b-it LSi
Life
Science
Informatics

LIMES
Life & Medical Sciences Institute

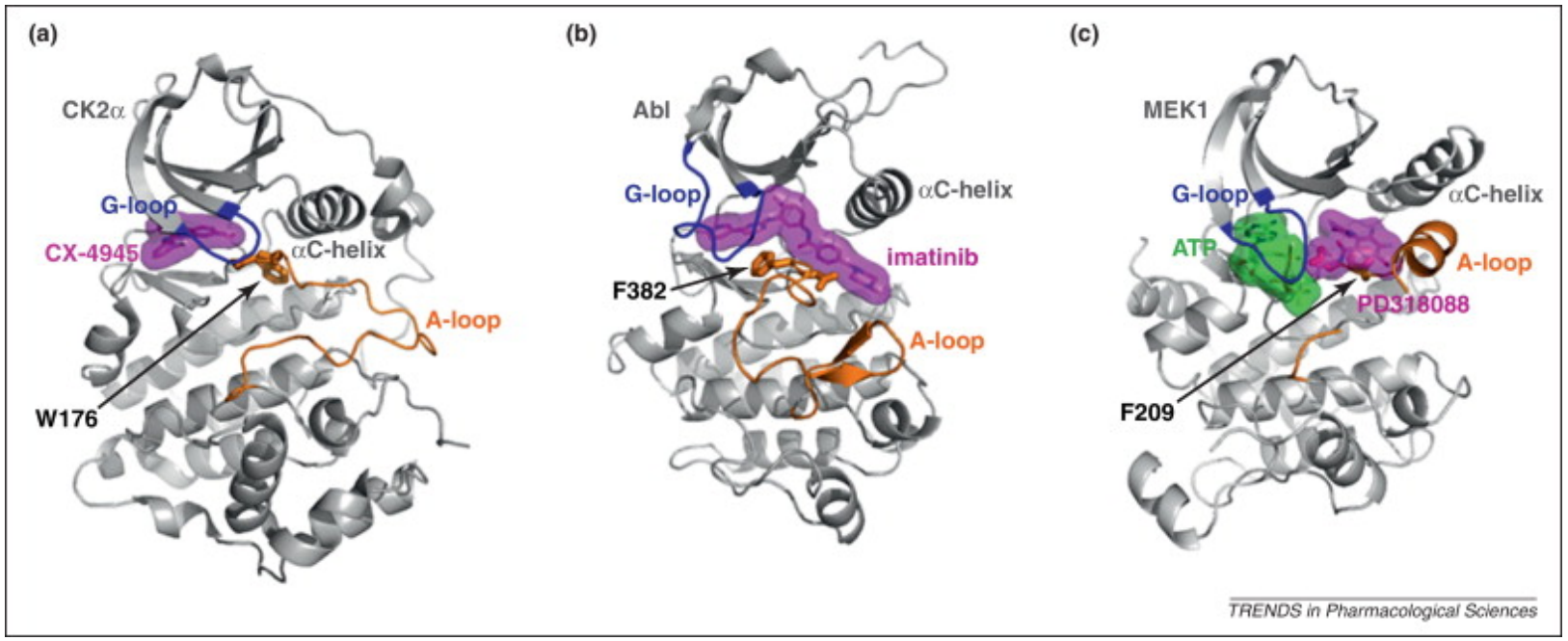
Emerging Big Data: Chemoinformatics-Driven View of Kinase Drug Discovery

Jürgen Bajorath
Life Science Informatics
University of Bonn

UNIVERSITÄT BONN

Protein Kinases

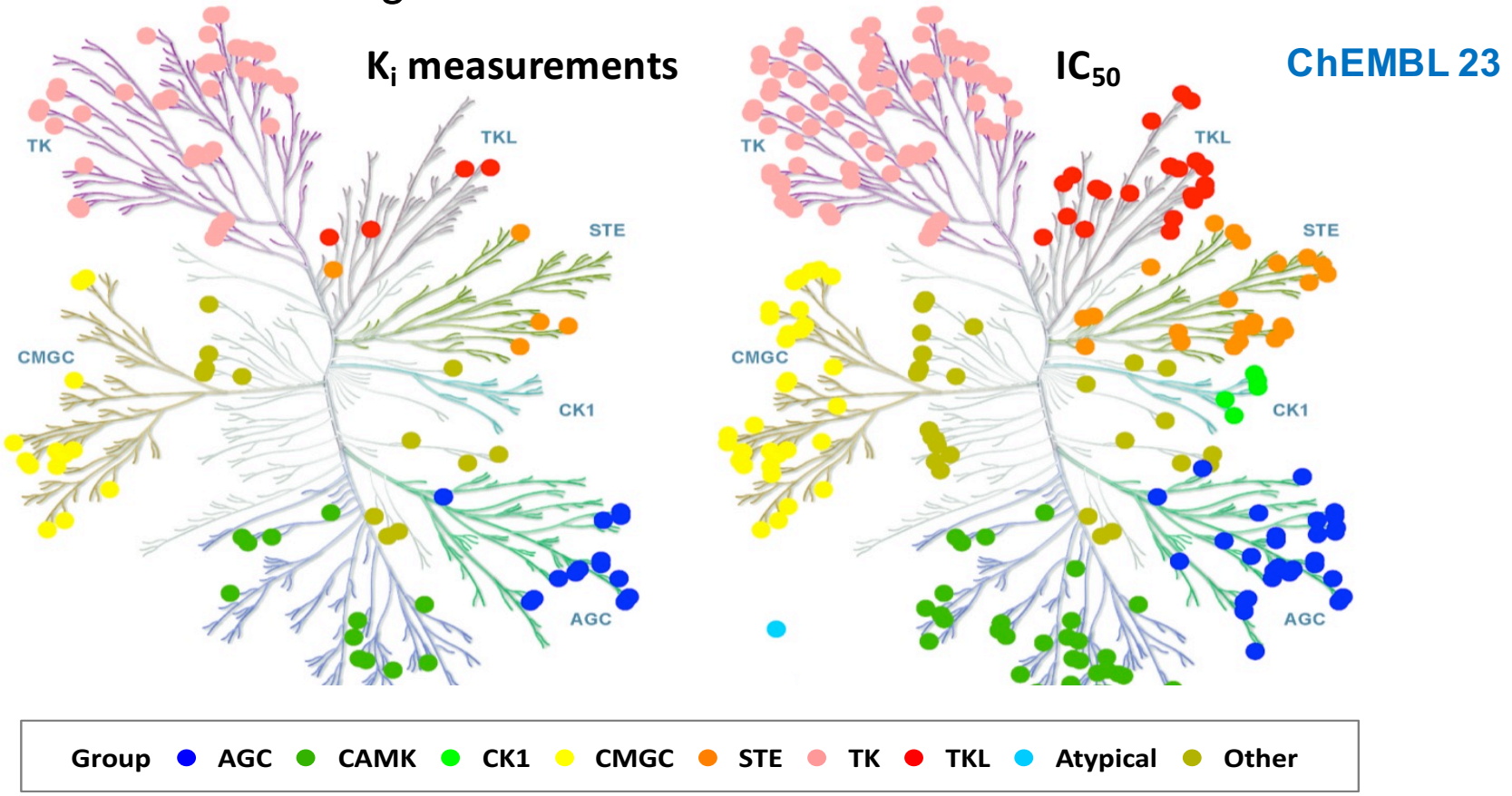
- **Kinases** and G protein coupled receptors currently are the most intensely investigated drug targets



- **Kinases** play a key role in native and aberrant signaling pathways and are prime targets in oncology, immunology/inflammation etc.

Inhibitor Coverage of the Kinome

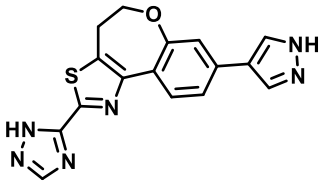
- Human kinome comprises 518 kinases
- Inhibitors with high-confidence data are available for 286 kinases



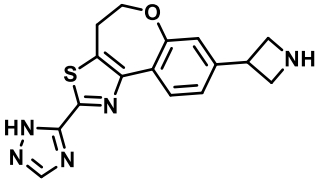
Kinase Inhibitors

- **Scaffolds** of structurally diverse and potent inhibitors

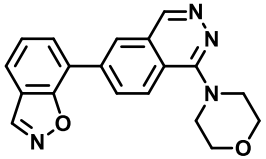
K_i subset



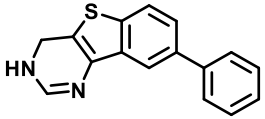
9 compounds
PI3K alpha
9.57 – 10.40



6 compounds
PI3K alpha
8.72 – 9.33

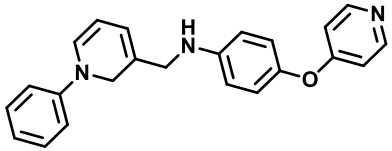


6 compounds
MAP kinase p38 alpha
9.22 – 9.70

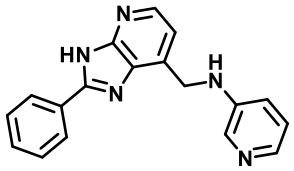


5 compounds
Ser/Thr protein kinase PIM3
9.00 – 9.52

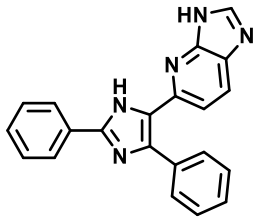
IC_{50} subset



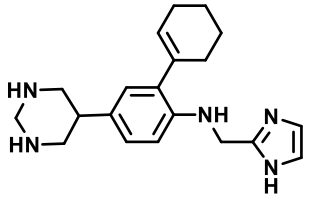
10 compounds
HGF receptor
8.05 – 9.00



9 compounds
GSK-3 beta
8.10 – 8.70



9 compounds
MAP kinase p38 alpha
8.14 – 8.60



6 compounds
CSF1R
8.60 – 9.52

Big Compound Data

- Compound activity data from biological screening and medicinal chemistry increasingly meet **Big Data** criteria

*The **'7 Vs'***

Volume
Velocity
Variety
Veracity
Variability
Visualization
Value

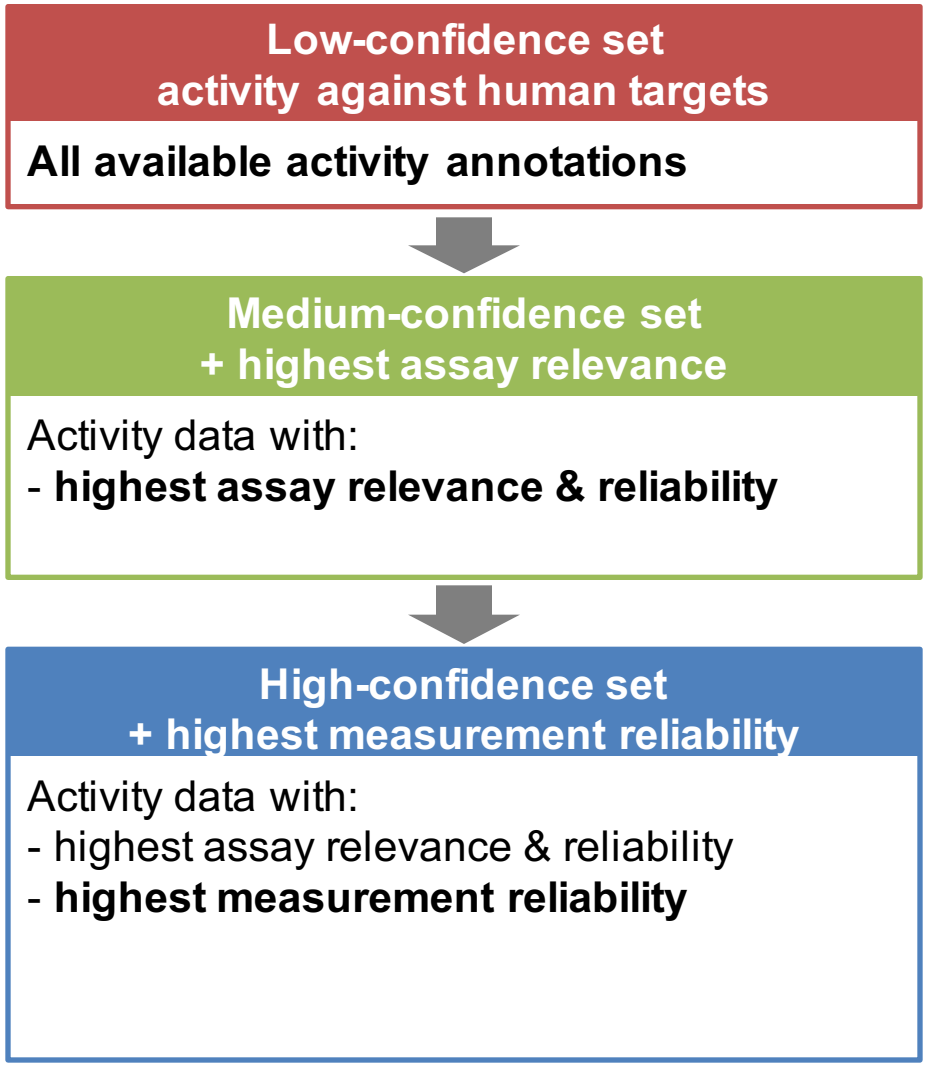
Heterogeneity (across databases)
Complexity (multi-layered data structures)
Confidence (experimental stringency)

(Hu & Bajorath, 2017)

(Van Rijmenam, 2013)

- Kinase inhibitors are a representative example

Activity Data Confidence Levels



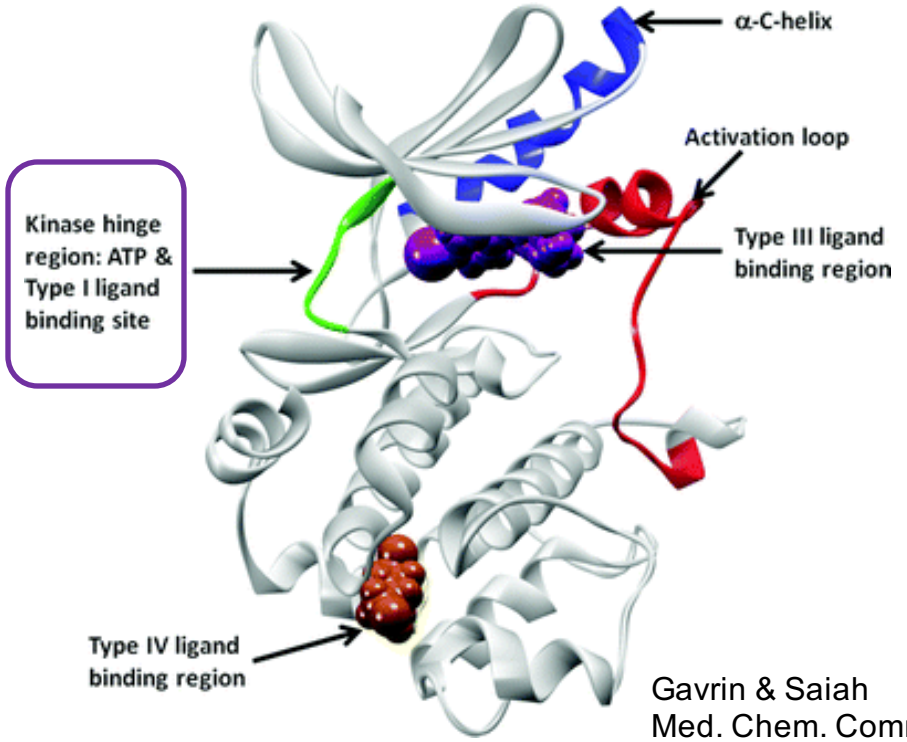
Data Growth

- Kinase inhibitors with **high-confidence** activity data

ChEMBL 23 **2017**	ChEMBL 18 **2015**
• **45,728** kinase inhibitors • 286 kinases • 12 kinase groups	• **18,951** kinase inhibitors • 266 kinases • 10 kinase groups

Kinase Inhibitors

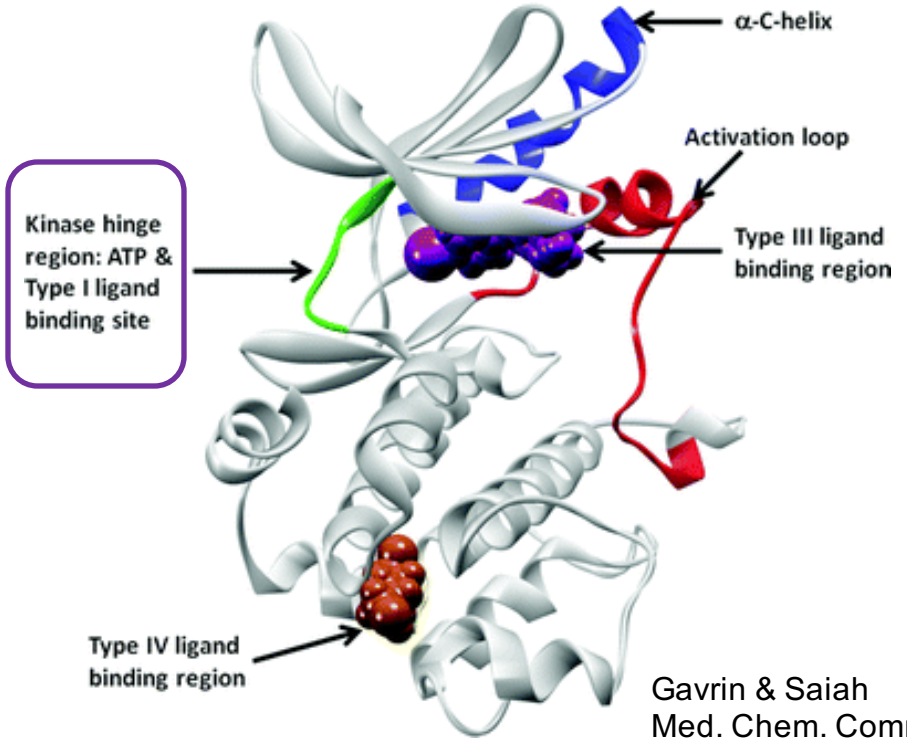
- More than 95% of current kinase inhibitors target the largely conserved **ATP** (cofactor) binding site (**type I**)



Gavrin & Saiah
Med. Chem. Commun. 2013

Kinase Inhibitors

- Type I inhibitors are expected to be **promiscuous** (active against multiple kinases)



Gavrin & Saiah
Med. Chem. Commun. 2013

Kinase Inhibitors

- **Promiscuity degrees** (high-confidence data)

ChEMBL 23 **2017**	ChEMBL 18 **2015**
• **45,728** kinase inhibitors • 286 kinases • 12 kinase groups	• **18,951** kinase inhibitors • 266 kinases • 10 kinase groups
• **33,157 (76.5%)** single-kinase inhibitors • 504 (**1.2%**) with activity against ≥ 5 kinases	• **14,892 (78.6%)** single-kinase inhibitors • 353 (**1.9%**) with activity against ≥ 5 kinases

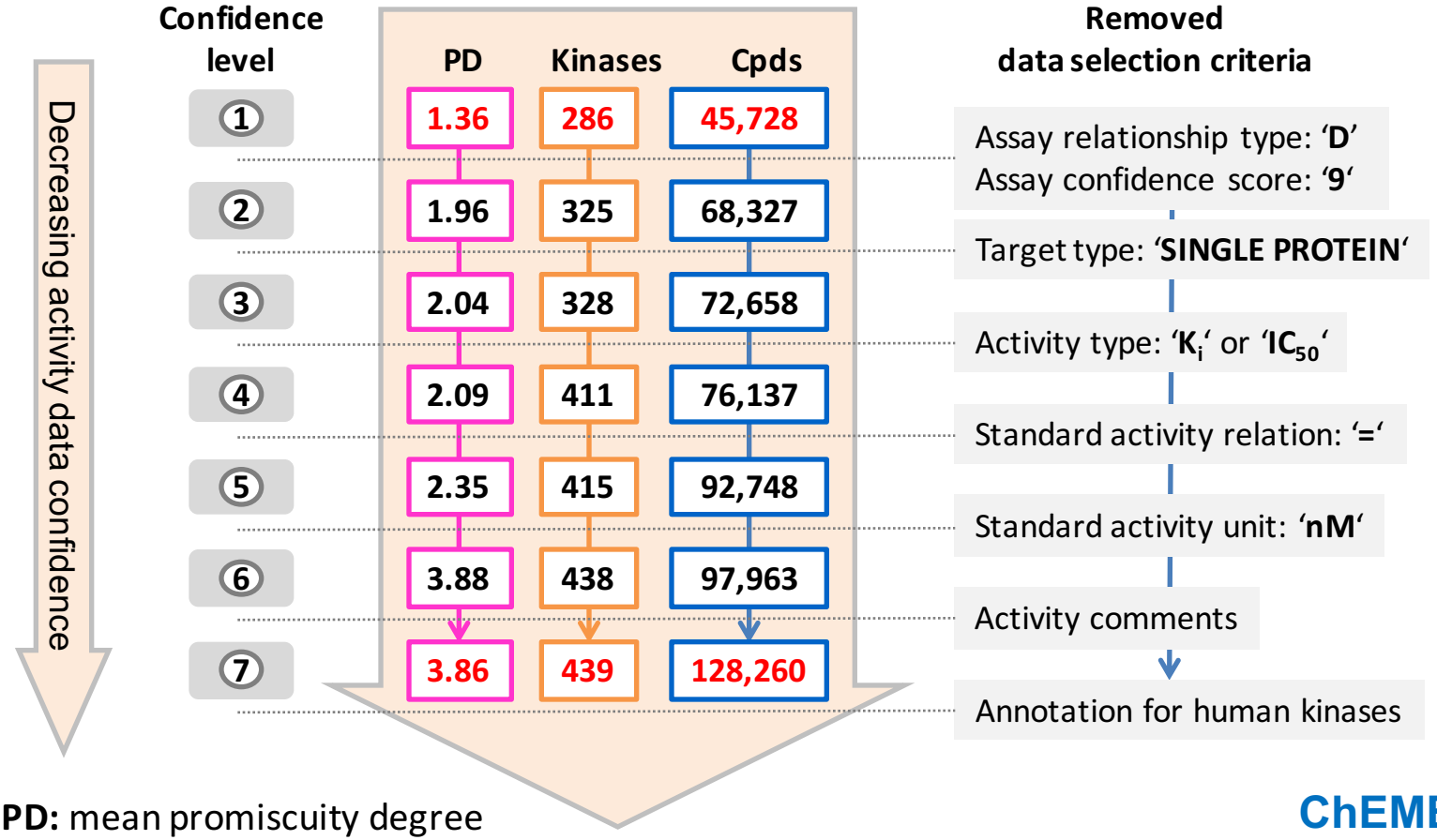
Kinase Inhibitors

- Low detectable promiscuity due to **data sparseness?**

ChEMBL 23 **2017**	ChEMBL 18 **2015**
• **45,728** kinase inhibitors • 286 kinases • 12 kinase groups	• **18,951** kinase inhibitors • 266 kinases • 10 kinase groups
• **33,157** (**76.5%**) single-kinase inhibitors • 504 (**1.2%**) with activity against ≥ 5 kinases	• **14,892** (**78.6%**) single-kinase inhibitors • 353 (**1.9%**) with activity against ≥ 5 kinases

Kinase Inhibitors - Promiscuity

- Iterative removal of confidence criteria leads to **increasing data volumes** and **kinome coverage**

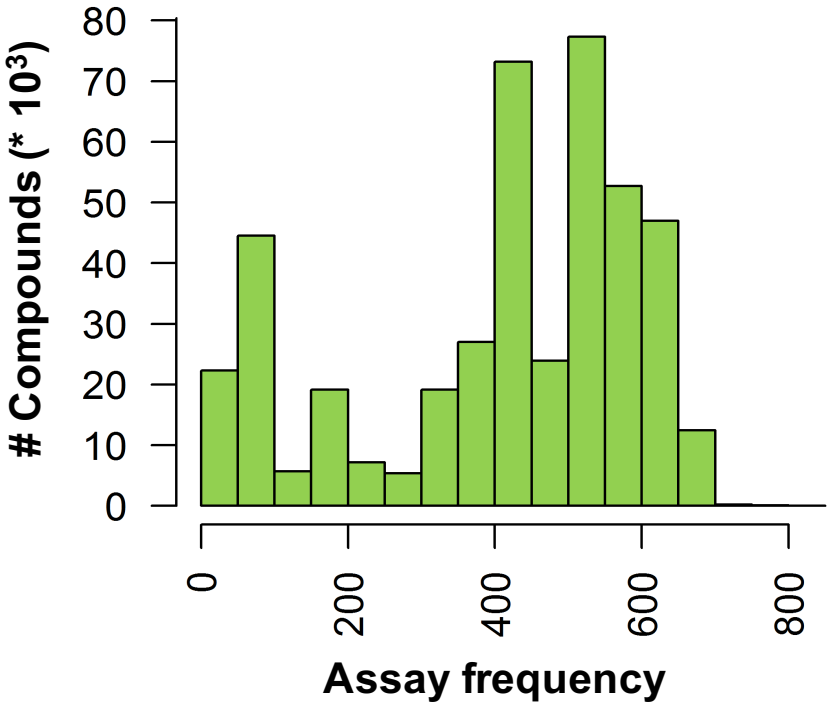


PD: mean promiscuity degree

ChEMBL 23

Extensively Assayed Compounds

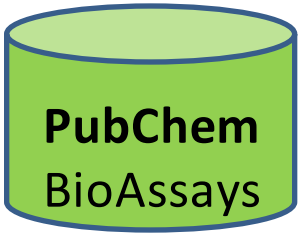
- **437,257** screening compounds assembled from **PubChem BioAssays** tested in both primary and confirmatory assays (> **800** targets)



Mean: 411
Median: 437

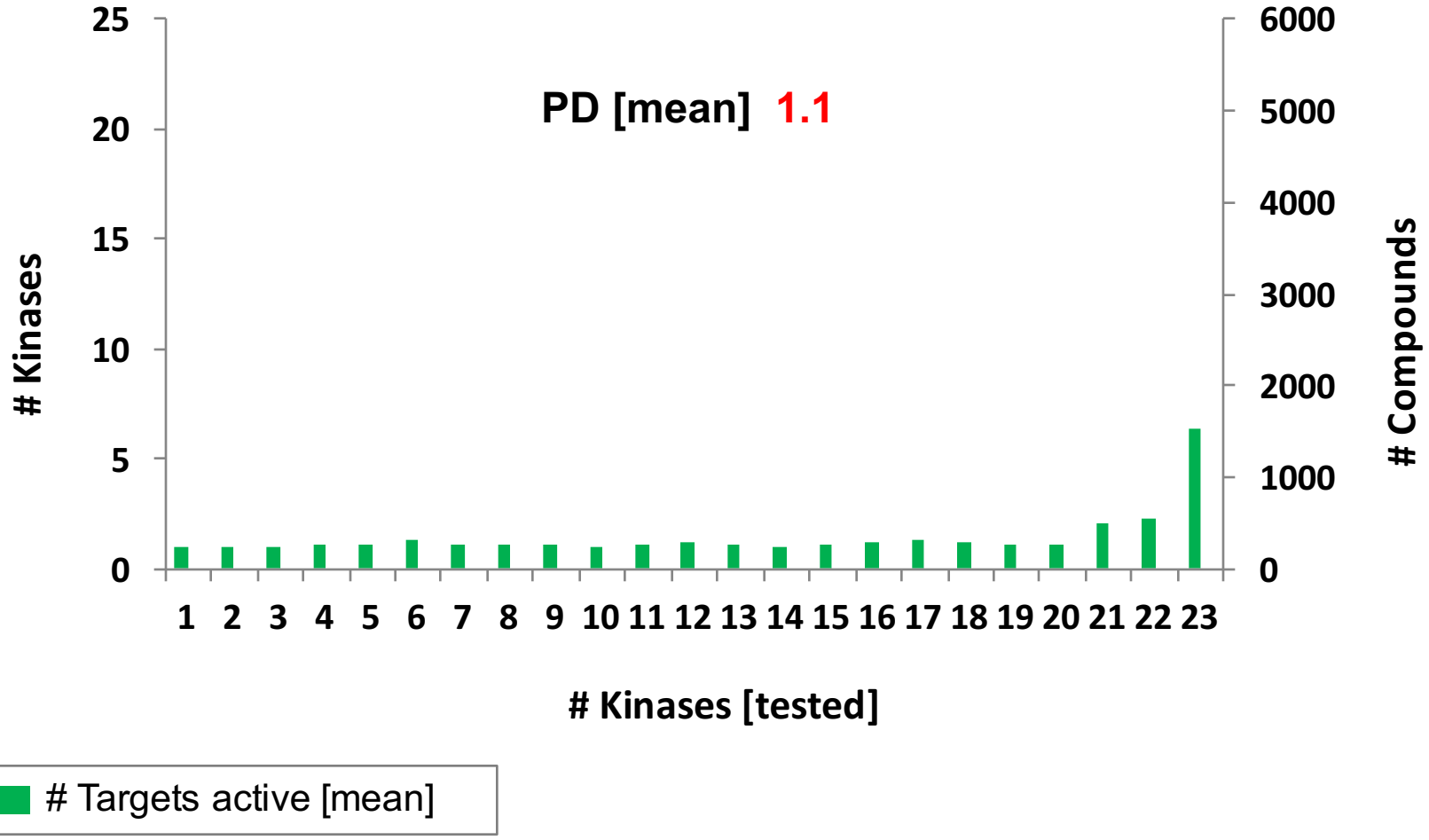
Kinase Inhibitors - Promiscuity

- Subset of **437,257 extensively assayed screening** compounds contains **28,172** inhibitors of a total of **43** human kinases

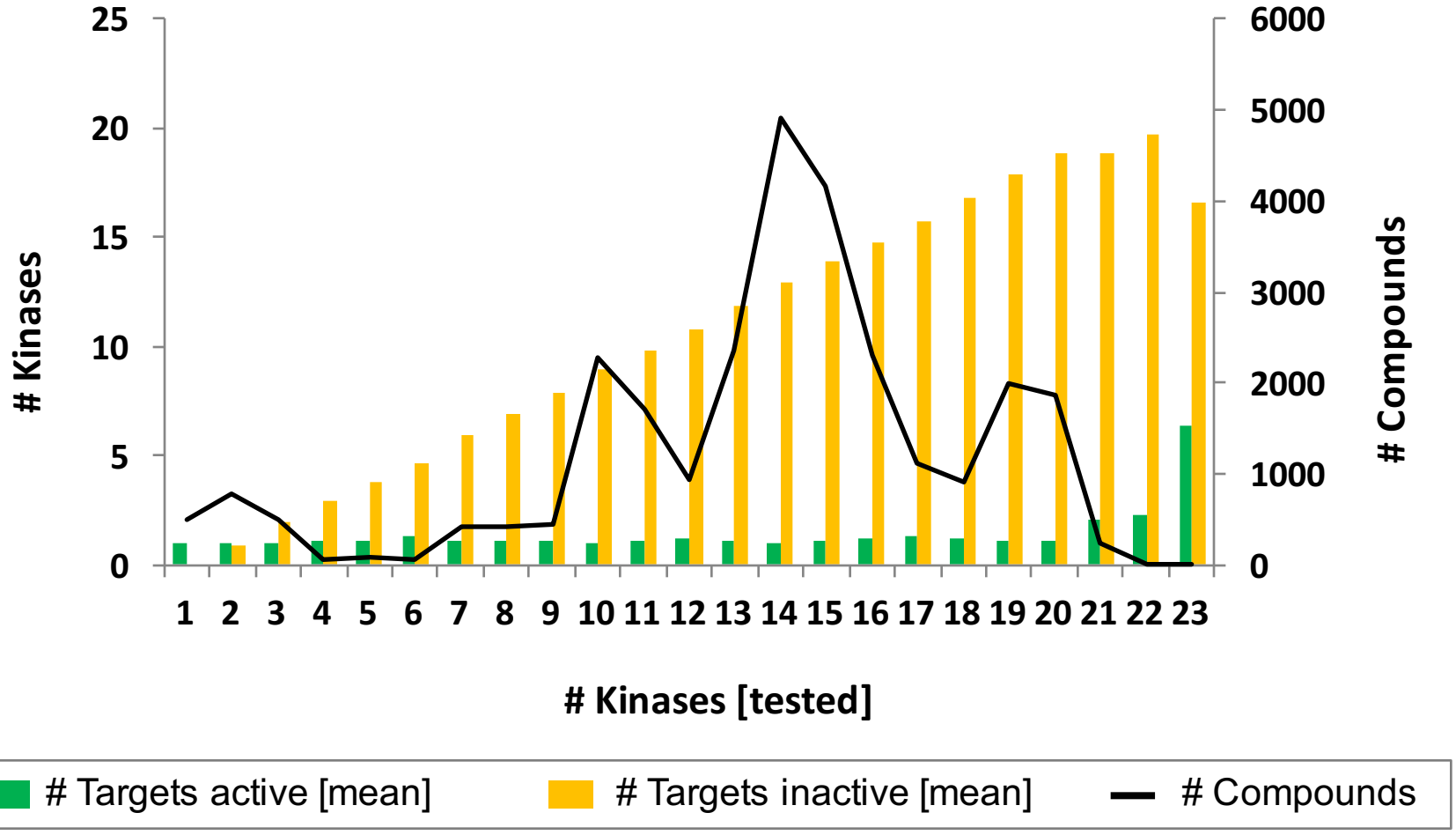


Extensively assayed screening compounds	
# Kinase inhibitors	**28,172**
# Kinase targets	**43**
Promiscuity degree (PD)	**1 – 11**
PD [mean]	**1.1**
# Kinases [tested]	**1 – 23** (mean **14**)

Kinase Inhibitors - Promiscuity

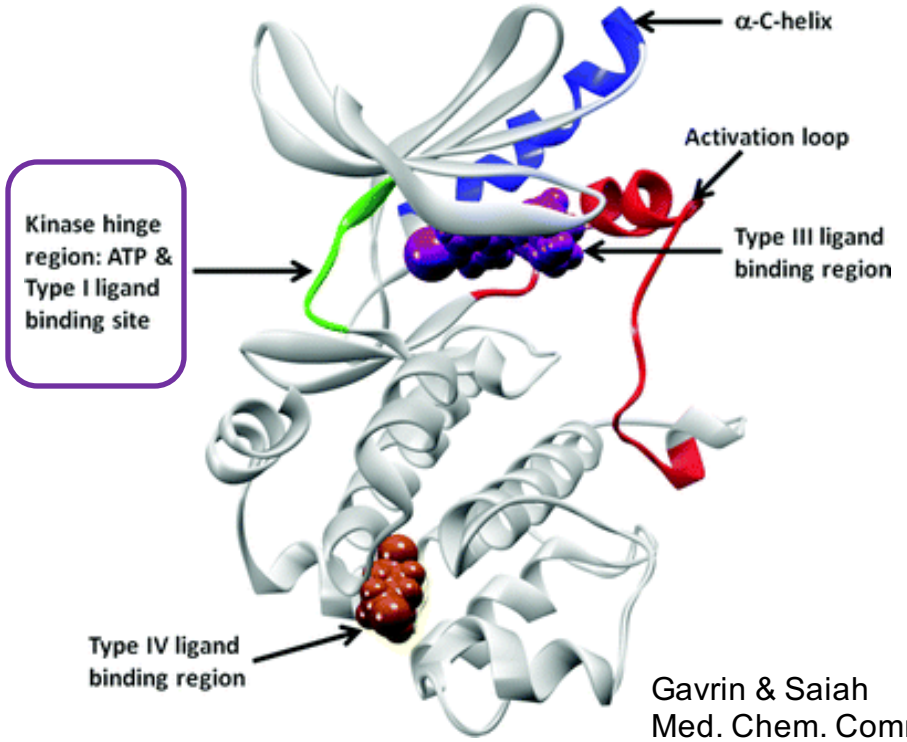


Kinase Inhibitors - Promiscuity



Kinase Inhibitors

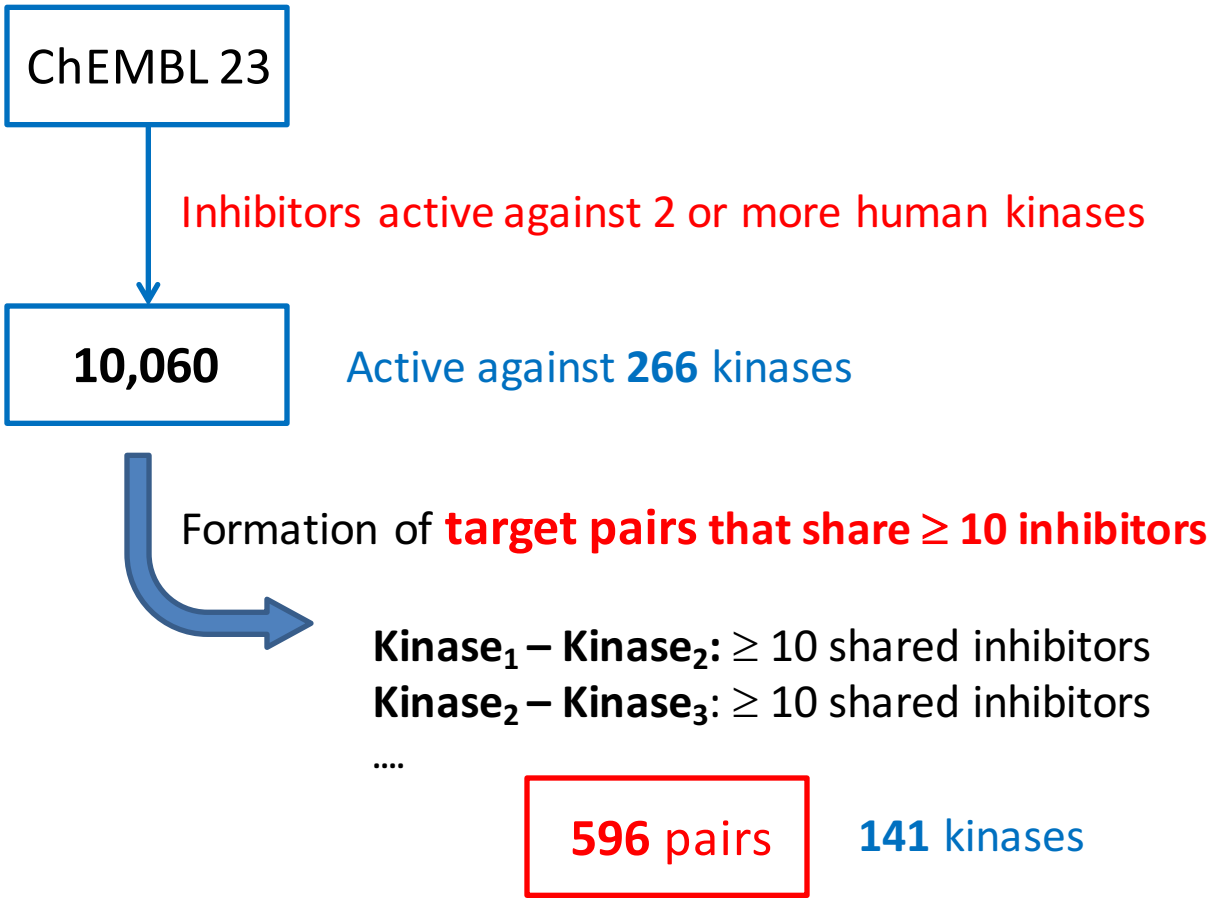
- Type I inhibitors are expected to have limited **selectivity** for individual kinases – **promiscuous** inhibitors should be **least selective**



Gavrin & Saiah
Med. Chem. Commun. 2013

Kinase Inhibitors - Selectivity

- **Promiscuous** human kinase inhibitors with high-confidence data



Kinase Pair Categories

- 596 protein kinase pairs sharing at least 10 inhibitors (median value of 18 compounds per pair)

- 3 different **pair categories** according to their phylogenetic distance[1-3]

(i)	**same family**	132 pairs
(ii)	**different families**	262 pairs
(iii)	**different groups**	202 pairs

[1]UniProt Consortium. *Nucleic Acids Res*. 2015, *43*, D204.
[2]Miranda-Saavedra, D.; Barton, G. J. *Proteins* 2007, *68,* 893-914.
[3]Manning, G. *et al. Science* 2002, *298*, 1912-1934.

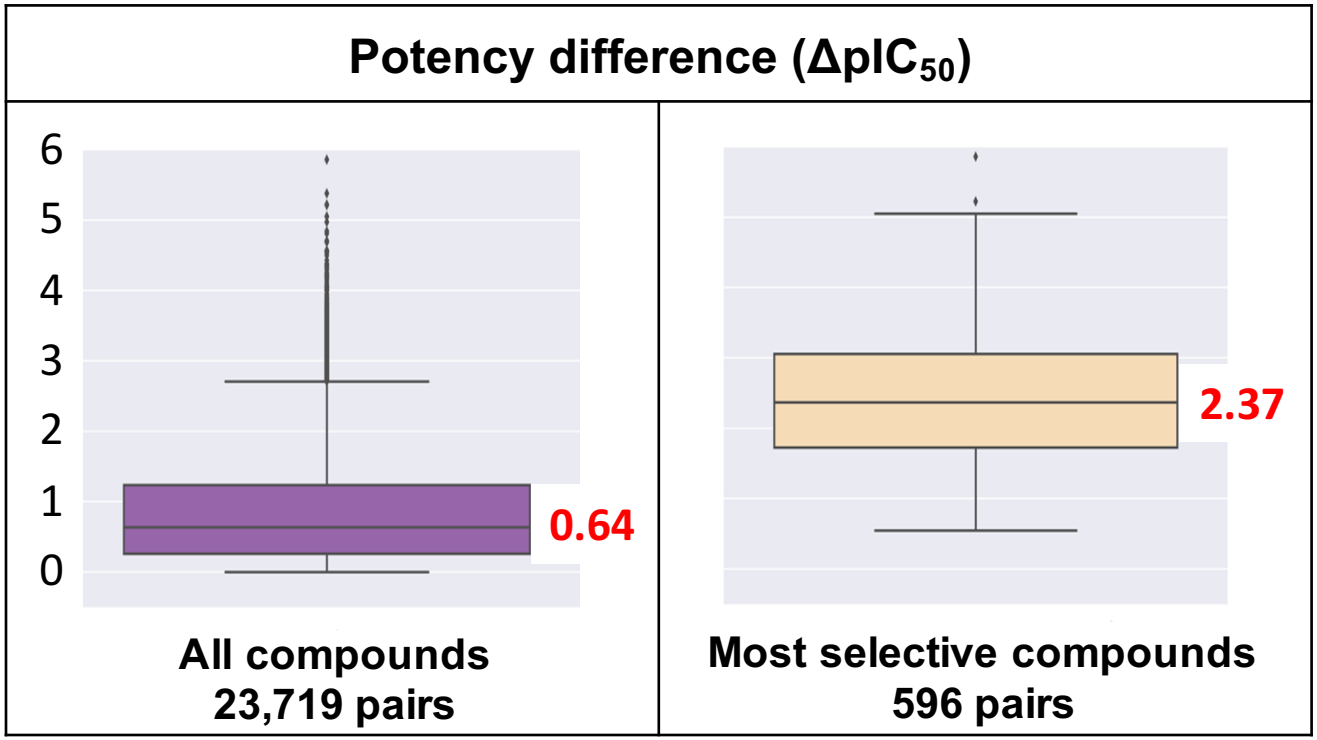
Selectivity Analysis

- By kinase pair-based analysis, all possible **selectivity relationships** for kinases and inhibitors were quantified

- For each kinase pair, **potency differences** (ΔpIC_{50}) between shared inhibitors were calculated as a **measure of selectivity**

Kinase Inhibitors - Selectivity

- Small global potency differences (median ΔpIC_{50} value 0.64)
- Different result for most selective inhibitors per pair (median 2.37)

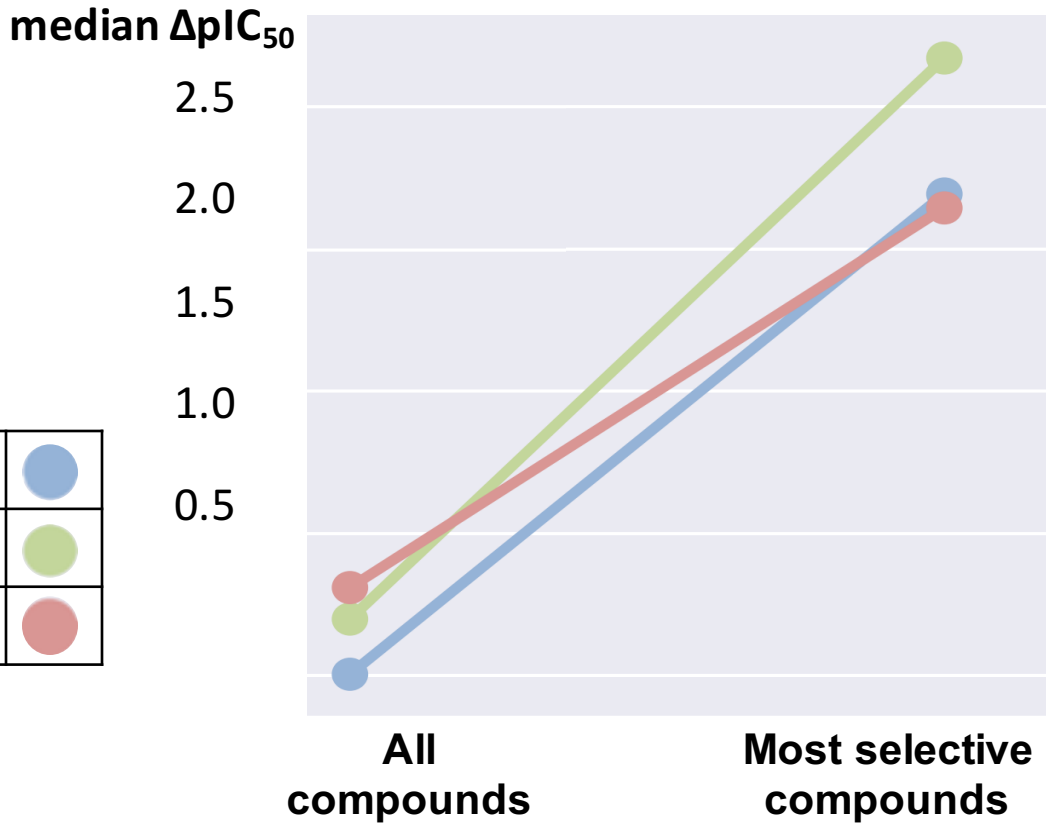


> 100-fold difference

Kinase Inhibitors - Selectivity

- Similar differences for pairs with increasing phylogenetic distances

(i)	same family	
(ii)	different families	
(iii)	different groups	



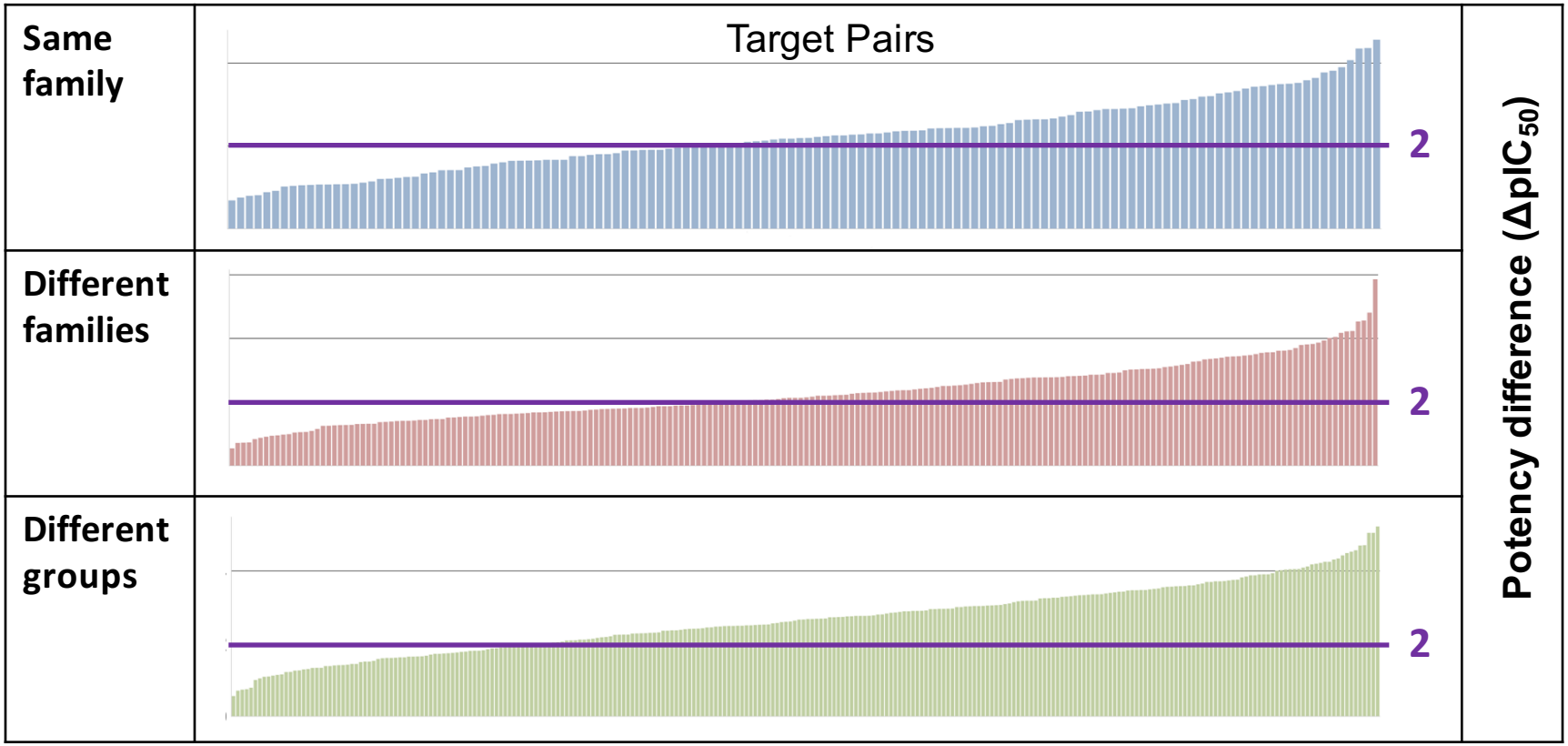
Selectivity Analysis

- ***Pair-based selectivity profiles:*** recording inhibitors with largest potency differences per kinase pair

- ***Compound-based selectivity profiles***: ordering available inhibitors according to increasing potency differences between paired kinases

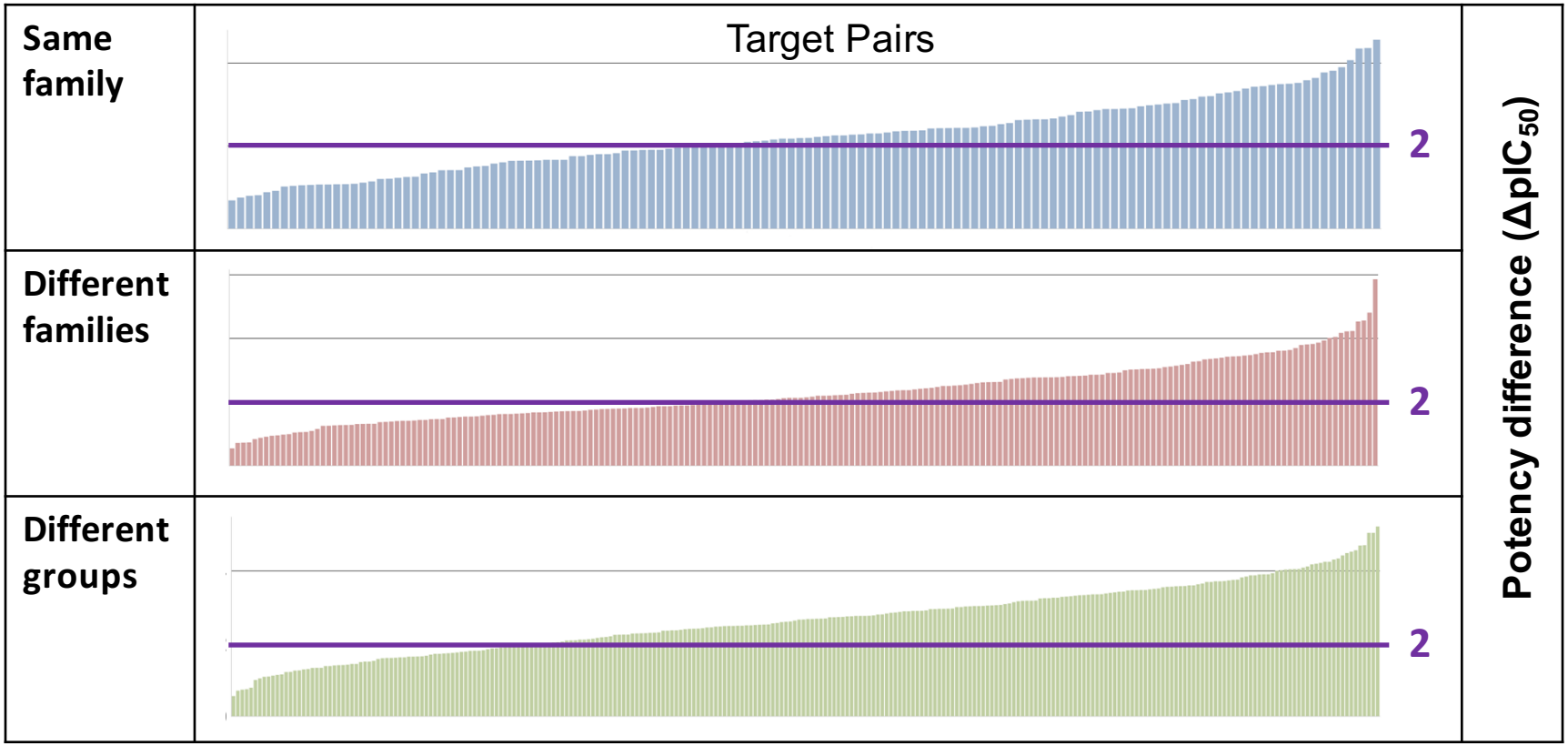
Category-Based Selectivity Profiles

- More than 50% of kinase pairs from each category have inhibitors with more than 100-fold potency differences



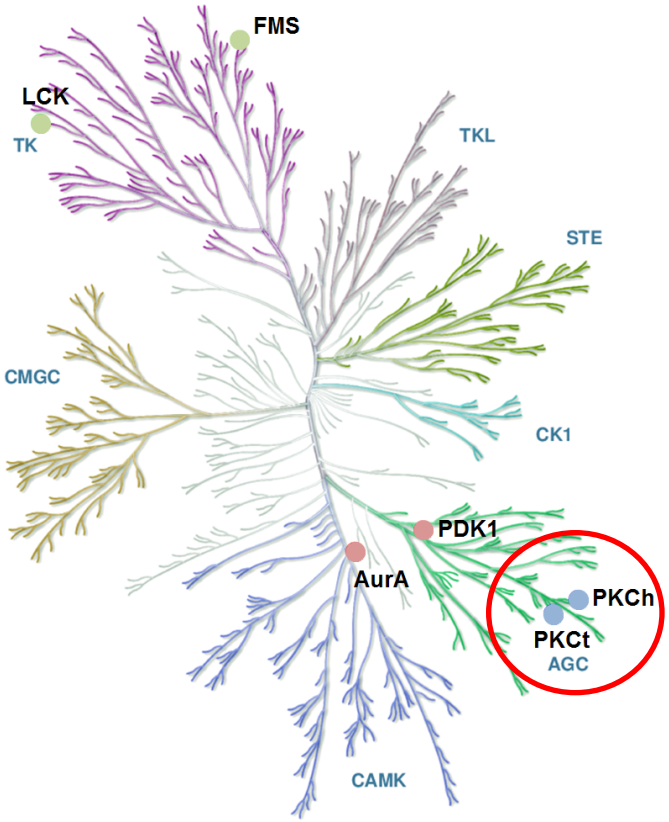
Kinase Inhibitors - Selectivity

- **Unexpected selectivity** among **promiscuous** kinase inhibitors

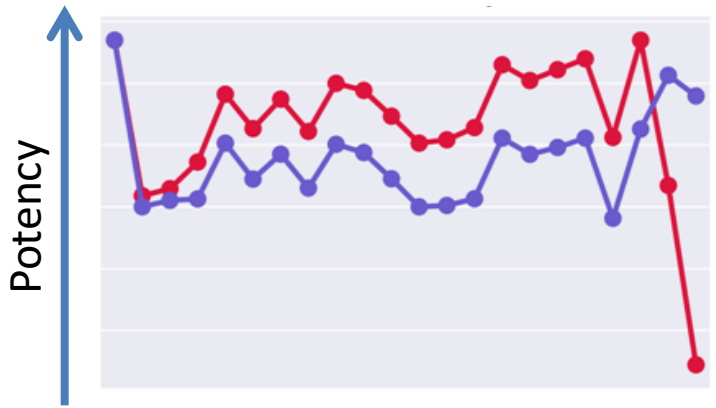


Compound-Based Selectivity Profiles

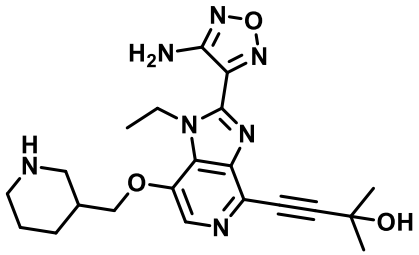
Kinases from the same family



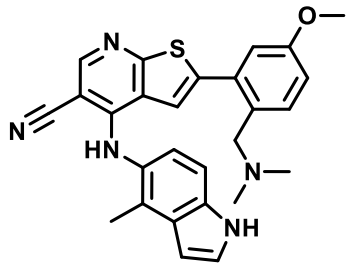
Protein kinase pair		Gatekeeper
PKCh ●	PKCt ●	M|M



Least selective

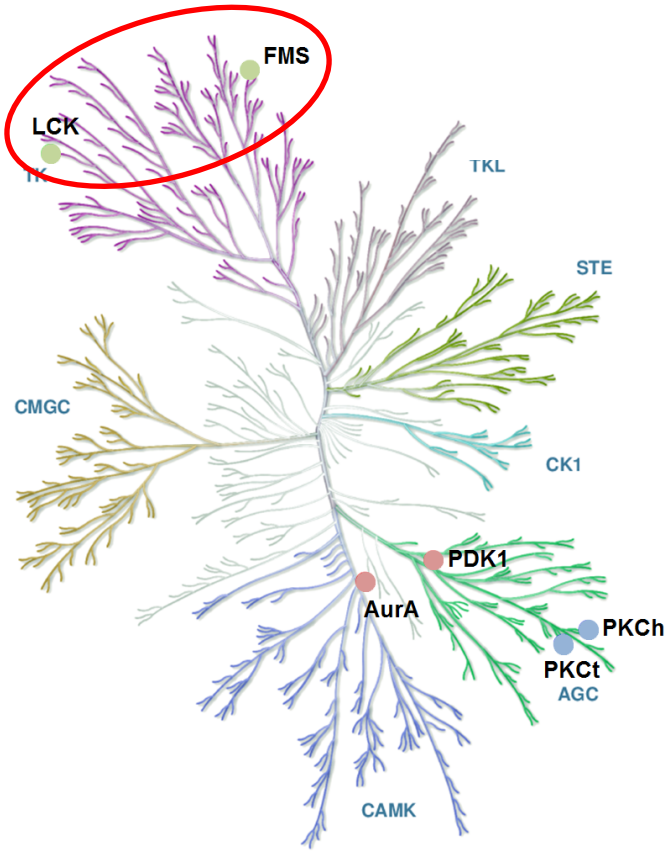


Most selective



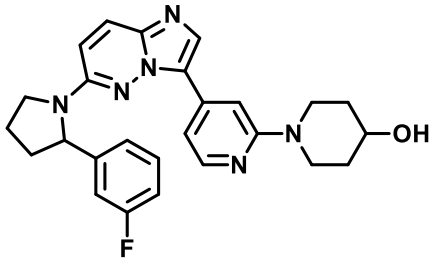
Compound-Based Selectivity Profiles

Kinases from different families

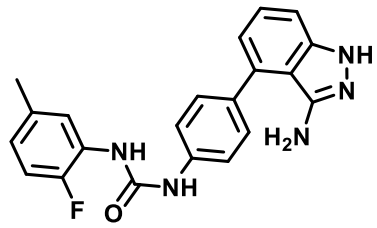


Protein kinase pair		Gatekeeper
FMS ●	LCK ●	T|T

Least selective

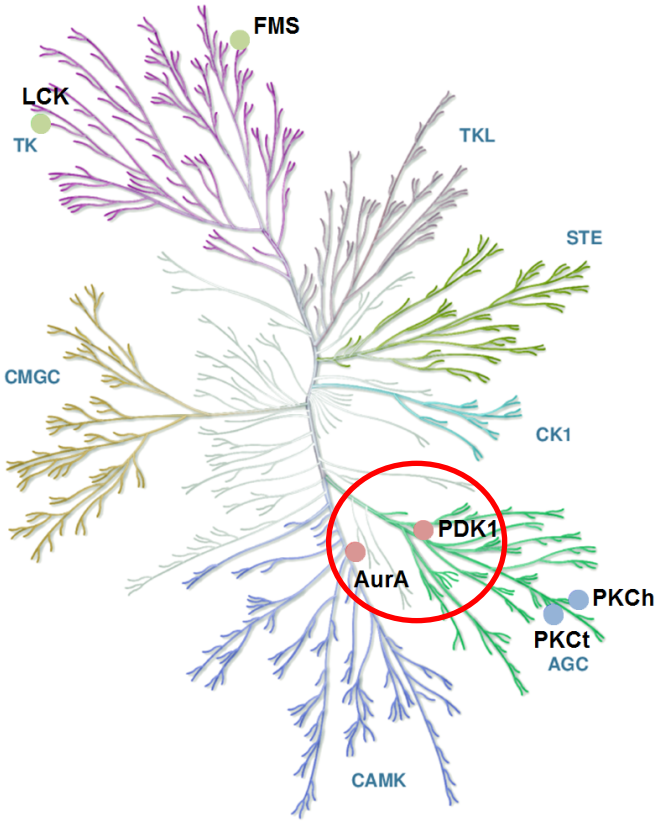


Most selective

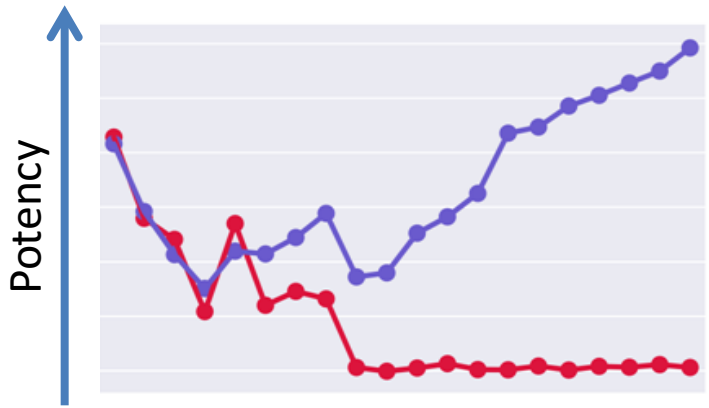


Compound-Based Selectivity Profiles

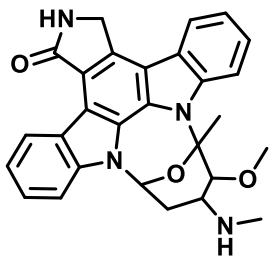
Kinases from different groups



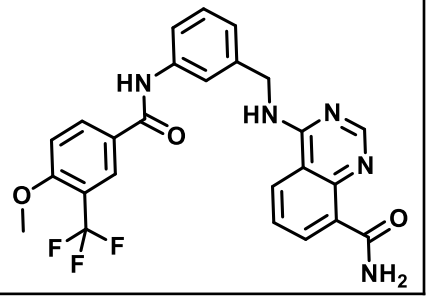
Protein kinase pair		Gatekeeper
PDK1 ●	AurA ●	L|L



Least selective



Most selective

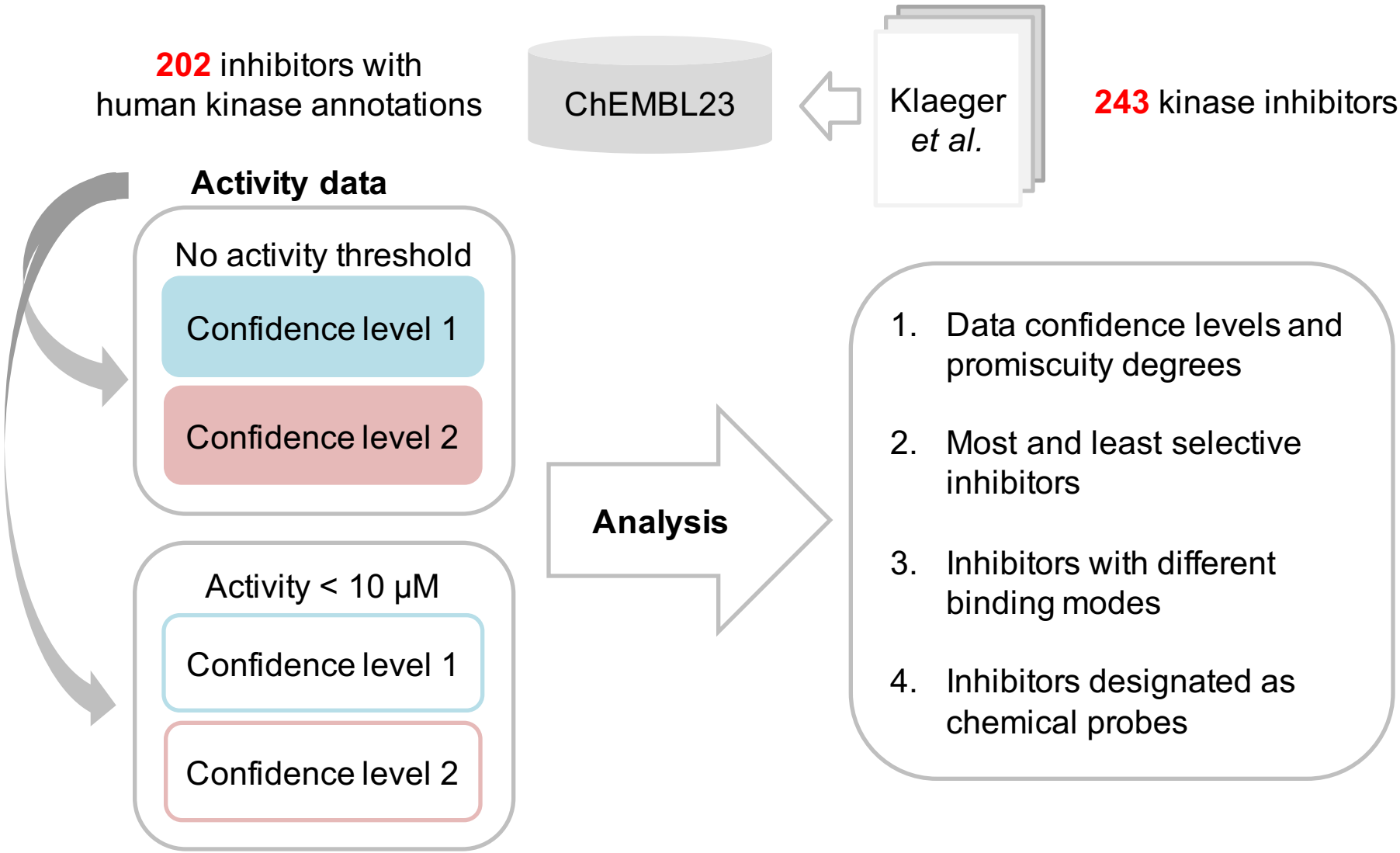


Profiling Clinical Kinase Inhibitors

- Most comprehensive kinase inhibitor profiling study currently available[1]
- Cell-based profiling of **243 clinical kinase inhibitors**
- "Kinobeads" assays, followed by qMS
- Cellular interactions with **253 human kinases** detected
- **Chemoinformatics perspective:**
  Reconciling observed promiscuity vs. selectivity trends on the basis of compound activity data from medicinal chemistry

[1]Klaeger *et al. Science* 2017, *358*, eaan4368.

Activity Data



Promiscuity Degrees (PDs)

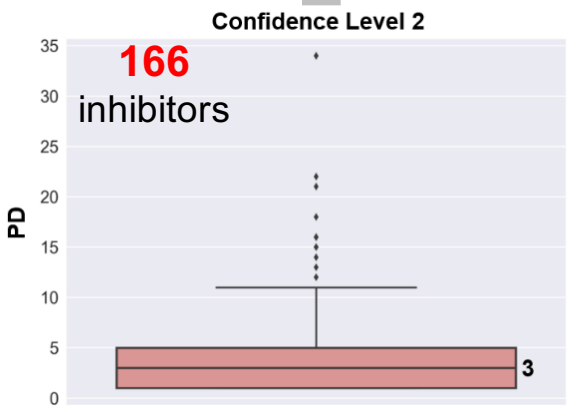
Highest Assay Confidence
+
Highest Measurement Confidence

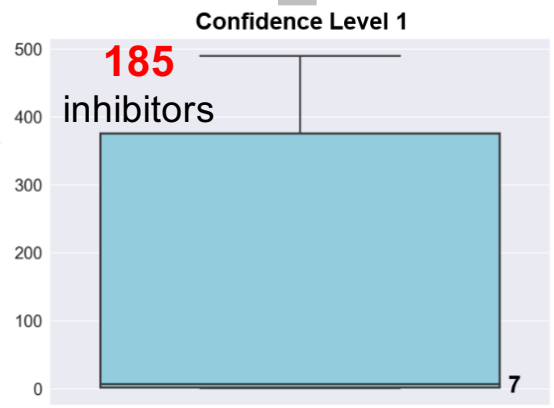
Highest Assay Confidence

No activity threshold

Confidence level 1
medium

Confidence level 2
high

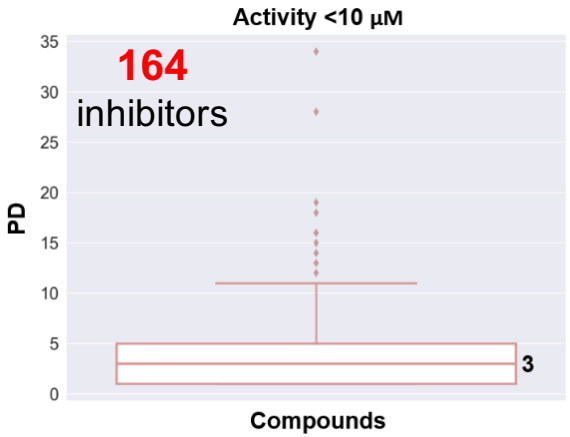


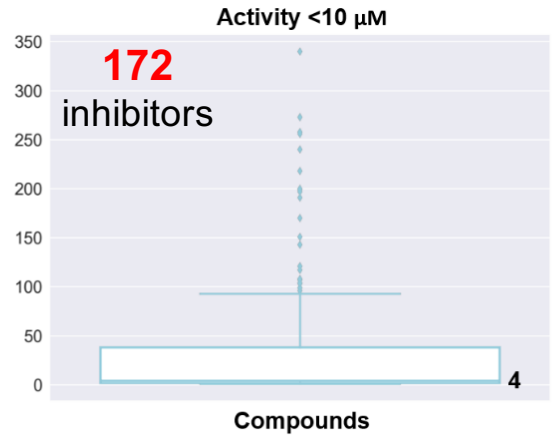


Activity < 10 µM

Confidence level 1

Confidence level 2





Promiscuity Degrees (PDs)

Strong influence of data confidence criteria

Small median PD values

No activity threshold

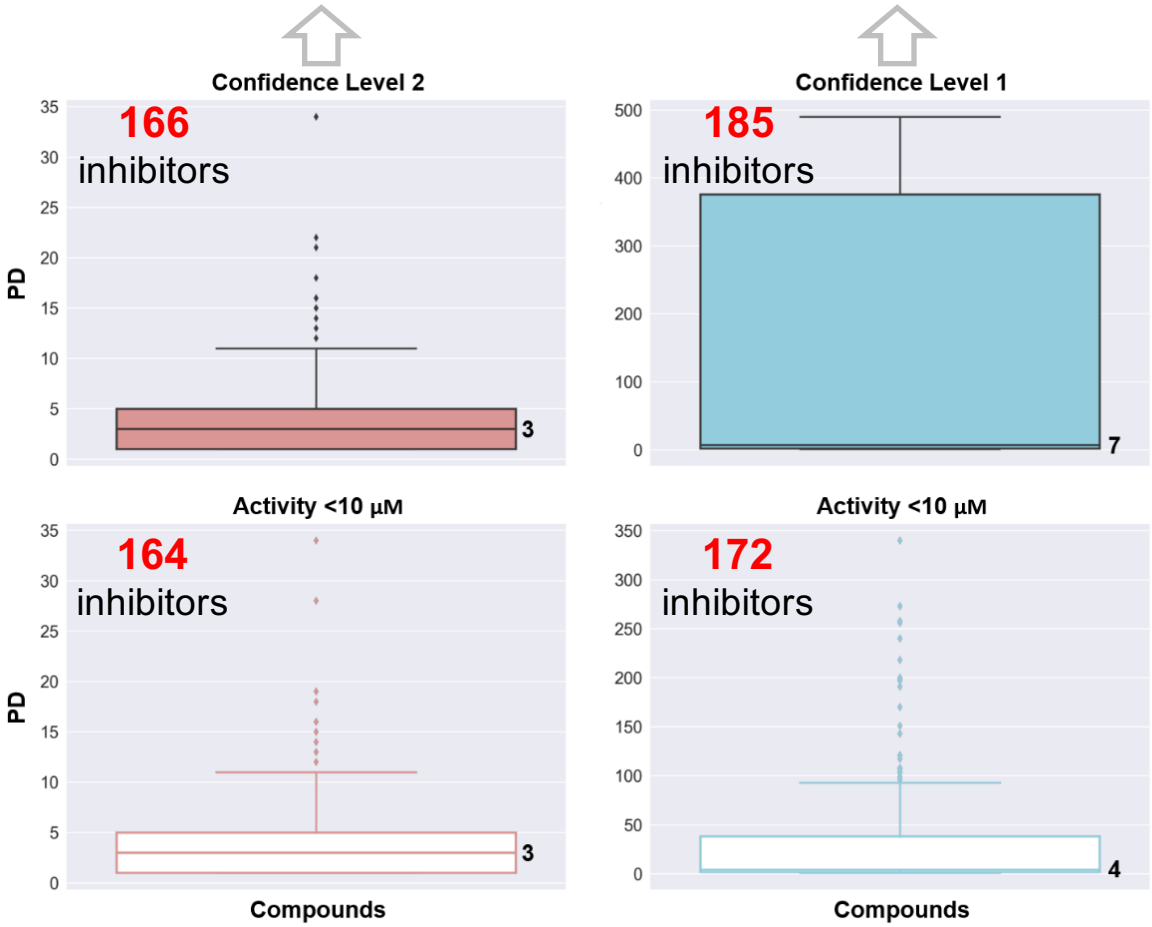
Confidence level 1
medium

Confidence level 2
high

Activity < 10 µM

Confidence level 1

Confidence level 2



Kinome Coverage

Cell-based profiling:
253 of 363 expressed human kinases

122 kinases
(< 10 µM: 122)

394 kinases
(< 10 µM: 379)

No activity threshold

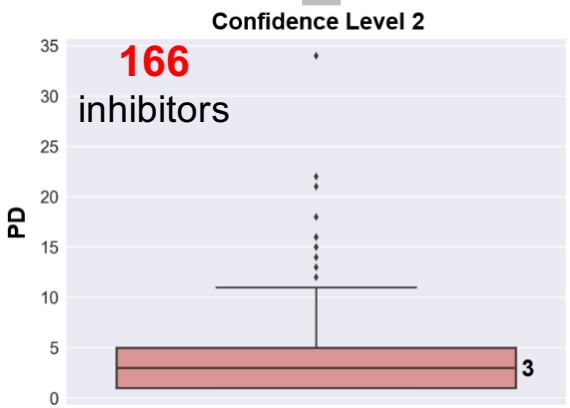
Confidence level 1
medium

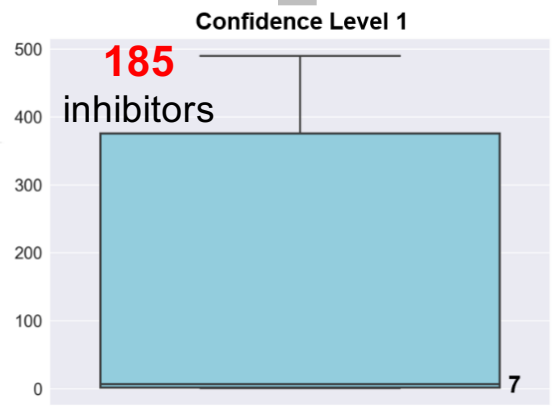
Confidence level 2
high

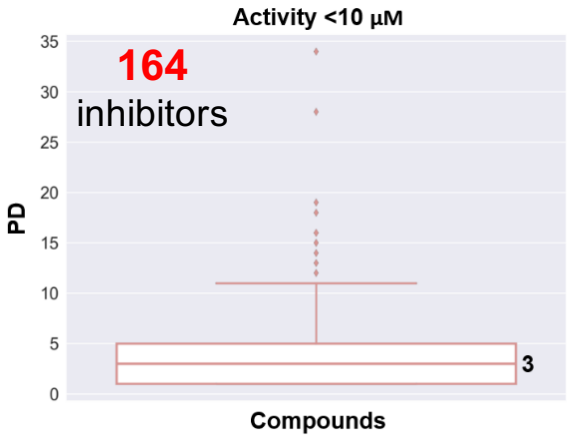
Activity < 10 µM

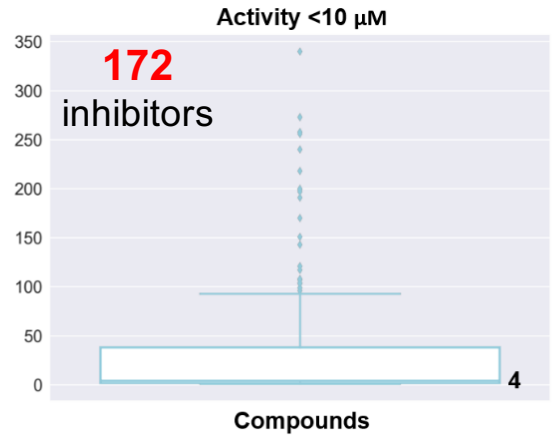
Confidence level 1

Confidence level 2

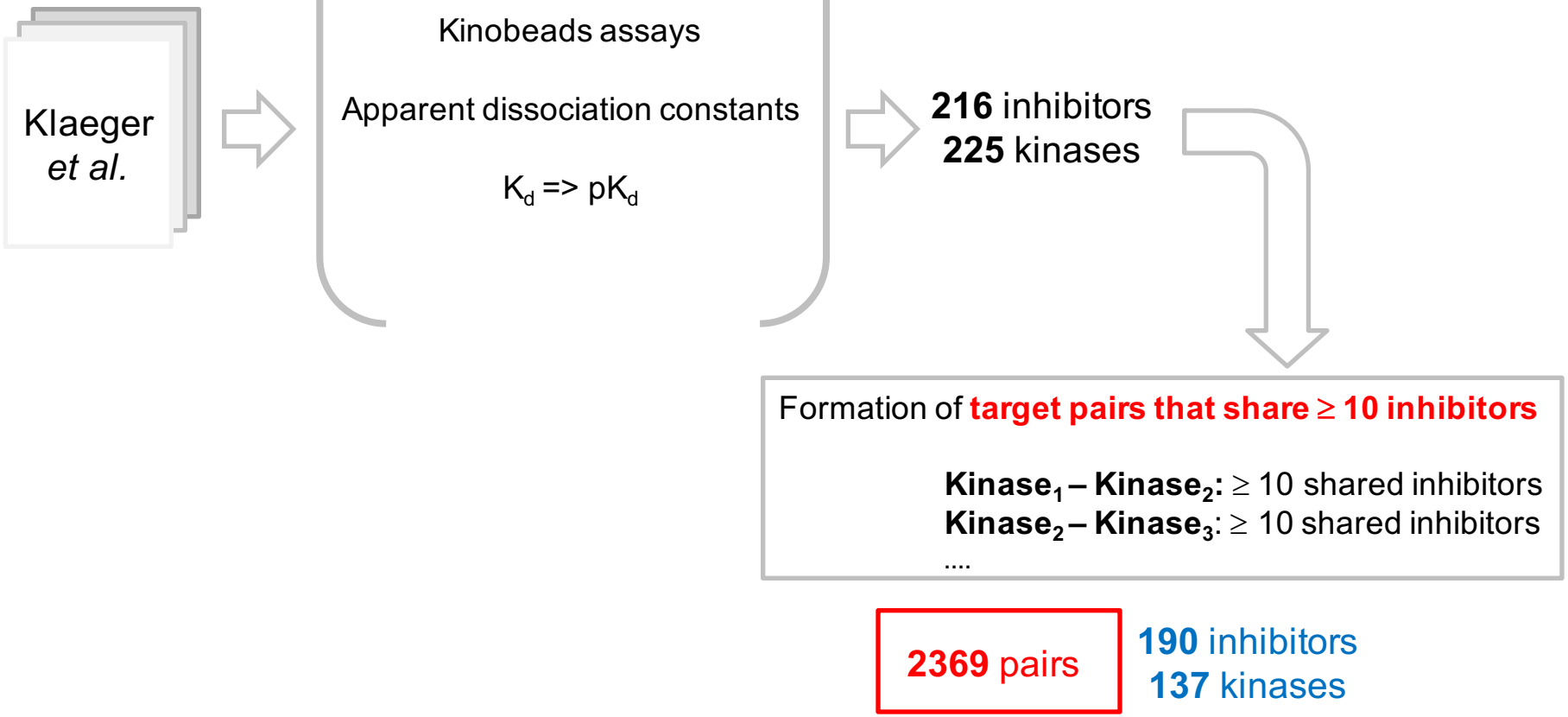








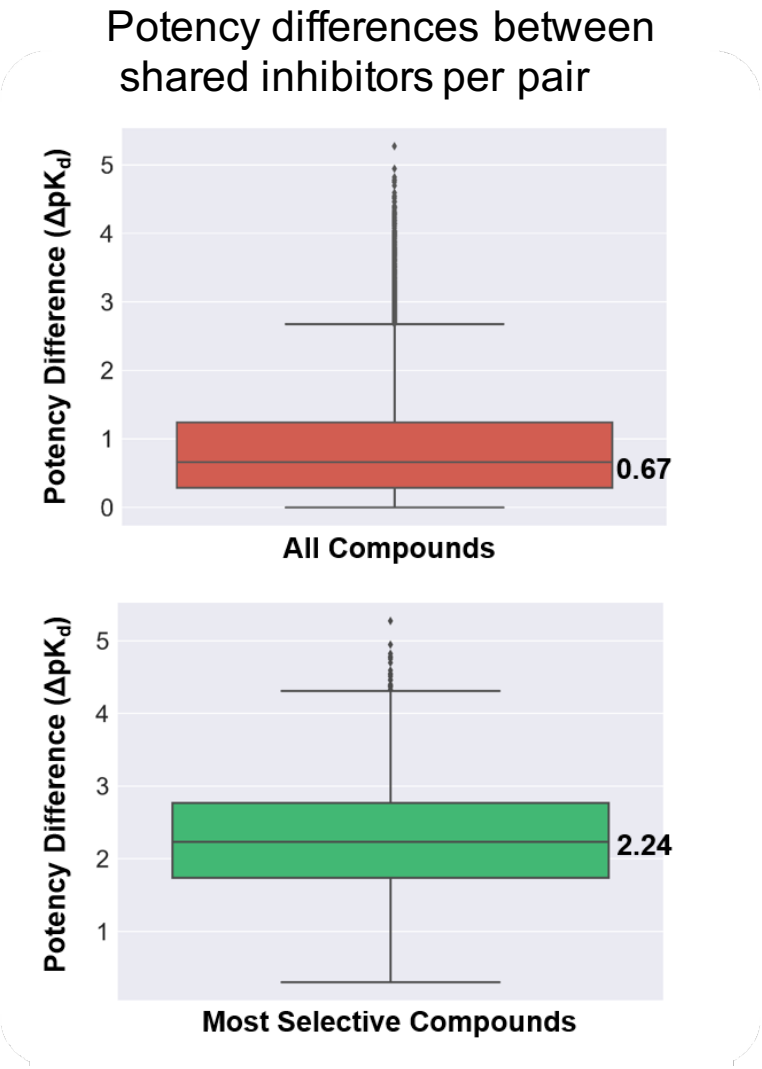
Kinase Pair-Based Selectivity Analysis



Kinase Inhibitors - Selectivity

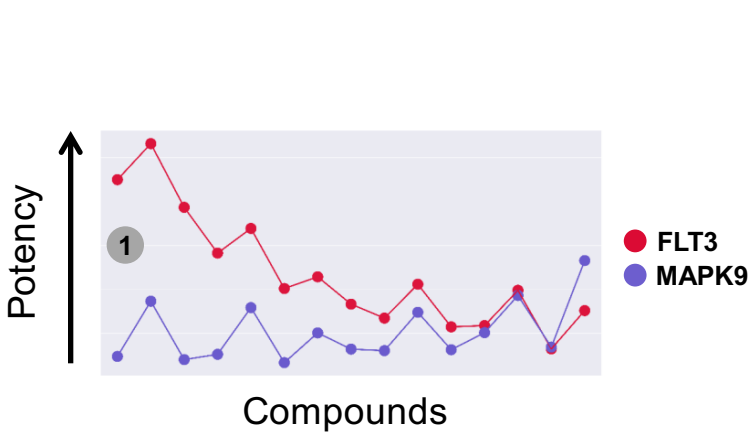
- Small global potency differences:
  median ΔpK_d value **0.67**
  (med. chem. data:
  median ΔpIC_{50} **0.69**)

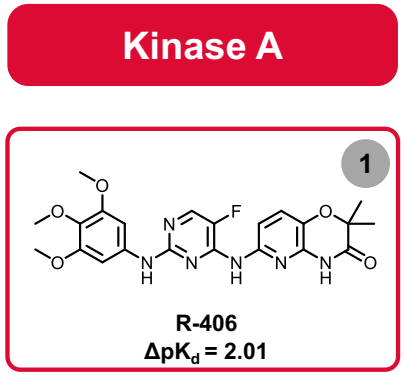
- Different picture for most selective inhibitors:
  median ΔpK_d value **2.24**
  (med. chem. data:
  median ΔpIC_{50} **2.37**)



Compound-Based Selectivity Profiles

- **Uni-directional** selectivity profiles revealed inhibitors with exclusive selectivity for one kinase over the other

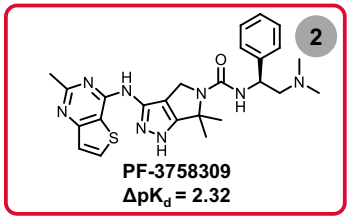
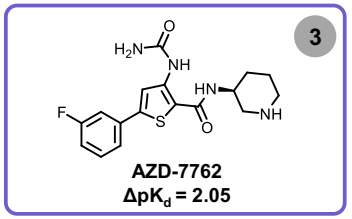
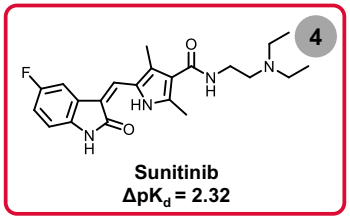
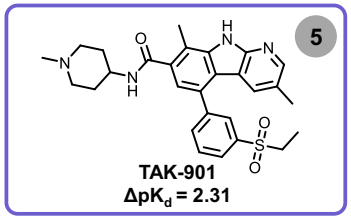
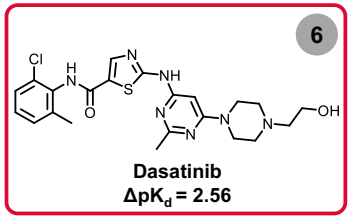
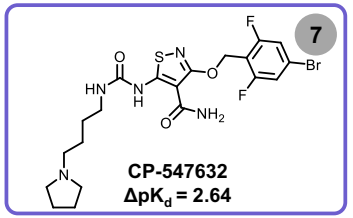




Compounds are ordered according to increasing potency differences for each kinase from the left to right and vice versa

Compound-Based Selectivity Profiles

- **Bi-directional** profiles uncovered inhibitors with inverted selectivity for paired kinases

		Kinase A	Kinase B
Same family			
Different families			
Different groups			

Conclusions

- Advent of the ‘big data‘ era in medicinal chemistry
- Kinase inhibitors as a representative example
- Most kinase inhibitors target the largely conserved ATP binding site
- Promiscuity and lack of selectivity are anticipated
- Systematic activity data analysis is strongly influenced by data confidence criteria
- Varying confidence criteria and inclusion of screening data put the data sparseness issue into perspective

Conclusions

- Data analysis reveals that ATP site-directed kinase inhibitors are less promiscuous and more specific than often assumed

- In part surprising agreement of computational selectivity analysis on the basis of activity data from medicinal chemistry and cell-based profiling