E3Net: A System for Exploring E3-mediated Regulatory Networks of Cellular Functions*[S]

Youngwoong Han‡§, Hodong Lee§¶‖, Jong C. Park**, and Gwan-Su Yi‡¶**‡‡

Ubiquitin-protein ligase (E3) is a key enzyme targeting specific substrates in diverse cellular processes for ubiquitination and degradation. The existing findings of substrate specificity of E3 are, however, scattered over a number of resources, making it difficult to study them together with an integrative view. Here we present E3Net, a web-based system that provides a comprehensive collection of available E3-substrate specificities and a systematic framework for the analysis of E3-mediated regulatory networks of diverse cellular functions. Currently, E3Net contains 2201 E3s and 4896 substrates in 427 organisms and 1671 E3-substrate specific relations between 493 E3s and 1277 substrates in 42 organisms, extracted mainly from MEDLINE abstracts and UniProt comments with an automatic text mining method and additional manual inspection and partly from high throughput experiment data and public ubiquitination databases. The significant functions and pathways of the extracted E3-specific substrate groups were identified from a functional enrichment analysis with 12 functional category resources for molecular functions, protein families, protein complexes, pathways, cellular processes, cellular localization, and diseases. E3Net includes interactive analysis and navigation tools that make it possible to build an integrative view of E3-substrate networks and their correlated functions with graphical illustrations and summarized descriptions. As a result, E3Net provides a comprehensive resource of E3s, substrates, and their functional implications summarized from the regulatory network structures of E3-specific substrate groups and their correlated functions. This resource will facilitate further in-depth investigation of ubiquitination-dependent regulatory mechanisms. E3Net is freely available online at http://pnet.kaist.ac.kr/e3net. *Molecular & Cellular Proteomics 11: 10.1074/mcp.O111.014076, 1–14, 2012.*

Ubiquitination is a regulatory process for the degradation of proteins influencing nearly all cellular processes. It is believed that the turnover of ~80% of cellular proteins is controlled by ubiquitination-mediated proteasomal degradation (1, 2). Damaged or abnormal proteins are selectively eliminated by this mechanism (3). In addition to being involved in proteasomal degradation of polyubiquitinated proteins, ubiquitination regulates intracellular signaling. Mono-ubiquitination of a receptor, such as epidermal growth factor receptors or notch receptors, is a prerequisite for efficient endocytic trafficking and signaling (4). In addition, ubiquitination can regulate protein functions by altering the properties or interaction mechanisms of the target protein. For example, ubiquitination-induced conformational change within the type 2 iodothyronine deiodinase acts as a switch that regulates its dimerization and catalytic activity, which control thyroid hormone action (5).

In the ubiquitination process, ubiquitin-protein ligase (E3)[1] plays a key role in regulating specific functions by recognizing substrate proteins for ubiquitination. The presence of various different E3s and their substrate specificity indicate that the degradation process, which is one of the important regulatory mechanisms in the cellular regulatory network, is specifically controlled by E3s. Thus, comprehensive knowledge about the substrate specificity of E3s can enhance the understanding of the regulatory mechanisms of cellular processes.

Not surprisingly, a rapidly growing body of published literature describes novel regulatory mechanisms of E3s for their substrates. For example, ~30,000 MEDLINE abstracts can be retrieved with the MeSH term "ubiquitin," and 17,000 abstracts can be retrieved with the MeSH term "ubiquitin-protein ligase OR e3 ligase" (in September 2011). However, these scattered data have not yet been fully integrated into a comprehensive E3-substrate network, and thus E3-mediated regulatory mechanisms cannot be analyzed in a systematic manner. Recently, specialized databases, such as UbiProt (6) and SCUD (7), provided information about E3s and their substrates identified from dozens of research articles including several high throughput proteomics studies. However, these databases only contain a small number of E3-substrate specificities (below 200, mostly for yeast), which are not sufficient to delineate the regulatory role of E3s in diverse cellular func-

From the ‡Department of Information and Communications Engineering, the ¶Department of Bio and Brain Engineering, and the **Department of Computer Science, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

[1] The abbreviations used are: E3, ubiquitin-protein ligase; GO-BP, Gene Ontology-Biological Process; GO-MF, Gene Ontology-Molecular Function; GO-CC, Gene Ontology-Cellular Component; PPI, protein-protein interaction; KEGG, Kyoto Encyclopedia of Genes and Genomes; APC, anaphase-promoting complex.

tions. The E3Miner (8), a text mining tool for an automatic extraction of E3-related data, could successfully fetch a large number of E3s and their specific substrates from MEDLINE abstracts, but the text mining method of E3Miner is not applicable to the valuable data in public databases including a catalogue of information on proteins such as UniProt.

Despite the importance of the ubiquitination-mediated regulatory mechanism on diverse cellular processes, there has been no systematic effort to build a system-wide analysis framework for the E3-mediated regulatory network with a wide range of cellular functional categories. This is mainly due to the lack of comprehensive E3-substrate specificity data. KEGG (9), one of the most comprehensive resources for cellular pathways, includes 201 biological pathways in humans, but the ubiquitination relation is found only in 14 pathways with 32 E3-substrate relations composed of 14 E3s and 26 substrates. With this small amount of information for E3-substrate relations, the complex E3-specific regulatory mechanism cannot be delineated correctly. An E3 can regulate multiple substrates in various functional categories, and each substrate can be regulated by multiple E3s. These relations can be organized as a network of E3-specific modular regulatory activities against multiple cellular pathways. For example, anaphase-promoting complex (APC)/cyclosome targets key regulators of the cell cycle such as cyclins and their related kinases (10). Our organized data show that APC/cyclosome-specific substrates are also targeted by more than 10 other E3s. An integrative view of these E3-substrate relations could elucidate cooperative or alternative regulations of the cell cycle via E3-mediated degradation. In the case of p53, a tumor suppressor regulating various anti-cancer mechanisms such as DNA repair, cell cycle arrest, and initiation of apoptosis (11), multiple E3s are involved in its ubiquitination and degradation. Our analysis results show that p53 is regulated by 20 different E3s, and 12 of them show statistically significant functional specificity of their substrate groups for particular cellular processes.

The complex cellular regulatory processes can be depicted more correctly with the addition of an E3-specific functional regulatory network by investigating the functional specificity of E3-specific substrates along their networks. A functional regulatory network of E3s and their specific substrates can be organized by analyzing common functions of individual substrates. For example, an E3-specific regulatory network of biological pathways can be constructed by integrating E3 groups regulating each pathway and pathway groups regulated by each E3. To meet this challenge, it is necessary to comprehensively collect the E3-substrate specificities and build a framework to integrate the network of E3-specific substrates and their specific functions.

We have developed E3Net in an attempt to provide a comprehensive collection of available E3s and their substrates, and a systematic framework for the analysis of E3-substrate networks and functional implications. We maximized the collection of E3s and their substrates by extracting the information from MEDLINE abstracts and rich textual descriptions in UniProt with a text-mining method and manual inspection and by integrating the curated data from high throughput experiments (12–31), UbiProt, and SCUD. This substantially enhanced E3-substrate specificity data enabled us to build comprehensive E3-mediated regulatory networks involved in diverse cellular processes from a functional enrichment analysis of E3-specific substrates with function terms in 12 different publicly available functional category resources. We also implemented interactive analysis and navigation tools in E3Net to provide a framework for systematic analysis of the E3-specific regulatory network. In this framework, the user can create an integrative view of all identified functional terms for each E3-specific substrate group and multiple E3s correlated with each functional term together.

In summary, E3Net will enhance the understanding and the possibility of further investigation of cellular regulatory mechanisms by incorporating an integrative view of the E3-mediated regulatory network. To illustrate the applicability of E3Net in the study of cellular regulatory processes, we show two case studies about an integrative view of E3-mediated regulatory modules specified in the cell cycle and a systematic search of E3-mediated regulatory modules for p53 with their functional implications.

<div align="center">EXPERIMENTAL PROCEDURES</div>

E3Net is composed of a database, analysis tools, and a web service interface. Fig. 1 presents a summary of the system configuration, and the details of the system are described below.

<div align="center">*Data Collection*</div>

E3s, Substrates, and Their Specificities—We extracted E3-substrate specificity data from the literature and UniProt database by using a text mining method. The text-mining process is designed to extract E3s, substrates, and their specificities from both MEDLINE abstracts and UniProtKB/Swiss-Prot comments, *i.e.* textual description in the comment field of UniProt protein records, which contain rich sentences about E3s, substrates, and their specificities. The detailed description of these text mining steps can be found in the supplemental materials. Currently, we target a MEDLINE corpus of 14,133 article abstracts and a UniProtKB/Swiss-Prot corpus composed of 14,167 records with keywords about E3s and substrates, such as "E3," "ubiquitin protein ligase," "ubiquitin ligase," and "ubiquitinated." In addition to E3-substrate specificities collected from textual resources, the data from high throughput experiments (12–31) and public ubiquitination databases (UbiProt (6) and SCUD (7)) are also integrated into our database. After the extraction of E3s, substrates, and their specificities, we further organized the information about E3 complexes and their individual subunit proteins, because the assembly of E3 complexes with other proteins is a prerequisite for ubiquitin ligase activity of many E3s. To organize E3 complexes, we identified E3 complexes in our data set by extracting the mentions of E3 complexes from MEDLINE abstracts and UniProt comments using a text-mining module. Among multiple subunits of an E3 complex, the protein that confers specificity with the substrates is marked as a substrate recognition component. For example, SCF complex is composed of Skp-, Cullin-, F box-, and RING box-containing proteins; the F box-containing protein is the substrate recognition component,
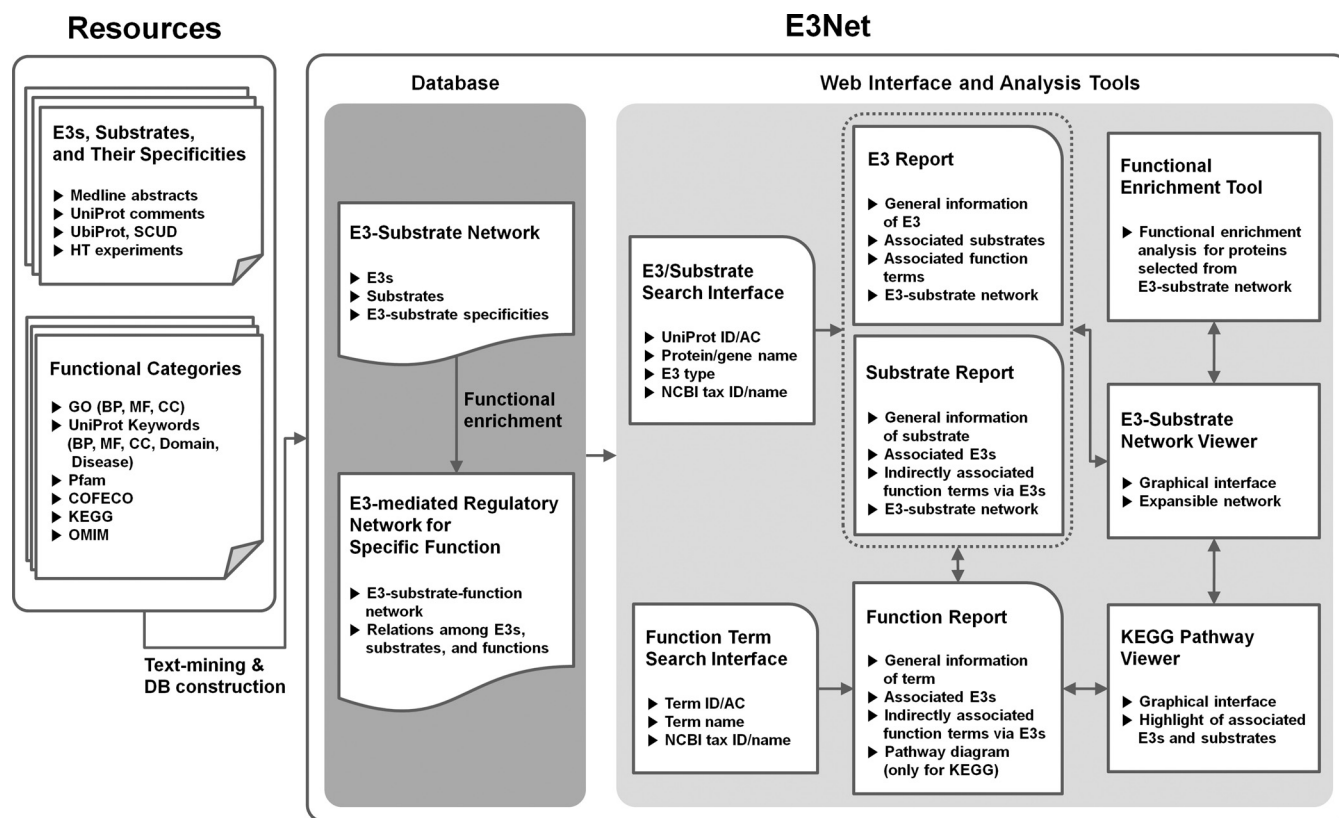
FIG. 1. **Schematic illustration of data collection, analysis, and representation of E3Net.** E3s, substrates, and their specificities are extracted from textual resources and public ubiquitination data. The functional enrichment analysis of extracted E3-specific substrate groups identifies functional specificity of substrate groups with function terms in various public functional category resources. The E3-substrate network and their specific functions are organized to build a comprehensive database of the E3-mediated regulatory network for specific functions. The web-based system of E3Net provides search interfaces for E3/substrate and specific function term and displays outputs with organized text and graphical illustrations. In addition, the set of embedded analysis and navigation tools in E3Net enables a user to explore further characteristics of the E3-mediated regulatory network of their interest.

which conjugates with the specific target substrate. In addition, we classified E3s with their E2-binding domains such as HECT, RING, and U box by using Pfam annotation. E3s that are not annotated in Pfam are assigned manually based on their descriptions in UniProt and the literature or remain as an unclassified category.

Functional Categories—To build a comprehensive reference set of functional information, we collected function terms from 12 existing functional category resources: 19,076 GO (32) biological process (GO-BP) terms, 8729 GO molecular function (GO-MF) terms, 2737 GO cellular component (GO-CC) terms, 415 biological process keywords in UniProt (33) (KW-BP), 238 molecular function keywords in UniProt (KW-MF), 130 cellular component keywords in UniProt (KW-CC), 35 domain keywords in UniProt (KW-DOM), 144 disease keywords in UniProt (KWDI), 11,912 protein families in Pfam (34) (Pfam), 28,780 protein complexes in COFECO (35) (COMP), 352 KEGG (9) pathways (KEGG), and 2626 OMIM (36) diseases (OMIM). In addition, we classified functional category resources into seven groups according to the type of information: molecular function (GO-MF and KW-MF), protein family (Pfam and KW-DOM), protein complex (COFECO), pathway (KEGG), cellular process (CO-BP and KW-BP), cellular localization (GO-CC and KW-CC), and disease (OMIM and KW-DI).

Construction of E3-mediated Regulatory Network

To build the E3-specific substrate network, we integrated E3-specific substrate groups from individual E3-substrate specificity data. To characterize the functions of those E3-specific substrate groups, we performed a functional enrichment analysis for each substrate group with the collected function terms. With significantly enriched function terms for each E3, we tried to construct a comprehensive network of cellular functions regulated by the specific E3s, which can fully delineate the relational characteristics between E3-specific substrate groups and cellular functions. The UniProt accession is used as a unified identifier to map the protein entries between E3-substrate networks and function terms. The degree of enrichment for a given substrate group and a specific function term is assessed quantitatively by a hypergeometric distribution as described in COFECO (35). The enrichment score is adjusted by Bonferroni correction to correct the occurrence of false positives. The function terms having a corrected p value below 0.01 are suggested as significant terms in a given functional category. For the pathway category, the selection threshold is loosened to a corrected p value below 0.05 to highlight more E3-mediated regulations associated with pathways. The enrichment scores of GO and UniProt keyword terms are calculated after the members of each term are reconstructed with those of all children terms under the hierarchical structure. The enrichment results of all E3-specific substrate groups are organized in a relational database such that the relations among E3s, substrates, and functions can be searched and traced. This relational structure presents an E3-substrate-function network that facilitates the delineation of the E3-mediated functional regulatory network starting from

E3s to the function terms or function terms to E3s. A user can acquire the summarized and visualized information of the network interactively with the analysis tools embedded in the web-based interface of the system.

Web Interface and Embedded Analysis Functions

Search Interface—E3Net provides two types of text search interfaces: E3/substrate search and function term search. The E3/substrate search interface supports the input queries of UniProt ID/AC, protein/gene name, E3 type (subunit composition and domain subclass), and NCBI taxonomy ID/name (for the search of whole entries in an organism) to attain an E3 or substrate specific report. Alternatively, a user can acquire a function-oriented report by using the input queries of term ID/name or NCBI taxonomy ID/name in the function term search interface.

Output and Analysis Interface—E3Net generates three forms of output interfaces: E3-, substrate-, and function-oriented reports. Basically, all identifiers for E3s, substrates, and function terms are interconnected with each other and hyperlinked to available internal or external data to allow a user to trace or compare the associated information comprehensively. The outputs can be generated and analyzed by the functions that summarize and filter the precalculated functional enrichment results and enable a user to carry out a functional enrichment analysis for the selected proteins. The output interface also has a network analysis function that visualizes and navigates the E3-substrate network and KEGG pathways interactively. The E3 or substrate report summarizes the general information of the query protein and its associated substrates or E3s and provides a graphical illustration of the E3-substrate network. All E3s, substrates, and their specificities are annotated with their source information, including sentences if available. The general information includes the name, ID, and annotated functions of a protein. The E3 report provides the associated substrates and their significantly enriched function terms in various types of functional categories in a unified table, and each function term is linked to its function report. The substrates on the E3 report can be grouped by individually or commonly enriched function terms. These commonly enriched function terms in different types of functional categories can be grouped as co-regulated functions. In addition, the E3 and substrate report contain a graphical interface providing an E3-substrate network view. In this interface, the network of the given E3 and substrate can be expanded to a larger E3-substrate network by sharing substrates or E3s. A user can select E3s and substrates of interest in this interface and independently perform a functional enrichment analysis for them. This functionality might be useful for users to derive their own interpretation when there are several possible contradictory interpretations or a few new findings that E3Net does not cover yet. Each E3-substrate network view is linked with the graphical interface of enriched KEGG pathways. The function report shows separated lists for function term-associated proteins and for E3s significantly enriched with the function term, and again, each E3 or substrate in the lists is linked to its E3 or substrate report. In the case of the KEGG pathway, the pathway diagram is visualized with the highlights of associated E3s and substrates in a graphical interface. A user can explore the KEGG pathway and associated E3-substrate network successively by using two interconnected graphical interfaces for KEGG pathways and E3-substrate networks.
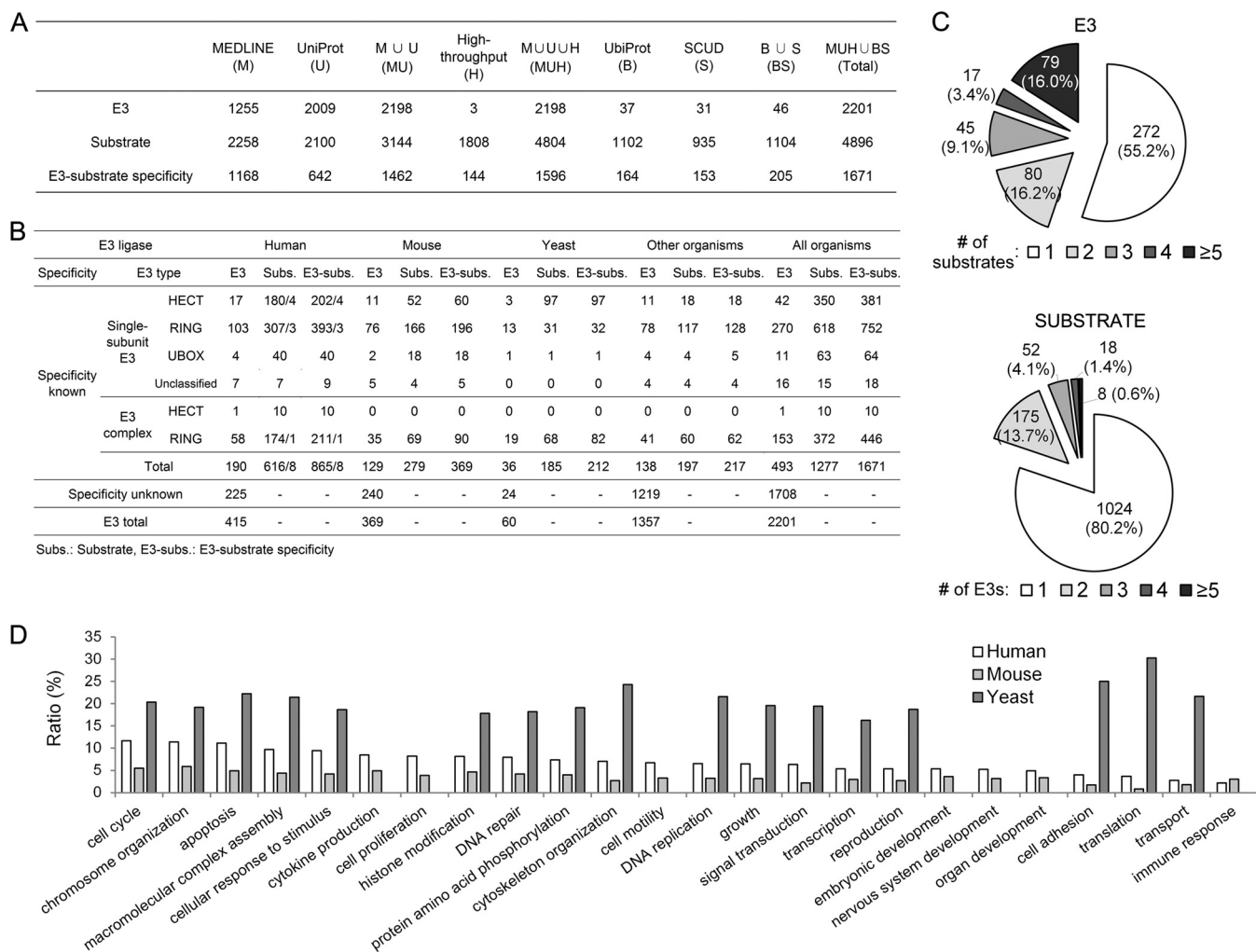
System Implementation

The database of E3Net is constructed using Oracle 10g, and the web interface is implemented with Java Server Pages and JavaScript on an Apache Tomcat 5.5 server, which is configured on a CentOS 5.5 Linux server. All of the internal functions are programmed in Java language, and visualization of the E3-substrate network is implemented using the library of Cytoscape Web (37).

RESULTS

Data Statistics of E3-substrate Network—We constructed a comprehensive database of E3s, substrates, and their specificities collected from textual resources (MEDLINE abstracts and UniProt comments), high throughput experiments, and public ubiquitination databases (UbiProt and SCUD). Through the data collection, we obtained 2201 E3s and 4896 substrates in 427 organisms and 1671 E3-substrate specificities between 493 E3s and 1277 substrates in 42 organisms (Fig. 2, *A* and *B*). Most of the collected E3s can be obtained by the text-mining process of UniProt comment, which covers 91.3% (2009 of 2201) of all E3s in E3Net. On the other hand, the substrate data are scattered over diverse resources without much redundancy, indicating the necessity of an integrative approach. 69.9% (1168 of 1671) of the collected E3-substrate specificities were extracted from MEDLINE abstract, which turned out to be the most valuable resource for collecting E3-substrate relations. As shown in Fig. 2*A*, MEDLINE abstract, UniProt comment, and high throughput experiment are three major and valuable resources to construct a comprehensive database of an E3-mediated regulatory network. UbiProt and SCUD contain relatively small numbers of E3s, substrates, and specificities, because many of their data are collected from the results of several high throughput experiments carried out for yeast proteins. Among 205 E3-substrate specificities collected from UbiProt and SCUD, 78 E3-substrate specificities (UbiProt: 55, SCUD: 50), which are not obtained from the E3Net text-mining process or high throughput experiments, are newly added into our database. To check whether newly added specificities are false negatives of our text-mining process, we inspected the reference articles of those specificities individually. For UbiProt, among 55 E3-substrate specificities newly added, 34 specificities are described in the full text of the articles, not in the abstracts. In addition, 12 specificities have no reference article. The remaining nine specificities are false negatives of our text-mining process. For SCUD, among 50 E3-substrate specificities newly added, 41 specificities are described in the full text of the articles. The remaining nine specificities are false negatives of our text-mining process.

In our database, E3s, whose specific substrates are known, are categorized into several subgroups according to their subunit composition (single-subunit E3 or E3 complex) and E2-binding domains (HECT, RING, or U box). As a result, 339 single-subunit E3s and 154 E3 complexes are identified to be involved in 1215 and 456 E3-substrate specificities, respectively, whereas the other E3s and substrates are dangled with no specific relations (Fig. 2*B*). The organized data are stored in a relational database. In human, 500–600 E3s are estimated to lead to the degradation of substrate proteins (38), whereas E3Net collects 415 human E3s, which cover 69–

A

	MEDLINE (M)	UniProt (U)	M∪U (MU)	High-throughput (H)	M∪U∪H (MUH)	UbiProt (B)	SCUD (S)	B∪S (BS)	MUH∪BS (Total)
E3	1255	2009	2198	3	2198	37	31	46	2201
Substrate	2258	2100	3144	1808	4804	1102	935	1104	4896
E3-substrate specificity	1168	642	1462	144	1596	164	153	205	1671

B

E3 ligase			Human			Mouse			Yeast			Other organisms			All organisms		
Specificity	E3 type		E3	Subs.	E3-subs.	E3	Subs.	E3-subs.	E3	Subs.	E3-subs.	E3	Subs.	E3-subs.	E3	Subs.	E3-subs.
Specificity known	Single-subunit E3	HECT	17	180/4	202/4	11	52	60	3	97	97	11	18	18	42	350	381
		RING	103	307/3	393/3	76	166	196	13	31	32	78	117	128	270	618	752
		UBOX	4	40	40	2	18	18	1	1	1	4	4	5	11	63	64
		Unclassified	7	7	9	5	4	5	0	0	0	4	4	4	16	15	18
	E3 complex	HECT	1	10	10	0	0	0	0	0	0	0	0	0	1	10	10
		RING	58	174/1	211/1	35	69	90	19	68	82	41	60	62	153	372	446
	Total		190	616/8	865/8	129	279	369	36	185	212	138	197	217	493	1277	1671
Specificity unknown			225	-	-	240	-	-	24	-	-	1219	-	-	1708	-	-
E3 total			415	-	-	369	-	-	60	-	-	1357	-		2201	-	-

Subs.: Substrate, E3-subs.: E3-substrate specificity

C

D

FIG. 2. **Data statistics of E3Net.** *A*, statistics of the collected data for each resource. *B*, statistics of the collected data for each type (subunit composition and domain subclass) and organism of E3s. Several numbers in the human substrate consist of two values: the first indicates the number of substrates in human and the second indicates the number of viral substrates. *C*, distribution of the E3-substrate connectivity. The *upper diagram* shows the distribution of the substrate connectivity for individual E3s, and the *lower diagram* shows the distribution of the E3 connectivity for individual substrates. *D*, function distribution of the collected substrates in E3Net for 24 selected GO-BP terms. The ratio indicates the proportion of substrates associated with a given term among all proteins associated with the term.

83% of estimated E3s. The majority of E3 complexes are Cullin-based RING-type E3s including SCF (Skp1-Cul1-F box), ECS (ElonginB/C-Cul2/5-SOCS box), BCR (BTB box-Cul3-Rbx1), and DCX (DDB-Cul4-X box) (39). We determined the specificity of the complexes with their substrate recognition components, *e.g.* F box-, SOCS box-, BTB box-, and X box-containing proteins. The only E3 complex that belongs to non-RING type E3 in our database is human E6-E6AP complex. E6AP (UBE3A) recognizes and ubiquitinates substrates with or without an adaptor protein. For several substrates, it constructs the E3 complex in conjunction with E6 to exhibit ubiquitin ligase activity. Our data includes 16 single-subunit E3s that do not fall into any major domain subclasses. These E3s are reported to have ligase activities via other domains, *e.g.* A20 Znf domain of A20 protein (40), C/H1 domain of p300 (EP300) (41), C-terminal caspase-like domain of Paracaspase

(42), amino acid 41–85 region of E4F1 (43), and Cys-plus-His-rich region of Kaposi's sarcoma-associated herpesvirus RTA (44). These E3s are categorized into the unclassified category. During the text-mining process, several virus E3s, such as immunodeficiency virus type 1 (HIV-1) Virion infectivity factor (Vif), forming a Cul5-based E3 complex are mapped to multiple UniProt accessions because they have many homologous proteins in similar virus types. To avoid redundancy, we take only one entry for the viral E3s. There can be a new consensus or variations to define the subclasses of E3 ligases because the research in this field is growing. We will try to update the major change of the classification of E3 ligases. Furthermore, E3Net is designed to support users to organize more specific subclasses with domain-specific information in addition to the major subclasses of E3 ligases. E3Net provides interfaces to explore associations between E3 ligases

and function terms as described under "Output and Analysis Interface" under "Experimental Procedures." A list of associated domains for any E3 can be found in the E3/substrate report page, and the list of E3s for each domain can be sorted out in the domain (function) term report page. For example, if a specific family of E3 ligases that have IBR domains is of interest, one can search for "IBR domain" in the function report page and obtain the list of E3 ligases like LUBAC complex, parkin, ARIH2, RNF144B, and so on. From this domain-associated information in E3Net, users may discover more specific domain families of E3 ligases and their domain-specific characteristics.

Our E3-substrate specificity data shows a high ratio of E3s ubiquitinating multiple substrates and a relatively low ratio of substrates regulated by multiple E3s. Fig. 2C shows that ~44.8% of E3s interact with multiple substrates, whereas ~19.8% of substrates interact with multiple E3s. In our data, the E3s showing the highest substrate connectivity are human SMURF1 and yeast RSP5, which interact with 92 and 90 substrates, respectively. The high throughput identification methods might provide a chance to get high substrate connectivity of the SMURF1 and the RSP5. For the most of E3s, however, their substrate information is come from heterogeneous individual experiments. Human CBL and CHIP (STUB1), for example, ubiquitinate 38 and 35 substrates, respectively, and these relations are collected from more than 130 different resources for each E3. The substrates showing high E3 connectivity are relatively small, but we can find a typical example like human p53, which is identified to be regulated by 23 different E3s. High substrate connectivity of an E3 indicates that E3 might be a key regulatory protein in diverse cellular processes.

Although E3Net integrates various resources, the collected substrates still cover a small section of estimated ubiquitination targets. Considering that the human genome encodes 20,000–25,000 proteins and ~80% of proteins are degraded (1, 2), 1590 human substrates collected in E3Net cover only 6−8% of the entire estimated substrates. We wanted to check whether we could build an E3-mediated regulatory network covering most cellular processes with this largest but still limited number of E3-substrate relations. We inspected the functional distribution of the collected substrates for the GO biological process (GO-BP). There are 6231 GO-BP terms that have one or more associated UniProt accessions in human, and human substrates in E3Net are enriched with 5172 terms. When we consider 19 GO-BP terms whose semantic distances from the root term (GO:0008150; biological_process) are one, our substrates are enriched with 17 of these terms. The two terms that are not enriched with the substrates include a very small member proteins: GO:0015976 (carbon utilization) associated with three proteins and GO:0007587 (sugar utilization) associated with four proteins. We then assessed the degree of associations between GO-BP terms and our substrate data in three major organisms including human,

mouse, and yeast. To do this, we selected 24 representative GO-BP terms in a manner that keeps the balance of the number of member proteins and minimize the overlap of associated proteins in the selected terms. Approximately 80% of substrates in three organisms could be tested in this analysis. Fig. 2D shows that the collected substrates are enriched with 24 representative GO-BP terms without severe bias. These results imply that, even with this partial substrate map, one can derive the characteristics of E3-mediated regulation networks over most parts of cellular processes.

Functional Association of E3-substrate Network—On the basis of comprehensively collected E3-substrate relations over biological processes, we could identify the diverse functional characteristics of E3-specific substrate groups. First we investigated the general aspects of the functional specificity associated with E3-specific substrate groups by using functional enrichment analysis. The functional specificity of E3-specific substrate groups could be highlighted by the comparison with other protein groups. For the comparison, we selected 66 human E3s having three or more specific substrates. For each of 66 E3s, two different random protein groups and E3-interacting protein (E3-PPI) groups were generated, where the amount of member proteins is the same as the corresponding E3-specific substrate group. Each random protein group contains sets of proteins randomly selected from different data sources: random substrate groups selected from all collected substrates in E3Net and random UniProt groups selected from UniProtKB/Swiss-Prot. Each E3-PPI group contains interacting partners of each E3 found in the combined protein-protein interaction database (45). These interacting partners are not necessarily the substrates of E3. A protein group is considered significant for a given functional category when it is enriched with one or more function terms in the functional category with a *p* value below the selection threshold. The ratio of significantly enriched E3-specific substrate groups is compared with the results of other protein groups. Each compared protein group is generated five times, and the average value of the ratio is used for the comparison. Fig. 3A shows the comparison results for 12 selected functional category resources. In this comparison, the selection threshold is tightened to a *p* value below 0.001 to highlight the difference in functional specificity between E3-specific substrate groups and random protein groups. In most functional categories, E3-specific substrate groups show a much higher ratio of significantly enriched groups than those of random protein groups; 44.2% of E3-specific substrate groups are significant on average over 12 functional category resources, which is superior to the value of other protein groups (E3-PPI: 29.1%, random substrate: 8.8%, and random UniProt: 5.7%). This predominance is maintained in various *p* value cut-offs from $p < 0.1$ to $p < 1.0E-10$. This result, especially for the comparison with E3-PPI, indicates that the members of many E3-specific substrate groups show significant functional specificity. The group of interacting pro-
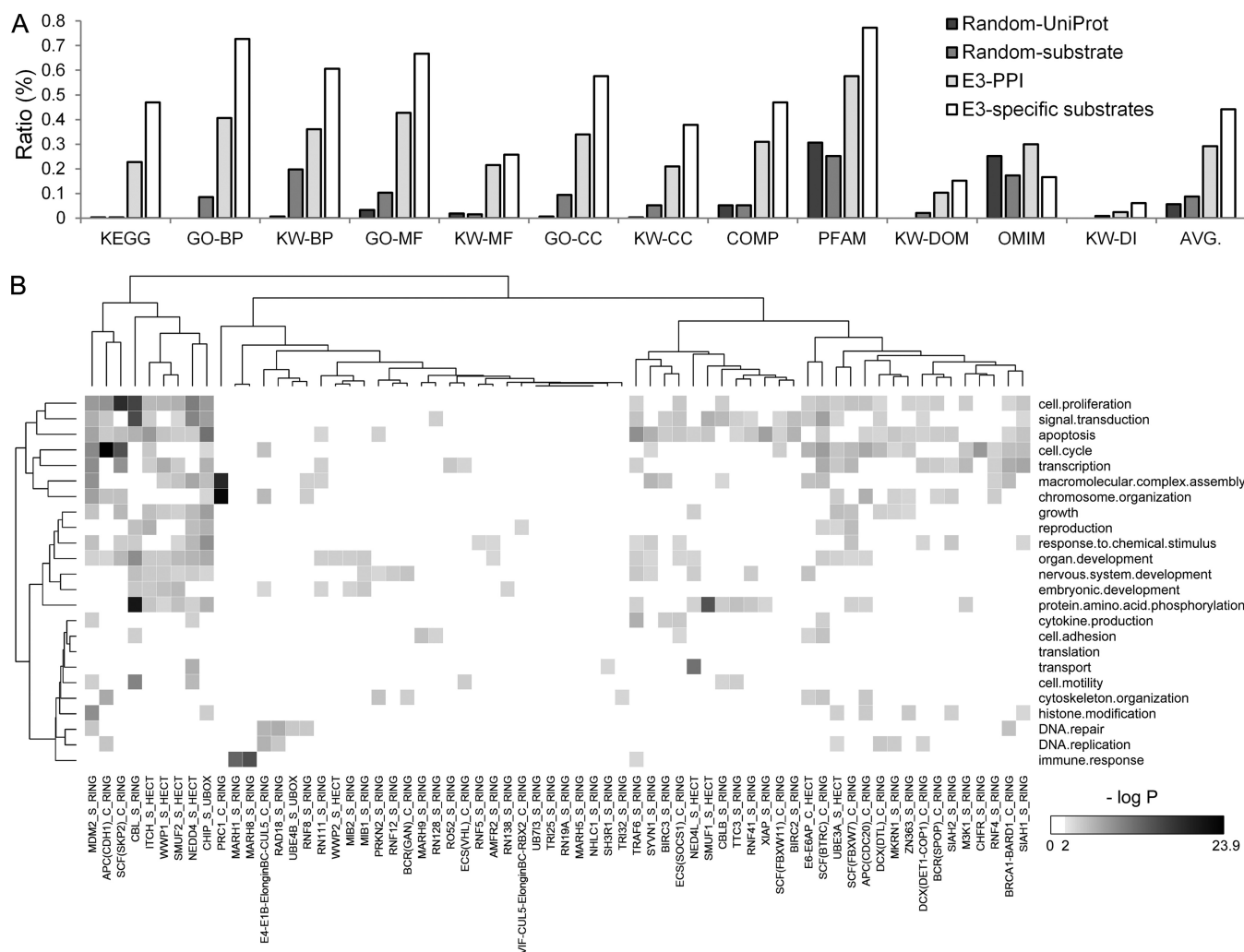
FIG. 3. **Modular features of E3-specific substrate group in their functions.** *A*, the proportion of significantly enriched ($p < 0.001$) protein groups with specific function terms in KEGG pathway (KEGG), GO biological process (GO-BP), biological process keywords of UniProt (KW-BP), GO molecular function (GO-MF), molecular function keyword of UniProt (KW-MF), GO cellular component (GO-CC), cellular component keyword of UniProt (KW-CC), protein complex (COMP), protein family (PFAM), domain keyword of UniProt (KW-DOM), disease (OMIM), and disease keyword of UniProt (KW-DI). *B*, the heat map representing the degree of association between 66 human E3-specific substrate groups and 24 selected GO-BP terms. The strength of the color in the heat map indicates negative logarithmic scale of p value calculated by enrichment analysis of given E3-specific substrate group and GO-BP term. Only significantly enriched results with p value below 0.01 are colored in the diagram.

teins is widely accepted as a functionally correlated module and is usually enriched highly with specific function terms. Therefore, our results suggest that the average functional correlation of E3-specific substrate groups is higher than what is expected for the protein-protein interaction groups. The superiority of the functional specificity of E3-specific substrate groups is increased when it is compared in the KEGG and GO categories, especially in GO-BP, which are well classified and more specific to cellular processes than other functional categories. Conversely, the ratio of significant group was not so different or even lower when the E3-specific substrate group is compared in the disease categories like KW-DI or OMIM. These functional categories are describing high order phenotypes whose terms may consist of genes in

heterogeneous cellular functions. The relatively high ratio of random UniProt groups in OMIM can be caused by the fact that a considerable number of terms in these categories have small member proteins, thereby decreasing the enrichment p value, even though a relatively small number of random proteins are matched to the function term.

We further inspected the characteristics of functional association of individual E3-specific substrate groups. There are a set of E3-specific substrate groups showing extremely high specificity to a particular function term. For example, the substrate group of human CBL is enriched with "tyrosine kinase" in Pfam (PF07714; $p = 3.94E-32$). Human CBL regulates 38 substrates, and 20 of them are enriched with the term. *Arabidopsis* COP1-SPA1 complex regulates eight sub-

strates, and all of them are enriched with a "red or far-red light signaling pathway (the phytochrome signaling pathway)" in GO-BP (GO:0010017; $p = 1.37E-17$). The yeast APC(CDC20) complex regulates 10 substrates, and all of them are enriched with "cell cycle phase" in GO-BP (GO:0022403; $p = 3.94E-9$), and nine of them are enriched with "M phase" (GO:0000279; 8.54E-8). In our data, 141 E3s have three or more specific substrates, and for 74 E3s, all of their substrates are significantly enriched with one or more particular function terms with a p value less than 0.001 (104 E3s with $p < 0.01$). The whole enrichment results for 296 E3s with a p value below 0.001 are listed in supplemental Tables S1–S7.

To get a comprehensive functional implication of E3-substrate network, we tried to associate almost all available functional categories with each E3-specific substrate group. Because we prepared the large compendium of functional categories, it is expected that multiple function terms would associated with each E3-specific substrate group. In fact, on the average, one can find 24 significantly enriched ($p < 0.001$) function terms for an E3-specific substrate group with three or more substrates in E3Net. In this inspection, we also notified that an E3-specific substrate group could be subdivided by their specific function terms. For example, the 28 of 38 human CBL substrates can be subdivided into three significantly enriched cellular functions. Twenty of them are enriched with "tyrosine kinase" in Pfam ($p = 3.94E-32$). Six substrates are enriched with "neurotrophin signaling pathway" in KEGG ($p = 1.30E-2$), sharing one substrate with "tyrosine kinase." Three of the remaining members are enriched with "cytokine receptor activity" in GO-MF (GO:0004896, $p = 3.47E-2$). The other combination of 18 substrates are enriched with "endocytosis" (11 substrates; $p = 6.60E-7$) and "focal adhesion" (11 substrates; $p = 2.82E-6$). These two groups overlap each other with four substrates, which are the members of the tyrosine kinase family. We can figure a compact functional repertoire of human CBL with this analysis. Further meaningful findings of functional implication will be possible with different approaches enabled by E3Net. Actually, the interface of E3Net supports this type of analysis for the functional implications of E3-specific substrate groups with a comprehensive view. One of the purposes of E3Net is to delineate the network of cellular functions regulated by the specific E3s.

To show an example of the comprehensive view of functional network in E3Net, we built a heat map summarizing the associations between 66 human E3-specific substrate groups having three or more substrates and function terms in GO-BP (Fig. 34B). In the heat map, each cell indicates the degree of functional specificity for given E3 and function term. Both row and column of the heat map are clustered by a hierarchical clustering method with the Manhattan distance and Ward's criterion using R package. This analysis shows how well E3Net will help to find the characteristics of functional implication for the E3-substrate network. Several E3-specific substrate groups show highly significant enrichment scores for a few particular function terms; the PRC1 complex, which includes several histone H2A variants as its substrates, is enriched with "chromosome organization" (GO:0051276; $p = 6.33E-22$) and "macromolecular complex assembly" (GO:0065003; $p = 6.43E-17$); the APC(CDH1) complex, which is known to regulate cell cycle progression through mitosis and G_1 phase (46), is enriched with "cell cycle" (GO:0007049; $p = 1.03E-21$); CBL, which ubiquitinates many protein kinases, is enriched with "protein amino acid phosphorylation" (GO:0006468; $p = 2.64E-19$), "cell proliferation" (GO:0008283; $p = 1.51E-13$), and "signal transduction" (GO:0007165; $p = 5.12E-13$); and the SCF(SKP2) complex, which ubiquitinates and degrades target proteins involved in cell cycle progression, is enriched with "cell proliferation " (GO:0008283; $p = 9.53E-16$) and "cell cycle" (GO:0007049; $p = 1.08E-11$). These E3-specific substrate groups are also significantly enriched with specific cellular localization terms in GO-CC; PRC1 is enriched with "nucleosome" (GO:0000786, $p = 4.43E-32$); APC(CDH1) is enriched with "nucleus" (GO:0005634; $p = 7.38E-7$) and its child term, "spindle" (GO:0005819; $p = 1.86E-14$); CBL is enriched with "plasma membrane" (GO:0005886; $p = 1.93E-9$); and SCF(SKP2) is enriched with "nucleus" (GO:0005634; $p = 7.85E-13$) and its child term, "nucleoplasm" (GO:0005654, $p = 5.51E-15$). The above cases show that the functional specificity of E3-specific substrate groups becomes clearer with the related results from two different categories: GO-BP terms for the cellular process and GO-CC terms for the cellular localization. We can find the E3s that are involved in many different cellular functions by using the comprehensive resource for functional categories in E3Net. E3s such as MDM2 and CHIP (STUB1) are noted to have relatively high associations with many function terms instead of showing extremely significant enrichment p value for a few function terms. Conversely, function terms including "cell cycle" (GO:0007049), "apoptosis" (GO:0006915), and "signal transduction" (GO:0007165), where ubiquitination-dependent regulation plays a vital role, appear to be highly associated with many E3s. Interestingly, we found that HECT-type E3s such as ITCH, WWP1, SMURF2, and NEDD4 are clustered together with similar function terms, although their substrate specificities are different, indicating high functional specificity of HECT-type E3s (supplemental Fig. S1). This is an example that the comprehensive integration of individual E3-substrate specificities and functional categories can be also useful to extract the common functional characteristics of the group of E3s.

Protein degradation plays an important role in regulating diverse cellular processes by determining the cellular concentrations of proteins. In the protein degradation process, a set of target proteins that shares E3 can be modularized into a group by means of their regulating E3. Our analysis suggests that a considerable number of E3s may regulate single or multiple cellular functions in a highly specific manner. These relations can be assembled as an E3-substrate-function net-
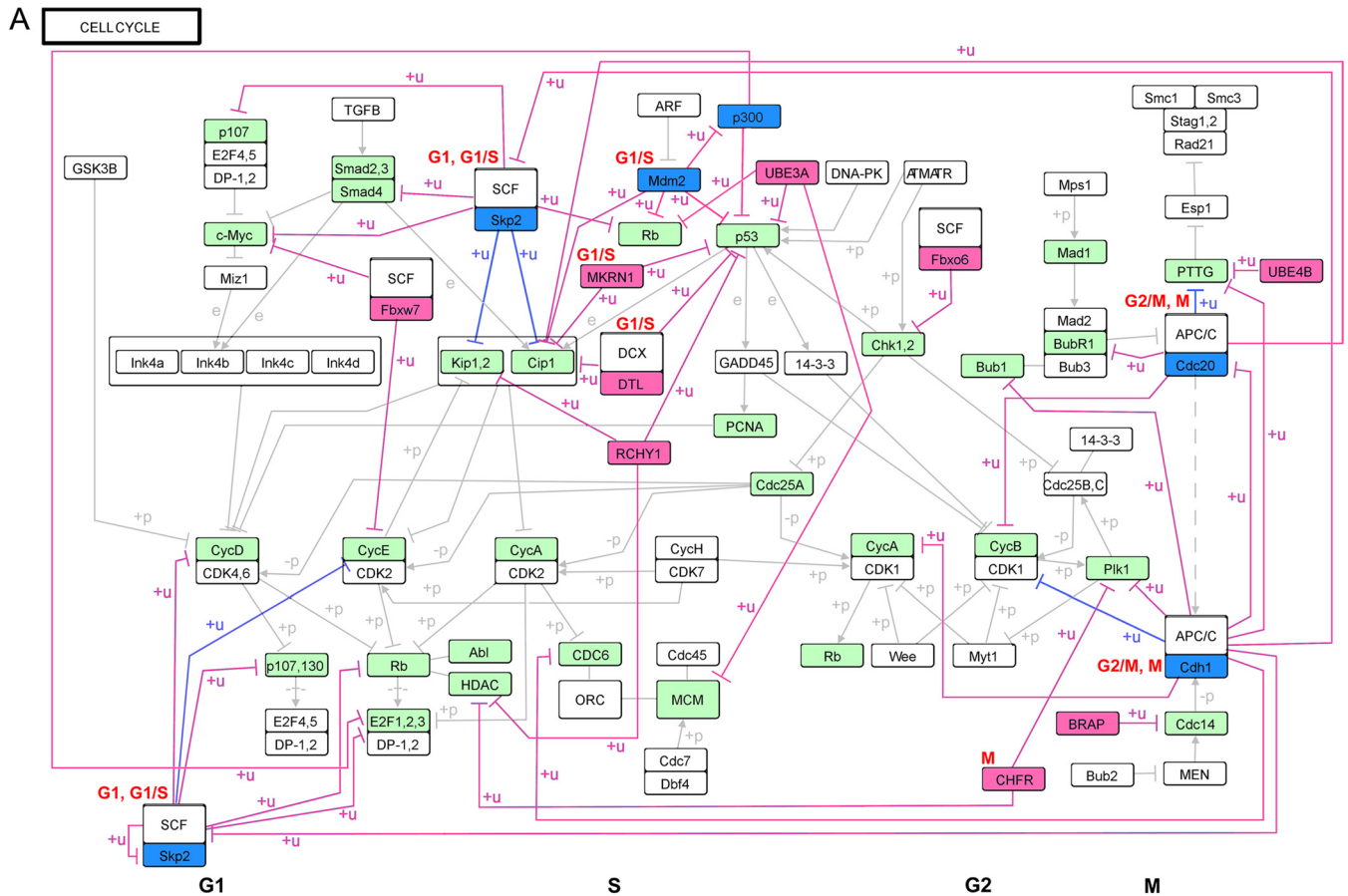
work, which provides a comprehensive map of the E3-mediated function regulations. The significant relations among E3s, substrates, and their associated function terms can be explored and summarized by using navigation and analysis tools embedded in E3Net. E3Net provides a valuable resource for in-depth investigation and systematic analysis of ubiquitination-dependent regulatory mechanisms. The data and analysis results in E3Net have great potential to capture the functional implications of an E3-specific substrate group and organize regulatory patterns of multiple E3-specific substrate groups in a comprehensive view. Exemplary analysis procedures using the embedded tools in E3Net are demonstrated in the following sections.

Analysis Framework Case 1: Integration of the Pathway-specific E3-mediated Regulation—The integrative analysis of E3-substrate networks and biological pathways can highlight significant E3-mediated regulatory modules involved in the pathways. Here we show an analysis framework of E3Net with an example of the cell cycle process. In the function term search interface of E3Net, a user can search the pathway of their interest easily with the keywords "human cell cycle," for example. The function report for KEGG then visualizes the cell cycle diagram, which contains five ubiquitination relations between three E3s and five substrates. In fact, ubiquitination is one of the major post-translational modifications in controlling progression between and within cell cycle phases (46). However, ubiquitination-dependent regulations are not described sufficiently in current KEGG cell cycle. For example, the SCF(SKP2) complex is known to regulate E2F1 during the S phase (46), but this relation is not found in the KEGG cell cycle. In this case, both the E3 and the substrate are involved in the pathway, but there is no ubiquitination relation between them. In other case, even E3 does not appear in the KEGG cell cycle. For example, the SCF(FBXW7) complex is known to regulate CCNE1 and MYC during the S phase (46), but the SCF(FBXW7) complex is not found in the KEGG cell cycle.

E3Net supports to find this valuable E3-pathway relation by providing integrated E3-substrate networks and their associated pathways. With E3Net, we analyzed 14 E3-specific substrate groups enriched with human cell cycle in KEGG ($p <$ 0.05), and 40 E3-substrate relations are identified in the KEGG cell cycle; 23 relations for five E3s are already involved in the cell cycle, and 17 relations for nine E3s are newly added. These findings are summarized in Fig. 4*A*. In this ubiquitination-enriched cell cycle, potential ubiquitination-mediated regulatory cores are found such as Cip1 (CDKN1A) and p53 (TP53); Cip1 and p53 are ubiquitinated by five and four cell cycle related E3s, respectively. The phase-specific regulatory characteristics of E3s in KEGG cell cycle can be organized by using the function of E3Net as shown in Fig. 4*B*. The function report for the KEGG cell cycle provides a list of significantly enriched E3-specific substrate groups, and each E3 is linked to the E3 report. The E3 report provides a list of significantly enriched function terms in various functional category re-

sources. The inspection of theses diverse sources of information enables better understanding of functional implications. The function terms in GO, for example, can specify the detailed parts of the cell cycle pathway. One can select the information about the cell cycle phases associated with a given E3 among many other function terms by using the filtering function of the E3Net reports, and integrate it with the search result of the pathway. Similar to the E3 report, the function report also provides an interface for users to reorganize enrichment results with the function terms of their interest. By using this interface in the function report of the KEGG cell cycle, we could capture the phase-specific characteristics of seven E3s; SCF(SKP2) complex, MKRN1, DCX(DTL) complex, and MDM2 are G_1 phase- or G_1/S transition-specific E3s, and APC(CDH1) complex, APC(CDC20) complex, and CHFR are M phase- or G_2/M transition-specific E3s. Such phase-specific regulations of E3s have been reported in a number of research articles (46, 47), but it is difficult to extract these characteristics in the current public pathway databases. In this way, one can enrich and highlight the ubiquitination-dependent regulations to the biological pathways. The integrative view of E3-substrate network and biological pathways provided by E3Net helps a user to organize pathway-specific E3-mediated regulatory modules and their functional implications.

Analysis Framework Case 2: Integration of the Protein-specific E3-mediated Regulation—A protein can be regulated by multiple E3s, and the functional specificities of these E3s can be used to enrich the functional implication of the protein. E3Net can facilitate this process with its embedded functions in the reports for E3, substrate, and function. The substrate report of E3Net provides the information about specific E3s for a selected substrate and a graphical interface visualizing the network of all E3s specifying the substrate. The network can be expanded to generate the other substrate of these E3s, and the associated function terms for each E3 can be obtained from the linked E3 report. In addition, the functional enrichment analysis for the user-selected proteins in the expanded network can be done directly in the substrate report interface. In this exemplary analysis for p53, the interface of substrate report shows 23 E3s regulating p53 and their substrate networks (Fig. 5). Moreover, the diverse function terms of each E3 can be inspected with the help of filtering function in the E3 report and the substrate report. In this inspection of p53-related E3 networks, we could categorize the most popular function terms of p53-related E3-substrate networks. We could summarize the function of nine p53 regulating E3s as transcription, DNA damage response, and apoptosis: MDM2, CHIP, ZN363 (RCHY1), and WWP1 for transcription; MDM2, DCX(DTL) complex, E4-E1B-ElonginBC-CUL5 complex, ICP0 (RL2), and MKRN1 for DNA damage response; and CHIP (STUB1), WWP1, MDM2, MKRN1, and SYVN1 for apoptosis. In this way, the enriched functions of E3s for a protein can be reorganized and summarized to some key categories. It is a possible function of E3Net to provide the abstract of func-

FIG. 4. **Enrichment and submodularization of cell cycle in terms of E3-mediated regulatory modules.** *A*, illustration of the ubiquitination-enriched pathway diagram for the KEGG cell cycle. E3s and substrates are highlighted with *blue* and *green*, respectively, and newly appended E3s and E3-substrate relations are highlighted with *pink*. Phase-specific E3s tagged with the corresponding cell cycle phases. *B*, a list of E3s whose substrate groups exhibit significant functional specificity with specific cell cycle phases.

Table B:

E3	Cell cycle phase		P-value
	GO ID	GO Name	
SCF(SKP2)_HUMAN	GO:0000082	G1/S transition of mitotic cell cycle	2.98E-11
	GO:0051318	G1 phase	2.32E-04
	GO:0031575	G1/S transition checkpoint	4.23E-03
MKRN1_HUMAN	GO:0031571	G1/S DNA damage checkpoint	1.59E-03
	GO:0031575	G1/S transition checkpoint	2.76E-03
DCX(DTL)_HUMAN	GO:0031571	G1/S DNA damage checkpoint	1.59E-03
	GO:0031575	G1/S transition checkpoint	2.76E-03
MDM2_HUMAN	GO:0031575	G1/S transition checkpoint	4.96E-03

E3	Cell cycle phase		P-value
	GO ID	GO Name	
APC(CDH1)_HUMAN	GO:0000087	M phase of mitotic cell cycle	7.31E-16
	GO:0031572	G2/M transition DNA damage checkpoint	1.83E-04
APC(CDC20)_HUMAN	GO:0000087	M phase of mitotic cell cycle	1.11E-04
	GO:0000086	G2/M transition of mitotic cell cycle	6.02E-03
CHFR_HUMAN	GO:0000087	M phase of mitotic cell cycle	8.16E-06

tional implication of the selected protein indirectly through the functional information of related E3s.

DISCUSSION

Despite numerous ubiquitination studies, current findings about E3s and substrates are still insufficient to cover the proteins involved in ubiquitination-dependent regulations. Our data collection scheme, including text mining of textual resources and integration of public ubiquitination resources, facilitates the construction of a comprehensive database for E3-substrate relations. The collected data, however, still cover only a partial map of the global E3-mediated regulatory network, especially regarding the amount of E3-substrate re-

lations. In the collected data, 22.4% (493 of 2201) of E3s and 26.1% (1277 of 4896) of substrates have specific substrates or E3s, respectively (Fig. 2B). The low coverage of E3-substrate relations is mainly caused by an abundance of undiscovered specificity. We hope that the situation will be improved by the rapid growth of published studies. Whenever it is necessary, our semi-automated data collection scheme can extract newly published data within a short time. To facilitate the continuing updating processes of the data, we carefully set up the text mining module, which manages the most time-consuming process to build the database of E3Net. The hardest part of text-mining process is to reduce the possible false positive and false negative results, which depreci-
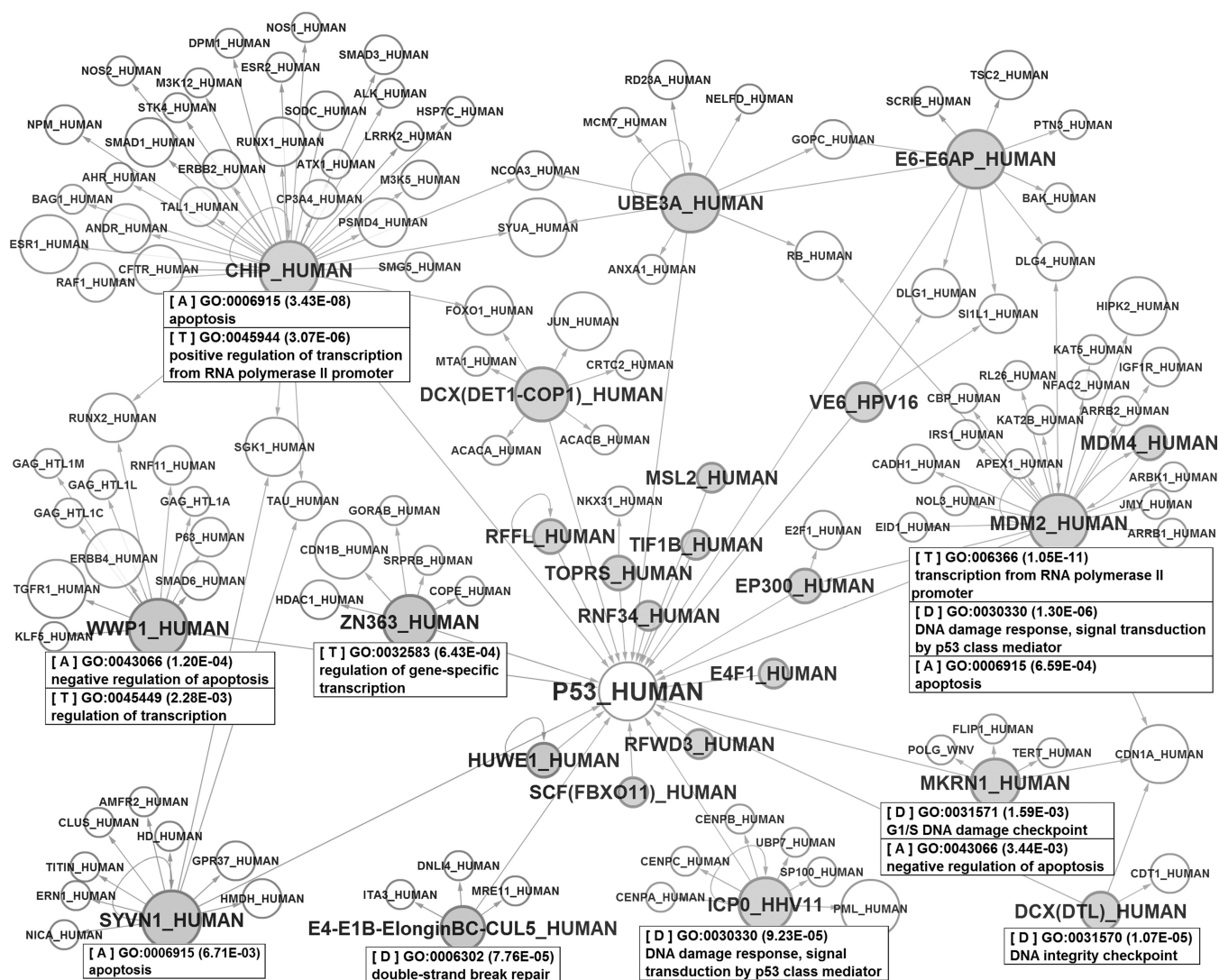
Fig. 5. **Illustration of the results from systematic searching and categorizing the functional regulations of p53.** The diagram shows the 2-hop E3-substrate network for p53, and each E3-mediated regulatory module is tagged in the box with the most significantly enriched GO-BP terms related to three major functions of p53, transcription (tagged as *T*), DNA damage response (tagged as *D*), and apoptosis (tagged as *A*). In the diagram, *gray* and *white nodes* indicate E3s and substrates, respectively. The size of each node depends on the number of its relations.

ate the quality of the collected data. To assess the confidence of the automatically mined data, manual inspection is necessary. Our text-mining module supports this process by providing evidence sentences along with the extracted E3s, substrates, and their specificities. During the text-mining procedure, we actually used this module to check the supporting sentences for each E3-substrate relation and filtered out ambiguously mined relations from textual resources. In particular, during the manual inspection process, we paid more attention to clarifying indirect ubiquitination patterns such as "P is degraded by E3," "P is a target of E3," or "P is catalyzed by E3" as shown in supplemental Table S2; these patterns were used to maximize the related data collection at the first step of text-mining step. Through the manual inspection process, we could eliminate unsuitably collected data such as the case that the E3

ligase regulates the substrate but the relation is not mediated by ubiquitination. This semi-automatic scheme can reduce manual validation efforts and provides a feasible data update method. We also tried to minimize false negatives by constructing alias data for the name of single E3/substrate proteins and E3 complexes, which were extracted from UniProt and the literature. With the alias data, we could maximize the data collection and reduce mining errors in protein name identification. Additionally, we attempted to reduce false negatives by employing different text-mining schemes for each of MEDLINE abstract and UniProt comment, because these two resources have different sentence structures. Through these efforts to maximize the data collection while maintaining reliability in a short period of data refining time, we achieved remarkable advances in quantity and quality of E3-substrate relations

compared with previous public databases, such as UbiProt and SCUD.

There can be additional concerns regarding the accuracy of the interpretation of E3Net about the functional specificity of E3 ligases having newly added discrepant information. Generally, adjusted functional interpretation for new findings can be derived automatically by the built-in functional enrichment tool of E3Net whenever the newly added substrates are updated. The adjustment reflects any change of the relative amount of proteins for specific function terms and gives the new functional specificities for all E3-specific substrate groups with their p values. However, E3Net does not provide the compromising conclusion directly for the contradictory evidences including the negation data of E3-substrate specificity. The inclusion of the accurate information for the negation data in E3Net should be a challenging task for the future work. It is not just a hard text mining issue to extract correct information from the negation data. A more critical problem is to assess the correct conclusion between contradictory interpretations. Contradictory claims should be analyzed very carefully with reliable evidence to reach a conclusion. The current E3Net system rather supports a semi-automatic way to examine the effect of discrepant findings on the functional implication after users define the contradictory E3-substrate relations carefully. Users can actively compare the functional specificities of selected E3-specific substrate groups with or without the suspected substrates by using the auxiliary functional enrichment tool embedded in the web interface of E3Net introduced under "Experimental Procedures."

The interesting question about E3 is whether it has specificity for cellular processes or functions in its regulation. With a largely increased map of E3-substrate relations and the integration of comprehensive functional information for them, we could show the possibility of E3Net to elucidate the functional specificity of E3s. A considerable number of E3s showed significant associations with cellular functions; 93.1% (459 of 493) of E3-specific substrate groups showed statistically significant enrichment score ($p < 0.01$). We introduced typical examples showing high functional specificity such as *Arabidopsis* COP1-SPA1 complex and yeast APC(CDC20) complex, whose substrates are all enriched with a particular function term with very significant p value. The remaining E3s, whose enrichment scores do not meet the criteria ($p < 0.01$), also have the potential to exhibit functional specificity with the increasing number of E3-substrate relations in the future. This can be partially examined by inspecting the pattern of currently available data. In our data, the ratio of E3s exhibiting functional specificity was higher in E3 groups having many substrates than E3 groups having relatively few substrates, with 78.7% (111 of 141) of E3s having three or more substrates and 97.3% (36 of 37) E3s having 10 or more substrates showed functional specificity with p values below 0.001. The next interesting question will be whether we can find the molecular elements that determine the functional specificity of E3. We have clustered the E3s with the pattern of their significantly enriched functions. Among the large subclasses of E3s that have designated signature domains such as HECT, RING, or U box, several HECT-type E3s showed similar patterns of associated functions (Fig. 3*B*). This similar pattern of functional specificity might not be originated from the HECT domain because it is known that the HECT domain does not directly bind to a substrate. However, the result at least implies that HECT-type E3s are supposed to share a substrate-binding mechanism that might contribute the regulation of substrates having similar functions. To discriminate this complex pattern of regulation, we may need further molecular information such as the structures of E3s and corresponding substrates, which demands extensive effort to build up. E3Net will be useful to minimize the number of targets to be investigated for this work by guiding the specific correlation among E3s, substrates, and functional categories.

In addition to E3 ligase, there is another ubiquitin ligase called E4 that is responsible for the extending of ubiquitin chains. Unlike E3 ligase, E4 ligase may not recognize specific determinants in substrates but recognize substrate-bound ubiquitin chains (48, 49). Currently, E4 ligase is not the target of E3Net because we thought the substrate specificity and the corresponding functional implication of E4 ligase looked unclear until now. Nevertheless, the information about some E4 ligases described with E3 ligases can be traced indirectly by inspecting the supporting sentences of E3 ligases provided by E3Net. In the future, if the regulation specificity mediated by E4 ligases can be build up, we may be able to include them in E3Net.

E3Net is an integrated information system constructed with massive amounts of E3-substrate relations and function terms in various functional category resources. The web interface and the embedded tools in E3Net facilitate the analysis of these relations comprehensively by providing a highly connected view of E3s, substrates, and functions. With this comprehensive view, a user can organize the E3-substrate network and the possible functional implications of E3 regulation. We expect that E3Net can be a valuable resource for in-depth investigation and systematic analysis of E3s, substrates, and their functions.

‡‡ To whom correspondence should be addressed. Tel.: 82-42-350-4318; Fax: 82-42-350-5358; E-mail: gsyi@kaist.ac.kr.

REFERENCES

1. Burger, A. M., and Seth, A. K. (2004) The ubiquitin-mediated protein degradation pathway in cancer: Therapeutic implications. *Eur. J. Cancer* **40,** 2217–2229

2. von Mikecz, A. (2006) The nuclear ubiquitin-proteasome system. *J. Cell Sci.* **119,** 1977–1984

3. Lecker, S. H., Goldberg, A. L., and Mitch, W. E. (2006) Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol.* **17,** 1807–1819

4. Haglund, K., and Dikic, I. (2005) Ubiquitylation and cell signaling. *EMBO J.* **24,** 3353–3359

5. Sagar, G. D., Gereben, B., Callebaut, I., Mornon, J. P., Zeöld, A., da Silva, W. S., Luongo, C., Dentice, M., Tente, S. M., Freitas, B. C., Harney, J. W., Zavacki, A. M., and Bianco, A. C. (2007) Ubiquitination-induced conformational change within the deiodinase dimer is a switch regulating enzyme activity. *Mol. Cell. Biol.* **27,** 4774–4783

6. Chernorudskiy, A. L., Garcia, A., Eremin, E. V., Shorina, A. S., Kondratieva, E. V., and Gainullin, M. R. (2007) UbiProt: A database of ubiquitylated proteins. *BMC Bioinformatics* **8:** 126

7. Lee, W. C., Lee, M., Jung, J. W., Kim, K. P., and Kim, D. (2008) SCUD: *Saccharomyces cerevisiae* Ubiquitination Database. *BMC Genomics* **9:** 440

8. Lee, H., Yi, G. S., and Park, J. C. (2008) E3Miner: A text mining tool for ubiquitin-protein ligases. *Nucleic Acids Res.* **36,** W416–W422

9. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **34,** D354–D357

10. Peters, J. M. (2006) The anaphase promoting complex/cyclosome: A machine designed to destroy. *Nat. Rev. Mol. Cell Biol.* **7,** 644–656

11. Lee, J. T., and Gu, W. (2010) The multiple levels of regulation by p53 ubiquitination. *Cell Death Differ.* **17,** 86–92

12. Andrews, P. S., Schneider, S., Yang, E., Michaels, M., Chen, H., Tang, J., and Emkey, R. (2010) Identification of substrates of SMURF1 ubiquitin ligase activity utilizing protein microarrays. *Assay Drug Dev. Technol.* **8,** 471–487

13. Merbl, Y., and Kirschner, M. W. (2009) Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 2543–2548

14. Hör, S., Ziv, T., Admon, A., and Lehner, P. J. (2009) Stable isotope labeling by amino acids in cell culture and differential plasma membrane proteome quantitation identify new substrates for the MARCH9 transmembrane E3 ligase. *Mol. Cell. Proteomics* **8,** 1959–1971

15. Yen, H. C., and Elledge, S. J. (2008) Identification of SCF ubiquitin ligase substrates by global protein stability profiling. *Science* **322,** 923–929

16. Seyfried, N. T., Xu, P., Duong, D. M., Cheng, D., Hanfelt, J., and Peng, J. (2008) Systematic approach for validating the ubiquitinated proteome. *Anal. Chem* **80,** 4161–4169

17. Manzano, C., Abraham, Z., López-Torrejón, G., and Del Pozo, J. C. (2008) Identification of ubiquitinated proteins in *Arabidopsis*. *Plant Mol. Biol.* **68,** 145–158

18. Zhou, T., Liang, B., Su, G. Y., Gong, W. L., Li, H. Y., Tian, L. F., He, K., Zhao, J., Man, J. H., Li, T., Li, W. H., Zhang, Z. Y., Wang, C. H., Li, A. L., Liu, H., Pan, X., Zhang, P. J., Jin, B. F., and Zhang, X. M. (2007) Identification of ubiquitin target proteins using cell-based arrays. *J. Proteome Res.* **6,** 4397–4406

19. Mayor, T., Graumann, J., Bryan, J., MacCoss, M. J., and Deshaies, R. J. (2007) Quantitative profiling of ubiquitylated proteins reveals proteasome substrates and the substrate repertoire influenced by the Rpn10 receptor pathway. *Mol. Cell. Proteomics* **6,** 1885–1895

20. Maor, R., Jones, A., Nühse, T. S., Studholme, D. J., Peck, S. C., and Shirasu, K. (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics* **6,** 601–610

21. Jeon, H. B., Choi, E. S., Yoon, J. H., Hwang, J. H., Chang, J. W., Lee, E. K., Choi, H. W., Park, Z. Y., and Yoo, Y. J. (2007) A proteomics approach to identify the ubiquitinated proteins in mouse heart. *Biochem. Biophys. Res. Commun.* **357,** 731–736

22. Gupta, R., Kus, B., Fladd, C., Wasmuth, J., Tonikian, R., Sidhu, S., Krogan, N. J., Parkinson, J., and Rotin, D. (2007) Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast. *Mol. Syst. Biol.* **3:** 116

23. Denis, N. J., Vasilescu, J., Lambert, J. P., Smith, J. C., and Figeys, D. (2007) Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. *Proteomics* **7,** 868–874

24. Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L., and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: Application in ubiquitin profiling and protein complex identification combined with *in vivo* cross-linking. *Mol. Cell. Proteomics* **5,** 737–748

25. Vasilescu, J., Smith, J. C., Ethier, M., and Figeys, D. (2005) Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. *J. Proteome Res.* **4,** 2192–2200

26. Mayor, T., Lipford, J. R., Graumann, J., Smith, G. T., and Deshaies, R. J. (2005) Analysis of polyubiquitin conjugates reveals that the Rpn10 substrate receptor contributes to the turnover of multiple proteasome targets. *Mol. Cell. Proteomics* **4,** 741–751

27. Matsumoto, M., Hatakeyama, S., Oyamada, K., Oda, Y., Nishimura, T., and Nakayama, K. I. (2005) Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **5,** 4145–4151

28. Kus, B., Gajadhar, A., Stanger, K., Cho, R., Sun, W., Rouleau, N., Lee, T., Chan, D., Wolting, C., Edwards, A., Bosse, R., and Rotin, D. (2005) A high throughput screen to identify substrates for the ubiquitin ligase Rsp5. *J. Biol. Chem.* **280,** 29470–29478

29. Kirkpatrick, D. S., Weldon, S. F., Tsaprailis, G., Liebler, D. C., and Gandolfi, A. J. (2005) Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. *Proteomics* **5,** 2104–2111

30. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21,** 921–926

31. Hitchcock, A. L., Auld, K., Gygi, S. P., and Silver, P. A. (2003) A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 12735–12740

32. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29

33. Magrane, M., and Consortium, U. (2011) UniProt Knowledgebase: A hub of integrated protein data. *Database* 2011, bar009

34. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res.* **38,** D211–D222

35. Sun, C. H., Kim, M. S., Han, Y., and Yi, G. S. (2009) COFECO: Composite function annotation enriched by protein complex data. *Nucleic Acids Res.* **37,** W350–W355

36. Amberger, J., Bocchini, C. A., Scott, A. F., and Hamosh, A. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM(R)). *Nucleic Acids Res.* **37,** D793–D796

37. Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., and Bader, G. D. (2010) Cytoscape Web: An interactive web-based network browser. *Bioinformatics* **26,** 2347–2348

38. Wilkinson, K. D., Ventii, K. H., Friedrich, K. L., and Mullally, J. E. (2005) The ubiquitin signal: Assembly, recognition and termination. Symposium on ubiquitin and signaling. *EMBO Rep.* **6,** 815–820

39. Petroski, M. D., and Deshaies, R. J. (2005) Function and regulation of Cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol.* **6,** 9–20

40. Shembade, N., Parvatiyar, K., Harhaj, N. S., and Harhaj, E. W. (2009) The ubiquitin-editing enzyme A20 requires RNF11 to downregulate NF-κB signalling. *EMBO J.* **28,** 513–522

41. Grossman, S. R., Deato, M. E., Brignone, C., Chan, H. M., Kung, A. L., Tagami, H., Nakatani, Y., and Livingston, D. M. (2003) Polyubiquitination of p53 by a ubiquitin ligase activity of p300. *Science* **300,** 342–344

42. Zhou, H., Wertz, I., O'Rourke, K., Ultsch, M., Seshagiri, S., Eby, M., Xiao, W., and Dixit, V. M. (2004) Bcl10 activates the NF-κB pathway through ubiquitination of NEMO. *Nature* **427,** 167–171

43. Le Cam, L., Linares, L. K., Paul, C., Julien, E., Lacroix, M., Hatchi, E., Triboulet, R., Bossis, G., Shmueli, A., Rodriguez, M. S., Coux, O., and Sardet, C. (2006) E4F1 is an atypical ubiquitin ligase that modulates p53 effector functions independently of degradation. *Cell* **127,** 775–788

44. Gould, F., Harrison, S. M., Hewitt, E. W., and Whitehouse, A. (2009) Kaposi's sarcoma-associated herpesvirus RTA promotes degradation of the Hey1 repressor protein through the ubiquitin proteasome pathway. *J. Virol.* **83,** 6727–6738

45. Han, Y., Sun, C. H., Kim, M. S., and Yi, G. S. (2009) Combined database system for binary protein interaction and co-complex association. *2009 International Association of Computer Science and Information Technology Spring Conference*, pp. 538–542, International Association of Computer Science and Information Technology

46. Fasanaro, P., Capogrossi, M. C., and Martelli, F. (2010) Regulation of the endothelial cell cycle by the ubiquitin-proteasome system. *Cardiovasc. Res.* **85,** 272–280

47. Argentini, M., Barboule, N., and Wasylyk, B. (2000) The contribution of the RING finger domain of MDM2 to cell cycle progression. *Oncogene* **19,** 3849–3857

48. Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H. D., Mayer, T. U., and Jentsch, S. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* **96,** 635–644

49. Crosas, B., Hanna, J., Kirkpatrick, D. S., Zhang, D. P., Tone, Y., Hathaway, N. A., Buecker, C., Leggett, D. S., Schmidt, M., King, R. W., Gygi, S. P., and Finley, D. (2006) Ubiquitin chains are remodeled at the proteasome by opposing ubiquitin ligase and deubiquitinating activities. *Cell* **127,** 1401–1413