Trends in Cell & Molecular Biology
Vol. 10, 2015

Original Communication

Comparative analysis of the human and zebrafish kinomes: focus on the development of kinase inhibitors

Nathan Wlodarchak[1], Rehan Tariq[1] and Rob Striker[1,2,*]

[1]Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison, Microbial Sciences Building, 1550 Linden Drive, Madison, WI 53706, USA;
[2]W. S. Middleton Memorial Veteran's Administration Hospital, 2500 Overlook Terrace, Madison, WI 53705, USA.

ABSTRACT

Targeting kinases with semi-selective kinase inhibitors is one of the most successful drug development strategies of the 21st century. Zebrafish have become an increasingly useful model for pharmaceutical development. Water-soluble compounds can be screened for zebrafish phenotypes in a high throughput format against a living vertebrate, and cell-signaling events can be imaged in transparent living fish. Despite zebrafish being a more relevant model than more distantly related systems such as the well-annotated kinome of yeast and drosophila, there is no comparative analysis of the human and zebrafish kinome. Furthermore most approved kinase inhibitors, often called 'DFG in' ATP competitive inhibitors, act on conserved active site residues in the kinase. Since the active site residues can be identified by examining the primary sequence, primary sequence identity can be a rough guide as to whether a particular inhibitor will have activity against another kinase. There is a need to evaluate the utility of zebrafish as a drug development model for active site inhibitors of kinases. Here we offer a systematic comparison of the catalytic domains of classical human kinases with the catalytic domains of all annotated zebrafish kinases. We found a high degree of identity between the catalytic domains of most human kinases and their zebrafish homologs, and we ranked 504 human kinase catalytic domains by order of similarity. We found only 23 human kinases with no easily recognizable homologous zebrafish catalytic domain. On the other hand we found 78 zebrafish kinase catalytic domains with no close human counterpart. These 'additional kinase active sites' could represent potential mediators of zebrafish toxicity that may not be relevant to human kinase inhibitors. We used two clinically approved human kinase inhibitors, one targeting a highly homologous target and one targeting a lesser homologous target, and we compared the known human kinase target structures with modeled zebrafish target structures. As expected, the homologous target had high structural identity, but even the less homologous target had high structural identity in residues contacted by the inhibitor. Overall this analysis should help guide researchers interested in studying human kinases and their inhibitors in more tractable systems.

KEYWORDS: kinome, drug development, kinase inhibitors, human kinases, zebrafish

INTRODUCTION

Protein phosphorylation is the primary signal transduction mechanism in eukaryotic cells. This is accomplished through the opposing action of kinases and phosphatases, and the proper regulation of these enzymes is critical for cell homeostasis [1-3]. Inhibiting kinases has become

*Corresponding author: rtstriker@wisc.edu

one of the most reliable pharmacologic strategies in the last decade with several new kinase targeting drugs currently in use [4]. Initially kinase inhibitors were intensively studied and approved as cancer drugs, but now the scope of their potential use has widened to immune regulation, anti-infectives, and cardiovascular diseases [4, 5]. Tofacitinib is currently approved for rheumatoid arthritis, and other Jak family inhibitors are entering clinical trials for rheumatoid arthritis and psoriasis [6]. Inhibitors for p38 MAP kinases are under investigation for treatment of atherosclerosis and chronic obstructive pulmonary disease (COPD) [7]. These and other successes have reinforced the search for small molecule kinase inhibitors which could selectively inhibit kinases with minimal side effects. There are 518 human kinases and inhibition of any of these kinases can potentially elicit a potent therapeutic effect because of the intrinsic cascading mechanism each kinase possesses [5]. Initially there was concern and skepticism that small molecules could confer selectivity for a single kinase. It is now clear that sufficient selectivity is achievable [5, 8]. Therapeutics inhibiting more than one kinase are sometimes desirable, and potentially more than 518 different effects could be achieved by inhibiting different combinations of kinases. Currently 25 kinase inhibitors are FDA approved, most of which are oral [4].

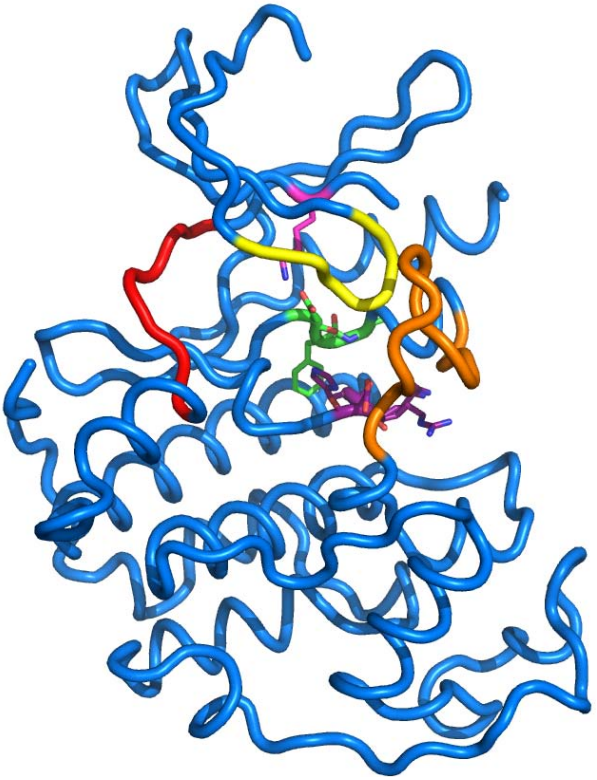

Figure 1. CDK2 (4EK3) as an example of a classical Hank's kinase with critical Hanks motifs highlighted. Hinge region in red, P-loop in yellow, activation loop in orange, invariant lysine in magenta sticks, HRD in purple sticks, and DFG in green sticks.

While the so-called 'atypical' kinases lack canonical primary sequence motifs, there is a common evolutionary and structural basis for protein kinases [9]. Protein kinases have a bi-lobed structure with catalysis occurring in the cleft between lobes [10]. The cleft contains key conserved features critical to phosphorylation: an ATP binding pocket, substrate binding residues, and a divalent cation—typically Mg^{2+} [10, 11]. Throughout the kinome there are well-recognized conserved motifs that collectively line the cleft, coordinate the metal cofactor and transfer the phosphate from ATP to the substrate [10, 11]. Steven Hanks and others identified several domains which were conserved among almost all kinases and were essential for their kinase activity [10, 12]. These domains became subsequently known as Hanks domains (Figure 1). As the ancestral kinases diversified, many different subfamilies evolved specializing in, for example, tyrosine phosphorylation or serine/threonine phosphorylation [13]. The overall human kinome has been divided into subfamilies based on structural and functional differences, and this classifying information is annotated in online databases [1, 14]. Prominent families include serine/threonine kinases such as AGC and CAMK as well as exclusive tyrosine kinases (TKs) and tyrosine kinase like (TKL) kinases which phosphorylate tyrosine as well as serine and threonine [15]. How these kinase families evolved has been examined in model organisms across a vast swath of divergent organisms from Saccharomyces to mice to humans, but more recent model organisms such as zebrafish have not been analyzed in detail [16].

Despite extensive diversification, the Hanks domains have remained largely unchanged, particularly inside the catalytic domain of the N-lobe between the glycine rich 'P-loop' and the DFG motif [10].

In fact, even small changes in these motifs lead to the categorization of kinases as likely pseudokinases [17]. Some classified pseudokinases retain catalytic activity, but non-canonical motifs at least suggest functions other than phospho-transfer [17, 18]. Although the Hanks domains in the cleft are highly conserved, some variation near the motifs may be present among species, particularly if their substrates diverged greatly. To better utilize zebrafish as a high throughput model for kinase inhibitor drug development, it is necessary to evaluate how similar the clefts are for both the human and zebrafish kinome.

Zebrafish (*Danio rerio*) is now a well-established model organism for developmental biologists [19]. It has many advantages including low cost, tissue transparency, fecundity, and short generation time [20]. These advantages have also made zebrafish an attractive model for drug development [21]. The zebrafish genome is sequenced and it has 71% homology to the human genome [22]. Genome duplication events in the teleost lineage complicate genetic comparisons between zebrafish and humans, since some genes were duplicated several times and others were not at all [23]. A common tool to probe the specific genetics of zebrafish has been the use of antisense morpholino oligomers to knock down specific genes; however, the specificity of this has been challenged recently [24]. The use of small molecule inhibitors may be a complementary approach if key inhibitor binding sites are well conserved between humans and zebrafish. The zebrafish kinome has yet to be extensively described, but if the kinome parallels the genome in homology, then the zebrafish may potentially serve as a functional model organism for kinase inhibitor development. While zebrafish are already a model for testing drug toxicity [21], expanding zebrafish utility as a kinase-specific model for toxicity and efficacy will require more detailed knowledge about the zebrafish kinome.

Understanding the relatedness of human and zebrafish kinases will improve the reliability of zebrafish as a model organism for studying kinase inhibitors. Testing drugs for human kinases with no clear zebrafish homolog might yield false negatives, and conversely, zebrafish kinases which have clear homologs to human kinases but with several specialized isoforms not found in humans may also provide data that is ultimately not helpful in the drug development process. Knowledge of these relationships would streamline drug development by identifying candidate kinase targets which are viable models in zebrafish and allowing more high-throughput testing to be done earlier in the drug development process. Here we did not examine the overall kinase homology *per se* but specifically analyzed how well the active sites match since most ATP competitive kinase inhibitors rely on critical residues in the Hank's domains.

METHODS

The human kinase domain sequences were retrieved from the KinBase database [1]. There are 531 non-redundant human kinases annotated in Kinbase, of which 25 are classified as atypical. Manning and colleagues have identified 15 additional atypical kinases and 13 pseudokinase domains, of which the former were excluded, and the latter were included in our analysis [1]. Approximately 350 zebrafish kinases were obtained from ensemble [25] and approximately 550 were obtained from zfin [26] using pfam 00069 for a total of ~900 zebrafish sequences from both sources. These kinases were visually examined, and redundant sequences were removed to arrive at a final list of 692. The shortest canonical human kinases from each family were used as a model to identify zebrafish kinase catalytic domains by performing sequence alignment and trimming zebrafish sequences where they align to the human catalytic domains. MultAlign [27] was used to align the kinase domains of the organisms up within their respective subfamilies in order to identify the homologous kinase domains in zebrafish kinases, and analysis was restricted to the kinase domain. Residues prior to the start of the P-loop, usually indicated by the presence of a nearby glycine rich region, and residues subsequent to the DFG motif were truncated, based on the alignment to the canonical kinase domain. This region will be subsequently referred to as the 'catalytic domain' (CD) (Figure 1).

Clustal Omega [28] was used to perform a protein BLAST analysis and generate a phylogenetic tree that depicted the relationship between the human and zebrafish kinase CDs. The data were organized

by pairing kinases grouped between species and unique CDs that have no orthologs between humans and zebrafish. The maximum matched percent identity of each zebrafish CD and of each human CD was extracted using Excel (Microsoft) and condensed into a master table (Table S1). A frequency distribution of the maximum percent identities was calculated using Prism (GraphPad software). A cutoff of 40% identity was chosen based on the frequency distribution of all maximum CD percent identities (Figure S1) as well as based on visual examination of the dendogram. Below 40% identity there was not clear pairing of human and zebrafish kinase active sites.

Crystal structures of FDA approved kinase inhibitors were downloaded from the PDB (3OG7 and 3LXK) [29, 30] and visualized in Pymol (Schrodinger). Predicted zebrafish kinase models were generated using I-Tasser [31]. Structural figures were also rendered in Pymol.

RESULTS AND DISCUSSION

The zebrafish and human kinase catalytic domains share high identity

The protein BLAST analysis revealed generally high percent identities between the zebrafish and human kinase catalytic domains (CD). Four hundred and eighty one human CDs matched to 614 zebrafish CDs with a percent identity greater than 40 (Figure 2). The extra zebrafish kinases are primarily due to additional isoforms for highly homologous forms of certain kinases—chiefly the PIM kinases and Aurora kinases (Table S1). Twenty four unique or divergent human CDs were identified (Figure 2, Table 1) and 78 unique zebrafish CDs were identified (Figure 2, Table 2). The average percent identity of CDs between both kinomes including unique kinases is 69% with a median identity of 75%, suggesting the identity of the kinome corresponds well to the total genomic identity [22].

Unique zebrafish CDs consist of many alternate isoforms of homologous kinases

There are 78 unique zebrafish CDs, defined as those with no human CD of greater than 40% identity (Table 2). Several of these CDs match as isoforms of zebrafish CDs which have percent identities greater than 40% (notably the PIM family and Aurora family); however, some of these CDs are isoforms of similar CDs with no strong human protein homolog (such as the TSSK family). This ‘soform variation’ may be a function

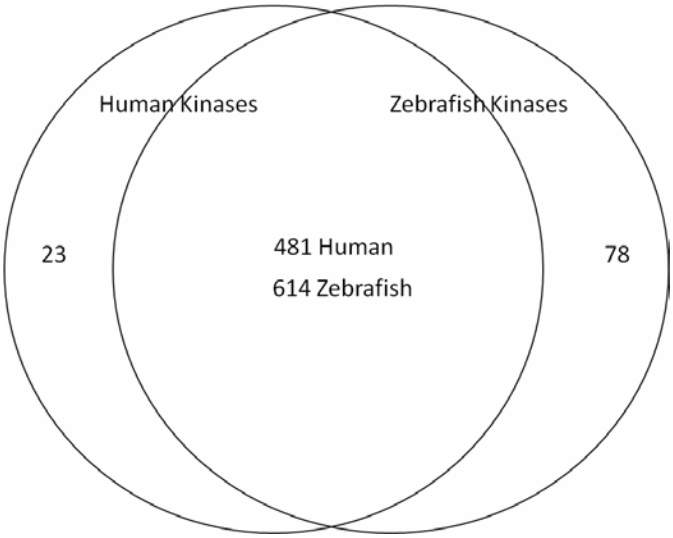

Figure 2. Unique vs. shared kinase catalytic domains. Of the total 692 non-redundant zebrafish sequences analyzed, 614 had greater than 40% identity to 481 human kinase catalytic domain sequences. Only 23 human sequences were less than 40% identity to any zebrafish sequence and 78 zebrafish sequences were less than 40% identity to any human sequence.

Table 1. Divergent or unique human kinase catalytic domains. Human catalytic domain sequences which do not match any zebrafish sequence with greater than 40% identity are shown below. Several of the unique sequences are members of distinct kinase families.

Unique human kinases	Percent identity	Zebrafish kinase match
COT	39	CHUK
FAM20A	24	MINK1
FAM20B	18	OXSR1B
FAM20C	26	PIM1
FJB1	27	PINK1
GCN2-P	31	MARK1
MLKL	37	TAOK1A
NME1A	21	SIK2A
NME1B	21	JAK1
NME3	20	PRKD3
NME4	21	JAK1
NME5	21	GUCY2F
NME6	21	POMK
NME7a	20	ZGC66101
PAN3	26	DCKL1B
SgK223	20	PRKD3
SgK269	25	PRKD3
SgK396	33	MAP3K12
SGK495	30	PTK6B
Slob	31	CDK13
TXNDC3	23	MYLKA
TXNDC6	19	CAMKVL
ULK4	39	STK36

of the diverged habitat of human and zebrafish (land vs. water) or some may be remnants of genome duplication events. These kinases are categorized as isoforms in this study based solely on maximum percent identity; therefore further functional characterization of these kinases is needed if they are to be investigated as drug target models.

Only 5% of human CDs are unique

The unique human CDs form a group of only 24 (Table 1). Interestingly, several of these CDs are related to each other, such as the FAM family, the NME family, and the SgK family. The FAM20 kinases are involved in phosphorylation of secreted substrates [32]. Given that humans and zebrafish live in fundamentally different environments, it is not surprising that pathways which interact with the environment (i.e. secretion) would be divergent. The NME family of kinases has been implicated in the function of tumor metastasis; however, the function of these kinases has been scarcely studied outside of humans [33]. Some pseudokinases such as sgk495 are represented, but unlike the FAM20 and NME kinase families, the SGK subfamily has several members with homologs between human and zebrafish (Table S1). The SGKs are generally involved in stress responses and cellular channel activation [34]. It is possible that some of these channel signaling processes have evolved to be highly divergent in humans, but several of these kinases are still under-explored.

There are seven typical human kinases that have no ortholog in mice [35]. Interestingly, while for many of these 7 there is likely not a strict zebrafish homolog, there is a CD with high identity to the human kinase ‘missing’ from the mouse kinome. For example neither the mouse kinome nor the zebrafish kinome has a CDK3 ortholog, but the CD of human CDK3 is 80% identical to that of zebrafish CDK2 (Table S1). Another example is that for human CK1alpha2: no ortholog is present in mice, but a high identity (98%) CD is present in zebrafish. The weakest zebrafish match among the 7 human CDs not present in mice is for human PKSH2 at only 74%. The differences in mouse and zebrafish kinomes suggest that organism selection for a particular drug-target model is critically dependent on information gleaned by the type of analysis presented in this study.

Pseudokinases are generally less conserved between zebrafish and humans

Forty-eight human kinases have been classified as ‘pseudokinases’ on the basis of variation in one of the three Hank’s domains (VAIK, HRD, or DFG) necessary for catalysis, and 28 have conserved homologues in yeast, worms, mice, and humans [17].

Table 2. Catalytic domains of divergent or unique zebrafish kinases. Zebrafish catalytic domain sequences which do not match any human sequence with greater than 40% identity are shown below. Most (72/78) unmatched kinases are mutants or splice variants of otherwise well matched partners. Six of the catalytic domains are likely truly unique.

Unique zebrafish kinases	Percent identity	Human kinase match
SICH211-138G9.2	39	AurB
SICH211-138G9.3	39	AurB
SICH211-272B8.7	38	AurB
SIDKEY-22I16.10	39	AurB
SIDKEY-34D22.3	38	AurB
SIDKEY-80C24.1	39	AurB
SIDKEY-80C24.5	39	AurB
SIDKEYP-67E1.2	39	AurB
SIDKEY-183G16.2	38	HUNK
SIDKEY-206F10.6	39	MARK1
SIDKEY-83M22.16	38	MARK1
PAK7	26	MLK1
SIDKEY-155D18.1	37	MSK2A
SIDKEYP-67E1.4	39	MSK2A
PXK	24	NuAK1
SIDKEY-197D18.1	39	NuAK1
SIDKEY-197D18.2	39	NuAK1
G6PD	19	PAK4
SICH211-249H16.8	38	PASK
SICH211-57M13.8	38	PASK
SIDKEY-211E20.1	38	PASK
SIDKEY-37M8.14	39	PASK
SICH211-119D14.4	38	PIM1
SICH211-168M18.3	33	PIM1
SICH211-196G2.7	34	PIM1
SICH211-196G2.8	38	PIM1
SICH211-198E20.11	38	PIM1
SIDKEY-236A14.6	38	PIM1
SIDKEYP-100A1.5	37	PIM1

Table 2 continued..

Unique zebrafish kinases	Percent identity	Human kinase match
SICH211-176G13.9	39	PIM2
SICH211-215M21.11	39	PIM2
SICH211-237D5.4	39	PIM2
SICH211-57M13.10	36	PIM2
SIDKEYP-110E4.3	39	PIM2
SICH211-147H1.4	39	PIM3
SICH211-196N4.5	32	PIM3
SICH211-196N4.7	35	PIM3
SICH211-196N4.8	32	PIM3
SICH211-214C11.2	35	PIM3
SICH211-214C11.4	32	PIM3
SICH211-214C11.7	31	PIM3
SICH211-214C11.8	32	PIM3
SICH211-57M13.3	39	PIM3
SICH211-57M13.5	39	PIM3
SICH211-57M13.6	39	PIM3
SICH211-57M13.7	39	PIM3
SIDKEY-222B8.5	34	PIM3
SIDKEYP-67E1.6	39	PIM3
SIDKEY-211E20.8	39	PKD1
SIDKEY-211E20.9	39	PKD1
SICH211-57M13.9	36	SIK
SICH211-10J20.2	38	TSSK1
SICH211-10J20.4	37	TSSK1
SICH211-215M21.13	37	TSSK1
SIDKEY-206F10.7	39	TSSK1
SIDKEY-217M5.9	38	TSSK1
SIDKEY-197D18.3	39	TSSK2
SICH211-160D20.5	38	TSSK3
SICH211-202N12.2	37	TSSK3
SICH211-207C6.4	37	TSSK3
SICH211-238G23.1	38	TSSK3
SICH211-57M13.1	36	TSSK3

Table 2 continued..

Unique zebrafish kinases	Percent identity	Human kinase match
SIDKEY-248F6.3	39	TSSK3
SIZFOS-754C12.2	38	TSSK3
SICH211-12E13.5	38	TSSK4
SICH211-12E13.6	38	TSSK4
PRKD3	27	TTBK2
PKN1B	23	TXNDC3

Although these 48 have atypical variation in key residues based on primary sequence, several have been shown to be catalytically active [17, 18]. For example, WNK1 does not have the classical VIAK motif in β-strand 3, but instead has a lysine residue (K233) that enters the active site from β-strand 2 and confers catalytic activity [36]. This suggests that structural and functional analysis is key to pseudokinase classification [18]. The CDs of twenty-two of these pseudokinases are also conserved between zebrafish and humans (Table 3). Of these 22, differences between zebrafish and human pseudokinases owing to the loss of critical Hanks domains (DFG, HRD, and VIAK) were only seen in three: Trib1, Trib2, and Sgk494A (Table 3). The zebrafish Trib kinases deviate more from the classical Hank's domains than do their human equivalents; however, Sgk494A has retained all of these domains. This suggests that there may be potential differences in structure and function between some of these pseudokinases which evolved after the human/zebrafish divergence. In addition to unique human pseudokinases, several groups which are represented in zebrafish have additional isoforms in humans (e.g. ANPa/b and RSKL1/2) further suggesting additional roles in humans that may be accomplished by fewer isoforms in zebrafish or not at all.

Structural identity is conserved in active sites of varying sequence identity

Although considerable attention is given to sequence identity in the literature, structural identity may be a more accurate predictor of functional conservation [8]. Fortunately, several of the FDA approved kinase inhibitors have been crystallized with their targets, and structures are available [4, 8]. The identity of the targets of these FDA approved drugs is variable, but analysis of these structures can provide information as to whether the most critical of contacts is preserved.

The FDA approved kinase inhibitor vemurafinib is a BRAF kinase inhibitor [4, 29]. Human BRAF kinase has 99% identity to zebrafish BRAF kinase in the catalytic active site (Table S1). Vemurafinib is coordinated in the BRAF active site by hydrogen bonding to the hinge region backbone and the activation loop backbone as well as through several stacking interactions and hydrophobic interactions (Figure 3). The BRAF kinase CDs are essentially identical between human and zebrafish; therefore the drug is almost certain to bind to zebrafish BRAF as it would in human BRAF.

Tofacitinib was FDA approved for arthritis and it inhibits human Jak3 kinase [4, 30]. Jak3 has one canonical CD and one pseudokinase domain. The canonical kinase domain has a 68% identity to zebrafish Jak2A (Table S1). Despite the lower sequence identity in the CDs, they still maintain high structural similarity (Figure 4). Tofacitinib binds to the active site of Jak3's canonical CD similar to most ATP competitors by hydrogen bonding to the hinge region backbone (Figure 4). It is also stabilized by hydrogen bonding to the P-loop and through hydrophobic interactions at the base and back of the active site. Although the CD residues appeared unchanged from the primary sequence, the higher overall divergence may have created conformational changes not evident from primary sequence alone. The I-Tasser-predicted structure shows the residues pointing toward the catalytic site in the CD and overall conformation are mostly unchanged (Figure 4). There is a small kink introduced in human Jak3 by substituting an alanine for glycine at aa966 (152), and there is a cysteine to leucine conversion at aa909 (142). Neither of these changes appears to affect the coordination of the drug in the active site, suggesting that despite a larger amount of evolutionary divergence compared to BRAF, the Jak3 active sites are structurally conserved and would be accurate models for testing new drugs. The structures of many human kinases are known, and several have been co-crystallized with drugs [8]. The sequence analysis presented here, combined

Table 3. Pseudokinase domain divergence. Of the 48 known human pseudokinase catalytic domains, 22 have various degrees of identity to zebrafish proteins. Pseudokinases with no clear homologous non-redundant partner are left unmatched. Modified critical Hanks domains (DFG, HRD, or VIAK) are indicated based on Boudeau and colleagues [17] and alignment of the matching pairs.

Human pseudokinase	Modified domains	% identity	Zebrafish pseudokinase	Modified domains
ANPA	HRD	84	NPR1A	HRD
ANPB	HRD	--	--	--
CASK	DFG	99	CASKA	DFG
CCK4	DFG	--	--	--
CYGD	HRD	60	GC3	HRD
CYGF	HRD	77	GC2	HRD
EPHA10	DFG, HRD, VIAK	--	--	--
EPHB6	DFG, HRD, VIAK	--	--	--
GCN2-P	HRD, VIAK	--	--	--
HER3 (ERBB3)	HRD	75	ERBB3A	HRD
HSER	HRD	--	--	--
ILK	DFG, HRD	86	ILK	DFG, HRD
IRAK2	DFG, HRD	--	--	--
JAK1-P	HRD	70	JAK1-P	HRD
JAK2-P	HRD	83	JAK2B-P	HRD
JAK3-P	HRD	60	JAK3-P	HRD
KSR1	VIAK	--	--	--
KSR2	VIAK	88	KSR2	VIAK
MLKL	DFG, HRD	--	--	--
NRBP1	DFG, HRD, VIAK	95	NRBP1	DFG, HRD, VIAK
NRBP2	DFG, HRD, VIAK	--	--	--
PSKH2	HRD	--	--	--
RSKL1	DFG, VIAK	--	--	--
RSKL2	DFG, VIAK	--	--	--
SCYL1	DFG, HRD, VIAK	--	--	--
SCYL2	DFG, HRD, VIAK	89	SICH211-244B2.1	DFG, HRD, VIAK
SCYL3	DFG, HRD, VIAK	73	SCYL3	DFG, HRD, VIAK
SGK071	HRD	--	--	--
SGK196	DFG, HRD, VIAK	60	POMK	DFG, HRD, VIAK
SGK223	DFG	--	--	--
SGK269	DFG	--	--	--

Table 3 continued..

Human pseudokinase	Modified domains	% identity	Zebrafish pseudokinase	Modified domains
SGK307	DFG, HRD	--	--	--
SGK396	HRD	--	--	--
SGK494	DFG, HRD, VIAK	59	SGK494A	NONE
SGK495	DFG	--	--	--
SLOB	DFG, HRD, VIAK	--	--	--
STLK5 (STRADA)	DFG, HRD, VIAK	72	STRADA	DFG, HRD, VIAK
STKL6 (STRADB)	DFG, HRD	--	--	--
SURTK106	DFG	--	--	--
TBCK	DFG, HRD, VIAK	62	TBCK	DFG, HRD, VIAK
TRB1	DFG	55	TRIB1	HRD, VIAK
TRB2	DFG	58	TRIB3	DFG, HRD, VIAK
TRB3	DFG	--	--	--
TYK2-P	HRD	64	TYK2-P	HRD
ULK4	DFG, VIAK	--	--	--
VACAMKL	HRD	91	CAMKVA	HRD
VRK3	DFG, HRD, VIAK	46	VRK3	DFG, HRD, VIAK

with structural information from published and modeled drug-kinase interactions, provide potentially insightful tools for drug development.

Limitations

There are several limitations to both our analysis and the general use of zebrafish as an alternative to more costly mammalian models for kinase inhibitor drug development. Testing drugs in zebrafish requires a certain amount of water solubility. Solvents such as DMSO, used to deliver insoluble drugs have toxicity effects on their own, reducing the effectiveness of a freshwater organism model. Fortunately, most kinase inhibitors are less than 1000 Daltons in size and have some polar moieties, allowing them to have sufficient water solubility [8], and all 25 of the currently FDA approved kinase inhibitors are orally available [4].

The zebrafish genome is rapidly improving in annotation but is certainly not as comprehensively annotated as the human genome. For example, twice the zebrafish kinases from zfin were recovered as were in ensemble, but there could still be unannotated kinases that were not included in this study. Very few human kinases were not matched to a zebrafish kinase in this analysis, which could indicate the coverage is sufficient and additional zebrafish kinases would likely be used for biological processes specific to the zebrafish.

Although this work has presented divergent or unique zebrafish CDs based on a histogram of identity and an alignment, the specific utility of zebrafish for drug development for a specific human kinase depends on a number of factors beyond the cut-off presented here. The zebrafish kinome consists of several isoforms of kinases with moderate-to-high identity to their human counterpart. These isoforms are mostly uncharacterized and may be unique to zebrafish and have unique functions. Attempts at drug development targeting kinases which have several other isoforms may be clouded by off target effects. Careful scrutiny should be made in using

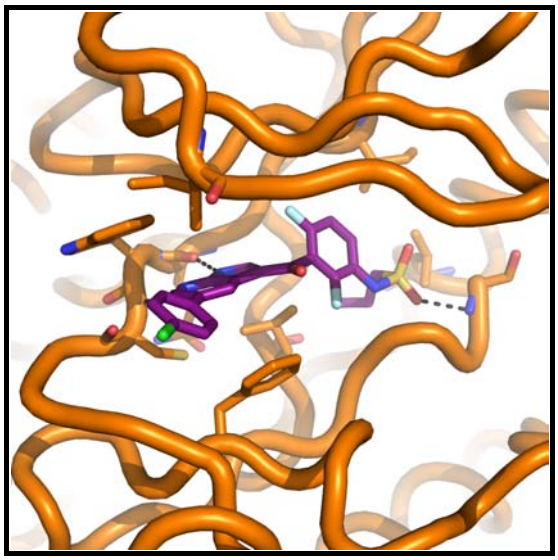

Figure 3. Vemurafenib in human BRAF kinase. The structure of human BRAF with bound Vemurafenib (3OG7) is shown in orange cartoon tube with critical coordinating residues shown in orange sticks. Vemurafenib is shown in purple sticks and hydrogen bonds to the drug are shown as black dashes. Vemurafenib is an FDA approved kinase inhibitor which selectively blocks human BRAF kinase activity and is indicated to treat some melanomas [4, 29]. The zebrafish BRAF kinase catalytic domain has 92% identity to human BRAF. Alignment of the zebrafish sequence to the human structure (3OG7) indicates that all drug coordinating residues in human BRAF are identical to zebrafish BRAF.

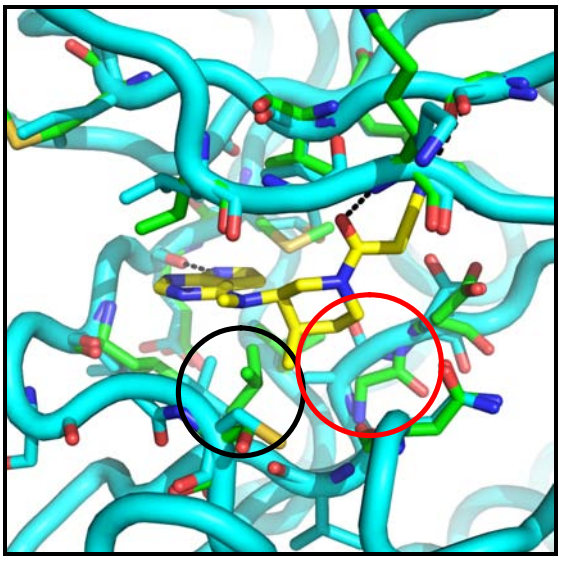

Figure 4. Tofactinib in human JAK3 kinase. The structure of human JAK3 with bound Tofactinib is shown in cyan cartoon tube with critical coordinating residues shown in cyan sticks (3LXK). Tofactinib is shown in yellow sticks. Zebrafish JAK2B residues are shown as green sticks. Tofactinib hydrogen bonds are shown as black dashes. Tofactinib is an FDA approved kinase inhibitor which selectively blocks human JAK3 kinase activity and is indicated to treat rheumatoid arthritis [4, 30]. The zebrafish JAK2A kinase catalytic domain has 68% identity to human JAK3. Threading the zebrafish sequence to the human JAK3 structure (3LXK) indicates that all but two residues in the human catalytic pocket JAK3 are identical to zebrafish JAK2A: C909-Hs is equivalent to L142-Dr (circled in black) and A966-Hs is equivalent to G152 (circled in red).

zebrafish as a model for a particular kinase for which zebrafish possesses multiple isoforms.

This analysis focused only on primary sequences. Although the structures of the zebrafish kinases are predicted to be similar to their human counterparts based on primary sequences and known similar structures, subtle nuances may still create situations where certain inhibitor interactions are disrupted or strengthened. The quaternary interactions of human kinases have been elucidated thoroughly in some cases such as cyclins and CDKs [37], but this is generally lacking in zebrafish. It is possible that species-specific accessory proteins could cause appreciable conformational changes in the structures of these kinases and alter potential binding patterns. As discussed above, the role of pseudokinases in zebrafish is not well known, and allosteric activation by or of pseudokinases is generally underexplored [17, 18]. Additional structural information will be needed in the future to address these concerns.

CONCLUSIONS

Overall, this work should facilitate decisions on when zebrafish represent a useful choice to study human kinase inhibitors. The utility clearly varies from high for inhibitors targeting highly homologous kinase CDs (BRAF) and subtly divergent CDs (Jak3), to low for more highly divergent kinases (FAM20 kinases) or if several divergent isoforms (PIM kinases) exist.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Anna Huttenlocher for editorial advice on the manuscript,

and Kelly Pittman for advice on selecting programs to create the BLAST matrix. We would also like to thank Caitlin Pepperell for early advice on the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare. Funding for N. Wlodarchak and R. Striker is from the Hartwell Foundation and the Institute for Clinical and Translational Research (ICTR). The project described was supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SUPPLEMENTARY FIGURE AND TABLE

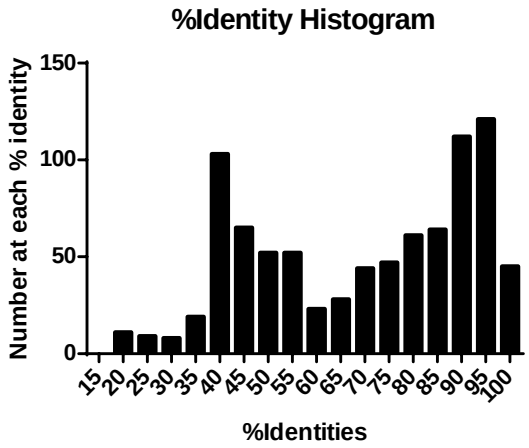

Figure S1. Frequency distribution of all kinase % identities. The percent identity of all human and zebrafish kinase catalytic domains was plotted by frequency. A significant break at 40% identity was observed, suggesting this as a cutoff to delineate unique or divergent catalytic domains.

Table S1. Highest percent identity match of all human kinases with all zebrafish kinases.

Human kinase	% identity	Zebrafish kinase
AAK1	89	AAK1A
	88	AAK1B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
ABL1	90	ABL2
ABL2	94	ABL2
ACK	62	TNK1
ACTR2	83	ACVR2AA
	82	ACVR2AB
ACTR2B	68	ACVR2B
AKT1	94	AKT1
AKT2	95	AKT2
	73	AKT2L
AKT3	99	AKT3
	99	AKT3A
ALK	45	TEC
ALK1	81	ACVRL1
ALK2	83	ACVR1L
	82	ACVRL1
ALK4	96	ACVR1B
	96	ACVR1BA
	95	ACVR1BB
ALK7	89	TGFBR1A
AMPKA1	96	PRKAA1
AMPKA2	72	PRKAA2
ANKRD3	89	RIPK4
	72	ANKK1
ANPa	84	NPR1A
ANPb	81	NPR1A
ARAF	86	ARAF
AurA	84	AURKA
AurB	79	AURKB
	40	SICH211-280A23.2
	40	SICH73-21B16.2
	40	SICH73-21B16.3
	39	SIDKEYP-67E1.2
	39	SIDKEY-80C24.1

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
AurB	39	SIDKEY-80C24.5
	39	SIDKEY-22I16.10
	39	SICH211-138G9.2
	39	SICH211-138G9.3
	38	SIDKEY-34D22.3
	38	SICH211-272B8.7
AurC	77	AURKB
	46	SIDKEY-44G23.9
	41	SICH211-122L14.3
	41	SICH211-122L14.7
	41	SICH73-132G21.6
	41	SIRP71-84D9.6
	41	SICH73-313G15.1
	41	SIRP71-84D9.4
	41	SIDKEY-40J3.1
	41	SIRP71-1L17.1
AXL	62	TYRO3
BARK1	98	ADRBK2
BARK2	98	ADRBK2
BIKE	81	BMP2K
BLK	83	BLK
BMPR1A	91	BMPR1AA
	89	BMPR1AB
BMPR1B	94	BMPR1BA
	92	BMPR1BB
BMPR2	88	BMPR2B
	71	BMPR2A
BMX	59	TEC
BRAF	99	BRAF
BRK	51	PTK6B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
BRSK1	96	SIDKEY-16P21.7
BRSK2	97	SIDKEY-16P21.7
	97	SIDKEY-16P21.7
	96	SICH211-255P10.4
BTK	65	TEC
BUB1	43	CHEK1
BUBR1	40	PHKG1B
CAMK1A	83	CAMK1A
CAMK1B	67	PNCK
CAMK1D	89	CAMK1DA
	86	CAMK1DB
	86	CAMK1DBA
	83	CAMK1A
CAMK1G	79	CAMK1GB
	74	CAMK1GA
CAMK2A	97	CAMK2A
	95	CAMK2D1
	43	SICH211-151H10.1
CAMK2B	96	CAMK2B1
	96	CAMK2B2
CAMK2D	96	CAMK2D2
CAMK2G	96	CAMK2G1
CAMK4	93	CAMK4
CaMKK1	80	CAMKK1
	80	CAMKK1A
CaMKK2	73	CAMKK1B
	72	SICH73-62L21.1
CAMLCK	77	MYLK3
CASK	99	CASKA
	99	CASKAB
	98	CASKB
CCK4	41	MUSK

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
CCRK	82	CDK20
CDC2	86	CDK1
CDC7	69	CDC7
CDK10	89	CDK10
CDK11a	98	CDK11B
CDK19	99	CDK19
CDK2	89	CDK2
CDK3	80	CDK2
CDK4	75	CDK4
	53	CDK21
CDK5	97	CDK5
CDK6	92	CDK6
CDK7	92	CDK7
CDK8	99	CDK8
CDK9	93	CDK9
CDKL1	92	CDKL1
CDKL2	59	CDKL1
CDKL3	56	CDKL1
CDKL4	80	CDKL1
CDKL5	90	CDKL5
CHED	94	CDK13
CHK1	70	CHEK1
CHK2	62	CHEK2
CK1A	87	CSNK1A1
CK1A2	90	CSNK1A1
CK1D	98	CSNK1DA
	98	CSNK1DB
CK1E	97	CSNK1E
CK1G1	93	CSNK1G1
CK1G2	96	CSNK1G2A
	93	CSNK1G2B
CK1G3	92	CSNK1G1
CK2a1	98	CK2A1
	98	ZGC86598

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
CK2a1	98	CSNK2A1
	91	CSNK2A2B
CK2a2	89	CK2A2A
	89	CSNK2A2A
CLIK1	83	STK35
	81	STK35L
	48	SICH211 63O20.7
CLIK1L	84	PDIK1L
CLK1	69	CLK4B
CLK2	88	CLK2A
	85	CLK2B
CLK3	68	CLK2A
CLK4	70	CLK4B
	68	CLK4A
COT	39	CHUK
CRIK	55	LATS1
CRK7	95	CDK12
CSK	89	CSK
CTK	60	CSK
CYGD	60	GC3
CYGF	77	GC2
	72	GUCY2F
	61	GC3
DAPK1	88	DAPK1
DARK2	82	DAPK2
	82	DAPK2A
	79	DAPK2B
DAPK3	94	DAPK3
	55	SIDKEY-240H12.4
DCLK1	91	DCKL1B
	91	DCLK1A
DCLK2	83	DCLK2A
	83	DCLK2B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
DCLK2	83	DCLK2C
DCLK3	54	DCLK3
DDR1	46	NTRK2B
DDR2	47	ABL2
DLK	92	MAP3K12
DMPK1	78	SICH211-89P3.3
DMPK2	79	CDC42BPAA
DRAK1	83	STK17A
	75	ZGC153952
	67	STK17B
DRAK2	71	STK17A
Dusty	89	DSTYK
DYRK1A	99	DYRK1AA
	99	DYRK1AB
	97	DYRK1B
DYRK1B	96	DYRK1B
DYRK2	97	DYRK2
	89	ZGC172180
DYRK3	43	SIDKEY-238N5.3
DYRK4	75	DYRK4
EGFR	84	ERBB2
EPHA1	66	EPHB4A
EPHA10	57	EPHA4B
EPHA2	78	EPHA2A
EPHA3	89	EPHA4B
	86	EK1
EPHA4	96	EPHA4A
EPHA5	89	EPHA7
EPHA6	88	EPHA4A
EPHA7	97	EPHA7
EPHA8	80	EPHA7
EPHB1	91	EPHB2B
EPHB2	97	EPHB2B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
EPHB3	87	EPHB3A
	79	EPHA2A
EPHB4	88	EPHB4A
	88	EPHB4B
EPHB6	55	EPHB3A
ERBB2	90	ERBB2
ERBB3	75	ERBB3A
	70	ERBB3B
ERBB4	95	ERBB4A
Erk1	89	MAPK3
Erk2	96	MAPK1
	92	MAPK3
Erk3	90	MAPK6
Erk4	68	MAPK4
Erk5	90	MAPK7
Erk7	77	MAPK15
FAK	93	PTK2.1
FAM20A	24	MINK1
FAM20B	18	OXSR1B
FAM20C	26	PIM1
FER	70	FES
FES	76	FES
FGFR1	91	FGFR1B
	88	FGFR1A
FGFR2	94	FGFR2
FGFR3	82	FGFR4
FGFR4	77	FGFR2
FGR	83	YRK
FJB1	27	PINK1
FLT1	70	FLT1
FLT3	54	KITB
FLT4	72	FLT4
FMS	68	CSF1RA
FRK	54	PTK6B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
Fused	84	STK36
FYN	95	FYNB
GAK	84	GAK
GCK	75	MAP4K6
GCN2	67	EIF2AK4
GCN2A	31	MARK1
GPRK4	94	GRK4
GPRK5	74	GRK5
GPRK6	91	GRK5L
GPRK7	80	GRK7A
	76	GRK7B
GSK3A	91	GSK3AA
	91	GSK3AB
GSK3B	98	GSK3B
	93	SIDKEYP-80C12.7
Haspin	65	SIDKEYP-26A9.2
HCK	82	LYN
HGK	91	TNIKA
HH498	94	TNNI3K
HIPK1	92	HIPK1A
HIPK2	98	HIPK2
HIPK3	92	HIPK3B
	94	HIPK3A
HIPK4	55	HIPK3A
HPK1	68	MAP4K5
HRI	45	EIF2AK1
HSER	47	ERBB2
HUNK	77	HUNK
	40	SICH73-381F5.2
	40	SIDKEY-93L1.4
	40	SIDKEY-93L1.6
	38	SIDKEY-183G16.2
ICK	87	MAK

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
IGF1R	89	IGF1RA
	89	IGF1RB
IKKa	75	CHUK
IKKb	79	IKBKB
IKKe	77	IKBKE
ILK	86	ILK
INSR	91	INSRA
	91	INSRB
IRAK1	63	IRAK1
IRAK2	48	IRAK1
IRAK3	52	IRAK1
IRAK4	54	IRAK4
IRE1	88	ERN1
	74	ERN2
IRE2	65	ERN1
IRR	79	INSRB
ITK	66	ITK
JAK1-P	70	JAK1-P
JAK1	77	JAK1
JAK2-P	83	JAK2B-P
JAK2	88	JAK2B
	84	JAK2A
JAK3	68	JAK2B
JAK3-P	60	JAK3-P
JNK1	95	MAPK8A
JNK2	92	MAPK9
JNK3	99	MAPK10
	96	MAPK8B
KDR	74	KDRL
	67	KDR
KHS1	93	MAP4K5
	77	MAP4K2L
	77	MAP4K2
KHS2	93	MAP4K3A

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
KHS2	92	MAP4K3B
	78	MAP4K6
KIS	71	UHMK1
KIT	71	KITB
	69	KITA
KSR1	77	KSR2
KSR2	88	KSR2
LATS1	96	LATS1
LATS2	93	LATS2
LCK	78	LCK
LIMK1	94	LIMK1A
LIMK2	87	LIMK2
LKB1	96	STK11
LMR1	46	PIM3
LMR2	51	NEK4
LMR3	40	ACVRL1
LOK	76	SLKA
LRRK1	44	LRRK2
LRRK2	72	LRRK2
LTK	45	ACVR2AB
LYN	89	LYN
LZK	86	MAP3K12
MAK	92	MAK
MAP2K1	94	MAP2K1
MAP2K2	96	MAP2K2A
	93	MAP2K2B
MAP2K3	60	POMK
MAP2K4	96	MAP2K4A
MAP2K5	89	MAP2K5
MAP2K6	91	MAP2K6
MAP2K7	88	MAP2K7
MAP3K1	93	MAP3K1
MAP3K2	94	MAP3K2
MAP3K3	95	MAP3K3

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
MAP3K4	99	MAP3K4
MAP3K5	95	MAP3K5
	91	SICH211-1I11.3
	89	MAP3K15
MAP3K6	73	SICH211-1I11.3
MAP3K7	81	MAP3K15
MAPKAPK2	88	MAPKAPK2A
	73	MAPKAPK2B
MAPKAPK3	78	MAPKAPK3
MAPKAPK5	92	MAPKAPK5
MARK1	39	SIDKEY-206F10.6
	38	SIDKEY-83M22.16
MARK2	98	MARK2A
	96	MARK2B
MARK3	97	MARK3B
	96	MARK1
	93	MARK3
	93	MARK3A
MARK4	95	SIDKEY-31M14.7
	93	MARK4A
MAST1	97	MAST1A
	96	MAST1B
MAST2	95	MAST2
MAST3	94	MAST1A
MAST4	96	MAST4
MASTL	86	MASTL
MEKK15	75	MAP3K19
MELK	80	MELK
MER	62	TYRO3
MET	89	MET
MINK	92	TNIKB
MISR2	48	CHEK1

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
MLK1	26	PAK7
	67	PIM3
MLK2	67	PIM3
MLK3	67	PIM3
MLK4	67	PIM3
MLKL	37	TAOK1A
MNK1	83	MKNK1
MNK2	84	MKNK2A
	83	MKNK2B
MOK	43	CK2A1
MOS	60	MOS
MSK1-P	81	RPS6KA5-P
MPSK1	65	STK16
MSK2	82	RPS6KA4
MSSK1	81	SRPK1A
MST1	95	STK3
MRCKA	95	SICH211-89P3.3
	93	CDC42BPAA
MRCKB	90	CDC42BPB
MSK1	89	RPS6KA4
	89	RPS6KA5
MSK2-P	72	RPS6KA4-P
	39	SIDKEYP-67E1.4
	37	SIDKEY-155D18.1
MST2	99	STK3
MST3	97	STK24B
	93	STK24A
	90	STK25A
MST4	93	MST4
	93	STK26
MUSK	80	MUSK
MYO3A	83	MYO3A
MYO3B	80	MYO3B

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
MYT1	50	PKMYT1
NDR1	98	STK38A
	96	STK38B
NDR2	95	STK38L
NEK1	89	NEK1
	56	SIDKEY-192D15.1
	52	SIDKEY-201L21.1
	51	SIDKEY-125E8.1
	48	SIDKEY-192D15.2
NEK10	80	NEK10
NEK11	57	PIM3
NEK2	82	NEK2
NEK3	42	ZGC113355
	42	NEK12
NEK4	86	NEK4
NEK5	74	NEK1
NEK6	87	NEK6
NEK7	89	NEK7
NEK8	85	NEK8
NEK9	47	NEK1
NIK	56	SICH211-253B1.4
	54	SIDKEY-260C8.4
NIM1	65	SICH211-22D5.2
	40	SIDKEY-183G16.1
NLK	100	NLK2
	92	NLK1
NME1A	21	SIK2A
NME1B	21	JAK1
NME3	20	PRKD3
NME4	21	JAK1
NME5	21	GUCY2F

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
NME6	21	POMK
NME7a	18	STK10
NRBP1	95	NRBP1
NRBP2	74	NRBP1
NRK	60	TNIKB
NuAK1	92	NUAK1A
	91	NUAK1B
	69	SICH211-117C9.1
	39	SIDKEY-197D18.2
	39	SIDKEY-197D18.1
	24	PXK
NuAK2	80	NUAK2
OBSCN-P	57	OBSCN-P
OBSCN	50	SICH211-195M20.1
OSR1	94	OXSR1A
	92	OXSR1B
p38a	96	MAPK14B
	95	MAPK14A
p38b	91	MAPK11
p38d	78	MAPK13
p38g	83	MAPK12A
	81	MAPK12B
	79	MAPK13
	69	ZGC171775
P70S6K	98	RPS6KB1B
	94	RPS6KB1A
P70S6KB	84	RPS6KB1B
PAK1	98	PAK1
PAK2	96	PAK2A
	96	PAK2B
PAK3	97	PAK1
PAK4	93	PAK4
	19	G6PD

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PAK5	91	PAK4
PAK6	89	PAK6B
PAN3	26	DCKL1B
PASK	44	SIDKEY-37M8.3
	40	SIDKEY-37M8.13
	40	SIDKEY-11F7.5
	40	SIDKEY-278K10.4
	39	SIDKEY-37M8.14
	38	SIDKEY-211E20.1
	38	SICH211-57M13.8
	38	SICH211-249H16.8
PBK	70	PBK
PCTAIRE1	93	CDK16
PCTAIRE2	96	CDK16
PCTAIRE3	90	CDK16
PDGFRA	58	PDGFRB
PDGFRB	68	PDGFRB
PDK1	88	PDPK1B
	83	PDPK1A
PEK	57	EIF2AK3
PFTAIRE1	99	CDK14
PFTAIRE2	85	CDK15
PHKG1	78	PHKG1A
	76	PHKG1B
PHKG2	75	PHKG2
PIK3R4	44	CHEK1
PIM1	76	PIM1
	54	SICH211-231M1.1
	53	SICH211-193E5.3
	53	SIDKEY-45J10.1

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM1	53	SIDKEY-45D16.8
	52	SICH211-194C3.4
	52	SIDKEY-101N15.3
	52	SIDKEY-11E6.2
	52	SIDKEY-11E6.4
	52	SIDKEY-278G4.6
	52	SICH211-244K5.5
	52	SIDKEY-10F21.3
	51	SIDKEY-278G4.4
	47	SICH211-155D24.5
	46	SICH211-113J14.2
	46	SICH211-222N4.5
	46	SICH73-272H16.1
	45	SICH73-272H16.3
	44	SIDKEYP-104H9.7
	44	SICH211-126G16.5
	44	SIDKEY-258F14.4
	44	SICH211-126G16.2
	44	SIDKEY-205D4.2
	43	SIDKEY-58B18.10
	41	SICH211-255L14.1
	41	SICH211-255L14.3
	41	SICH211-255L14.5
	41	SIDKEY-236A14.13
	40	SICH211-165A10.4

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM1	40	SIDKEY-40H17.6
	40	SICH211-165A10.5
	38	SIDKEY-236A14.6
	38	SICH211-196G2.8
	38	SICH211-119D14.4
	38	SICH211-198E20.11
	37	SIDKEYP-100A1.5
	34	SICH211-196G2.7
	33	SICH211-168M18.3
PIM2	45	SIDKEY-58B18.9
	44	SIDKEY-31M5.5
	44	SIDKEY-31M5.7
	43	SIDKEY-31M5.3
	41	SIDKEY-149I17.9
	41	SICH211-10J20.5
	40	SICH211-10J20.1
	39	SIDKEYP-110E4.3
	39	SICH211-215M21.11
	39	SICH211-176G13.9
	39	SICH211-237D5.4
	36	SICH211-57M13.10
PIM3	74	PIM3
	59	PIM2
	54	SIDKEY-278G4.9
	54	SIDKEY-193D10.6

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM3	54	SIDKEY-278G4.1
	53	SICH211-193E5.4
	53	SIDKEY-43A14.3
	53	SIDKEY-43A14.4
	53	SICH211-207C6.10
	53	SIDKEY-193D10.8
	53	SICH211-231M1
	53	SIDKEY-278G4.3
	53	SICH211-244K5.3
	53	SIDKEY-278G4.8
	53	SICH211-231M1.3
	53	SIDKEY-278G4.2
	53	SIDKEY-11E6.5
	53	SIDKEY-11E6.3
	53	SICH211-231M1.4
	53	SIDKEY-193D10.2
	53	SIDKEY-193D10.7
	53	SIDKEY-193D10.5
	53	SIDKEY-278G4.7
	53	SIDKEY-278G4.10
	53	SICH211-244K5.4
	53	SICH211-231M1.5
	52	SICH211-207C6.9
	52	SIDKEY-45D16.9
	52	SICH211-244K5.2

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM3	52	SIDKEY-193D10.11
	52	SIDKEY-193D10.3
	52	SIDKEY-265A11.5
	51	SIDKEY-278G4.5
	51	SIDKEY-11E6.1
	51	SIDKEY-236A14.11
	51	SIDKEY-236A14.10
	51	SIDKEY-236A14.5
	51	SIDKEY-236A14.5
	51	SIDKEY-236A14.11
	51	SIDKEY-1P9.2
	50	SIDKEY-43A14.5
	50	SIDKEY-236A14.4
	50	SIDKEY-236A14.4
	46	SIDKEY-58B18.3
	45	SICH73-129A22.11
	45	SICH211-214C11.9
	45	SICH211-196N4.2
	45	SICH211-196N4.1
	45	SIDKEY-179A6.3
	44	SIDKEY-179A6.1
	44	SICH211-196N4.3
	44	SICH211-214C11.6
	44	SIDKEY-179A6.2
	44	SIDKEY-43B14.9

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM3	44	SICH211-270P4.3
	44	SIDKEY-149I17.8
	44	SICH211-254P10.3
	44	SIDKEY-208K22.2
	43	SIDKEY-43B14.7
	43	SIDKEY-43B14.6
	43	SICH211-126G16.3
	43	SICH211-126G16.4
	43	SIDKEY-205O12.2
	43	SIDKEY-58B18.6
	43	SICH211-215M21.17
	43	SIDKEY-14A7.4
	43	SICH211-214C11.3
	43	SIDKEY-108D22.5
	42	SIDKEY-33C12.6
	42	SIDKEY-84O3.91
	42	SIDKEY-84O3.11
	42	SIDKEY-84O3.13
	42	SIDKEY-84O3.10
	42	SIDKEY-84O3.15
	42	SIDKEY-33C12.8
	42	SIDKEY-112G5.2
	42	SIDKEY-10B15.18
	42	SIDKEY-112G5.7
	42	SIDKEY-33C12.7

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PIM3	42	SIDKEY-84O3.12
	42	SIDKEY-84O3.14
	42	SIDKEY-112G5.5
	42	SIDKEY-19P15.5
	39	SICH211-57M13.6
	39	SICH211-57M13.3
	39	SICH211-57M13.5
	39	SICH211-57M13.7
	39	SICH211-147H1.4
	39	SIDKEYP-67E1.6
	35	SICH211-196N4.7
	35	SICH211-214C11.2
	34	SIDKEY-222B8.5
	32	SICH211-214C11.4
	32	SICH211-196N4.8
	32	SICH211-214C11.8
	32	SICH211-196N4.5
	31	SICH211-214C11.7
PINK1	67	PINK1
PKACA	95	PRKACAA
	91	PRKACAB
PKACB	94	PRKACBB
	90	PRKACBA
PKACG	80	PRKACAB
PKCA	91	PRKCA
PKCB	93	PRKCBB
	91	PRKCBA

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PKCD	87	PRKCDA
	87	PRKCDB
PKCE	97	PRKCEA
	96	PRKCEB
PKCG	77	PRKCG
PKCH	80	PRKCHB
	80	PRKCHA
PKCI	99	PRKCI
PKCT	88	PRKCQ
PKCZ	96	PRKCZ
PKD1	96	PRKD1
	39	SIDKEY-211E20.8
	39	SIDKEY-211E20.9
PKD2	73	ZGC175248
PKD3	94	PRKD1
PKG1	99	PRKG1B
	96	PRKG1A
	61	SIDKEY-121J17.5
PKG2	80	PRKG2
PKN1	86	PKN1A
PKN2	94	PKN2
	94	ZGC153916
	86	PKN1A
	75	PKN3
PKN3	70	PKN1A
PKR	53	PKZ
	47	EIF2AK2
PLK1	81	PLK1
PLK2	89	PLK2A
	81	PLK2B
PLK3	82	PLK3
PLK4	83	PLK4
PRKX	82	PRKX

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
PRKY	82	PRKX
PRP4	99	PRPF4BB
	95	PRPF4BA
PRPK	64	TP53RK
PSKH1	91	PSKH1
	66	SICH211-27E6.1
PSKH2	74	PSKH1
PYK2	72	PTK2BB
QIK	92	SIK2B
	86	SIK2A
QSK	93	ZGC66101
RAF1	95	RAF1A
	94	RAF1B
RET	88	RET
RHOK	86	GRK1A
	82	GRK1B
RIPK1	68	RIPK1L
RIPK2	68	RIPK2
RIPK3	57	PIM3
RNAseL	52	CHEK1
ROCK1	98	ROCK1
ROCK2	93	ROCK2A
	82	ROCK2B
RON	65	MET
ROR1	43	RYK
ROR2	43	NTRK2B
ROS	48	ABL2
RSK1	96	RPS6KA1
RSK1-P	87	RPS6KA1-P
RSK2-P	89	RPS6KA3A-P
RSK2	96	RPS6KA3A
	95	RPS6KA3B
RSK3	92	RPS6KA3A
RSK3-P	87	RPS6KA2-P

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
RSK4-P	86	RPS6KA6-P
RSK4	96	RPS6KA6
RSKL1	48	CHEK1
RSKL2	48	CHEK1
RYK	89	RYK
SBK	78	SICH211-183D21.3
	69	BSK146A
	68	BSK146B
	48	SICH211-171H4.3
	47	SBK3
SCYL1	48	CAMKVL
SCYL2	89	SICH211-244B2.1
SCYL3	73	SCYL3
SgK069	61	SIDKEY-8E10.3
SgK071	50	CHEK1
SGK085	76	MYLK4A
	74	MYLK4B
	67	MYLK2
SGK1	95	SGK1
	88	SICH211-195B13.1
	72	SGK2B
SgK110	50	SICH211-183D21.3
SgK196	60	POMK
SGK2	86	SGK2A
SgK223	20	PRKD3
SgK269	25	PRKD3
SGK288	64	ANKK1
SGK3	88	SGK3
SgK307	50	POMK
SgK396	33	MAP3K12
SgK493	34	AURKA
SGK494	59	SGK494A

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
SGK494	53	SGK494B
SGK495	30	PTK6B
SIK	91	SIK1
	36	SICH211-57M13.9
SKMLCK	66	MYLK4B
SLK	88	SLKA
	87	SLKB
	70	SIDKEY-81J8.6
Slob	31	CDK13
SMMLCK	81	MYLKA
	79	MYLKB
	62	SIDKEY-194J22.6
SNRK	98	SNRKA
	85	SNRKB
SPEG	57	SICH211-195M20.1
SPEG-P	67	SICH211-195M20.1
	62	SPEG
SRC	95	SRC
SRM	53	FYNB
SRPK1	91	SRPK1A
	88	SRPK1B
SRPK2	85	SRPK1A
SSTK	55	TSSK6
STK33	70	STK33
STLK3	87	OXSR1A
STLK5	72	STRADA
STLK6	47	STRADA
SURTK106	41	PIM3
SYK	80	SYK
TAK1	90	MAP3K7
TAO1	96	TAOK1A
TAO2	92	TAOK2A

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
TAO2	92	TAOK2B
TAO3	88	TAOK3A
TBCK	62	TBCK
TBK1	88	TBK1
TEC	75	TEC
TESK1	52	ZGC113162
TESK2	85	TESK2
TGFBR1	99	TGFBR1A
	99	TGFBR1B
TGFBR2	88	TGFBR2
	52	SIDKEY-101K6.5
TIE1	77	TEK
TIE2	90	TEK
TLK1	86	TLK1B
	86	TLK1A
TLK2	87	TLK2
TNIK	97	TNIKB
	94	TNIKA
	61	MINK1
TNK1	49	TNK1
TRAD	57	TRIOA
TRB1	55	TRIB1
TRB2	58	TRIB3
TRB3	49	TRIB3
TRIO	70	TRIOA
	67	TRIOB
TRKA	73	NTRK2B
TRKB	90	NTRK2B
TRKC	95	NTRK3A
TSSK1	40	SICH211-215M21.25
	40	SICH211-215M21.23
	40	SICH211-220I15.5

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
TSSK1	40	SICH211-215M21.21
	40	SICH211-220I15.3
	39	SIDKEY-206F10.7
	38	SIDKEY-217M5.9
	38	SICH211-10J20.2
	37	SICH211-215M21.13
	37	SICH211-10J20.4
TSSK2	41	SIDKEY-14A7.2
	39	SIDKEY-197D18.3
TSSK3	39	SIDKEY-248F6.3
	38	SICH211-160D20.5
	38	SICH211-238G23.1
	38	SIZFOS-754C12.2
	37	SICH211-207C6.4
	37	SICH211-202N12.2
	36	SICH211-57M13.1
TSSK4	38	SICH211-12E13.5
	38	SICH211-12E13.6
TTBK1	95	TTBK1B
TTBK2	86	SIDKEY-12H9.11
	27	PRKD3
TTK	67	TTK
TTN	74	TTNB
TXK	68	TEC
TXNDC3	23	PKN1B
TXNDC6	19	CAMKVL
TYK2	58	JAK1

Table S1 continued..

Human kinase	% identity	Zebrafish kinase
TYK2-P	64	TYK2-P
TYRO3	71	TYRO3
ULK1	85	ULK1B
	78	ULK1A
ULK2	90	ULK2
ULK3	79	ULK3
ULK4	39	STK36
VACAMKL	91	CAMKVA
	89	CAMKVB
	70	CAMKVL
VRK1	73	VRK1
VRK2	67	VRK2
VRK3	46	VRK3
Wee1	77	WEE1
	67	WEE2
Wee1B	65	WEE2
Wnk1	94	WNK1A
	94	WNK1B
	86	SIDKEYP-61G11.1
Wnk2	93	IM7152756
Wnk3	98	IM7152756
Wnk4	83	WNK1A
YANK1	87	STK32A
YANK2	87	STK32A
YANK3	81	STK32A
YES	95	YES1
	89	YRK
YSK1	95	STK25B
ZAK	82	ZAK
ZAP70	71	ZAP70

REFERENCES

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. and Sudarsanam, S. 2002, Science, 298(5600), 1912-1934.
2. Cohen, P. 2002, Nat. Rev. Drug Discov., 1(4), 309-315.
3. Hunter, T. 1995, Cell, 80(2), 225-236.
4. Hantschel, O. 2015, ACS Chem. Biol., 10(1), 234-245.
5. Zhang, C. and Bollag, G. 2010, Curr. Opin. Genet. Dev., 20(1), 79-86.
6. Patterson, H., Nibbs, R., McInnes, I. and Siebert, S. 2014, Clin. Exp. Immunol., 176(1), 1-10.
7. Fisk, M., Gajendragadkar, P. R., Maki-Petaja, K. M., Wilkinson, I. B. and Cheriyan, J. 2014, Am. J. Cardiovasc. Drugs, 14(3), 155-165.
8. Wang, Q., Zorn, J. A. and Kuriyan, J. 2014, Methods Enzymol., 548, 23-67.
9. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Buillard, V., Cerutti, L., Copley, R., Courcelle, E., Das, U., Daugherty, L., Dibley, M., Finn, R., Fleischmann, W., Gough, J., Haft, D., Hulo, N., Hunter, S., Kahn, D., Kanapin, A., Kejariwal, A., Labarga, A., Langendijk-Genevaux, P. S., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McAnulla, C., McDowall, J., Mistry, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Orengo, C., Petryszak, R., Selengut, J. D., Sigrist, C. J., Thomas, P. D., Valentin, F., Wilson, D., Wu, C. H. and Yeats, C. 2007, Nucleic Acids Res., 35, D224-228.
10. Hanks, S. K. and Hunter, T. 1995, FASEB J., 9(8), 576-596.
11. Stout, T. J., Foster, P. G. and Matthews, D. J. 2004, Curr. Pharm. Des., 10(10), 1069-1082.
12. Hanks, S. K. and Quinn, A. M. 1991, Methods Enzymol., 200, 38-62.
13. Hunter, T. 2014, Cold Spring Harb. Perspect. Biol., 6(5), a020644.
14. Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. 2002, Trends Biochem. Sci., 27(1), 54-56.
15. Manning, G., Plowman, G. D., Hunter, T. and Sudarsanam, S. 2002, Trends Biochem. Sci., 27(10), 514-520.
16. Miranda-Saavedra, D. and Barton, G. J. 2007, Proteins, 68(4), 893-914.
17. Boudeau, J., Miranda-Saavedra, D., Barton, G. J. and Alessi, D. R. 2006, Trends Cell Biol., 16(9), 443-452.

18. Zeqiraj. E. and van Aalten, D. M. 2010, Curr. Opin. Struct. Biol., 20(6), 772-781.
19. Paksa, A. and Raz, E. 2015, Curr. Opin. Cell Biol., 36, 80-85.
20. Phillips, J. B. and Westerfield, M. 2014, Dis. Model Mech., 7(7), 739-743.
21. Gibert, Y., Trengove, M. C. and Ward, A. C. 2013, Curr. Med. Chem., 20(19), 2458-2466.
22. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G. J., White, S., Chow, W., Kilian, B., Quintais, L. T., Guerra-Assuncao, J. A., Zhou, Y., Gu, Y., Yen, J., Vogel, J. H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S. F., Laird, G. K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Elliot, D., Threadgold, G., Harden, G., Ware, D., Begum, S., Mortimore, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Lloyd, C., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C. M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Lanz, C., Raddatz, G., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Schuster, S. C., Carter, N. P., Harrow, J., Ning, Z., Herrero, J., Searle, S. M., Enright, A., Geisler, R., Plasterk, R. H., Lee, C., Westerfield, M., de Jong, P. J., Zon, L. I., Postlethwait, J. H., Nüsslein-Volhard, C., Hubbard, T. J., Roest Crollius, H., Rogers, J. and Stemple, D. L. 2013, Nature, 496(7446), 498-503.
23. Glasauer, S. M. and Neuhauss, S. C. 2014, Mol. Genet. Genomics, 289(6), 1045-1060.
24. Kok, F. O., Shin, M., Ni, C. W., Gupta, A., Grosse, A. S., van Impel, A., Kirchmaier, B. C., Peterson-Maduro, J., Kourkoulis, G., Male, I., DeSantis, D. F., Sheppard-Tindell, S., Ebarasi, L., Betsholtz, C., Schulte-Merker, S., Wolfe, S. A. and Lawson, N. D. 2015, Dev. Cell, 32(1), 97-108.
25. Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Aken, B. L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S. M. J., Spudich, G., Trevanion, S. J., Yates, A., Zerbino, D. R. and Flicek, P. 2015, Nucleic Acids Res., 43, D662-D669.
26. Howe, D. G., Bradford, Y. M., Conlin, T., Eagle, A. E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S. A., Paddock, H., Pich, C., Ramachandran, S., Ruef, B. J., Ruzicka, L., Schaper, K., Shao, X., Singer, A., Sprunger, B., Van Slyke, C. E. and Westerfield, M. 2013 Nucleic Acids Res., 41, D854-860.
27. Corpet, F. 1988, Nucleic Acids Res., 16(22), 10881-10890.

28. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P. and Lopez, R. 2013, Nucleic Acids Res., 41, W597-600.
29. Bollag, G., Hirth, P., Tsai, J., Zhang, J., Ibrahim, P. N., Cho, H., Spevak, W., Zhang, C., Zhang, Y., Habets, G., Burton, E. A., Wong, B., Tsang, G., West, B. L., Powell, B., Shellooe, R., Marimuthu, A., Nguyen, H., Zhang, K. Y., Artis, D. R., Schlessinger, J., Su, F., Higgins, B., Iyer, R., D'Andrea, K., Koehler, A., Stumm, M., Lin, P. S., Lee, R. J., Grippo, J., Puzanov, I., Kim, K. B., Ribas, A., McArthur, G. A., Sosman, J. A., Chapman, P. B., Flaherty, K. T., Xu, X., Nathanson, K. L. and Nolop, K. 2010, Nature, 467(7315), 596-599.
30. Chrencik, J. E., Patny, A., Leung, I. K., Korniski, B., Emmons, T. L., Hall, T., Weinberg, R. A., Gormley, J. A., Williams, J. M., Day, J. E., Hirsch, J. L., Kiefer, J. R., Leone, J. W., Fischer, H. D., Sommers, C. D., Huang, H. C., Jacobsen, E. J., Tenbrink, R. E., Tomasselli, A. G. and Benson, T. E. 2010, J. Mol. Biol., 400(3), 413-433.
31. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. 2015, Nat. Methods, 12(1), 7-8.
32. Tagliabracci, V. S., Pinna, L. A. and Dixon, J. E. 2013, Trends Biochem. Sci., 38(3), 121-130.
33. Desvignes, T., Pontarotti, P., Fauvel, C. and Bobe, J. 2009, BMC Evol. Biol., 9, 256.
34. Pearce, D. 2003, Cell Physiol. Biochem., 13(1), 13-20.
35. Caenepeel, S., Charydczak, G., Sudarsanam, S., Hunter, T. and Manning, G. 2004, Proc. Natl. Acad. Sci., 101(32), 11707-11712.
36. Min, X., Lee, B. H., Cobb, M. H. and Goldsmith, E. J. 2004, Structure, 12(7), 1303-1311.
37. Endicott, J. A. and Noble, M. E. 2013, Biochem. Soc. Trans., 41(4), 1008-1016.