Supporting Information: Figures

AN UNPRECEDENTEDLY LARGE-SCALE KINASE INHIBITOR SET ENABLING THE ACCURATE PREDICTION OF COMPOUND-KINASE ACTIVITIES: A WAY TOWARDS SELECTIVE PROMISCUITY BY DESIGN?

Serge Christmann-Franck[*†$], *Gerard J.P. van Westen*[‡#], *George Papadatos*[‡], *Fanny Beltran Escudie*[†&], *Alexander Roberts*[†¥], *John P. Overington*[‡%], *Daniel Domine*[†@]

[†] Merck Serono, Chemin des Mines 9, 1202 Genève, Switzerland
[‡] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

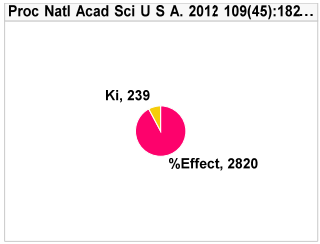
Corresponding Author
* serge.christmann-franck@hotmail.fr

Present Addresses:
[$] Novartis Institutes for Biomedical Research, Postfach, CH-4002, Basel, Switzerland
[#] Medicinal Chemistry, Leiden Academic Center for Drug Research, Leiden University, Einsteinweg 55, 2333CC, Leiden, The Netherlands
[&] Medicines for Malaria Venture, Route de Pré-Bois 20, 1215 Meyrin, Switzerland
[¥] EMD Serono, 45 Middlesex Turnpike, Billerica, MA 01821, USA
[%] Stratified Medical, 40 Churchway, London, NW1 1LM, UK
[@] Wega Informatik AG, Aeschengraben 20, 4051 Basel, Switzerland

Figure S1. Comparison across the data sources of the relative numbers of measurements and of their distribution according to the results types

Figure S1. Comparison across the data sources of the relative numbers of measurements and of their distribution according to the results types

Figure S1. Comparison across the data sources of the relative numbers of measurements and of their distribution according to the results types



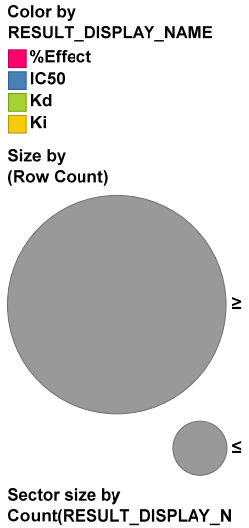
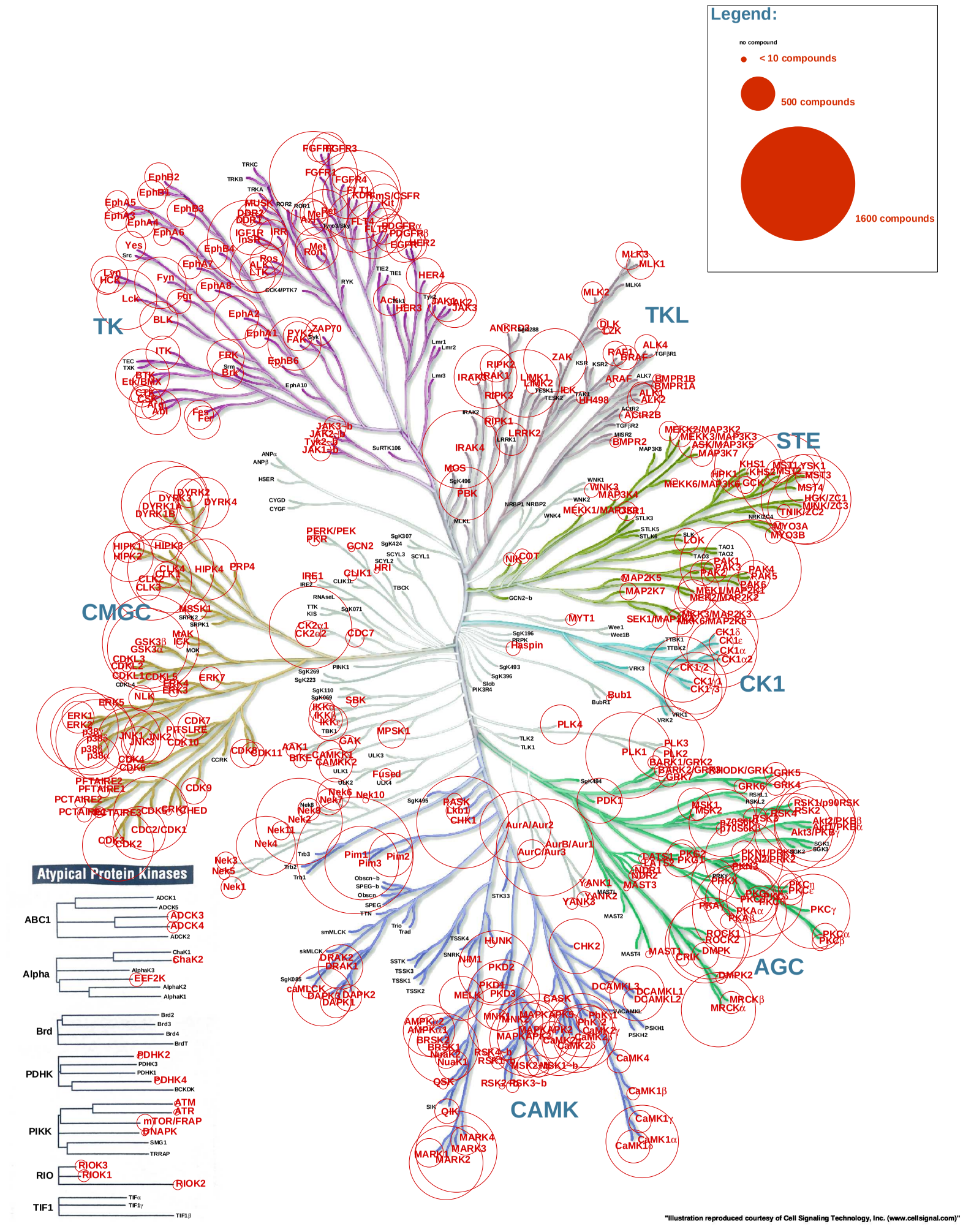


Figure S1. Comparison across the data sources of the relative numbers of measurements and of their distribution according to the results types

Figure S2. (high resolution version of Fig 3). Coverage of the kinome by the current dataset; the kinase names and circles sizes are proportional to the number of corresponding compounds. The picture was generated using the Kinome Render[4] and the kinome tree illustration is reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)


Legend:
no compound
< 10 compounds
500 compounds
1600 compounds
TK
TKL
STE
CMGC
CK1
AGC
CAMK
Atypical Protein Kinases
ABC1
Alpha
Brd
PDHK
PIKK
RIO
TIF1
"Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)"

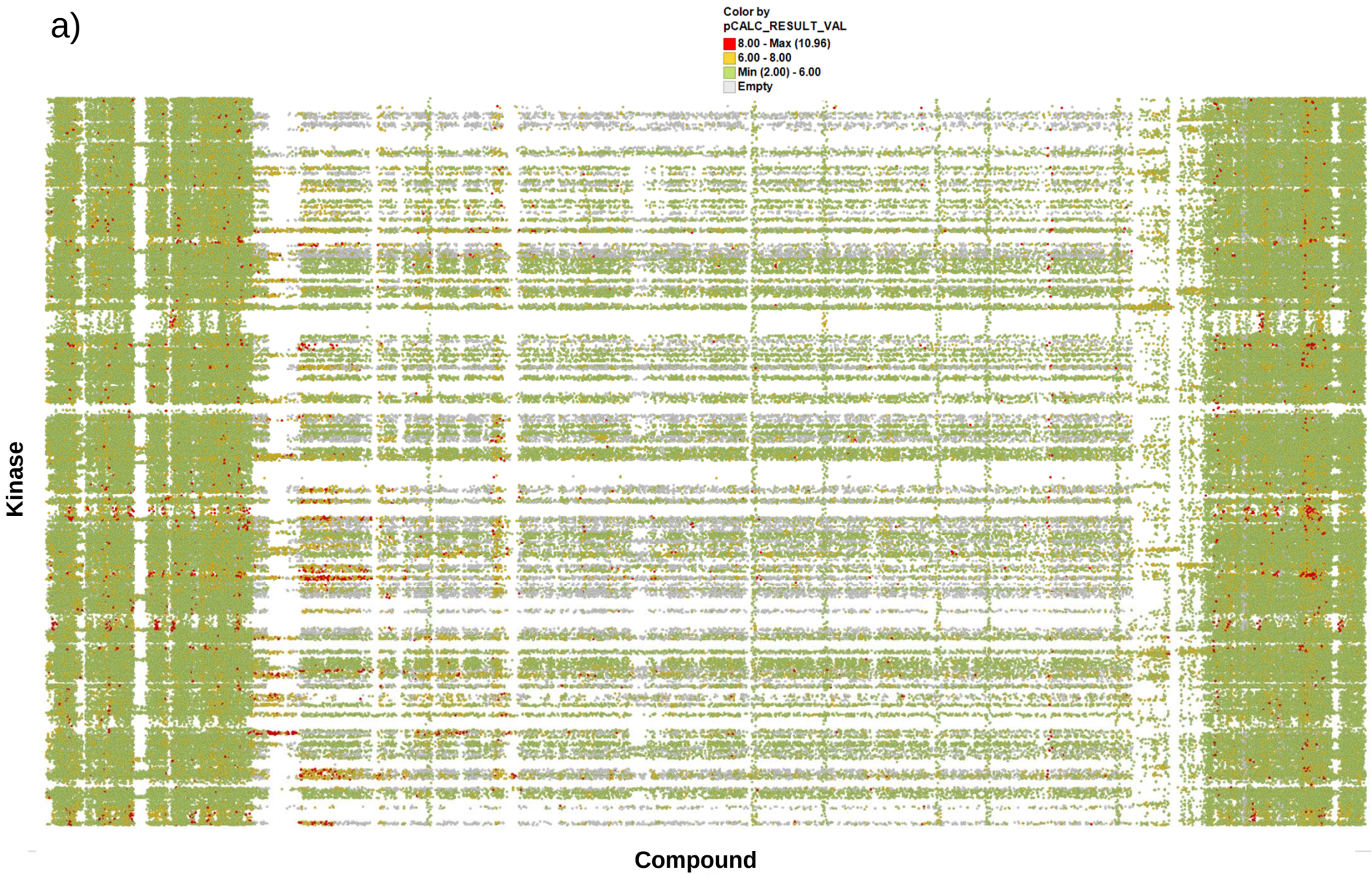


Figure S3. Coverage across data sources, at the Kinase level: (a) complete dataset

b)

Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

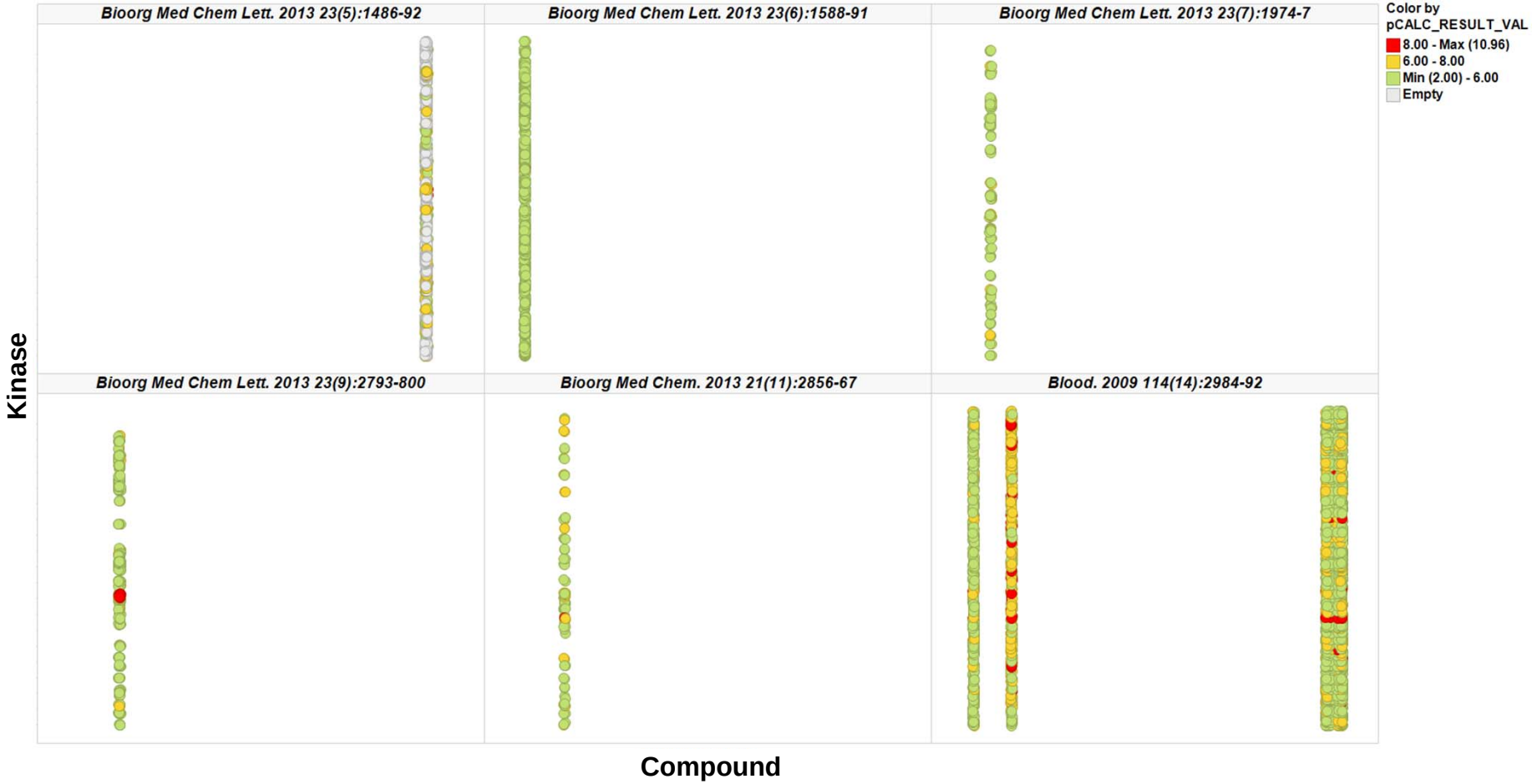


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

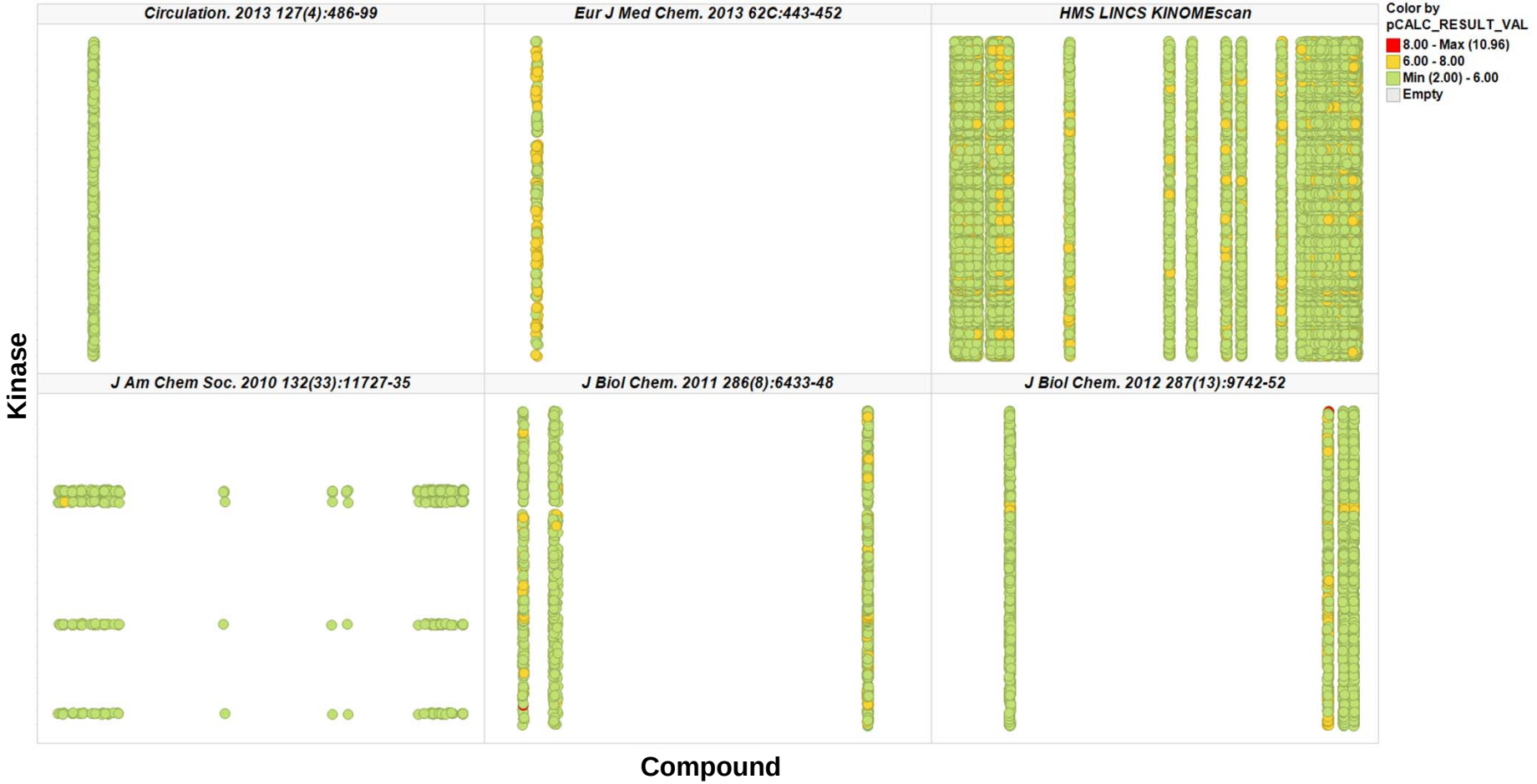


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

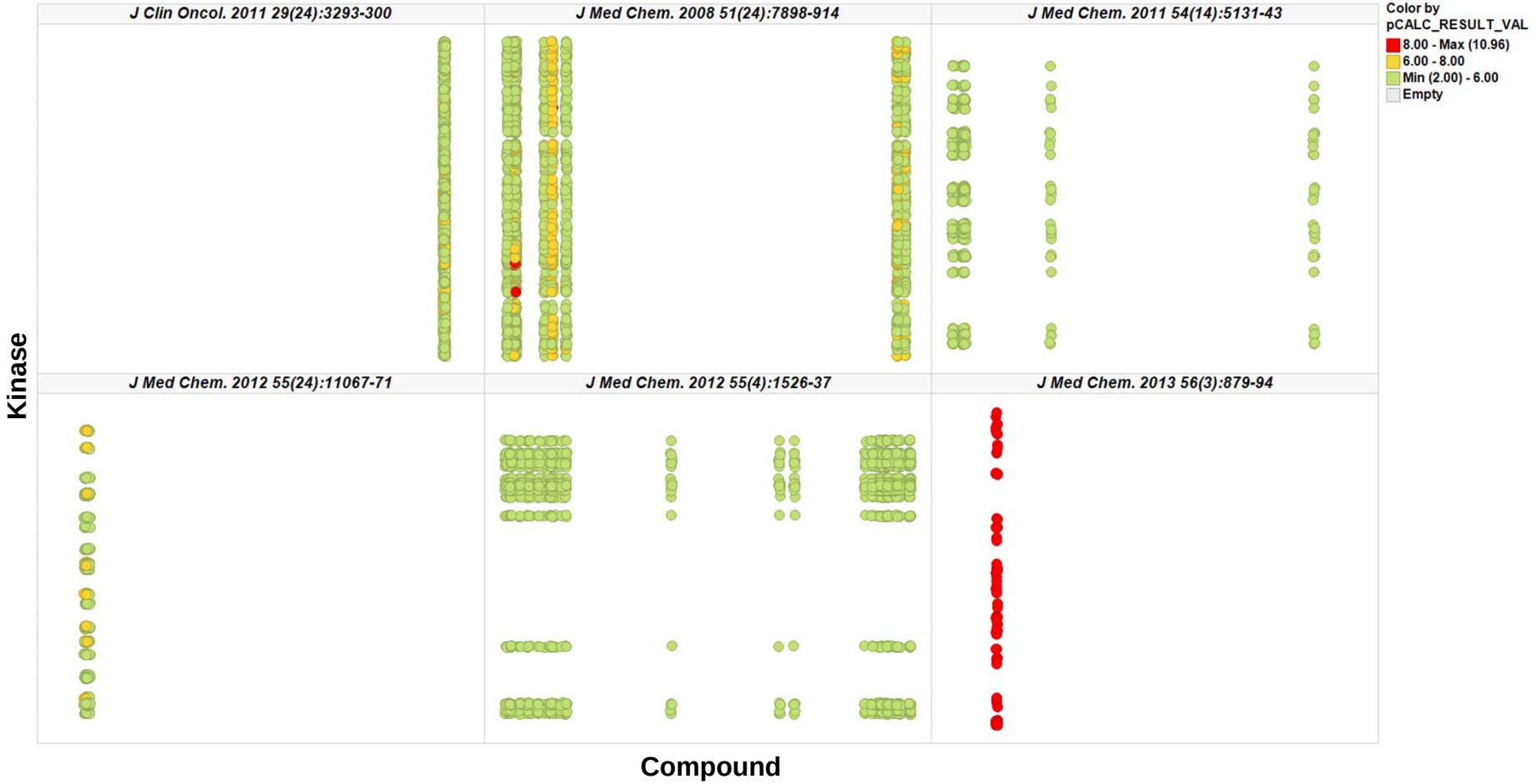


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

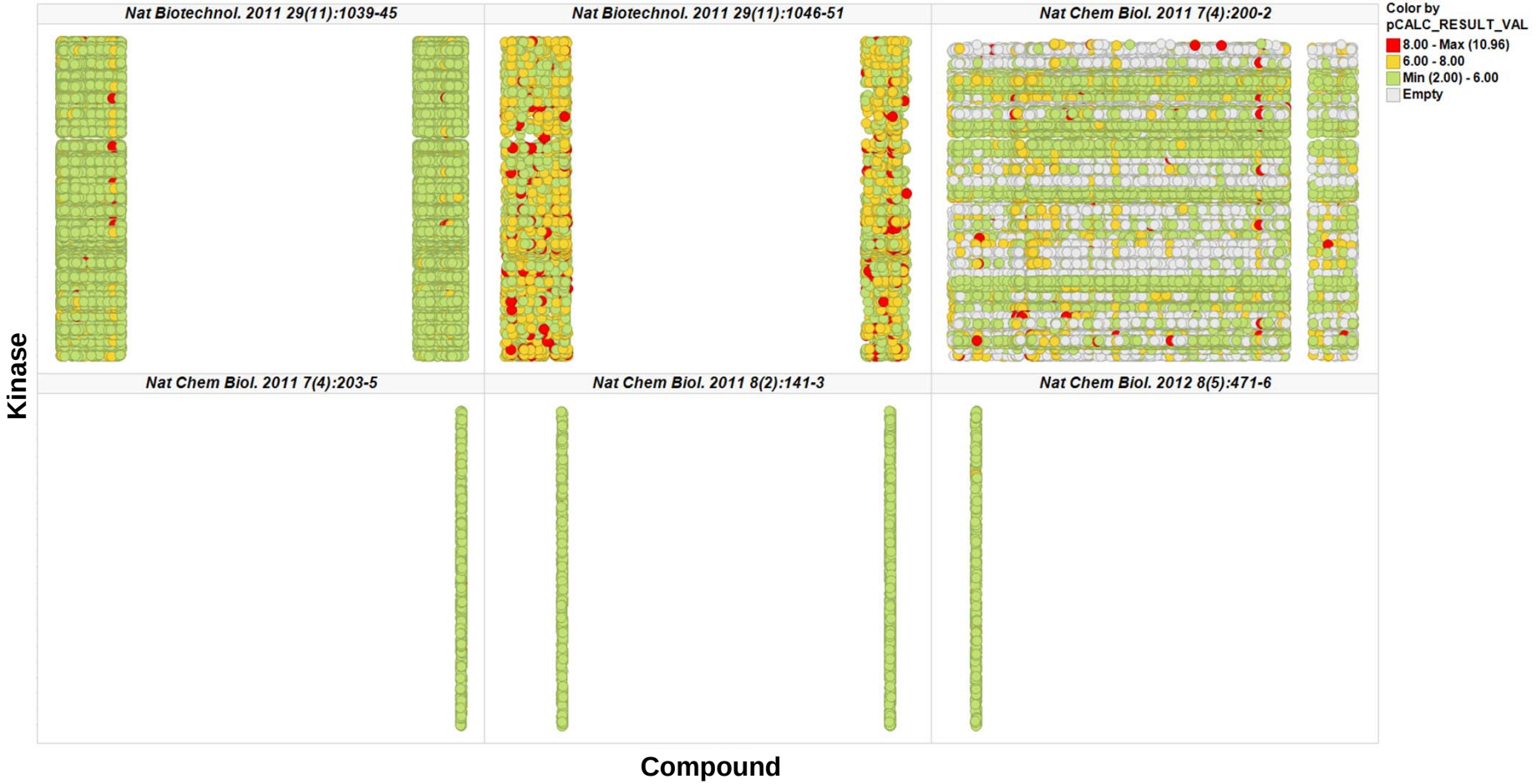


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

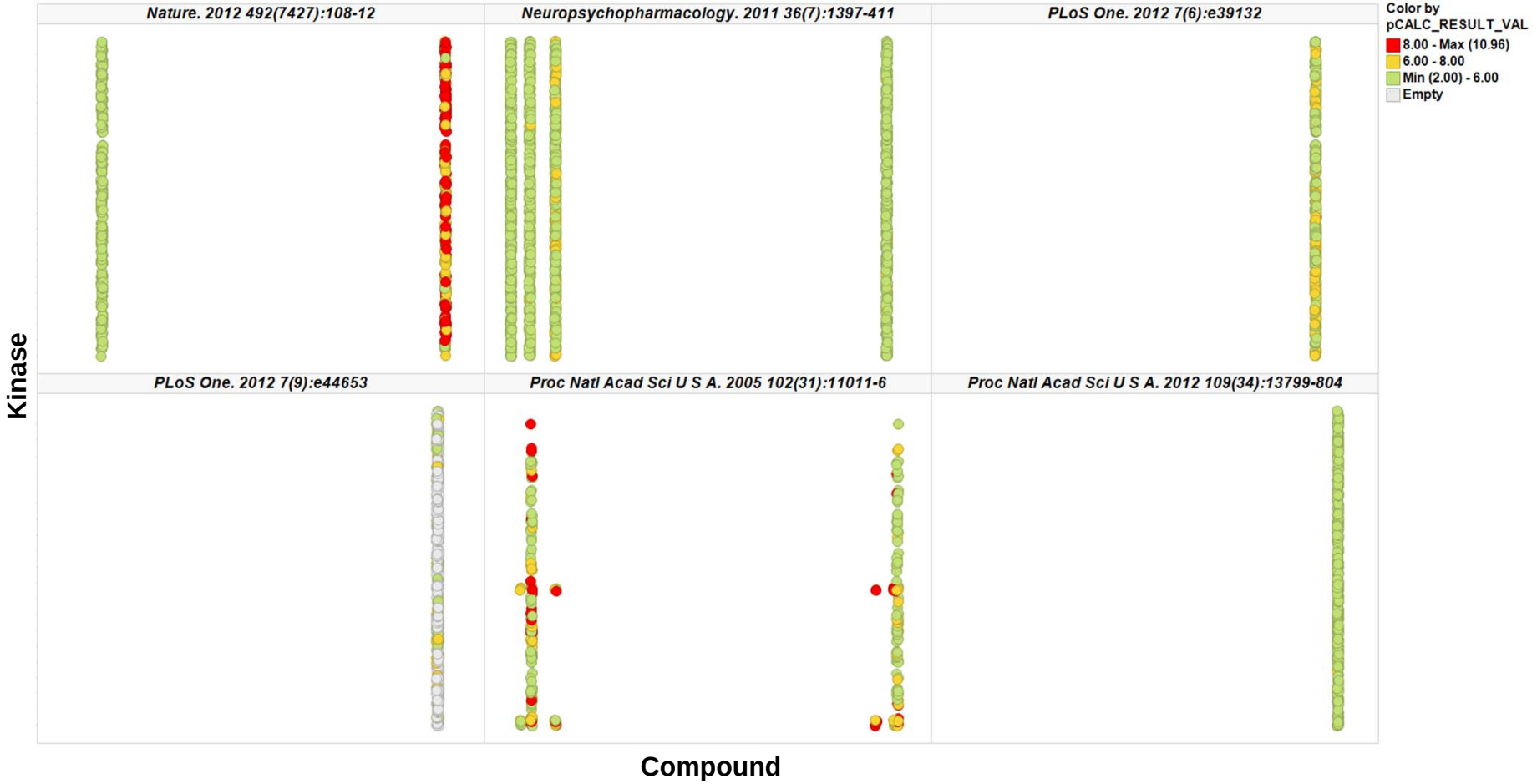


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

b)

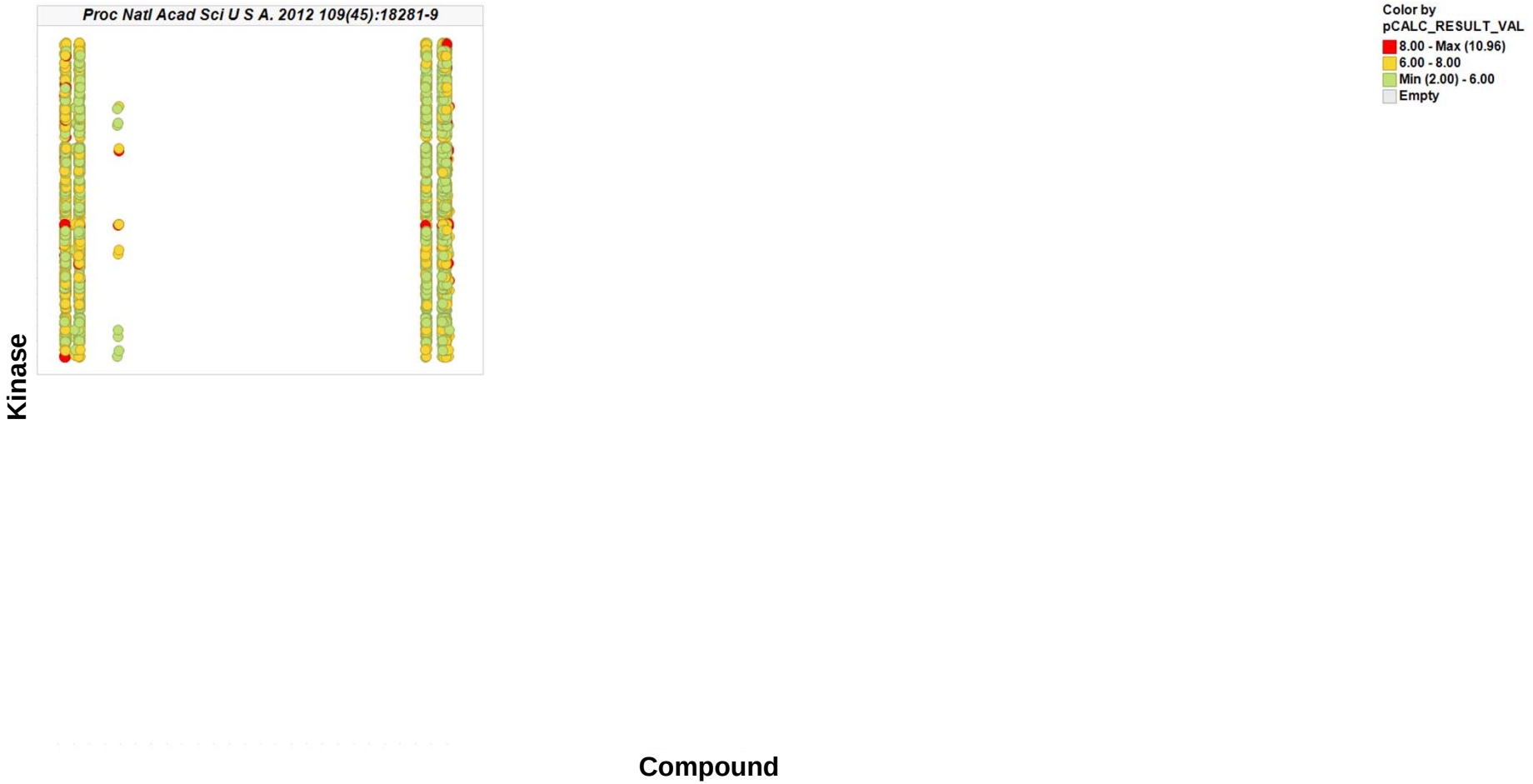


Figure S3. Coverage across data sources, at the Kinase level: (b) per data source

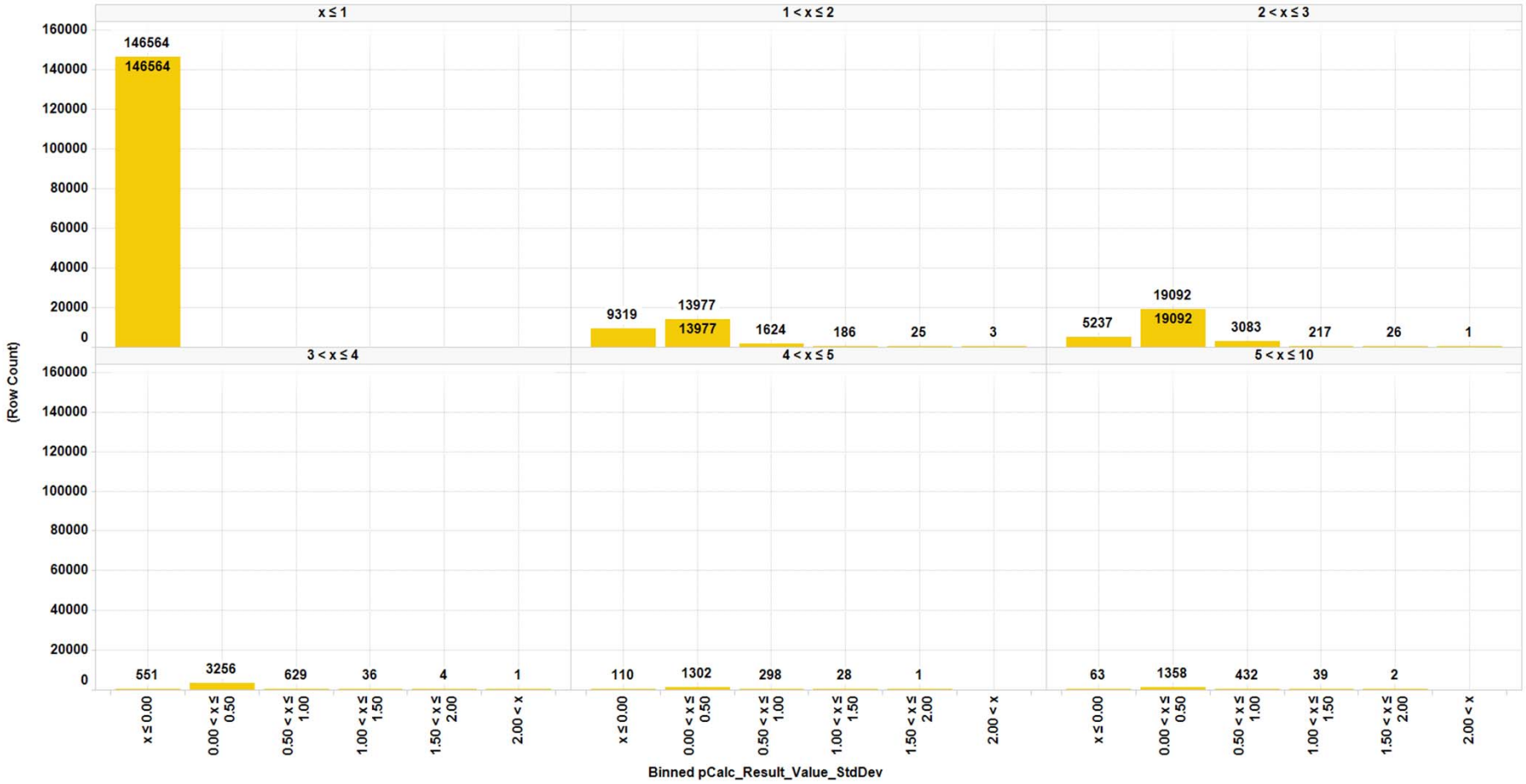


Figure S4. Distribution of the binned pStandardized_Result_Value_SD, split by the number of measurements per Kinase-compound pair

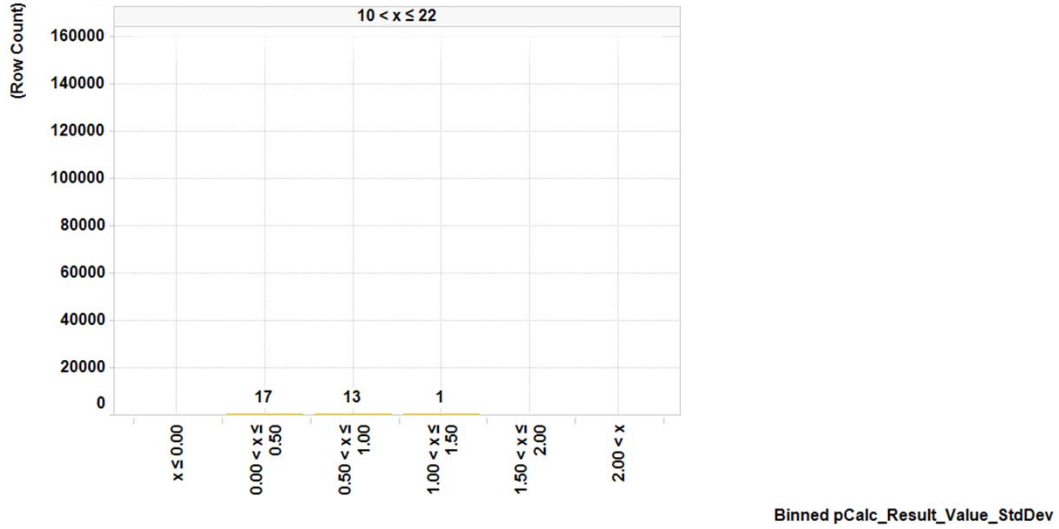


Figure S4. Distribution of the binned pStandardized_Result_Value_SD, split by the number of measurements per Kinase-compound pair

Figure S5. Coverage of the kinome by compounds from the current dataset PKIs exhibiting a STD_RESULT_VALUE of 100 nM or less; the Kinases names and circles sizes are proportional to the number of corresponding measurements. The picture was generated using the Kinome Render[4] and the kinome tree illustration is reproduced courtesy of Cell Signaling Technology, Inc (www.cellsignal.com)

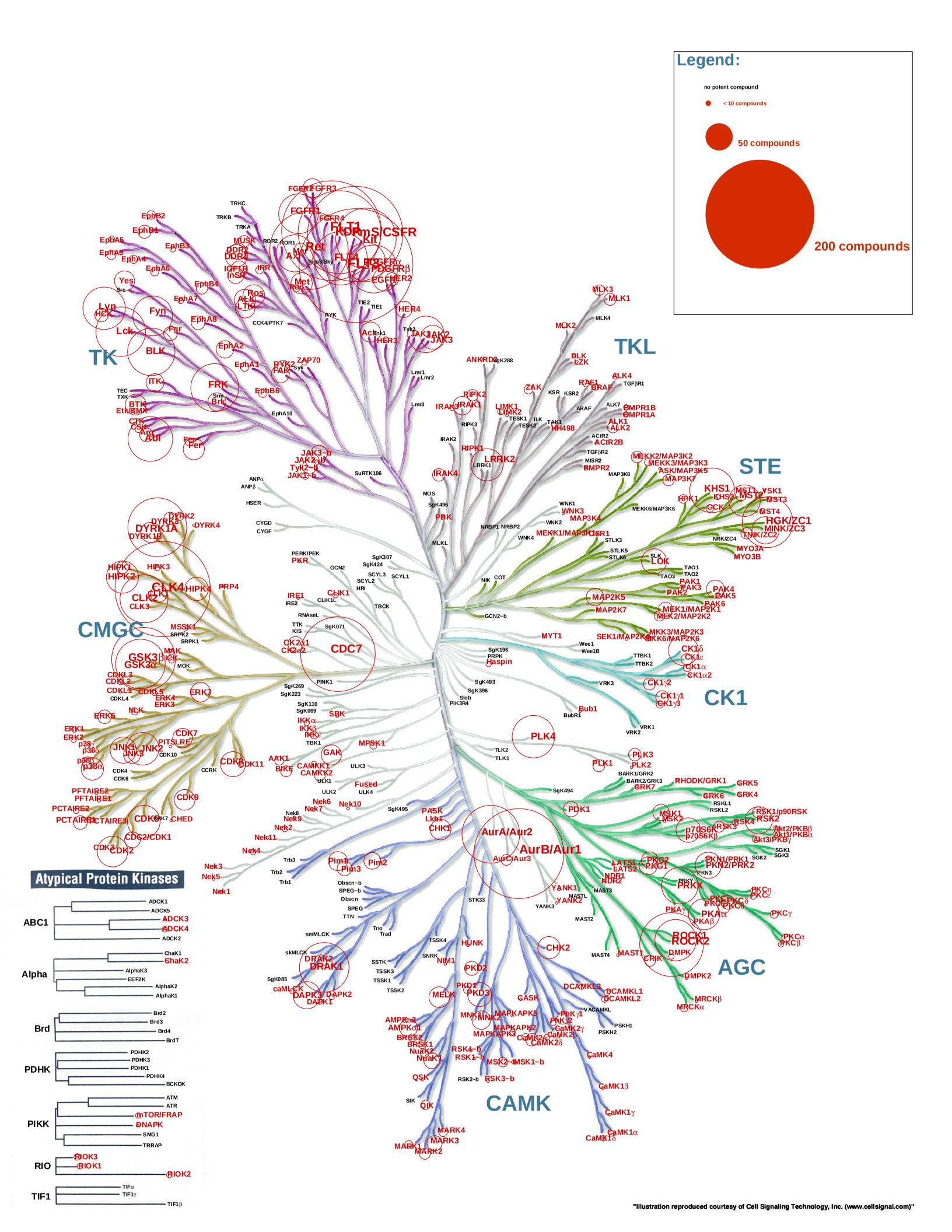
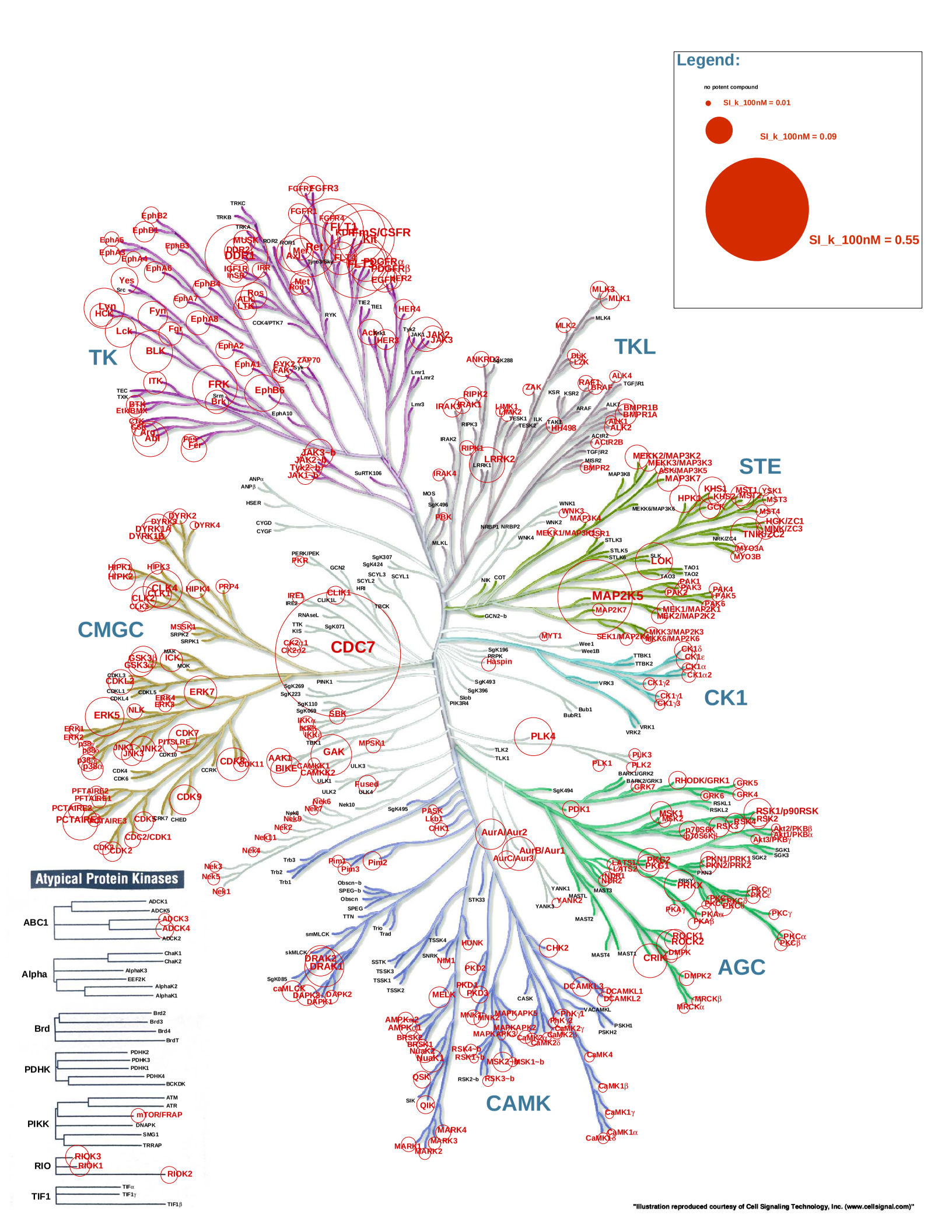

Legend:
no potent compound
< 10 compounds
50 compounds
200 compounds
TK
TKL
STE
CMGC
CK1
AGC
CAMK
Atypical Protein Kinases
ABC1
Alpha
Brd
PDHK
PIKK
RIO
TIF1
"Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)"

Figure S6. Spread of the SI_k_100nM values across the kinome, for the kinases with at least 100 compounds tested; the kinase names and circles sizes are proportional to the SI_k_100nM values. The picture was generated using the Kinome Render[4] and the kinome tree illustration is reproduced courtesy of Cell Signaling Technology, Inc (www.cellsignal.com)


Legend:
no potent compound
SI_k_100nM = 0.01
SI_k_100nM = 0.09
SI_k_100nM = 0.55
TK
TKL
STE
CMGC
CDC7
CK1
AGC
CAMK
Atypical Protein Kinases
ABC1
Alpha
Brd
PDHK
PIKK
RIO
TIF1
"Illustration reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)"

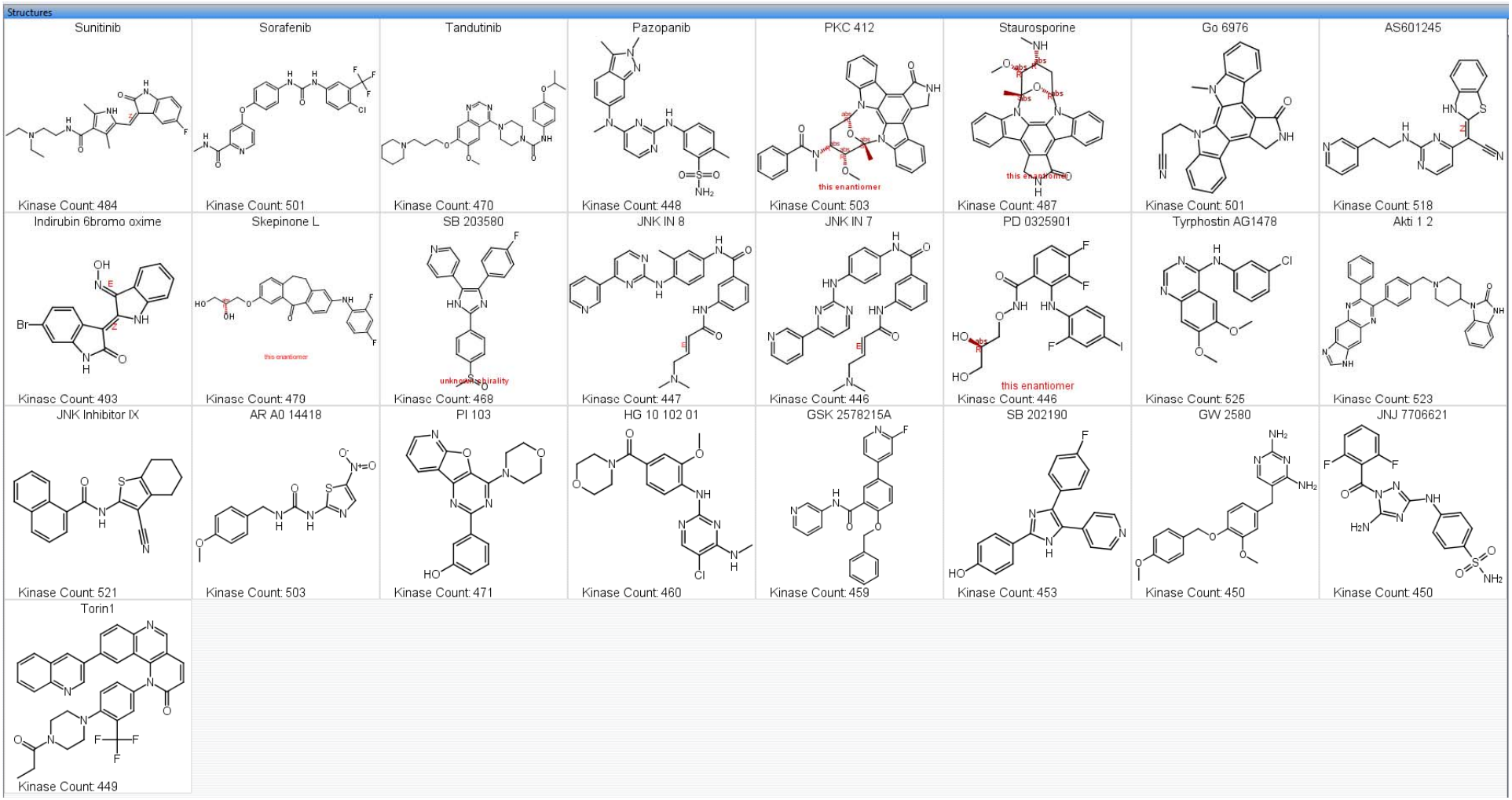


Figure S7. Structures of the 25 PKIs exhibiting the highest number of Kinases

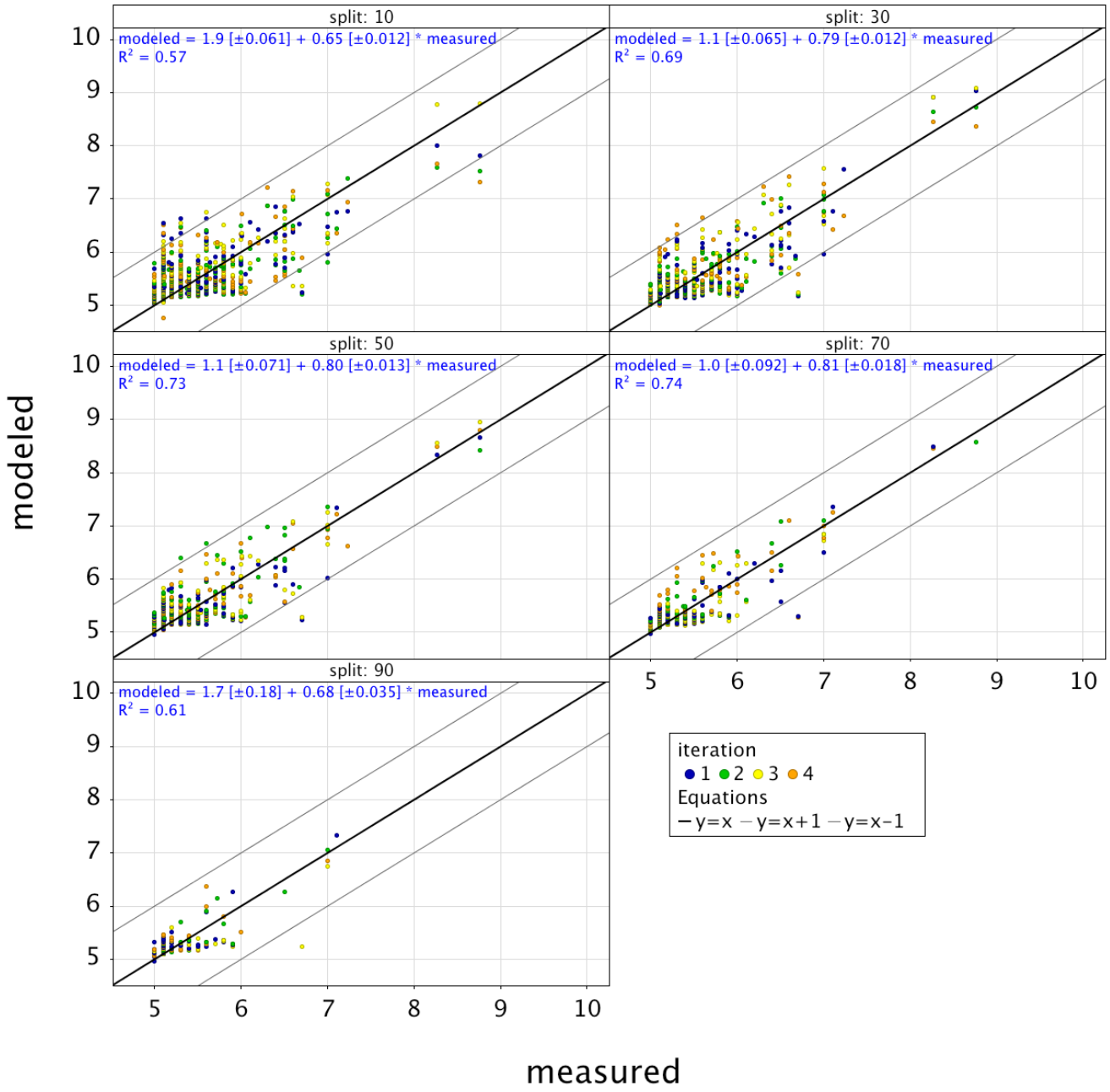


Figure S8. Example of a well modeled kinase: CAMK2B

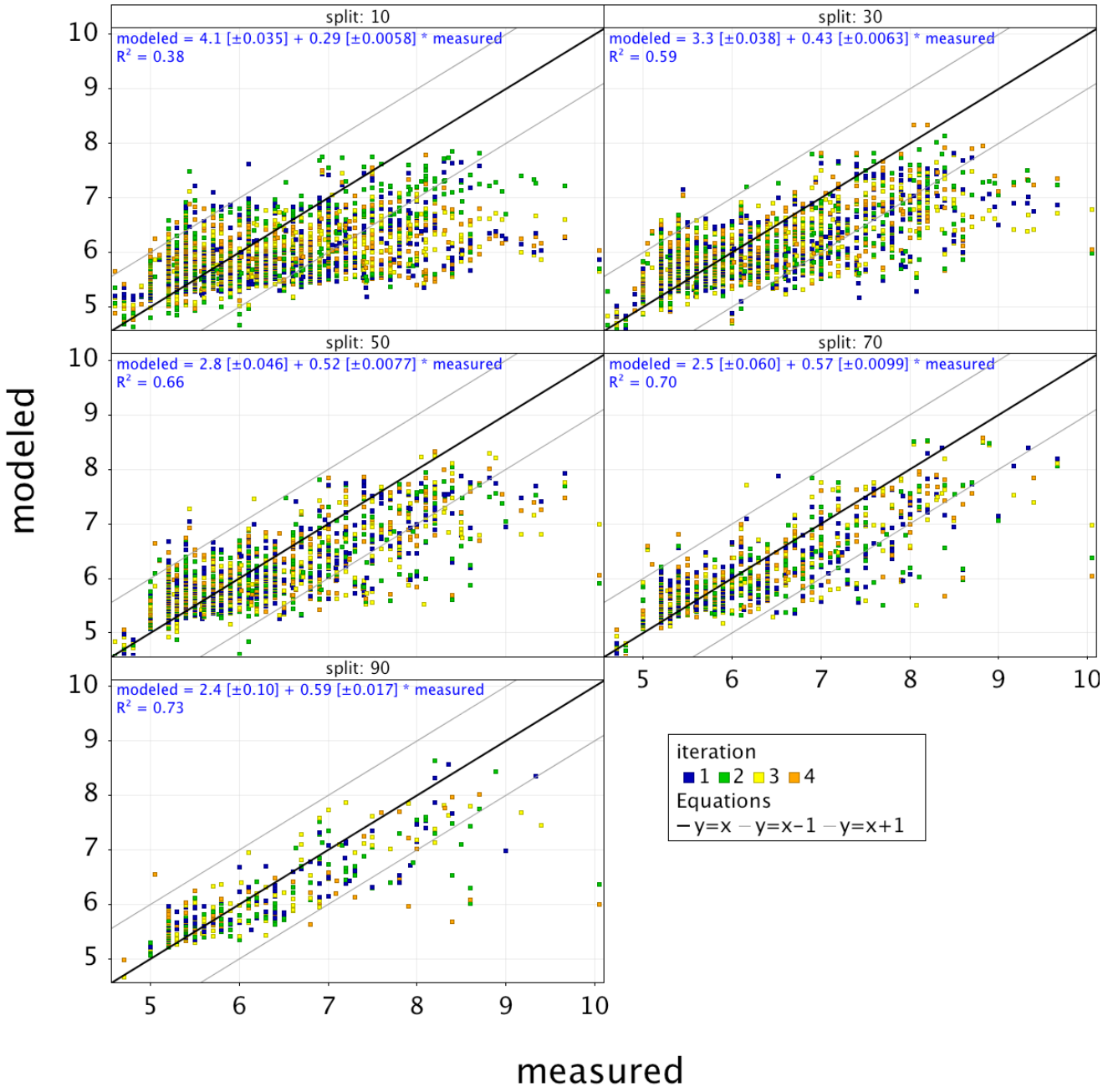


Figure S9. Example of a poorly modeled kinase: GSK3B

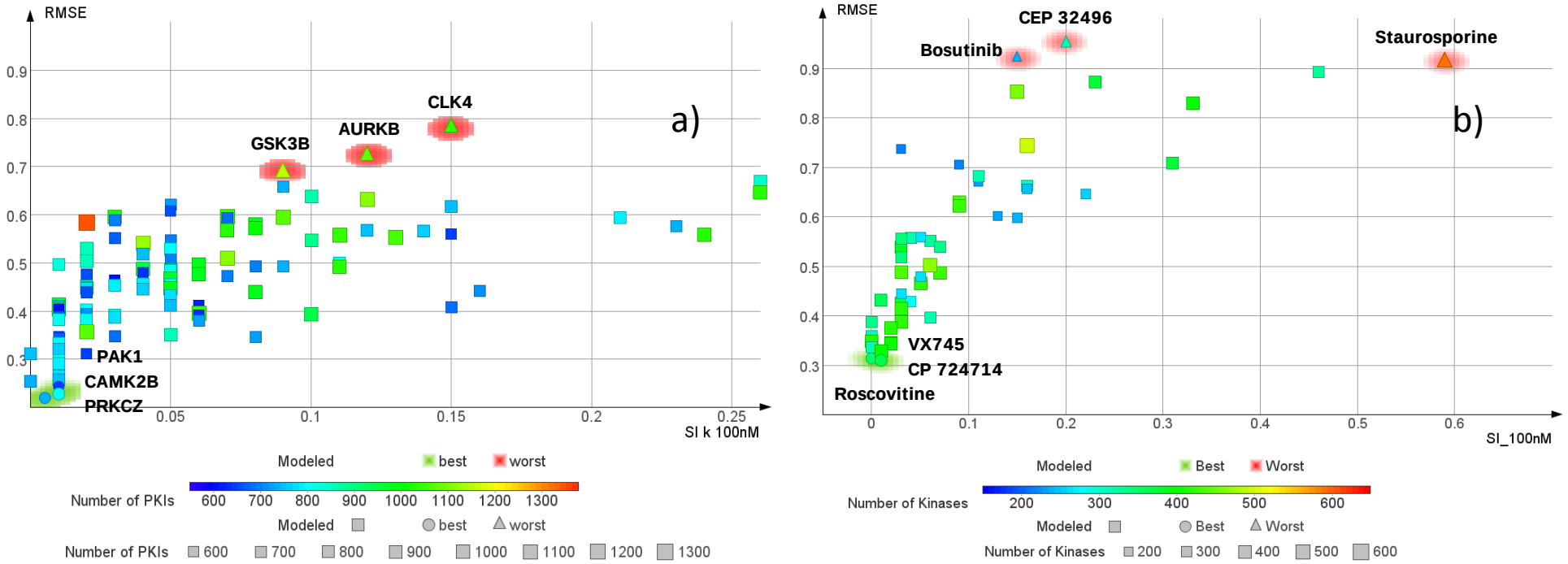


Figure S10. Comparison between RMSE and SI_(k)_100nM values for the Kinases

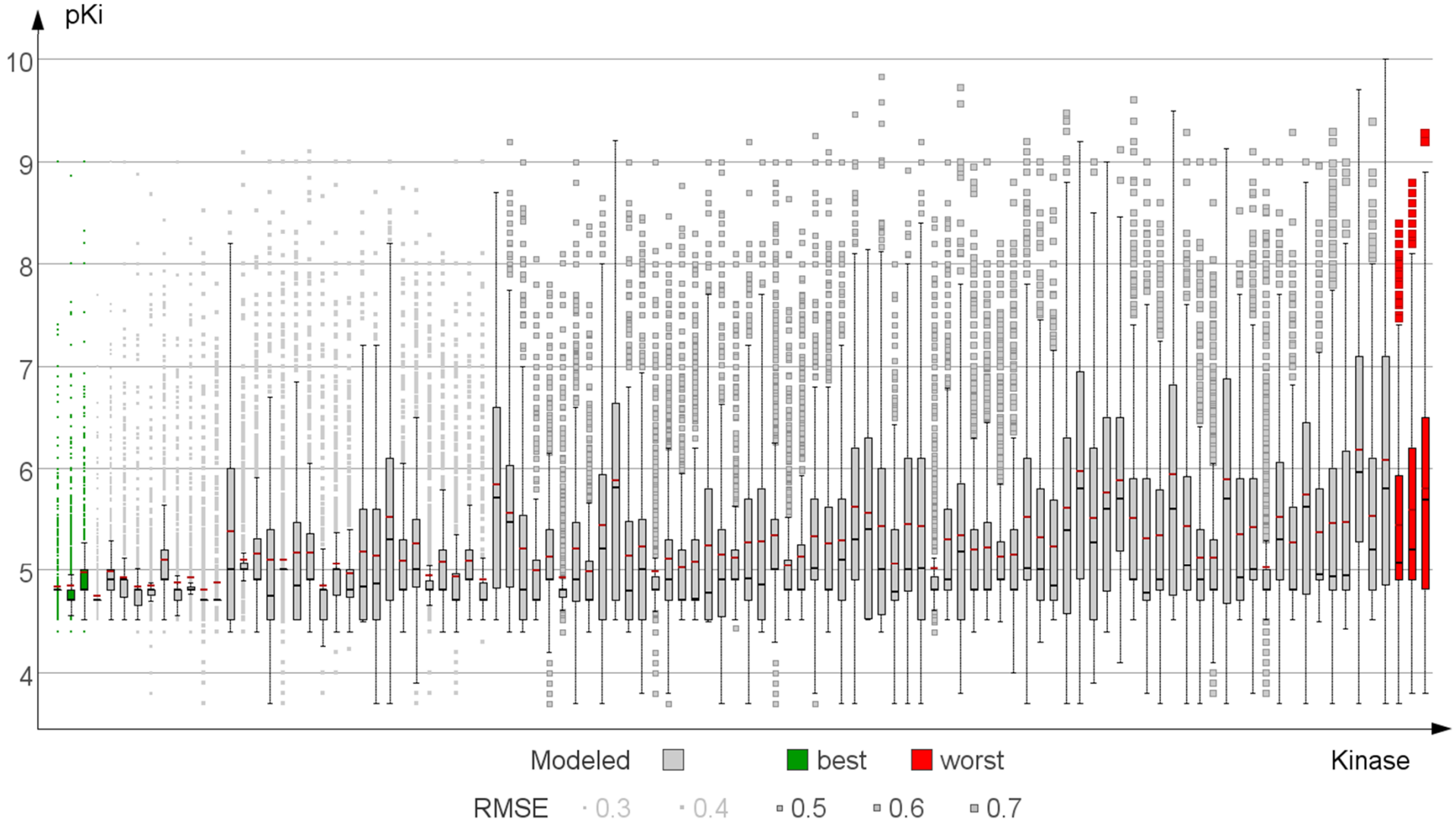


Figure S11. Influence of the spread of pKi values on the RMSE for the Kinases. The Kinases (x axis) are ordered by increasing RMSE value. The boxplots depict the first and third quartiles of the data, and the red and black lines correspond to the mean and median pKi values respectively

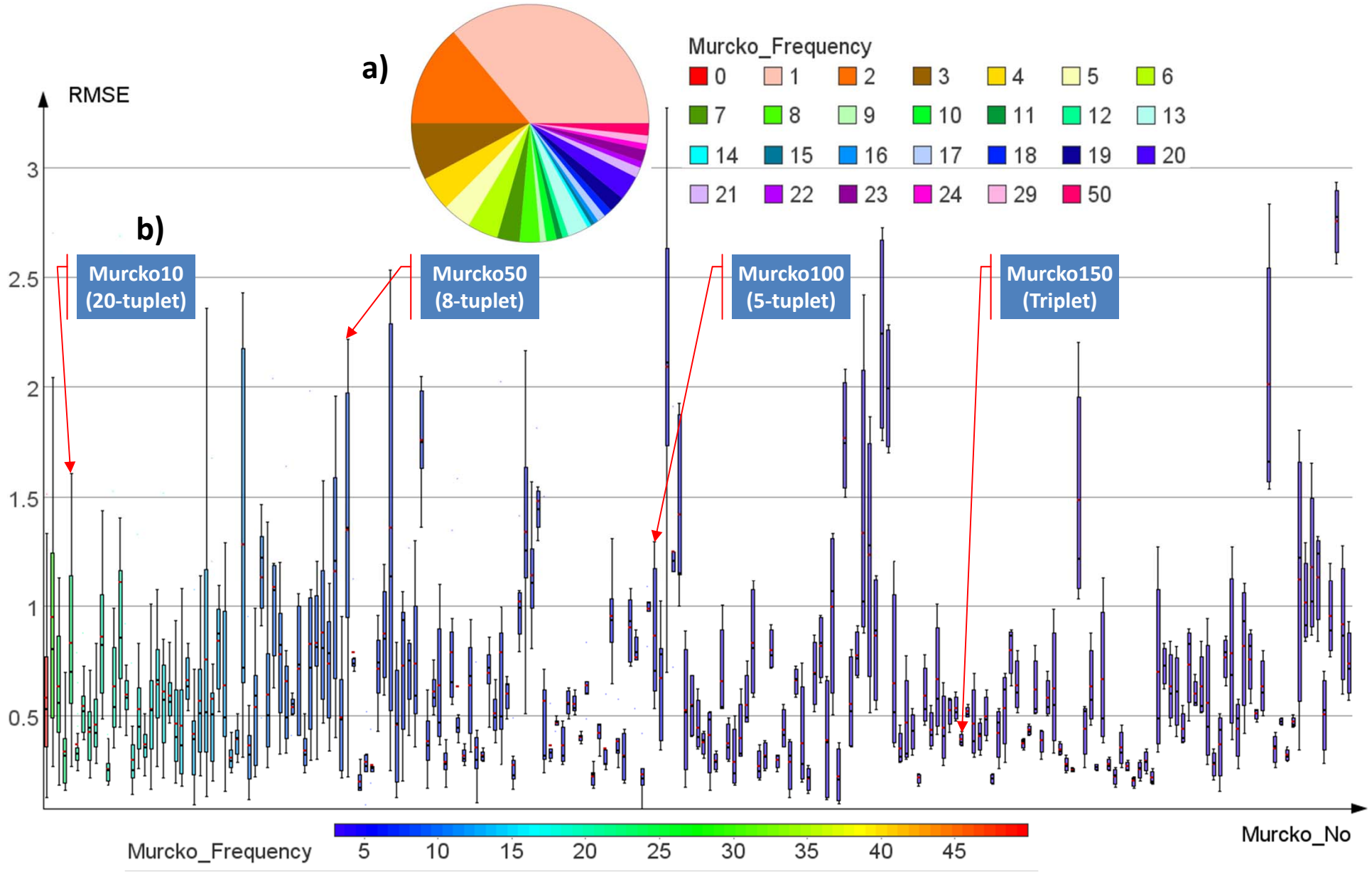


Figure S12. a) (inset): relative populations of singletons, pairs, etc…
b) Boxplot of the RMSE vs the Murcko fragment No (the Murcko fragments have been numbered by decreasing population: Murcko N1 is the most populated one with 22 compounds)

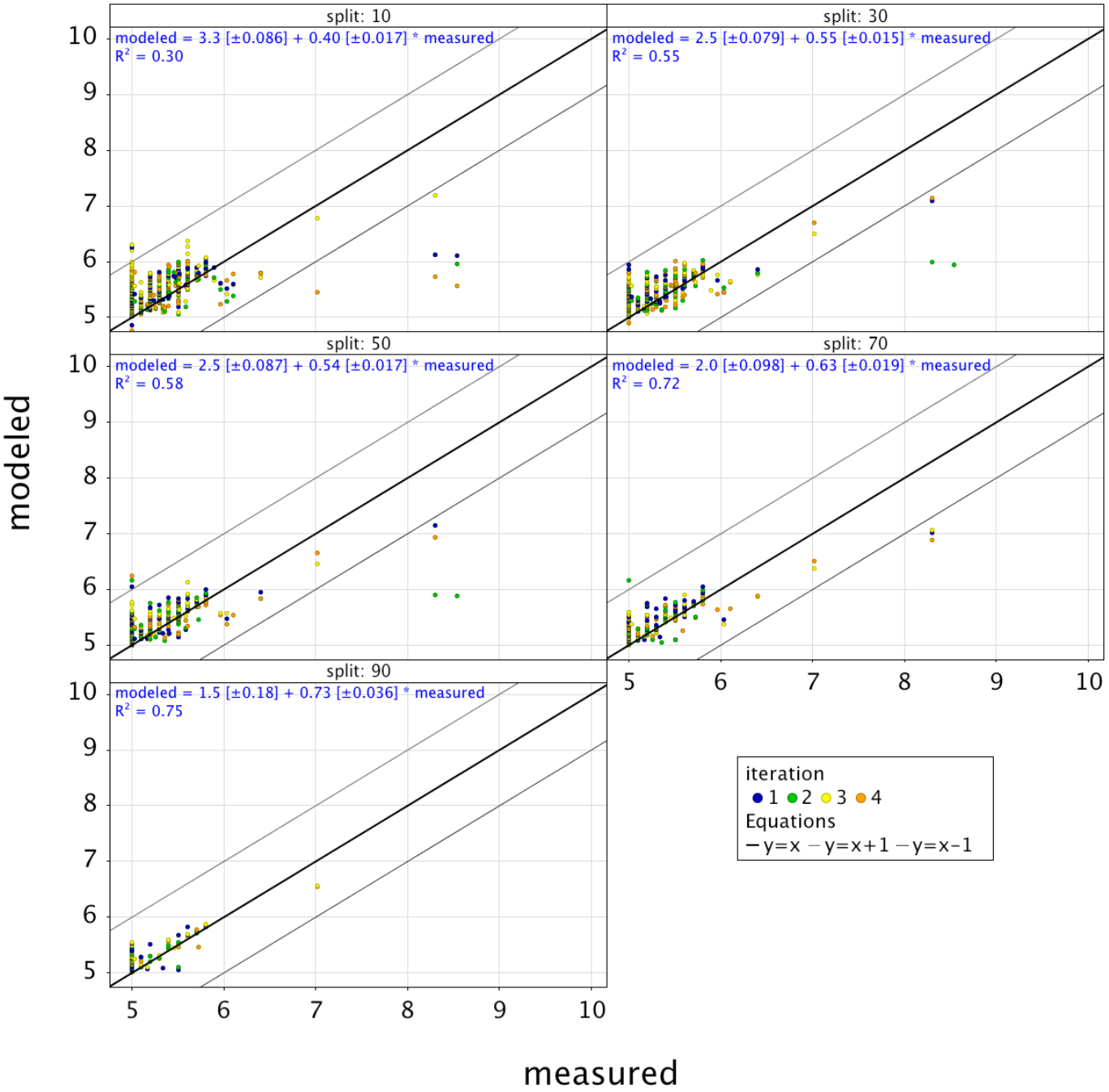


Figure S13. Example of a well modeled compound: VX745

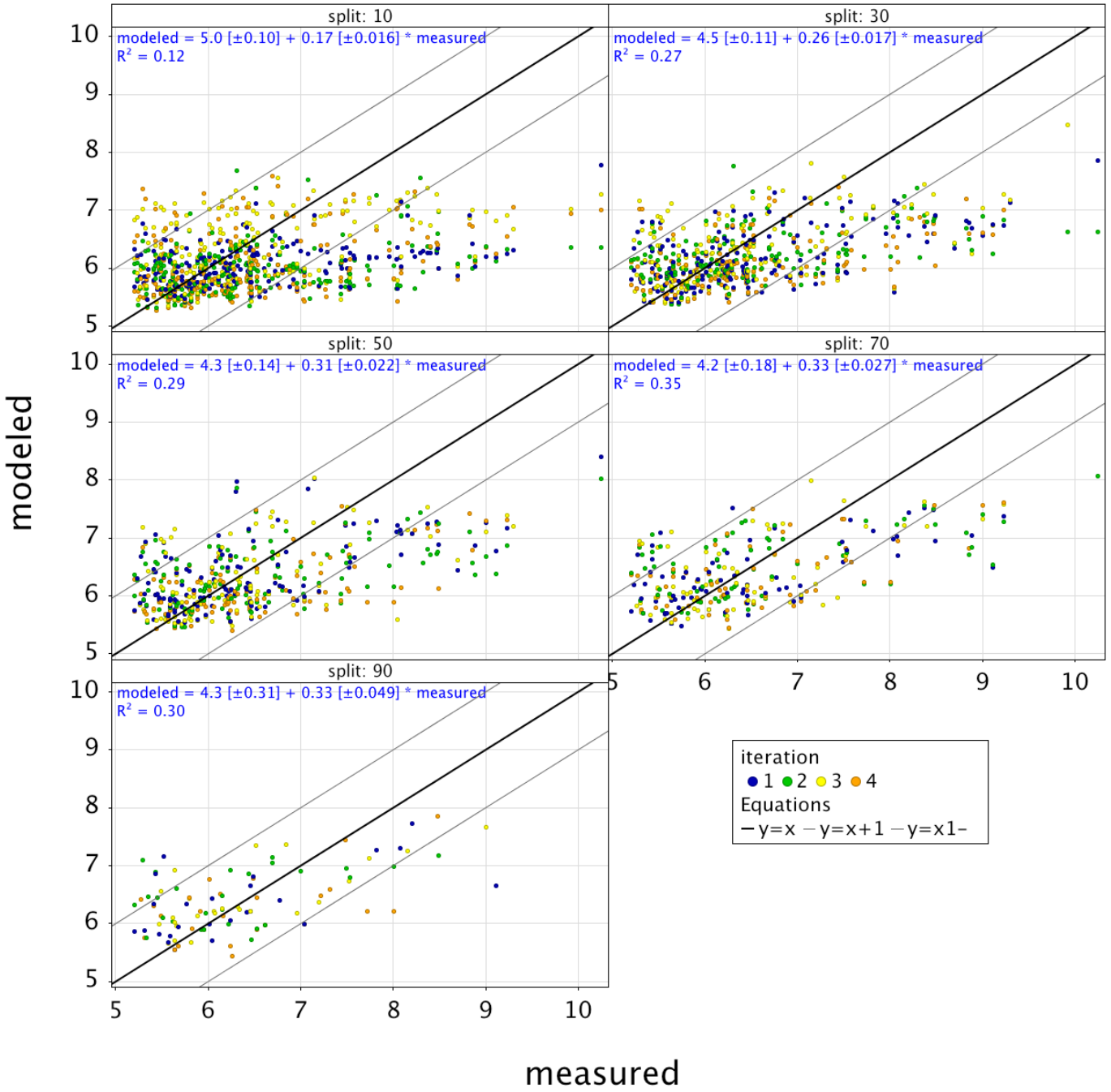


Figure S14. Example of a poorly modeled compound: Bosutinib

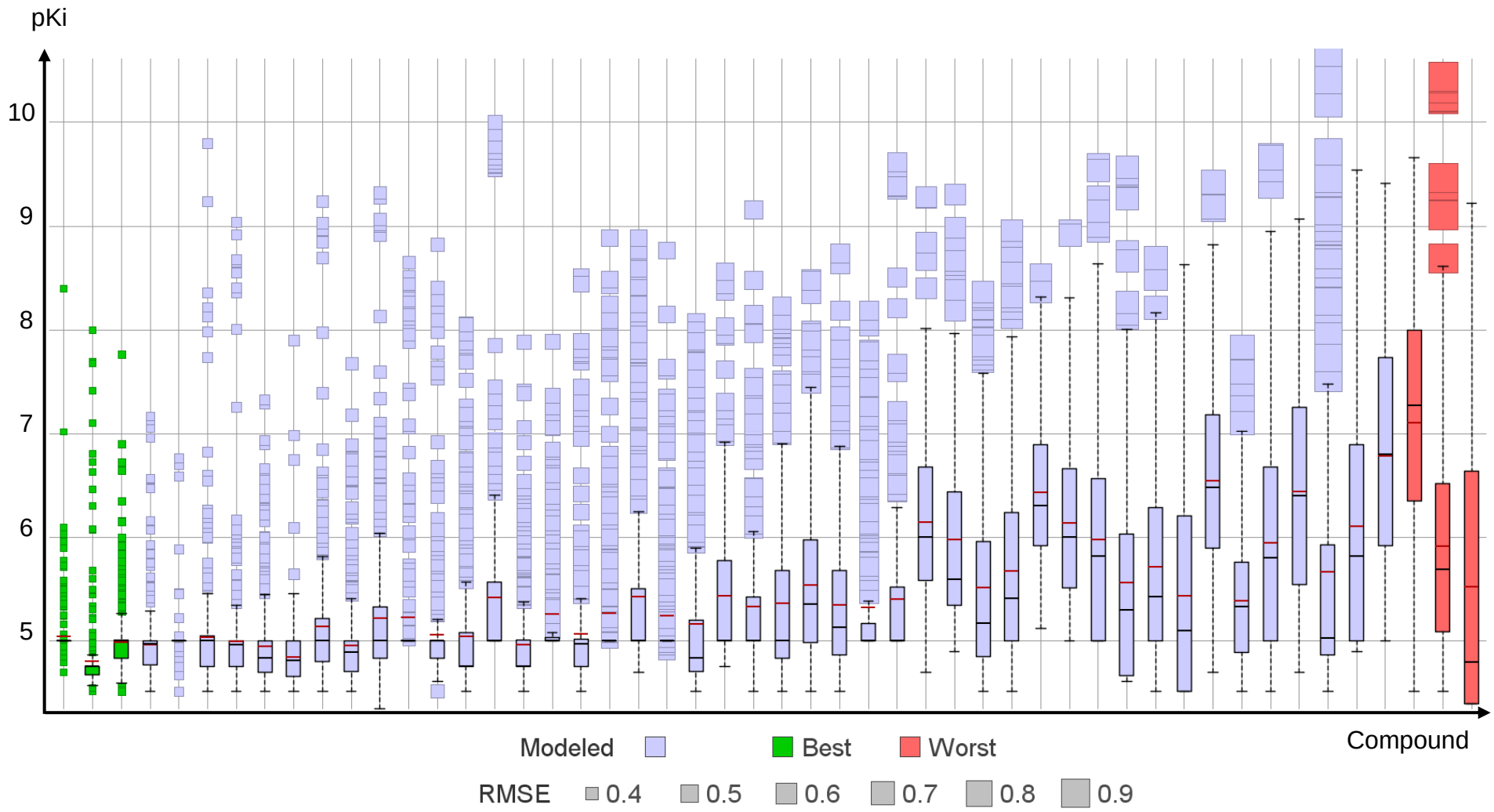


Figure S15. Influence of the spread of pKi values on the RMSE for the PKIs. The PKIs (x axis) are ordered by increasing RMSE value. The box plots depict the first and third quartiles of the data, and the red and black lines correspond to the mean and median pKi values respectively

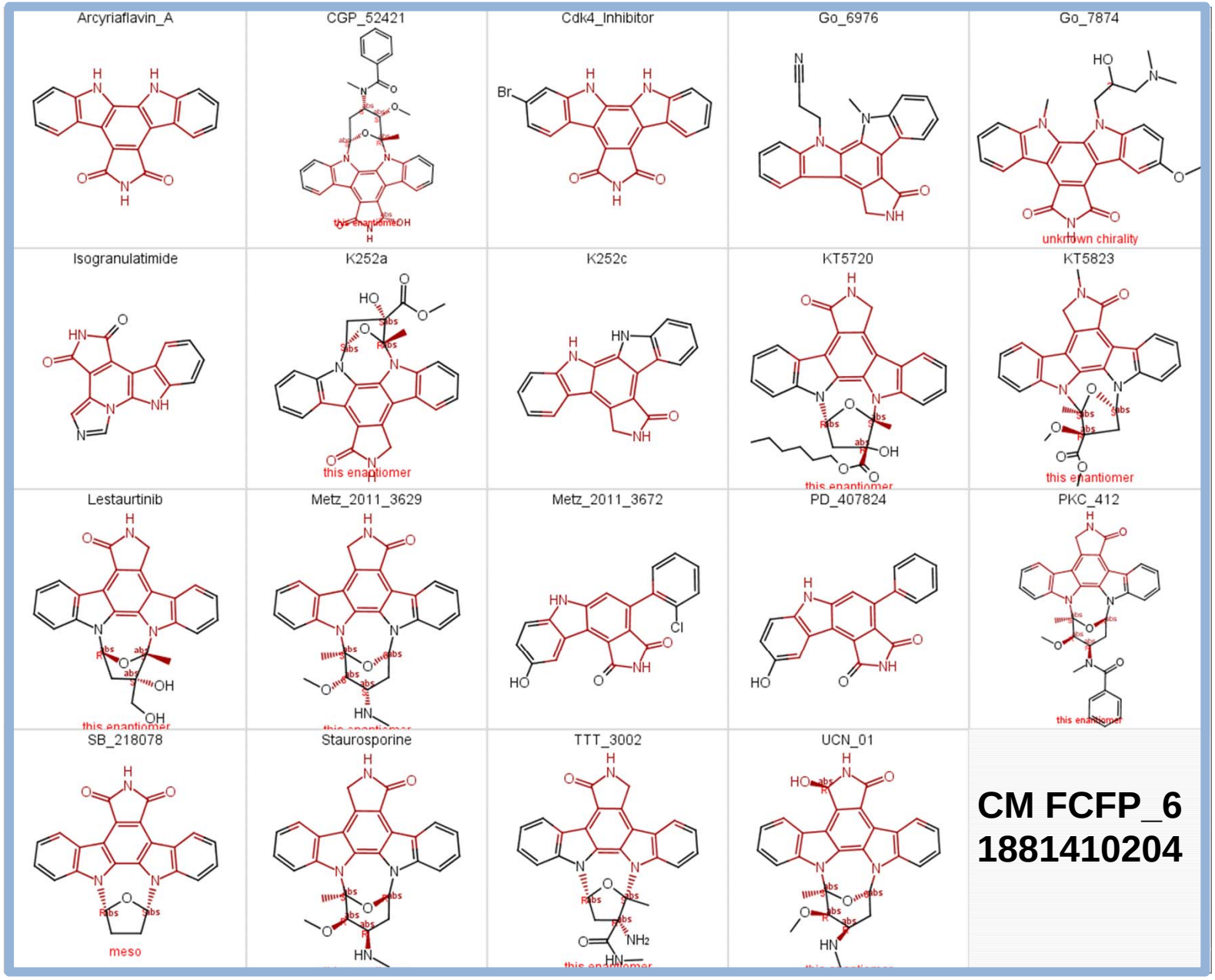


Figure S16. Examples of compounds retrieved using the substructures showing the highest correlation to affinity. The boxes colors correspond to the chemical classes.

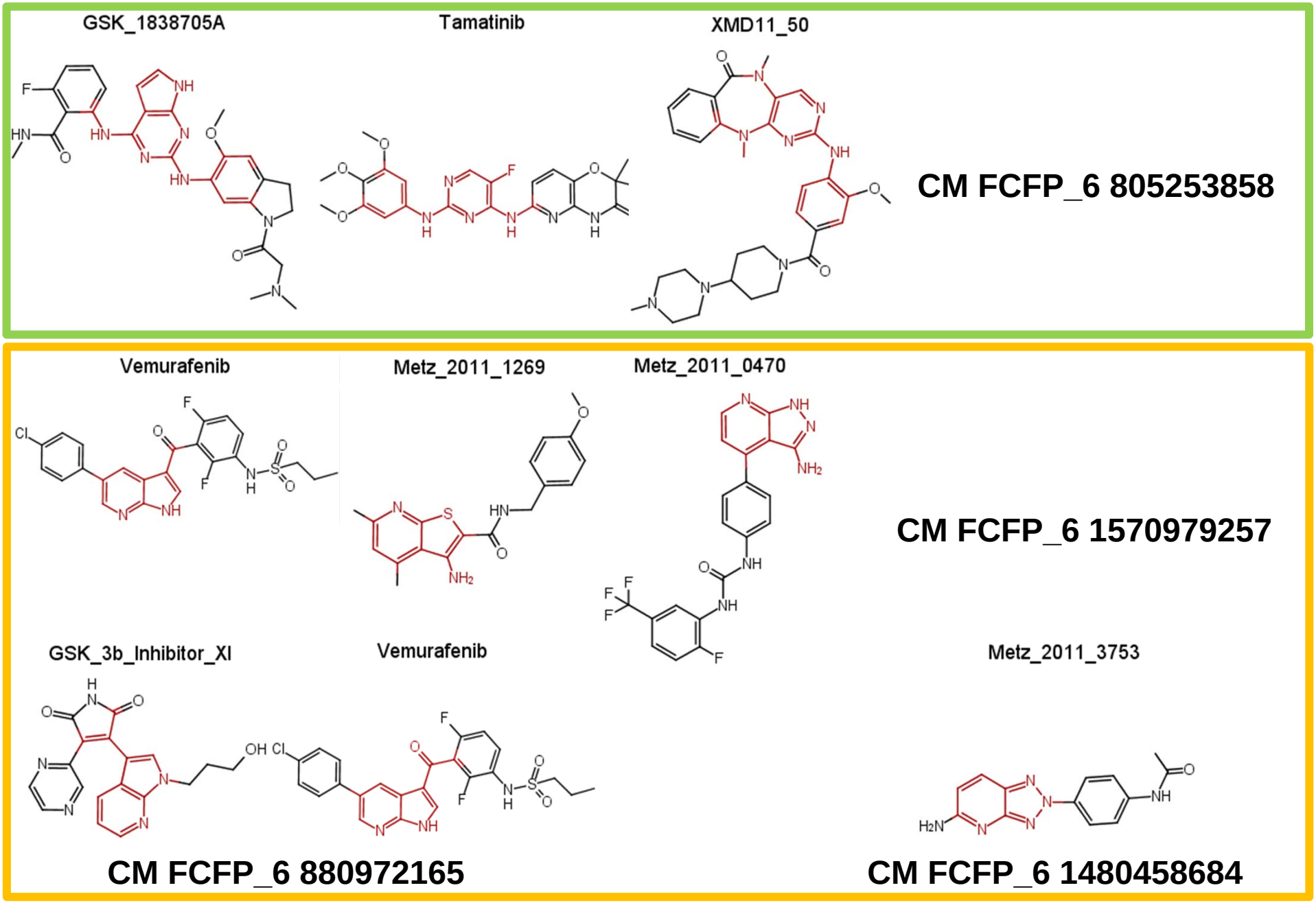


Figure S16. Examples of compounds retrieved using the substructures showing the highest correlation to affinity. The boxes colors correspond to the chemical classes.