b-it LSi
Life
Science
Informatics

LIMES
Life & Medical Sciences Institute

Analyzing Molecular Promiscuity from a Ligand and Target Perspective

Jürgen Bajorath
Life Science Informatics
University of Bonn

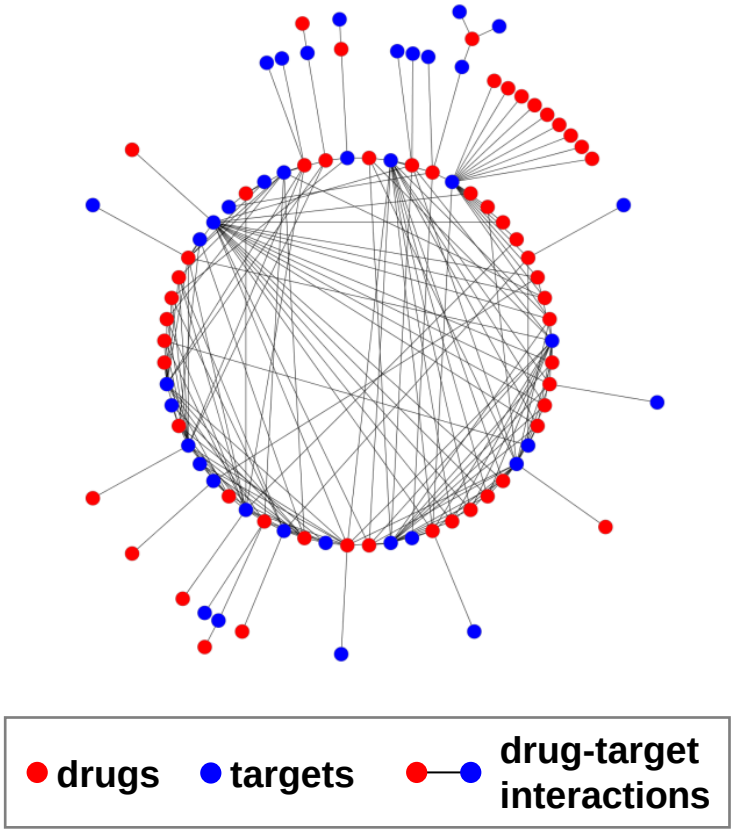
universitätbonn

Polypharmacology and Promiscuity

- Polypharmacology:

  an emerging theme in drug discovery

  - therapeutic efficacy of drug molecules often results from interactions with multiple targets
  - paradigm: ATP-site directed kinase inhibitors used in oncology



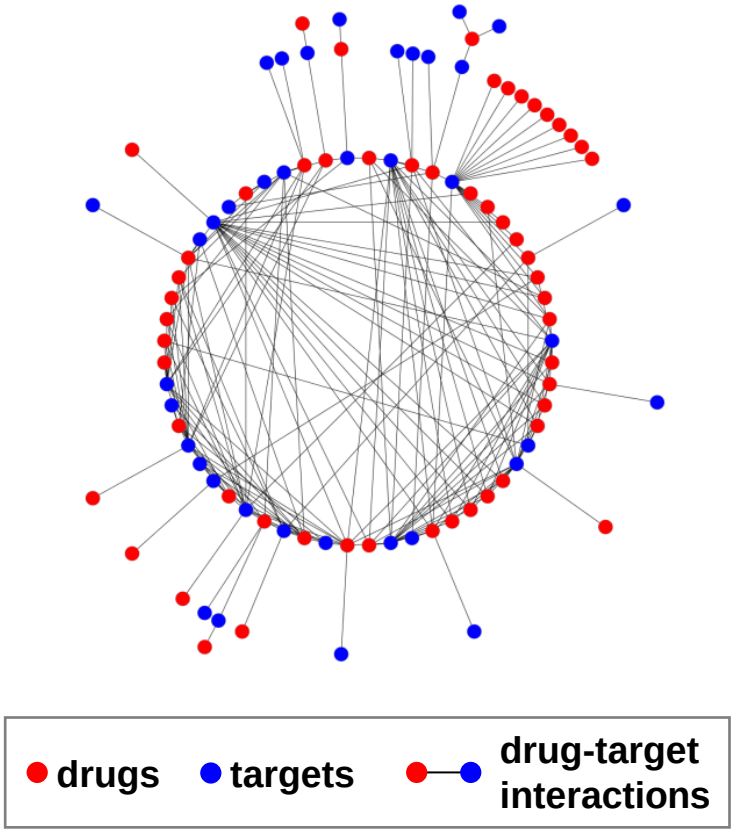
Polypharmacology and Promiscuity

- Promiscuity:

  molecular basis of polypharmacology

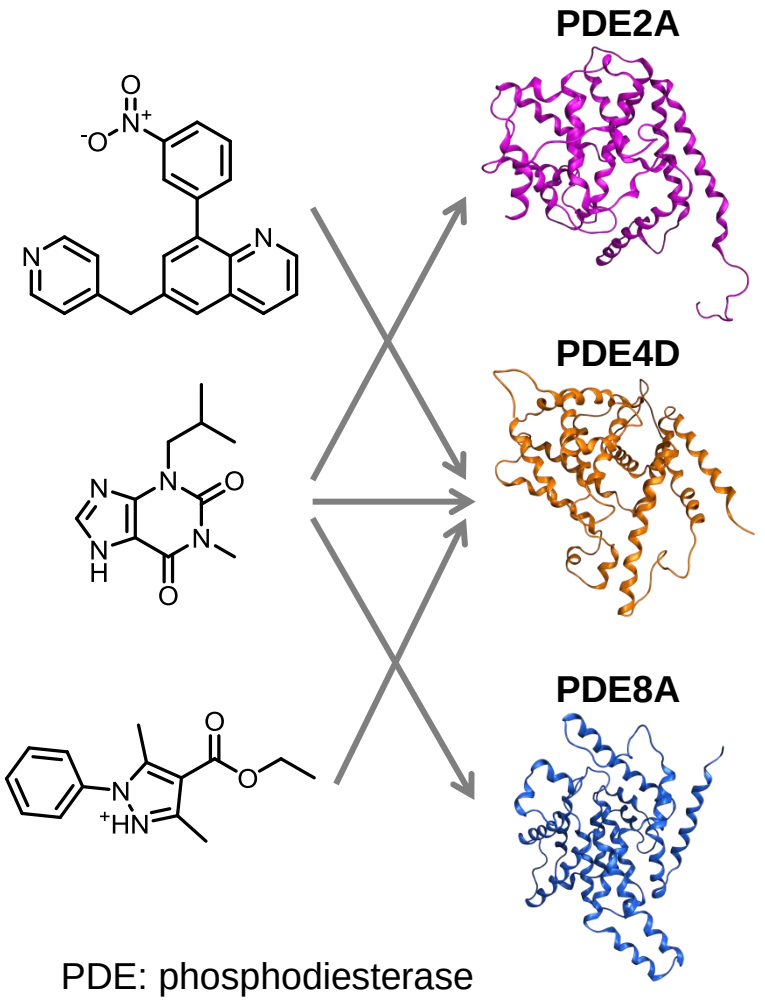
  - originally defined as the ability of small molecules to specifically interact with multiple targets

  - triggering a departure from the single-target concept in drug discovery



Different Views of Promiscuity

- Ligand-centric view
  - ability of small molecules to specifically interact with multiple targets
- Target-centric view
  - ability of proteins to bind different classes of compounds



PDE: phosphodiesterase

Large-Scale Promiscuity Analysis

- Systematically evaluate all currently available compounds and activity data from medicinal chemistry and biological screening

- Advent of ’Big Data’ in medicinal chemistry provides an unprecedented basis – and considerable challenges

Entering the 'Big Data' Era in Chemistry

Public Database	Organization	CPDs/Structures (Million, Aug./Nov. 2015)
ZINC 14	UCSF	**23**
ZINC 15		**CPDs ≤ 1000 Da** **Collected: 220 (!)** **'Drug-like' purchasable: >120 (!!)**

‘Big Data’ Criteria

5 ‘V‘s*

Volume

Velocity

Variety

Veracity

Value

Complexity

Heterogeneity

Confidence

*Lusher, S. J. *et al. Drug Discov. Today.* **2014**, *19*, 859

Data Sets with Varying Confidence Levels

Compound data sets from ChEMBL18

All available activity data — Set 1: 1,291,676

↓

Target organism (*homo sapiens*) — Set 2: 936,924

↓

Assay relationship type (***Direct***); **Assay confidence score** (*9*) — Set 3: 605,206

↓

Target class (*single protein*) — Set 4: 605,056

↓

Measurement types (K_i *or* IC_{50}) — Set 5: 179,161

↓

Standard activity relation (=) — Set 6: 150,379

↓

Standard activity unit (*nM*) — Set 7: 150,211

↓

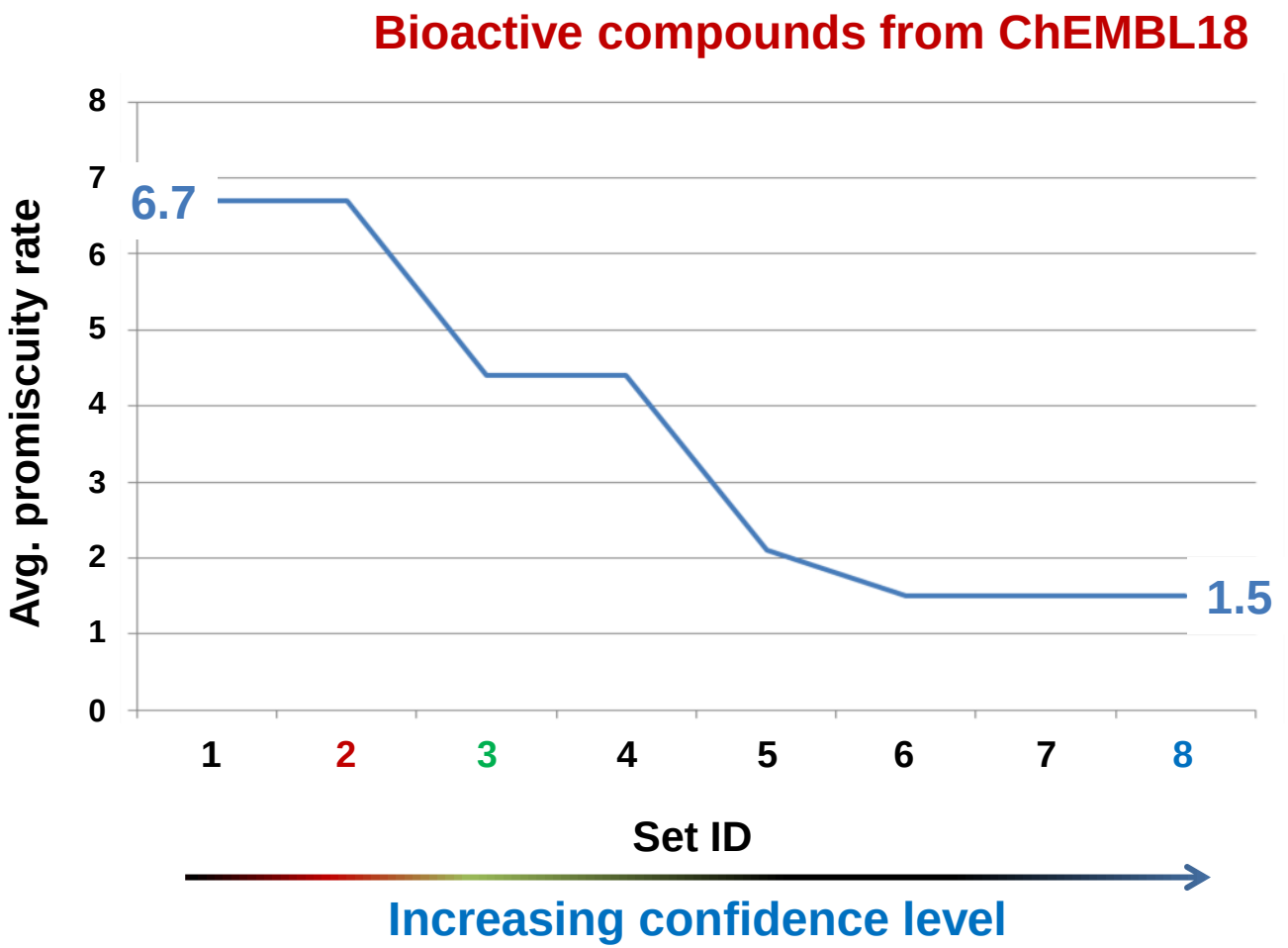
Activity comments (*remove 'inactives'*) — Set 8: 148,373

Increasing stringency
Increasing data confidence level

Promiscuity in Light of Data Confidence

- Promiscuity analysis illustrates the impact of data confidence criteria
- Ligand-centric view

Promiscuity vs. Data Confidence

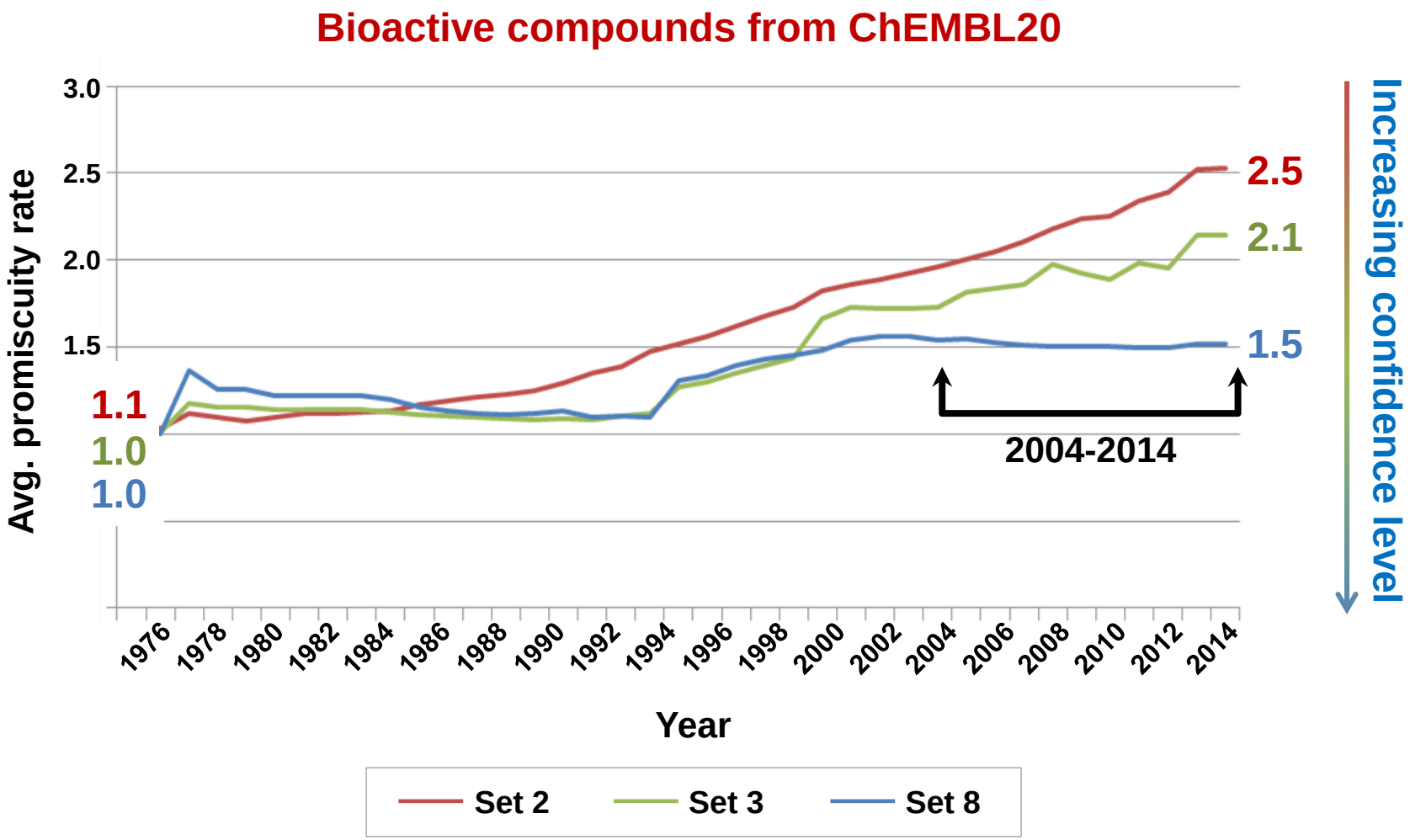


Hu, Y.; Bajorath, J. *J. Chem. Inf. Model.* **2014**, *54*, 3056

Balanced View on Promiscuity

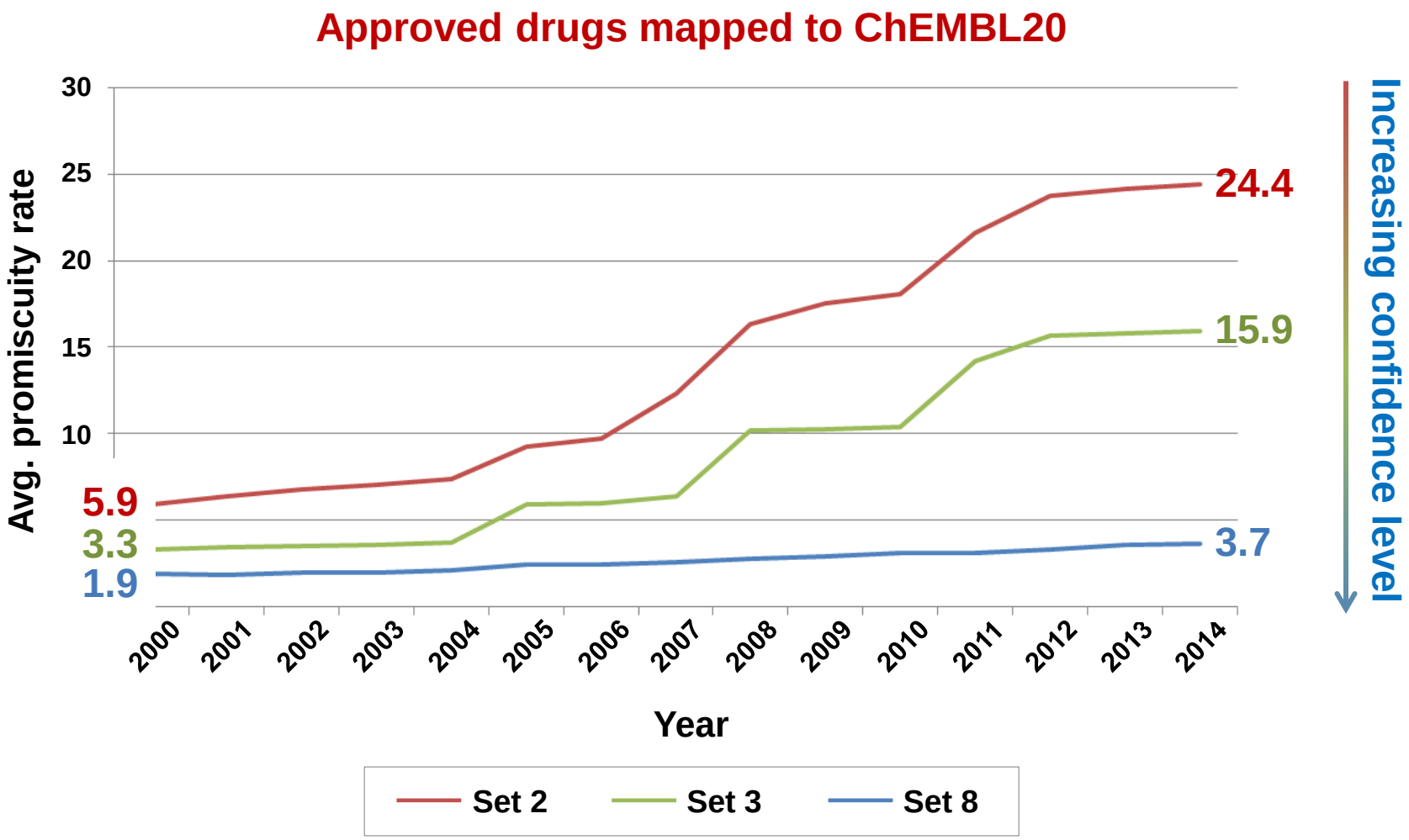
- Data sparseness principally results in conservative estimates
- Given current activity data volumes, statistically sound trends are anticipated
- Focusing on high-confidence data limits false-positive annotations

Compound Promiscuity Over Time

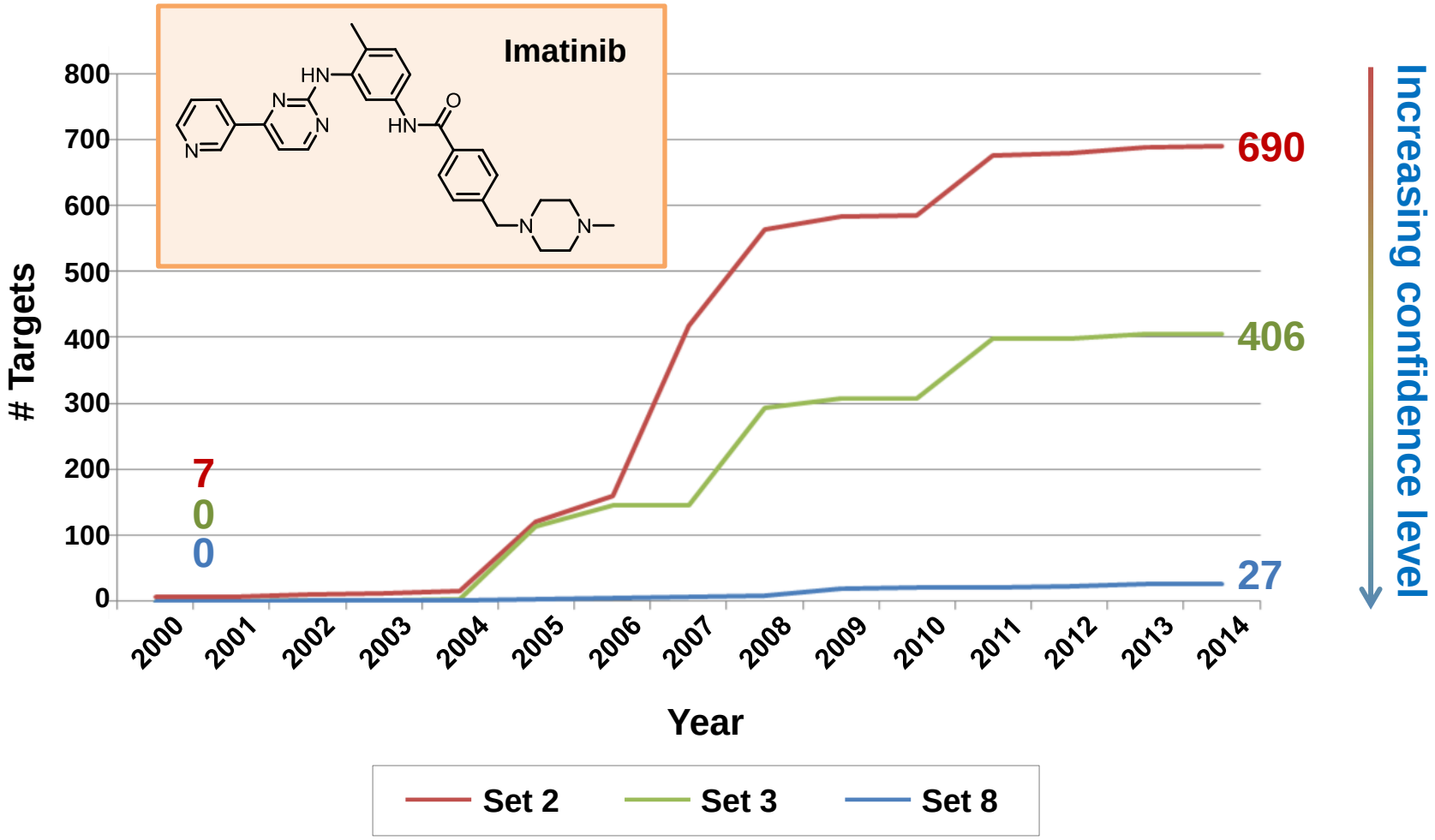


Hu, Y.; Jasial, S.; Bajorath, J. *F1000Research* **2015**, *4*, 118

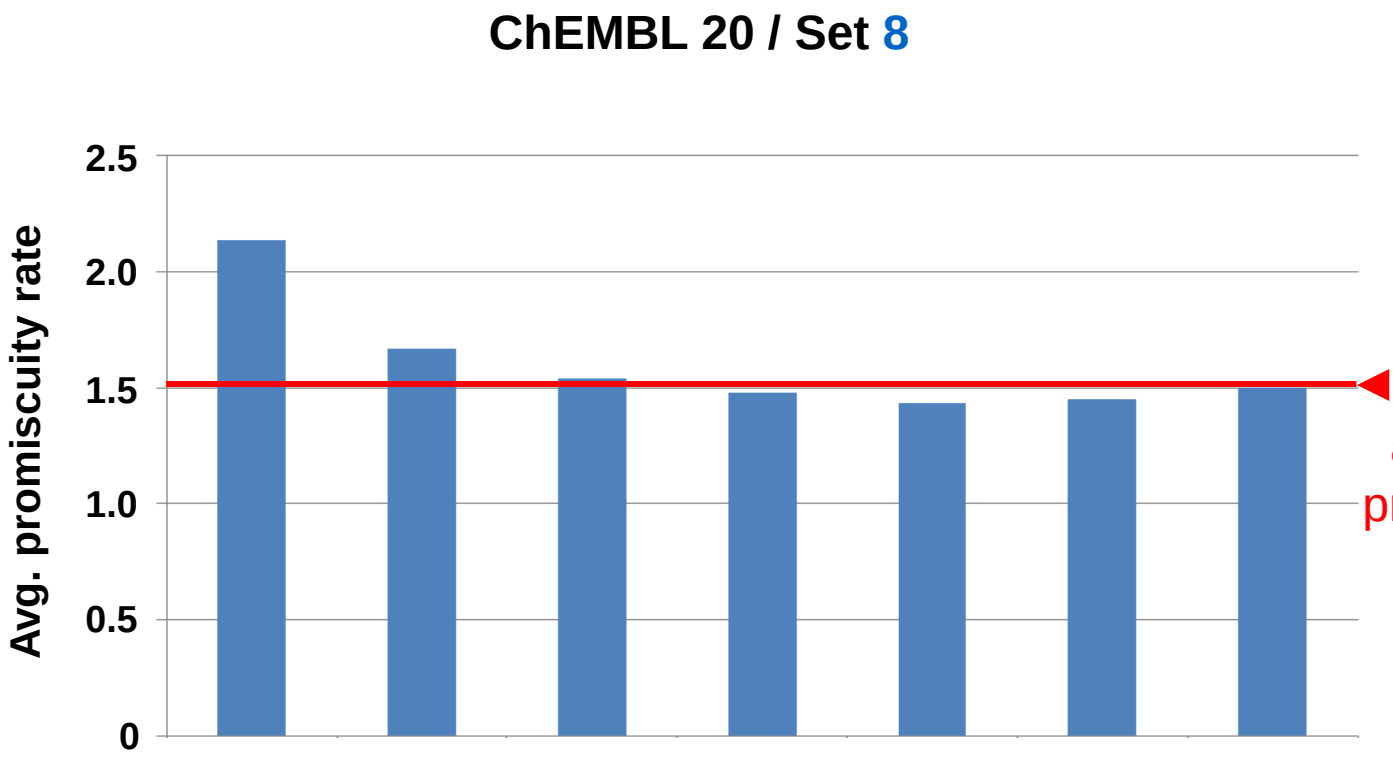
Drug Promiscuity Over Time



Promiscuity of Imatinib Over Time



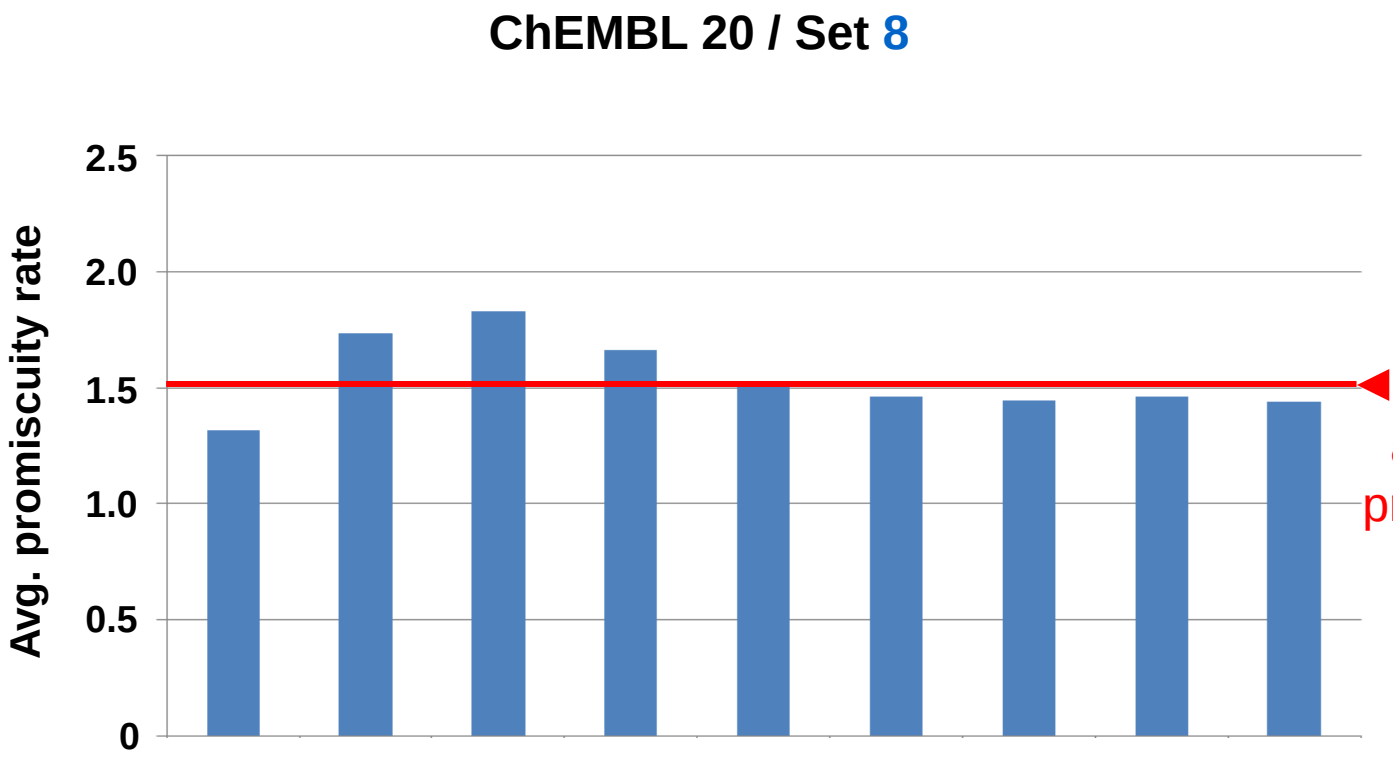
Promiscuity vs. Molecular Weight



MW Range	≤200	300	400	500	600	700	>700
# Compounds	2294	16,123	46,914	51,884	24,572	6839	5367

Hu, Y.; Jasial, S.; Bajorath, J. *F1000Research* **2015**, *4*, 118

Promiscuity vs. Lipophilicity

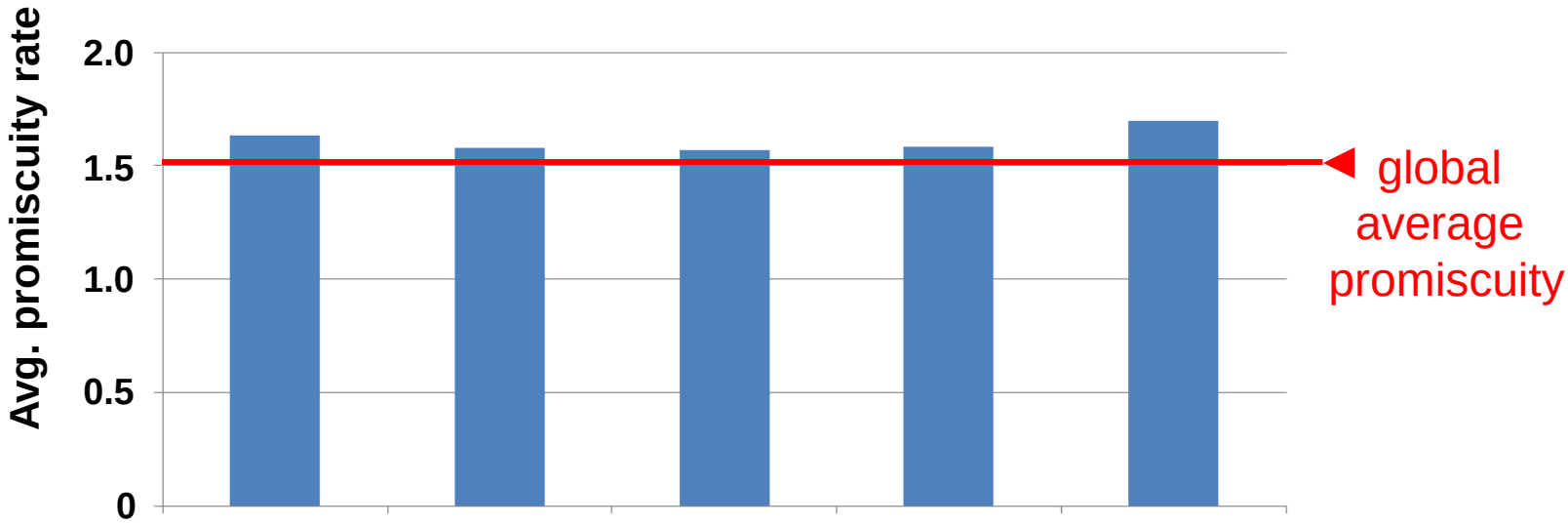


logP	≤-4	-2	0	2	4	6	8	10	>10
# Compounds	826	1169	4044	22,501	59,630	48,746	14,384	2240	453

Hu, Y.; Jasial, S.; Bajorath, J. *F1000Research* **2015**, *4*, 118

Promiscuity Across Target Families

ChEMBL 20 / Set 8



Family	GPCR class A	Ion channels	Kinases	Nuclear receptors	Proteases
# Targets	167	80	278	30	149
# Compounds	48,977	11,532	22,254	5619	19,319

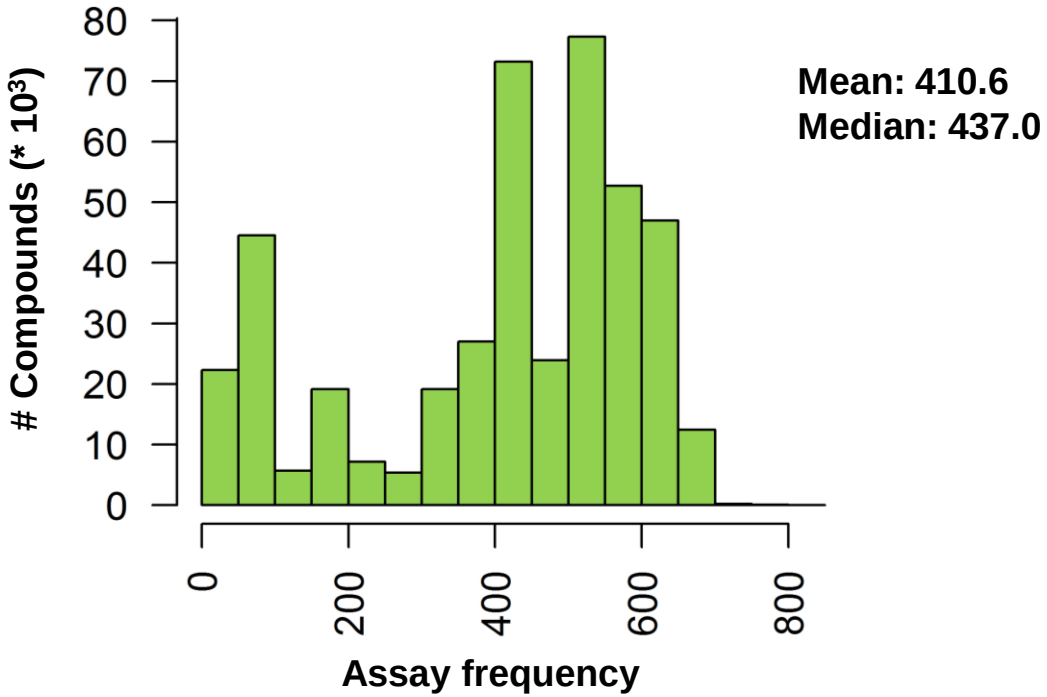
Promiscuity of Screening Hits

- Assay frequency or compounds inactivity information is typically not taken into account when assessing the promiscuity

- Extension of promiscuity analysis by identifying most extensively assayed public domain compounds

Extensively Assayed Compounds

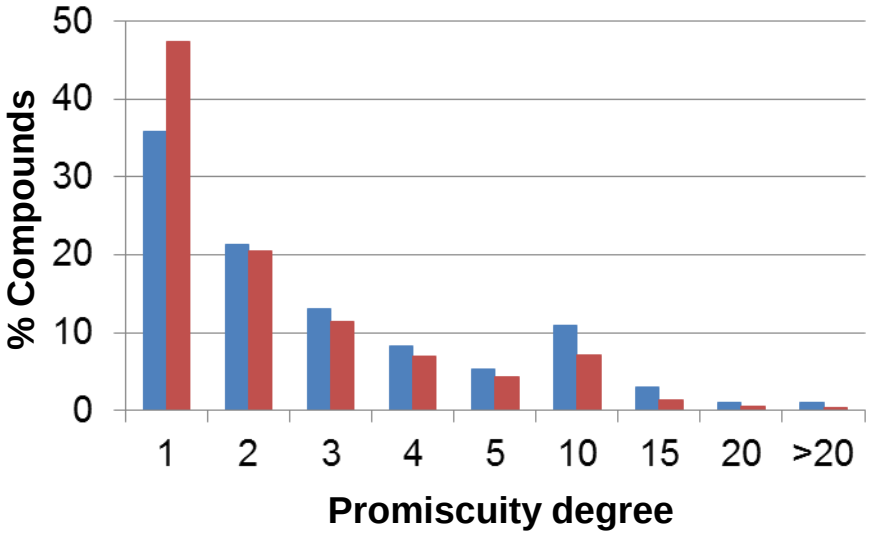
- 437,257 compounds assembled from PubChem BioAssays tested in both primary and confirmatory assays (>800 targets)



Jasial, S.; Hu, Y.; Bajorath, J. *PLoS One* **2016**, 11, e0153873

Extensively Assayed Compounds

- 267,418 compounds active in primary assays
- 196,607 compounds active in confirmatory assays



	Promiscuity	
	Primary	**Confirmatory**
Median	2.0	2.0
Mean	3.4	2.6

primary
confirmatory

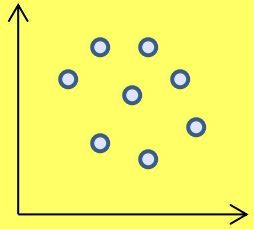
Target Promiscuity

- Target-centric view
- Derived from compound activity data

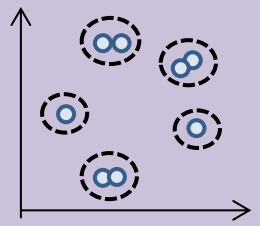
Hu, Y.; Bajorath, J. *PLoS One* **2015**, *10*, e0126838

Target Promiscuity Indices (**TPI**s)

Biologically Relevant Chemical Space

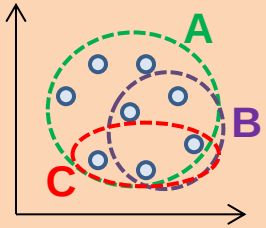


Scaffold Space



compounds represented by the same scaffold

Activity Space



compounds active against the same target

Target Promiscuity Indices (**TPI**s)

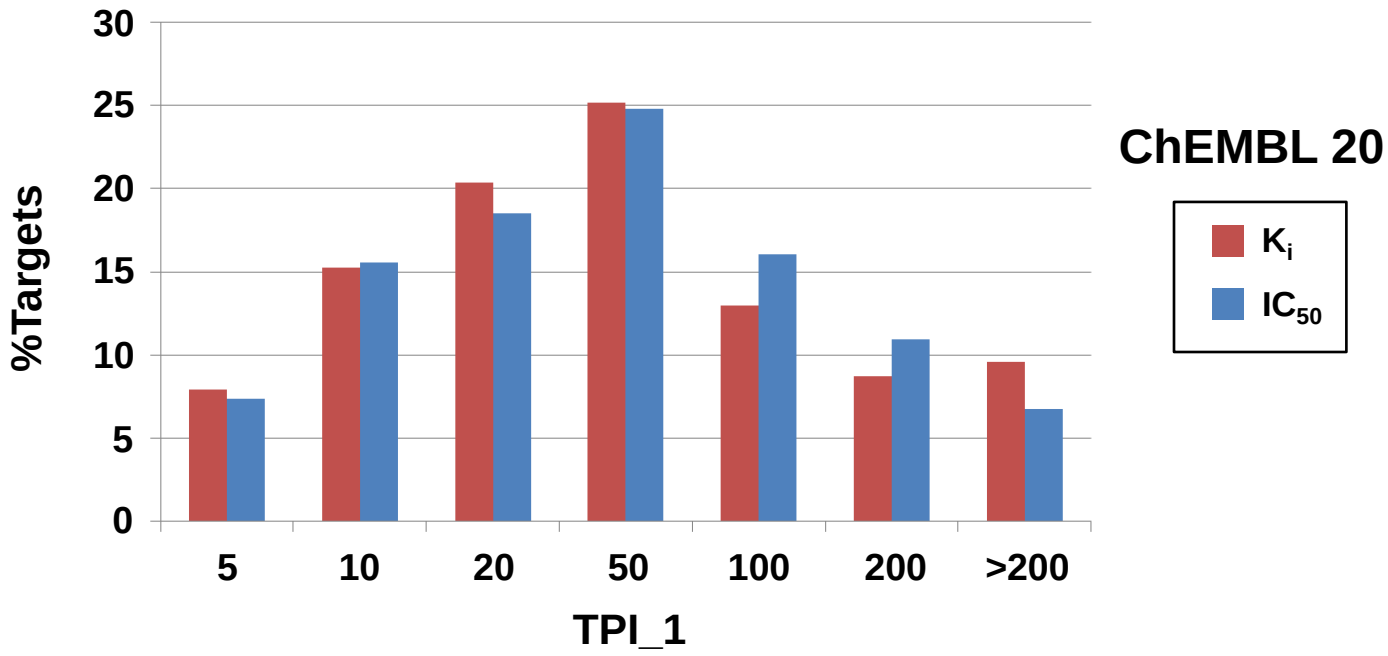
- **TPI_1**: *first-order target promiscuity index*
  - calculated as the number of unique scaffolds of all compounds active against a given target
  - *indicates the ability of a target to interact with structurally diverse compounds (i.e., scaffold hopping potential)*

- **TPI_2**: *second-order target promiscuity index*
  - average degree of promiscuity of all compounds active against the target
  - *reflects the tendency of a target to interact with specific or promiscuous compounds*

Hu, Y.; Bajorath, J. *PLoS One* **2015**, *10*, e0126838

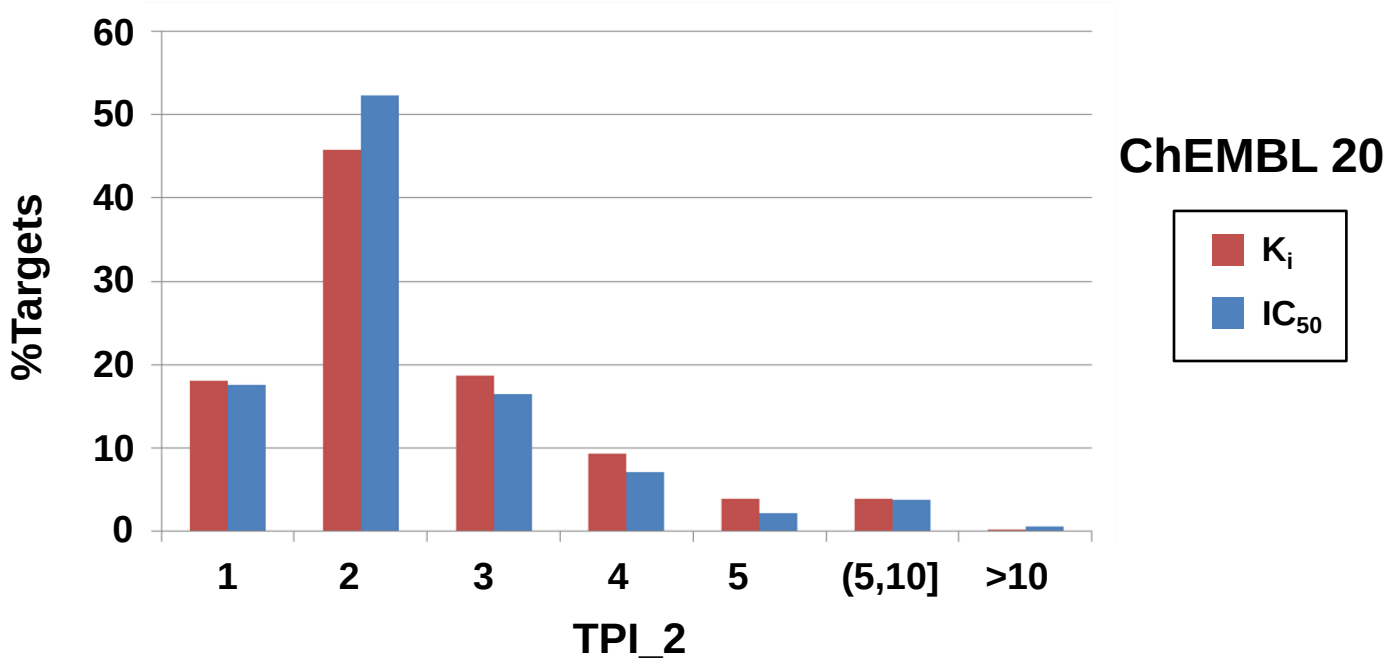
Distribution of TPI_1

- The average TPI_1 value over all targets is 77 (K_i data) and 61 (IC_{50}): Most targets bind structurally diverse compounds



Distribution of TPI_2

- Only ~18% of all targets interact with compounds having no other reported activity (‘pseudo-specific’ compounds) (TPI_2 value: 1)
- Most targets bind varying numbers of promiscuous compounds



Hu, Y.; Bajorath, J. *PLoS One* **2015**, *10*, e0126838

Targets with Varying TPI Patterns

- Targets interact with compounds
  - structurally diverse (>120 distinct scaffolds)
  - with no other reported activities

TPI pattern	Target name	#Cpds	TPI_1	TPI_2
High TPI_1 Low TPI_2	Leukotriene A4 hydrolase	217	124	1.01
	C-X-C chemokine receptor type 3	372	129	1.00

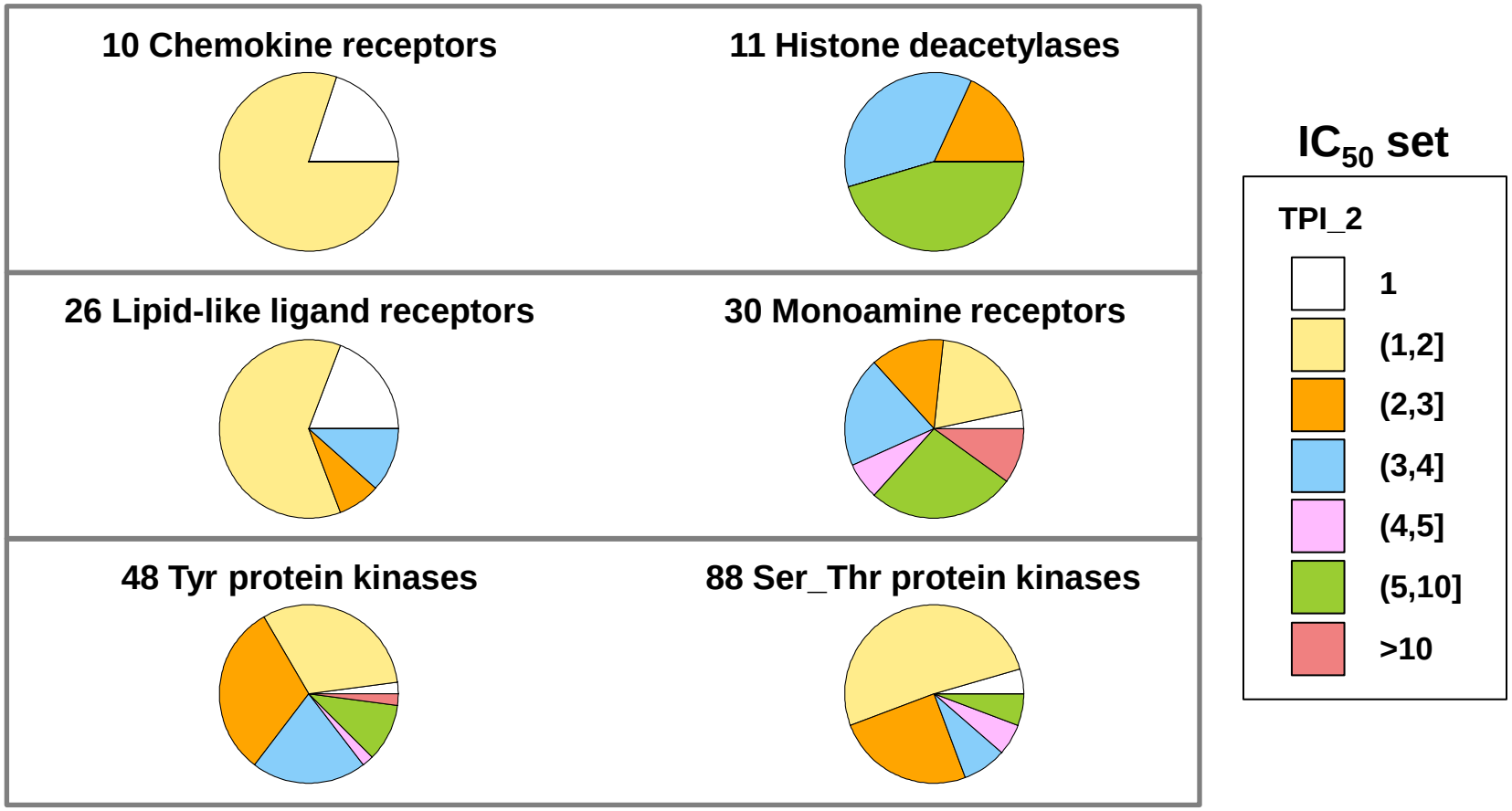
Targets with Varying TPI Patterns

- Targets interact with compounds
  - structurally homogeneous
  - preferentially promiscuous

TPI pattern	Target name	#Cpds	TPI_1	TPI_2
Low TPI_1 High TPI_2	Group IID secretory phospholipase A2	10	4	4.70
	Matrix metalloproteinase 16	12	6	6.42

Target Family Promiscuity Profiles

- TPI_2 values establish promiscuity profiles of target families



Conclusions

- Promiscuity
  - molecular basis of polypharmacology
  - ‘big data’ era enables and challenges large-scale promiscuity analysis
  - promiscuity must be viewed in light of data confidence

- Ligand-Centric Promiscuity
  - promiscuity degree of bioactive compounds is generally low
  - comparably low degree for ligands of major target families
  - drugs often have higher promiscuity

- Target-Centric Promiscuity
  - most targets recognize structurally diverse compounds
  - most targets bind both pseudo-specific and promiscuous compounds
  - different promiscuity patterns are observed