The human phosphotyrosine signaling network: Evolution and hotspots of hijacking in cancer

Lei Li, Chabane Tibiche, Cong Fu, et al.

Genome Res. published online December 22, 2011
Access the most recent version at doi:10.1101/gr.128819.111

The human phosphotyrosine signaling network: Evolution and hotspots of hijacking in cancer

Lei Li,[1,2,9] Chabane Tibiche,[3,9] Cong Fu,[3,9,10] Tomonori Kaneko,[1] Michael F. Moran,[2,4,5] Martin R. Schiller,[6] Shawn Shun-Cheng Li,[1,7,11] and Edwin Wang[3,8,11]

[1]*Department of Biochemistry and the Siebens-Drake Research Institute, Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario N6A 5C1, Canada;* [2]*Molecular Structure and Function Program, Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G 1X8, Canada;* [3]*Computational Chemistry and Bioinformatics Group, Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec H4P 2R2, Canada;* [4]*Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada;* [5]*McLaughlin Centre for Molecular Medicine and Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L7, Canada;* [6]*School of Life Sciences, University of Nevada, Las Vegas, Nevada 89154-4004, USA;* [7]*Children's Health Research Institute, London, Ontario N6C 2V5, Canada;* [8]*Center for Bioinformatics, McGill University, Montreal, Quebec H3G 0B1, Canada*

Phosphotyrosine (pTyr) signaling, which plays a central role in cell–cell and cell–environment interactions, has been considered to be an evolutionary innovation in multicellular metazoans. However, neither the emergence nor the evolution of the human pTyr signaling system is currently understood. Tyrosine kinase (TK) circuits, each of which consists of a TK writer, a kinase substrate, and a related reader, such as Src homology (SH) 2 domains and pTyr-binding (PTB) domains, comprise the core machinery of the pTyr signaling network. In this study, we analyzed the evolutionary trajectories of 583 literature-derived and 50,000 computationally predicted human TK circuits in 19 representative eukaryotic species and assigned their evolutionary origins. We found that human TK circuits for intracellular pTyr signaling originated largely from primitive organisms, whereas the inter- or extracellular signaling circuits experienced significant expansion in the bilaterian lineage through the "back-wiring" of newly evolved kinases to primitive substrates and SH2/PTB domains. Conversely, the TK circuits that are involved in tissue-specific signaling evolved mainly in vertebrates by the back-wiring of vertebrate substrates to primitive kinases and SH2/PTB domains. Importantly, we found that cancer signaling preferentially employs the pTyr sites, which are linked to more TK circuits. Our work provides insights into the evolutionary paths of the human pTyr signaling circuits and suggests the use of a network approach for cancer intervention through the targeting of key pTyr sites and their associated signaling hubs in the network.

[Supplemental material is available for this article.]

An important feature that distinguishes multicellular metazoans from unicellular organisms is that the former possess elaborate regulatory and signaling systems for divergent functions (Putnam et al. 2007; King et al. 2008; Manning et al. 2008; Pincus et al. 2008; Lim and Pawson 2010). It has been suggested that the development of complex regulatory systems, such as molecular networks that are mediated by tyrosine phosphorylation, plays an important role in the appearance of multicellularity and the coordination of complex morphogenetic events in eumetazoans (Weiss and Littman 1994; Tan et al. 2009b). Therefore, understanding of the evolutionary paths of cellular regulatory networks is important for the evolution of animal complexity and for the achievement of a system-level understanding of human development and the pathophysiology of complex diseases (Boran and Iyengar 2010).

[9]These authors contributed equally to this work.
[10]Present address: State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130023, P. R. China.
[11]Corresponding authors.
E-mail sli@uwo.ca.
E-mail edwin.wang@cnrc-nrc.gc.ca.
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.128819.111.

Tyrosine-kinase-mediated phosphotyrosine (pTyr) signaling has been used as a model to promote the understanding of the evolution of signaling networks and cell–cell communications in multicellular animals (King et al. 2003; Nichols et al. 2006; Grimson et al. 2008). The core machinery, or the tyrosine kinase (TK) circuit in pTyr signaling, consists of a TK that functions as the "writer" to phosphorylate a Tyr residue in a protein substrate, an SH2/pTyr-binding (PTB) domain that acts as a "reader" to recognize the modification, and a protein tyrosine phosphatase (PTP) that plays the role of an "eraser" to terminate the kinase signal (Pincus et al. 2008). Previous studies have focused on the evolution of individual components of the pTyr signaling circuit, but it is clear that the evolution of these components is highly dependent on the other components and their formation of functional circuits, which are further integrated into a more complex pTyr signaling network (Nichols et al. 2006; King et al. 2008; Pincus et al. 2008; Tan et al. 2009b; Gough and Foley 2010). Therefore, it is important to understand how the elaborate cell–cell communication and tissue-specific signaling mechanisms that are found in the human pTyr signaling network functionally evolved from the TK circuits. The examination of the TK circuit as a functional unit may yield insights that are unattainable by studying the individual components separately.

To our knowledge, no study has previously directly addressed how human pTyr signaling circuits and networks evolved from premetazoans and unicellular metazoans. In this study, we investigated

the evolution of the human pTyr signaling system by analyzing the evolution of the human TK circuits. We classified the TK circuits into discrete signaling routes that are associated with intracellular, inter-/extracellular, and tissue-specific signaling and then examined the path of human TK circuit evolution by comparing the human TK circuits to orthologous circuits from 19 representative organisms and investigated (1) in which organisms the human TK circuit components originated and (2) which evolutionary paths were preferentially used for the formation of circuits that are responsible for intracellular, inter-/extracellular, and tissue-specific pTyr signaling. These 19 organisms are either well-known model organisms or key species in evolution. For example, *Monosiga brevicollis*, which is the closest-known relative to metazoans, is an organism that can provide clues regarding the genesis of the animal kingdom, while *Nematostella vectensis*, which is the simplest, most primitive animal with a tissue grade of organization, is an emerging model in which to study the evolution of the ancient cell–cell communication system. These analyses revealed key steps and distinct trajectories in the evolution of the different human pTyr signaling routes. We also show that cancer signaling preferably exploited promiscuous pTyr sites on multifunctional substrates.

Results

Assembly of the human TK circuit data sets and the identification of orthologous circuits in model organisms

The simplified TK circuit that was analyzed in this study is composed of a TK, a substrate containing an experimentally verified pTyr site, and a protein containing an SH2/PTB domain (Fig. 1A). Because of the generally broad specificity of a PTP (Moorhead et al. 2009) and the paucity of information on specific PTP–substrate interactions, we excluded PTPs from the TK circuits, even though they are an integral part of the tyrosine kinase signaling system (Pincus et al. 2008). We also excluded dual-specificity kinases (Lindberg et al. 1992) because they can phosphorylate both tyrosine and serine/threonine residues, which makes the signaling outcome uncertain. Moreover, no reliable algorithm is currently available for the prediction of PTP–substrate or dual-specific kinase–substrate relationships.

To analyze the human TK circuits, we assembled two data sets (see Methods): (1) a curated human TK circuit data set (583 circuits) (Supplemental Data File 1), and (2) a computationally predicted data set (~50,000 circuits) in which the pTyr (and, therefore, the substrate) is experimentally verified but the corresponding kinase and SH2/PTB domain are predicted using existing algorithms (Supplemental Data File 2).

To simplify the analysis, the circuit components (i.e., TK, SH2/PTB, and substrate) were grouped by protein families based on common binding properties and/or biological functions (e.g., CRK and CRKL are placed in the same group; see Methods) (Huang et al. 2008). Moreover, the pTyr sites from different members of the same protein family that are conserved (based on multiple sequence alignment by MAFFT) (Katoh et al. 2005) were treated as a single pTyr site to assemble the corresponding TK circuit.

For each human TK circuit, we identified the orthologs of the circuit components in 19 selected species (Wall et al. 2003; Elango et al. 2009; Cherry 2010). Because abundant information for the TK circuits is available in humans and yet this information is much sparser in other organisms (Tan et al. 2009a), we started from a human TK circuit and back-predicted the orthologous circuit in a target species using the Roundup database (see Methods). The selected 19 species are: *Saccharomyces cerevisiae*, *M. brevicollis*, *N. vectensis*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Bos taurus*, *Canis familiaris*, *Danio rerio*, *Gallus gallus*, *Gasterosteus aculeatus*, *Macaca mulatta*, *Monodelphis domestica*, *Mus musculus*, *Ornithorhynchus anatinus*, *Oryzias latipes*, *Pan troglodytes*, *Rattus norvegicus*, *Takifugu rubripes*, *Xenopus tropicalis*, and *Homo sapiens* (Fig. 1B).

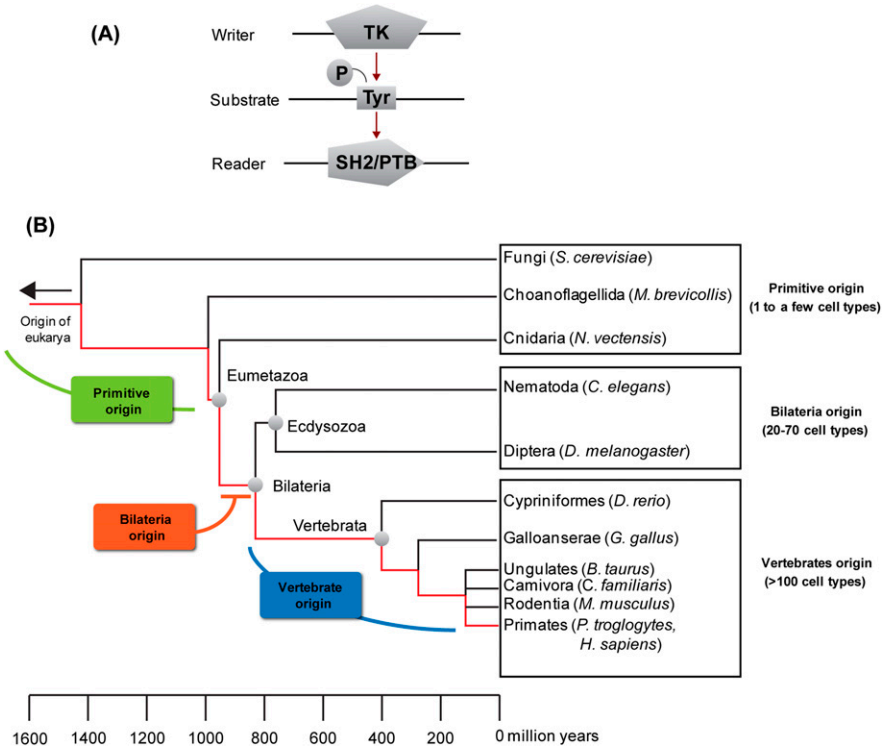

Figure 1. A schematic drawing of the tyrosine kinase (TK) signaling circuit and grouping of species used for the analysis of the human TK circuit evolution. (*A*) A simplified TK circuit that comprises a kinase that functions as the "writer," a substrate that contains a specific tyrosine (Tyr) site for phosphorylation, and an SH2 or PTB domain-containing protein that functions as the reader of the phosphorylated Tyr (pTyr). (*B*) A phylogenetic tree of representative species included in the analysis of the evolution of human TK circuit evolution. They were classified into three groups. (1) The primitive group, including *S. cerevisiae*, *M. brevicollis*, and *N. vectensis*, represents organisms branched from the human lineage before the emergence of bilaterians. (2) Ecdysozoans, including *C. elegans* and *D. melanogaster*, represent organisms between the branches of primitive organisms and vertebrates. (3) Vertebrates that contain the vertebrate animals we analyzed in this study, representing different branches of vertebrate evolution. The human ancestry line is shown in red. Based on this grouping, we defined origins of human protein orthologs as follows. If an ortholog is found in a primitive organism, it is assigned a primitive (or P)-origin. Similarly, if an ortholog is identified in an ecdysozoan organism but not a primitive organism, it is considered to have originated from the bilateria (or B-origin). Orthologs found only in vertebrates are assigned a V-origin. The branch times were estimated based on data from the TimeTree server (Hedges et al. 2006).

One approach that was used to investigate the evolutionary origin of human pTyr signaling utilized a comparison between the human TK circuits and the orthologous circuits in each selected species. However, some species contain only a small number of circuits that are orthologous to the 583 curated circuits, which makes it impossible to perform a statistically meaningful analysis on the evolution of human TK circuits on a per-species basis. To overcome this limitation, we classified the selected species into three groups that were based on previous studies that showed the statistical robustness of this scheme of species classification (Fig. 1B; Putnam et al. 2007; King et al. 2008). We classified the species into primitive, bilaterian, and vertebrate groups. The primitive group includes *S. cerevisiae*, *M. brevicollis*, and *N. vectensis*. The bilaterian lineage is represented by the ecdysozoa, *C. elegans*, and *D. melanogaster*. All of the vertebrate species are placed in the same group. These three groups correlate grossly with the degree of organismal complexity, which is measured by the number of cell types. Specifically, primitive organisms contain one or a few cell types, bilateria contain 20–70 cell types, and vertebrates are the most complex and contain more than 100 cell types (Vogel and Chothia 2006).

Based on this grouping scheme, we defined the origins of human protein orthologs as follows. If an ortholog is identified in an organism of the primitive group, we define the protein as a primitive (or P)-origin ortholog. If an ortholog is identified in a bilaterian organism but not in a primitive organism, it is assigned to a bilaterian (or B) origin. If an ortholog is identified only in a vertebrate, it is assigned to a vertebrate (or V) origin. We also assigned an evolutionary origin to each human TK circuit based on the earliest period in which all three components of the circuit co-existed. For example, the SRC-CD19(pTyr500)-PIK3R1 circuit (i.e., written in the order of writer-substrate-reader) is considered to have originated from vertebrates because the circuit was not complete until a CD19 ortholog appeared in the vertebrate *B. taurus*, despite the presence of SRC and PIK3R1 orthologs in *M. brevicollis* (a primitive organism).

Different signaling routes underlie intracellular, inter-/extracellular, and tissue-specific pTyr signaling

To examine the role of the kinases and substrates in the evolution of the signaling circuitry, the TK circuits were classified into four signaling routes based on the subcellular localization patterns of the components, which have been annotated in the Swiss-Prot database (Fig. 2A; Boeckmann et al. 2003). A TK in a circuit may be a transmembrane receptor tyrosine kinase (RTK) or a cytoplasmic tyrosine kinase (CTK), whereas a substrate may be a membranous or cytoplasmic protein. The SH2/PTB domain–containing proteins were not considered in the signaling route classification because they are primarily cytoplasmic; however, notable exceptions can shuttle between the cytoplasm and the nucleus (e.g., the STAT SH2 domains) (Croker et al. 2008) or become membrane-associated (e.g., PI3K). The CTK-cytoplasmic and CTK-membranous signaling routes process signals from a cytoplasmic TK to a cytoplasmic or membranous substrate, respectively. In contrast, the RTK-cytoplasmic and RTK-membranous signaling routes transmit signals from a receptor TK to a cytoplasmic or membranous substrate, respectively.

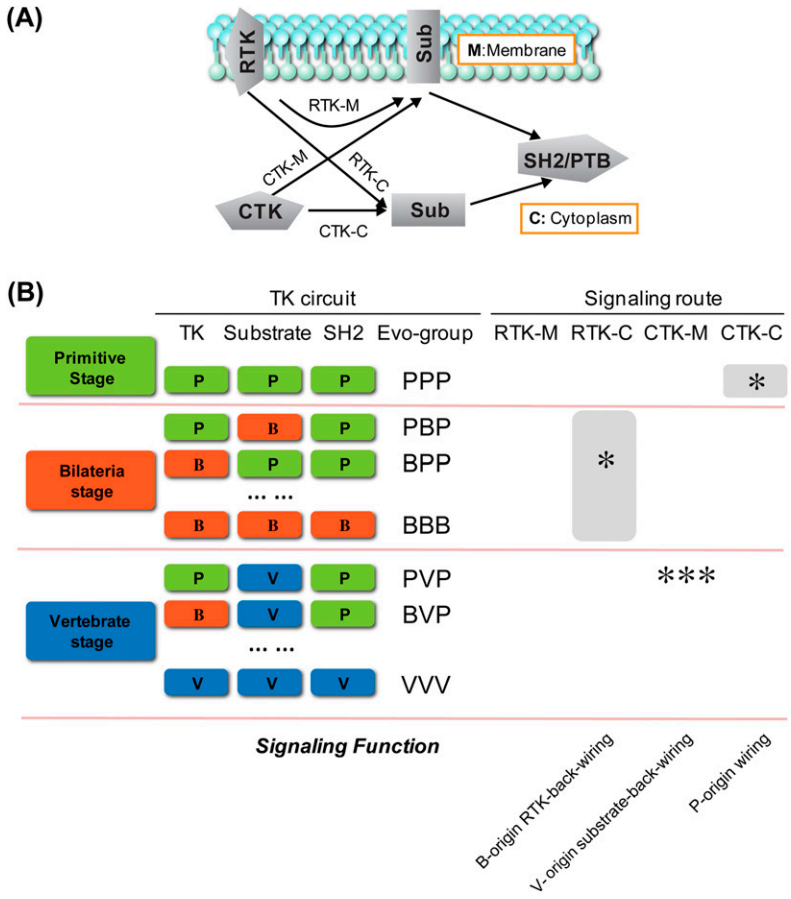

Figure 2. Evolutionary origins of tyrosine kinase (TK) circuits and signaling routes. (*A*) A diagram depicting the signaling routes for intracellular, inter- or extracellular, and tissue-specific communication via the TK circuits. A kinase or substrate may be a membranous or cytoplasmic protein. Depending on cellular locations of its components, a TK circuit may belong to one of the four signaling routes, namely, RTK-M, RTK-C, CTK-M, and CTK-C, where M and C denote membrane and cytoplasm, respectively. (*B*) Enrichment of different signaling routes at distinct evolutionary stages. TK circuits were divided into "evo-groups" according to the origins of the corresponding circuit components. The evo-group is used to assign the origin of a signaling route. The enrichment of a particular evo-group circuit in a signaling route provides information on the evolutionary origin of the latter. TK circuits of a particular origin that are enriched ($P < 0.05$) in a signaling route for both curated and predicted data sets were identified by gray rectangles with the significantly enriched evo-group identified by asterisks. Intracellular signaling (represented by the CTK-C signaling route) is significantly enriched with the PPP evo-group circuits (via the P-origin self-wiring path). Extracellular or intercellular signaling (represented by the RTK-C route) is significantly enriched with the BPP evo-group circuits in which newly evolved (B-origin) TKs "wire back" to ancient (P-origin) substrates (via the B-origin RTK-back-wiring path). In contrast, tissue-specific CTK-M signaling is enriched with PVP circuits that feature vertebrate-origin substrate wiring-back to ancient (P-origin) TKs and SH2 domain-containing proteins (via the V-origin substrate-back-wiring path). (*) $P < 0.05$; (***) $P < 0.001$, randomization test.

This signaling route classification scheme was used to segregate the pTyr circuits into pathways that are associated with intracellular (e.g., CTK-cytoplasmic) or extra-/intercellular signaling (e.g., RTK-membranous, RTK-cytoplasmic, and CTK-membranous). The CTK-cytoplasmic route, which accounts for 28.6% of the curated and 34.3% of the predicted circuits, is dedicated to intracellular signaling through the phosphorylation of cytoplasmic substrates by a cytoplasmic TK. A CTK may also phosphorylate a membrane substrate when it is activated by neighboring cells or environmental cues as in the CTK-membranous route, which accounts for 27.1% of the curated and 7.9% of the predicted circuits. A typical example of this signaling route is the phosphorylation of CD247 by FYN in the T-cell receptor (TCR) signaling pathway (Weiss and Littman 1994). In contrast, membrane-embedded RTKs couple extracellular stimuli to intracellular signaling cascades via the phosphorylation of either cytoplasmic (via the RTK-cytoplasmic route, which comprises 29.3% of the curated and 45.6% of the predicted circuits) or membranous substrates (via the RTK-membranous route, which comprises 14.9% of the curated and 12.2% of the predicted circuits). The RTK-membranous route includes RTK autophosphorylation in which the TK and substrate are located on the same protein (Deribe et al. 2010; Lemmon and Schlessinger 2010). We found that 82.8% of the curated RTK-membranous circuits belong to the RTK autophosphorylation category; this finding probably reflects a bias of the available experimental data on RTK signaling (Lemmon and Schlessinger 2010). However, RTK autophosphorylation comprises only 0.03% of the predicted RTK-membranous circuits. Due to this discrepancy between the curated and predicted TK circuit data sets, we excluded the RTK-membranous signaling routes from further analyses.

Signal transduction in vertebrates occurs in a cell- or tissue-specific manner. To determine which signaling route is tissue-specific, we assigned each component of a TK circuit as either ubiquitously or tissue-specifically expressed, based on the Human Protein Atlas database (Berglund et al. 2008b). Because most TKs and SH2/PTB domain–containing proteins (over 92%) are ubiquitously expressed (see Supplemental Data File 3), we assigned tissue-specificity to a TK circuit based on the tissue-specificity of the corresponding substrate and then examined the tissue-specificity of the different signaling routes. Approximately 62.8% of the CTK-membranous circuits in the curated database are tissue-specific, while only 1.2% of the RTK-cytoplasmic and 20.1% of the CTK-cytoplasmic circuits are tissue-specific. A similar trend was observed for the predicted data set; 27.3% of the CTK-membranous circuits are tissue-specific, while 8.0% of the RTK-cytoplasmic and 7.6% of the CTK-cytoplasmic circuits are tissue-specific (Table 1). We show that tissue-specific circuits are significantly enriched in the CTK-membranous route (both Pp and Pc $< 1.0 \times 10^{-4}$, where Pc and Pp represent the *P*-values that were obtained from randomization tests using the curated and predicted TK data sets, respectively). These results strongly suggest that TK circuits that belong to the CTK-membranous route are preferentially used in tissue-specific pTyr signaling.

Because a membrane protein is the substrate of the CTK-membranous circuit, we next examined whether specific membrane proteins are favored for tissue-specific pTyr signaling. Indeed, we found that >50% of the substrates in either the curated or the predicted CTK-membranous circuits are receptors, and of these, only a small fraction are RTKs according to Gene Ontology (Dennis et al. 2003). This result suggests that CTKs may regulate nonkinase receptor signaling events by pathway crosstalk. For example, the direct phosphorylation of an ionotropic glutamate receptor by the SRC or FYN kinase regulates its localization and activity, thereby modulating excitatory neurotransmission in the brain (Dingledine et al. 1999). Therefore, in addition to being largely responsible for tissue-specific pTyr signaling, the CTK-membranous circuits play an important role in regulating signaling pathways that are mediated by nonkinase membrane receptors.

Table 1. Tissue-specific TK circuits in different signaling routes

	Signaling route		
	RTK-C	CTK-M	CTK-C
Curated data set			
Tissue-specific circuits	2	76	33
Total circuits	167	121	164
Percentage of tissue-specific circuits	1.2%	62.8% ($P < 1.0 \times 10^{-4}$)[a]	20.1%
Predicted data set			
Tissue-specific circuits	1,562	938	1,126
Total circuits	19,548	3,436	14,816
Percentage of tissue-specific circuits	8.0%	27.3% ($P < 1.0 \times 10^{-4}$)[a]	7.6%

RTK-C indicates RTK-cytoplasmic; CTK-M, CTK-membranous; and CTK-C, CTK-cytoplasmic.
[a]The *P*-values obtained from the randomization test. All of the TKs, SH2, and PTB protein families are widely expressed across all tissues, but a fraction (i.e., f%) of the substrate families are expressed in a tissue-specific manner. To test if CTK-M circuits are enriched with tissue-specific circuits, we randomly assigned f% of the total substrates as tissue-specific and then calculated the fraction (f%-random) of the CTK-M circuits that are tissue-specific. We tested the null hypothesis, f%-random ≥ f%-real (the fraction of tissue-specific CTK-M circuits in the original data set). Ten thousand times of randomizations were used for calculating the *P*-value. The same procedure was applied to test the tissue-specificity of other types of circuits.

Distinct evolutionary paths for the intracellular, inter-/extracellular, and tissue-specific human pTyr signaling circuits

What are the evolutionary origins and trajectories of the different signaling routes? To address this question, we examined whether the human TK circuits that belong to a given signaling route are statistically enriched at a specific period of evolution by performing randomization tests (see Methods). These analyses indicate that the CTK-cytoplasmic signaling route is significantly enriched with primitive-origin circuits (Pc $< 1.0 \times 10^{-4}$; Pp $= 2.3 \times 10^{-2}$) (Fig. 2B; Supplemental Fig. S2). In contrast, the RTK-cytoplasmic signaling route is enriched with bilateria-origin circuits (Pc $= 9.4 \times 10^{-3}$; Pp $= 1.5 \times 10^{-2}$) (Fig. 2B; Supplemental Fig. S2). The majority of tissue-specific CTK-membranous circuits are found in vertebrates, although this enrichment is not statistically significant (Fig. 2B; Supplemental Fig. S2). Therefore, human TK circuits for intracellular pTyr signaling originated mainly from primitive organisms, whereas those for inter-/extracellular pTyr signaling originated mainly from the bilaterian lineage.

Although the above analysis identified the evolutionary origins of the CTK-cytoplasmic and RTK-cytoplasmic signaling routes, it did not provide information of the evolutionary method of the corresponding TK circuits. To address this question, the TK circuits were divided into different evolutionary groups, or evo-groups, according to the origins of the circuit components. Specifically, an evo-group is represented by a three-letter acronym in which the three letters denote the origins (i.e., P, B, and V) of the TK, the substrate, and the SH2/PTB domain of the TK circuit, respectively. For example, a BVP evo-group features a TK of bilaterian

origin, a substrate of vertebrate origin, and an SH2/PTB domain of primitive origin (Fig. 2B; Supplemental Fig. S2). It should be emphasized that the evo-groups and the signaling routes are different methods that can be used to represent the same TK circuits. While the former method emphasizes the evolutionary origin, the latter stresses the function of the circuit.

Each component of a TK circuit has three possible evolutionary origins based on our species grouping scheme (Fig. 1B); therefore, by enumeration, there are 27 possible evo-groups (i.e., PPP, PBP and VVV) that can be used to segregate the human TK circuits. We next examined which evo-group circuits were preferentially used in a given signaling route by performing randomization tests for both the curated and predicted circuit data sets (see Methods). The CTK-cytoplasmic, RTK-cytoplasmic, and CTK-membranous signaling routes are significantly enriched with TK circuits that belong to distinct PPP, BPP, and PVP evo-groups, respectively (Fig. 2B; Supplemental Fig. S2). Of note, the PPP circuits are significantly enriched in the CTK-cytoplasmic signaling route ($Pc < 1.0 \times 10^{-4}$; $Pp < 2.3 \times 10^{-2}$, randomization tests). This indicates that functional coupling among primitive-origin TK components (or the "self-wiring" of components that emerged during the same evolutionary period) is the preferred evolutionary path for intracellular pTyr signaling. However, PPP circuits were also found in other signaling routes, which suggests that the prototypes of the human pTyr signaling machineries developed at the earliest stage of metazoan evolution. In contrast to the role of PPP circuits in pTyr signaling, BPP circuits are found to be statistically enriched in the RTK-cytoplasmic signaling route ($Pc < 1.6 \times 10^{-3}$; $Pc < 1.7 \times 10^{-2}$; randomization tests). This finding elucidates an evolutionary path for inter-/extracellular pTyr signaling in which bilaterian RTKs "wire back" to and phosphorylate ancient substrates (i.e., B-origin RTK-back-wiring) (Fig. 2B). These results are consistent with our findings that the human intracellular pTyr signaling originated mainly in primitive species, and inter-/extracellular signaling emerged largely in the bilaterian lineage.

Circuits of the evo-group PVP are significantly enriched in the CTK-membranous signaling route (Pc or $Pp < 1.0 \times 10^{-4}$) (Fig. 2B; Supplemental Fig. S2). Because this route is enriched with tissue-specific circuits (Table 1), a likely evolutionary path for tissue-specific pTyr signaling is through V-origin substrate-back-wiring (Fig. 2B; Supplemental Fig. S2). Specifically, a newly evolved (vertebrate-origin) substrate is phosphorylated by an existing (primitive-origin) tyrosine kinase, which, in turn, is recognized by a primitive SH2/PTB domain. Therefore, although the vertebrate circuits (i.e., PVP, BVP, and VVV) as a group are not significantly enriched in the CTK-membranous signaling route, the TK circuits of the PVP evo-group must have played an important role in the evolution of tissue-specific signaling. The evolutionary landscape of the TCR signaling pathways provides a typical example of how the TK circuits of the PVP evo-group are used in cell-specific pTyr signaling and the pTyr signaling network expansion (Supplemental Fig. S3).

To test whether our results regarding the significant enrichment of specific evo-groups in the different signaling routes could tolerate potential mistakes in protein ortholog identification, we performed a sensitivity assay that is similar to the assay that was described by Cui et al. (2006). Assuming a false identification rate of 20% for orthologs (a highly unlikely scenario) by the Roundup database, we randomly assigned an evolutionary origin (P, B, or V) to 20% of the TKs, substrates and SH2/PTB proteins in the curated and predicted data sets. Next, we repeated the same statistical analysis and found that the corresponding *P*-values were <0.05. These sensitivity assays demonstrate that the conclusions that we drew from the circuit analysis are robust and can tolerate potential errors in the identification of orthologs.

Taken together, our analysis indicates an evolutionary hierarchy for the different pTyr signaling functions; intracellular signaling emerged first in primitive organisms, and this signaling was followed by the emergence of inter- and extracellular signaling in bilateria. Tissue specific–signaling evolved mainly in the vertebrates. As summarized in Figure 2B, the distinct evolutionary paths, "self-wiring," "B-origin RTK-back-wiring," and "V-origin substrate-back-wiring," that were used by the CTK-cytoplasmic (intracellular signaling), RTK-membranous (intercellular signaling), and CTK-membranous (tissue-specific signaling) circuits, respectively, suggest a general mechanism for the evolution of the human pTyr signaling network; new circuits were formed by the back-wiring of a newly emerged kinase or substrate to the existing components. This provides an economic yet efficient means by which to wire and expand the pTyr signaling network by "recycling" existing circuit components and bestowing upon them new roles through the functional coupling to newly evolved kinases or substrates.

The pTyr sites that are associated with multiple TK circuits are hotspots for cancer signaling

Aberrant pTyr signaling drives many of the fundamental biological processes that accompany tumor initiation and progression. Indeed, genes that encode all of the human TKs, and most (56.7%) of the human SH2-containing proteins have been identified as cancer-driving genes in COSMIC, a database of somatically acquired mutations in human cancer (Forbes et al. 2011). Using an integrative analysis of the human signaling network with tumor genomic data, we previously showed that tyrosine kinases are significantly enriched in the cancer signaling network (Cui et al. 2007). Therefore, we wanted to determine whether and how the pTyr signaling network is perturbed in cancer at the TK circuit level.

Recent advances in mass spectrometry have led to the identification of numerous phosphorylation sites, including pTyr sites, in cancer samples (Moran et al. 2006; Olsen et al. 2006; Nita-Lazar et al. 2008; Macek et al. 2009; Thingholm et al. 2009; Harsha and Pandey 2010). The availability of cancer-specific phosphoproteomic data provides a unique opportunity to investigate how TK circuits are wired in cancer cells. To characterize the role of human TK circuits in cancer-cell signaling, we extracted the pTyr proteome data that were determined from 191 lung cancer and 48 normal lung samples (Rikova et al. 2007). We considered pTyr sites that were identified in cancer samples but not in normal samples to be cancer pTyr sites. The predicted circuits were employed to analyze the relationship between the circuits and cancer pTyr sites because the curated data set contains only limited annotated pTyr sites that are insufficient for statistical analysis. Of the 998 cancer pTyr sites contained in 9635 circuits in the data set (see Supplemental File 4), 34% were found to be detected in two or more tumor samples. We used a frequency cutoff of two (similar trends were obtained when the cutoff was set to four or seven) to distinguish between high-frequency cancer (HFC) pTyr sites (i.e., a site is detected in two or more tumor samples) and low-frequency cancer (LFC) pTyr sites (i.e., a site is detected in only one tumor sample). By using this criterion, we identified 341 HFC pTyr sites that were associated with 4245 TK circuits and 657 LFC pTyr sites that were associated with 5390 TK circuits.

Are HFC and LFC pTyr sites different in their capacity to form TK circuits? To address this question, we compared the number of

TK circuits that are mediated by an HFC or a LFC pTyr site/substrate. The average number of circuits (12.4) that are linked to an HFC pTyr site is significantly larger than those (8.20) that are linked to a LFC pTyr site ($P = 1.53 \times \geq 10^{-8}$; Wilcoxon rank sum test). This difference may be attributed to the increased capacity of an HFC pTyr site to be coupled to more TKs or SH2/PTB domains. We found that an HFC pTyr site, on average, links to more TK writers and more SH2/PTB readers than does a LFC pTyr site (the number of writers per HFC/LFC pTyr site was 3.63/2.73, $P = 2.23 \times 10^{-9}$, and the number of readers per HFC/LFC pTyr site was 3.08/2.65, $P = 2.54 \times 10^{-3}$; Wilcoxon rank sum test). To extract the general trend, we divided the cancer pTyr sites into four groups based on their frequencies of detection, and each group contained >10% of the total cancer pTyr sites. We then plotted the average number of TKs, SH2s/PTBs, and TK circuits against the frequency of detection of a pTyr site in the cancer samples (Fig. 3A). It is apparent that the increasing frequency of the detection of a pTyr site in the cancer samples correlated with an increasing number of kinases, SH2/PTB domains, and TK circuits that are connected to the site. Therefore, promiscuous pTyr sites that are shared by many TKs, SH2/PTB domains, and circuits are potential "hotspots" for cancer signaling. Presumably, these highly connected pTyr sites are used to process a variety of cellular information and are used for signal exchange and integration.

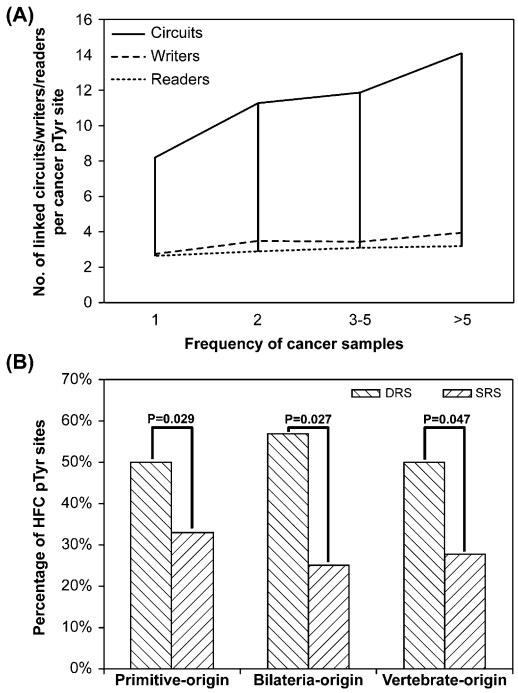

Figure 3. Cancer TK circuits preferably make use of multifunctional substrates and highly connected pTyr sites. (*A*) Cancer TK circuits are highly connected. Shown in the graph is a correlation between the frequency of a pTyr site to be detected from cancer samples (*x*-axis) and the number of SH2/PTB domains, TKs, or circuits connected to that site, based on the predicted data set. The cancer pTyr sites are divided into five groups (based on frequency of detection) to ensure that each group has at least 10% of the total cancer pTyr sites. (*B*) TK circuits identified with high frequency in cancer samples are significantly enriched in multifunctional substrates of the primitive (P), bilaterian (B), and vertebrate (V) origin. The substrates corresponding to the high-frequency cancer (HFC) pTyr sites are divided into either the SRS (singular-role substrate) or the DRS (dual-role substrate) group according to the absence or presence of a kinase and/or an SH2/PTB domain in the same substrate. Bar graph: comparison in the number (shown in percentage) of HFC pTyr sites in the SRS and DRS groups. The substrates are classified according to their evolutionary origins. DRSs of the P-, B-, and V-origin are significantly enriched in cancer signaling. *P*-values were calculated using χ^2 test.

Because all tyrosine kinases are potentially oncogenic and because cancer-associated mutations are found in most SH2/PTB proteins, we reasoned that these proteins may play a key role in the formation of cancer TK circuits. In addition to the Tyr phosphorylation site(s), a substrate may also contain a kinase and/or an SH2/PTB domain. This latter type of substrate can function as a writer and/or reader in other TK circuits; therefore, it has dual or multiple potential functions. To distinguish between these two types of substrates, we defined a substrate as a dual-role substrate (or DRS) if it contains a kinase domain and/or an SH2/PTB domain in addition to the pTyr site. We named a substrate a single-role substrate (or SRS) if it does not contain a kinase or an SH2/PTB domain. Next, we examined the distribution of cancer pTyr sites in the two types of substrates. We found that HFC pTyr sites are significantly enriched in the DRS relative to the SRS type (50.6% vs. 32.6%, $P = 6.78 \times 10^{-3}$; χ^2 test). This indicates that DRSs are preferentially recruited for cancer pTyr signaling. Therefore, cancer cells preferentially target signaling hubs that are nucleated by a DRS; these cells exploit the additional kinase and/or SH2/PTB domain(s) that are contained in the DRS and the potential of the corresponding pTyr sites to mediate the formation of a large number of TK circuits.

The different roles of the DRSs and SRSs in cancer pTyr signaling were further characterized by examining their respective distribution patterns during the different evolutionary periods (Fig. 3B). To accomplish this characterization, we assigned each cancer TK circuit an evolutionary origin (i.e., primitive, bilateria, or vertebrate) and calculated the number of corresponding DRS- or SRS-containing circuits. We found that TK circuits that originated in all three of the origins are statistically enriched with DRSs that harbor HFC pTyr sites compared with the SRSs ($P = 0.029$, 0.027, and 0.047 for the primitive, bilaterian, and vertebrate groups, respectively; χ^2 test). Therefore, cancer pTyr signaling selectively recruits DRSs regardless of evolution.

Taken together, our data show that cancer signaling preferentially employs pTyr sites that are coupled to more TK circuits and those that contain DRSs.

Discussion

In this study, we used a network approach to examine the evolutionary history of the human pTyr signaling. We deconvoluted the pTyr network into basic functional units, or TK circuits. Our analysis revealed that intracellular, intercellular, and tissue-specific pTyr signaling routes possessed distinct evolutionary origins, evolved in a stepwise manner, and, furthermore, took distinct evolutionary paths that produced the different pTyr signaling functions.

Intracellular signaling first appeared in primitive metazoans by "self-wiring" (Fig. 2B), which implies that the circuit components, including CTKs, substrates, and SH2/PTB domains, were intact in the selected primitive species; these components were probably also present in the common ancestors of these primitive species. It is intuitively plausible that intracellular pTyr signaling was the first to emerge in metazoan evolution because the ancestral species either are unicellular or contain a few cell types; therefore, the need for extracellular signaling and intercellular communication is not as urgent as in more complex metazoans. This hypothesis is consistent with the finding that extra-/intercellular pTyr signaling is statistically enriched with TK circuits that evolved in the bilaterian

lineage. An inter- and extracellular signaling circuit features a bilateria-origin kinase that phosphorylates a primitive substrate, which, in turn, recruits a primitive protein that contains an SH2/PTB domain. This type of evolutionary pathway, which is termed B-origin RTK-back-wiring (Fig. 2B), is apparently used to expand the human pTyr signaling network for inter- and extracellular communication in bilaterian animals (e.g., Ecdysozoa) and their immediate ancestors.

In contrast, tissue-specific pTyr signaling is characterized by the presence of additional circuits that first emerged in the vertebrate lineage. Intriguingly, we found that tissue-specific signaling circuits are formed via the phosphorylation of a vertebrate-origin substrate by a primitive tyrosine kinase, which is followed by the recruitment of a primitive SH2/PTB domain (V-origin substrate-back-wiring path) (Fig. 2B). Therefore, instead of reinventing the toolkit of pTyr signaling, new circuits formed in the late stages of metazoan evolution, which exploited pre-existing circuit components by coupling them to newly evolved substrates. This mechanism of the pTyr signaling network expansion is highly economical and is an effective means of engendering novel signaling functions to an ancient regulatory protein.

It is noteworthy that these conclusions are drawn from the statistical analysis of circuit enrichment for a particular signaling function, which does not imply the absence of any circuit for tissue-specific or intercellular communication in a primitive organism. On the contrary, all three types of circuits can be found in at least one target species of the primitive group, which suggests that the prototypes of pTyr signaling were developed at the earliest stage of metazoan evolution when the toolkit was in place. It should be noted that our observations are based on the current genome sequences of a number of organisms that originated either before or after the emergence of metazoans from single-celled eukaryotic ancestors. When more genome sequences and more experimentally determined TK circuits become available in the future, it would be interesting to confirm and expand on these observations. Although our analysis is based on 19 representative species, we expect the main conclusions of the current work to stand the test of time. In support of this assertion, we added nine Primitive-stage species in the analysis and found that the enrichment patterns of TK circuits (Fig. 2B) were not changed . Therefore, the models we present here should provide a useful framework for the determination of the origins of pTyr signaling and the origins of analogous multicomponent signaling platforms (circuits).

We also provide a system-level understanding of pTyr signaling in cancer. We showed that the pTyr sites that are linked to multiple kinases, SH2/PTB domains, and TK circuits form hotspots for cancer cell signaling. Previous studies suggest that hub kinases are preferentially involved in cancer signaling (Miller et al. 2008, Tan et al. 2009a). Taken together, these subnetworks that are defined by kinase hubs and pTyr site hubs are preferentially used by cancer signaling. Therefore, targeting these subnetworks may provide an effective strategy for cancer intervention (Olsen et al. 2006).

In summary, our study provides insights into the evolution of the human pTyr signaling network, and the implications of this study will impact the understanding of normal cellular functions and the mechanism of tumorigenesis.

Methods

Definition of TK circuits

We defined a TK circuit as a three-component system composed of (1) a TK that functions as a writer to add a phosphate moiety to a Tyr residue, (2) a substrate that contains the Tyr residue that is phosphorylated by the TK, and (3) a protein containing an SH2 or PTB domain that functions as the reader of the pTyr residue. Since the pTyr on a substrate is necessary for the TK circuit to be functional, only experimentally verified pTyr sites were used to build the collection of human TK circuits analyzed herein.

Construction of a curated human TK circuit data set

We manually collected the experimentally determined human TK–substrate interactions and substrate–SH2/PTB domain interactions from the literature (see Supplemental Materials), as well as the Phospho.ELM and PhosphoSitePlus databases (Diella et al. 2004; Hornbeck et al. 2004). We grouped the closely related proteins (e.g., CRK and CRKL) into families because they have similar binding properties and biological functions (Huang et al. 2008). The TK and SH2 protein families were identified based on the Kinome database and the relevant literature (Enright et al. 2002; Boeckmann et al. 2003; Huang et al. 2008). The substrate families were identified based on the Ensembl family database (Enright et al. 2002). Once a pTyr site was identified in one member of a protein family, the corresponding site in other members of the same family were identified by sequence alignment using the program MAFFT (version 6.846, E-INS-i option with default parameters) (Katoh et al. 2005). For example, Tyr221 in CRK aligns with Y207 in CRKL, and both have been shown to be phosphorylated (Hornbeck et al. 2004). We treated these two pTyr sites as a single pTyr site because it is highly probable that they are phosphorylated by the same kinase(s) and read by the same SH2/PTB domain(s) and thereby have similar functions in the signaling network. If a TK–substrate and a substrate–SH2/PTB interaction involve the same pTyr site, they are integrated into a TK circuit. All three components in ~80% of the curated TK circuits are experimentally verified, whereas either the TK or the SH2/PTB domain is missing from the remaining circuits. In the latter case, we completed the circuits by using appropriate bioinformatic pipelines to predict the corresponding tyrosine kinases or SH2/PTB domains, respectively. TKs were predicted using NetworKIN and NetPhorest (Linding et al. 2007; Miller et al. 2008). The kinase identified by both programs was considered as the writer for the pTyr site. Similarly, we predicted the SH2 reader using NetPhorest and SMALI (Li et al. 2008; Miller et al. 2008) and predicted the PTB readers using NetPhorest (Miller et al. 2008). The predicted writers and readers were then integrated with the substrate containing the pTyr to construct the TK circuits. Approximately 20% of the TK circuits were constructed in this manner.

The curated data set contains 583 circuits covering 20 of 29 TK families, 26 of 38 SH2 families, six PTB families, and 92 substrate families bearing 243 pTyr sites in humans. It should be noted that the number of pTyr sites did not exactly match the number of TK circuits. For instance, Tyr72 in the protein BLNK could be phosphorylated by either SYK or INSR (insulin receptor) and bind to either the VAV or GRB2 SH2 domains upon phosphorylation. Therefore, this pTyr site in BLNK is involved in four TK circuits.

Computationally predicted human TK circuits

To increase coverage of the pTyr signaling network, we constructed a data set comprising over 50,000 computationally predicted human TK circuits. Each of the predicted circuits contains an experimentally determined pTyr site as annotated in the PhosphoSitePlus database (Hornbeck et al. 2004), a TK predicted to phosphorylate this site, and an SH2 or a PTB domain predicted to bind to the pTyr. For each pTyr site, its TK writers were predicted using NetPhorest and GPS2.1 with a high threshold of false-positive rate 2% (Xue et al.

2008). If a TK was covered by both programs, the TK writers predicted by both methods were retained. If the TK was covered by a single program, this program was employed for the prediction. The SH2 readers were predicted by both NetPhorest and SMALI (Li et al. 2008; Miller et al. 2008), and the PTB readers were predicted by NetPhorest. The pTyr sites that have predicted writers and readers were retained in the data set. Protein families were constructed as described above. The aligned pTyr sites from different members within a substrate family were considered as a single pTyr site if they share at least one predicted TK writers and at least one SH2/PTB readers. Otherwise, the aligned pTyr sites were considered as different pTyr sites since they have dissimilar specificity, suggesting that they have different functions. A circuit was constructed by integrating a single pTyr site, a single writer, and a single reader. Finally, using the computational approach we constructed 50,475 circuits, in which half of the curated circuits are included.

Assignment of evolutionary origins of the human TK circuits

We assigned the evolutionary origins of the human TK circuits based on the orthologous circuits in 19 eukaryotes mentioned above (Fig. 1B). For each human TK circuit, we used the Roundup database (Wall et al. 2003) to identify orthologs of the circuit components in these selected species. We next classified these organisms into three different evolutionary periods (P, B, and V; see main text). We then annotated the evolutionary origins of the three components (i.e., TKs, substrates, and SH2s/PTBs) of each human TK circuit. The evolutionary period in which a component first appeared was considered the origin of the corresponding protein or protein family. For instance, if a TK (e.g., ABL) was found from *M. brevicollis* to *H. sapiens*, it is considered to be originated in the primitive period. Finally, the origin of a circuit was assigned to the evolutionary period in which the last circuit component appeared. For example, in the circuit SRC(TK)/CD19(substrate, pTyr500)/ PIK3R1(SH2), the orthologs of the human SRC and PIK3R1 were found in *M. brevicollis* (a primitive organism), but the ortholog of the human CD19 was first seen in *B. taurus* (a vertebrate). This circuit was assigned a vertebrate (or V) origin. The evolutionary annotations of the circuits are documented in Supplemental Data Files 1 and 2. In an independent analysis, we used the InParanoid database (Berglund et al. 2008a) for ortholog identification, repeated the same circuit analyses as described in the main text, and obtained similar results (data not shown). It should be noted that both human tyrosine kinases and SH2 domain–containing proteins include multiple domains, some of which are well co-conserved with TK and SH2 domains, respectively (Pincus et al. 2008). We employed this feature to determine the orthologs of tyrosine kinases and SH2 proteins. A protein was defined as the ortholog of a query tyrosine kinase/SH2 domain–containing protein that not only was identified by the ortholog database but also contained at least one same domain plus TK/SH2 domain as the query protein.

Determination of the tissue specificity of the TK circuits

We annotated the expression of each circuit component as tissue-specific or ubiquitous (widely expressed) based on the Human Protein Atlas (Berglund et al. 2008b), which documents the expression patterns of human proteins in ~65 (64–66) different tissues/cell types. If a protein is expressed in more than half of the tissues/cells (33), we consider it as a ubiquitous protein; otherwise, we consider it as a tissue-specific protein. We considered a protein family as ubiquitous if it contains at least one ubiquitous protein. Otherwise, the family was annotated as tissue-specific. Our observation that CTK-membranous circuits are enriched in tissue-specific proteins is still correct when other thresholds are employed, such as tissue-specific proteins are those expressed in 20 tissues (data not shown).

Statistical analysis

Fisher's exact tests, χ^2 tests, Wilcoxon rank-sum tests, and randomization tests were used to evaluate the statistical significance of our observations. The detailed procedures for network circuit randomization tests have been described previously (Wang and Purisima 2005). For the randomization test of the tissue-specific circuits, we use CTK-membranous as an example. To test if CTK-membranous circuits are enriched with tissue-specific circuits, we randomly assigned f% of the total substrates as tissue-specific and then calculated the fraction (f%-random) of the CTK-membranous circuits that are tissue-specific. We tested the null hypothesis, f%-random $\geq$ f%-real (the fraction of tissue-specific CTK-membranous circuits in the original data set). Ten thousand times of randomizations were used for calculating the *P*-value. The same procedure was applied to test the tissue-specificity of other types of circuits. To perform randomization tests for the TK circuits, we built a network using the curated TK circuits (for testing in the curated data set) or the predicted TK circuits (for testing in the predicted data set) and then randomly swapped the evolutionary origins (P, B, and V) to the circuit components.

To test the enrichment of a signaling route in an evolutionary stage or period, for each round of the randomly shuffling of the evolutionary origins (P, B, and V) of the circuit components, we tested the null hypothesis, EN $\geq$ RN, where EN and RN are the expected and the real numbers of the circuits for the signaling route in that evolutionary period, respectively. The *P*-values were calculated based on 10,000 times of randomizations.

To test the enrichment of an evo-group in a signaling route, we applied the same randomization testing procedures, but tested the null hypothesis, EN $\geq$ RN, where EN and RN are the expected and the real numbers of the evo-group circuits in that signaling route, respectively.

Acknowledgments

This project was supported by grants from the Canadian Cancer Society (S.S.C.L.), Genome Canada through the Ontario Genomics Institute (S.S.C.L.), NRC's Genome and Health Initiative (E.W.), Canada Institutes of Health Research (E.W.), and the National Institutes of Health (M.R.S., NIH grant number GM079689). S.S.C.L. holds a Canada Research Chair in Functional Genomics and Cellular Proteomics, and M.F.M. holds a Canada Research Chair in Molecular Therapeutics.

Authors' contributions: E.W. and S.S.C.L. conceived the project. L.L. and E.W. designed the research. L.L., E.W., C.F., C.T., and T.K. performed research and analyzed data. L.L., C.T., and M.R.S. performed the computational experiments. L.L., E.W., T.K., S.S.C.L., M.R.S., and M.F.M. wrote the paper.

References

Berglund AC, Sjolund E, Ostlund G, Sonnhammer EL. 2008a. InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res* **36:** D263–D266.

Berglund L, Bjorling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA, Persson A, Ottosson J, Wernerus H, Nilsson P, et al. 2008b. A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteomics* **7:** 2019–2027.

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31:** 365–370.

Boran AD, Iyengar R. 2010. Systems approaches to polypharmacology and drug discovery. *Curr Opin Drug Discov Devel* **13:** 297–309.
Cherry JL. 2010. Highly expressed and slowly evolving proteins share compositional properties with thermophilic proteins. *Mol Biol Evol* **27:** 735–741.
Croker BA, Kiu H, Nicholson SE. 2008. SOCS regulation of the JAK/STAT signalling pathway. *Semin Cell Dev Biol* **19:** 414–422.
Cui Q, Yu Z, Purisima EO, Wang E. 2006. Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol* **2:** 46. doi: 10.1038/msb4100089.
Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, Zhang S, Liu L, Lu M, O'Connor-McCourt M, et al. 2007. A map of human cancer signaling. *Mol Syst Biol* **3:** 152. doi: 10.1038/msb4100200.
Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** 3. doi: 10.1186/gb-2003-4-5-p3.
Deribe YL, Pawson T, Dikic I. 2010. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17:** 666–672.
Diella F, Cameron S, Gemund C, Linding R, Via A, Kuster B, Sicheritz-Ponten T, Blom N, Gibson TJ. 2004. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* **5:** 79. doi: 10.1186/1471-2105-5-79.
Dingledine R, Borges K, Bowie D, Traynelis SF. 1999. The glutamate receptor ion channels. *Pharmacol Rev* **51:** 7–61.
Elango N, Hunt BG, Goodisman MA, Yi SV. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci* **106:** 11206–11211.
Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30:** 1575–1584.
Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, et al. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39:** D945–D950.
Gough NR, Foley JF. 2010. Focus issue: systems analysis of protein phosphorylation. *Sci Signal* **3:** eg6. doi: 10.1126/scisignal.3137eg6.
Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455:** 1193–1197.
Harsha HC, Pandey A. 2010. Phosphoproteomics in cancer. *Mol Oncol* **4:** 482–495.
Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22:** 2971–2972.
Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B. 2004. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4:** 1551–1561.
Huang H, Li L, Wu C, Schibli D, Colwill K, Ma S, Li C, Roy P, Ho K, Songyang Z, et al. 2008. Defining the specificity space of the human SRC homology 2 domain. *Mol Cell Proteomics* **7:** 768–784.
Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33:** 511–518.
King N, Hittinger CT, Carroll SB. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* **301:** 361–363.
King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature* **451:** 783–788.
Lemmon MA, Schlessinger J. 2010. Cell signaling by receptor tyrosine kinases. *Cell* **141:** 1117–1134.
Li L, Wu C, Huang H, Zhang K, Gan J, Li SS. 2008. Prediction of phosphotyrosine signaling networks using a scoring matrix-assisted ligand identification approach. *Nucleic Acids Res* **36:** 3263–3273.
Lim WA, Pawson T. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142:** 661–667.
Lindberg RA, Quinn AM, Hunter T. 1992. Dual-specificity protein kinases: will any hydroxyl do? *Trends Biochem Sci* **17:** 114–119.
Linding R, Jensen LJ, Ostheimer GJ, van Vugt MA, Jorgensen C, Miron IM, Diella F, Colwill K, Taylor L, Elder K, et al. 2007. Systematic discovery of in vivo phosphorylation networks. *Cell* **129:** 1415–1426.
Macek B, Mann M, Olsen JV. 2009. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* **49:** 199–221.
Manning G, Young SL, Miller WT, Zhai Y. 2008. The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc Natl Acad Sci* **105:** 9674–9679.
Miller ML, Jensen LJ, Diella F, Jorgensen C, Tinti M, Li L, Hsiung M, Parker SA, Bordeaux J, Sicheritz-Ponten T, et al. 2008. Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* **1:** ra2. doi: 10.1126/scisignal.1159433.
Moorhead GB, De Wever V, Templeton G, Kerk D. 2009. Evolution of protein phosphatases in plants and animals. *Biochem J* **417:** 401–409.
Moran MF, Tong J, Taylor P, Ewing RM. 2006. Emerging applications for phospho-proteomics in cancer molecular therapeutics. *Biochim Biophys Acta* **1766:** 230–241.
Nichols SA, Dirks W, Pearse JS, King N. 2006. Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci* **103:** 12451–12456.
Nita-Lazar A, Saito-Benz H, White FM. 2008. Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics* **8:** 4433–4443.
Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127:** 635–648.
Pincus D, Letunic I, Bork P, Lim WA. 2008. Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc Natl Acad Sci* **105:** 9680–9684.
Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, et al. 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317:** 86–94.
Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, et al. 2007. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131:** 1190–1203.
Tan CS, Bodenmiller B, Pasculescu A, Jovanovic M, Hengartner MO, Jorgensen C, Bader GD, Aebersold R, Pawson T, Linding R. 2009a. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci Signal* **2:** ra39. doi: 10.1126/scisignal.2000316.
Tan CS, Pasculescu A, Lim WA, Pawson T, Bader GD, Linding R. 2009b. Positive selection of tyrosine loss in metazoan evolution. *Science* **325:** 1686–1688.
Thingholm TE, Jensen ON, Larsen MR. 2009. Analytical strategies for phosphoproteomics. *Proteomics* **9:** 1451–1468.
Vogel C, Chothia C. 2006. Protein family expansions and biological complexity. *PLoS Comput Biol* **2:** e48. doi: 10.1371/journal.pcbi.0020048.
Wall DP, Fraser HB, Hirsh AE. 2003. Detecting putative orthologs. *Bioinformatics* **19:** 1710–1711.
Wang E, Purisima E. 2005. Network motifs are enriched with transcription factors whose transcripts have short half-lives. *Trends Genet* **21:** 492–495.
Weiss A, Littman DR. 1994. Signal transduction by lymphocyte antigen receptors. *Cell* **76:** 263–274.
Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. 2008. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics* **7:** 1598–1608.

Received July 10, 2011; accepted in revised form December 2, 2011.