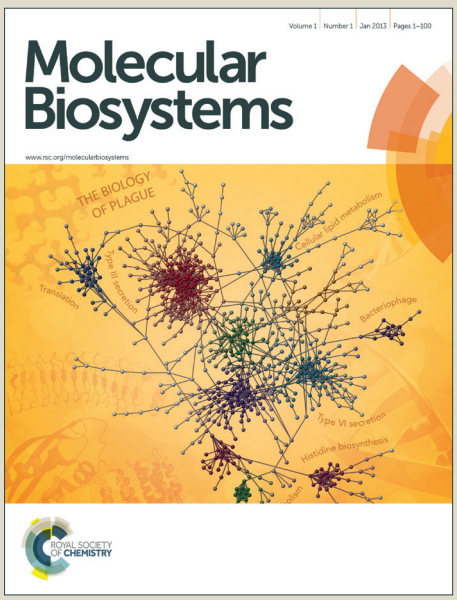
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Presence and Utility of Intrinsically Disordered Regions in Kinases

Jaymin J Kathiriya[1,6], Ravi Ramesh Pathak[1,6], Eric Clayman[1], Bin Xue[7], Vladimir N. Uversky[3,4,5,6], and Vrushank Davé[1,2]

[1]Morsani College of Medicine, Department of Pathology and Cell Biology, [2]Department of Molecular Oncology, H. Lee Moffitt Cancer Center and Research Institute, [3]Department of Molecular Medicine, [4]USF Health Byrd Alzheimer's Research Institute, University of South Florida, Tampa, FL, 33612, [5]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia, [6]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, and [7]Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida, 33620.

Running Title: Intrinsic disorder drive kinase-kinase interactions

Author Contribution: [6]J.J.K. and R.R.P contributed equally to this work. J.J.K, R.R.P., E.C and B.X performed all the experiments. The project was conceived and designed by J.J.K, R.R.P., and V.D. and written by J.J.K. and V.D. with inputs from R.R.P. and V.N.U.

To whom correspondence should be addressed: Vrushank Davé, MS/Ph.D. Assistant Professor. Dept. of Pathology & Cell Biology. Morsani College of Medicine, MDC 64. University of South Florida. 12901 Bruce B Downs Blvd, Tampa, FL 33612. Office Phone: 813-974-0930. Lab Phone: 813-974-6102 Email: vdave@health.usf.edu

Keywords: Kinase interactions; oncogenesis; SRC and AKT signaling; SMAD signaling; transcription; protein-protein interaction networks; intrinsically disordered proteins; targeted therapy; network medicine.

ABSTRACT

Since aberrant cell signaling pathways underlie majority of pathophysiological morbidities, kinase inhibitors are routinely used for pharmacotherapy. However, most kinase inhibitors suffer from adverse off-target effects. Inhibition of one kinase in a pathogenic signaling pathway elicits multiple compensatory feedback signaling loops, reinforcing the pathway rather than inhibiting it, leading to chemoresistance. Thus, development of novel computational strategies providing predictive evidence to inhibit specific set of kinases to mitigate an aberrant signaling pathway with minimum side-effects is imperative. First, our analyses reveal that many kinases contain intrinsically disordered regions, which may participate in facilitating protein-protein interactions at the kinome level. Second, we employ kinome-wide approach to identify intrinsic disorder and streamline a methodology that adds to the knowledge of therapeutically targeting kinase cascades to treat diseases. Further, we find that within the kinome network, some kinases with intrinsically disordered regions have high topological score, likely acting as kinome modulators. Third, using network analysis, we demonstrate that 5 kinases emerge as topologically most significant, forming kinome sub-networks, comprising of other kinases and transcription factors that are known to serves as drivers of disease pathogenesis. To support these findings, we have biologically validated the interplay between kinome modulators SRC and AKT kinases and uncovered their novel function in regulating transcription factors of the SMAD family. Taken together, we identify novel kinome modulators driven by intrinsic disorder, and biologically validate the thesis that therapeutic disruption of the function of kinome modulators engaged in regulatory cross-talk between disparate pathways can lead to reduced oncogenic potential in cancer cells.

INTRODUCTION

Kinases and phosphatases control phosphorylation-dephosphorylation cycles of proteins, regulating a myriad of biological processes, including cell-growth, differentiation and behavior.[1] In human kinome, there are 518 kinases comprising of 10 groups based on their sequence and structural features.[2] Kinases phosphorylate more than 90% of cellular proteins at least once during their lifetime,

altering their activity, sub-cellular localization, turn-over and macromolecular interactions, thereby, affecting intracellular signaling pathways.[3] Thus, dysregulation of kinase function underlie diseases and pathological conditions, including metabolic and neurological disorders, infectious diseases, and most importantly, cancer.[4, 5] Therefore, protein kinase inhibitors (PKIs) are widely used to inhibit aberrant kinase activity for therapy. Unfortunately, poor PKI selectivity and low efficacy associated with acquired resistance due to compensatory signaling within or across kinase pathways have limited their use in the clinic.

The major drawbacks in using known kinase inhibitors are as follows. First, the ATP cleft targeted by type I PKIs display very high promiscuity due to the conserved ATP binding site and requires a high dose to be effective, causing severe toxicities.[6-9] Second, while the competitive non-ATP type II and covalent PKIs are specific and efficient, they have met with adverse side effects, including toxicity, due to irreversible covalent binding to unanticipated kinases.[10-14] In contrast, purely allosteric PKIs, generally small molecules, targeting allosteric region outside the catalytic domain of the kinase have high selectivity and therefore are being intensely sought.[15-20] Thus, identifying allosteric PKIs remains an active area of research that continues to evolve.[21] Therefore, identification of novel targeting approaches to critical kinases within the kinome is imperative to reduce toxicity and poor efficacy.[14]

Apart from the above limitations, the pleotropic action of PKIs affects activity of multiple kinases, perturbing kinome-level functions. It has been postulated that these off-target effects can be circumvented by targeting protein-protein interaction (PPI) interfaces to abrogate pathogenic kinase-kinase interactions (KKIs).[22, 23] Indeed, since PPI surfaces are unique on each signaling kinase, targeting PPIs by peptides/peptidomimetics can be highly selective, which reduces oncogenic signaling.[22-27] Taken together, these observations indicate that identification and targeting of functionally important allosteric flexible regions coupled with disruption of PPIs at the kinome level is required to efficiently dampen pathogenic signaling cascades driven by aberrant kinase activity.

Functionally important allosteric flexible stretches of protein are often embedded in intrinsically disordered regions (IDRs), playing important regulatory roles in protein structure and function.[28] IDRs

also facilitate PPI networks by virtue of their structural plasticity, giving higher functional adaptability to a protein and, in turn, to the entire PPI network.[29, 30] By the same token, IDRs in kinases provide structural adaptability and versatility while increasing specificity, allowing stringent substrate discrimination.[31, 32] Indeed, IDRs, particularly adjacent to phosphorylation sites on kinase substrates (which will be a kinase in KKI network), enhance reciprocal accessibility and adaptability, facilitating kinase action.[33] These IDRs in kinases likely enhance their structural and functional repertoire, enabling multiple disparate interactions with diverse kinase substrates (or other kinases); a property central to forming intricate, yet efficient, regulatory signaling kinase interaction networks.[34-37] Thus, it is plausible that pathogenic KKI networks can be disrupted by therapeutically targeting IDRs, making them ideal drug targets.[38-41]

Unfortunately, since KKI networks themselves remain poorly defined, the role of IDRs in mediating KKIs and their functions in kinome networks remains unknown. While dynamic nature of the kinome and its reprogramming in response to single kinase inhibitors have been studied, the concept of entire human kinome as a singular entity for targeting in cancer therapy has only recently emerged.[42-48] Indeed, the emerging concept that cancer treatment will require a cocktail of kinase inhibitors together with other drugs reinforces the thesis that cancers will have to be targeted at the kinome level.[42, 44, 49] Since intrinsic disorder facilitate PPIs, we hypothesize that identifying IDRs will uncover important kinases forming hubs in KKI networks that will prove vital to understanding the inner-workings of KKI networks, which by extension, can be used to develop targeted drugs.

Therefore, in the present study we have performed a system wide analysis of the human kinome utilizing protein intrinsic disorder as a tool to reveal IDR driven important kinase hubs, their KKI networks, and defined their distinct roles in the pathogenesis of cancer and other diseases. Our study reveals that 417 of 504 human kinases (83%) have IDRs, prompting us to build a KKI network to elucidate the role of IDRs at the kinome level. Further investigation of KKIs reveals a unique subset of kinases involved in progression of specific diseases. We also discover a subset of kinases that emerges as critical hubs driving KKIs via phosphorylation mediated activation, a cardinal feature that drives pathogenesis of cancer and other diseases. Based on our predictive model, we propose and validate a

multiple kinase sub-network in lung cancer cells and elucidate the relationship between critical kinome modulators SRC (proto-oncogene tyrosine-protein kinase Src), a member of the TK group, and AKT (RAC-alpha serine/threonine-protein kinase), a member of the AGC group. Our analysis also reveals a new role for SRC in the regulation of SMAD (also known as MADH or **m**others **a**gainst **d**ecapentaplegic **h**omolog) activity that influences proliferation of lung cancer cells. We particularly show that while SRC and SMAD do not interact physically, their presence in our KKI network was sufficient to assert direct functional role in cell processes. Because such functional interactions cannot be revealed by proteomic approaches, we believe, this strategy also provides a novel method to identify functional molecular cross-talk, which is not based on physical association, yet may be critical in drug targeting. In summary, we have performed a first of its kind systems analysis of the human kinome, utilizing intrinsic disorder as an operating function, and demonstrated that IDRs facilitate KKIs. Using this approach, we propose and validate a new functional interaction, demonstrating SRC dependent SMAD inactivation, making this network a viable drug target in treating cancer with high SRC activity.

MATERIALS AND METHODS

Derivation of Kinases and Disorder Prediction

List of 518 kinases was compiled as published by Manning et al.[2] Kinase domain information and FASTA sequences were retrieved from UniProt (www.uniprot.org).[50] Proteins without confirmed kinase domains in the UniProt were not considered for the disorder prediction. Disorder analysis for the 504 kinases was performed using the PONDR-FIT software.[51] The software assigns a disorder score to each amino acid residue of a protein. Residues with disorder scores of greater than 0.5 were considered to be residues with structure breaking propensities, or intrinsically disordered residues. We defined an IDR as a long disordered region with a stretch of at least 25 such intrinsically disordered residues.

Validation of Disorder Prediction

IDRs prediction was performed using two different methods. First, crystal structures of 11 different kinases, one from each kinase group, were visualized using PyMOL. There was a complete overlap of predicted IDRs and missing regions from the crystal structures. Second, MobiDB[52, 53], a database tool of

protein disorder and mobility annotation was used to compare predicted IDRs and all crystal/NMR structures of a given kinase deposited in Protein Data Bank (PDB) database (www.rcsb.org)[54]. A consensus structured regions (i.e., combination of all deposited crystal structures of a given kinase) were overlapped with predicted IDRs. The predicted IDR were interrogated for concordance with experimentally validated protein structures. 100% concordance was interpreted as an IDR completely overlapping with absence of structure in the protein. For example, an IDR of 100 amino acid length with 20 of its residues overlapping a consensus structure was considered to be having 80% concordance with PDB.

Derivation of Kinome PPIs and network analysis

Experimentally validated protein-protein interaction (PPI) data for the 518 kinases comprising 10 groups was compiled using manual data curation and various softwares including Database of Interacting Proteins (DIP)[55-57], Interologous Interaction (I2D) Database[58, 59], InnateDB[60, 61], IntAct[62], MatrixDB[63, 64], The Molecular INTeraction Database (MINT)[65-68], Molcon (http://www.ebi.ac.uk/Tools/webservices/psicquic/view/main.xhtml), The Microbial Protein Interaction Database (MPIDB)[69] and BioGRID[70, 71]. For intragroup associations, PPI information for each kinase from an individual group was overlapped by VENN selections with PPI information of all other kinases from the same group (e.g. AGC group kinase with other AGC group kinases). Intergroup associations were identified by performing VENN selections of individual kinases from each of the 10 groups with kinases from every group (e.g. AGC group kinase with all kinases from remaining 9 groups). Individual PPI networks for intra and intergroup associations were constructed and visualized using Cytoscape.[72, 73] Network analysis was performed to identify topologically significant hubs from the PPI networks using Network Analyzer[74] and CentiScaPe plug in tools.[75]

Disease enrichment analysis of the kinome

Kinases from the different groups were analyzed using the Core Analysis option in the IPA software (Ingenuity® Systems, www.ingenuity.com). Disease enrichment profiles were generated for each group of kinases by selecting “Diseases and Disorders” from the “Customize Chart” function under the “Disease

and Function" output. Heat-maps representing distribution of each disease across kinase group were generated using P-Values for individual diseases and groups (Supplementary Table S2).

Cell Culture, transfection, and gene assays

H1650 (ATCC# CRL-5883) and H292 (ATCC# CRL-1848) cells were cultured in RPMI medium (Invitrogen) with 10% fetal bovine serum and 5% mixture of penicillin G, streptomycin and Plasmocin (Invitrogen) in a 5% CO_2 incubator at 37 °C. Transient transfections into H292 cells were done using the PEI method.[76] Briefly, 6-well plates at 30–50% confluence were transfected with up to 4µg of myr-AKT (a kind gift from Dr. Sellers, Harvard Medical School, Dana-Farber Cancer Institute) or (CA-SRC Addgene Plasmid 13660). 2µg SBE4-Luc (Addgene Plasmid 16495) was used as an artificial reporter of SMAD activity. Activity response was normalized with pBV-Luc empty vector (Addgene Plasmid 16539). Total DNA was normalized with corresponding amounts of pCDNA as a negative control. Two days after transfection, luciferase assays were performed using 50 µl of the supernatant. The light units were assayed by luminometry (MLX, Microtiter Luminometer, DYNEX, McLean, Virgina, USA)

Cell proliferation assay with inhibitor treatments

Cell proliferation assays were performed using Cell Counting Kit-8 (Fluka, Biochemika). 1.2 x 10^4cells were plated in each well of a 96 well plate and cultured in RPMI growth medium as described above. 16 hours after plating the cells, growth medium was removed. Cell culture medium containing 500nM of Src specific inhibitor Dasatinib (BMS-354825; Selleck chemicals, Houston, USA) or 100nM of AKT specific inhibitor MK-2206 2HCl (Selleck Chemicals) or both was added to appropriate wells. DMSO was used as a toxicity test for negative controls. 24 hours after drug treatment, the cell numbers in triplicate wells were measured as a function of absorbance (450 nm) of reduced WST-8 (2-(2-methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium, monosodium salt).

Western blots analysis

Proteins were isolated after either the inhibitor treatment (Dasatinib) or transfections (with CA-SRC) and separated by SDS-PAGE on a 10% gel and electroblotted to nitrocellulose membrane (0.1µm; Inivtrogen). Blots were blocked with either 5% TBST (10 mM Tris, pH 8, 150 mM NaCl, 0.1% Tween

20) or 5% milk and incubated in primary antibody to p-AKT (#9271; Cell Signaling), p-SMAD2/3 (11769; Santa Cruz) or β-actin (A5060; Sigma-Aldrich) overnight at 4°C. Primary antibodies to p-AKT and β-actin were used at 1:1000 dilutions. p-SMAD2/3 antibody was used at 1:500 dilutions. Peroxidase-conjugated AffiniPure Goat Anti-Rabbit IgG secondary antibody (Jackson ImmunoResearch Laboratories, INC.) was used at 1:10,000 concentration. Blots were developed by chemiluminescence (Pierce Biotechnology) and autoradiographed. Cells were inhibited with 500nM of Dasatinib - BMS-354825 (specific inhibitor of SRC) for 15 mins following which, proteins were isolated and analyzed as mentioned above.

RESULTS

Kinases engage in transient interactions to phosphorylate their target proteins, eliciting myriad cellular signaling events. Because hyper-activation of kinases causes aberrant signaling in several pathologies, inhibiting kinases has emerged as a therapeutic strategy.[14] However, PKIs have met with limited success because of their target promiscuity and toxicity. Recognizing IDRs as a structural feature influencing kinase activity that can be exploited as drug targets[33, 77-81], we performed structural and functional analyses of the kinome, investigating how IDRs influence KKIs and KKI networks that may involve non-kinase proteins (as shown in the pipeline; Figure 1). We designed a strategy to identify functionally most relevant kinases using three parametric filters: intrinsic disorder, topological significance via network analysis, and disease enrichment analysis. We identified 5 kinases that were consider being most significant, thus called “kinome modulators”, which likely act as hub proteins[82], herein called hub-kinases, central to modulating KKI network structure and function at the kinome level (Figure 1).

Prevalence of intrinsic disorder in the Kinome

We analyzed 504 kinases and computationally predicted intrinsic disorder in their structure using PONDR-FIT.[28, 51] Our analyses revealed that 417 of the 504 kinases (83%) contain IDRs (Figure 2A;

Supplementary Table S1). Intrinsic disorder was observed in kinases amongst all 10 kinase groups. At least 70% of kinases in each group have IDRs with the exception of RGC group comprising of only five kinases[2] (Figure 2A). To validate the prediction of IDRs in kinases, a kinase with solved structure that had missing segments was examined. A hallmark of an IDR is its inability to yield discernible electron density required to resolve its three-dimensional structure.[83] Upon querying the Protein Data Bank (PDB) database, SRPK2, a Serine/Arginine rich protein-specific kinase of CMGC (cyclin-dependent kinase [**C**DK], mitogen-activated protein kinase [**M**APK], glycogen synthase kinase [**G**SK3], CDC-like kinase [**C**LK]) group implicated in replication of Hepatitis B Virus,[84, 85] was found to have its crystal structure coordinates deposited (PDB-ID: 2X7G) with several regions missing, likely due to lack of crystallization. Our disorder prediction revealed that SRPK2 has two long IDRs, one at its N-terminus (amino acids 1-62) and the other in the middle of its kinase domain (amino acids 242-516) (Supplementary Table S1). Likewise, analysis of crystal structures of 10 other kinases revealed regions that did not crystalize corresponded to our prediction of IDRs (Supplementary Figure S1). Using PyMol,[86] a tool used for crystal structure visualization, the rendering of coordinates of the solved crystal structure of all eleven kinases revealed that our predicted IDRs in these kinases indeed matched with the missing residues in these structures, providing a proof of principle to our prediction of IDRs (Figure 2B).

To further support our IDR prediction, we correlated the degree of concordance between 100 random IDRs derived from 43 different kinases and their existing structural features present in a comprehensive database tool called MobiDB[52, 53]. This database predicts protein disorder and mobility annotations (using what methods??), and superimpose them on to the structural features derived from X-ray and NMR studies. The tool combines all known crystal structures of a given protein to generate consensus experimentally validated structured regions. We analyzed 100 random IDRs from 43 different kinases and plotted the percentage of the amino acid residues which were in concordance with the experimentally validated crystal structures (see materials and methods). The analysis revealed that 66% of the IDRs had a complete overlap with unstructured regions derived from PDB structures, while 85% of the IDRs had 80% or higher overlap. (Figure 2D; Supplementary Table S8). Having established the

validity of the IDR prediction by correlating the absence of IDRs in known crystal structures, we used this method and predicted that 83% of the kinases contained IDRs, suggesting that IDRs are common in kinases.

Aberrant KKIs Underlie Cancer Pathogenesis

Intrinsic disorder in proteins is functionally important.[87] However, Intrinsic disorder and its function at the kinome level remain undetermined. We probed the entire kinome and identified kinases that are rich in IDRs. Using this information, we performed diseases enrichment analysis using ingenuity pathway analysis (IPA) and revealed that subsets of kinases were associated with specific diseases (Figure 3A, Supplementary Table S2). Since aberrant cell signaling by dysfunctional kinases is associated with cancer and myriad other pathologies,[88] it is plausible that IDRs within these kinases may participate in driving the pathogenesis. Therefore, we hypothesized that dysfunctional kinases that are rich in IDRs drive aberrant signaling via KKIs driving disease pathogenesis. To test this hypothesis, we examined whether individual kinases and kinase groups that are enriched in IDRs may represent a group or groups of disease conditions. Indeed, our analysis revealed cancer to be the most significant disease driven by aberrant kinase function (Figure 3A). Three kinase groups (TK – **T**yrosine **K**inase, AGC – containing PK**A**/PK**G**/PK**C**, and Atypical kinases) are most significantly associated with cancer. Of the three groups, the TK-group of kinases are most significantly associated with cancer (77 of 90 kinases, Supplementary Table S2), indicating that tyrosine kinases are central to cancer pathogenesis, supporting previous findings in the clinic[89-91]. TK-group of kinases were also found to participate in various other diseases and disorders (Figure 3A), suggesting that there may be a strong association between certain pathogenic process with this unique sub-set of kinases.

Dysregulation in intricate gene and protein networks underlie disease pathogenesis.[92] Because kinases are highly implicated in a number of diseases and disorders, we hypothesized that enrichment of cancer as a top disease may be due to dysregulated KKI networks. To test this hypothesis, we first developed PPI networks of kinases interacting with other kinases from the same kinase group (intra-group KKIs). TK group had the highest number of interactions; 87% of TK-group kinases interacted amongst

themselves, resulting in a total of 292 interactions (Figure 3B & C), indicating that functions of TK-group kinases are widely influenced by interactions amongst themselves. For the entire kinome, 612 intra-group interactions were identified, where both the participating kinases in a given interaction belonged to the same group (Figure 3C & Supplementary Table S3). Remarkably, 84% of kinases participating in these intra-group interactions contained IDRs (Supplementary Tables S1 and S3), supporting the notion that IDRs in intra-group KKIs that contribute to biological function when disrupted may lead to diseases.

Kinases engage in Inter- and Intra-Group Interactions Utilizing Disordered Regions

The kinase groups are categorized based on their functional and structural similarities.[2] We tested the hypothesis that intra-group kinases, having similar domain structures, are likely to provide complementary surfaces that may interact with each other, which in turn, would lead to a higher number of interactions as compared to interactions between kinases from different groups (inter-group). Surprisingly, our results show that approximately 2.5 times more (1498 interaction) inter-group KKIs were identified as opposed to 610 intra-group interactions, indicating that KKIs utilize distinct and divergent regions of the kinases to engage in physical and functional interactions. Taken together, for both types of interactions, we identified 2108 interactions that comprised 385 kinases (Supplementary Table S3). Since 83% of these 385 kinases consisted of IDRs (Supplementary Table S3), the likelihood of IDRs within these kinases contributing to these interactions is high. We tested this possibility by analyzing all 501 kinases for molecular recognition features (MoRFs)[93], which are short regions within IDRs that facilitate PPI via disorder-to-order transition. We predicted a total of 2129 MoRFs within 501 kinases (Supplementary Table S9). We also observed a positive relationship between percentage disorderliness of a kinase and the number of predicted MoRFs.(Figure 4B), providing a possible mechanism of how IDRs can contribute to the 2108 KKIs. These KKIs had IDRs on interacting surfaces of both kinases, or an IDR on one kinase surface (Supplementary Table S3). Detailed analysis revealed that 90% of the interactions had at least one disordered region, further suggesting that IDRs may play a role in forming KKI networks (Figure 4A, Supplementary Table S3).

Our analysis revealed that 77% of the kinome participates in KKI (Set-A; Supplementary Table S3), which are associated with vital biological processes. By the same token, the remaining 23% of kinases in the kinome (Set-B; Supplementary Table S3) does not participate in KKIs; however they play important roles in cellular processes and disease pathogenesis. Therefore, to identify and distinguish functional roles of these two distinct set of kinases, we subjected these two sets to differential disease enrichment analysis using Ingenuity Pathway Analysis (IPA) software. Indeed, cancer was the most enriched disease driven by KKIs, while the set of kinases that did not engage in KKIs were involved in diseases that did not feature "cancer" in the top five disease groups (Figure 4C). Taken together, our results show that a distinct group of kinases drives pathogenesis of specific diseases, and that there is a considerable crosstalk between kinases likely involving IDRs in KKIs. More importantly, 90% of KKIs comprise of IDRs (Figure 4A, Supplementary Table S3), and aberrant interactions in these KKIs, in part, underlie pathogenesis of cancer. Our results provide preliminary but critical insights that will help design molecules that target KKIs for pharmacotherapy of cancer.

Topological analysis reveals SRC and AKT as two of the most significant kinome modulators

Having established that KKIs play a major role in the pathogenesis of cancer, critical hub kinases driving KKIs were sought as they may be used as drug targets. From graph theory, for kinases to be defined as hubs they have to interact with a high number of other kinases, thereby radiating a network at the systems level. It is possible that identification of such network involving multiple kinase as well as non-kinase hub proteins may become ideal for therapeutic targeting.[94] Therefore, we sought to identify these hub kinases (HKs). Our approach comprised of a three-pronged strategy that involved all10 kinase groups (Figure 5A). In the first set X, we identified 92 HKs that are highly interacting intra-group kinases (Supplementary Table S4). In the second set Y, we identified 84 HKs that are highly interacting inter-group kinases (Supplementary Table S4). While the third set Z of 76 HKs were derived by applying topological analysis, using degree centrality as a measure,[95, 96] to each group's KKIs (Supplementary Table S4)(Figure 5Ai). We further obtained a fourth set containing 40 HKs common to all the above three sets (Figure 5Aii). This is an enriched set of HKs that likely engage in critical cellular functions and

influence the kinome. Further, a secondary topological analysis (Figure 5Aiii) was performed for the 40 HKs, revealing a final list of 5 most significant kinases that we call kinome modulators (KMs). These KMs were identified based on their degree centrality. Figure 5B depicts the kinase group, the degree centrality and the total number of interactions for the 5 KMs. These 5 KMs participated in a total of 174 KKIs and involved 127 unique kinases that make up approximately 25% of the kinome (Supplementary Table S5). A TK-group member, SRC, emerged as the most significant KM amongst the 5 KMs. Supporting the significance of IDRs in kinases, 4 of the 5 KMs showed IDR enrichment, highlighting the importance of IDRs in imparting versatile functions not only to the function of the kinase protein, but to the entire kinome.

Since kinases are considered as signal transducers, apart from extensive cross-talk mediated by KKIs, interaction of kinases with non-kinase substrate proteins also have crucial effects on a variety of cellular processes. Therefore, interaction of non-kinase proteins within the sub-network formed by 5 KMs was sought at the proteome level. Analyses of non-kinase proteins facilitating functional interactions of the 5 KMs at the proteome level identified a proteome-wide interactome consisting of 1200 interactions with 963 unique proteins (Fig 5C; Supplementary Table S5). Since cancer showed up as the most affected diseases driven by KKIs, causing uncontrolled growth, proliferation, dysregulation in cell death and survival pathways,[97, 98] we hypothesized that proteins interacting with KMs may involve proteins participating in providing robustness to the cell survival machinery. To test the hypothesis, we subjected the interactome of the 5 KMs to disease and functional enrichment. Indeed, disease enrichment profile revealed cancer as the most enriched disease, while functional enrichment identified “Cell Death and Survival” as the most enriched function, indicating that the KMs exert cancer pathogenesis via its interactome comprising of 963 proteins (Supplementary Tables S6 and S7). Network analysis on this interactome using IPA revealed significant number of non-kinase proteins that likely synergize with the KMs and help drive cancer pathogenesis. 8 of these proteins that interacted with at least 4 of the 5 KMs (Figure 5D) were considered to be functionally most significant. Interestingly, 2 of these proteins were transcription factors SMAD3 and STAT3 (**S**ignal **T**ransducer and **A**ctivator of **T**ranscription **3**) critical in

driving oncogenesis[99-102] (Figure 5D, highlighted). Thus, the interactions of kinases with non-kinase proteins revealed a putative network that bridged kinases with target transcription factors, providing an ideal entrée to identifying target molecules in cancer treatment. Indeed, STAT3 is a well-established target for cancer therapy.[102, 103] Likewise role of SMAD3 in cancer is well established[100]. Taken together, our analysis identified 5 KMs, which emerged from 518 kinases, as interacting directly with 25% of the kinome, influencing cancer pathogenesis via KKIs. Importantly, our studies demonstrated that multiple KMs interact with important transcription factors, expanding the repertoire of possible kinome sub-networks that may influence critical cellular processes.

Two regulatory hub kinases, SRC and AKT relay inhibition of SMAD activity

Our KKI network analysis revealed a physical interaction between the two KMs, SRC and AKT kinases, representing TK and AGC groups respectively[104]. These two kinase groups play a significant role in driving pathogenesis of myriad diseases including cancer[91, 105] (Figure 3A). Therefore, we hypothesized that SRC and AKT, being KMs, engage in the formation of a functional sub-network. To test this hypothesis, we interrogated the existence of such a sub-network that would underlie a physiological process *in vivo*. To achieve this goal, first, we independently inhibited SRC and AKT activity by dasatinib and MK-2206 respectively in a highly proliferative H292 human lung cancer cell line. While both reduced cell proliferation up to ~50%, simultaneous inhibition of both SRC and AKT further reduced proliferation of the H292 human lung cancer cells significantly by ~70% (Figure 6A). Our results reinforce the concept that a sub-network driving cell proliferation was being modulated by the two KMs SRC and AKT kinases, targeting of which abrogated cell proliferation. Second, to further confirm our hypothesis that AKT and SRC form a functional sub-network, we used H1650 lung cancer cell line with very high AKT (due to loss of PTEN) and SRC activity (due to EGFR exon 20 deletion).[106] We examined the SRC-AKT crosstalk in H1650 lung cancer cells. Indeed, simultaneous inhibition of these two KMs significantly reduced proliferation of cells despite having hyperactive AKT and SRC (Figure 6B), providing a novel clinical rationale that a combination of SRC and AKT inhibitors may be a superior therapeutic strategy for treating lung cancer patients.

Although inhibition of SRC and AKT in our KKI network revealed functional relatedness between these two KMs, we also wanted to test whether SRC could influence the two transcription factors STAT3 and SMAD3 present in our network (Figure 5C). While SRC mediated STAT3 activation is known,[107, 108] the relationship between SRC and SMAD3 remains unexplored. Interestingly, although our analysis revealed that SMAD3 is a part of our sub-network, it does not interact with SRC kinase directly. However, since SRC is a part of the sub-network, we hypothesized that activity of SMADs was indirectly influenced by SRC. To test this hypothesis, we overexpressed constitutively active form of SRC (CA-SRC) and examined the role of SMADs by measuring the luciferase activity of an artificial promoter-reporter plasmid comprising multiple canonical SMAD Binding Elements (SBE-promoter)[109] (Figure 6C). CA-SRC significantly inhibited SBE-promoter activity by 60% (Figure 5C, lane 3), suggesting an indirect effect of SRC on SMAD mediated transcriptional activity. To define how SRC activity was relayed onto SMAD, most likely via an intermediary protein, we performed the following experiment. Given that inhibition of SMAD by AKT is well established[110, 111] and SRC directly regulates AKT in our sub-network,[104] we tested the hypothesis that inhibition of SMADs by CA-SRC may be, in part, mediated via AKT signaling. Supporting this concept, we discovered that dasatinib mediated inhibition of SRC activity indeed reduced p-AKT levels in H292 and H1650 lung cancer cells (Figure 6D). To further confirm this result, we over-expressed a constitutively active form of AKT called myr-AKT (myristoylated-AKT). Increasing concentration of myr-AKT significantly decreased SMAD-mediated transcriptional activity by ~60% (Figure 6E) consistent with previous reports.[110, 111] While overexpression of CA-SRC increased p-AKT levels, it decreased p-SMAD2/3 levels (Figure 7A), indicating that SRC is likely mediating its inhibitory activity on SMAD2/3 via activation of AKT. Taken together, we validated the proposed model of a sub-network comprising of KMs and transcription factor SMAD2/3 (Figure 7B), confirming our newly predicted relationships between the KMs and their target proteins despite lack of physical interactions. Taken together, we demonstrate that top KMs are able to regulate each other in a sub-network, driving tumorigenic properties of lung cancer cells via transcription factors. This

coordination between kinase hubs and transcription factors can be exploited to elicit therapeutic benefits in lung cancer patients.

DISCUSSION AND CONCLUSIONS

Recent large scale systems level proteomic studies have revealed that most pathological conditions are driven by perturbation in PPIs at the network level, including interactions amongst kinases.[112-114] As a result, current approaches of targeting one kinase in a signaling pathway are not sufficient to inhibit pathogenic signaling. Moreover, sustained use of kinase inhibitors not only show toxicity due to off-target activities, but also elicits de novo signaling feedback loops leading to dynamic rewiring of kinase signaling cascades,[46, 47] reinforcing aberrant signaling rather than suppressing it. Thus, there is an unmet need to understand the signaling networks at the kinome level. For this purpose, identifying and targeting hub kinases or kinome modulators within the kinome, for therapeutic ends is imperative.

In the present study, using structural informatics on the entire human kinome, we have discovered that IDRs may play a significant role in expanding KKI repertoire. It is plausible that the inherent structural and functional plasticity of the IDRs that are present in the proteins may readily rewire signaling associated with pathogenesis of various diseases as speculated before[33, 38, 51, 115]. Our analysis will prompt further investigation to establish causative relationship between IDRs and KKIs. Moreover, the dynamic rewiring of the kinome signaling via KKIs following chemotherapy is an area of active research[46, 47, 116-118]. These aberrant yet robust alternative signaling pathways activated following chemotherapy may, in part, emanate from such rewiring potential of KKI via IDRs, causing chemoresistance. Such hitherto unexplored mechanism needs to be considered to understand and mitigate chemoresistance. These IDRs are therefore potential drug target regions residing and functioning within critical hub kinases. Our analysis also reveals an unprecedented amount of cross-talk within the kinases themselves, contributing to diseases pathogenesis and progression. Combining computational and systems biology approaches with extensive data-mining, we have identified and validated the biology of a KKI

sub-network comprising SRC and AKT. Our analysis uncovered a new role of SRC kinase in modulating SMAD activity via AKT in lung cancer cells. Such functional kinase-transcription factor interactions with no direct physical interactions are thus readily detected by our approach, filling a significant technical gap, which will complement proteomic approaches that are designed to identify only direct PPI based functional networks.

IDRs and their ability to confer pliancy to the functional proteome potentiate PPIs and increases functional repertoire of the proteins in various signaling cascades.[29, 30] However, the presence of IDRs and their potential roles in the human kinome had remained hitherto unexplored. Given the versatility of kinases and their ability to affect virtually all signaling pathways, herein we performed structural analysis of the human kinome. Our analyses revealed that 83% of human kinases have IDRs, supporting the notion that presence of IDRs may render kinases highly versatile yet functionally specific. Our studies also discovered hub kinases containing IDRs with a functional advantage, allowing them to recognize their interaction partners and also in interacting with multiple proteins.[119]

To test the hypothesis that kinases engage in KKI and reciprocate functional modulation, which in turn may optimize their roles and activities either via PTMs and/or PPIs, several PPI databases were probed. We derived a kinome network of experimentally validated KKIs to reveal extensive interactions across as well as within each of the ten kinase groups. We found that KKIs are widespread and plays regulatory roles in the kinome, which will strengthen our understanding of the rewiring within the kinome during and after therapeutic treatments. Furthering our hypothesis that IDRs drive KKIs, we also discovered that 90% of the KKIs (1906 of 2108 kinases) occur when at least one of the two interacting kinases has an IDR, suggesting functioning of the KKI network may involve IDRs and, by extension, kinase cascades in pathophysiologically relevant signaling events. For example, our analysis revealed that while a distinct set of 116 kinases exclusively interact with non-kinase proteins, it is the KKIs within the remaining 385 kinases that underlie cancer pathogenesis. Further, detailed examination of our functional enrichment of the kinases also revealed that distinct kinases were associated with specific diseases, providing functional and clinical relevance to studying these kinases and their relevance in the kinome.

The propensity of hub proteins and genes to drive a network of diseases has necessitated a network medicine approach to tackle diseases reflecting complex intracellular and intercellular networks.[120] Cancer therapy suffers from the dynamic ability of the kinome to bypass single kinase inhibition by activating other kinases. Thus, our study proposes that an understanding of kinase network-modules derived from experimentally validated KKI networks is necessary to discover components of the compensatory signaling responses causing drug-resistance. We derived and validated a SRC and AKT driven network, supporting the notion that KMs are functionally interconnected. One way by which KMs likely interact is via IDRs. Interestingly, a common IDR in the AKT C-terminus functionally interacts with SRC and also participates in inhibition of SMAD phosphorylation.[104, 111] Thus, it is plausible that inhibition of SRC inactivates AKT via this IDR, which in turn will increase SMAD activity, as reflected in our studies. Having identified a new role for SRC in influencing AKT-SMAD axis, we also showed that SRC and AKT activity concertedly increased the proliferation of cancer cells. Our results support this concept since SRC inhibitor can significantly inhibit proliferation of lung cancer cells harboring hyper-activated AKT and SRC.

In summary, we provide evidence that systems level exploration of the human kinome through structural analysis leads to identification of new molecular relationships among apparently distinct kinase-driven pathways. Using our approach, similar kinase sub-network modules can be validated using *in vitro* as well as *in vivo* models to identify new roles of kinases. Additionally, our study also reveals new target regions in the form of IDRs that can be utilized to design small molecule drugs/peptidomimetics to disrupt kinase hubs. Given that while our present analysis is limited to a small subset of experimentally validated KKIs, we believe that the same approach can be utilized on a larger set of predicted kinase-substrate and kinase-kinase relationships, providing new therapeutic opportunities. In addition, studies that reveal high confidence predictions of kinase-substrate relationships[121] and drug side-effects[122] can be utilized to predict relationships within the KKIs and ascertain IDRs in kinase networks mediating disease pathogenesis providing cues to novel drug development strategies.

Acknowledgements: We thank Ms. Prerna Malaney for critically reading this manuscript.

REFRENCES:

1. G. Manning, G. D. Plowman, T. Hunter and S. Sudarsanam, *Trends in biochemical sciences*, 2002, 27, 514-520.
2. G. Manning, D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam, *Science*, 2002, 298, 1912-1934.
3. C. Petretti and C. Prigent, *Biology of the cell / under the auspices of the European Cell Biology Organization*, 2005, 97, 113-118.
4. G. E. Tusnady, Z. Dosztanyi and I. Simon, *Nucleic acids research*, 2005, 33, D275-278.
5. H. Yang, V. Guranovic, S. Dutta, Z. Feng, H. M. Berman and J. D. Westbrook, *Acta crystallographica. Section D, Biological crystallography*, 2004, 60, 1833-1839.
6. J. Bain, L. Plater, M. Elliott, N. Shpiro, C. J. Hastie, H. McLauchlan, I. Klevernic, J. S. Arthur, D. R. Alessi and P. Cohen, *The Biochemical journal*, 2007, 408, 297-315.
7. M. W. Karaman, S. Herrgard, D. K. Treiber, P. Gallant, C. E. Atteridge, B. T. Campbell, K. W. Chan, P. Ciceri, M. I. Davis, P. T. Edeen, R. Faraoni, M. Floyd, J. P. Hunt, D. J. Lockhart, Z. V. Milanov, M. J. Morrison, G. Pallares, H. K. Patel, S. Pritchard, L. M. Wodicka and P. P. Zarrinkar, *Nature biotechnology*, 2008, 26, 127-132.
8. B. W. Jester, K. J. Cox, A. Gaj, C. D. Shomin, J. R. Porter and I. Ghosh, *Journal of the American Chemical Society*, 2010, 132, 11727-11735.
9. Z. A. Knight and K. M. Shokat, *Chemistry & biology*, 2005, 12, 621-637.
10. L. Garuti, M. Roberti and G. Bottegoni, *Current medicinal chemistry*, 2010, 17, 2804-2821.
11. M. S. Cohen, C. Zhang, K. M. Shokat and J. Taunton, *Science*, 2005, 308, 1318-1321.
12. E. L. Kwak, R. Sordella, D. W. Bell, N. Godin-Heymann, R. A. Okimoto, B. W. Brannigan, P. L. Harris, D. R. Driscoll, P. Fidias, T. J. Lynch, S. K. Rabindran, J. P. McGinnis, A. Wissner, S. V. Sharma, K. J. Isselbacher, J. Settleman and D. A. Haber, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102, 7665-7670.
13. S. Peroukides, T. Makatsoris, A. Koutras, A. Tsamandas, A. Onyenadum, C. Labropoulou-Karatza and H. Kalofonos, *World journal of gastroenterology : WJG*, 2011, 17, 2349-2352.
14. J. Zhang, P. L. Yang and N. S. Gray, *Nature reviews. Cancer*, 2009, 9, 28-39.
15. J. R. Simard, M. Getlik, C. Grutter, V. Pawar, S. Wulfert, M. Rabiller and D. Rauh, *Journal of the American Chemical Society*, 2009, 131, 13286-13296.
16. C. Pargellis, L. Tong, L. Churchill, P. F. Cirillo, T. Gilmore, A. G. Graham, P. M. Grob, E. R. Hickey, N. Moss, S. Pav and J. Regan, *Nature structural biology*, 2002, 9, 268-272.
17. K. J. Cox, C. D. Shomin and I. Ghosh, *Future medicinal chemistry*, 2011, 3, 29-43.
18. S. F. Barnett, D. Defeo-Jones, S. Fu, P. J. Hancock, K. M. Haskell, R. E. Jones, J. A. Kahana, A. M. Kral, K. Leander, L. L. Lee, J. Malinowski, E. M. McAvoy, D. D. Nahas, R. G. Robinson and H. E. Huber, *Biochemical Journal*, 2005, 385, 399-408.
19. V. Calleja, M. Laguerre, P. J. Parker and B. Larijani, *PLoS biology*, 2009, 7, e17.
20. P. Liu, H. Cheng, T. M. Roberts and J. J. Zhao, *Nat Rev Drug Discov*, 2009, 8, 627-644.
21. A. C. Dar and K. M. Shokat, *Annual review of biochemistry*, 2011, 80, 769-795.
22. M. R. Arkin and J. A. Wells, *Nat Rev Drug Discov*, 2004, 3, 301-317.
23. M. R. Arkin and A. Whitty, *Current opinion in chemical biology*, 2009, 13, 284-290.

24. A. G. Tzakos, D. Fokas, C. Johannes, V. Moussis, E. Hatzimichael and E. Briasoulis, *Molecules*, 2011, 16, 4408-4427.
25. A. Mullard, *Nat Rev Drug Discov*, 2012, 11, 172-174.
26. T. Berg, Aachen, Germany, 2009.
27. F. Rechfeld, P. Gruber, J. Hofmann and J. Kirchmair, *Current topics in medicinal chemistry*, 2011, 11, 1305-1319.
28. A. C. Ferreon, J. C. Ferreon, P. E. Wright and A. A. Deniz, *Nature*, 2013, 498, 390-394.
29. P. M. Kim, A. Sboner, Y. Xia and M. Gerstein, *Molecular systems biology*, 2008, 4, 179.
30. P. Malaney, R. R. Pathak, B. Xue, V. N. Uversky and V. Dave, *Scientific reports*, 2013, 3, 2035.
31. J. A. Ubersax and J. E. Ferrell Jr, *Nat Rev Mol Cell Biol*, 2007, 8, 530-541.
32. J. Liu, J. R. Faeder and C. J. Camacho, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106, 19819-19823.
33. L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic and A. K. Dunker, *Nucleic acids research*, 2004, 32, 1037-1049.
34. P. E. Wright and H. J. Dyson, *Journal of molecular biology*, 1999, 293, 321-331.
35. R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed and P. E. Wright, *Proceedings of the National Academy of Sciences of the United States of America*, 1996, 93, 11504-11509.
36. A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger and J. E. Villafranca, *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 1998, 473-484.
37. F. Karush, *Journal of the American Chemical Society*, 1950, 72, 2705-2713.
38. V. N. Uversky, C. J. Oldfield and A. K. Dunker, *Annual review of biophysics*, 2008, 37, 215-246.
39. M. Anurag and D. Dash, *Molecular bioSystems*, 2009, 5, 1752-1757.
40. S. J. Metallo, *Current opinion in chemical biology*, 2010, 14, 481-488.
41. V. N. Uversky, *Expert opinion on drug discovery*, 2012, 7, 475-488.
42. L. Zhang and R. J. Daly, *Critical reviews in oncogenesis*, 2012, 17, 233-246.
43. R. J. Daly, *Critical reviews in oncogenesis*, 2012, 17, vi-vii.
44. Z. A. Knight, H. Lin and K. M. Shokat, *Nature reviews. Cancer*, 2010, 10, 130-137.
45. O. Fedorov, S. Muller and S. Knapp, *Nature chemical biology*, 2010, 6, 166-169.
46. L. M. Graves, J. S. Duncan, M. C. Whittle and G. L. Johnson, *The Biochemical journal*, 2013, 450, 1-8.
47. J. S. Duncan, M. C. Whittle, K. Nakamura, A. N. Abell, A. A. Midland, J. S. Zawistowski, N. L. Johnson, D. A. Granger, N. V. Jordan, D. B. Darr, J. Usary, P. F. Kuan, D. M. Smalley, B. Major, X. He, K. A. Hoadley, B. Zhou, N. E. Sharpless, C. M. Perou, W. Y. Kim, S. M. Gomez, X. Chen, J. Jin, S. V. Frye, H. S. Earp, L. M. Graves and G. L. Johnson, *Cell*, 2012, 149, 307-321.
48. S. Chu, M. Holtz, M. Gupta and R. Bhatia, *Blood*, 2004, 103, 3167-3174.
49. E. ter Haar, W. P. Walters, S. Pazhanisamy, P. Taslimi, A. C. Pierce, G. W. Bemis, F. G. Salituro and S. L. Harbeson, *Mini reviews in medicinal chemistry*, 2004, 4, 235-253.
50. T. U. Consortium, *Nucleic acids research*, 2013, 41, D43-D47.
51. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker and V. N. Uversky, *Biochimica et biophysica acta*, 2010, 1804, 996-1010.
52. T. Di Domenico, I. Walsh, A. J. Martin and S. C. Tosatto, *Bioinformatics*, 2012, 28, 2080-2081.

53. T. Di Domenico, I. Walsh and S. C. Tosatto, *BMC bioinformatics*, 2013, 14 Suppl 7, S3.
54. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic acids research*, 2000, 28, 235-242.
55. I. Xenarios, D. W. Rice, L. Salwinski, M. K. Baron, E. M. Marcotte and D. Eisenberg, *Nucleic acids research*, 2000, 28, 289-291.
56. I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S. M. Kim and D. Eisenberg, *Nucleic acids research*, 2002, 30, 303-305.
57. I. Xenarios, E. Fernandez, L. Salwinski, X. J. Duan, M. J. Thompson, E. M. Marcotte and D. Eisenberg, *Nucleic acids research*, 2001, 29, 239-241.
58. K. R. Brown and I. Jurisica, *Genome biology*, 2007, 8, R95.
59. K. R. Brown and I. Jurisica, *Bioinformatics*, 2005, 21, 2076-2082.
60. D. J. Lynn, G. L. Winsor, C. Chan, N. Richard, M. R. Laird, A. Barsky, J. L. Gardy, F. M. Roche, T. H. Chan, N. Shah, R. Lo, M. Naseer, J. Que, M. Yau, M. Acab, D. Tulpan, M. D. Whiteside, A. Chikatamarla, B. Mah, T. Munzner, K. Hokamp, R. E. Hancock and F. S. Brinkman, *Molecular systems biology*, 2008, 4, 218.
61. K. Breuer, A. K. Foroushani, M. R. Laird, C. Chen, A. Sribnaia, R. Lo, G. L. Winsor, R. E. Hancock, F. S. Brinkman and D. J. Lynn, *Nucleic acids research*, 2013, 41, D1228-1233.
62. S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni and H. Hermjakob, *Nucleic acids research*, 2014, 42, D358-363.
63. E. Chautard, L. Ballut, N. Thierry-Mieg and S. Ricard-Blum, *Bioinformatics*, 2009, 25, 690-691.
64. E. Chautard, M. Fatoux-Ardore, L. Ballut, N. Thierry-Mieg and S. Ricard-Blum, *Nucleic acids research*, 2011, 39, D235-240.
65. A. Chatr-aryamontri, A. Ceol, L. M. Palazzi, G. Nardelli, M. V. Schneider, L. Castagnoli and G. Cesareni, *Nucleic acids research*, 2007, 35, D572-574.
66. A. Zanzoni, L. Montecchi-Palazzi, M. Quondam, G. Ausiello, M. Helmer-Citterich and G. Cesareni, *FEBS letters*, 2002, 513, 135-140.
67. L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico, L. Castagnoli and G. Cesareni, *Nucleic acids research*, 2012, 40, D857-861.
68. A. Ceol, A. Chatr Aryamontri, L. Licata, D. Peluso, L. Briganti, L. Perfetto, L. Castagnoli and G. Cesareni, *Nucleic acids research*, 2010, 38, D532-539.
69. J. Goll, S. V. Rajagopala, S. C. Shiau, H. Wu, B. T. Lamb and P. Uetz, *Bioinformatics*, 2008, 24, 1743-1744.
70. C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, *Nucleic acids research*, 2006, 34, D535-539.
71. C. Stark, B. J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski and M. Tyers, *Nucleic acids research*, 2011, 39, D698-704.
72. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, 2003, 13, 2498-2504.

73. M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker and G. D. Bader, *Nature protocols*, 2007, 2, 2366-2382.
74. Y. Assenov, F. Ramirez, S. E. Schelhorn, T. Lengauer and M. Albrecht, *Bioinformatics*, 2008, 24, 282-284.
75. G. Scardoni, M. Petterlini and C. Laudanna, *Bioinformatics*, 2009, 25, 2857-2859.
76. O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix and J. P. Behr, *Proceedings of the National Academy of Sciences of the United States of America*, 1995, 92, 7297-7301.
77. Y. Shan, M. P. Eastwood, X. Zhang, E. T. Kim, A. Arkhipov, R. O. Dror, J. Jumper, J. Kuriyan and D. E. Shaw, *Cell*, 2012, 149, 860-870.
78. F. D. Smith, S. L. Reichow, J. L. Esseltine, D. Shi, L. K. Langeberg, J. D. Scott and T. Gonen, *eLife*, 2013, 2, e01319.
79. J. A. Ubersax and J. E. Ferrell, Jr., *Nat Rev Mol Cell Biol*, 2007, 8, 530-541.
80. X. B. U. V. N., *Intrinsically Disordered Proteins*, 2013, 1.
81. R. R., *Thorax*, 2013, 68.
82. Y. Pritykin and M. Singh, *PLoS computational biology*, 2013, 9, e1003243.
83. A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner and Z. Obradovic, *Journal of molecular graphics & modelling*, 2001, 19, 26-59.
84. H. Daub, S. Blencke, P. Habenberger, A. Kurtenbach, J. Dennenmoser, J. Wissing, A. Ullrich and M. Cotten, *Journal of virology*, 2002, 76, 8124-8137.
85. Y. Zheng, X. D. Fu and J. H. Ou, *Virology*, 2005, 342, 150-158.
86. Schrodinger, LLC, unpublished work.
87. A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva and Z. Obradovic, *Biochemistry*, 2002, 41, 6573-6582.
88. C. London, *Veterinary and comparative oncology*, 2004, 2, 177-193.
89. C. H. Huang and B. C. Powers, *Clinical Medicine Insights. Oncology*, 2012, 6, 137-147.
90. P. M. Ellis, W. Morzycki, B. Melosky, C. Butts, V. Hirsh, F. Krasnoshtein, N. Murray, F. A. Shepherd, D. Soulieres, M. S. Tsao and G. Goss, *Curr Oncol*, 2009, 16, 27-48.
91. A. Arora and E. M. Scholar, *The Journal of pharmacology and experimental therapeutics*, 2005, 315, 971-979.
92. K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabasi, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, 8685-8690.
93. F. M. Disfani, W. L. Hsu, M. J. Mizianty, C. J. Oldfield, B. Xue, A. K. Dunker, V. N. Uversky and L. Kurgan, *Bioinformatics*, 2012, 28, i75-83.
94. Z. Spiro, I. A. Kovacs and P. Csermely, *Journal of biology*, 2008, 7, 20.
95. S. Maslov and K. Sneppen, *Science*, 2002, 296, 910-913.
96. E. Estrada, *Proteomics*, 2006, 6, 35-40.
97. S. Cory, A. Strasser, T. Jacks, L. M. Corcoran, T. Metz, A. W. Harris and J. M. Adams, *Cold Spring Harbor symposia on quantitative biology*, 1994, 59, 365-375.

98. G. McGill and D. E. Fisher, *Frontiers in bioscience : a journal and virtual library*, 1997, 2, d353-379.
99. S. Kawamata, K. Matsuzaki, M. Murata, T. Seki, K. Matsuoka, Y. Iwao, T. Hibi and K. Okazaki, *Inflammatory bowel diseases*, 2011, 17, 683-695.
100. E. Tarasewicz and J. S. Jeruss, *Cell Cycle*, 2012, 11, 2443-2451.
101. K. R. Reddy, Y. Guan, G. Qin, Z. Zhou and N. Jing, *The Prostate*, 2011, 71, 1796-1809.
102. N. Jing and D. J. Tweardy, *Anti-cancer drugs*, 2005, 16, 601-607.
103. M. Zhao, B. Jiang and F. H. Gao, *Current medicinal chemistry*, 2011, 18, 4012-4018.
104. T. Jiang and Y. Qiu, *The Journal of biological chemistry*, 2003, 278, 15789-15793.
105. L. R. Pearce, D. Komander and D. R. Alessi, *Nat Rev Mol Cell Biol*, 2010, 11, 9-22.
106. T. S. Peat, J. A. Christopher and J. Newman, *Acta crystallographica. Section D, Biological crystallography*, 2005, 61, 1662-1669.
107. J. Turkson, T. Bowman, R. Garcia, E. Caldenhoven, R. P. De Groot and R. Jove, *Molecular and cellular biology*, 1998, 18, 2545-2552.
108. R. Garcia, T. L. Bowman, G. Niu, H. Yu, S. Minton, C. A. Muro-Cacho, C. E. Cox, R. Falcone, R. Fairclough, S. Parsons, A. Laudano, A. Gazit, A. Levitzki, A. Kraker and R. Jove, *Oncogene*, 2001, 20, 2499-2513.
109. L. Zawel, J. L. Dai, P. Buckhaults, S. Zhou, K. W. Kinzler, B. Vogelstein and S. E. Kern, *Molecular cell*, 1998, 1, 611-617.
110. A. R. Conery, Y. Cao, E. A. Thompson, C. M. Townsend, Jr., T. C. Ko and K. Luo, *Nature cell biology*, 2004, 6, 366-372.
111. I. Remy, A. Montmarquette and S. W. Michnick, *Nature cell biology*, 2004, 6, 358-365.
112. A. del Sol, R. Balling, L. Hood and D. Galas, *Current Opinion in Biotechnology*, 2010, 21, 566-571.
113. E. Michalsky, M. Dunkel, A. Goede and R. Preissner, *BMC bioinformatics*, 2005, 6, 122.
114. S. Dutta and H. M. Berman, *Structure*, 2005, 13, 381-388.
115. M. M. Babu, R. van der Lee, N. S. de Groot and J. Gsponer, *Current opinion in structural biology*, 2011, 21, 432-440.
116. T. J. Stuhlmiller, H. S. Earp and G. L. Johnson, *Clinical pharmacology and therapeutics*, 2014, 95, 413-415.
117. T. R. Wilson, J. Fridlyand, Y. Yan, E. Penuel, L. Burton, E. Chan, J. Peng, E. Lin, Y. Wang, J. Sosman, A. Ribas, J. Li, J. Moffat, D. P. Sutherlin, H. Koeppen, M. Merchant, R. Neve and J. Settleman, *Nature*, 2012, 487, 505-509.
118. S. A. Rosenzweig, *Biochemical pharmacology*, 2012, 83, 1041-1048.
119. C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal and L. M. Iakoucheva, *PLoS computational biology*, 2006, 2, e100.
120. A. L. Barabasi, N. Gulbahce and J. Loscalzo, *Nature reviews. Genetics*, 2011, 12, 56-68.
121. R. H. Newman, J. Hu, H. S. Rho, Z. Xie, C. Woodard, J. Neiswinger, C. Cooper, M. Shirley, H. M. Clark, S. Hu, W. Hwang, J. S. Jeong, G. Wu, J. Lin, X. Gao, Q. Ni, R. Goel, S. Xia, H. Ji, K. N. Dalby, M. J. Birnbaum, P. A. Cole, S. Knapp, A. G. Ryazanov, D. J. Zack, S. Blackshaw, T. Pawson, A. C. Gingras, S. Desiderio, A. Pandey, B. E. Turk, J. Zhang, H. Zhu and J. Qian, *Molecular systems biology*, 2013, 9, 655.
122. M. Kuhn, M. Al Banchaabouchi, M. Campillos, L. J. Jensen, C. Gross, A. C. Gavin and P. Bork, *Molecular systems biology*, 2013, 9, 663.

FIGURE LEGENDS

Figure 1: Flowchart depicting approach leading to identification of Kinome Modulators.

(A1) 504 human kinases subjected to PONDR-FIT to predict intrinsic disorder in the human kinome. Long stretches of Intrinsically Disordered Regions (IDRs) were identified in 83% of the human kinome (417 of 504 kinases). Kinome dendrogram illustration is reproduced courtesy of Cell Signaling Technology, Inc. (www.cellsignal.com)**. (A2)** Disease Enrichment Analysis using Ingenuity Pathway Analysis (IPA) revealed Cancer as the most enriched disease in addition to specific kinase groups driving a specific set of diseases and disorders. **(B)** To understand the mechanisms behind enrichment of different diseases, we analyzed the KKIs of the entire kinome and built protein-protein interaction (PPI) network. **(C)** A closer look at the KKIs revealed two sets of interactions. **(C1)**: Interactions between kinases from two different groups, and **(C2):** Interactions between kinases from the same group. Combining the two sets, we constructed a vibrant KKI network of 385 kinases. These 385 kinases were most enriched in cancer pathogenesis. Interestingly, 90% of the 2100 identified interactions were mediated by disordered kinases. **(D)** Topological analysis was performed to identify hub kinases essential to the kinase interaction network. We identified 5 topologically most significant “kinome modulators”, 4 of which are intrinsically disordered kinases. These kinome modulators interact with each other, augmenting high cross-talk within the entire kinome. We proposed and validated the interplay between SRC and AKT kinases and reveal a new role of SRC in modulation of SMAD activity via p-AKT.

Figure 2: Wide-spread Prevalence of Intrinsic Disorder in the Kinome. (A) PONDR-FIT analysis predicts Intrinsically Disordered Regions in 83% Kinases. PONDR-FIT analysis was used to predict Intrinsically Disordered Regions (IDRs) in 417 of 504 (83%) human kinases categorized in 10 groups. A region was considered an IDR if a stretch of 25 or more consecutive amino acids had a disorder score of 0.5 or more. The colors of the bars represent color coding of different Kinase Groups followed throughout the paper. Of note, RGC group has only 5 kinases. **(B&C)** IDRs are unable to crystallize. The protein structure of SRPK2 was obtained from PDB database (pdb: 2X7G). PyMOL visualization reveals missing structured regions from the crystallized parts of the protein. Numbers in horizontal cartoon are AA

residue number indicative of predicted IDRs and crystallized structured regions. The structure shows missing N-term IDR and missing part of the Kinase Domain (KD) represented by red dotted line. (**D**) 85% of IDRs are absent in solved structures of their respective proteins. We analyzed 100 IDRs, which were randomly picked from a total of 43 proteins to query their presence in their solved structures. We reveal that 66 out of 100 IDRs (66%) had complete or concordance while 85 out of 100 IDRs (85%) had 80% or higher concordance with solved structures. We utilized MobiDB (see materials and methods) to analyze the solved structures. MobiDB database combines all deposited crystal structures (through NMR and X-ray crystallography) to generate consensus structured regions. Red-filled circles represent 66 IDRs that had complete concordance (100%); green-filled circles represent 19 IDRs that had high concordance (>80%); black-filled circles represent 15 IDRs with poor concordance (<80%).

Figure 3: Intra-group interaction analysis highlights Tyrosine Kinase Group. (A) Disease Association analysis enriches Cancer as the most significant disease driven by 339 kinases. The heat map shows group-wise participation of kinases in various diseases. The analysis was done using Ingenuity Pathway Analysis (IPA). IPA core analysis revealed significant enrichment of Cancer across all Kinase groups, specifically in TK and AGC groups. **(B)** Protein-Protein Interaction (PPI) data mining revealed a high number of "Intra-group" PPIs. Here, a network is created using 292 interactions found among 77 Kinases out 90 Kinases belonging to TK groups – the largest of all 10 groups. Software used: Cytoscape. **(C)** List of kinase groups participating in the number of intra-group interactions. "Kinases" refers to the number of kinases in each group that participate in intra-group interactions.

Figure 4: Aberrant KKI underlie cancer pathogenesis.

(A) Kinome level PPI analysis enriches TK group as the highest interacting kinase group. Further analysis of PPIs within (intra-) and between (inter-) kinase groups identified a total of 2108 interactions within 393 of 504 kinases. This data is used to create a Group-Group Interaction (GGI) model. Size of the circle corresponds to the total number of interactions (Inter- and Intra-group) the kinases from a specific group participates in. Color of the circle represents per cent DO-DO/DO-O (red) or O-O (blue) interactions. (B) Proteins with increased disorderliness are rich in molecular recognition features (MoRFs). We analyzed

501 proteins and calculated their degree of disorderliness (in percentage) as per the following formula: (AA with disorder score ≥ 0.5 ÷ total number of AA in the protein) X 100. We used MoRFPred to predict MoRFs in 501 proteins, which were then normalized to the protein size (number of MoRFs ÷ Total length of protein). Scatterplot graph with linear trend-line generated in Excel program determined positive relationship between the degree of disorderliness and number of MoRFs. **(C)** Two sets of kinases were created: Set-A) Kinases that interact with other kinases; Set-B) kinases that do not interact with other kinases. Disease enrichment analysis was performed on the two sets using IPA and the top 5 disease profile is shown. Disease enrichment analysis of two sets of kinases reveals Cancer to be most enriched by kinases that interact with other kinases. On the other hand, kinases that interact exclusively with non-kinase proteins do not significantly participate in pathogenesis of cancer. –log (P-value) represents the number of molecules in a given pathway that meet cut criteria, divided by the total number of molecules that belong to the disease pathway. Dotted orange line represents the threshold value for enrichment for each disease and disorder

Figure 5: Topological Analysis reveals SRC and AKT as two of the most significant Kinome Modulators. (A) Workflow for identifying most significant Kinome Modulators. (i) 3 Sets were created. Top 10 interacting kinases from each kinase group's Set-X) intragroup interactions, Set-Y) intergroup interactions, and Set-Z) Topologically significant kinases. (ii) 40 kinases common to the three lists were used for the secondary topological analysis (iii) to identify 5 most significant hubs called "Kinome Modulators" (KMs) based on degree centrality. **(B)** The 5 most significant KMs were identified. The table shows the top 5 kinases and their degree centrality scores. **(C)** Identification of non-kinase proteins interacting with KMs. Interactome of the KMs (red) reveal non-kinase proteins (green) that interact with multiple Kinome Modulators. **(D)** List of non-kinase proteins against the number of Kinome Modulators each protein is interacting with. 2 Transcription factors (SMAD3 and STAT3) are highlighted.

Figure 6: Validation of regulatory interplay between two hub kinases, SRC and AKT, reveals previously unidentified SRC driven p-AKT mediated inhibition of SMAD activity. (A&B) Lung cancer cells respond better to dual inhibitions of SRC and AKT. Dual inhibition of SRC and AKT in (A)

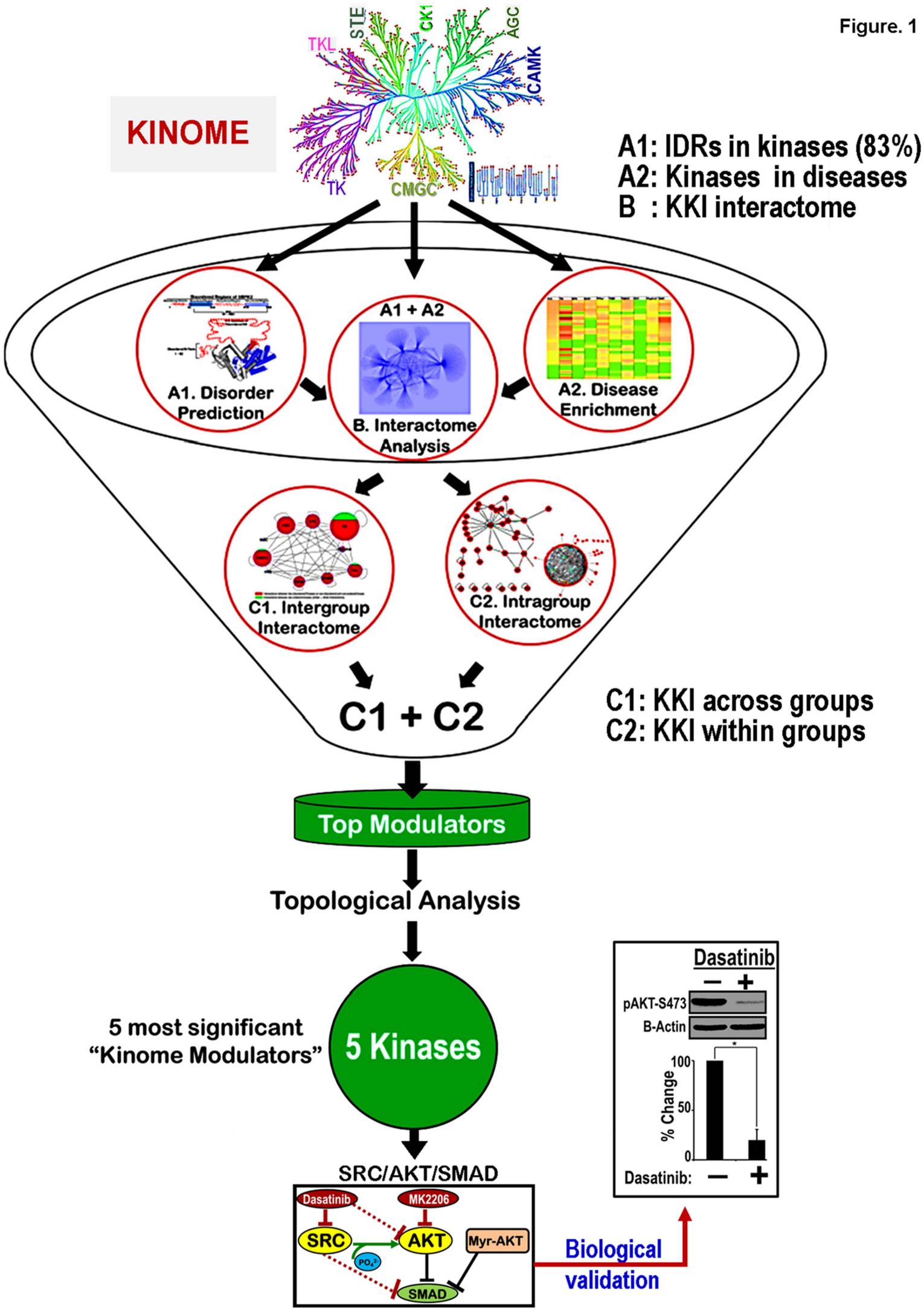
H292 cells or (B) H1650 cells with dasatinib and MK-2206 respectively shows more effective inhibition of proliferation as judged by percent increase in cell number. Error bars represent SE of n = 6 **(C)** SRC inhibits SMAD activity. Increasing amounts of CA-SRC (0µg, 2µg, and 4µg) significantly reduces SMAD activity as assessed by artificial SMAD reporter SBE4-Luc. Experiments were repeated twice in triplicate. Error bars represent SE of n = 3. **(D)** SRC affects phosphorylation of AKT. Inhibition of SRC with 500nM dasatinib in H292 cells or H1650 cells for 15 minutes significantly reduces p-AKT levels. Expression was normalized to β-Actin levels. Error bars represent SE of three independent experiments. **(E)** AKT inhibits SMAD activity. Luciferase activity shows decreased SMAD activity via artificial SMAD reporter SBE4-Luc in response to increasing amount of myr-AKT (0µg, 2µg, and 4µg) after transfection in H292 cells. Expression values were normalized to transfections without myr-AKT. Error bars represent SE of n = 6. (**A**-**E**: * = P-value < 0.05, ** = p-values < 0.001).

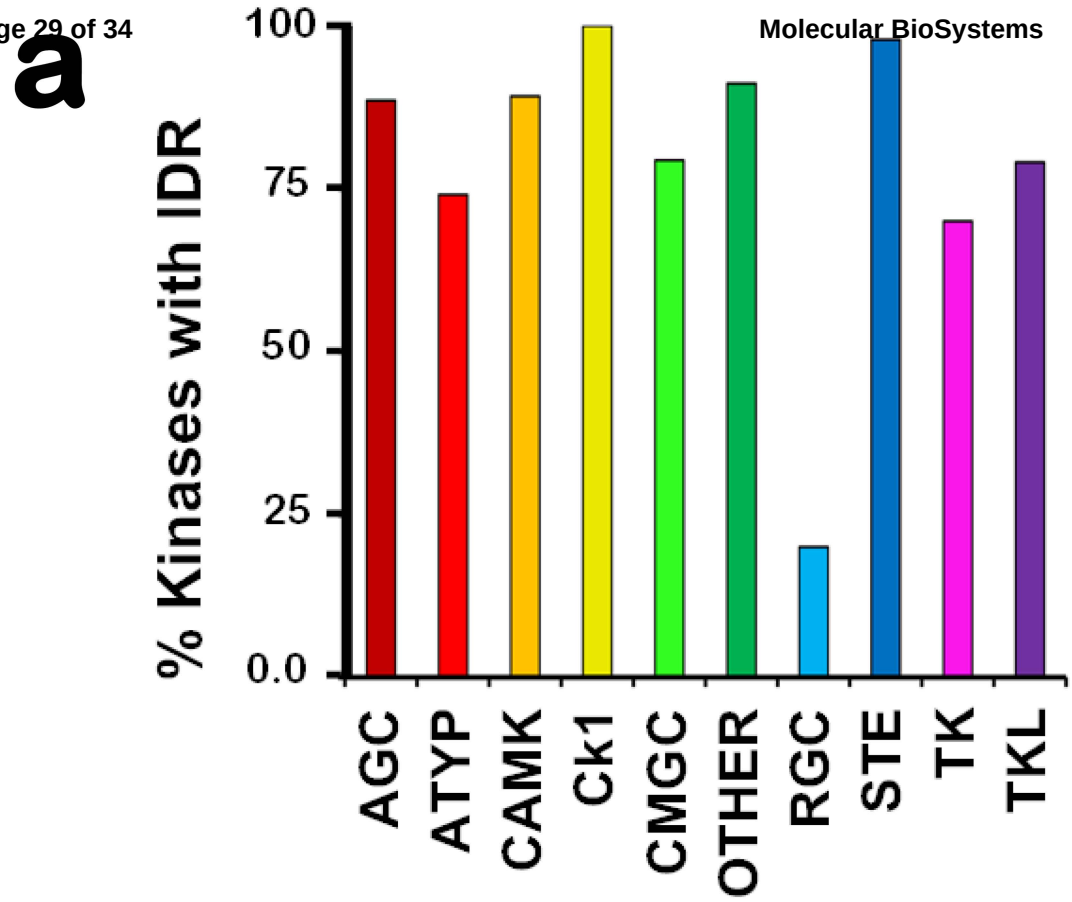
Figure 7: SRC inhibits SMAD activity via AKT modulation. (A) SRC inhibits activated SMAD via p-AKT. Western blots show increased p-AKT and reduced p-SMAD2/3 in response to transient transfection of H292 cells with increasing amounts of CA-SRC. Experiments were performed in duplicates. **(B)** Proposed model shows SRC affecting p-AKT levels and the downstream proliferative mechanism via SMAD-3 transcription factor.

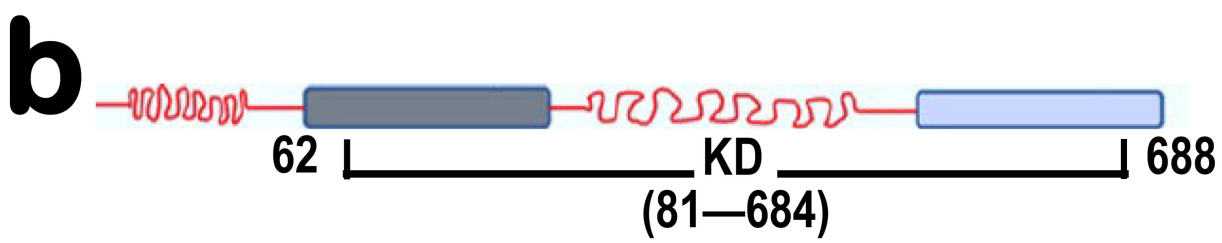
FIGURE LEGEND FOR GRAPHICAL ABSTRACT

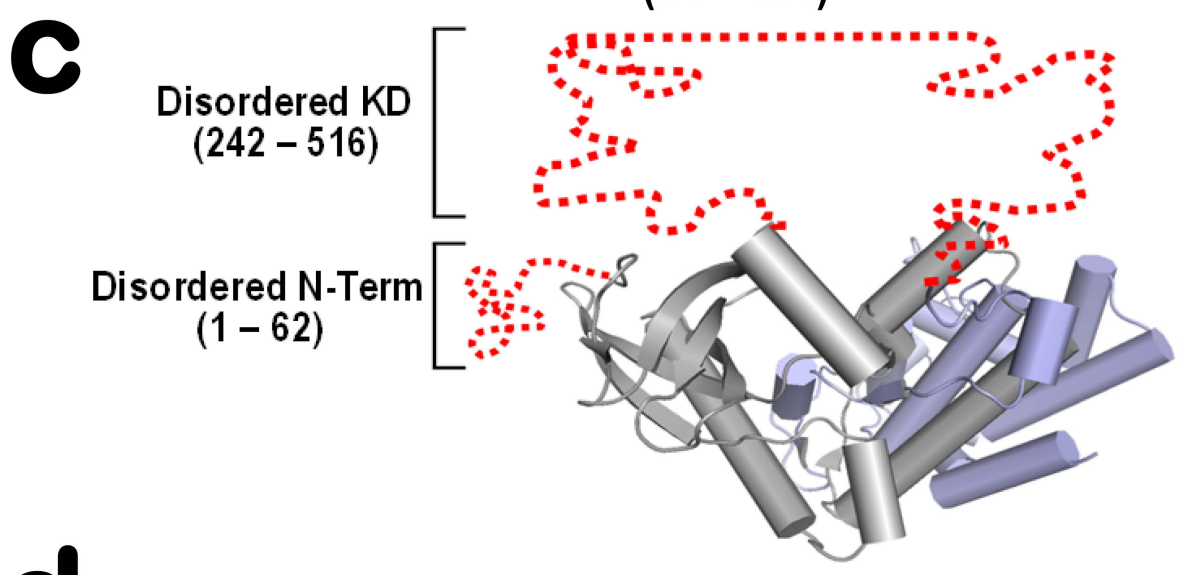
Table of Content Entry: We reveal presence of intrinsically disordered regions in human kinome and build a kinase-kinase interaction network identifying novel SRC-SMAD relationship.

Figure. 1









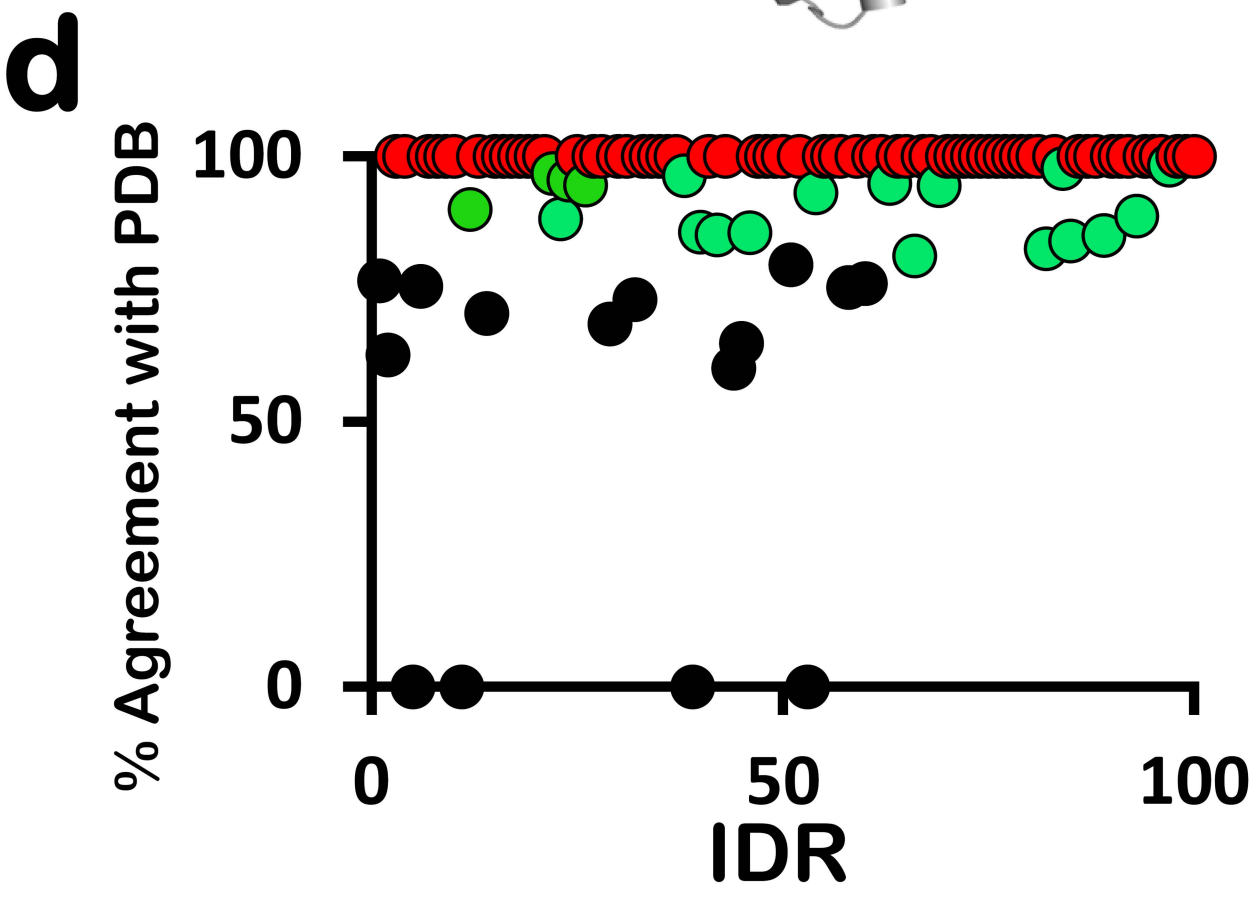
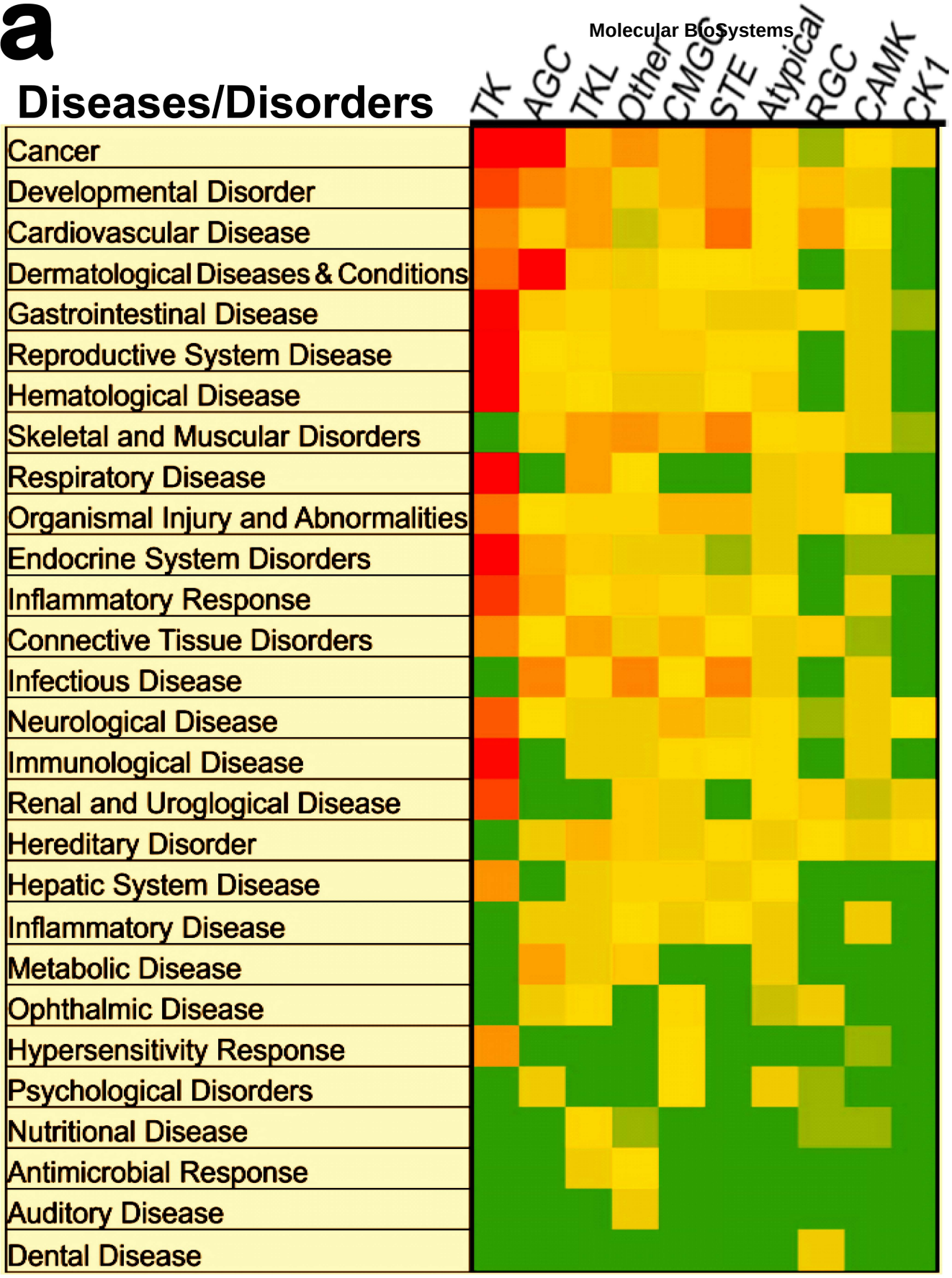
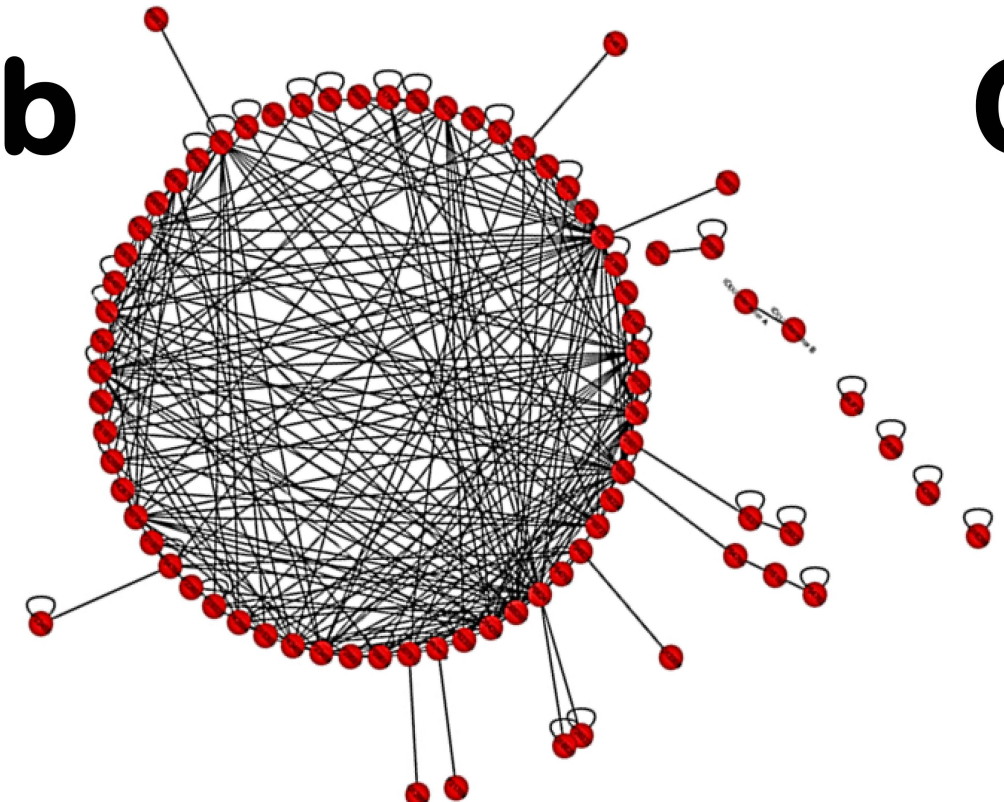


Figure 3

a



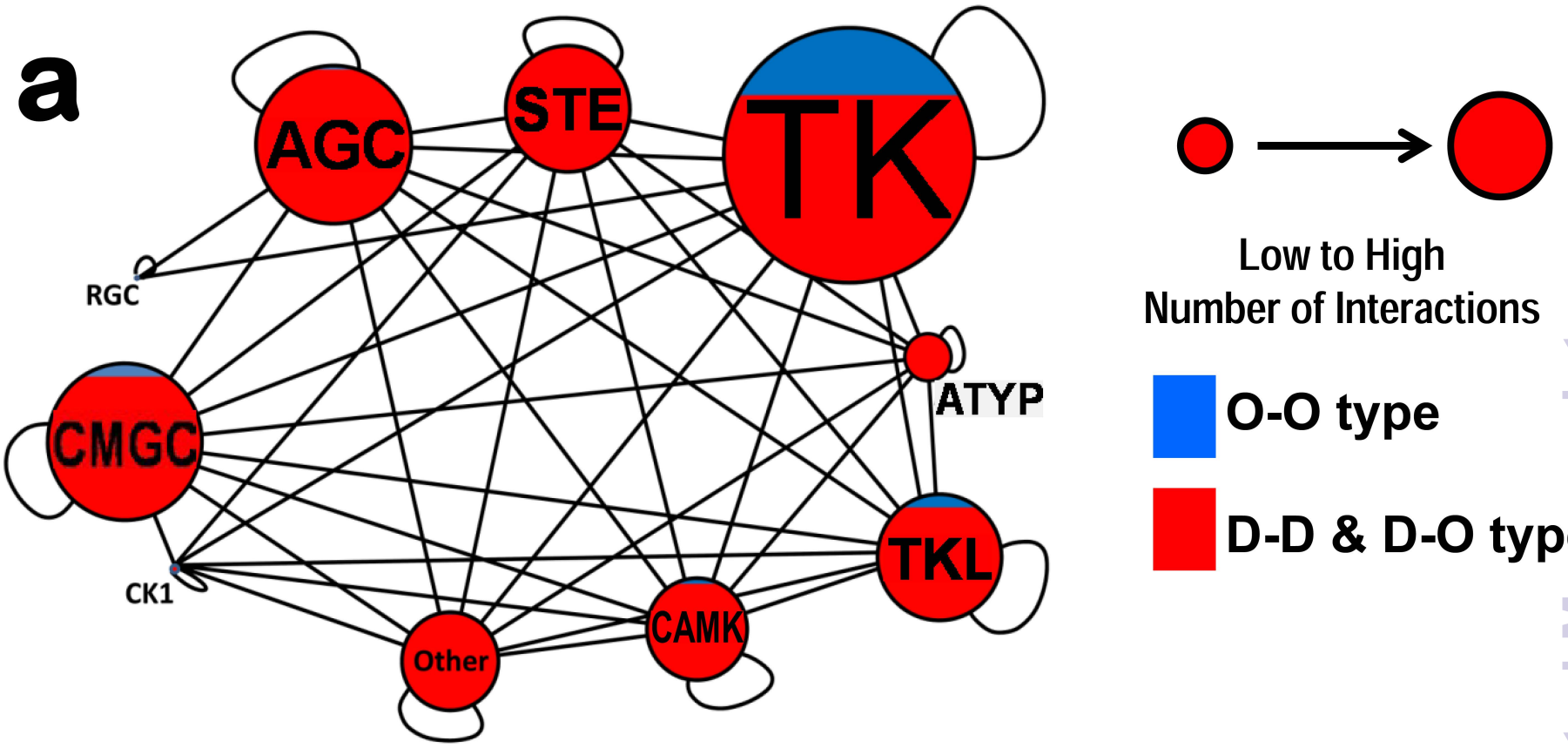
b

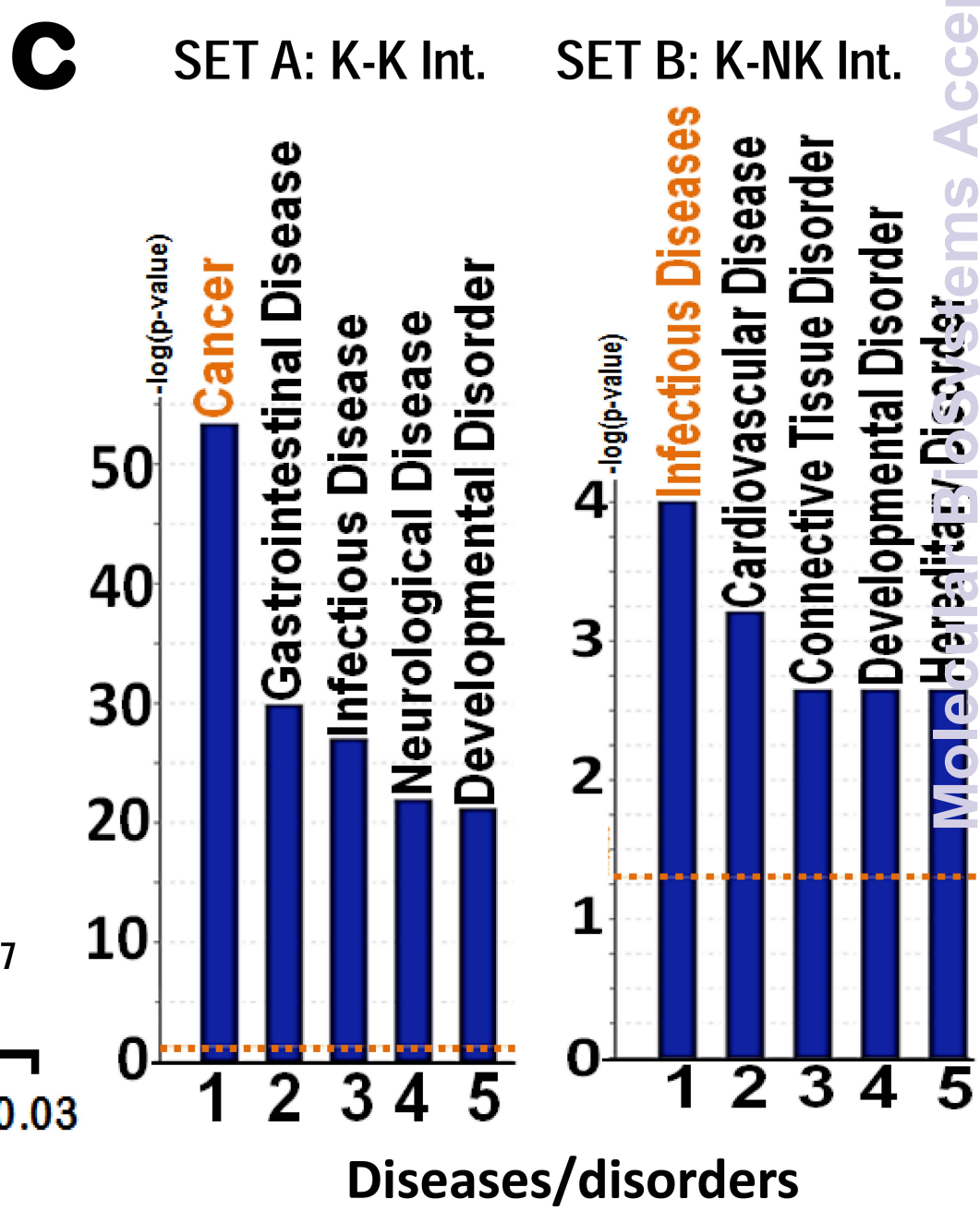


C

Groups	Interactions	Kinases
TK	292	78
AGC	55	39
TKL	55	34
Other	39	31
CMGC	51	41
STE	49	33
Atypical	8	10
CAMK	60	39
CK1	3	4

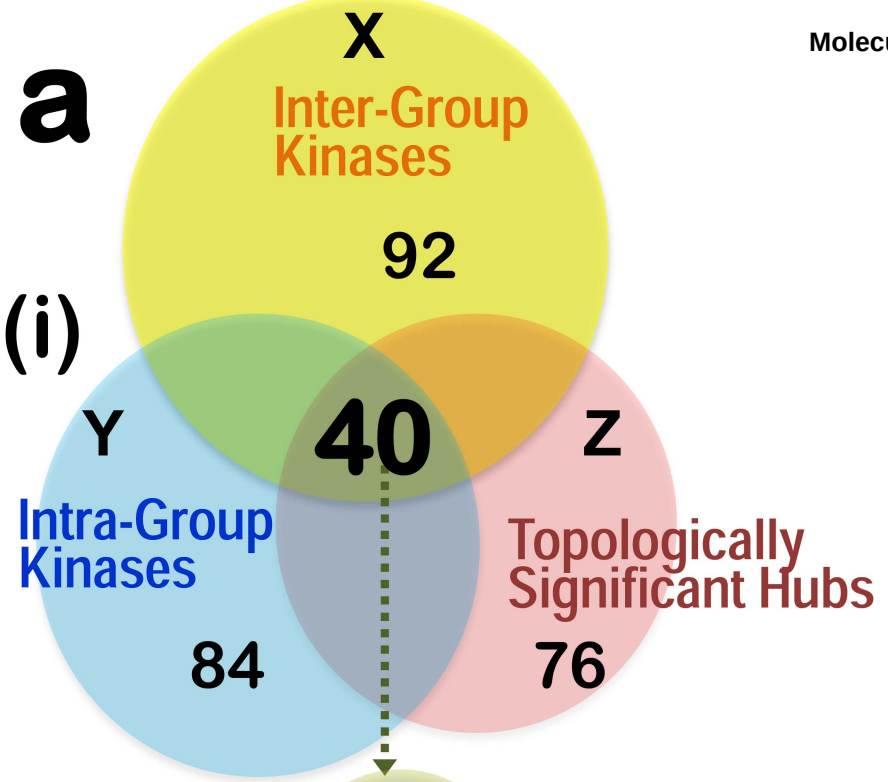
Figure. 4





a

(i)



(ii) 40 Common Kinases

Topological Analysis II

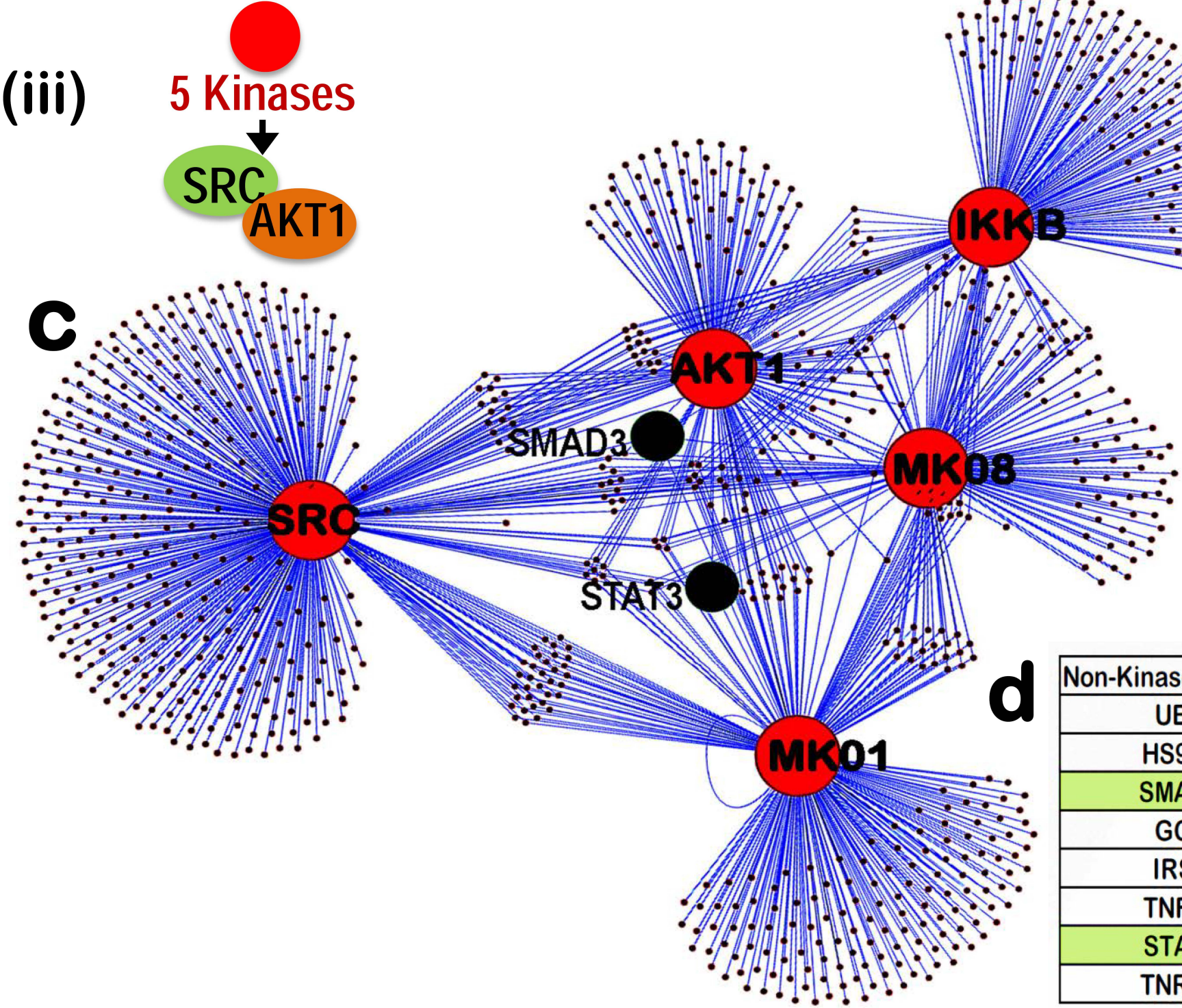
(iii) 5 Kinases

SRC AKT1

b

Kinase	Group	Degree Centrality	Interactions
SRC	TK	0.515	61
MK01	CMGC	0.424	42
AKT1	AGC	0.394	32
MK08	CMGC	0.394	22
IKKB	Other	0.394	27

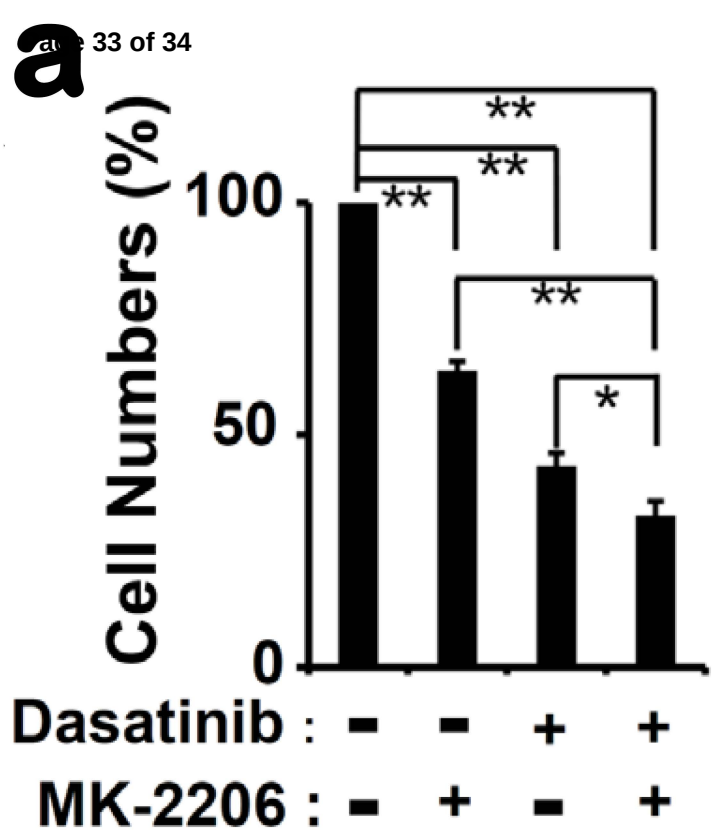
c

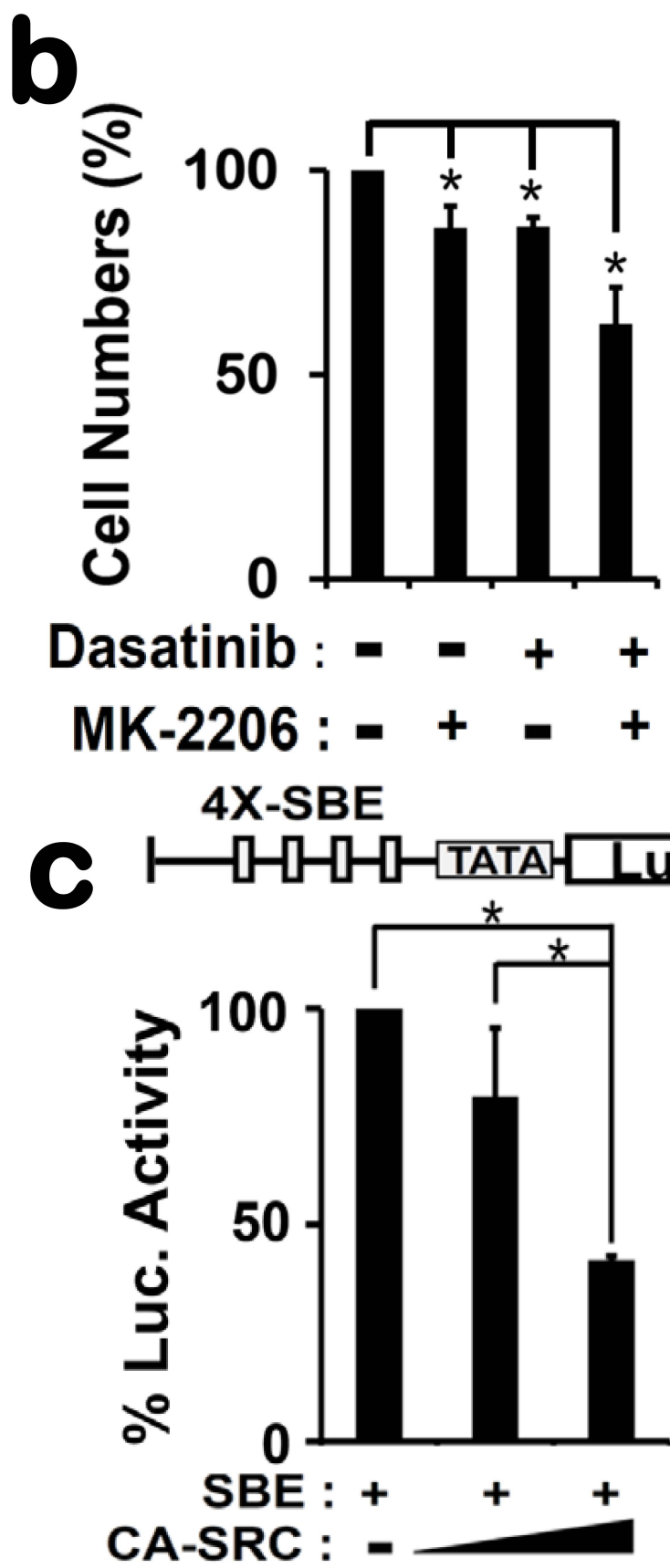


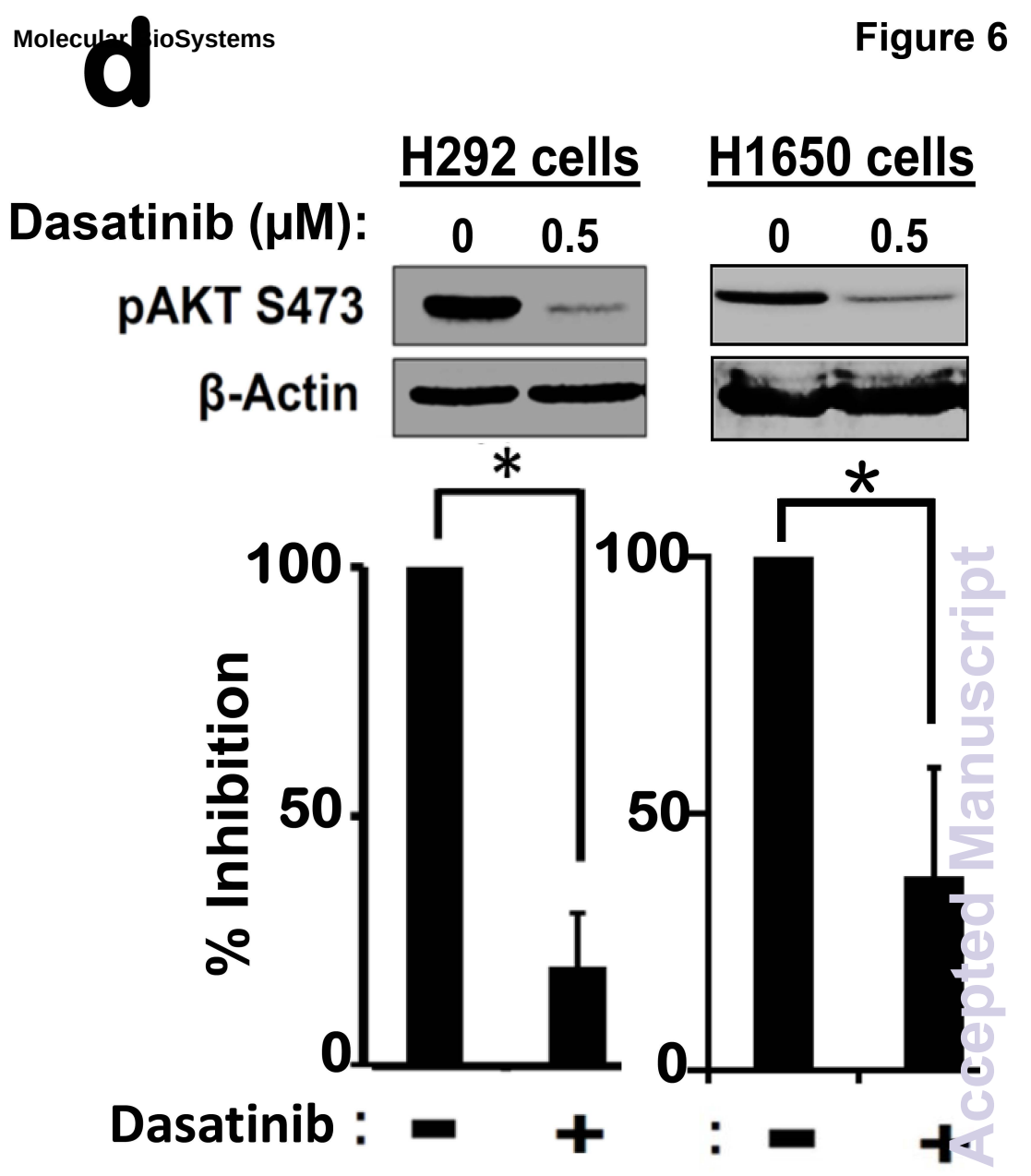
d

Non-Kinase Proteins
UBC
HS90A
SMAD3
GCR
IRS1
TNF11
STAT3
TNR1A

Figure 6







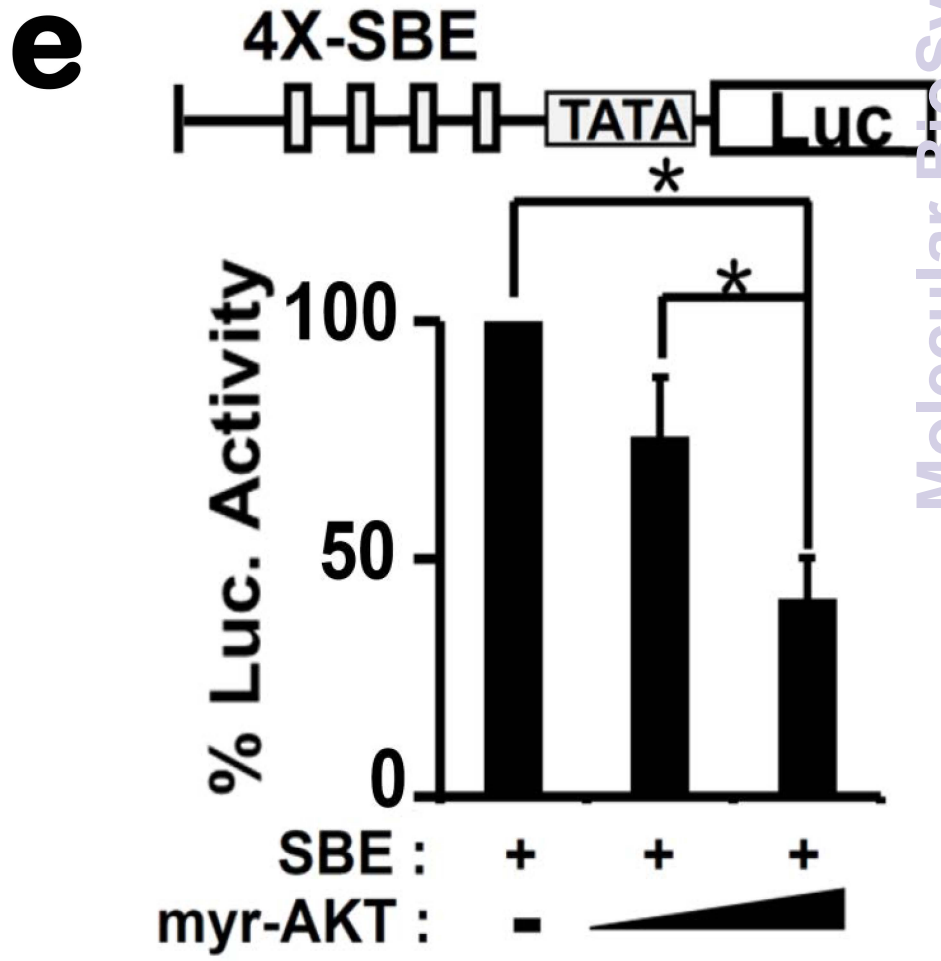
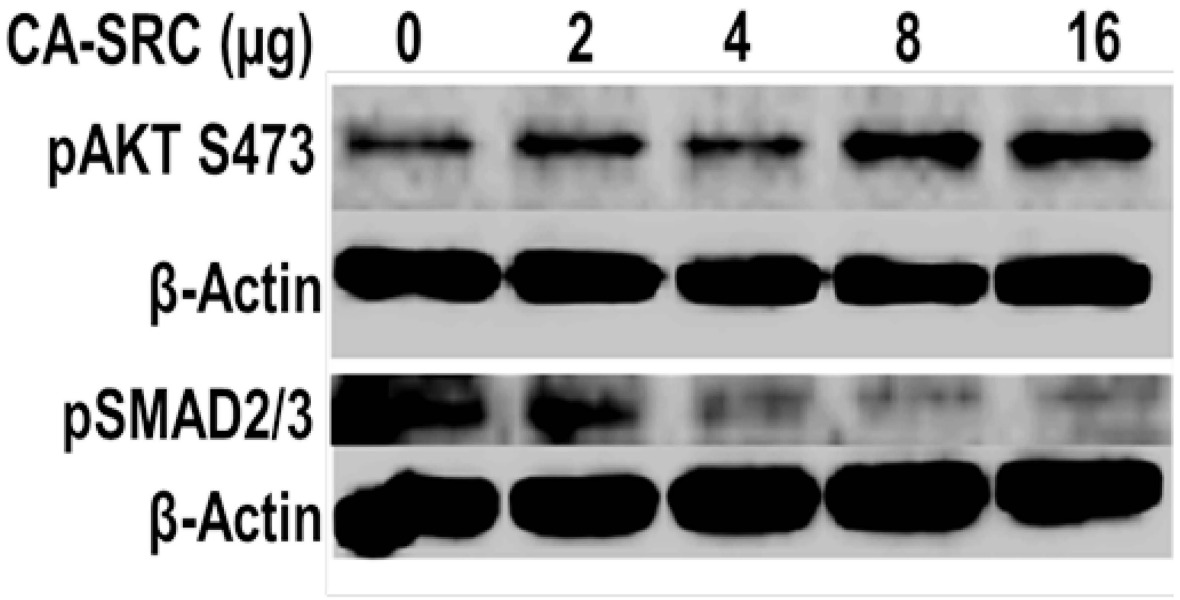


Figure 7

a



b