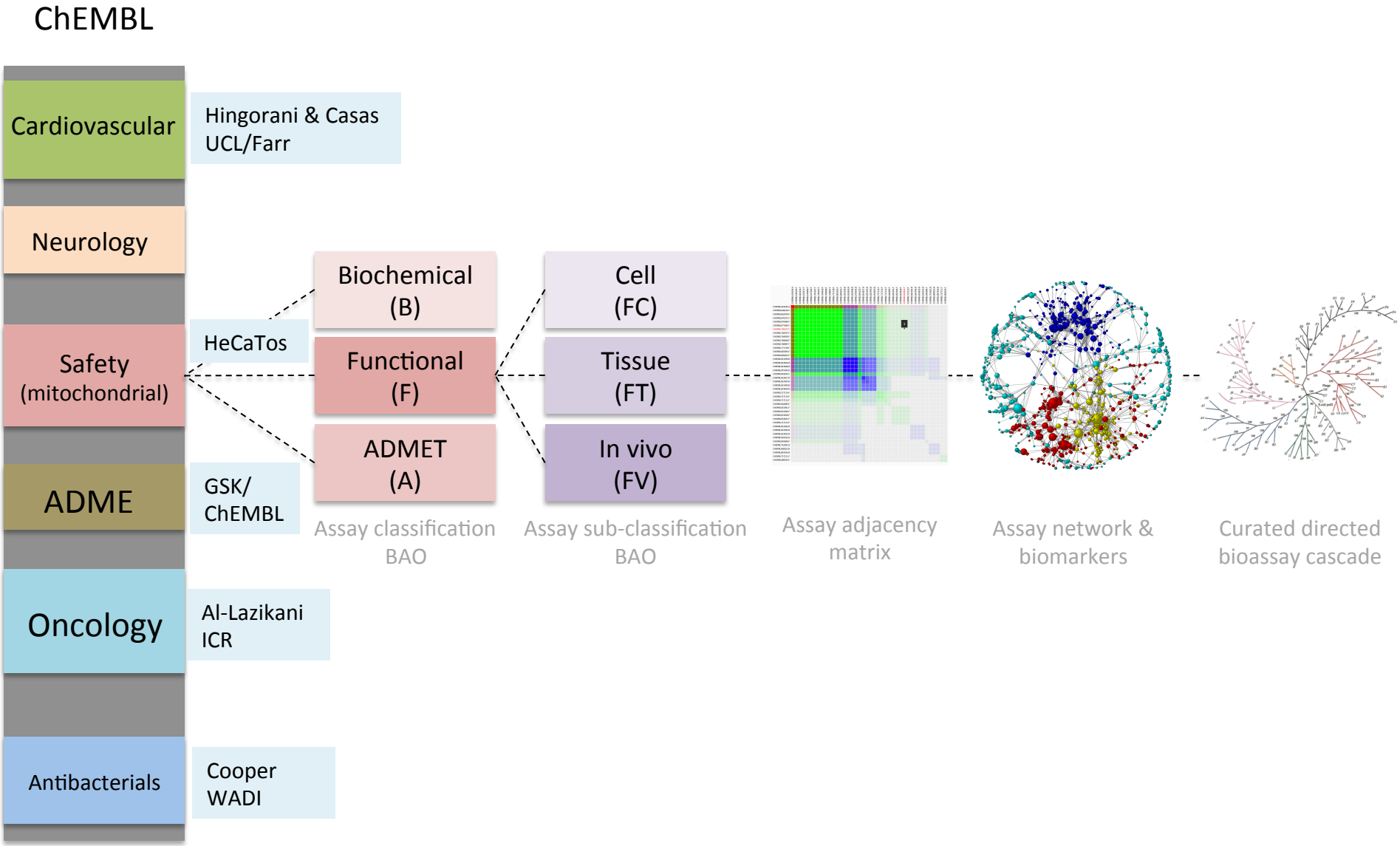
...

Mechanism

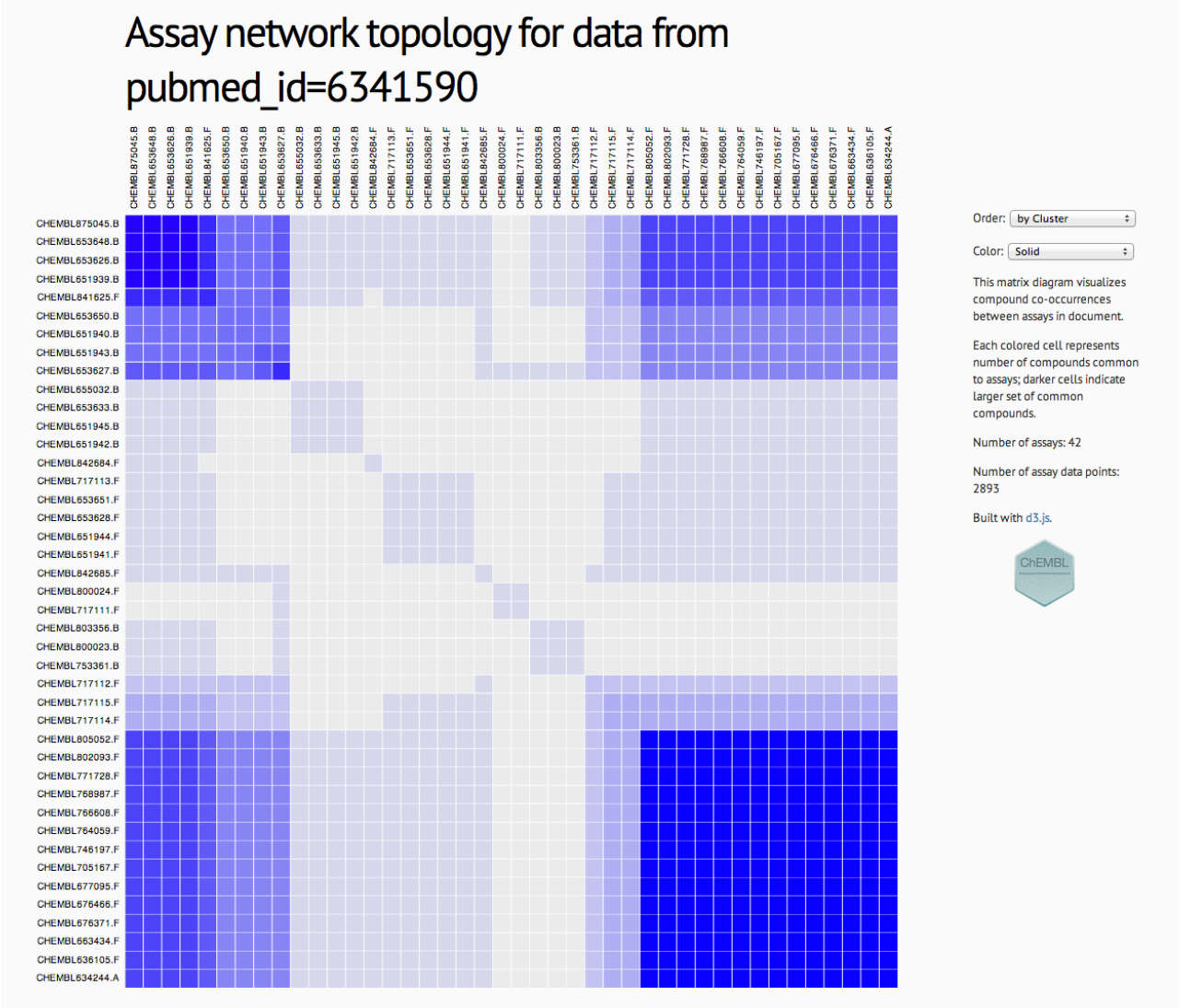
Target

Disease

ChEMBL Bioassay Annotation Strategy

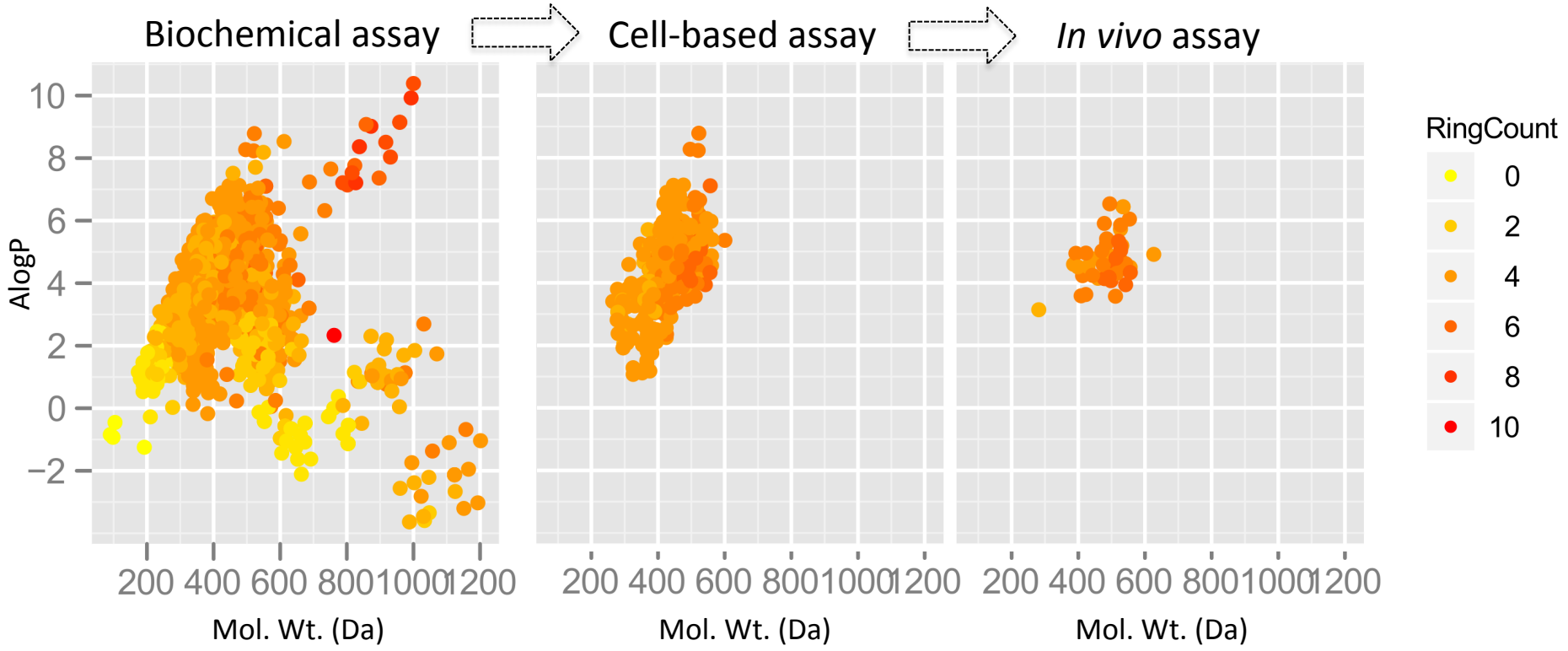


Assay Networks – Adjacency Matrix



EGFR Assay Cascades from ChEMBL

Physicochemical properties for cSrc EGFR pathway inhibitors



F. Krueger (unpublished)

SureChEMBL

- EMBL-EBI acquired the SureChem product from Digital Science
  - 15 million chemical structures
  - Automatically extracted chemical structures from full-text patent
- Research community wants open access to patent data
  - Patent literature 2-3 years ahead of published literature
  - Better competitive position
- Plan to provide ongoing free, Open resource to entire community

SureChEMBL Data Coverage

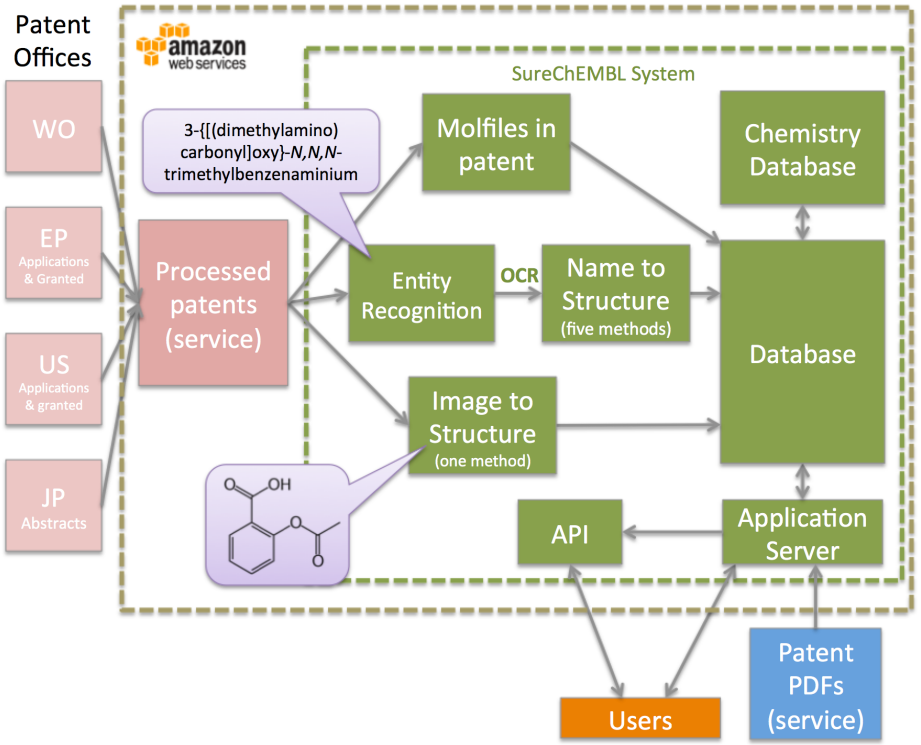
	Data	Description & Languages	Years
EP applications	Bib. data Full text	DocDB + Original Original (EN, DE, FR)	from 1978
EP granted	Bib. data Full text	DocDB + Original Original (EN, DE, FR)	From 1980
WO applications	Bib. data Full text	DocDB + Original Original (EN, DE, FR, ES, RU)	From 1978 From 1978
US applications	Bib. data Full text	DocDB + Original Original (EN)	From 2001 From 2001
US granted	Bib. data Full text	DocDB + Original Original (EN)	From 1920 From 1976
JP applications	Bib. data Full text	DocDB PAJ - English abstracts/titles	From 1973 From 1976
JP granted	Bib. data	DocDB	From 1994
90+ countries	Bib. data	DocDB	From 1920

SureChEMBL Chemistry Data Coverage

- Exemplified structures from patent title, description, abstract and claims
- Structures from text 1976 onwards
- Structures from images 2007 onwards
- USPTO have provided 'Complex Work Units' since 2001
  - CWU file types include MOL and CDX
  - CWUs processed as part of pipeline

SureChEMBL Plans

- UniChem/ChEMBL integration
- Enhanced entity extraction
  - Tag commonly used identifiers UniProt, ChEMBL, ChEBI, PubChem, PDB, Research codes
  - Proteins, diseases, animal models, cell lines, assays
- New search methods
  - Protein sequence-based searching, including biotherapeutics
- IMI Open PHACTS funding
  - refactor SureChEMBL API
  - RDF Conversion and semantic tagging -> *EBI RDF Platform*
- Status
  - Database running and updated
  - Legacy API running
  - Front-end running
  - Legacy user account system removed

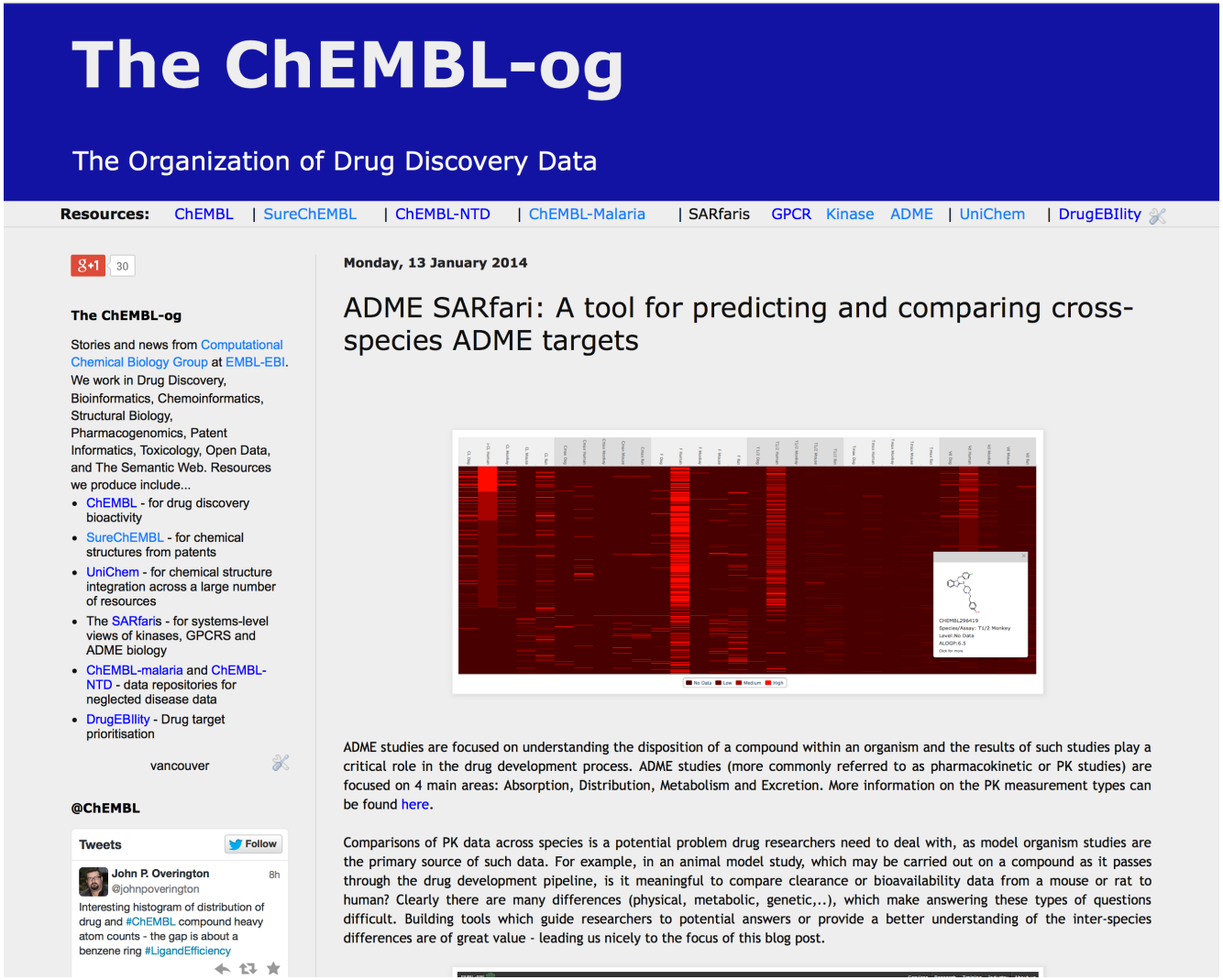


Acknowledgements

ChEMBL Database
Anne Hersey
Anna Gaulton
Mark Davies
Michal Nowotka
George Papadatos
Jon Chambers
Louisa Bellis
Rita Santos
Felix Krueger
Gerard Van Westen
Ruth Akhtar
Francis Atkinson
Patricia Bento
Ramesh Donadi
John Paul Overington

Institute of Cancer Research
Bissan Al-Lazikani
Paul Workman

FIMM, Helsinki
Krister Wennerberg



http://chembl.blogspot.com