

# Building and exploring an integrated human kinase network: Global organization and medical entry points.

## Abstract

Biological matter is organized in functional networks of different natures among which kinase–substrate and protein–protein interactions play an important role.

Large public data collections allowed us to compile an important corpus of interaction data around human protein kinases.

One of the most interesting observations analyzing this network is that coherence in kinase functional activity relies on kinase substrate interactions primarily and not on which protein complexes are formed around them.

Further dissecting the two types of interactions at the level of kinase groups (CMGCs, Tyrosine kinases, etc.) we show a prevalence of intra-group interconnectivity, which we can naturally relate to current scenarios of evolution of biological networks.

Tracking publication dates we observe high correlation of kinase interaction research focus with general kinase research.

We find a similar bias in the targets of kinase inhibitors that feature high redundancy.

Finally, intersecting kinase inhibitor specificity with sets of kinases located at specific positions in the kinase network, we propose alternative options for future therapeutic strategies using these compounds.

## Biological significance.

Despite its importance for cellular regulation and the fact that protein kinases feature prominent targets of modern therapeutic approaches, the structure and logic of the global, integrated protein phosphorylation network have not been investigated intensively.

To focus on the regulatory skeleton of the phosphorylation network, we contemplated a network consisting of kinases, their substrates, and publicly available physical protein interactions.

Analysis of this network at multiple levels allowed establishing a series of interesting properties such as prevalence of kinase substrate interactions as opposed to general protein–protein interactions for establishing a holistic control over kinases activities.

Kinases controlling many or a few only other kinases, in addition to non-kinases, were distributed in cellular compartments differently. They were also targeted by kinase inhibitors with distinct success rates. Non-kinases tightly regulated by a large number of kinases were involved in biological processes both specific and shared with their regulators while being preferably localized in the nucleus.

Collectively, these observations may provide for a new perspective in the elaboration of pharmacological intervention strategies. We complemented our study of kinase interactions with a perspective of how this type of data is generated in comparison with general research about those enzymes.

Namely, what was the temporal evolution of the research community attention for interaction versus non-interaction-based kinase experiments.

## Introduction:

The organization of biological matter into functional networks with elements of modularity has been identified as key to warrant the accomplishment of a great variety of biochemical and cellular functions with a limited set of gene products.

In many instances, molecular networks have been very well studied, such as in intermediate metabolism, protein interactions and phosphorylation. In eukaryotes, and with the added layer of dedicated tyrosine phosphorylation, protein phosphorylation has evolved as a primary intracellular signaling strategy.

It bears many intrinsic specificities over other posttranslational modifications, including reversibility, energetic convenience and the ability of changing polarity of protein surfaces, leading to allosteric changes as well as governing protein interaction.

As phosphorylation has many biochemical and cellular changes both as input and as outcome, it induces a network that intersects with a plethora of biological processes.

Continuous efforts and methodological developments have greatly augmented our knowledge of how human kinases interact with other proteins or small molecules. More than 11,000 protein–protein physical interactions (PPIs) involving at least one kinase can be retrieved from public databases that compile the individual efforts of a whole community.

Different technologies were used to unravel these data with a strong contribution of affinity purification-mass spectrometry (AP-MS) and the yeast 2-hybrid (Y2H) system.

Known kinase interactions were obtained thanks to a multitude of publications, which did not necessarily focus on kinases but nonetheless included some of those enzymes.

Recently, dedicated campaigns mapped human protein kinase interactors specifically and on a large-scale. For instance, following a tandem AP–MS approach and using a tetracycline-inducible strep-hemagglutinin tag [7], the protein complexes formed around 32 commonly expressed kinases, i.e. kinases found in most cell types, were mapped in HEK293 cells [11].

Another study employing the same experimental protocol charted the complexes involving CMGC kinases in an unprecedented whole kinase group-wide effort [12].

Since kinases act primarily by regulating other proteins through their enzymatic activity, to map kinase PPIs is not sufficient for understanding their functional relationships. PPIs inform us on their collaboration with other proteins to form protein complexes or molecular machines, whereas kinase–substrate interactions (KSIs) picture what is regulated by their action.

Human KSIs have been studied more sparsely than PPIs so far, although data exist from several databases and new large-scale efforts were undertaken.

CEASAR, a protein microarray-based strategy unraveled 3,656 KSIs for 289 protein kinases [13], bringing the number of known KSIs to more than 8500.

Furthermore, computer-inferred KSIs such as the 75,000+ KSIs predicted by NetworkKIN [14], exploiting experimental phospho-proteomic data, known kinase substrate specificity, and kinase physical and genetic interactions, might provide a useful complement after stringent filtering before more experimental KSI are measured.

The regulatory function of kinases has been observed to be deregulated in many diseases, in particular cancer. Accordingly, a large number of kinase inhibitors have been developed by drug discovery laboratories and pharmaceutical companies, which are mostly small molecules or antibodies.

To relate kinases with their PPI partners as well as their substrates calls for further annotating such a network with the perturbation entry points available for disease therapy or chemical biology.

Several large in vitro kinase screens have provided comprehensive and quantitative drug–protein interaction data (DPIs) for a lot of cancer kinase inhibitors.

In this category, 72 inhibitors were screened against 442 kinases by Ambit Biosciences and another set of 178 inhibitors against 300 kinases by the Peterson's group. Furthermore, specialized databases collect DPIs from a broad range of scientific reports, e.g. DrugBank (78 DPIs).

Finally, chemical proteomics has emerged as a very interesting, more physiologically correct alternative to in vitro screens, where immobilized compounds serve as bait in affinity purifications to identify kinase inhibitor protein targets in an unbiased and cell type-dependent manner. This methodology has been applied successfully to small molecules inhibiting protein kinase activity.

Altogether, the availability of kinase PPIs, DPIs, as well as KSIs in unprecedented, large numbers created a unique opportunity for assembling a kinase-centered network combining these three kinds of interactions and to perform a global study of kinases in their environment.

We hence collected and integrated data from the various sources mentioned above and computed such a network.

We started our analysis by examining how protein interaction information was correlated with classical kinase research. Investigating the global topology of PPI and KSI networks we could obtain new insights in how they differ as well as refine previous hypotheses regarding the existence of global kinase communication ways.

We finally investigated how existing kinase inhibitors actually cover different classes of kinases and how KSIs might help exploring new therapeutic approaches.

## Construction of the network:

The list of human kinases with kinase group assignment was downloaded from UniProt web site and was comprised of 508 kinases.

Assignment to kinases families (sub-groups) was obtained from kinase.com web site.

Binary PPIs were collected from several public repositories: IntAct, MINT, InnateDB, DIP, HPRD, MatrixDB.

Binary PPIs were generated following the spoke model in the case of AP-MS data. PPIs were also obtained from two repositories describing protein complexes: CORUM and NCI PID. We used the matrix model to obtain binary interactions in this case.

Experimentally determined KSIs were obtained from IntAct, PhosphoSitePlus, and Phospho.ELM databases along with Newman et al. supplementary material. Computational inferences of KSI were retrieved from NetworkKIN website.

DPIs were compiled from DrugBank (inhibitors of human protein kinases only), in vitro kinase screens, and several chemical proteomics publications.

All interaction datasets downloaded from public databases were obtained from the versions available on the 10th of January 2014.

Only human-human interactions derived from physical association methods were retained. UniProtKB/Swiss-Prot human protein sequences were downloaded on the same day and all protein accession codes mapped to Swiss-Prot and updated to the current primary accession code to avoid duplications that can typically account for 2–5% of the proteins when public databases are merged directly.

Management of the PPI downloads was operated by a database system implemented in-house using Postgres and Perl.

DPIs and KSIs were downloaded manually and further processed by dedicated Perl and R scripts. Computational KSIs were filtered by comparing with experimental data, see Results and discussion.

DPIs were also submitted to filtering. In vitro screen data were imposed a maximum cut-off of 1000 nM (Refs [15,27]) and 50% of remaining activity. Chemical proteomics data, except Ref, were taken as filtered in the original publications (Refs [20,23–25]).

We further extracted a list of strong interactors by first requiring more stringent thresholds (100 nM for Ref. [15]; 200 nM for Ref. [27]; 25% remaining activity for Ref. [16]; no addition for Refs [11,14–16]).

In addition, for each compound in each dataset separately, targets had to be either outliers with the strongest affinity or hits with an affinity not worse than 5 times the most potent target (References [15,16,27]) or 10 times the most potent target (References [20,23–25]).

Outliers were defined by the `boxplot.stats` function of R (default parameters). DPIs from DrugBank were not considered as strong in the absence of affinity measures but we added to the strong target list the primary targets reported in Davis et al. irrespective of the filters above.

## Random networks

We generated two types of random networks to assess the significance of different topological features.

A first operation consisted in selecting random nodes from our reference network reduced to PPIs only (Section 3.4). This selection was done to achieve similar node degree distribution and repeated 1000 times.

PPI degrees of all kinases in the reference network were extracted and the probability of selecting non-kinase nodes randomly was adjusted to obtain a set of nodes, same number as kinases in the reference network, having the same degree distribution.

If kinases with degree  $d$  represented  $x\%$  of the kinases, then the probability to draw a non-kinase node with degree  $d$  was set to  $x\%$ . A second type of random network was generated for evaluating the hierarchical nature of KSI-mediated kinase-kinase interactions

In this case, we did not generate random selections of subnetworks as above but instead generated 200 full networks whose nodes followed the original degree distribution. This was achieved using the `randomNodeGraph` function of the `graph` R package, which implements the algorithm of M.E.J. Newman.

Self-connected nodes were removed and a correction factor applied to compensate for the resulting slight reduction in the number of edges.

## Results-Reference network

We assembled a comprehensive reference interaction network by incrementally adding the three kinds of interactions we considered in this work (Fig. 1A), starting with all PPIs we retrieved from public sources;

We eliminated redundancy and PPIs of proteins with themselves that cannot be recognized as real or technical artefacts, e.g. in AP-MS where a bait is almost always detected in its pulldown.

A total of 104,964 PPIs linking 13,347 proteins was collected.

In the next step, we added KSIs considering experimental KSIs first. In case an experimental KSI overlapped a PPI, the interaction was considered KSI. We then further proceeded with computational KSIs (compKSIs) obtained from NetworkKIN, which makes available a large set of such predictions (75,275 compKSIs).

As many new experimental KSIs have been released since NetworkKIN inferences were computed, we decided to exploit experimental KSIs to filter computer predictions.

NetworkKIN compKSIs were associated with three scores (context, motif, and ranking scores) which were potentially correlated.

We therefore submitted the whole collection of 75,275 triples to principal component analysis (PCA), a classical statistical procedure to reduce data dimensionality in the presence of redundancy. T

The result (Fig. 1B) showed that the first two principal components described 97% of data variability (bottom left bar plot). That is, the 2-dimensional projection of the data captured information available in the three NetworkKIN scores almost completely.

We observed that motif and ranking scores were highly redundant (co-linear in the plot, Fig. 1B bottom right), meaning that the context score and e.g. the motif score were sufficient to describe confidence in compKSIs. Plotting the 943 experimental KSIs overlapping NetworkKIN predictions in red, we could clearly observe a rather localized, parallelogram-shaped area.

A simple filter could hence be designed: we imposed two independent thresholds to the context and motif scores such that 90% of the experimentally supported compKSIs were above each limit.

This filter retained 14,905 compKSIs out of the original 75,275. Altogether, we added 8,539 experimental KSIs, linking 405 kinases to 2,774 substrates, and compKSIs raised these numbers to 23,444 KSIs, linking 409 kinases to 3,941 substrates.

DPIs from distinct sources were filtered separately due to the diversity of the screening methods; see Materials and methods. This resulted in 8,046 DPIs representing clear affinities between 476 kinases and 205 compounds. The reference network we assembled (Fig. 1C) included 14,176 proteins among which 487 kinases.

## Results-Accumulation of knowledge in kinase interaction research

Collecting PPI data from multiple sources, we kept track for each of the PubMed ID (PMID) and the year of publication. In combination with a resource provided by the NCBI that links genes to publications (gene2pubmed, <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/>), we could investigate how the community focused its attention to kinases comparing dates of general publications with publications disclosing kinase PPIs from 1975 to 2013.

By summing the number of publications of each type for every kinase, we observed a strong correlation between general knowledge and knowledge about interaction partners (Fig. 2A).

This means that popular and well-studied kinases were predominantly selected for performing PPI experiments, network biologists aiming at further characterizing “well-known” kinases.

This was certainly due to the rather recent emergence of PPI mapping technology, which was naturally directed towards questions already identified as the most important ones instead of embarking in systematic exploration. Interestingly, this trend might change soon as unbiased, systematic studies are emerging.

We then wanted to know whether information about kinase PPIs was acquired following different modes such as very intensive early studies followed by no activity afterwards or constant but less intense attention. We therefore considered cumulative time courses counting how many general versus PPI-disclosing publications were released until a certain time point (Fig. 2B).

The end-points of those curves were the dots of Fig. 2A (full totals). We clustered the cumulative time courses using the K-means algorithm and found no distinct modes. For instance, K-means applied to the generation of 3 clusters showed that the only differences were in the total amount of information available.

The shapes of the cumulative time courses were highly similar otherwise. To apply K-means specifying more clusters did not change this result.

We hence concluded that not only the level of attention in kinase PPI research followed closely the one of general kinase research but this relationship remained rather constant over time. It will be interesting to see if further development of PPI mapping technologies will change this in the future.

We asked one further question to our data, which was whether a bias existed at a finer level. Given the prevalent role of kinases in cancer, we classified kinases as cancer kinases and non-cancer kinases.

Two different sources were used: the COSMIC cancer gene census, which identified 41 kinases; KEGG cancer pathways, which contained 72 kinases (25 shared with COSMIC census). The level of attention to cancer versus non-cancer kinases in general research as opposed to PPI mapping was measured by the total number of publications available for the respective kinase classes.

In Fig. 2C (first two columns), one can read that PPI kinase research was much less biased towards cancer than general kinase research.

In the case of PPIs, kinase interactions could be measured without choosing a kinase as bait. That is, information was acquired without a kinase being in the original research focus.

Consequently, we repeated our analysis only considering publications disclosing kinase interactions where the kinase was the bait and we found the same trend (Fig. 2C, third column).

We finally wanted to assess whether the trend by which kinase PPI data were highly correlated with general information about kinases was not biased by protein abundance. Namely, one could imagine that abundant kinases were easier to analyze and to discover. Therefore, the correlation observed would mostly reflect this practical difficulty.

For the purpose of this study, we defined as low abundant or seldom expressed the proteins found in 3 or less of the 11 cell line proteomes recently published by the Mann's group [44], where more than 10,000 proteins were identified in each proteome.

We indeed found a lower correlation ( $R^2 = 0.33$  instead of 0.60, Fig. 2A) for such proteins, definitely indicating the presence of a bias but the correlation remained clearly positive and significant. The cancer versus non-cancer kinase analysis (Fig. 2C, last column) stayed unchanged.

## Results- How do kinases connect to the human interactome

visually suggests that kinases (red and yellow) and their substrates (blue) tend to occupy more central positions. We wanted to investigate this global aspect more precisely and considered kinase PPI connectivity, i.e. how many PPIs a protein has with other proteins.

Other topological measures exist [45], for instance centrality or the clustering coefficient, but they are usually correlated and for the sake of simplicity it was more convenient to limit ourselves to one measure.

Compared to an average human protein, a kinase was more connected with other proteins (Fig. 2A) to participate in more protein complexes most likely.

It has been already observed that commonly expressed proteins tend to have more PPIs [46,47]. Here, we defined commonly expressed as detected in all of the 11 proteomes by Geiger et al. [44]. This definition certainly reflected both an important function required by most cells and an abundance level not too low.

We could repeat the observation of higher connectivity in our network and further remark that within the kinase family the same trend existed, i.e. common kinases participated in more PPIs.

Turning our attention to kinase substrates, we performed a similar analysis and found that they also tended to be more connected with other proteins through PPIs (Fig. 2B).

To ensure that this trend would not be caused by the two thirds of computationally derived KSIs we repeated the analysis for experimental KSIs only and observed the same bias. As expected, commonly expressed substrates had even higher connectivity.

No correlation was observed between kinase and their substrate PPI connectivity (Fig. 2C, adjusted  $R^2 = 3E-5$ ; same result for experimental KSI only).

## Results - A kinase–kinase network.

After having considered kinases within the whole human interactome, it was a natural subsequent question to ask our data whether kinases – as a whole – evolved their direct interactions to collaborate with each other (PPIs) and to operate internal control and regulation (KSIs).

It has been suggested that through PPIs kinases might establish some significantly dense sub-network in yeast [48] or in human [12] (restricted to CMGC kinase data in the latter case).

We hence tried to assess the existence of this phenomenon at the level of the whole kinase family and distinguishing for the first time PPIs and KSIs, which represent very different forms of interaction.

Starting with PPIs, we first adopted a global perspective by marking all the proteins in interaction with one kinase at least. This represents what the whole kinase family can interact with.

Among the latter proteins, we counted the numbers of kinases and non-kinases and computed the ratio of these two counts obtaining 0.08. This ratio was close to what was reported for CMGCs only ( $44/652 = 0.07$ ).

The number of non-kinase proteins in contact with at least one kinase is thus much larger than the number of kinases ( $1/0.08 = 12.5$  times). Nonetheless, 8% was much more than the 2.5% of human genes coding kinases. To properly evaluate whether this ratio was unexpectedly large required careful statistical modeling.

As a matter of fact, we saw that kinases had more PPIs than an average human protein, which could artificially bias the observed ratio. Selecting 487 proteins – the number of kinases in our network – randomly that followed the same connectivity distribution as kinases (Materials and methods) we confirmed higher inter-kinase PPI density (Fig. 3D top).

To gain deeper insight in the topology of the kinase–kinase PPI network, we repeated a similar analysis but at the level of individual kinases. It turned out that the average ratio kinase versus non-kinase PPIs was not increased (Fig. 3D bottom).

This revealed that PPI-mediated collaboration across kinases is not the rule although a certain number of kinases are sufficiently well-connected to other kinases via PPIs to reverse this bias when considering all the kinases as a single entity as we did above in our global analysis.

The 10 most PPI-connected kinases were SRC (32 PPIs with other kinases; 228 in total), EGFR (29;357), FYN (25;244), MAP3K7 (21;107), RAF1 (18;82), ERBB2 (18;180), ERBB3 (18;105), LYN (18;80), ABL1 (17;209), and IKBKB (16;102). Among this list LYN, RAF1, ERBB3, MAP3K7 and IKBKB had a very large number of PPIs with other kinases, typically more than 20% of their interactions.

. For comparison, the more PPI-connected kinases in general were EGFR, IKBKE (341 PPIs), FYN, SRC, ABL1, PLK1 (185), ERBB2, MAP3K3 (178), GSK3B (174), and PRPF4B (173).

Considering KSIs, we performed similar calculations as above and we found that the part of the proteome globally covered as substrates of kinases was significantly enriched in kinases. Reducing the data to the sole experimentally derived KSIs confirmed this result (same figure). At the individual kinase level, we also found significant overrepresentation of kinases as substrates of other kinases (Fig. 3E bottom; same result for experimental KSIs).

Intra-kinase control or regulation is thus distributed across all kinases and is not the fact of a few global controllers only, a clear difference with PPIs.

The kinase–kinase network combining PPIs and KSIs is depicted in Fig. 3F, where we notice that commonly expressed proteins again seemed to concentrate at the center and kinases of the same groups to be closer to each other.

The actual higher connectivity of commonly expressed kinases and substrates is reported in Fig. 3A and B. To evaluate whether inter-group connectivity preference truly exists, we decomposed the interactions (Fig. 3G and H).

Taking into account the group sizes, the total number of kinases in the network, and the number of interactions between two groups, or within a single group, we could identify stronger than expected by chance interconnections (hypergeometric test).

Most significant PPIs interconnections took place within single groups. In the case of KSIs, we obtained a lot of more significant interconnections across two groups although the ones within a group were among the strongest. As we already did, we confirmed this result considering experimental KSIs only (data not shown).



The observation of significant interaction among kinases through KSIs at the level of individual kinases (Fig. 3E bottom) suggested an evolution of KSIs between kinases to globally coordinate the activity of these enzymes.

We considered the kinase–kinase network spanned by KSIs only, which we name KSI kinase network, and observed that it had a scale-free, or small world, topology [45,49] (Fig. 4A, black dots) that is commonly observed in biological networks: a few highly-connected hubs and a majority of less connected nodes. Because kinases have a dual role in this network, being either substrates or enzymes, we wanted to explore the existence of a relationship between the network topology and a prevalent enzymatic or substrate role.

Selecting kinases that play the role of the enzyme towards other kinases in >80% of their KSIs, we recognized immediately important signaling cascades (Fig. 4B). We asked our compiled data whether the complete integration of the effect of all signaling and regulation pathways by kinases on themselves would preserve a global structure.

The observation that “as enzyme” KSIs and “as substrate” KSIs also induced a scale-free topology (Fig. 4A, red and blue dots) already indicated that a specific structure exists at the functional role level.

In order to further ascertain its existence, we compared how kinases that control many other kinases, and are themselves less controlled, connected with other kinases in the real KSI network and in artificial networks with correct topology but random enzymatic/ substrate roles.

Sorting the kinases from full controllers (always phosphorylating other kinases) to fully controlled (never phosphorylating other kinases) and counting how many times one kinase has kinase substrates that are less controllers than itself (Fig. 4C), we found a significant difference in the real data (Fig. 4D); see also Materials and methods.

Kinase controllers thus influence the rest of the kinase activities significantly. Kinases that had less than 5 substrates known (not only kinases) were excluded from the simulation to limit the impact of absent data, which left a KSI network of 357 kinases and 2698 interactions. Identical results were obtained using experimental KSIs only (data not shown).

## Results - Kinases and non-kinases occupying special positions in the phosphorylation network:

The demonstrated existence of a global organization with kinases through KSIs called for the selection and inspection of specific sets of kinases, defined by how they were positioned with respect to the rest of the network. For this purpose, we decided to further increase the requirement on data reliability by only considering experimental KSIs and to limit our study to kinases for which at least 10 such KSIs were available, i.e. a rather precise appreciation for their connectivity was possible, which left us with 168 kinases.

A first set we defined were kinases seldom or never regulated by other kinases (in >80% of their KSIs they play the role of the enzyme). We identified 143 such proteins and named this set controllers. Then, considering KSIs between kinases, we defined kinase controllers that had >80% of their KSIs towards other kinases (KSIs towards non-kinase substrates did not count for this second set).

We found 57 kinase controllers, which were all included in the controller set (Fig. 5) meaning that kinase controllers are not specialized in this function and hence connect the kinase network to the rest of the interactome.

A dual set to the kinase controllers were the 23 controlled kinases that play the role of a substrate in >80% of their KSIs with other kinases.

Some of them were also controllers having a large number of non-kinase substrates (Fig. 5). We further defined a set of 46 kinases that have a comparable number of KSIs as substrates or enzymes (enzymatic role in between 1/3 and 2/3 of their KSIs).

We named them kinase integrators to indicate a potential function integrating and propagating multiple signals within the kinase network.

As it was the case for the controlled kinases, part of the integrators turned out to be controllers as well (a majority in this case). Highly-controlled non-kinases were defined as the proteins with the largest number of distinct kinases phosphorylating them (top 5%, 122 proteins).

To contrast the different sets we submitted them to GO term [50] analysis. Controllers were localized predominantly in the cytosol, different membranes, Golgi, and also in the nucleus but clearly less than in the cytosol.

Kinase controllers and integrators, and controlled kinases were in membranes and the cytosol. Highly-controlled non-kinases were predominantly located in the cytosol though also present in the nucleus (reverse distribution compared to controllers); they were specifically associated with transcription factor complexes and nuclear euchromatin.

In terms of function, the four kinase sets associated with all types of kinase activities and signaling cascades with a nice decrease from controllers (strongest) down to controlled kinases (weakest).

Controllers and kinase controllers were also enriched in transmembrane receptors, growth factor binding. Highly-controlled non-kinases had a strong and specific association with RNA and DNA binding, and transcription. Immune response and innate immunity pathways were strongly represented in controllers and kinase integrators (TLR signaling).

Highly-controlled non-kinases were also involved in immune system processes. Apoptosis and cell proliferation very significantly included the latter non-kinases as well as the controller set.

Finally, complementing the preferential nuclear localization and nucleic acid binding specificity of the highly-controlled non-kinases, we found a strongly association with DNA damage signaling and DNA repair.

Compiling the genes annotated to be tumor suppressor or oncogenes in GSEA MSigDB [51], or listed in COSMIC cancer gene census [42] we found an intersection with 24 highlycontrolled genes ( $P < 4E-14$ , hypergeometric test).

## Results - Kinase inhibitors, the modulating medical entry points.

Given the implication of kinases in a large number of disorders [52], numerous kinase inhibitors have been developed to modulate these enzymes as curative agents.

Early and extremely successful examples such as imatinib (Gleevec®), aimed at treating chronic myeloid leukemia, were introduced as targeted therapies against a specific kinase (the fusion protein BCR-ABL in this example).

Later investigations unraveled the existence of several targets beyond BCR-ABL for this compound, even including a non-kinase (NQO2) [25], and more generally kinase inhibitors tend to be promiscuous compounds [15,20,25].

It is therefore difficult to provide a clear picture of which kinases are specifically and potently targeted by which molecules, separating many tool compounds, approved drugs, and molecules in clinical development or only disclosed in patent applications

For the purpose of this study we preferred to capture trends reflecting the interest of the drug development community and, to a certain extent, the difficulty in targeting certain kinases.

Our reference network contained compiled data from large in vitro screens and chemical proteomics that were filtered according to criteria indicating clear inhibition of a kinase by a compound.

Here, we needed to define reasonable, more stringent criteria to relate each compound to “strong” targets, i.e. targets for which potency compared to other targets of the same molecule was significantly stronger thus making the compound a likely selective inhibitor for these selected targets (Materials and methods).

We obtained a strong target list comprised of 218 kinases interacting with 170 compounds and forming 1261 DPIs. Reporting the number of times, a kinase is a strong target for distinct molecules (Fig. 6A) shows a very biased attention of drug development across kinase groups and specific enzymes. A similar figure was reported previously.

With respect to the four kinase sets we defined previously (Fig. 5) we observed numerous examples of kinases not targeted by any compound suggesting many unexplored therapeutic options for future inhibitor development.

Each had a comparable number of strongly targeted kinases but taking into account the multiplicity of compounds strongly targeting each kinases (heights of bars in Fig. 6A), we observed significant differences: controllers that were not kinase controllers were strongly enriched ( $P < 3.4E-6$ , Fisher exact) as well as kinase integrators ( $P < 1.3E-8$ ), most of which as controllers.

Controlled kinases were not found to be favored drug targets. We further broke down target biases to the kinase family level and identified PDGFR ( $P < 1.1E-15$ ), Trk ( $P < 3.2E-4$ ), NEK( $6.9E-4$ ), VEGFR ( $P < 1.3E-3$ ), Abl ( $P < 4.9E-3$ ), and EGFR ( $P < 5.7E-3$ ).

To illustrate how a kinase network such as the one we assembled might actually support the implementation of new paradigms in the application kinase inhibitors in cancer therapy, we drafted a strategy to circumvent aberrant behavior of mutated non-kinases.

The tumor suppressor adenomatous polyposis coli (APC) has been found to be involved in several tumors when mutated (gastric and colon cancers, medulloblastoma, hepatocellular carcinoma, etc.) [54,55].

As indicated in Fedorov et al. [53] APC is difficult to modulate through small molecules and signaling around this protein could be a preferable target for therapy. From our reference network we collected all PPIs relating APC to its functional context (Wnt and JNK signaling, cell migration) and retrieved 114 proteins, all non-kinases (Fig. 6B).

We subsequently found 85 kinases having experimental KSIs with at least one of these proteins (35/144 APC neighbors). The two kinases CDK2 and SRC had the largest number of substrates and covered together 18/144 of the APC partners.

One could thus propose to disrupt with maximum efficacy the environment of APC by composing a combined therapy with SRC and CDK2 inhibitors. Dual CDK1 and CDK2 inhibitors would also be an option to explore.

## Discussion and conclusions