BIOINFORMATICS DISCOVERY NOTE Vol. 28 no. 3 2012, pages 301–305
doi:10.1093/bioinformatics/btr676

Systems biology

Advance Access publication December 6, 2011

A simple statistical test to infer the causality of target/phenotype correlation from small molecule phenotypic screens

Xin Wei[1,*], Ann F. Hoffman[2], Shannon M. Hamilton[2], Qing Xiang[2], Yang He[2], W. Venus So[1], Sung-Sau So[3] and David Mark[2]

[1]Pharmaceutical Research and Early Development Informatics/Statistics, [2]Discovery Technologies, [3]Medicinal Chemistry, F. Hoffmann-La Roche, Nutley, NJ 07110, USA

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Cell-based phenotypic screens using small molecule inhibitors is an important technology for early drug discovery if the relationship between the disease-related cellular phenotype and inhibitors' biological targets can be determined. However, chemical inhibitors are rightfully believed to be less specific than perturbation by biological agents, such as antibody and small inference RNA. Therefore, it is often a challenge in small molecule phenotypic screening to infer the causality between a particular cellular phenotype and the inactivation of the responsible protein due to the off-target effect of the inhibitors.

Results: In this article, we present a Roche in-house effort of screening 746 structurally diverse compounds for their cytotoxicity in HeLa cells measured by high content imaging technology. These compounds were also systematically profiled for the targeted and off-target binding affinity to a panel of 25 pre-selected protein kinases in a cell-free system. In an effort to search for the kinases whose activities are crucial for cell survival, we found that the simple association method such as the chi-square test yields a large number of false positives because the observed cytotoxic phenotype is likely to be the result of promiscuous action of less specific inhibitors instead of true consequence of inactivation of single relevant target. We demonstrated that a stratified categorical data analysis technique such as the Cochran–Mantel–Haenszel test is an effective approach to extract the meaningful biological connection from the spurious correlation resulted from confounding covariates. This study indicates that, empowered by appropriate statistical adjustment, small molecule inhibitor perturbation remains a powerful tool to pin down the relevant biomarker for drug safety and efficacy research.

Contact: xin.wei@roche.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on September 11, 2011; revised on November 29, 2011; accepted on December 1, 2011

1 INTRODUCTION

In the past, drug discovery heavily relied upon cell-based phenotypic screening in which the drug candidates that induce the desired cellular phenotype are discovered first and the molecular mechanism of action (MMOA) is identified later (Hart, 2005). The completion of the human genome project and the emergence of 'omic' type technologies such as RNA microarrays greatly increase the expectation of disease 'target' oriented drug discovery. However, the productivity of the pharmaceutical industry keeps declining despite all these technological innovations. As a result, we have recently seen phenotypic screening regain the momentum in delivering promising drug candidates. The challenge, however, lies on the elucidation of molecular mechanism by which the candidate compounds elicit the cellular phenotype of interest, which is always important for understanding disease and medicine and serve as the basis for the further pre-clinical development (Swinney and Anthony, 2011).

In recent years, pharmaceutical industries have keen interest in developing therapeutic small molecules targeting protein kinases, a gene family implicated in cancer and inflammation. However, it is proven difficult to determine the crucial targets responsible for pathological phenotype in a protein kinase-focused phenotypic screen due to the off-target behaviors of small molecules. Many kinase inhibitors demonstrate broad cross-reactivity to the family of protein kinases which share a conserved ATP binding pocket prone to competitive interaction (Correll *et al.*, 2006). In this study, we described the profiling of 746 structurally diverse compounds for their *in vitro* kinase inhibiting properties and cytotoxicity, a cellular phenotype regarded as a typical surrogate for early safety and efficacy assessment. We devised a three-step statistical framework to facilitate the identification of protein kinases whose inactivation is responsible for compound-induced cytotoxicity. General association methods that correlate blockage of kinases with cytotoxicity such as the chi-square test was first performed to generate a potential target list. Then a logistic regression model is conducted to confirm the suspected covariate that confounds with the inhibitors' targeted effect, therefore, may interfere with causal inference. Finally, a stratified categorical analysis such as the Cochran–Mantel–Haenszel (CMH) test is conducted across the entire level of confounding covariates to identify the causal target whose significance is independent of an inhibitor's cross-reactivity. We are also able to use literature knowledge and experimental data to validate the resulting targets unveiled by the CMH test. This study demonstrates the necessity for a rigorous statistical correction in order to make accurate inference from data sources influenced by latent confounding

*To whom correspondence should be addressed.

factors, such as a phenotypic screen querying small molecule compounds.

2 EXPERIMENTS AND DATA ANALYSIS

See Supplementary Materials for detailed description of experiments. The Chi-square test, fisher exact test, Cochran-Mantel-Haenszel test and the logistic regression are conducted with SAS 9.1.3 (SAS institute, 2002).

3 METHODS AND RESULTS

3.1 Selection of kinases and compounds

Due to the prohibitive cost of exhaustively profiling all the kinases in human genome, we decided to build a small protein kinase panel that represents the Human Kinome by maximizing sequence diversity. Diversity is measured by pairwise distance measurement of the kinase catalytic domain sequences. Pairwise distance matrix (p-distance) was generated using MEGA 3 (Kumar et al. 2004). The kinase with the highest number of neighbors (defined as p-distance < 0.6) was selected. This kinase and its neighbors were then removed from the distance matrix, which was then scanned again for the next kinase with the highest number of neighbors. The above steps were repeated until all kinases were removed from the matrix. The top 30 kinases and their closest neighbors were considered for prioritization by other factors (e.g. availability of enzyme/peptide substrate for setting up the kinase assay and structural diversity of the ATP-binding site). Twenty kinases were selected in this iterative process. Five additional ones were added subsequently from kinase therapeutic projects.

746 compounds were selected for in vitro kinase binding and cytotoxicity profiling based on in-house availability, structural diversity and their known kinase binding properties. The pIC50 values of this compounds library for the 25 protein kinases were merged with the maximum cytotoxicity index within the dose range for each compound and visualized in a heat map shown in supplementary figure 1. We also showed that the 25-kinase panel well represents the diversity of the complete human kinome as the relative level of inhibitor cross-reactivity measured from this small panel is in good concordance with that measured by the more complete Ambit kinase panel (Supplementary Fig. S2). The cytotoxicity index is defined from multiple cytotoxicity parameters by the automated algorithm provided by vendor of high-content screening technology (Hoffman and Garippa, 2007).

3.2 Identification of the association between the inactivation of individual kinase and cytotoxicity by two-way contingency table

We summarize the evolution of our analysis strategy in a flowchart as shown in Supplementary Figure S3. The activity data of *in vitro* kinase inhibition is often truncated for the less potent compounds that typically do not have numerical IC50 values. In order to make use of this censored information, we propose to apply categorical data analysis methodologies to this dataset instead of linear correlation of continuous variables. Specifically, the kinase inhibitors are coded as 'inhibitory' or 'non-inhibitory' and 'toxic' or 'non-toxic' according to the pre-defined cutoff value for their IC50 in kinase inhibition assay and cytotoxicity index, respectively. Here we first chose IC50 = 1 μM as threshold for a potent inhibition, which is primarily based on the distribution of binding affinity and the fact that maximum cytotoxicity index of compounds are primarily induced at 10 μM, the highest dose of titration series. We use cytotoxicity index = 50% as the cutoff that sets apart the toxic and non-toxic compounds because cells above this value demonstrate typical cytotoxic phenotype with high confidence. Given this type of dichotomization, the status of kinase inhibition can be formulated in a series of two-way contingency tables with cytotoxicity status as follows:

Kinase k	Non-toxic	Toxic	Row
Non-inhibitory	n_{00k}	n_{01k}	$n_{0.k}$
Inhibitory	n_{10k}	n_{11k}	$n_{1.k}$
Column	$n_{.0k}$	$n_{.1k}$	$n_{..k}$

It is well known that the test statistics follow the chi-square distribution under null hypothesis that an inhibitor of kinase k does not have elevated risk to induce cell death compared with non-inhibitors.

$$\chi^2_k = \sum_i \sum_j \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}} \ (i = 0, 1, j = 0, 1)$$

where $\hat{\mu}_{ijk} = \frac{n_{i.k} \times n_{'jk}}{n_{..k}}$ is the expected frequency under the null-hypothesis (Agresti, 2002). If the observed proportion of toxic inhibitors significantly exceeds the expected value under null hypothesis then the it can be concluded that inactivation of kinase k is strongly associated with cytotoxic phenotype therefore the function of kinase $k$is likely to be crucial to the homeostasis of the cell. Table 1 lists 14 protein kinases with significant P-value from the chi-square test (Column 2) that could be considered as candidate for house-keeping protein kinases in HeLa cells.

3.3 Inhibitor promiscuity is a confounding variable that influences the inference on kinase/cytotoxicity association

We use logistic regression model to show that the promiscuity of inhibitors is another significant contributing factor to cytotoxicity. The rationale, result and interpretation of our logistic model are provided in Supplementary Table S1). Moreover, the boxplot in Figure 1b shows that the inhibitors of each kinase have very different median promiscuity measured by the number of targets they bind to. Specifically, Figure 1c showed the heat map of binding profiles for several toxic potent inhibitors of c-Src and KDR, both of which are identified as crucial kinases for cell homeostasis by the chi-square test. Potent c-Src inhibitors also block the activity of KDR and PDGFR with lesser but still strong affinity while KDR inhibitors show significant cross-reactivity to c-Src, PDGFR and ABL. The influence of inhibitor cross-reactivity as possible confounding factor on the association study between kinase and cytotoxicity is demonstrated in Figure 1a. Thus, the observed linkage between the blockage of a specific kinase and cytotoxicity could possibly be caused by the cross-reactivity of inhibitors instead of indispensability of individual kinase target.

3.4 Use CMH test to account for bias introduced by inhibitor promiscuity

As shown in Figure 1a, the observed cytotoxicity could either be induced by the loss of activity of single important kinase or blockage of multiple pathways by non-specific inhibitors or the combination of both. To extract the causality of cytotoxicity by inactivation of individual target from confounding covariate(s) (cross-reactivity), we introduce the CMH test. The essence of this simple statistical test is to perform a stratified contingency table analysis for the relationship between response and predictor variables (cytotoxicity and kinase inhibition) while controlling binding promiscuity. Specifically, a master 2 $\times$ 2 table for the whole compound library against one kinase is split up into a series of partial tables consisting of inhibitors with different levels of promiscuity designated as level p, as shown in Supplementary Table S2.

Table 1. *P*-value from chi-square and CMH test for 14 protein kinases

Kinases	Chi-square test (*P*-value)	Ratio	CMH test (*P*-value)
Aurora A[a]	<0.0001	28/62	0.0022
c-Src	<0.0001	43/144	0.2249
KDR	<0.0001	39/145	0.6726
P56LCK	<0.0001	33/109	0.1557
FGFR	<0.0001	25/75	0.5785
ABL	<0.0001	31/113	0.8761
PDGFR	<0.0001	36/149	0.3621
MST2	<0.0001	18/46	0.2241
CHK2	<0.0001	12/30	0.3447
Par-1b[a]	<0.0001	7/11	0.0474
EGFR	0.0036	8/29	0.9889
EphrinR	0.0037	6/19	0.7967
ROCK-II	0.0090	6/21	0.6415
c-RAF	0.0306	17/98	0.3873

Ratio: the number of toxic inhibitors / total number of inhibitors. [a]Protein kinases remain significant after CMH correction.

Here we classify the kinase inhibitors into three categories according to the number of kinases a compound binds to: specific (binds to 1–3 kinases), less specific (binds to 4–8 kinases) and promiscuous (binds to >8 kinases). Supplementary Table S3 shows that the percentage of toxic Src inhibitors increases nearly 2-fold from less promiscuous to highly promiscuous strata, indicating that this increase is largely the result of increased inhibitor promiscuity. Conversely, Aurora A kinase has similar percentage of toxic inhibitors in two strata with different promiscuity, suggesting that the high ratio of toxicity is likely to be the true reflection of kinase function and independent of inhibitor cross-reactivity. Formally, a CMH test statistic can be computed according to the following equation:

$$\mathrm{CMH} = \frac{[\sum_p (n_{00p} - \hat{\mu}_{00p})]^2}{\sum_p var(n_{00p})} (p = 1, 2, 3),$$

$$\text{Where } \hat{\mu}_{00p} = \frac{n_{.0p} \times n_{0.p}}{n_{..p}},\ var(n_{00p}) = \frac{n_{0.p} \times n_{1.p} \times n_{.0p} \times n_{.1p}}{n_{..p}^2 \times (n_{..p} - 1)}$$

It can be shown that this statistic follows a large sample chi-square distribution with one degree of freedom under the null hypothesis that there is no association between cytotoxicity and blockage of kinase activity after controlling for target promiscuity. If the association remains strong after combining the information from all distinct strata of covariates, a unified statistical significance can then be calculated to indicate that the induction of cell cytotoxicity is independent of inhibitor promiscuity and likely to be the consequence for the inactivation of the target kinase (Agresti, 2002). Applying the CMH test to the list of kinases essentially eliminates the majority of targets identified as important by the chi-square test. Only Aurora A kinase and Par-1b (MARK2) still demonstrates a significant CMH *P*-value (table 1, column 4), suggesting that the kinase/phenotype association is independent of confounding factor and therefore these two kinases is essential for the cell homeostasis. A straightforward way to validate the utility of CMH test in this setting is to test the cytotoxicity of highly selective inhibitors whose biological effects can be more reliably attributed to the inactivation of their target kinases. For c-Src and KDR, the top two kinases that are significant in the chi-square test but invalidated by CMH test, we identified four specific inhibitors against them out of the whole library that do not bind to any other kinases with high affinity. These compounds' cytotoxicity index is fairly low as shown in Table 2, essentially confirming that the association of c-Src and KDR inactivation with cell cytotoxicity is primarily caused by the intrinsic promiscuity of their inhibitors instead of the indispensability of kinase function. On the other hand, the most selective inhibitors for Aurora-A kinase and Par-1b, which are the only two targets

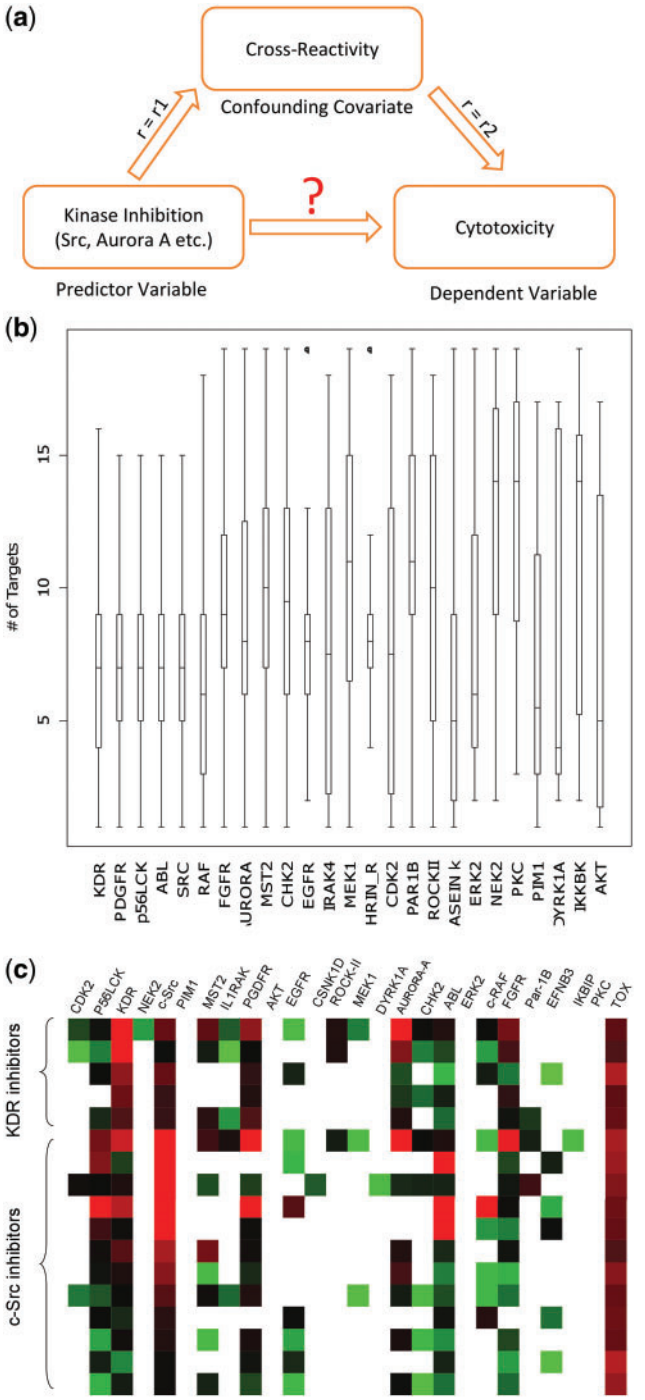

Fig. 1. (**a**) The relationship between kinase inhibition, toxicity and inhibitor cross-reactivity and their corresponding statistical meanings. $r1$ and $r2$ are the correlation coefficients of association. (**b**) The cross-reactivity of inhibitors for various kinases (line in the box indicates the median of cross-reactivity for inhibitors of a particular kinase). (**c**) Toxic inhibitors of KDR and c-Src show high degree of cross-reactivity.

vindicated by the CMH test, are much more toxic as measured by their borderline toxicity just slightly below the cutoff. The Aurora A kinase inhibitor only hits one target without any detectible off-target activity within our 25 kinase panel. The most potent and selective Par-1b (IC50 $< 1\ \mu$M) we identified

Table 2. Specific inactivation of Aurora-A and Par-1B kinases induce higher degree of toxicity than that of c-Src and KDR.

Kinase	CMH test	Compound	No. of targets[a]	Toxicity index
c-Src	0.2249	RO-XXXX	1	35.0
KDR	0.6726	RO-XXXX	1	9.91
		RO-XXXX	1	32.0
		RO-XXXX	1	18.8
Aurora-A	0.0022	RO-XXXX	1	46.51
Par-1b	0.0474	RO-XXXX	1	49.10

[a]Only count binding with high affinity (IC50 $< 1\,\mu$M).

so far, however, also binds seven other kinases with low affinity ($1\,\mu\text{M} <$ IC50 $< 20\,\mu$M). Based on the low risk of weak inhibitors on cell survival revealed by logistic regression model, it is likely that the cytotoxicity is primarily elicited by the loss of Par-1b activity. In addition, among the 34 potent Aurora A kinase inhibitors that do not make the toxicity threshold, nearly half of them (16/34) demonstrate border-line toxicity index between 40% and 50%. For the only four potent Par-1b inhibitors whose toxicity index is <50%, three of them show border-line toxicity.

We also perform simulation study to demonstrate that CMH test reduces the false discovery rate in a wide range of situations with various levels of $r1$ and $r2$, the correlation coefficient shown in Figure 1a. Specifically, the low $r1$ simulates the likes of Aurora A whose inhibitors with toxicity distribute across the different promiscuity strata more evenly than Src and KDR. $r2$, as the association coefficient of confounding and dependent variables, simulates the various mechanisms by which inhibitor off-target effect leads to cell death. A low $r2$ would be the result of the additive effect of multiple isolated signaling pathways blocked by one compound. The concomitant inactivation of synergistic targets by non-specific inhibitors, on the other hand, results in more pronounced cytotoxicity, reflected by a higher correlation coefficient ($r2$). In either case, these results demonstrate that CMH test is capable of reducing false negatives resulted from confounding factor while not losing significant amount of statistical power when the confounding covariates are less of concern (Supplementary Fig. S4). It shows that the CMH test is a powerful method to distinguish the causal biological relationship from unrelated confounding factors.

4 DISCUSSION

Besides gene knock-out models and small interference RNA technology, gene function can be studied by chemical genetics using small molecule intervention in the format of phenotypic screening (Spring, 2005). Nevertheless, chemical perturbation studies are vulnerable to the bias caused by the cross-reactivity of compound. This is particularly true for kinase inhibitors because the targets of inhibitors share a highly conserved substrate binding pocket. In this study, we show that this issue can be at least partially tackled by following a three-step statistical correction: association identification $\rightarrow$ confounding factor confirmation $\rightarrow$ confounding factor adjustment in order to infer causality of cytotoxicity by the inhibition of house-keeping kinases (Supplementary Fig. S3). Stratifying compound library based on the overall inhibitor cross-reactivity is an effective strategy to reduce the type I errors for target identification arising from grossly promiscuous action of non-specific inhibitors. Yet, it does not entirely solve the causality issue for a relatively specific inhibitor, which binds to both the real target and a small number of non-relevant kinases. Currently, this adjustment invalidates the majority of kinases that have been listed to be significant for cell homeostasis by regular association methods such as the chi-square test. Consistent with this result, our survey of biological literature also indicates that many kinases on the list indeed should not be essential to the basic homeostasis of cells. For instance, c-Src knock-out mice exhibit largely normal phenotype except in osteoblast most likely owing to the compensatory roles played by Lyn and Yes, the two close homologs of c-Src (Klinghoffer *et al.*, 1999; Soriano *et al.*, 1991). On the other hand, KDR gene knockout in mice is embryonic lethal due to the defect in early hematopoiesis and endothelial development (Shalaby *et al.*, 1997). However, it is in doubt whether these angiogenesis-related functions would be crucial for the survival of HeLa cells, a human cervical carcinoma cell line. More revealingly, all the highly specific potent inhibitors against c-Src and KDR identified in this compound library only show mild toxic phenotype measured by cytotoxicity index. Conversely, the specific potent inhibitors for Aurora A and Par-1b kinase, both of which remain significant after the CMH adjustment, lead to much higher degrees of toxicity. Besides the strong experimental evidences, Aurora A kinase is well known for its essential role in mitosis and the kinase null mice die at very early embryonic stage due to the compromised spindle assembly (Lu *et al.*, 2008). Danusertib, a potent inhibitor for Aurora A, is currently in Phase I clinical trials for the treatment of various advanced solid tumors (Steeghs *et al.*, 2009). The experimental evidence for the role of Par-1b is less clear because the most potent and selective Par-1B inhibitor (IC50 $< 1\,\mu$M) we identified in this study also inhibits seven other kinases with low affinity. The combinatorial knock-out of Par-1b and Par-1a, its closely homologous gene, is embryonic lethal, raising the possibility of inhibitor's cross-reactivity to Par-1a that is not included in our kinase panel (Lennerz *et al.*, 2010).

This approach has limitations. It foremost demands the extensive off-target profile be generated for all compounds before it can be statistically adjusted, which is not always available. However, our analysis offers compelling evidence that a snapshot of cross-reactivity index from a restricted panel of targets can serve as a good surrogate for a more comprehensive screen, which might have prohibitive cost. Meanwhile, the public database such as PubChem may also serve as an important resource for information on the interaction profiles of interested chemicals to various biological entities (Wang *et al*, 2010). Secondarily, our approach requires large numbers of compounds with high diversity to be screened in order to gain sufficient statistical power. The pool of potent inhibitors with toxicity also needs to be large enough so that they can be stratified in order for adjustment of covariates. Another downside of protein family focused screen is that the interaction of inhibitors with random targets outside the targeted protein family is overlooked. This limitation can only be circumvented by using more compounds in that the general pattern drawn from a large library is less likely to be influenced by the opportunistic off-target effects outside the gene family that have been screened.

ACKNOWLEDGEMENTS

We appreciate the useful discussion with Drs Seng-lai Tan and Jim Rosinski.

Funding: The study has been funded by F. Hoffmann-La Roche, Ltd.

Conflict of Interest: none declared.

REFERENCES

Agresti,A. (2002) *Categorical Data Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Correll,P.H. *et al.* (2006) Molecular regulation of receptor tyrosine kinases in hematopoietic malignancies. *Gene* **374**, 26–38.

Hart,C.P. (2005) Finding the target after screening the phenotype. *Drug Discov. Today*, **10**, 513–519.

Hoffman,A.F. and Garippa,R.J. (2007) A pharmaceutical company user's perspective on the potential of high content screening in drug discovery. *Methods Mol. Biol.*, **356**, 19–31.

Klinghoffer,R.A. *et al.* (1999) Src family kinases are required for integrin but not PDGFR signal transduction. *EMBO J.*, **18**, 2459–2471.

Kumar,S. *et al.* (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinformatics*, **5**, 150–163.

Lennerz,J.K. *et al.* (2010) Loss of Par-1a/MARK3/C-TAK1 kinase leads to reduced adiposity, resistance to hepatic steatosis, and defective gluconeogenesis. *Mol. Cell. Biol.*, **30**, 5043–5056.

Lu,L. *et al.* (2008) Aurora A is essential for early embryonic development and tumor suppression. *J. Biol. Chem.*, **283**, 31785–31790.

SAS Institute Inc (2002) *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, NC.

Shalaby,F. *et al.* (1997) A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis. *Cell*, **89**, 981–990.

Soriano,P *et al.* (1991) Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell*, **64**, 693–702.

Spring,D.R. (2005) Chemical genetics to chemical genomics: small molecules offer big insights. *Chem. Soc. Rev.*, **34**, 472–482.

Steeghs,N. *et al.* (2009) Phase I pharmacokinetic and pharmacodynamic study of the aurora kinase inhibitor danusertib in patients with advanced or metastatic solid tumors. *J. Clin. Oncol.*, **27**, 5094–5101.

Swinney,D.C. and Anthony,J. (2011) How were new medicines discovered? *Nat. Rev. Drug Discov.*, **10**, 507–519.

Wang,Y. *et al*. (2010) An overview of the PubChem BioAssay resource. *Nucleic Acids Res*., **38**, D255–D266.