Cell Reports
Resource

# The Protein Interaction Landscape of the Human CMGC Kinase Group

Markku Varjosalo,[1,5] Salla Keskitalo,[1,5] Audrey Van Drogen,[1] Helka Nurkkala,[4] Anton Vichalkovski,[1] Ruedi Aebersold,[1,2,3] and Matthias Gstaiger[1,2,*]
[1]Department of Biology, Institute of Molecular Systems Biology
[2]Competence Center for Systems Physiology and Metabolic Diseases
ETH Zurich, 8093 Zurich, Switzerland
[3]Faculty of Science, University of Zurich, 8006 Zurich, Switzerland
[4]Institute of Biotechnology, University of Helsinki, 00014 UH, Finland
[5]Present address: Institute of Biotechnology, University of Helsinki, 00014 UH, Finland
*Correspondence: gstaiger@imsb.biol.ethz.ch
http://dx.doi.org/10.1016/j.celrep.2013.03.027

## SUMMARY

**Cellular information processing via reversible protein phosphorylation requires tight control of the localization, activity, and substrate specificity of protein kinases, which to a large extent is accomplished by complex formation with other proteins. Despite their critical role in cellular regulation and pathogenesis, protein interaction information is available for only a subset of the 518 human protein kinases. Here we present a global proteomic analysis of complexes of the human CMGC kinase group. In addition to subgroup-specific functional enrichment and modularity, the identified 652 high-confidence kinase-protein interactions provide a specific biochemical context for many poorly studied CMGC kinases. Furthermore, the analysis revealed a kinase-kinase subnetwork and candidate substrates for CMGC kinases. Finally, the presented interaction proteome uncovered a large set of interactions with proteins genetically linked to a range of human diseases, including cancer, suggesting additional routes for analyzing the role of CMGC kinases in controlling human disease pathways.**

## INTRODUCTION

Reversible phosphorylation of proteins is a key mechanism to regulate the activity of enzymes and control the localization and stability of proteins and their interactions with other proteins. These events in turn control information processing by cellular signal pathways to regulate metabolism, cell division, apoptosis, and more. Protein phosphorylation is catalyzed by 518 members of the human protein kinase family, representing one of the largest protein families in human cells (Manning et al., 2002). Deregulation of human kinases has been causally linked to a variety of human malignancies (Lahiry et al., 2010; Manning, 2009), including cancer, and kinases represent 20% of all putative drug targets (Futreal et al., 2004; Hopkins and Groom, 2002). Understanding the mechanisms that control kinase activity and substrate specificity is therefore of central interest to biomedical research. Studies using classical biochemical approaches have shown that kinases exert their function in the context of kinase protein complexes formed by specific associations with a range of other proteins, including regulatory subunits, kinase inhibitors, scaffold proteins, and proteins that target the complex to specific subcellular sites, as well as substrates. In spite of their eminent biological importance and in contrast to the situation in yeast, where several large-scale affinity purification and mass spectrometry (AP-MS) studies have provided insight into how the yeast kinome is connected to other proteins (Breitkreutz et al., 2010; Gavin et al., 2006), our general knowledge of human protein kinase complexes is very limited. This is due to the focus of small-scale studies on kinases that have been genetically linked to essential processes or diseases, and the lack of global studies on human kinase complexes (Edwards et al., 2011).

Investigators have used an array of techniques to determine protein-protein interactions (PPIs) and define protein complexes. Currently, however, AP-MS is the only method that is capable of isolating and identifying protein complexes from human cells under near-physiological conditions (Gingras et al., 2007). Recent advances in AP protocols, MS instrumentation, and computational tools resulted in the identification of high-confidence interaction (HCI) proteomes of different human protein groups (Behrends et al., 2010; Breitkreutz et al., 2010; Choi et al., 2011; Glatter et al., 2009; Sardiu et al., 2008; Sowa et al., 2009).

Here we present a systematic protein interaction study of the human CMGC (cyclin-dependent kinase [CDK], mitogen-activated protein kinase [MAPK], glycogen synthase kinase [GSK3], CDC-like kinase [CLK]) group of protein kinases. This evolutionarily conserved group consists of 62 members (http://uniprot.org), which are assigned to nine families. The CDKs and MAPKs are the two largest and best-studied CMGC groups. Besides control of the cell cycle, CDKs are known to control the activity of human tumor suppressors and thus have been of prime interest in molecular cancer research. The MAPKs regulate a variety of cellular processes and participate extensively in the control of cell-fate decisions across all eukaryotic phyla. Other interesting CMGC families include the dual-specificity tyrosine-regulated kinases (DYRK) and the serine-arginine

protein kinases (SRPK); however, these kinases are much less studied and only sparse information on how they engage in interactions with other proteins is available in public databases.

In this work, we systematically determined the protein interactions of all of the 57 CMGC kinases expressed in human tissues (http://genesapiens.org). CMGC kinase group members frequently have related interaction profiles that are enriched for specific Gene Ontology (GO) groups that link the CMGC kinase families to various biochemical processes in transcription, RNA processing, cellular communication, the regulation of the cell cycle, and more. The most prominent protein group among CMGC-binding proteins is the kinases themselves, suggesting that the human kinome as a whole forms large regulatory networks. Here we present a combined computational and experimental approach for identifying kinase-substrate candidates. Finally, the identified biochemical context points to a multitude of molecular links between CMGC kinases and proteins associated with human diseases. We note that several groups of proteins that are linked to similar disease phenotypes tend to cluster around specific CMGC kinase complexes. We found an overrepresentation of cancer-associated proteins (CAPs) in CMGC kinase complexes. Therefore, this work represents an important resource to direct future studies aimed at identifying approaches to interfere with human pathologies linked to perturbed CMGC interaction proteomes.

## RESULTS AND DISCUSSION

### Systematic Analysis of Human CMGC Kinome Complexes

Despite their central role in the control of biological processes and diseases, human protein kinases have not previously been subjected to a systematic analysis. Here, we applied a recently developed, integrated experimental and computational AP-MS method to characterize the human CMGC kinase interactome (Breitkreutz et al., 2010; Choi et al., 2011; Glatter et al., 2009). The method is schematically illustrated in Figures 1A–1C. We generated isogenic HEK293 cell lines for each of the 57 studied kinases, and the kinase in question was inducibly expressed with a Twin-Strep-tag and hemagglutinin (SH) tag (Table S1) to a level that approximately matched that of the intrinsic kinase (Glatter et al., 2009; Figure S1). Kinase complexes were purified by double-AP and tryptic peptide samples were subjected to liquid chromatography-MS (LC-MS) in technical replicates (overall reproducibility > 86%).

The double-AP yielded less background contaminants but had similar sensitivity compared with the single Strep purification (Figures S1C and S1D). We used the generated LC-MS data to identify the sample proteins, which we then further classified as high-confidence interacting proteins (HCIPs) or nonspecific interactors using the recently developed SAINT algorithm (Choi et al., 2011; Figures 1A–1C). Overall, the AP-MS analysis of the 57 CMGC group members resulted in a high-confidence network of 481 proteins and 652 kinase-protein interactions (for details on filtering, see Tables S1 and S2).

We applied two independent methods—coimmunoprecipitation (coIP) in HEK293T cells and bimolecular fluorescence complementation (BiFC) assays in HeLa cells—to validate the specificity of the obtained CMGC interactome data set (Figures S2 and S3). Twenty-three tested interactions were validated with either coIP or BiFC, and 19 of 23 (83%) were validated with both of the tested methods, which corresponds well to the validation rates reported recently in other HCIP studies using AP-MS (Behrends et al., 2010; Sowa et al., 2009).

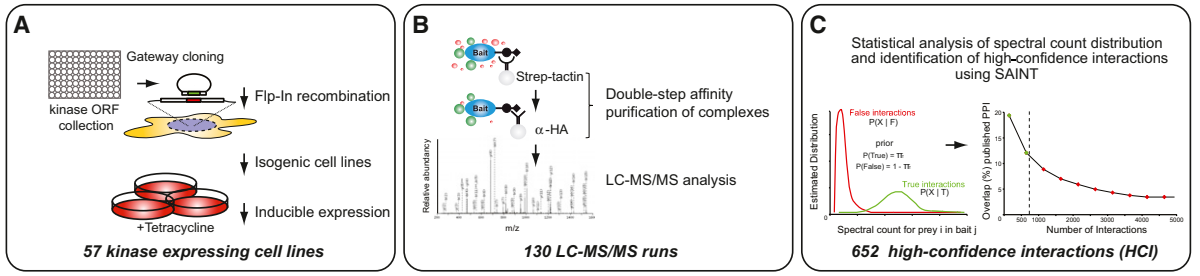
On average, CMGC kinases undergo complex formation with 11 HCIPs (Figure 1D). This number is similar to what has been reported for other human baits in large-scale AP-MS studies (Behrends et al., 2010; Sowa et al., 2009). The number of HCIs per bait, however, varies significantly across the different CMGC families. SRPKs showed more than three times the number of binding partners per bait compared with other CMGC kinases, whereas CDKL and RCK kinases revealed only a few binding partners (Figure 1E). The largest CMGC family, CDK, interacts with almost half of the CMGC interaction network components (219), but the average number of interactions per bait in the CDK family does not deviate from the average.

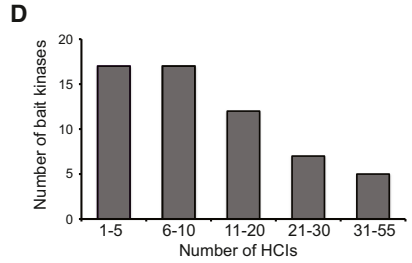
The functional diversity of the CMGC kinase-binding proteins is illustrated in a network graph in Figure 1F, where interacting proteins are grouped based on their GO classifications (simplified GO Biological Processes [GO-BP]; Keshava Prasad et al., 2009). Since the CMGC group includes kinases with established roles in cell-cycle regulation and signaling, it is not surprising that many CMGC HCIPs have functions in cell communication (80 HCIPs), cell growth (24), and cell cycle (19). An unexpectedly large number of HCIPs are involved in transcription and RNA-related processing (147 [31% of all the HCIPs]; Figure 1F). Subsequent hierarchical clustering based on GO-BP (Figure 1G) and GO Molecular Functions (GO-MF; Figure S4) revealed that CDK family members, as expected, preferentially bind to proteins associated with cell-cycle regulation. As many as 51% of CMGC kinases have a minimum of three HCIPs associated with cell communication or signal transduction. The other large ontology group in the CMGC interactome is "regulation of nucleic acid metabolism," which covered 53% of HCIPs linked mainly to complexes of SRPKs and CLKs. The GO-BP analysis links many poorly studied kinases to distinct cellular functions. For example, proteins that bind to DYRK1A and DYRK1B have roles in the regulation of cell growth, a function that was not previously reported for these kinases.

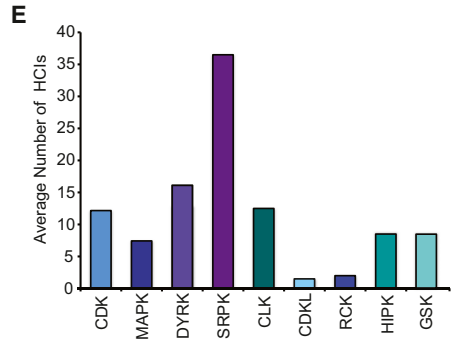
### Comparison with Previously Identified CMGC Kinase Interactions, and the Modular Topology of the CMGC Interactome

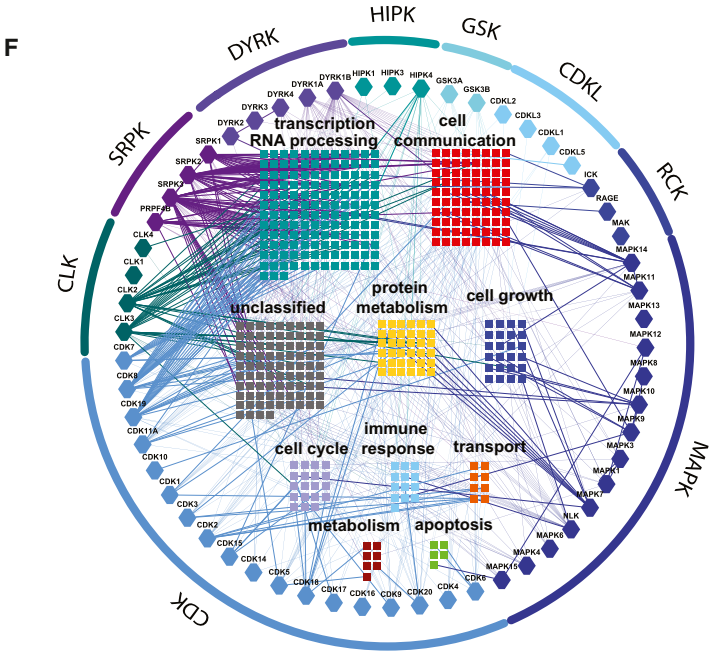
Overall, we identified 531 protein interactions for CMGC kinases and 121 interactions already annotated in public interaction databases (Cowley et al., 2012; Turner et al., 2010). We also estimated the fraction of public interactions not covered by our study. However, an objective evaluation of PPI data coverage is complicated by the heterogeneous source and the unknown false-positive rate of public PPI data. To address these issues, we used public data annotated in the PINA database (Cowley et al., 2012), which allowed us to filter the public interactome for the number of independent reports supporting a given interaction as a proxy for data robustness. When we considered all physical PPIs (including yeast two-hybrid data) supported by at least two independent reports, our study covered 16% of

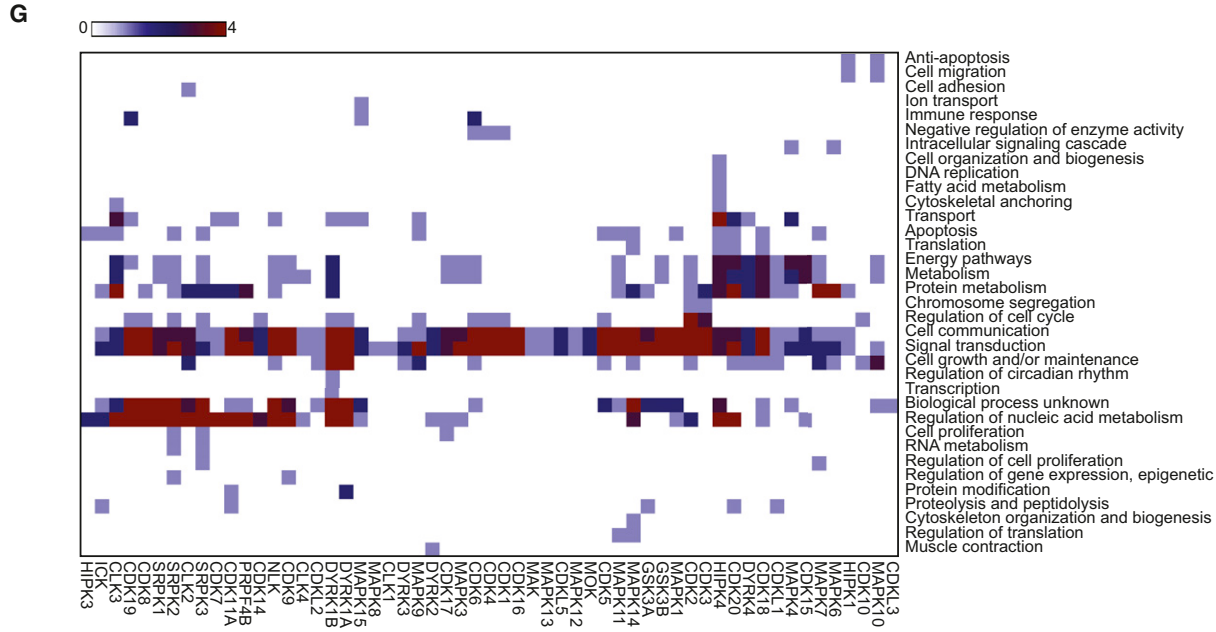
Cell PRESS
Open ACCESS




(legend on next page)

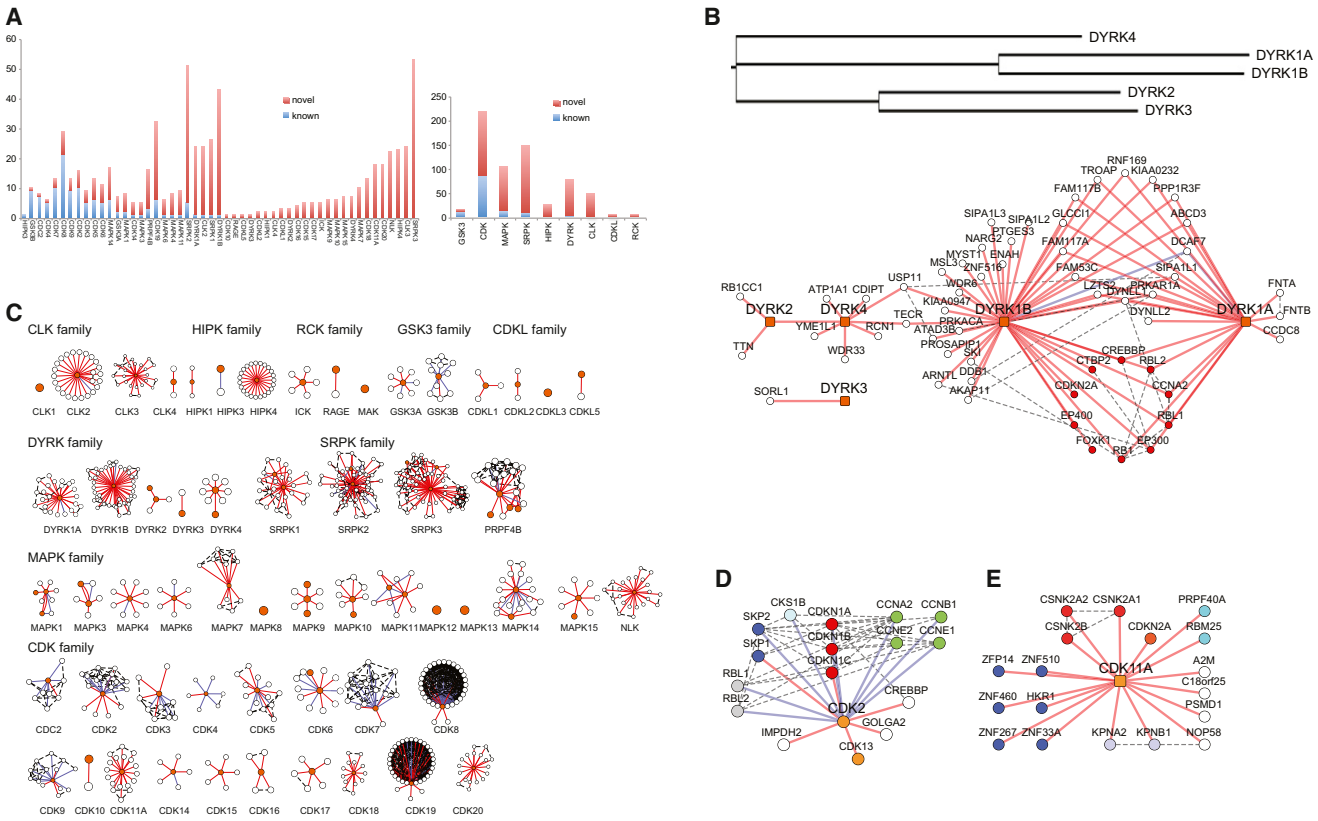



**Figure 2. Relationship of the CMGC AP-MS Data to Previously Published PPI Information**

(A) Overview of HCIs identified in CMGC kinases. Interactions identified in this study are shown in blue, and red indicates the PPIs reported in public databases.

(B) Comparison of DYRK interactomes. Known interactions between the HCIPs are shown with a dotted line. Sequence similarity between DYRK family members is illustrated as a dendrogram (upper panel).

(C) Overview of HCIs of individual CMGC kinases grouped with their corresponding families.

(D) A subnetwork of the CDK2 interactome illustrates the interactions with cyclins (green), CDK inhibitors (red), regulatory subunit (light blue), components of the SCF (blue), and Rb family members (gray).

(E) CDK11A interacts with CKII subunits (red), RNA-binding proteins (turquoise), zinc finger proteins (blue), and importin subunits (lilac).

See also Figures S2 and S7, and Tables S2 and S7.

public PPI data. However, the coverage moved up to 49% when we considered only robust PPIs that are supported by more than six publications (for details see Table S2). We also found that the public interactions that overlapped with our study were published on average 10.7 times, in contrast to the set of public interactions not found in our study, which were referenced only 2.9 times on average. This illustrates the overall robustness of the presented PPI data.

The overlap between our data and already reported interactions varies significantly across the different CMGC members (Figure 2A). This can be explained by the fact that various families have been studied to different extents and by different methods in the past.

The DYRKs are a particularly rich source for interaction information. The DYRK subnetwork is composed of 60 proteins and 78 interactions, of which only two could be found in public databases (Figure 2B). Most of the interactions we found constitute complexes containing the highly related class I family members DYRK1A and DYRK1B. DYRK1A and DYRK1B complexes have 20 proteins in common, and these proteins are not

**Figure 1. Proteomic Analysis of CMGC Kinase Complexes**

(A–C) Schematic overview of (A) the generation of the 57 isogenic HEK293 cell lines for inducible expression of SH-tagged CMGC bait kinases, (B) double-AP strategy and LC-MS/MS analysis of the purified kinase complexes, and (C) identification of HCIs from AP-MS data using statistical filtering.

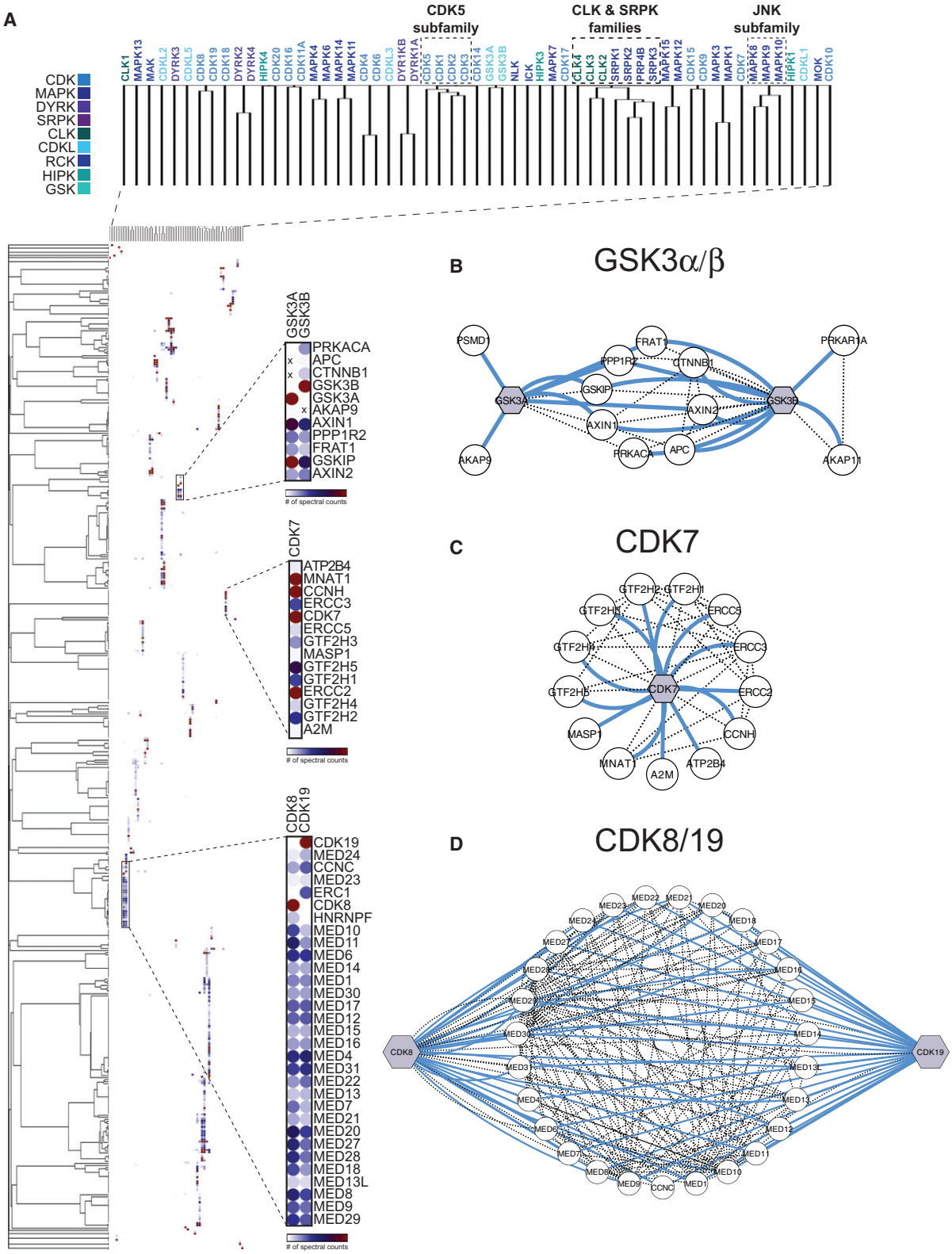
(D) Distribution of the number of HCIs identified across CMGC kinases (baits).

(E) Average number of HCIs for each CMGC kinase family.

(F) Network model of the CMGC kinase interaction landscape. The CMGC kinase families are organized based on their sequence similarity and the HCIPs are grouped according to their simplified GO process annotation.

(G) Hierarchical clustering of CMGC kinases and GO-BP terms associated with the corresponding HCIPs. The color gradient illustrates the number of HCIPs assigned to each particular GO term.

See also Figure S1 and Table S1.




*(legend on next page)*

found in complexes with the class II family members DYRK2, DYRK3, and DYRK4. For example, DYRK1A and DYRK1B complexes both contain LZTS2, a putative tumor suppressor that suppresses Wnt signaling by promoting nuclear exclusion of CTNNB1 (Thyssen et al., 2006). We also found interesting differences between DYRK1A and DYRK1B. For example, DYRK1B complexes contain MYST1 and MSL3, two proteins from the human MSL complex that are involved in chromatin remodeling, and ARNTL, a bHLH-PAS transcription factor that is essential for the control of circadian rhythm (Bunger et al., 2000). Furthermore, DYRK1B kinase complexes contain the oncogene c-SKI, which binds the SMAD complex and suppresses the transforming growth factor β (TGFβ) target gene promoter (Akiyoshi et al., 1999; Javelaud et al., 2011). In contrast, DYRK1A was found in complexes containing the farnesyltransferases FNTA and FNTB. The interactomes of the class II members DYRK2, DYRK3, and DYRK4 are quite different from each other. They form fewer interactions and only two proteins (TECR and USP11) are shared with class I DYRK complexes. The observed differences are consistent with the idea that class I and class II DYRKs diverged early in evolution (Aranda et al., 2011) and have undergone functional diversification by acquiring new protein interactions.

The public interaction data indicated that a number of kinases display a highly interlinked modular topology, suggesting that these proteins are organized in larger complexes. This was apparent within the SRPK, CLK, and CDK families (Figure 2C). Among the 131 proteins associated with SRPK and the related CLK family members, we found proteins that partition in several distinct RNA processing complexes, including the Nop56p-associated pre-rRNA complex; the U2, U4/U6, and U5 spliceosome complexes; and the EJC/TREX complex (Hayano et al., 2003; Hegele et al., 2012). Among the well-studied CDK family members, we could identify expected complexes with regulatory subunits, cyclins, and CDK inhibitors. Figure 2D illustrates the interactions found for CDK2. For poorly studied CDK family members, we found a higher fraction of previously unknown additional interactions. CDK11A, for example, revealed interactions with casein kinase subunits, a panel of zinc finger transcription factors, and nuclear transport proteins, which point to previously unrecognized roles for CDK11A in the control of transcription and nuclear transport (Figure 2E).

## Relationships among CMGC Interaction Proteomes

To analyze the relationships among the interactomes of different CMGC group members, we performed hierarchical clustering of CMGC kinases and their HCIPs (Figure 3A). We found that individual CMGC kinases can form unique complexes that are highly specific for that particular kinase, such as CDK7 or nemo-like kinase (NLK). We also noticed that related kinases often share a significant fraction of HCIPs. Closer inspection of these related interactomes revealed interesting differences that may point to different biochemical roles for related CMGC group members. For example, GSK-3 α (GSK3A) and GSK-3 β (GSK3B) share many HCIPs (specifically GSKIP, FRAT1, PPP1R2, AXIN1, AXIN2, and PRKACA) that are bound even with similar relative abundances as estimated by spectral counting (Figure 3B). However, only GSK3B could be detected in complexes with β-catenin (CTNNB1) and adenomatous polyposis coli (APC; Figure 3B), indicating a nonredundant function of GSK3B in Wnt signaling. This finding is consistent with the key role assigned to GSK3B in regulating Wnt/β-catenin signaling. In the absence of Wnt signaling, GSK3B binds to the "destruction complex," which in addition contains AXIN1, APC, and casein kinase I (CKI). As part of this complex, GSK3B phosphorylates β-catenin, leading to its subsequent ubiquitination and proteasomal degradation (Dajani et al., 2003). The GSK3B example illustrates how subtle difference in protein complex formation may result in functional diversification. It also suggests that such specific differences in the presented PPI data may be helpful for functionally dissecting other highly related CMGC group members that are less well characterized than GSK3B.

## CDK-Containing Transcription Complexes

Three members of the CDK family (CDK7, CDK8, and CDK19) form complexes with established roles in basic RNA pol II transcription. We noted that the identified HCIPs of CDK7 (CDK8 and CDK19) were highly connected based on public PPI data, indicating the formation of large complexes engaged in transcription. Since complex formation with these three kinases has been well studied in the past, we used these examples to benchmark the robustness and sensitivity of our experimental approach. CDK7 is known to form a trimeric complex with cyclin H (CCNH) and MNAT1 (Figure 3C), also referred to as the CDK-activating kinase or CAK (Kaldis, 1999). CAK phosphorylates other CDKs within the activation segment (T-loop) and acts as a component of the general polymerase II transcription factor TFIIH, where it phosphorylates the C-terminal domain of the RNA pol II large subunit (Roy et al., 1994; Figure 3C). The helicase xpd/ERCC2 plays a structural role in tethering the CDK7-cyclin H-MNAT1 trimer to the core subunits of TFIIH (Chen et al., 2003; Coin et al., 1999). Our analysis revealed all ten subunits of TFIIH and CAK, of which the CAK subunits represent the most abundant protein interacting with CDK7 (Gibbons et al., 2012).

**Figure 3. Hierarchical Clustering of the CMGC HCIPs**

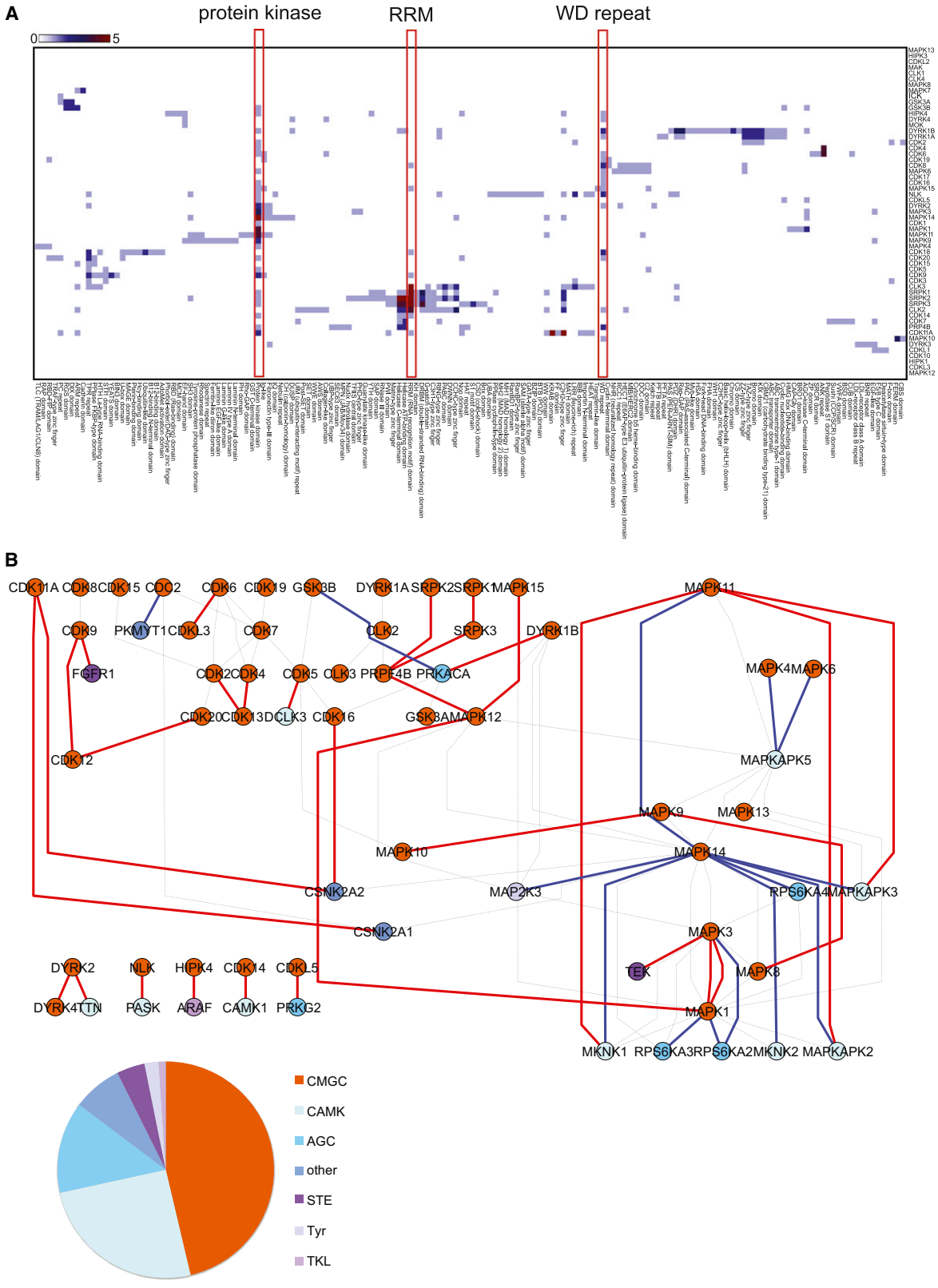
(A) Heatmap generated from hierarchical clustering of the 481 HCIPs and the 57 CMGC kinases reveals clusters of related kinases that share interactions. The color of each individual HCIP (rows) corresponds to its relative abundance, quantified using normalized spectral counts, in each CMGC kinase complex (columns).

(B) GSK3A and GSK3B form overlapping but also distinct interactions, reflecting their different roles in Wnt signaling.

(C) CDK7, a component of the general transcription factor TFIIH, has a unique interactome "fingerprint" comprised exclusively of interactions with the components of the TFIIH complex.

(D) CDK8 and CDK19 form a highly related cluster that corresponds to the transcriptional MED complex. The MED complex associates in a mutually exclusive manner with either CDK8 or CDK19. Blue edges indicate interactions identified in this study and known interactions are shown by black dotted lines.

See also Figure S3.




(legend on next page)

CDK8 and CDK19 and their regulatory subunit cyclin C (CCNC) are components of the Mediator (MED) coactivator complex (Sato et al., 2004). The precise composition of this subunit, as well as potential differences in MED complex formation, has not been described yet. The MED complex supports transcriptional activation via binding to the activation domain of transcription factors and to the general pol II transcription machinery. The MED core is composed of a head module, a middle module, and a tail module (Guglielmi et al., 2004), and can bind to RNA polymerase II (pol II) to form a holoenzyme. Alternative forms of the MED complex have been proposed that are free of pol II but include a kinase module, Cyclin C, and two additional subunits: MED12 and MED13 (Guglielmi et al., 2004). The two additional core complex subunits, MED25 and MED26, which function specifically in transcriptional activation, are not present in the kinase-associated MED (Ding et al., 2009; Yang et al., 2004). Consistent with this model, we found that both CDK8 and CDK19 form mutually exclusive complexes with all subunits of the alternative form of MED complex lacking MED25 and MED26. In addition, we found that CDK8 and CDK19 associate with the MED subunits in similar relative amounts, as estimated from normalized spectral counts (Figure 3D), indicating that the two paralogous kinases do not undergo differential MED complex formation. The presented data represent the most complete set of interactions between the CDK8/19 and MED subunits currently available, and illustrate the robustness and sensitivity of the experimental and computational workflow used in this study.

## Topology of CMGC Kinase-Kinase Interactions

We next clustered the occurrence of specific structural domains in the proteins found in CMGC kinase complexes across all CMGC kinases (Figure 4A). The most frequently found domains were the RNA recognition motif (RRM) domain (n = 55) and protein kinase domain (n = 45). The majority of proteins containing an RRM domain were associated with SRPKs and CLKs. Remarkably, almost half of the CMGC kinases were interacting with proteins that contain a protein kinase domain, suggesting that human CMGC kinases have a previously unidentified preference for interacting with other protein kinases, which is reminiscent of the kinase-kinase network reported in a global analysis of yeast kinase complexes (Breitkreutz et al., 2010).

To illustrate the kinase-kinase interactions in more detail, we combined all detected kinase-kinase interactions with published interaction information for the identified kinases in a network graph (Figure 4B). Remarkably, the majority of the kinase-kinase interactions (n = 44 interactions) represent CMGC kinases interacting with other CMGC kinases. The second-largest kinase group is the CAMK group (n = 24 interactions), followed by the AGC group (n = 13 interactions). Most of the previously reported kinase-kinase interactions clustered around the MAPK family. The presented kinase-kinase network provides insights into the organization of kinases into large regulatory kinase-kinase networks, and the mechanism of multisite substrate phosphorylation by kinase-kinase complexes.

## Inferring Regulatory Networks within the CMGC Interaction Proteome

Protein phosphorylation requires the formation of transient or stable kinase-substrate complexes. The obtained AP-MS data may therefore include new kinase-substrate relationships. In order to narrow down the potential kinase-substrate space, we pursued the following strategy: First we compiled all phosphopeptides identified from the CMGC AP-MS data and all phosphopeptides from CMGC network nodes available from public data (Hornbeck et al., 2012). Second, we predicted upstream kinases for these phosphosites using the NetworKIN algorithm to construct a hypothetical kinase-substrate network (Linding et al., 2007). Finally, we merged this predicted network with the CMGC AP-MS data to identify those kinase-protein interactions that intersect with the predicted kinase-substrate interaction and thus may represent high-confidence kinase-substrate interactions.

In total, we identified 1,315 phosphopeptides (false discovery rate [FDR] < 1%) corresponding to 503 unique phosphosites in 106 unique proteins from the CMGC AP-MS data. Of the 503 phosphosites identified, 380 had been reported previously (Figure 5A; Table S3). A large fraction (188/503) of the identified phosphosites was detected in protein kinases, and a fraction mapped to the activation loop region in the CMGC kinases (Figure S5).

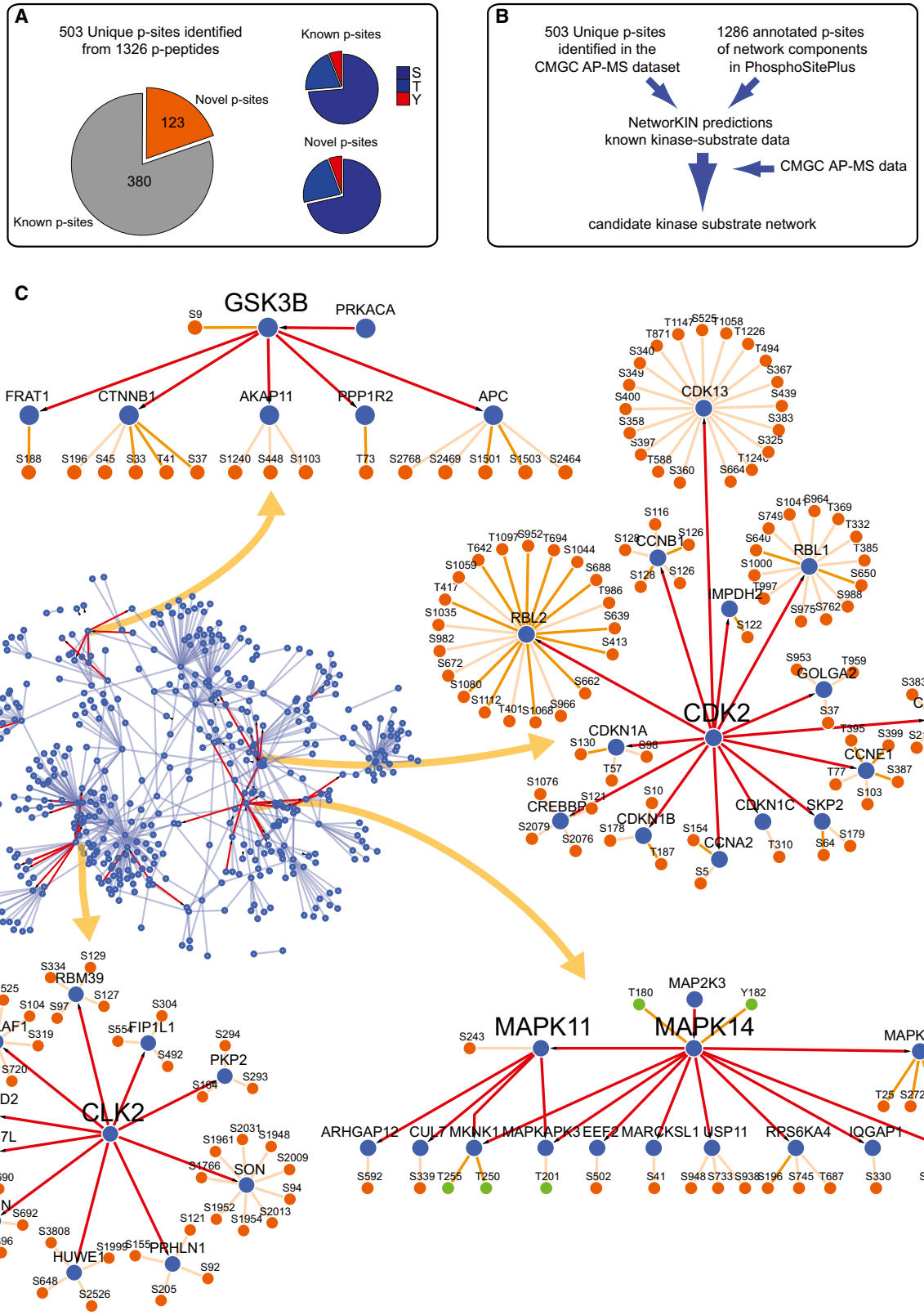
When we combined our data with public phosphosite information, we obtained a total of 1,789 phosphosites for the 481 CMGC network components (Figure 5B; Hornbeck et al., 2012). Using NetworKIN, we could predict candidate upstream kinases for these sites (Table S4). Overall, our PPI data intersected with 47 predicted kinase-substrate interactions. In addition to the kinase-substrate interaction predicted by NetworKIN, our AP-MS data overlapped with 23 experimental kinase-substrate interactions (Hornbeck et al., 2012), resulting in a total overlap of 56 interactions between 17 kinases and their candidate substrates (Table S4). From the intersecting candidate kinase-substrate network (Figure 5C) we identified several regulatory modules where a kinase could be assigned to a set of candidate substrates (Figure 5C). For example, the relatively poorly studied kinase CLK2 is associated with ten candidate substrates, the majority of which can be linked to RNA processing and transcription. Furthermore, our analysis revealed several GSK3B sites on three components of the Wnt signaling pathway,

**Figure 4. CMGC-Kinase-Kinase Highway**

(A) Hierarchical clustering of protein domains present in the CMGC interacting proteins. The domains found most frequently within CMGC interacting proteins are highlighted. The color gradient corresponds to the number of unique proteins that contain a particular domain and interact with an individual CMGC kinase.

(B) Network model illustrating physical interactions between CMGC kinases and other kinase family members. Node color corresponds to the kinase families shown in the pie chart. The pie chart shows the distribution of interactions between various kinase family members and CMGC kinases. Dashed lines illustrate interactions from the public databases, and solid edges represent known (blue edges) and additional (red edges) kinase-kinase interactions found in the AP-MS data set.

See also Figure S4.




*(legend on next page)*

including the evolutionary conserved residues S33, S37, and T41 on β-catenin (Liu et al., 2002); the APC phosphorylation on S1501 (Ferrarese et al., 2007); and S188 of the proto-oncogene FRAT1 (Hornbeck et al., 2012). Besides Wnt signaling components, we also found Protein phosphatase inhibitor 2 and Protein kinase A-anchoring protein 11 (AKAP11) as candidate GSK3B substrates. The cyclic AMP-dependent protein kinase (PRKACA) and its regulatory subunit PRKR1A were also found in GSK3B complexes, and PRKACA has been proposed to phosphorylate GSK3B on S9 (Hornbeck et al., 2012).

The largest candidate kinase-substrate networks from our analysis were found for CDK2, which suggests 14 substrates (eight that have already been annotated as in vivo or in vitro substrates, and six that represent additional substrate candidates [CREBBP, CDKN1C, CCNB1, GOLGA2, CCNE2, and CDK13]).

The same approach also identified kinases predicted to phosphorylate other kinases in their T-loop (also referred to as the activation loop) region and thus could act as upstream activating kinases in regulatory networks. The examples identified revealed a MAPK signaling module in which MAPK family members appear to be sequentially phosphorylated at their T-loop site by other MAPK family members (Figure 5C).

Consistent with our findings, MAP2K3 (also known as MKK3) is known to activate MAPK14/p38-MAPKα via T-loop phosphorylation of T180 and Y182 of MAPK14 (Dérijard et al., 1995; Raingeaud et al., 1995). Following its activation, MAPK14 can phosphorylate a number of substrates, including other MAPK family members (Cuadrado and Nebreda, 2010). Our analysis showed that MKNK1 and MAPKAP3 form stable complexes with the two p38 kinases (MAPK14 and MAPK11) and are predicted to be phosphorylated in their T-loop region by the associated p38 kinases. By applying the combined AP-MS/NetworKIN analysis, in addition to known p38 substrates (e.g., EEF2 and RPS6KA4), we identified additional candidate substrates. These include the candidate MAPK14 substrates USP11 (a deubiquitinating enzyme implicated in NF-kappa-B activation; Yamaguchi et al., 2007) and IQGAP1 (implicated as a scaffold for MAPK signaling; Roy et al., 2004, 2005), as well as the MAPK11 candidate substrates ARHGAP12 (a Rho-type GTPase-activating protein) and Cul7 (an Skp, Cullin, F-box-containing complex [SCF] E3 ubiquitin ligase component that is genetically linked to 3-M syndrome; Huber et al., 2005).

## An SRPK-Substrate Network Is Linked to RNA Processing

Substrate prediction tools such as NetworKIN and Scansite do not cover SRPKs because only a few substrates are known for the SRPK family and prediction is complicated by a mechanism of progressive substrate phosphorylation of multiple serines (Aubol et al., 2003). To identify candidate substrates for SRPK1 and SRPK2, we used two criteria. The first is the ability to reconstitute substrate phosphorylation in vitro, and the second is the physical association of the kinase with its substrate in vivo. We performed systematic in vitro kinase (IVK) assays with SRPK1 and SRPK2 on protein microarrays containing >9,000 unique human proteins (Figure 6A) and combined the results from these experiments with the SRPK1/2 AP-MS data (Figure 6B). Using a Z-score cutoff of 0.25, we identified 155 and 160 IVK substrates for SRPK1 and SRPK2, respectively (Table S5). We found differences between SRPK1 and SRPK2, but also 72 shared substrates (Figure 6B). Furthermore, the identified IVK substrates were highly enriched for known splicing proteins ($p = 1.80E-15$; Figure S6; Table S6). Among the three annotated substrates for SRPK1 and SRPK2 that were present on the microarray, we could confirm SRSF1 and RBM8A as in vitro substrates. Besides SRSF1, we identified 26 proteins linked to messenger RNA (mRNA) splicing, 22 of which could be assigned to distinct functional groups involved at different steps of mRNA splicing (Hegele et al., 2012; Figure 6C). We noticed a significant overlap between the proteins identified by IVK experiments and AP-MS. The integration of SRPK1 and SRPK2 IVK-substrate relationships with the AP-MS data resulted in a network model that consists almost exclusively of proteins involved in mRNA splicing or RNA processing (Figure 6D, light orange), 11 of which have been described recently as components of the exon junction complexes (EJC) interactome (Singh et al., 2012). These data suggest that SRPK protein binding and kinase activity may be required to coordinate distinct molecular events involved in RNA processing.

## Disease Phenotypes Linked to the CMGC Interaction Proteome

The systematic mapping and high-throughput sequencing of human disease loci provides insights into the role of cellular proteins in human pathogenesis. However, the mechanistic basis of how genomic lesions are translated into disease phenotypes is poorly understood. We therefore integrated the information on disease loci with the CMGC kinase-protein interaction data to better understand how kinases may control disease pathways, and to uncover modules of biochemically related proteins linked to a specific disease phenotype.

Using annotated genetic disease information (OMIM and COSMIC), we queried the CMGC AP-MS data set for the presence of disease-associated proteins (DAPs) and studied the topology of a protein interaction network formed between kinases and DAPs. Altogether, we identified 91 DAPs that physically interacted with CMGC kinases, forming a network of 143 interactions (Figure 7A). By far the most prominent disease group we found in the CMGC interaction proteome was cancer. Other

**Figure 5. Candidate CMGC Kinase-Substrate Network Predicted by a Combined Proteomics and Computational Approach**

(A) CMGC AP-MS analysis revealed 503 unique phosphosites, 123 of which have not been detected before.

(B) Computational strategy to predict upstream kinases for the phosphosites present in the CMGC network components. Phosphosites found in this study were combined with annotated phosphosites to predict candidate kinase relationships using NetworKIN. The predictions were filtered for kinase-substrate pairs that also form stable complexes based on the CMGC AP-MS data set.

(C) Network diagrams illustrating kinase-substrate interactions (red edges) intersecting with the AP-MS data set (blue edges). Nodes in the enlarged network graphs illustrate kinases (blue), substrates (blue), and phosphorylation sites (orange). T-loop phosphorylation sites are indicated as green nodes. Edges pointing toward the phosphorylation sites indicate known (orange) or predicted (opaque orange) phosphorylation events.

See also Figure S5 and Tables S3 and S4.

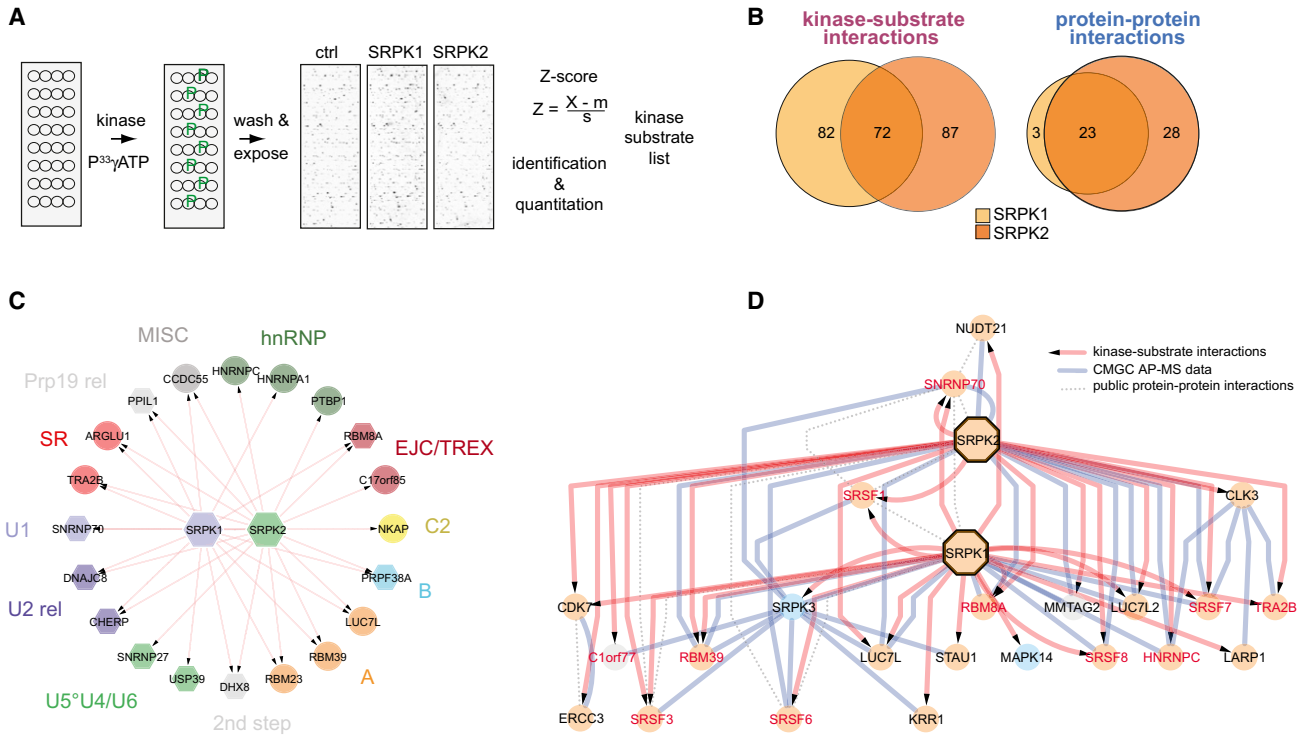



**Figure 6. SRPK1 and SRPK2 Physically and Functionally Interact with Different Classes of mRNA-Processing Proteins**

(A) Identification of SRPK1 and SRPK2 IVK substrates using protein microarrays.

(B) Venn diagrams illustrating substrates and binding partners shared by SRPK1 and SRPK2.

(C) SRPK1 and SRPK2 IVK-substrate network involving distinct functional groups engaged during various steps of mRNA splicing.

(D) Integration of data from IVK assays and AP-MS data reveals a functional network involved in various steps of RNA processing (light orange nodes). Components of the EJC interactome are indicated in red.

See also Figure S6 and Tables S5 and S6.

diseases linked to CMGC complexes include mental retardation, microencephaly, retinitis pigmentosa, xeroderma pigmentosum, and 3-M syndrome. If a group of proteins constitute a specific complex that is linked to a particular disease pathway, it is conceivable that mutations in the corresponding complex subunits might result in a similar disease phenotype. We therefore analyzed the CMGC interaction proteomes for cases in which a particular disease phenotype was clustered around specific kinase complexes. Despite the limited availability of genetically mapped disease information and the likely incompleteness of the AP-MS interaction data, we could identify five clusters in which a group of proteins binding to a particular CMGC kinase mapped to a specific disease phenotype (Figure 7A). Specifically, all proteins in the CMGC network (PRPF8, SNRNP200, PRPF31, and PRPF6) associated with retinitis pigmentosa, a progressive retinal dystrophy, were exclusively found in complexes with PRPF4B. Likewise, we found Cul7, CCDC8, and OBSL1—three proteins that are associated with 3-M syndrome, a primordial growth retardation disorder (Hanson et al., 2011)—in complexes with the p38 MAPKs MAPK11 and MAPK14.

Overall, 43 kinase-associated proteins identified in the AP-MS data set are genetically linked to various forms of human cancer. The observed frequency is about three times higher than expected from a random network of similar topology (Figure 7A, inset). Several CMGC kinases qualified as "cancer hubs" with multiple interactions to cancer-linked proteins. Similarly to the examples described above, some of these "cancer hubs" were associated with a set of cancer proteins pointing to a particular type of cancer. For example, and in agreement with previous results, we identified three proteins (ERCC2, ERCC3, and ERCC5) in CDK7 complexes that are linked to xeroderma pigmentosum, an autosomal recessive genetic disorder that is linked to severe DNA repair defects and causes severe UV sensitivity with increased risk for skin cancer (Giglia-Mari et al., 2004; Jeronimo et al., 2007). Moreover, all of the five proteins (MLLT1, MLLT3, AFF1, AFF3, and AFF4) linked to acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) in the CMGC network were identified in complexes with CDK9 (Figure 7B). In ALL and AML, the genes of these CDK9-binding proteins undergo translocation with the histone methyltransferase MLL. It is believed that these diseases are caused by epigenetic changes that lead to overproliferation of immature white blood cells (Benedikt et al., 2011).

Other CMGC kinases associate with multiple cancer proteins that are linked to various cancer types. CDK6, for example, has been found in complexes with the cyclin D2 (CCND2), cyclin D3 (CCND3), and CDK inhibitor p18-INK6 (CDKN2C) linked to various tumor types, including non-Hodgkin lymphomas

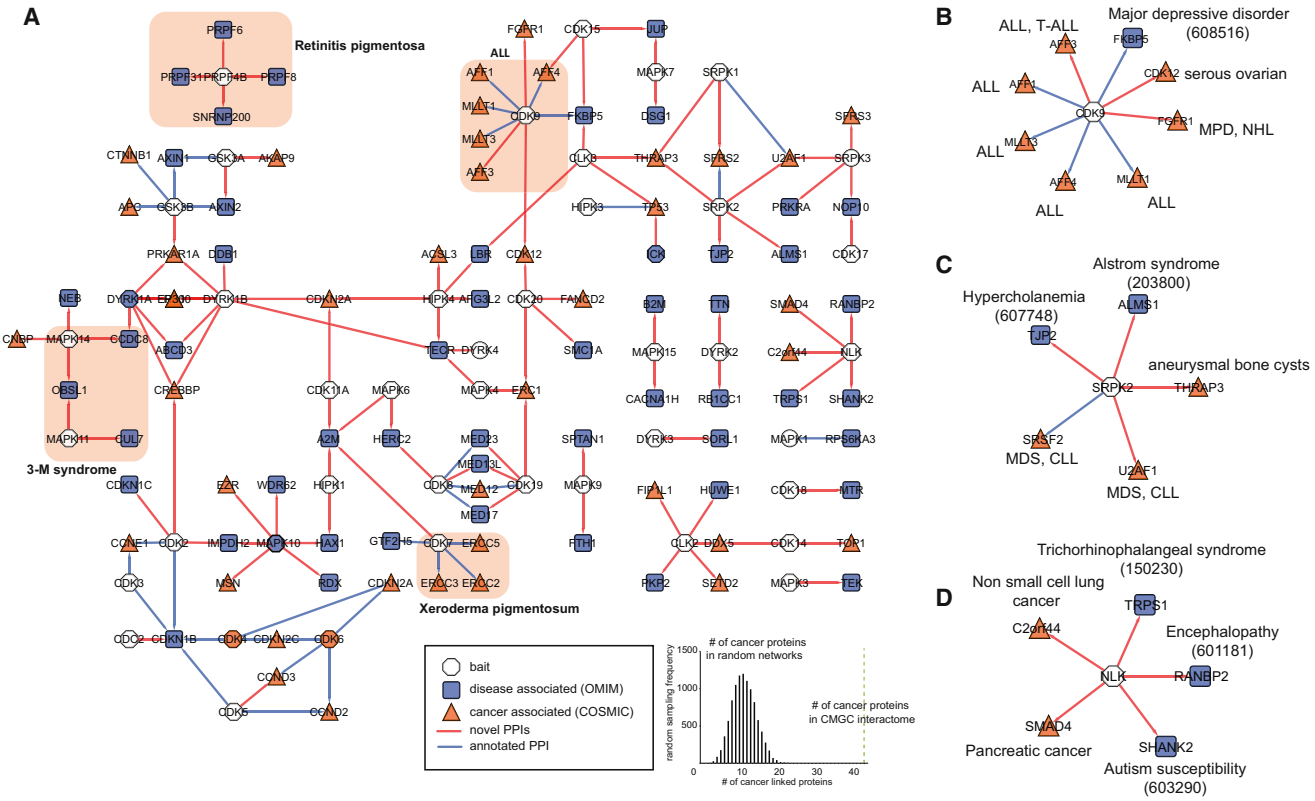



**Figure 7. Complex Formation between CMGC Kinases and Human Disease Proteins**

(A) Physical interactions between CMGC kinases and proteins associated with human diseases. Known interactions are indicated by blue edges, and additional interactions are highlighted in red and. Modules containing CMGC kinase-associated proteins genetically linked to the same disease phenotype indicated in the figure are highlighted as orange squares. Inset at the bottom illustrates enrichment of cancer proteins in the CMGC interactome (green dotted line; n = 43; $p < 0.001$) compared with the distribution of CAPs occurring in 10,000 random networks (481 proteins, 652 interactions).

(B) CDK9 interacts almost exclusively with proteins mutated in ALL.

(C) SRPK2 complex formation with human DAPs.

(D) DAPs in complexes with NLK. Numbers in parentheses refer to OMIM IDs.

(CCND2) and gliomas (CDKN2C). Likewise, GSK3A and GSK3B interact with a set of proteins linked to colon cancer (APC and CTNNB1), hepatocellular carcinoma (AXIN1), and thyroid carcinoma (PRKAR1A; Forbes et al., 2008). Importantly, besides these established cancer protein interactions, our data revealed previously unrecognized interactions with CAPs and less studied kinase subfamily members such as CLK2 and SRPK2. CLK2, for example, forms interactions with three CAPs, including SETD2, a histone lysine N-methyltransferase that was recently linked to renal carcinoma (Dalgliesh et al., 2010). Likewise, SRPK1 and SRPK2 interact with THRAP3, SRSF2, and U2AF1, and both SRSF2 and U2AF1 are linked to chronic lymphocytic leukemia (CLL), the second most common adult leukemia (Figure 7C). In addition, NLK forms complexes with five DAPs, including SMAD4, a protein that has been found to be significantly mutated in human pancreatic cancer (Figure 7D; Biankin et al., 2012; Hahn et al., 1996). Because most of these interactions between CMGC group members and DAPs have not yet been reported in public interaction databases, the presented kinase-DAP interactome represents a rich source for developing novel hypotheses about the control of disease pathways by CMGC kinases.

## Concluding Remarks

In this work we have provided a comprehensive analysis of the CMGC kinase interaction proteome. Our study reveals 531 protein interactions for the CMGC kinase group and confirms 121 previously described interactions, suggesting that the previously annotated interaction proteome for CMGC kinases, and thus for the human kinome in general, is far from being complete. The interactions we found are not equally distributed across the CMGC kinase families but are enriched in subgroups that are poorly represented in the current literature, such as the CLK, HIPK, and DYRK families. This indicates that the protein-interaction information for the CMGC kinase group that is available in public databases is biased toward well-studied kinases. This bias, together with limited interaction information, restrains the development of models for regulatory pathways, highlighting the need for systematic AP-MS studies to fill the gaps in our knowledge about poorly studied protein groups.

Comparisons of orthologous protein-interaction data across species can provide insights into evolutionary conservation and changes in protein interactions and complexes. We compared the CMGC interaction data from our study with orthologous

interaction data from budding yeast because yeast represents the only eukaryotic organism for which proteome-wide interaction data are currently available. We found conservation around complexes of the CDK, MAPK, DYRK, SRPK, and CLK families, but not for the other CMGC families (Figure S7; Table S7). Approximately half of the orthologous interactions represented in the current yeast interactome involve CDKs. They constitute highly conserved complexes with cyclins and CDK regulatory subunits, TFIIH, or transcriptional MED complexes involved cell-cycle control and transcription, two basic biological functions. Also, the orthologous interactions within many MAPK family members are well represented in existing yeast interaction data and indicate an expansion of orthologous human MAPK complexes compared with yeast. In contrast, orthologous interactions with other family members are poorly represented in the known yeast interactome or are not present at all (GSK3A, GSK3B, HIPK4, and PRPF4B). These differences may indicate evolutionary changes in the human interaction landscape, but to some extent they may also reflect limitations in protein interaction data coverage. The presented data will facilitate efforts to understand the functional diversification of a related group of kinases by means of alternative complex formation. We noted diverse but also highly related complex formation among closely related CMGC group members, which indicates functional diversification but also redundancy of certain CMGC kinases. With the contextual information presented here, it is now possible to assign poorly characterized kinases to a specific functional context and expand the potential roles of some of the established CMGC group members. The CMGC interaction proteome may include regulators of kinase activity as well as novel kinase substrates. Obtaining reliable kinase-substrate information represents a major challenge in the development of novel regulatory pathway models to elucidate cellular regulation in health and disease. To this end, we could show that at least a fraction of kinase-associated proteins can also serve as kinase substrates. Despite obvious restrictions imposed by complex stability during AP, the limited dynamic range of MS analysis, and the limited signaling states represented in exponentially growing HEK293 cells, we were able to identify, besides well-established kinase substrates, a number of candidate kinase substrates by using an integrated proteomics and computational approach. In parallel, by using a protein microarray approach, we also showed that a significant fraction of proteins identified in complexes of the SRPK group also act as kinase substrates in vitro.

Kinases represent 20% of current drug targets and have been genetically linked to a number of human disease phenotypes, including cancer (Hopkins and Groom, 2002). Compounds that either enhance or inhibit disease-relevant kinase-protein interactions may provide a promising but as yet poorly explored route to modulate kinase networks that are perturbed in human diseases. The CMGC interaction proteome revealed 108 physical interactions between CMGC group members and human DAPs not found in public databases. The distribution of disease phenotypes across the interaction landscape revealed several specific diseases that were clustered around particular kinases and thus may provide clues to the role of protein complexes in controlling the emergence of specific disease phenotypes. Proteins associated with human cancer were enriched in the CMGC kinase network, which may hint at a role of the CMGC interaction proteome in controlling tumor cell growth. Therefore, the presented CMGC interaction proteome represents a knowledge base for instructing future functional experiments to uncover the molecular mechanisms that control human tumor growth by specific CMGC kinases, which in turn may provide new opportunities for pharmacological intervention using drugs that modulate the kinase interaction proteome.

## EXPERIMENTAL PROCEDURES

### Cell Culture

SH-tagged human CMGC kinases stably and inducibly expressing Flp-In 293 T-Rex cell lines were each expanded to ~5 × $10^7$ cells and 1 μg/ml tetracycline was added for 24 hr to induce expression of SH-tagged bait proteins.

### AP of Protein Complexes and MS Analysis

For AP, the 5 × $10^7$ cells were lysed in 4 ml of lysis buffer (0.5% NP40, 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 50 mM NaF, 1.5 mM $NaVO_3$, 5 mM EDTA, supplemented with 0.5 mM PMSF and protease inhibitors; Sigma). The cleared lysate was loaded on spin columns (Bio-Rad) containing 200 μl Strep-Tactin beads (IBA GmbH) and the beads were washed three times with 1 ml of lysis buffer. Proteins were eluted from the Strep-Tactin beads with 2 mM biotin, followed by incubation of the eluate with 100 μl anti-HA agarose (Sigma) for 2 hr on a rotation shaker. The anti-HA agarose was washed three times with 1 ml of lysis buffer without protease inhibitor and detergent. Purified protein complexes were eluted with 0.2 M glycine, pH 2.5, and subsequently neutralized with 100 mM $NH_4HCO_3$. Cysteine bonds were reduced with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) and alkylated with 10 mM iodoacetamide. The proteins were then trypsinized to peptides, and the peptides were purified with C18 microspin columns (Harvard Apparatus).

### MS Analyses

The tryptic peptide samples were analyzed on a hybrid LTQ Orbitrap XL mass spectrometer (Thermo Scientific) using Xcalibur version 2.0.7 coupled to an Eksigent NanoLC-2D HPLC nanoflow system (dual pump system with one analytical column; Eksigent) via a nanoelectrospray ion source using a liquid junction (Thermo Scientific). Each purification sample was analyzed in technical replicates, and for each replicate 7% of the sample was loaded onto a 15 cm (∅75 μm) fused silica analytical column (PicoFrit; New Objective) packed with C18 reversed-phase material (Magic C18 AQ 3 μm; Michrom BioResources). The peptides were eluted with 40 min gradient (constant flow rate of 300 nl/min) ranging from 5% to 35% solvent B, followed by a 10 min gradient from 35% to 80% solvent B. After every two technical replicate samples, 100 fmol of control peptide ([Glu1]-Fibrinopeptide B human (Sigma) was analyzed twice by LC-MS/MS, allowing the standardized monitoring of the LC-MS/MS system performance, and detection and possible exclusion of carryover protein (namely, bait protein from the previous sample). Acquired MS2 scans were searched against the UniProtKB/Swiss-Prot protein database (release 12.0) using the XTandem search algorithm (Craig and Beavis, 2004) with the k-score plug-in (MacLean et al., 2006) for identification of the CMGC interactome, or with Mascot (Matrix Science) for phosphopeptide identification.

### Identification of HCIs

To obtain high-confidence protein-interaction data from AP-MS raw data, we used Significance Analysis of Interactome (SAINT), which determines the statistical significance of observed interactions using protein abundance based on label-free quantification (Choi et al., 2011). The empirical frequency threshold was set to 0.09 and the iProb threshold value was set to >0.9 for HCIs. The detailed SAINT input and obtained output files are listed in Table S1. The filtered CMGC high-confidence data set was further analyzed and visualized using Cytoscape 2.8.3. Interaction information from the iRefWeb and PINA database was used to annotate known protein interactions (Cowley et al., 2012; Razick et al., 2011; Turner et al., 2010).

### Protein Kinase-Substrate Arrays

To identify substrates for kinases, we performed IVK assays on Protoarray human protein microarrays (version 5.0; Invitrogen). For the IVK assays on the protein microarray, the array was first blocked and then incubated with the corresponding kinase (50 nM) in the presence of radiolabeled [$\gamma$-33P] ATP. The array was then washed to remove the unbound $\gamma$-33P, dried, and exposed to X-ray film. The acquired image of the array was analyzed using the ProtoArray Prospector software bundle (Invitrogen). The raw data were subjected to background substraction, signal scatter compensation, and outlier detection, and the *Z* factor cutoff value was set at $\geq$0.4. Phosphorylated proteins with a *Z* score > 0.25 were considered as potential substrates.

For further details regarding the methods and materials used in this work, see the Extended Experimental Procedures.

## ACCESSION NUMBERS

The interaction data from this article have been deposited in the IntAct database under accession number IM-17935 (http://www.ebi.ac.uk/intact/), and the peptide raw data have been deposited in the Peptide Atlas under accession number PASS00226 (http://www.peptideatlas.org/).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, seven tables, and Extended Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.03.027.

## LICENSING INFORMATION

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ACKNOWLEDGMENTS

We thank Professor Jussi Taipale for critically reading the manuscript, and Atul Sethi for bioinformatics support. This work was supported by the European Union 7th Framework project PROSPECTS (Proteomics Specification in Space and Time, grant HEALTH-F4-2008-201648), the SystemsX.ch project PhosphonetX and ERC advanced grant Proteomics v3.0 (grant 233226) from the European Union to R.A., and European Union 7th Framework Marie Curie Actions IEF grant "Cancer Kinome" (grant 236839) to M.V.

Received: December 14, 2012
Revised: March 1, 2013
Accepted: March 18, 2013
Published: April 18, 2013

## REFERENCES

Akiyoshi, S., Inoue, H., Hanai, J., Kusanagi, K., Nemoto, N., Miyazono, K., and Kawabata, M. (1999). c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. J. Biol. Chem. *274*, 35269–35277.

Aranda, S., Laguna, A., and de la Luna, S. (2011). DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. FASEB J. *25*, 449–462.

Aubol, B.E., Chakrabarti, S., Ngo, J., Shaffer, J., Nolen, B., Fu, X.D., Ghosh, G., and Adams, J.A. (2003). Processive phosphorylation of alternative splicing factor/splicing factor 2. Proc. Natl. Acad. Sci. USA *100*, 12601–12606.

Behrends, C., Sowa, M.E., Gygi, S.P., and Harper, J.W. (2010). Network organization of the human autophagy system. Nature *466*, 68–76.

Benedikt, A., Baltruschat, S., Scholz, B., Bursen, A., Arrey, T.N., Meyer, B., Varagnolo, L., Müller, A.M., Karas, M., Dingermann, T., and Marschalek, R. (2011). The leukemogenic AF4-MLL fusion protein causes P-TEFb kinase activation and altered epigenetic signatures. Leukemia *25*, 135–144.

Biankin, A.V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J., et al.; Australian Pancreatic Cancer Genome Initiative. (2012). Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature *491*, 399–405.

Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduva, V., Larsen, B., Lin, Z.Y., Breitkreutz, B.J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. Science *328*, 1043–1046.

Bunger, M.K., Wilsbacher, L.D., Moran, S.M., Clendenin, C., Radcliffe, L.A., Hogenesch, J.B., Simon, M.C., Takahashi, J.S., and Bradfield, C.A. (2000). Mop3 is an essential component of the master circadian pacemaker in mammals. Cell *103*, 1009–1017.

Chen, J., Larochelle, S., Li, X., and Suter, B. (2003). Xpd/Ercc2 regulates CAK activity and mitotic progression. Nature *424*, 228–232.

Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. Nat. Methods *8*, 70–73.

Coin, F., Bergmann, E., Tremeau-Bravard, A., and Egly, J.M. (1999). Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. EMBO J. *18*, 1357–1366.

Cowley, M.J., Pinese, M., Kassahn, K.S., Waddell, N., Pearson, J.V., Grimmond, S.M., Biankin, A.V., Hautaniemi, S., and Wu, J. (2012). PINA v2.0: mining interactome modules. Nucleic Acids Res. *40*(Database issue), D862–D865.

Craig, R., and Beavis, R.C. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics *20*, 1466–1467.

Cuadrado, A., and Nebreda, A.R. (2010). Mechanisms and functions of p38 MAPK signalling. Biochem. J. *429*, 403–417.

Dajani, R., Fraser, E., Roe, S.M., Yeo, M., Good, V.M., Thompson, V., Dale, T.C., and Pearl, L.H. (2003). Structural basis for recruitment of glycogen synthase kinase 3beta to the axin-APC scaffold complex. EMBO J. *22*, 494–501.

Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., et al. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature *463*, 360–363.

Dérijard, B., Raingeaud, J., Barrett, T., Wu, I.H., Han, J., Ulevitch, R.J., and Davis, R.J. (1995). Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms. Science *267*, 682–685.

Ding, N., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Boyer, T.G. (2009). MED19 and MED26 are synergistic functional targets of the RE1 silencing transcription factor in epigenetic silencing of neuronal gene expression. J. Biol. Chem. *284*, 2648–2656.

Edwards, A.M., Isserlin, R., Bader, G.D., Frye, S.V., Willson, T.M., and Yu, F.H. (2011). Too many roads not taken. Nature *470*, 163–165.

Ferrarese, A., Marin, O., Bustos, V.H., Venerando, A., Antonelli, M., Allende, J.E., and Pinna, L.A. (2007). Chemical dissection of the APC Repeat 3 multistep phosphorylation by the concerted action of protein kinases CK1 and GSK3. Biochemistry *46*, 11902–11910.

Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., and Stratton, M.R. (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr. Protoc. Hum. Genet. *Chapter 10*, Unit 10.11.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dümpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2012). Subunit architecture of general transcription factor TFIIH. Proc. Natl. Acad. Sci. USA *109*, 1949–1954.

Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J., et al. (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat. Genet. *36*, 714–719.

Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. Nat. Rev. Mol. Cell Biol. *8*, 645–654.

Glatter, T., Wepf, A., Aebersold, R., and Gstaiger, M. (2009). An integrated workflow for charting the human interaction proteome: insights into the PP2A system. Mol. Syst. Biol. *5*, 237.

Guglielmi, B., van Berkum, N.L., Klapholz, B., Bijma, T., Boube, M., Boschiero, C., Bourbon, H.M., Holstege, F.C., and Werner, M. (2004). A high resolution protein interaction map of the yeast Mediator complex. Nucleic Acids Res. *32*, 5379–5391.

Hahn, S.A., Schutte, M., Hoque, A.T., Moskaluk, C.A., da Costa, L.T., Rozenblum, E., Weinstein, C.L., Fischer, A., Yeo, C.J., Hruban, R.H., and Kern, S.E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science *271*, 350–353.

Hanson, D., Murray, P.G., O'Sullivan, J., Urquhart, J., Daly, S., Bhaskar, S.S., Biesecker, L.G., Skae, M., Smith, C., Cole, T., et al. (2011). Exome sequencing identifies CCDC8 mutations in 3-M syndrome, suggesting that CCDC8 contributes in a pathway with CUL7 and OBSL1 to control human growth. Am. J. Hum. Genet. *89*, 148–153.

Hayano, T., Yanagida, M., Yamauchi, Y., Shinkawa, T., Isobe, T., and Takahashi, N. (2003). Proteomic analysis of human Nop56p-associated pre-ribosomal ribonucleoprotein complexes. Possible link between Nop56p and the nucleolar protein treacle responsible for Treacher Collins syndrome. J. Biol. Chem. *278*, 34309–34319.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

Hopkins, A.L., and Groom, C.R. (2002). The druggable genome. Nat. Rev. Drug Discov. *1*, 727–730.

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*(Database issue), D261–D270.

Huber, C., Dias-Santagata, D., Glaser, A., O'Sullivan, J., Brauner, R., Wu, K., Xu, X., Pearce, K., Wang, R., Uzielli, M.L., et al. (2005). Identification of mutations in CUL7 in 3-M syndrome. Nat. Genet. *37*, 1119–1124.

Javelaud, D., van Kempen, L., Alexaki, V.I., Le Scolan, E., Luo, K., and Mauviel, A. (2011). Efficient TGF-β/SMAD signaling in human melanoma cells associated with high c-SKI/SnoN expression. Mol. Cancer *10*, 2.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J., et al. (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol. Cell *27*, 262–274.

Kaldis, P. (1999). The cdk-activating kinase (CAK): from yeast to mammals. Cell. Mol. Life Sci. *55*, 284–296.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. *74*, 5383–5392.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database—2009 update. Nucleic Acids Res. *37*(Database issue), D767–D772.

Lahiry, P., Torkamani, A., Schork, N.J., and Hegele, R.A. (2010). Kinase mutations in human disease: interpreting genotype-phenotype relationships. Nat. Rev. Genet. *11*, 60–74.

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., et al. (2007). Systematic discovery of in vivo phosphorylation networks. Cell *129*, 1415–1426.

Liu, C., Li, Y., Semenov, M., Han, C., Baeg, G.H., Tan, Y., Zhang, Z., Lin, X., and He, X. (2002). Control of beta-catenin phosphorylation/degradation by a dual-kinase mechanism. Cell *108*, 837–847.

MacLean, B., Eng, J.K., Beavis, R.C., and McIntosh, M. (2006). General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. Bioinformatics *22*, 2830–2832.

Manning, B.D. (2009). Challenges and opportunities in defining the essential cancer kinome. Sci. Signal. *2*, pe15.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science *298*, 1912–1934.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658.

Raingeaud, J., Gupta, S., Rogers, J.S., Dickens, M., Han, J., Ulevitch, R.J., and Davis, R.J. (1995). Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine. J. Biol. Chem. *270*, 7420–7426.

Razick, S., Mora, A., Michalickova, K., Boddie, P., and Donaldson, I.M. (2011). iRefScape. A Cytoscape plug-in for visualization and data mining of protein interaction data from iRefIndex. BMC Bioinformatics *12*, 388.

Roy, R., Adamczewski, J.P., Seroz, T., Vermeulen, W., Tassan, J.P., Schaeffer, L., Nigg, E.A., Hoeijmakers, J.H., and Egly, J.M. (1994). The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. Cell *79*, 1093–1101.

Roy, M., Li, Z., and Sacks, D.B. (2004). IQGAP1 binds ERK2 and modulates its activity. J. Biol. Chem. *279*, 17329–17337.

Roy, M., Li, Z., and Sacks, D.B. (2005). IQGAP1 is a scaffold for mitogen-activated protein kinase signaling. Mol. Cell. Biol. *25*, 7940–7952.

Sardiu, M.E., Cai, Y., Jin, J., Swanson, S.K., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2008). Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. Proc. Natl. Acad. Sci. USA *105*, 1454–1459.

Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks, C.A., Jin, J., Cai, Y., Washburn, M.P., et al. (2004). A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. Mol. Cell *14*, 685–691.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750–764.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389–403.

Thyssen, G., Li, T.H., Lehmann, L., Zhuo, M., Sharma, M., and Sun, Z. (2006). LZTS2 is a novel beta-catenin-interacting protein and regulates the nuclear export of beta-catenin. Mol. Cell. Biol. *26*, 8857–8867.

Turner, B., Razick, S., Turinsky, A.L., Vlasblom, J., Crowdy, E.K., Cho, E., Morrison, K., Donaldson, I.M., and Wodak, S.J. (2010). iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database (Oxford) *2010*, baq023.

Yamaguchi, T., Kimura, J., Miki, Y., and Yoshida, K. (2007). The deubiquitinating enzyme USP11 controls an IkappaB kinase alpha (IKKalpha)-p53 signaling pathway in response to tumor necrosis factor alpha (TNFalpha). J. Biol. Chem. *282*, 33943–33948.

Yang, F., DeBeaumont, R., Zhou, S., and Näär, A.M. (2004). The activator-recruited cofactor/Mediator coactivator subunit ARC92 is a functionally important target of the VP16 transcriptional activator. Proc. Natl. Acad. Sci. USA *101*, 2339–2344.