nature chemical biology

BRIEF COMMUNICATION

PUBLISHED ONLINE: 20 FEBRUARY 2011 | DOI: 10.1038/NCHEMBIO.530

Navigating the kinome

James T Metz, Eric F Johnson, Niru B Soni, Philip J Merta, Lemma Kifle & Philip J Hajduk*

Although it is increasingly being recognized that drug-target interaction networks can be powerful tools for the interrogation of systems biology and the rational design of multitargeted drugs, there is no generalized, statistically validated approach to harmonizing sequence-dependent and pharmacology-dependent networks. Here we demonstrate the creation of a comprehensive kinome interaction network based not only on sequence comparisons but also on multiple pharmacology parameters derived from activity profiling data. The framework described for statistical interpretation of these network connections also enables rigorous investigation of chemotype-specific interaction networks, which is critical for multitargeted drug design.

Rationally designing drugs that modulate the activity of multiple protein targets (targeted polypharmacology) is an established and growing trend in drug discovery programs across the pharmaceutical industry[1–3]. The use of drug-target networks to inform and accelerate these efforts holds great promise[4–8], but their utility in truly prospective drug design has been limited. We present a critical statistical analysis of kinomics screening data across 172 different protein kinases, establishing rigorous criteria for understanding both the information content and the reliability of the derived pharmacology relationships. We then implement these criteria to not only interpret global pharmacology relationships between kinases but also to identify specific chemotypes that have the potential to bias either toward or away from polypharmacology between any two protein kinases. The ability to rationally navigate and analyze both the biological and chemical data should dramatically improve our ability to identify drug leads with a high likelihood of targeting disease-related kinases while avoiding kinases associated with clinical liabilities.

There has been an explosion in the recent literature of techniques and analyses for deducing relationships between protein targets using the similarity (either chemical or biochemical) of the compounds to which they bind[9,10] rather than their underlying sequence similarity. It has been recognized that our ability to rationally leverage pharmacology data across large numbers of clinically relevant proteins has the potential to dramatically increase discovery productivity by enhancing both the efficacy and safety of new drug candidates[1,2]. Targeted polypharmacology also holds significant potential in the identification of treatments for highly complex diseases that are resistant to the reductionist approaches (for instance, modulating a single, isolated molecular target) that dominate most of the pharmaceutical industry. However, optimizing multiple activities, minimizing off-target liabilities and trying to balance drug-like properties is an exceedingly difficult task, and robust new tools and approaches for leveraging network pharmacology in lead selection and optimization are required.

The multitargeted design of drugs that inhibit two or more protein kinases has been the object of intense pharmaceutical research over the last decade[3]. The high homology between protein kinases (especially within the ATP-binding site), makes targeted polypharmacology against multiple kinases highly achievable, but this carries the associated risk of serious selectivity issues that can translate into toxic side effects. This is typically addressed with empirical screening of large numbers of compounds across large numbers of kinases to determine the biochemical profile of lead candidates and has thereby resulted in rich datasets of ligand activity that can be probed to understand kinase pharmacology. Using proprietary databases, a number of "kinome networks" have been described that relate kinases via the similarity in their ligand-binding profiles[11–13] rather than via their sequence similarity. Such networks can highlight those kinases that are most likely to be simultaneously inhibited by a common ligand, which can aid in opportunistic discovery efforts or in the design of tailored selectivity panels for off-target liabilities. Unfortunately, no criteria have been described for understanding the information content in such relationships (which relates to their reliability) or for determining how to create subnetworks of desired pharmacological patterns (for example, targeting kinases A and B but not C and D) in the context of (or even in spite of) their overall relationships.

We have collected a dataset of more than 150,000 kinase inhibitory values, comprising more than 3,800 compounds tested

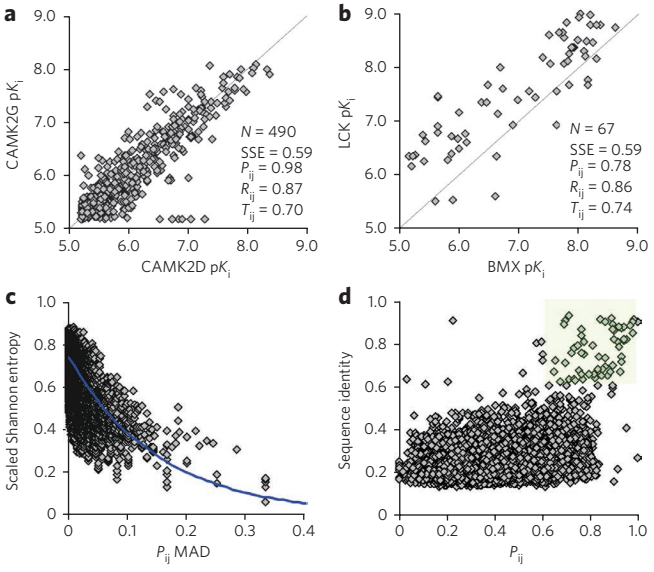

Figure 1 | Statistical analysis of kinome pharmacology data. (**a**,**b**) Basic pharmacology parameters derived from activity data (in pK_i units) are shown for (**a**) CAMK2D and CAMK2G and (**b**) BMX and LCK. (**c**) Median absolute deviations in P_{ij} values upon introduction of simulated experimental noise are shown as a function of the scaled Shannon entropy (SSE), in which it can be observed that SSE values below 0.4 lead to unacceptably large median errors (<MAD> ≥ 0.1) in P_{ij}. (**d**) Relevance of the derived pharmacology values was determined by comparison with sequence identities, as shown for P_{ij}, for which it can be observed that 93% of kinases with greater than 60% sequence identity also have P_{ij} values greater than 0.6.

Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois, USA. *e-mail: philip.hajduk@abbott.com

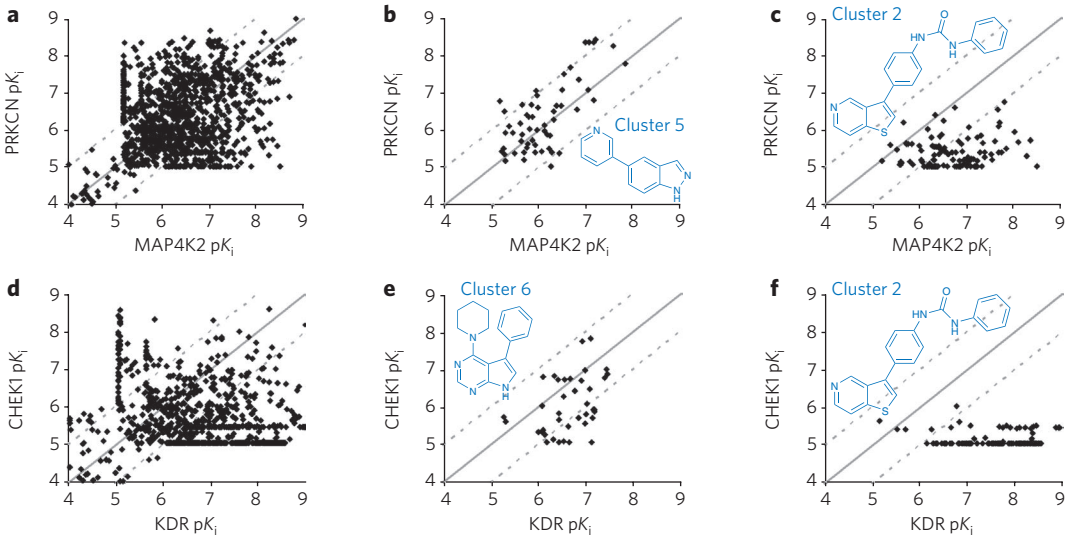

Figure 2 | Chemotype analyses of kinase pharmacology data. (**a**,**b**) PRKCN and MAP4K2 show a general pharmacological relationship ($P_{ij} \geq 0.6$) (**a**), which can be maintained with certain subseries, as illustrated by Chemotype 5 (**b**). (**c**) However, the overall relatedness can be abolished with other subseries, as illustrated for Chemotype 2 (**c**, where all pharmacology parameters fall below critical thresholds). (**d**–**f**) Similarly, although no global pharmacology relationship exists between KDR and CHEK1 (**d**), certain subseries can drive a pharmacology connection (**e**, as illustrated by Chemotype 6) whereas others can achieve specificity (**f**, as illustrated by Chemotype 2).

against 172 different protein kinases (see **Supplementary Results**, **Supplementary Table 1**, **Supplementary Figs. 1** and **6**). General analyses of trends in the data indicate that certain kinases can be potently inhibited by a large number of compounds, whereas others are apparently exquisitely selective (see **Supplementary Fig. 1**). In addition, the ability of a compound to potently inhibit multiple kinases slightly increases with the number of hydrogen bond donors and acceptors but does not correlate with size or hydrophobicity (see **Supplementary Fig. 2**). Examples of pK_I values (the negative base-10 log of the K_I values) for two pairs of highly related kinases are shown in **Figure 1a**,**b**, along with the derived pharmacology values P_{ij} (the Hopkins pharmacology interaction strength)[5], R_{ij} (the Pearson correlation coefficient) and T_{ij} (the activity profile Tanimoto)[13]. Questions that must be satisfactorily addressed before engaging in prospective drug discovery using these data relate to data quality, robustness and relevance. In other words, when a prediction is made about the pharmacological relationship between any two kinases, how reliable, how meaningful and how stable (with respect to new data) is the connection? The problems of data incompleteness and the significant changes in interaction networks that can occur as new data become available are well documented and represent a serious limitation to productive use of pharmacology networks in drug discovery[4,14]. To address these questions, we use the information content (as measured by the scaled Shannon information entropy, SSE)[15,16] in the pK_I values between the two kinases to assess the reliability and robustness of the network. The level of information available between any two kinase pairs will depend not only on the number of data points but also on how the data are spread across the potency range. As an example, the two kinase pairs in **Figure 1a**,**b** have the same level of information (SSE = 0.59), despite having nearly an order of magnitude difference in the number of data points (481 versus 67). To calibrate the level of information required to make a robust pharmacology connection between kinases, we assessed the influence of random experimental noise on the resulting pharmacology parameters. When the scaled Shannon entropy drops below about 0.4, the associated errors in the Hopkins Pharmacology Interaction Strength parameter, P_{ij}[5], become (on average) larger than 0.1 unit (see **Fig. 1c**). An alternative view of the relationship between the information content and the reliability of the derived pharmacology parameters is the change in the parameter values when only fractions of the data are used. When the SSE value is greater than 0.4, even discarding half of the compounds relating two kinases results, on average, in small (<0.05) changes in the pharmacology parameters. As a result, we propose that kinases can be confidently classified as 'related' or 'unrelated' only when the information content extant in the underlying activity data exceeds this minimum threshold (SSE ≥ 0.4). Kinase 'relatedness' is, of course, a smoothly varying function, but values corresponding to exceptionally strong correlations can be derived from a comparison with sequence identities. As shown in **Figure 1d**, >90% of kinases that share sequence identities of >60% (which has been proposed as the level of identity correlated to structure-activity relationship similarity)[12] also have P_{ij} values ≥ 0.6. Such values of P_{ij} are also rare, composing only the top 10% of the P_{ij} distribution. Note that in the original description, P_{ij} values of as low as 0.1 were used to propose pharmacology relationships across a broad range of target families[5]. Such a value is not useful for kinome pharmacology, as most values exceed this number (see **Supplementary Table 2**). Similar analyses with R_{ij} and T_{ij} result in lower limits of 0.45 and 0.55, respectively, to establish strong pharmacology relationships between kinases.

Given statistical measures for both reliability (SSE ≥ 0.4) and relevance ($P_{ij} \geq 0.6$; $R_{ij} \geq 0.45$; $T_{ij} \geq 0.55$), we can now construct a kinase interaction map based on both sequence and ligand-binding activity that gives the fullest possible picture of kinase relationships and is highly resistant to new data. The structure of this map is given in the **Supplementary Information (Supplementary Fig. 3)**. In contrast to the sequence-only network (shown in **Supplementary Fig. 3a**), in which families of kinases exist as independent network 'islands,' the network based on both sequence and pharmacology similarity is highly condensed and inter-related. In fact, the average number of connections for any kinase increases from eight on the basis of sequence identity alone to seventeen using both sequence and pharmacology similarity. The nearest-neighbor network for KDR is shown in **Supplementary Figure 3c**, illustrating the 39 kinases that show either sequence or pharmacology similarity. It is also important to note that the calculated pharmacology measures change in a very systematic and expected manner as various

thresholds are modified (for example, potency windows or potency limits, as illustrated in **Supplementary Fig. 4**), suggesting that the resulting networks are quite robust and not strongly dependent on these variables.

Although such maps can be highly useful for understanding trends in polypharmacology, they reflect only the overall likelihood that a compound active against one kinase will also be active against a connected kinase. Not all compounds between connected kinases will be active against both kinases, and kinases without a robust connection can still have many active compounds in common. In order to impact medicinal chemistry design decisions, what is required is an understanding of specific chemotypes or functional groups that either create or destroy a specific connection, as defined by the therapeutic endpoint. This requires analyzing subsets of the data as defined by chemical structure with the goal of achieving a specific sub-network of connections (and lack of connections)[3]. We have clustered the 3,800 compounds used in this analysis, resulting in more than 600 clusters that can be represented by a common substructure (see **Supplementary Table 1**). Given these clusters, the pharmacology relationships between any two kinases can be evaluated on a chemotype-by-chemotype basis. An example of this process is given in **Figure 2a–c**, where the connection between two kinases that are globally related can either be maintained or reduced by pursuing specific chemotypes, thus enabling multitargeted optimization. Likewise, as shown in **Figure 2e–f**, specific chemotypes can also drive a pharmacology connection between two kinases that are not globally related. Although subset analyses invariably involve analyzing smaller and smaller sets of data, it is important to note that the same level of statistical rigor can be applied to the subset analysis as was applied to the global analysis by using the principles of information content (Pipeline Pilot XML script for calculating kinase interaction parameters from the provided data has been provided; see **Supplementary Methods**, **Supplementary Fig. 5** and **Supplementary Dataset 1**).

In summary, we have completed a robust statistical analysis of kinome profiling data that establishes quantitative criteria for understanding both the reliability of a proposed pharmacology connection and its relevance to protein polypharmacology. As a result, multiple pharmacology interaction parameters can be simultaneously interrogated and can be robustly combined with sequence information to construct comprehensive pharmacology networks that enable kinome-wide analyses of potential drug polypharmacology. Although this work has focused exclusively on kinase activity data, the principles can be extended to proteome-wide compound profiling data. As demonstrated here, the ability to confidently explore these networks using individual chemotypes is a critical next step in systematically leveraging such information to rationally design multitargeted drugs while minimizing toxic liabilities mediated by unwanted kinase activity.

Received 17 September 2010; accepted 20 January 2011; published online 20 February 2011

References

1. Roth, B.L., Sheffler, D.J. & Kroeze, W.K. *Nat. Rev. Drug Discov.* **3**, 353–359 (2004).
2. Frantz, S. *Nature* **437**, 942–943 (2005).
3. Morphy, R. *J. Med. Chem.* **53**, 1413–1437 (2010).
4. Mestres, J., Gregori-Puigjane, E., Valverde, S. & Sole, R.V. *Mol. Biosyst.* **5**, 1051–1057 (2009).
5. Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S. & Hopkins, A.L. *Nat. Biotechnol.* **24**, 805–815 (2006).
6. Keiser, M.J. *et al. Nat. Biotechnol.* **25**, 197–206 (2007).
7. Keiser, M.J. *et al. Nature* **462**, 175–181 (2009).
8. Oprea, T.I., Tropsha, A., Faulon, J.L. & Rintoul, M.D. *Nat. Chem. Biol.* **3**, 447–450 (2007).
9. Vogt, I. & Mestres, J. *Mol. Inf.* **29**, 10–14 (2010).
10. Metz, J.T. & Hajduk, P.J. *Curr. Opin. Chem. Biol.* **14**, 498–504 (2010).
11. Vieth, M. *et al. Biochim. Biophys. Acta* **1697**, 243–257 (2004).
12. Vieth, M., Sutherland, J.J., Robertson, D.H. & Campbell, R.M. *Drug Discov. Today* **10**, 839–846 (2005).
13. Bamborough, P., Drewry, D., Harper, G., Smith, G.K. & Schneider, K. *J. Med. Chem.* **51**, 7898–7914 (2008).
14. Mestres, J., Gregori-Puigjane, E., Valverde, S. & Sole, R.V. *Nat. Biotechnol.* **26**, 983–984 (2008).
15. Godden, J.W. & Bajorath, J. *J. Mol. Graph. Model.* **18**, 73–76 (2000).
16. Godden, J.W. & Bajorath, J. *QSAR Comb. Sci.* **22**, 487–497 (2003).

Author contributions

J.T.M. performed all of the statistical analyses and created the network visualizations; E.F.J. designed the initial kinome screening panel and supervised the enzymology; N.B.S., P.J.M. and L.K. performed all kinase assays; P.J.H. supervised the research and wrote the manuscript.

Competing financial interests

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturechemicalbiology/.

Additional information

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to P.J.H.