npj | Systems Biology and Applications

www.nature.com/npjsba

ARTICLE OPEN

Pathways on demand: automated reconstruction of human signaling networks

Anna Ritz[1,4], Christopher L Poirel[1,5], Allison N Tegge[1], Nicholas Sharp[1,6], Kelsey Simmons[2], Allison Powell[2], Shiv D Kale[2] and TM Murali[1,3]

Signaling pathways are a cornerstone of systems biology. Several databases store high-quality representations of these pathways that are amenable for automated analyses. Despite painstaking and manual curation, these databases remain incomplete. We present PathLinker, a new computational method to reconstruct the interactions in a signaling pathway of interest. PathLinker efficiently computes multiple short paths from the receptors to transcriptional regulators (TRs) in a pathway within a background protein interaction network. We use PathLinker to accurately reconstruct a comprehensive set of signaling pathways from the NetPath and KEGG databases. We show that PathLinker has higher precision and recall than several state-of-the-art algorithms, while also ensuring that the resulting network connects receptor proteins to TRs. PathLinker's reconstruction of the Wnt pathway identified CFTR, an ABC class chloride ion channel transporter, as a novel intermediary that facilitates the signaling of Ryk to Dab2, which are known components of Wnt/β-catenin signaling. In HEK293 cells, we show that the Ryk–CFTR–Dab2 path is a novel amplifier of β-catenin signaling specifically in response to Wnt 1, 2, 3, and 3a of the 11 Wnts tested. PathLinker captures the structure of signaling pathways as represented in pathway databases better than existing methods. PathLinker's success in reconstructing pathways from NetPath and KEGG databases point to its applicability for complementing manual curation of these databases. PathLinker may serve as a promising approach for prioritizing proteins and interactions for experimental study, as illustrated by its discovery of a novel pathway in Wnt/β-catenin signaling. Our supplementary website at http://bioinformatics.cs.vt.edu/~murali/supplements/2016-sys-bio-applications-pathlinker/ provides links to the PathLinker software, input datasets, PathLinker reconstructions of NetPath pathways, and links to interactive visualizations of these reconstructions on GraphSpace.

npj Systems Biology and Applications (2016) **2,** 16002; doi:10.1038/npjsba.2016.2; published online 3 March 2016

INTRODUCTION

A major focus in systems biology is the identification of the networks of reactions that guide the propagation of cellular signals from receptors to downstream transcriptional regulators (TRs). Over the past two decades, databases have been developed to store the interactions present in signaling pathways,[1–5] facilitating their retrieval for computational analyses. While these databases have been iteratively improved over the years, they are still largely built through extensive and time-consuming manual curation. Further, the proteins and interactions within the same signaling pathway may vary considerably from one database to another.

Inspired by these challenges, we sought to develop a computational approach to automatically reconstruct signaling pathways from a background network of molecular interactions (the interactome). We conceptualized the problem as follows (Figure 1): given as input only the receptors and the transcription factors/regulators (TRs) in a specific signaling pathway, can we analyze the interactome to recover the pathway with high accuracy? Several earlier methods have addressed a computationally similar problem of connecting a set of sources or "causes" (akin to receptors) to a set of targets or "effects" (akin to TRs) through a compact sub-network of the interactome.[6–18] However, most of these methods are routinely evaluated on data in budding yeast. To tackle the increased complexity of human signaling pathways, we sought to develop an algorithm with two desirable characteristics. First, the method must be able to compute a reconstruction that captures a large subset of the interactions in the curated signaling pathway. Ideally, it should have a tunable parameter that smoothly determines the size of the solution. Second, to reflect the process of signal transduction, the receptors must be connected to the downstream TRs in the reconstructed pathway.

We develop PathLinker, an algorithm that satisfies both criteria. PathLinker finds the k highest scoring paths from any receptor to any TR, where k is a user-defined parameter (Figure 1). As the value of k increases, the solution smoothly increases to capture more interactions in the curated pathways. By design, every interaction in the reconstruction lies on some path from a receptor to a TR. Thus, PathLinker satisfies both criteria for a reconstruction algorithm.

We apply PathLinker to a comprehensive set of 15 signaling pathways in the NetPath database[3] and 32 pathways in the KEGG database,[5] both of which are manually curated. Compared with several other approaches,[15–20] we show that PathLinker is the only

[1]Department of Computer Science, Virginia Tech, Blacksburg, VA, USA; [2]Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA and [3]ICTAS Center for Systems Biology of Engineered Tissues, Virginia Tech, Blacksburg, VA, USA.
Correspondence: TM Murali (murali@cs.vt.edu)
[4]Current address: Department of Biology, Reed College, Portland, OR, USA.
[5]Current address: RedOwl Analytics, San Francisco, CA, USA.
[6]Current address: Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.
Received 9 July 2015; revised 26 November 2015; accepted 27 November 2015

npj nature partner journals

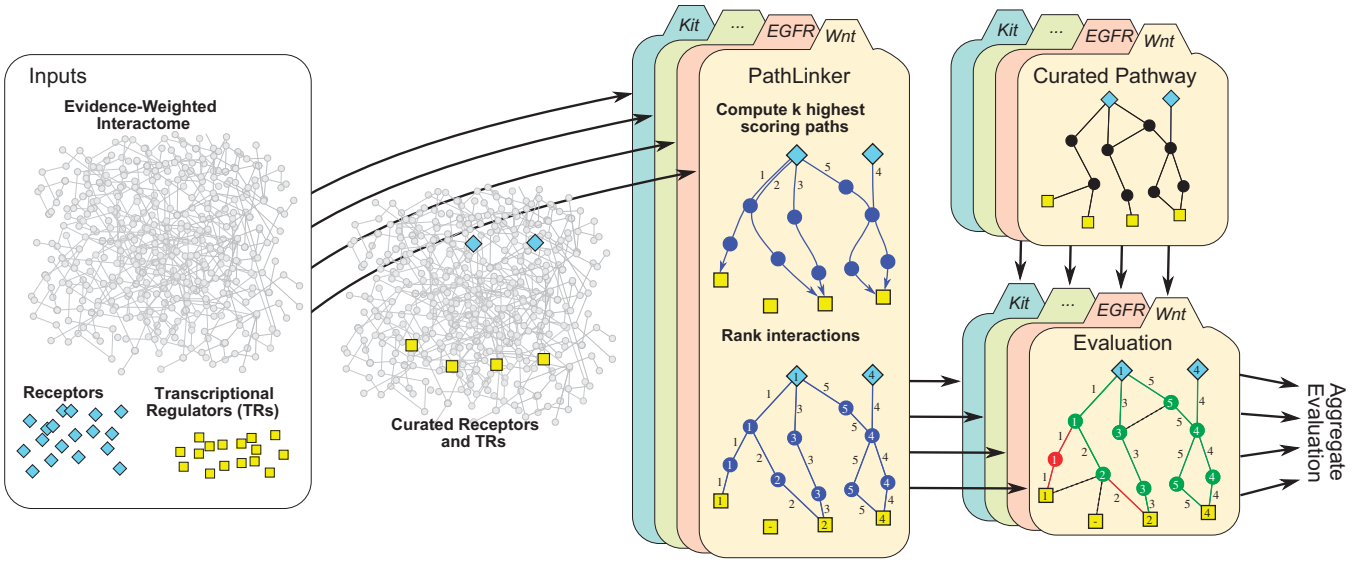


Figure 1. Overview of the PathLinker algorithm. Given an interactome, we identify a set of receptors and a set of TRs for a particular curated pathway (e.g., Wnt). We apply PathLinker to reconstruct the pathway, ranking proteins and interactions by their first occurrence in the *k* shortest paths from any receptor to any TR. Using the curated pathway as a ground truth, we evaluate the performance of PathLinker. We combine the ranked lists for multiple curated pathways to obtain an aggregate evaluation.

Table 1. Method abbreviations

Abbreviation	*Algorithm name/type*	*Reference*
PathLinker	*k* shortest paths from any receptor to any TR	This paper
ShortestPaths	Shortest paths from every receptor to every TR	
RWR	Random walk with restarts	Haveliwala *et al.*[20]
ResponseNet	Network flow	Yeger-Lotem *et al.*[17]
PCSF	Prize collecting Steiner forest	Tuncbag *et al.*[15]
ANAT	Shortest paths/Steiner trees	Yosef *et al.*[18]
IPA	Ingenuity Pathway Analyzer	Ingenuity Pathway Analysis (IPA)[16]
BowTieBuilder	Approximation to the Steiner tree connecting receptors and TRs	Supper *et al.*[19]

Abbreviation: TR, transcriptional regulator.

method that can reconstruct this pathway with high recall while also ensuring connectivity between receptors and TRs. To further highlight PathLinker's effectiveness, we examine results for the Wnt pathway in detail. One of the highest scoring paths computed by PathLinker in the Wnt pathway reconstruction suggests that cystic fibrosis transmembrane conductance regulator (CFTR) and its interactions with receptor-like tyrosine kinase (Ryk) and Dab, mitogen-responsive phosphoprotein, homolog 2 (Dab2), both of which are known members of the Wnt pathway, comprise a novel signaling mechanism from Wnts to β-catenin. We experimentally validate this role for CFTR using loss of function short interfering RNA (siRNA)-based silencing.

RESULTS

We first evaluated the ability of PathLinker and other algorithms to reconstruct a diverse collection of 15 signaling pathways in the NetPath database (Supplementary Section S1). We then experimentally validated a novel prediction from PathLinker on the Wnt signaling pathway.

Pathway reconstructions from the NetPath database

Comparison to other algorithms. We compared PathLinker with six other network-based algorithms (Table 1), including shortest path (ShortestPaths, BowTieBuilder[19]), random walk with restarts (RWR[20]), network flow (ResponseNet[17]), Steiner forest (PCSF[15]), ANAT,[18] and a greedy seed-based method (Ingenuity Pathway Analyzer (IPA[16]). Brief descriptions of these methods and the user-defined parameters we selected appear in Supplementary Section S2.

For each pathway reconstruction, we used the interactions in the NetPath pathway as the set of positives and a subsampled set of interactions not present in the NetPath pathway as the set of negatives (Supplementary Section S3). For each algorithm, we aggregated the reconstructions of these pathways to measure the precision and recall (Figure 2a and Supplementary Section S3). We observed that ANAT, PCSF, ResponseNet, ShortestPaths, and BowTieBuilder achieved values of recall < 0.1. While IPA returned sub-networks with larger recall values, the precision was never above 0.2. RWR achieved the best precision for recall values between 0.05 and 0.13, and PathLinker and RWR were comparable for all other values of recall.

To determine the source of the false positive interactions in PathLinker compared with RWR, we asked if the false positives were "close" to the pathway as represented in the NetPath database. First, we recomputed precision of all algorithms after ignoring interactions that involved at least one true positive node in the NetPath pathway ("pathway-adjacent negatives") before subsampling the negatives (Figure 2b). This modification increased the precision for all the algorithms, with PathLinker clearly dominating all the other methods at values of recall between 0.2 and 0.6. To further investigate this trend, we computed each interaction's distance from any protein in the pathway, where a distance of zero indicated a true positive and a distance of one indicated a pathway-adjacent negative (Figure 2c and Supplementary Section S3). At a recall of 0.2, RWR contained a larger proportion of true positives (purple regions) than PathLinker, while the proportion of true positives was similar at recall 0.4 and 0.6. However, the larger proportion of interactions that were at a distance of 1 from the pathway (dark blue regions) across all three values of recall indicates that PathLinker's false positives were closer to the pathway than RWR's false positives.

To compare PathLinker and RWR using the criterion where we required receptors and TRs to be connected in the reconstruction, we assessed how quickly PathLinker and RWR recovered the curated receptors and TRs. For PathLinker and RWR,

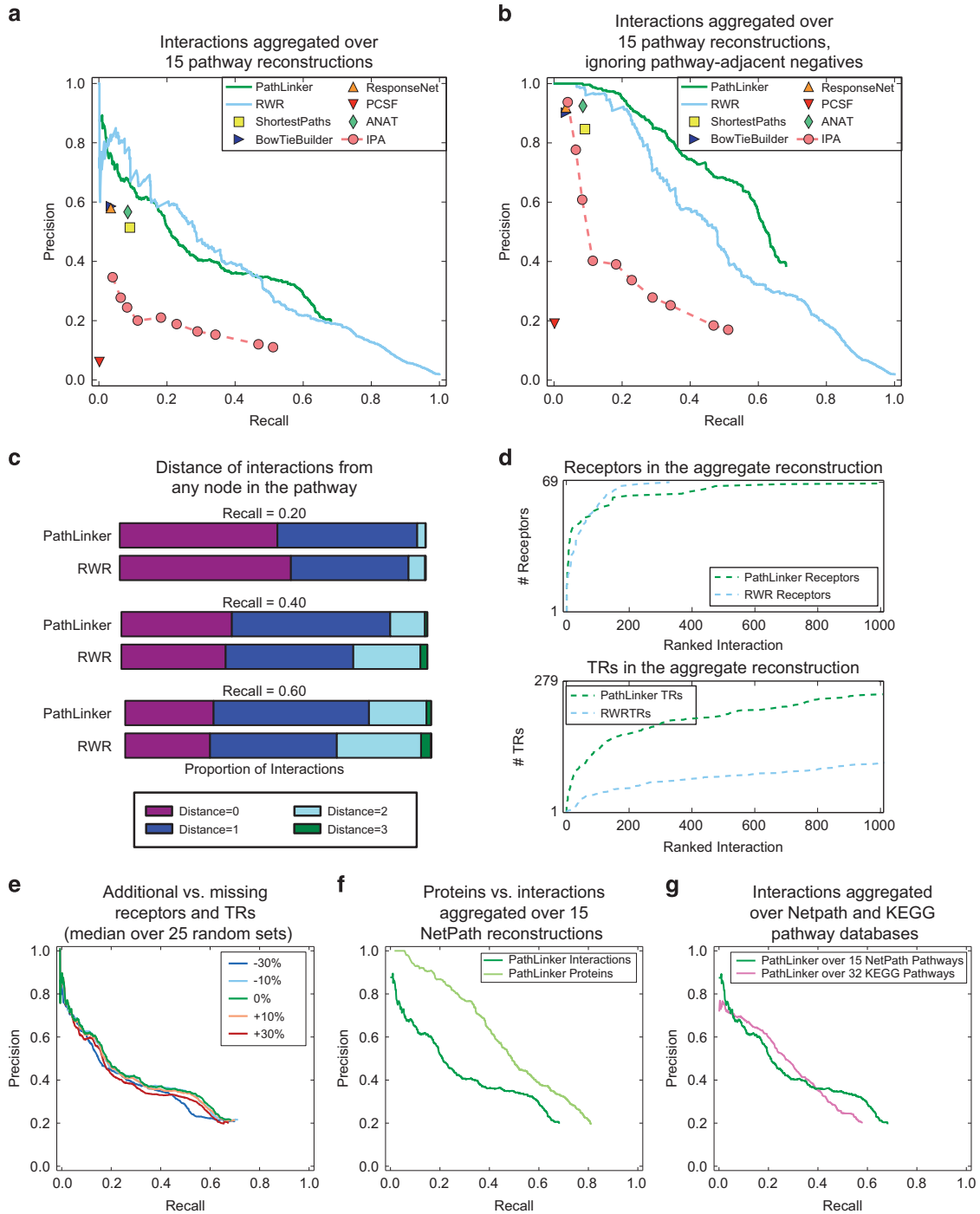


Figure 2. Evaluation of pathway reconstructions aggregated over 15 NetPath pathways. (**a**) Precision and recall of the interactions in pathway reconstructions computed by PathLinker and other algorithms. (**b**) Precision and recall of PathLinker and RWR without considering interactions adjacent to the pathway (distance = 1). (**c**) Distances of each interaction from the pathway for PathLinker and RWR at recalls of 0.2, 0.4, and 0.6. (**d**) Rank of receptors (top) and TRs (bottom) in the first 1,000 interactions from PathLinker and RWR reconstructions (rank for all interactions in Supplementary Figure S2). (**e**) Median values of precision and recall of PathLinker when oversampling and undersampling receptors and TRs. (**f**) Precision and recall of PathLinker when recovering proteins compared with interactions. (**g**) Precision and recall of PathLinker when reconstructing 15 NetPath pathways compared with 32 KEGG pathways. RWR, random walk with restarts.

we recorded the index of the first interaction that contained each receptor or each TR. Figure 2d shows the results for the first 1,000 ranked interactions, and Supplementary Figure S2 shows the full ranking. PathLinker and RWR recovered receptors at about the same rate, although PathLinker's long tail indicated that the last few receptors were difficult for PathLinker to retrieve. Conversely, PathLinker successfully recovered 90% of the TRs in the pathways in the first 1,000 ranked interactions, compared with only 38% recovered by RWR.

Evaluation of PathLinker's performance. We assessed PathLinker's performance in several additional ways to investigate its robustness to the inputs and its effectiveness for other pathway

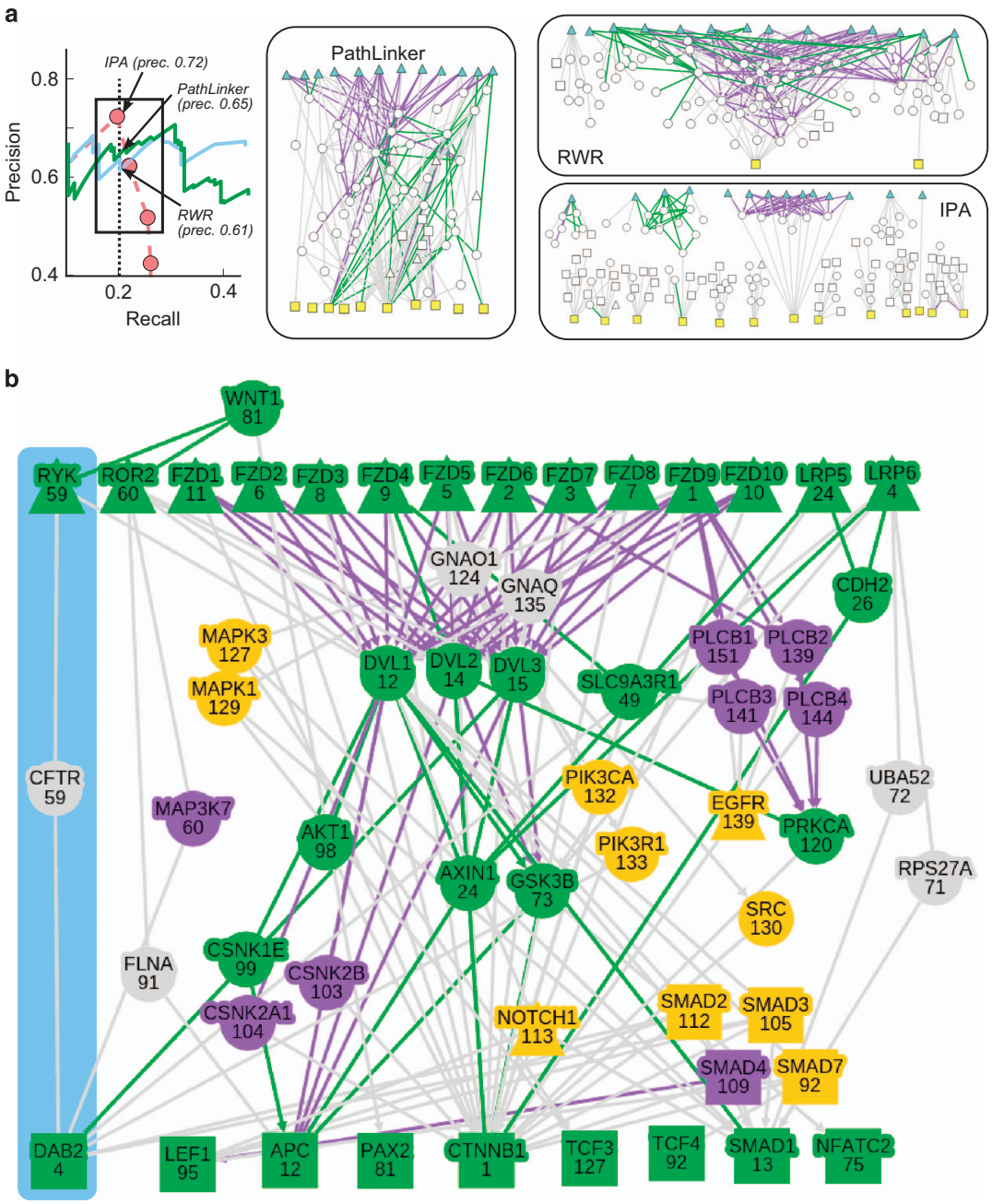
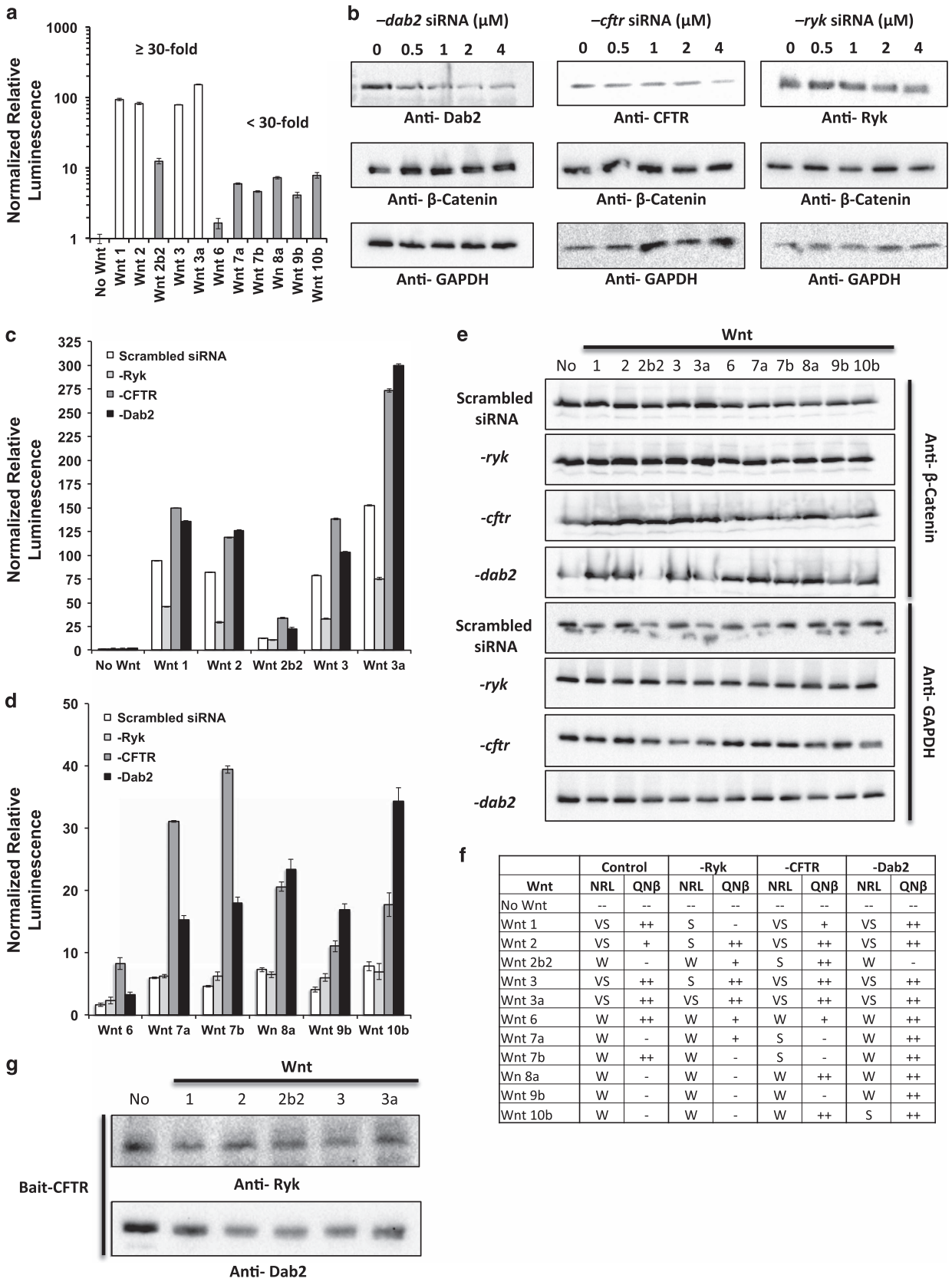


Figure 3. Visualizations of Wnt pathway reconstructions. (**a**) Visualizations of PathLinker, RWR, and IPA pathway reconstructions at a recall of 0.2. The displayed networks correspond to this value of recall (black arrows in the precision/recall curve.) Blue triangles: Wnt receptors; yellow squares: Wnt TRs, green edges: NetPath interactions, purple edges: KEGG interactions that are not present in NetPath. (**b**) Network formed by the 200 highest scoring paths in PathLinker's reconstruction of the Wnt pathway. The number in each node denotes the index of the first path in which that protein appears. Triangles: receptors; squares: TRs, green nodes/edges: NetPath proteins/interactions, purple nodes/edges: KEGG proteins/interactions that are not present in NetPath, orange nodes: proteins known to be involved in Wnt signaling crosstalk. The blue region highlights the novel Ryk–CFTR–Dab2 path, which we experimentally validate in this paper. IPA, Ingenuity Pathway Analyzer; RWR, random walk with restarts.

databases. First, we added (incorrect) receptors/TRs to the input or removed correct receptors/TRs from the input and compared the resulting reconstructions (Figure 2e and Supplementary Section S3). When we deleted 30% of the receptors and 30% of the TRs from the input, the mean precision at recall of 0.3 and 0.6 dropped by 11% (from 0.42 to 0.38) and 27% (from 0.28 to 0.22), respectively, compared with the precision values with the correct inputs (Supplementary Figure S3). The results were similar for random additions of 30% of the receptors and 30% of the TRs.

Second, we evaluated the performance of recovering proteins in the reconstructions. At similar values of recall, PathLinker's precision for protein recovery was much higher than that for interaction recovery (Figure 2f). In fact, the precision values of all algorithms improved considerably (comparing Figures 2a,b with Supplementary Figure S4). When excluding proteins that have an interaction with at least one protein in the pathway, all algorithms have nearly perfect precision (Supplementary Figure S4).

Our analysis thus far relied on 15 pathways from a single database. Our last three assessments estimated the effect of interactions present only in NetPath and extended the scope of the analysis to a larger set of NetPath pathways and to the KEGG database. First, we estimated the reliance of our reconstructions on NetPath-only interactions by applying PathLinker to an



	Control		-Ryk		-CFTR		-Dab2	
Wnt	NRL	QNβ	NRL	QNβ	NRL	QNβ	NRL	QNβ
No Wnt	--	--	--	--	--	--	--	--
Wnt 1	VS	++	S	-	VS	+	VS	++
Wnt 2	VS	+	S	++	VS	++	VS	++
Wnt 2b2	W	-	W	+	S	++	W	-
Wnt 3	VS	++	S	++	VS	++	VS	++
Wnt 3a	VS	++	VS	++	VS	++	VS	++
Wnt 6	W	++	W	+	W	+	W	++
Wnt 7a	W	-	W	+	S	-	W	++
Wnt 7b	W	++	W	-	S	-	W	++
Wn 8a	W	-	W	-	W	++	W	++
Wnt 9b	W	-	W	-	W	-	W	++
Wnt 10b	W	-	W	-	W	++	S	++

Figure 4. Experimental validation of CFTR's effect on Wnt-mediated signaling. (**a**) Normalized TCF/LEF promoter-driven luciferase activity in the presence and absence of 11 different secreted Wnt (sWnt) proteins via transient expression. White bars signify a 30-fold greater activation in comparison to the No Wnt control. Gray bars signify less than 30-fold activation in comparison to the No Wnt control. (**b**) Efficacy of dose-dependent siRNA-mediated silencing of Ryk, Dab2, and CFTR on respective cellular protein levels and intracellular concentration of β-catenin as determined by western blot. (**c**, **d**) Normalized TCF/LEF promoter-driven luciferase activity post silencing of Ryk, Dab2, CFTR, or control scrambled siRNA in the presence or absence of 11 different sWnt proteins via transient expression. Graph is divided into two groups to better visualize differences between control scrambled siRNA and Ryk-, CFTR- or Dab2-specific siRNA silencing. (**e**) Intracellular concentration of β-catenin via western blot post silencing of Ryk, Dab2, or CFTR in the presence or absence of 11 different sWnt proteins via transient expression. Please refer to Supplementary Figure S9 for quantification of these data. (**f**) Summary of the correlation between Luciferase activity (**c**, **d**) and band intensity (**e**) under different experimental conditions. "++", $\geqslant$1.3-fold; "+", 1.3-fold$>x\geqslant$1-fold; "−", $<$1-fold; NRL, normalized relative luminescence; QNβ, qualification of normalized β-catenin intensity; VS, very strong ($\geqslant$30-fold); S, strong (30-fold$>x\geqslant$15-fold); W, weak ($<$15-fold). (**g**) Co-immunoprecipitation of endogenous Ryk and Dab2 using endogenous CFTR as a bait in the presence or absence of Wnts 1, 2, 2b2, 3, or 3a.

interactome that excluded these interactions. Only 4% of the interactions in the interactome were present in at least one NetPath pathway; further, 35% of these interactions were supported solely by NetPath (Supplementary Table S5). To evaluate the resulting reconstruction, we used the 65% of NetPath interactions that remained in the interactome as positives. While the proportion of positives in the interactome dropped from 4% to 2.6%, PathLinker's performance was comparable to that in the original interactome (Supplementary Figure S5). Next, we applied PathLinker and RWR to an expanded set of 29 NetPath pathways that contained at least one receptor and at least one TR, i.e., we removed the criterion that at least three paths should connect receptors to TRs in each pathway. We observed similar trends in performance on the expanded set as on the original set of 15 pathways (Supplementary Figure S6). When we ignored pathway-adjacent negatives, the precision of the reconstructions for the expanded set was smaller than for the original set. Nevertheless, PathLinker still clearly dominated over RWR (Supplementary Figure S6). Finally, we assessed the performance of PathLinker on another signaling pathway database. Accordingly, we computed aggregate precision and recall over the reconstructions of 32 KEGG signaling pathways that contained at least three paths from receptors to TRs, removing disease pathways from consideration (Figure 2g). The aggregate precision-recall curves for NetPath and KEGG pathways were comparable, with PathLinker performing slightly better on NetPath pathways at very low ($<$0.05) and high ($>$0.4) values of recall.

Wnt pathway reconstructions. We visualized the topologies of the Wnt pathway reconstructions from the PathLinker, RWR, and IPA at a recall of 0.20 (Figure 3a and Supplementary Table S6). We selected these three methods since every other approach achieved a recall of at most 0.13 for the Wnt pathway reconstructions (Supplementary Figure S7). In addition to the true positive interactions from NetPath (green edges), all three reconstructions contained interactions that are present in KEGG but missing from NetPath (purple edges). IPA had a slightly higher precision than PathLinker and RWR; however, the reconstruction contained 13 connected components, and only 3 TRs were connected to receptors. RWR's reconstruction contained two connected components and only two TRs. In contrast, PathLinker produced a reconstruction with many receptor-to-TR paths that contain NetPath and KEGG interactions, including 10 of the 13 TRs.

To more carefully explore the highest ranked paths in the PathLinker reconstruction, we examined the network formed by the top 200 paths computed by PathLinker using the receptors and TRs in the Wnt pathway in NetPath (Figure 3b). For this analysis, we added two receptors that were missing from the earlier precision-recall analysis (Supplementary Section S1). The PathLinker network included 16 proteins not previously known to be in the NetPath or KEGG representations of the Wnt pathway (gray or orange nodes in Figure 3b). Fifteen of these proteins are either involved in Wnt crosstalk, have been shown to be involved in β-catenin signaling in non-human models, or are involved in general post-translational protein modifications (Supplementary Section S5).

The remaining protein, CFTR, was the highest ranked of all proteins not previously known to be in Wnt pathway in the NetPath or KEGG databases. It appeared in the 59th path computed by PathLinker (Figure 3b). PathLinker indicated that CFTR acted as a signal transducer from Ryk, a receptor tyrosine kinase involved in Wnt signaling and organismal development,[21–24] to Dab, mitogen-responsive phosphoprotein, homolog 2 (Dab2), a known negative regulator of β-catenin signaling.[25,26] As Wnt signaling is associated with several types of cellular differentiation and specification, the closing of membrane channels to facilitate morphological changes is biologically relevant.[27]

Exploring the role of CFTR in Wnt signaling

We designed a series of experiments to determine the role of Ryk, CFTR, and Dab2 in Wnt/β-catenin-mediated signaling as predicted by PathLinker (blue region in Figure 3b). We utilized a quantitative TCF/LEF luciferase reporter assay and measurement of cellular β-catenin levels to determine if silencing of Ryk, CFTR, or Dab2 has a specific effect on Wnt/β-catenin signaling. We employed the Wnt plasmid library[28] to transiently express 11 different secreted Wnt proteins (referred to hereby as Wnt) in HEK293 cells. Transient expression of Wnts has been previously shown to induce the expression of luciferase enzyme driven by a synthetic, tandem TCF/LEF promoter when co-transfected into HEK293 cells.[28] We were able to determine and verify the extent of TCF/LEF-promoted luciferase activity by each of the 11 Wnt proteins tested (Figure 4a). Transient expression of Wnt 1, 2, 3, and 3a resulted in robust TCF/LEF-promoted luciferase activity ($\geqslant$30-fold), while Wnt 2b2, 6, 7a, 7b, 8a, 9b, and 10b promoted such activity to a much lesser extent ($<$30-fold) in comparison to control samples not treated with Wnt.

We then determined the efficacy of transient silencing of CFTR, Dab2, and Ryk by siRNA in HEK293 cells via western blot in a dose-dependent manner (Figure 4b). In the No Wnt control cells, cellular levels of β-catenin were not noticeably perturbed by siRNA silencing of CFTR and Ryk, but increased as cellular protein levels of Dab2 decreased. In these No Wnt control cells, we determined there were no significant changes in TCF/LEF-promoted luciferase activity in the absence of Ryk, Dab2, or CFTR. In the absence of Dab2 or CFTR, both TCF/LEF-promoted luciferase activity (Figures 4c,d) and β-catenin levels determined by western blot (Figure 4e) significantly increased for cells stimulated by nearly all Wnts (the exception being Wnt2b2 in the absence of Dab2 for measurement of β-catenin) in comparison to control scrambled siRNA-treated cells. Conversely, in the absence of Ryk, there was (i) significant ablation in TCF/LEF-promoted luciferase activity and (ii) decreased levels of cellular β-catenin in the presence of only Wnt 1, 2, 3, or 3a in comparison to control scrambled siRNA-treated cells. We noted no significant difference of TCF/LEF-promoted luciferase reporter

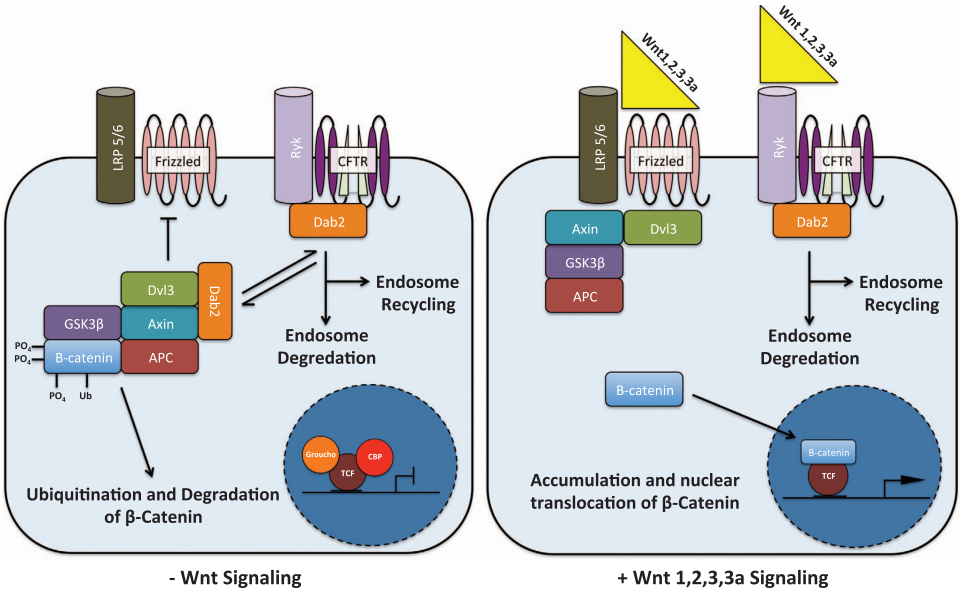


Figure 5. Suggested model for Ryk–CFTR–Dab2-mediated amplification of Wnt 1-, 2-, 3-, and 3a-specific signaling. In the absence of Wnt 1, 2, 3, and 3a, a subset of Dab2 is associated with either homeostatic recycling of CFTR or formation and maintenance of the β-catenin destruction complex. In the presence of these Wnts, Dab2 is recruited to the Ryk–CFTR membrane complex thereby allowing Axin and Dvl3 to be recruited to the LRP5/6-Frizzled membrane complex and facilitating the phosphorylation and degradation of Axin. Freed β-catenin is subsequently able to accumulate and translocate into the nucleus to catalyze gene-specific transcription.

activity or levels of cellular β-catenin for cells expressing Wnt 6, 7a, 7b, 8a, 9b, 10b in comparison to the control scrambled siRNA.

Cellular β-catenin levels determined by western blot were in accord with the activation of TCF/LEF promoter when stimulated by the respective Wnt (Figure 4e). In the presence of a stimulatory Wnt (specifically Wnt 1, 2, 3, and 3a), an increase in β-catenin levels in comparison to the No Wnt Control correlated with increased TCF/LEF-promoted luciferase activity (Figure 4f and Supplementary Figure S9). In instances where normalized relative luminescence was ablated, quantification of β-catenin was marginal or diminished as well (Figure 4f).

Utilizing endogenous CFTR as a bait, we were able to co-immunoprecipitate both Ryk and Dab2 in No Wnt control cells (Figure 4g). These interactions were qualitatively diminished in HEK293 cells transiently expressing Wnt 1, 2, 2b2, 3, and 3a. We hypothesize that Wnt-mediated receptor endocytosis triggers CFTR to the degradation pathway rather than membrane recycling, resulting in decreased cellular levels of CFTR and potentially Ryk and Dab2. Further studies on cellular trafficking of the Ryk–CFTR–Dab2 complex will provide insight into these results.

DISCUSSION

Reconstructing multiple pathways

We have considered two distinct types of algorithms: those that returned a single sub-network, producing a point on the precision-recall curve (ShortestPaths, ResponseNet, PCSF, and ANAT, BowTieBuilder, and IPA and those that provided a ranked list of interactions, producing precision-recall curves (PathLinker and RWR). In the case of IPA, since changing parameters yielded networks with substantially different precision and recall, we present results for this algorithm for nine parameter values. Since the single sub-network approaches had the goal of computing compact sub-networks that connected sources to targets, they were able to reconstruct pathways with high precision but only with low recall. Only the algorithms that offered a ranked list of interactions, PathLinker and RWR, reached a recall of $\geqslant 0.6$. These results showed that an important component of a pathway reconstruction algorithm was a parameter, such as *k*, whose increase caused a smooth variation and expansion of the resulting network. While both RWR and PathLinker had this property, only PathLinker offered an additional guarantee of connecting receptors to TRs (Figure 2d and the networks in Figure 3a). We conclude that PathLinker reconstructions captured the structure of signaling pathways much better than IPA and RWR, despite comparable performance in terms of precision and recall.

Several previous studies have focused on recovering only the proteins within a pathway, a methodology commonly used to predict the biological processes of which a protein may be a member.[29] All algorithms improved considerably when evaluating the proteins in the pathway reconstructions (Figure 2f), demonstrating that reconstructing the interactions within a pathway is a more challenging problem than that of recalling the proteins in the pathway. In addition, false positive interactions in reconstructions that are "near" the curated pathway may indeed represent valid interactions that have not yet been added to the pathway through the curation process (Figures 2b,c). High-confidence predictions adjacent to the pathway may be ideal candidates for further experimental studies aimed at expanding known signaling pathways.

Novel role of the Ryk–CFTR–Dab2 path in Wnt/β-catenin signaling

Wnt proteins are essential components of higher order eukaryotic development, cellular homeostasis, and wound healing. The canonical Wnt signaling pathway has been shown to be specific for a subset of Wnts, while other Wnts are known to signal through alternate means (reviewed in the study by MacDonald *et al.*[30]). Using 11 of the 19 known Wnts, we further this understanding by showcasing how the tested Wnts differentially activate the TCF/LEF promoter via β-catenin to significantly varying degrees. We show that Wnts 1, 2, 3, and 3a are capable of $\geq$ 30-fold activation of the TCF/LEF promoter, and do so in part via a novel Ryk–CFTR–Dab2 pathway that further regulates the cellular levels of β-catenin.

Ryk is a predicted tyrosine-protein kinase containing an extracellular WIF domain that has been previously shown to directly bind to Wnt 1 and Wnt 3a, though its signaling

mechanism was unknown.[23] Silencing of Ryk by siRNA in mice results in defects in axon guidance and neurite outgrowth in response to Wnt 3a induction.[22] The interaction between Ryk and CFTR was first determined in the CFTR interactome[30] and was not directly pertinent to the study's dissection of the Hsp90 co-chaperone, Aha1, and CFTR interaction.[31] We validated the Ryk–CFTR and CFTR–Dab2 interaction via co-immunoprecipitation. CFTR functions intrinsically as a membrane chloride ion channel protein and known point mutations result in impaired functionality resulting in the clinical manifestation of cystic fibrosis.[32] CFTR is impacted by intracellular calcium (reviewed in the study by Antigny *et al*.[33]), an alternate product of certain non-canonical Wnt signaling pathways.[33,34] Dab2 is involved in endosomal recycling and degradation of CFTR and is a well-known regulatory component of receptor-mediated endocytosis.[35,36] Dab2 also functions as a negative regulator of the β-catenin destruction complex.[26,37,38] Even though prior groups had previously identified these functionalities independently, there was no evidence or speculation for the role of CFTR in Wnt/β-catenin-mediated signaling particularly by Ryk or Dab2.

We present a model incorporating the Ryk–CFTR–Dab2 pathway as an amplifier of Wnt 1-, 2-, 3-, and 3a-specific β-catenin signaling (Figure 5). Our results suggest the recruitment of Dab2 to the Ryk–CFTR membrane complex in the presence of specific Wnt proteins. This process further impedes the formation of the β-catenin destruction complex, thereby freeing additional β-catenin to further amplify TCF/LEF promoter transcription. It is currently unknown if Wnt signaling via Ryk modifies the sodium transport function of CFTR in preparation for context specific cellular processes or if Wnt-specific signaling facilitates the degradation of CFTR. Further molecular characterizations are required to provide insight into the novel role of CFTR in facilitating Wnt 1-, 2-, 3-, and 3a-specific signaling.

In conclusion, we have presented PathLinker, an algorithm that automates the reconstruction of human signaling pathways by connecting the receptors and TRs for a pathway through a physical and regulatory interaction network. Based on our comprehensive analysis on 15 NetPath pathways, PathLinker achieved much higher recall (while maintaining reasonable precision) than several other methods. Furthermore, it was the only method that could control the size of the reconstruction while ensuring that receptors were connected to TRs in the result. PathLinker's reconstruction of the Wnt pathway indicated that CFTR facilitates the signaling from Ryk to Dab2. In HEK293 cells, we validated this path experimentally and showed its specificity for 4 of the 11 Wnts tested (Wnt 1, 2, 3, and 3a). Based on these results, we propose a model that suggests Dab2 is recruited to the Ryk–CFTR membrane complex in response to a defined Wnt stimulus that ultimately amplifies Wnt 1, 2, 3, and 3a canonical signaling. In summary, PathLinker provides a promising framework for reconstructing a well-studied signaling pathway given relatively little information about its components. It may serve as a powerful approach for discovering the structure of poorly studied processes and prioritizing both proteins and interactions for experimental study.

MATERIALS AND METHODS

PathLinker

The problem of pathway reconstruction takes as input (i) a weighted directed interactome G containing physical and regulatory interactions between pairs of proteins, (ii) the receptors S in a signaling pathway of interest, and (iii) the TRs T in the same pathway. A reconstruction of a pathway P consists of a sub-network of G that connects the receptors in P to the TRs in P using proteins and interactions in G.

Given an interactome $G=(V, E)$, where every edge e in E has an associated weight w_e between 0 and 1, a receptor set S, a TR set T, and a user-defined parameter k, PathLinker computes the k highest scoring loopless paths that begin at any receptor in S and terminate at any TR in T. We define the score of a path to be the product of the edge weights along the path. We add an artificial source s with a directed edge (s, x) for each node $x \in S$ and an artificial sink t with a directed edge (y, t) for each node $y \in T$. We assign the following cost to each edge (u, v):

$$c_{uv} = \begin{cases} -\log(w_{uv}) & \text{if } u, v \in V \setminus \{s, t\} \\ 0 \text{ if } u = s \text{ or } v = t. \end{cases}$$

Let the cost of a path be the sum of the costs of the edges in the path. Therefore, the least costly $s \rightsquigarrow t$ path is equivalent to the path from S to T that maximizes the path score. PathLinker computes the k highest scoring paths in this modified graph by incorporating a novel integration of Yen's algorithm[39] with the A* heuristic (Supplementary Section S6). This technique is up to 41 times faster than Yen's algorithm by itself (Supplementary Figure S8) and is thus capable of handling the complexity of human interaction networks and signaling pathways.

We compute a pathway reconstruction G_k for each value of k by taking the union of the k highest scoring paths. By construction, the interactions in the k shortest paths are a subset of those in the $(k+1)$ shortest paths, thereby ensuring that our reconstructions vary smoothly with k. For precision and recall calculations, we compute $k = 20{,}000$ paths and rank each node and edge by the index of the first path in which it appears. This value of k reflects the high degree of redundancy (edge reuse) among paths in signaling networks.

Data sets

We constructed a directed human protein interactome from numerous protein–protein interaction and signaling pathway databases.[3–5,40] The resulting network contained 12,046 nodes and 152,094 directed edges, where multiple types of evidence supported many of the edges. We weighted each edge in the network using a Bayesian approach that computes interaction probabilities based on the sources of evidence.[17] We identified sets of signaling receptors and TRs from previously published lists of human receptors[41] and TRs.[42,43] We selected 15 NetPath pathways and 32 KEGG pathways that each contained at least one receptor, at least one TR, and were connected by at least three paths (Supplementary Tables S3 and S4). For more information, refer to Supplementary Section S1.

Experimental methods

We conducted experiments in HEK293 cells using the public Wnt plasmid Library[28] and validated siRNA. We present detailed methods in Supplementary Section S7.

ACKNOWLEDGEMENTS

The National Institute of General Medical Sciences of the National Institutes of Health grant R01-GM095955 (TMM), National Science Foundation (NSF) grant DBI-1062380 (TMM), National Research Service Award F32-ES024062 (ANT), and Environmental Protection Agency grant EPA-RD-83499801 (TMM) supported this work. We also acknowledge funding from the ICTAS Center for Systems Biology of Engineered Tissues at Virginia Tech (TMM). We thank Brendan Avent and Peter Burnham for their help in preparing the manuscript.

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

1. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
2. Schaefer, C. F. *et al.* PID: the pathway interaction database. *Nucleic Acids Res.* **37**, D674–D679 (2009).
3. Kandasamy, K. *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* **11**, R3 (2010).
4. Paz, A. *et al.* SPIKE: a database of highly curated human signaling pathways. *Nucleic Acids Res.* **39**, D793–D799 (2011).
5. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
6. Bailly-Bechet, M. *et al.* Finding undetected protein associations in cell signaling by belief propagation. *Proc. Natl Acad. Sci. USA* **108**, 882–887 (2011).

7. Gitter, A., Klein-Seetharaman, J., Gupta, A. & Bar-Joseph, Z. Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.* **39**, e22 (2011).
8. Huang, S. S. & Fraenkel, E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci. Signal.* **2**, ra40 (2009).
9. Komurov, K., White, M. A. & Ram, P. T. Use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data. *PLoS Comput. Biol.* **6**, e1000889 (2010).
10. Mohammadi, S., Subramaniam, S. & Grama, A. Inferring the effective TOR-dependent network: a computational study in yeast. *BMC Syst. Biol.* **7**, 84 (2013).
11. Scott, J., Ideker, T., Karp, R. M. & Sharan, R. Efficient algorithms for detecting signaling pathways in protein interaction networks. *J. Comput. Biol.* **13**, 133–144 (2006).
12. Silverbush, D. & Sharan, R. Network orientation via shortest paths. *Bioinformatics* **30**, 1449–1455 (2014).
13. Steffen, M., Petti, A., Aach, J., D'Haeseleer, P. & Church, G. Automated modelling of signal transduction networks. *BMC Bioinformatics* **3**, 34 (2002).
14. Stojmirovic, A. & Yu, Y. K. ITM Probe: analyzing information flow in protein networks. *Bioinformatics* **25**, 2447–2449 (2009).
15. Tuncbag, N. *et al.* Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *J. Comput. Biol.* **20**, 124–136 (2013).
16. Ingenuity Pathway Analysis (IPA). IPA Network Generation Algorithm. http://www.ingenuity.com/wp-content/themes/ingenuity-qiagen/pdf/ipa/IPA-netgen-algorithm-whitepaper.pdf (2005).
17. Yeger-Lotem, E. *et al.* Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. *Nat. Genet.* **41**, 316–323 (2009).
18. Yosef, N. *et al.* Toward accurate reconstruction of functional protein networks. *Mol. Syst. Biol.* **5**, 248 (2009).
19. Supper, J. *et al.* BowTieBuilder: modeling signal transduction pathways. *BMC Syst. Biol.* **3**, 67 (2009).
20. Haveliwala, T. H. Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Trans. Knowl. Data Eng.* **15**, 784–796 (2003).
21. Bovolenta, P., Rodriguez, J. & Esteve, P. Frizzled/RYK mediated signalling in axon guidance. *Development* **133**, 4399–4408 (2006).
22. Keeble, T. R. *et al.* The Wnt receptor Ryk is required for Wnt5a-mediated axon guidance on the contralateral side of the corpus callosum. *J. Neurosci.* **26**, 5840–5848 (2006).
23. Lu, W. G., Yamamoto, V., Ortega, B. & Baltimore, D. Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. *Cell* **119**, 97–108 (2004).
24. Yoshikawa, S., McKinnon, R. D., Kokel, M. & Thomas, J. B. Wnt-mediated axon guidance via the Drosophila derailed receptor. *Nature* **422**, 583–588 (2003).
25. Jiang, Y., He, X. & Howe, P. H. Disabled-2 (Dab2) inhibits Wnt/beta-catenin signalling by binding LRP6 and promoting its internalization through clathrin. *EMBO J.* **31**, 2336–2349 (2012).
26. Jiang, Y., Luo, W. & Howe, P. H. Dab2 stabilizes Axin and attenuates Wnt/beta-catenin signaling by preventing protein phosphatase 1 (PP1)-Axin interactions. *Oncogene* **28**, 2999–3007 (2009).
27. Chen, Y. *et al.* Aquaporin 2 promotes cell migration and epithelial morphogenesis. *J. Am. Soc. Nephrol.* **23**, 1506–1517 (2012).
28. Najdi, R. *et al.* A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. *Differentiation* **84**, 203–213 (2012).
29. Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* **9**(Suppl 1), S4 (2008).
30. MacDonald, B. T., Tamai, K. & He, X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev. Cell* **17**, 9–26 (2009).
31. Wang, X. D. *et al.* Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* **127**, 803–815 (2006).
32. Gadsby, D. C., Vergani, P. & Csanady, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483 (2006).
33. Antigny, F., Norez, C., Becq, F. & Vandebrouck, C. CFTR and Ca2+ signaling in cystic fibrosis. *Front. Pharmacol.* **2**, 67 (2011).
34. De, A. Wnt/Ca2+ signaling pathway: a brief overview. *Acta Bioch. Bioph. Sin.* **43**, 745–756 (2011).
35. Fu, L. W. *et al.* Dab2 is a key regulator of endocytosis and post-endocytic trafficking of the cystic fibrosis transmembrane conductance regulator. *Biochem. J.* **441**, 633–643 (2012).
36. Cihil, K. M. *et al.* Disabled-2 protein facilitates assembly polypeptide-2-independent recruitment of cystic fibrosis transmembrane conductance regulator to endocytic vesicles in polarized human airway epithelial cells. *J. Biol. Chem.* **287**, 15087–15099 (2012).
37. Hocevar, B. A. *et al.* Regulation of the Wnt signaling pathway by disabled-2 (Dab2). *EMBO J.* **22**, 3084–3094 (2003).
38. Jiang, Y., Prunier, C. & Howe, P. H. The inhibitory effects of Disabled-2 (Dab2) on Wnt signaling are mediated through Axin. *Oncogene* **27**, 1865–1875 (2008).
39. Yen, J. Y. Finding the k shortest loopless paths in a network. *Manag. Sci.* **17**, 712–716 (1971).
40. Aranda, B. *et al.* PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods* **8**, 528–529 (2011).
41. Almen, M. S., Nordstrom, K. J., Fredriksson, R. & Schioth, H. B. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**, 50 (2009).
42. Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140**, 744–752 (2010).
43. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).

Supplementary Information accompanies the paper on the *npj Systems Biology and Applications* website (http://www.nature.com/npjsba)

Supplementary Information for Pathways on Demand: Automated Reconstruction of Human Signaling Networks

Anna Ritz[1,4], Christopher L. Poirel[1,5], Allison N. Tegge[1], Nicholas Sharp[1,6], Kelsey Simmons[2], Allison Powell[2], Shiv D. Kale[2], and T. M. Murali [*1,3]

[1]*Department of Computer Science, Virginia Tech, Blacksburg, VA, USA.*
[2]*Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA.*
[3]*ICTAS Center for Systems Biology of Engineered Tissues, Virginia Tech, Blacksburg, VA, USA.*

[4]*Current Address: Department of Biology, Reed College, Portland, OR, USA.*
[5]*Current Address: RedOwl Analytics, 171 2nd Street, 3rd Floor, San Francisco, CA, USA.*
[6]*Current Address: Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.*

Supplementary Information

1 Datasets

Human interactome. We constructed a directed human protein interactome from numerous protein-protein interaction and signaling pathway databases. The interactome consisted of nodes representing proteins, bi-directed edges representing physical interactions, and directed edges representing regulatory/signaling interactions. The interactome included 40,447 physical interactions between protein pairs downloaded using PSICQUIC [9] from the following databases: BIND, DIP, InnateDB, IntAct, MINT, MatrixDB, and Reactome. We ignored interactions from PSICQUIC that were computationally predicted, functional, or from unspecified experimental methods (Supplementary Table S1). We identified signaling interactions from three pathway databases: 382 signaling interactions and 3,414 physical interactions from NetPath [10], 20,154 signaling interactions and 2,286 physical interactions from KEGG [11], and 12,093 signaling interactions and 41,314 physical interactions from SPIKE [12]. The signaling pathway databases often annotated interactions differently. For

[*]murali@cs.vt.edu

example, a NetPath physical interaction may be represented in KEGG as a signaling interaction. We used this information to replace 2,856 physical interactions by the more informative directed signaling interaction. The resulting network contained 12,046 nodes and 152,094 directed edges, where many of the edges were supported by multiple types of evidence. Note that by construction, the NetPath and KEGG signaling pathways were subgraphs of the human interactome. However, we did not annotate these interactions with the identities of the pathways of which they were members. We used UniProtKB protein identifiers for all analyses.

Table S1: Physical interaction experimental evidence codes from PSICQUIC that we ignored during interactome construction.

Proteomics Standard Initiative Molecular Interaction (PSI-MI) ID	Description
MI:0036	domain fusion
MI:0046	experimental knowledge based
MI:0063	interaction prediction
MI:0064	interologs mapping
MI:0085	phylogenetic profile
MI:0087	predictive text mining
MI:0105	structure based prediction
MI:0363	inferred by author
MI:0364	inferred by curator
MI:0686	unspecified method coexpression
MI:0045	experimental interaction detection

Weighting the human interactome. We weighted each edge in the network using a Bayesian approach that computes interaction probabilities [6]. This method assigns a high probability to an interaction that is supported by evidence that connects proteins co-annotated to the same set of user-specified biological processes. The weighting scheme takes as input the human interactome annotated with experimental evidence sources and a set of GO terms. We used the experimental evidence codes supplied by PSICQUIC, KEGG edges (divided into interaction types), NetPath edges, and SPIKE edges as sources of evidence in the interactome. We selected the GO term "regulation of signal transduction" and eight other terms that were (i) children of the "signal transduction" and (ii) annotated more than 50 genes (Supplementary Table S2). From these GO terms, we established the set of positives as all pairs of proteins co-annotated to the same GO term. We also established the set of negatives as pairs that were not co-annotated to the same GO term, sub-sampling this set so that it was 10 times as large as the positive set. We computed the probability that each source of evidence connects pairs of proteins co-annotated to the same GO term and used these data to compute the probability of each edge. Many evidence probabilities were close to 1. To mitigate the effect of these evidence types on our algorithms, we set a threshold of 0.75 on all probabilities [6].

Table S2: GO terms considered for evidence-based weighting. The first eight terms are children of the "signal transduction" GO term.

# Genes	GO Term	GO Description
155	GO:0030522	intracellular receptor signaling pathway
387	GO:0002764	immune response-regulating signaling pathway
73	GO:0030968	endoplasmic reticulum unfolded protein response
252	GO:0097190	apoptotic signaling pathway
80	GO:0007602	phototransduction
1898	GO:0007166	cell surface receptor signaling pathway
1023	GO:0035556	intracellular signal transduction
179	GO:0023014	signal transduction by phosphorylation
1728	GO:0009966	regulation of signal transduction

Receptor and TR lists. We identified a set of 2,124 signaling receptors from a previously-published list of human signal receptors [13]. In addition, we manually included three members of the CD3-TCR complex (CD3D, CD3E, and CD3G), which serve as receptors for the T Cell Receptor pathway that were not present in the published list. We retrieved a set of 2,286 human TRs reported in two studies: i) all TRs listed by Ravasi *et al.* [14] and ii) high-quality TRs from Vaquerizas *et al.* [15]. The latter classified TRs as 'a', 'b', 'c', 'x', and 'other'. We took only TRs classified as 'a', 'b', or 'other' because TRs in these classes have experimental evidence of regulatory function in a mammalian organism or were manually curated to be TRs. We identified the receptors and TRs in each signaling pathway by taking the intersection of the proteins in the pathway with the list of receptors and list of TRs.

The precision and recall results were determined solely by running PathLinker and other algorithms with the receptor and TR lists described above. When we carefully examined the NetPath receptors for the Wnt pathway, we observed that two Frizzled receptors, FZD4 and FZD6, were missing from the literature-determined lists. For analysis to identify potential hypotheses for followup in the lab, we manually added these receptors to the PathLinker inputs and re-ran PathLinker.

NetPath pathways. We identified 15 NetPath pathways that met the following criteria: i) the pathway contained at least one receptor, ii) the pathway contained at least one TR, and iii) the minimum cut between the receptors and TRs was at least three in the NetPath pathway (i.e., three edges must be removed from the pathway to disconnect the receptors from the TRs) (Supplementary Table S3). The first two criteria ensured that each pathway had a natural beginning and end to the signal propagation. The third criteria ensured the pathway was sufficiently connected. We included the third criterion because several pathways had a minimum cut of zero; such curated pathways were likely highly incomplete as there was no connection (path) from any signaling receptor to a downstream TR. We did not consider the Notch pathway since its receptors have intracellular domains that are also TRs. We downloaded NetPath SBML Level 2 Version 1 files from `http://www.netpath.org`. These files represent interactions as a set of reactants, products, and modifiers; we treated each (modifier,reactant) pair as a pairwise interaction. We treated interactions denoted as 'phys-

ical' or 'interaction' as bi-directed and all other types were directed (e.g., 'phosphorylation,' 'methylation,' and 'acetylation').

Table S3: NetPath pathways used for analysis. Recoverable receptors/TRs are those remaining after removing incoming edges to receptors and outgoing edges from TRs (see "Evaluation Framework").

Pathway	#Nodes	#Edges	Min Cut	# Receptors	# TRs	# Recoverable Receptors	# Recoverable TRs
BDNF	72	139	4	5	4	5	4
EGFR1	231	1456	30	6	33	6	33
IL1	43	178	7	3	5	3	5
IL2	67	242	16	3	12	3	12
IL3	70	176	5	2	9	2	9
IL6	53	162	6	4	14	4	14
IL7	18	52	5	2	3	2	3
Kit Receptor	76	207	5	6	8	6	8
Leptin	55	135	8	3	15	3	15
Prolactin	68	199	10	4	10	4	10
RANKL	57	142	4	2	12	2	12
TCR	154	504	8	7	20	6	20
TGFβ Receptor	209	863	32	5	77	5	77
TNFα	239	913	15	4	44	4	44
Wnt	106	428	7	14	14	14	13

KEGG pathways. The KEGG database contains 276 human pathways divided into six categories: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, and Human Diseases. We focused on Environmental Information Processing, Cellular Processes, and Organismal Systems since these groups contained signaling related pathways. We ignored pathways in the Metabolism and Genetic Information Processing categories since they were not related to signaling. We did not consider the Human Diseases category either, since our goal in this work was to focus on normal physiological processes. Each of these categories contains several subgroups of pathways. We considered only those subgroups related to signaling. Of the remaining 54 KEGG pathways, we analyzed the 32 pathways that met the following criteria: i) the pathway contained at least one receptor, ii) the pathway contained at least one TR, and iii) the minimum cut between the receptors and TRs was at least three in the KEGG pathway. (Supplementary Table S4). We parsed the KEGG KGML pathway files, an XML-style file format specific to KEGG pathways. Our parser follows the description of the KEGG Markup Language (KGML) available at `http://www.kegg.jp/kegg/xml/docs/`. We parsed KEGG entries that corresponded to genes, proteins, and complexes (`gene` and `group` types). We collected UniProtKB identifiers from the original KGML files. We retained only "reviewed" UniProtKB identifiers, as defined by the UniProtKB database. If a single KEGG identifier mapped to multiple reviewed UniProtKB identifiers, then we duplicated the information for

each UniProtKB identifier.We parsed the protein-protein relations (`PPRel`), treating interactions as bi-directed edges if they were denoted as 'binding/association' or 'dissociation', or if they are components of the same complex. We treated all other interaction types (e.g, 'activation', 'inhibition', 'phosphorylation') as directed edges. KEGG contained information about interactions between protein families, e.g., `Wnt` and `Fzd`. In this case, we considered each (`Wnt`,`Fzd`) protein pair as a separate interaction.

Table S4: KEGG pathways used for the aggregate precision and recall computation.

Name	KEGG ID	Name	KEGG ID
Adherens junction	hsa04520	Adipocytokine signaling pathway	hsa04920
Apoptosis	hsa04210	Axon guidance	hsa04360
Chemokine signaling pathway	hsa04062	Circadian entrainment	hsa04713
Dopaminergic synapse	hsa04728	Endocytosis	hsa04144
ErbB signaling pathway	hsa04012	Focal adhesion	hsa04510
FoxO signaling pathway	hsa04068	GnRH signaling pathway	hsa04912
HIF-1 signaling pathway	hsa04066	Hippo signaling pathway	hsa04390
Insulin signaling pathway	hsa04910	Jak-STAT signaling pathway	hsa04630
Prolactin signaling pathway	hsa04917	MAPK signaling pathway	hsa04010
Melanogenesis	hsa04916	Natural killer cell mediated cytotoxicity	hsa04650
Neurotrophin signaling pathway	hsa04722	NF-kappa B signaling pathway	hsa04064
Notch signaling pathway	hsa04330	Osteoclast differentiation	hsa04380
TGF-beta signaling pathway	hsa04350	Thyroid hormone signaling pathway	hsa04919
Tight junction	hsa04530	Toll-like receptor signaling pathway	hsa04620
VEGF signaling pathway	hsa04370	Wnt signaling pathway	hsa04310
Leukocyte transendothelial migration	hsa04670	Signaling pathways regulating pluripotency of stem cells	hsa04550

We found that there were relatively few TRs from Ravasi *et al.* and Vaquerizas *et al.* that appeared in KEGG pathways. On average, there were about twice the number of TRs from these lists that appeared in NetPath pathways compared to KEGG (18.6 and 9.8, respectively). KEGG pathways contained on average 11.9 proteins that were not in the TR lists but had no outgoing edges in the interactome, which may be considered alternate "targets" for PathLinker. The number of such proteins was much smaller for NetPath pathways (4.1 on average). For the KEGG analysis, we included proteins that have no outgoing edges as end-points for PathLinker, in addition to the TRs.

2 Algorithms for Comparison

We briefly describe each algorithm and discuss the parameters we use (Supplementary Figure S1). Unless otherwise specified, we run all methods on a weighted, directed network.

RWR [3] is a random walk with restarts, also known as a teleporting random walk or topic-based PageRank. At each step, a walker moves to a neighbor with probability (1-q) and "restarts" at one of the receptors with probability q. In practice, the interactome we use is aperiodic (since there is at least one cycle of length 2 and at least one cycle of length 3),

but not necessarily irreducible. To ensure irreducibility, we add edges from each node to all other nodes in the interactome with a small teleportation probability of $1/(|V| \times 10^6)$. We use the well-known power iteration method to efficiently compute the stationary distribution of the random walk. We compute the flux score for edge (u, v) by multiplying the visitation probability of u by the edge weight and normalizing by the weighted out degree of u. We rank the interactions and proteins of a reconstruction by decreasing order of flux score.

ANAT [4] returns a sub-network connecting receptors to TRs that allows a trade-off between shortest paths and minimum Steiner trees with a parameter α. We ran the `steinprt` software package for ANAT to compute one sub-network for each signaling pathway. We selected $\alpha = 0$ since it achieved higher precision than all other values of α on NetPath pathways.

PCSF [5] solves a Prize-Collecting Minimum Steiner Forest problem using a message passing algorithm, which returns a single sub-network. We introduced a source node connected to all receptors, set all TRs as terminal nodes, and ran the `msgsteiner` software package to identify a set of Steiner trees. PCSF takes two parameters: p, the value of the prize for each terminal and ω, the penalty on the number of trees. Different parameter values produced similar precision and recall on NetPath pathways; we set $p = 5$ and $\omega = 0.10$.

ResponseNet [6] uses a min-cost network flow approach to identify a sub-network that connects receptors to TRs. We implemented ResponseNet in Python and solved the linear program using CPLEX. ResponseNet requires a parameter γ that controls the number of interactions that carry flow. Different parameter values produced similar precision and recall on NetPath pathways; we set $\gamma = 20$. In principle, we could rank interactions and proteins of a reconstruction by decreasing order of flow. However, since ResponseNet typically yielded non-zero flow on a small number of edges, we included any node with incoming positive flow and any edge with positive flow in the output network.

The Ingenuity Pathway Analyzer (IPA) contains many algorithms that identify subsets of their interactome. IPA's Network Generation algorithm identifies a sub-network that links user-specified nodes [7]. We implemented this algorithm for comparison, calling it IPA. It operates on an unweighted network, and requires a parameter $n_{\max}$ that determines the size of the computed networks. We ran IPA on an unweighted version of the interactome using multiple values of $n_{\max}$, since different parameter values returned sub-networks with different values of precision and recall.

ShortestPaths computes shortest paths between receptors and TRs. Specifically, for every receptor r and every TR t, we identify the shortest path between r and t. When there are multiple shortest paths between r and t, we include all of them. We output a network composed of the union of all shortest paths computed for all receptor-TR pairs. Note that this algorithm is a variation of ANAT with $\alpha = 0$.

BowTieBuilder uses a heuristic approach to compute a Steiner tree connecting receptors to TRs [8]. First, BowTieBuilder initializes the reconstructed pathway P to include the set of receptors and TRs, and sets all receptors and TRs as *unvisited*. Next, BowTieBuilder compute a distance matrix D containing the length of the shortest path from every receptor r to every TR t. BowTieBuilder then iteratively selects the shortest path in D that connects an *unvisited* node and a *visited* node. If there is no such path, it identifies the shortest path between any two *unvisited* nodes. The algorithm adds this path

to the network P and marks all the nodes along the path as *visited*. BowTieBuilder then updates the matrix D to include the length of the shortest path from any receptor or TR to nodes along the added path. BowTieBuilder repeats these steps until all receptor and TR nodes are marked as *visited*. The network P represents the reconstructed pathway.

Parameters used by the algorithms

Algorithm	Parameter	Meaning
RWR	q	Teleportation probability
ANAT	α	Tradeoff between global (Steiner tree) and local (shortest path) solution
PCSF	ω	Penalty for adding a new tree
	p	Prize for each node
ResponseNet	γ	Number of interactions that carry flow
IPA	$n_{\max}$	Maximum sub-network size

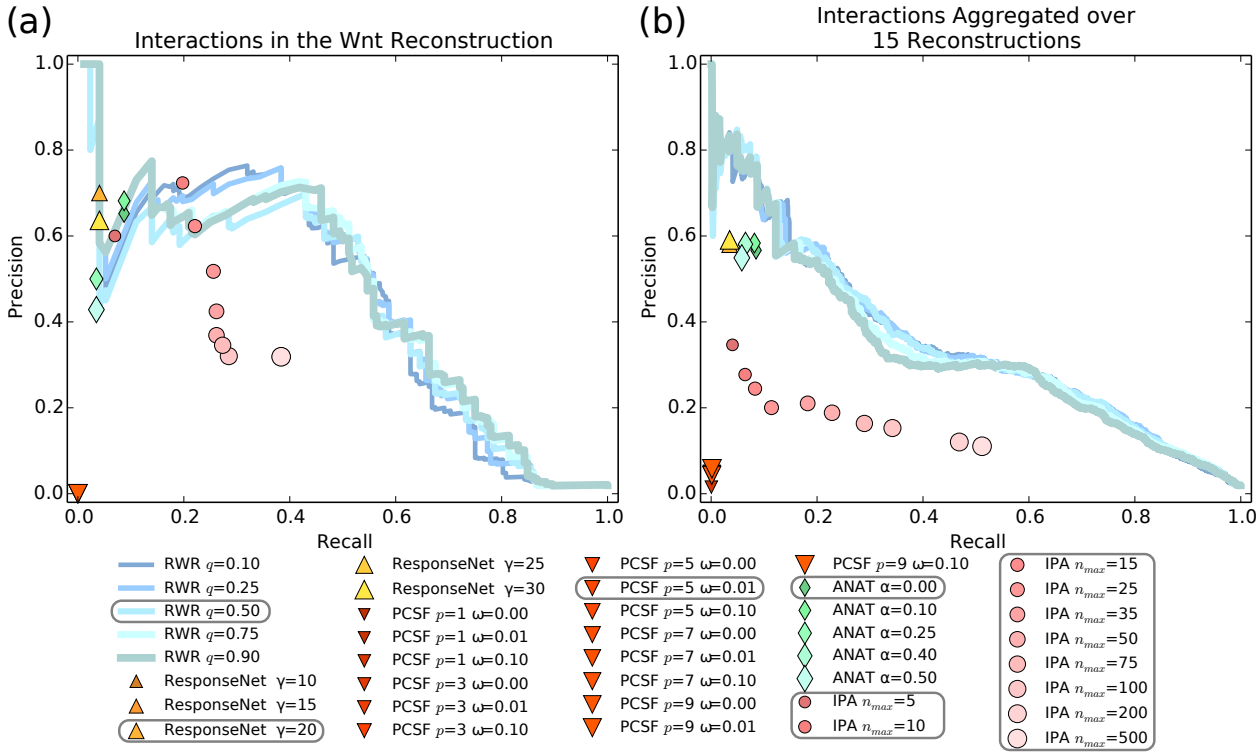


Figure S1: (Top) Description of user-defined parameters for algorithms. (Bottom) Precision and recall of interactions in the (a) Wnt pathway reconstruction and (b) 15 aggregated NetPath reconstructions with variation in internal parameters. Gray rounded rectangles denote parameter values that we used in the precision-recall analysis in the main manuscript.

3 Evaluation Framework

Single pathway. Given a curated pathway and the weighted interactome G, we performed the following steps to compute precision and recall. We identified the receptors and TRs in the curated pathway using the receptor and TR lists. We called these *pathway receptors* and *pathway TRs*. We removed edges incoming to the pathway receptors and edges outgoing from pathway TRs from G. We performed this step before running PathLinker to ensure that each path contained exactly one receptor and exactly one TR. We performed this step for all the other algorithms as well, since we found that it improved their precision. We applied each algorithm to G, using the pathway receptors as the sources S and the pathway TRs as the targets T. We ranked the interactions (or proteins) in the solution returned by each algorithm. For single sub-network solutions, we took the entire set of interactions. For PathLinker, we ranked each interaction by the first path in which it appeared (increasing order). For RWR, we ranked the interactions by edge flux score (decreasing order).

We identified the set P of positive interactions as those present in the curated pathway (ignoring direction). We identified a set N of negative interactions as follows. Ideally, we would have liked to use a curated dataset of negative examples. However, we are not aware of a database that contains interactions that are not in any signaling pathway. Therefore, we adopted the longstanding convention in the computational biology community of sampling negative examples randomly from the universe [16–19], which in our application was the set of all interactions in the interactome. We randomly sub-sampled a negative set N of edges (ignoring direction) from the background interactome in the ratio of 50 negatives to one positive, ensuring that N did not contain any edges in P. We acknowledge that the choice of 50 is arbitrary and that each algorithm's performance will depend on this number. However, since we only used N in the estimation of precision, the choice of 50 does not affect the output of the individual algorithms but only their relative performance. In the analyses where we ignored KEGG positives or ignored pathway-adjacent negatives, we removed these interactions from G *before* subsampling N.

We computed the precision and recall using the positive set P, the negative set N, and the ranked interactions X. Let X_i denote the set of the first i interactions. The precision and recall for X_i were

$$\text{Precision}_i = \frac{|X_i \cap P|}{i} \quad \text{and} \quad \text{Recall}_i = \frac{|X_i \cap P|}{|P|}. \tag{1}$$

We applied a similar method for computing the precision and recall when we reconstructed the proteins in a curated pathway.

Multiple pathways. We computed the precision and recall for a set of m signaling pathways $p_1, p_2, \ldots, p_m$. After computing the precision and recall for each pathway individually, we had m distinct collections of ranked edges, positive edges, and negative edges, denoted as $X^{(j)}$, $P^{(j)}$, and $N^{(j)}$, respectively. We aggregated the ranked lists by appending the pathway name to the edge, i.e., we computed,

$$X = \bigcup_{j=1}^{m} \left[((e, p_j), k) \text{ for } e, k \in X^{(j)} \right],$$

where e was an edge in pathway p_j and k was the rank of that edge in $X^{(j)}$. Finally, we sorted the elements in X by the value k. We similarly appended the pathway name to the positives and negatives:

$$P = \bigcup_{j=1}^{m} \left[(p, p_j) \text{ for } p \in P^{(j)}\right] \quad \text{and} \quad N = \bigcup_{j=1}^{m} \left[(n, p_j) \text{ for } n \in N^{(j)}\right].$$

We used these three aggregated collections to compute precision and recall for X, P, and N using Equation (1). We computed aggregate precision and recall for nodes in a similar manner.

Quantifying Distance in the Interactome To calculate the distance from an edge in a reconstructed pathway to the signaling pathway (such as the Wnt signaling pathway in NetPath), we defined a measure δ based on the shortest path length. We first describe $\delta(n)$, the distance from a node u to the signaling pathway. We computed the shortest path length $d(u, v)$ from u to every node v in the pathway using Dijkstra's algorithm; we ignored direction in this calculation. Let V_P be the set of nodes in the signaling pathway (the positive set). We defined

$$\delta(u) = \min_{v \in V_P} d(u, v),$$

where $\delta(u) = 0$ if node u is in the signaling pathway. Let E_p be the set of edges in the signaling pathway. We defined $\delta(u, v)$, the distance from edge (u, v) to the signaling pathway, as

$$\delta(u, v) = \begin{cases} 0 & \text{if } (u, v) \in E_p \\ \min\left(\delta(u), \delta(v)\right) + 1 & \text{otherwise.} \end{cases}$$

Intuitively, $\delta((u, v))$ is 0 if (u, v) is in the pathway. Otherwise, it is the length of the shortest path connecting the edge to the pathway. Note that $\delta(u, v) = 1$ for an edge (u, v) that is not a member of the pathway, even if u and v are proteins in the pathway. For a ranked list of edges in a pathway reconstruction, we visualized the distribution of these distances δ as a bar chart.

Sampling receptors and TRs. We define a *sampling percentage* ρ relative to the pathway receptors S and pathway TRs T. For example, when $\rho = -30\%$, we omit 30% of the receptors and 30% of the TRs. When $\rho = 30\%$, we add 30% new receptors and 30% new TRs. When $\rho = 0\%$, we use the correct receptors and TRs. We considered $\rho = [-50\%, -30\%, -10\%, 0\%, 10\%, 30\%, 50\%]$. For each non-zero value of ρ and for each NetPath pathway P, we randomly generate 25 sets of receptors and TR and apply PathLinker to each set. For each value of ρ, we compute the median precision-recall curve by partitioning the recall values into 1,000 bins.

4 Precision and Recall Analyses

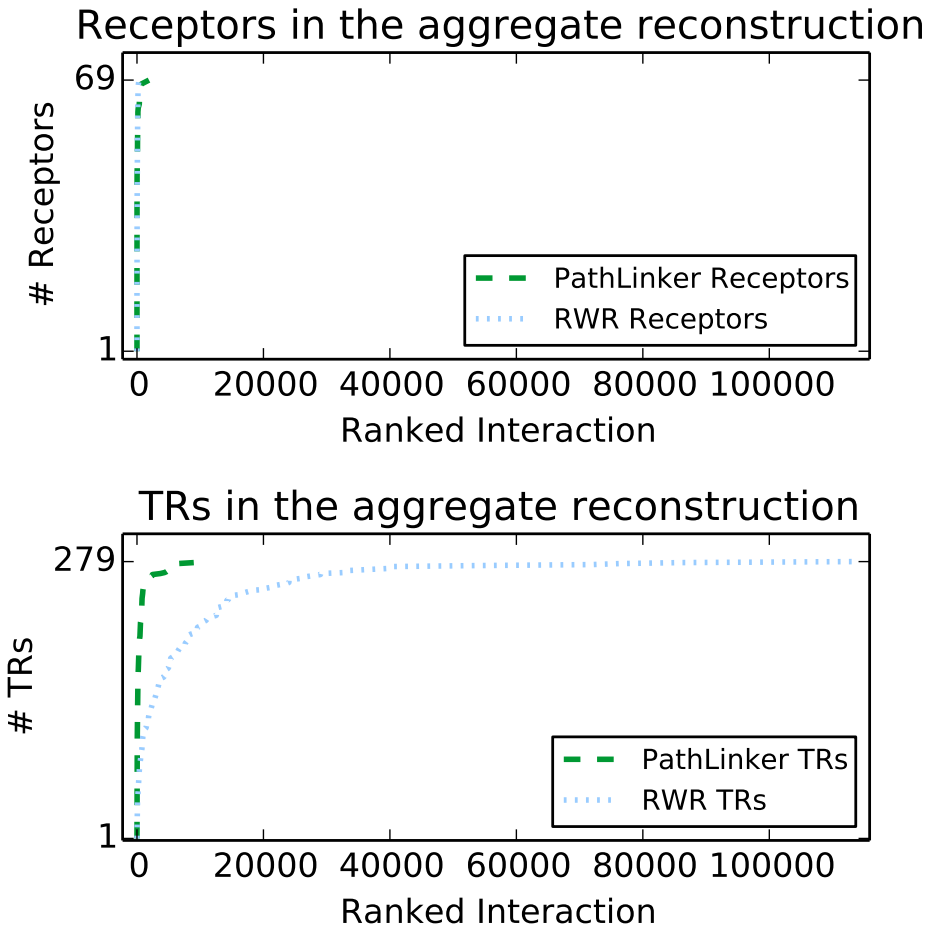
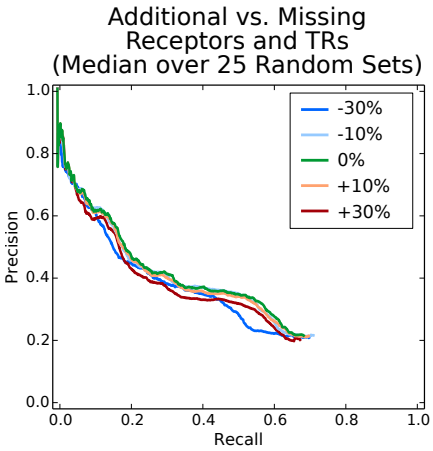


Figure S2: Rank of the first occurence of receptors (left) and TRs (right) in interactions from PathLinker and RWR reconstructions.



ρ	Recall=0.3 Mean (Std. Dev.)	Recall=0.6 Mean (Std. Dev.)
-30%	$0.38\ (1.90 \times 10^{-2})$	$0.22\ (1.32 \times 10^{-2})$
-10%	$0.41\ (1.05 \times 10^{-2})$	$0.26\ (1.83 \times 10^{-2})$
0%	0.42	0.28
+10%	$0.40\ (5.58 \times 10^{-3})$	$0.27\ (6.99 \times 10^{-3})$
+30%	$0.37\ (1.47 \times 10^{-2})$	$0.24\ (1.22 \times 10^{-2})$

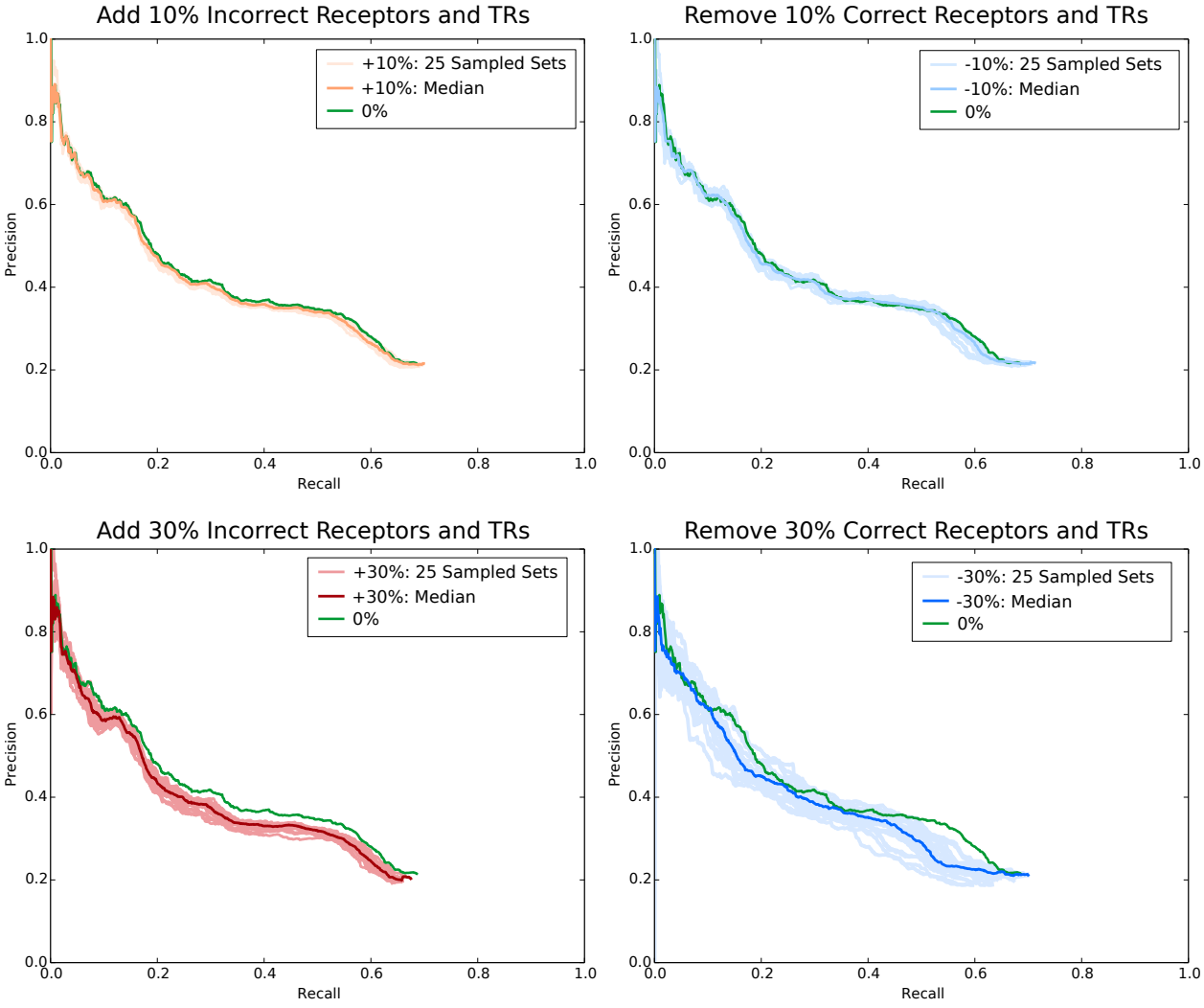


Figure S3: Precision and recall of 25 PathLinker reconstructions sampled for every value of ρ in $[-30\%, -10\%, 0\%, 10\%, 30\%]$.

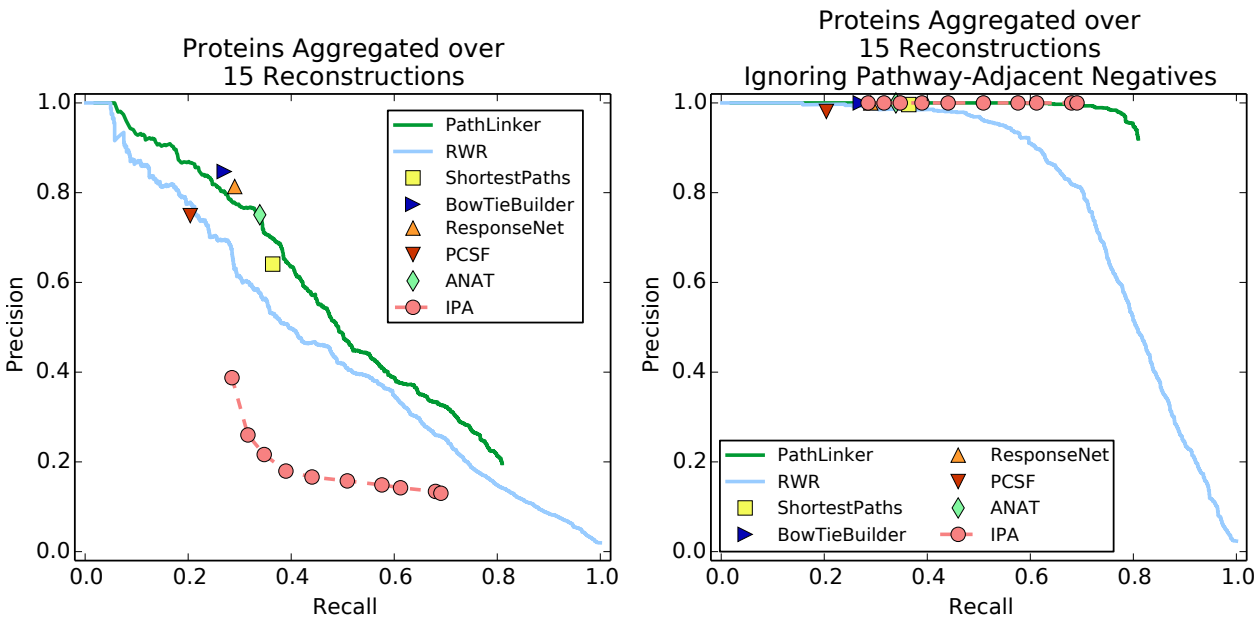


Figure S4: Precision and recall of the proteins in the 15 NetPath pathway reconstructions over all algorithms. The on left subsampled negatives to be 50 times the number of positives. The plot on right removed pathway-adjacent negatives (proteins that shared an edge with a protein in the pathway) before subsampling.

Table S5: Proportion of positive proteins and interactions in the interactomes. A protein or an interaction is a positive if it appears in any of the 15 NetPath pathways.

	Proteins			Interactions		
Interactome	# Pos	Total	Fraction	# Pos	Total	Fraction
Original interactome	1,254	12,063	0.104	3,795	95,617	0.040
Interactome without Netpath-only interactions	1,235	12,044	0.102	2,493	94,315	0.026

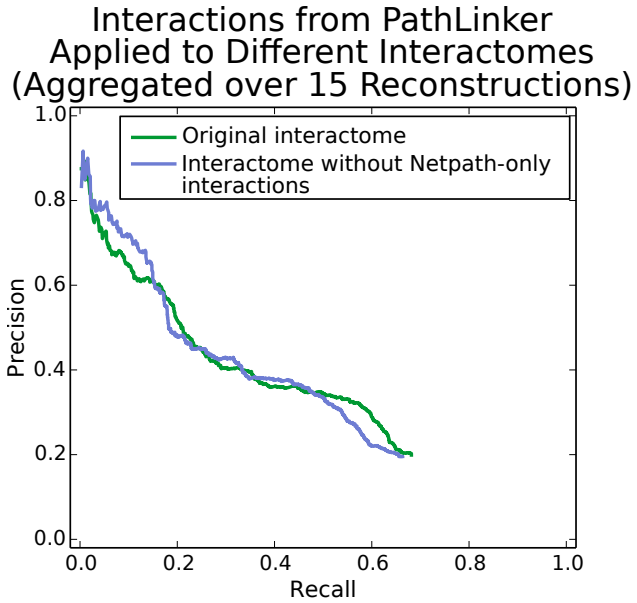


Figure S5: Performance of PathLinker on the original interactome compared to the interactome without NetPath-only interactions.

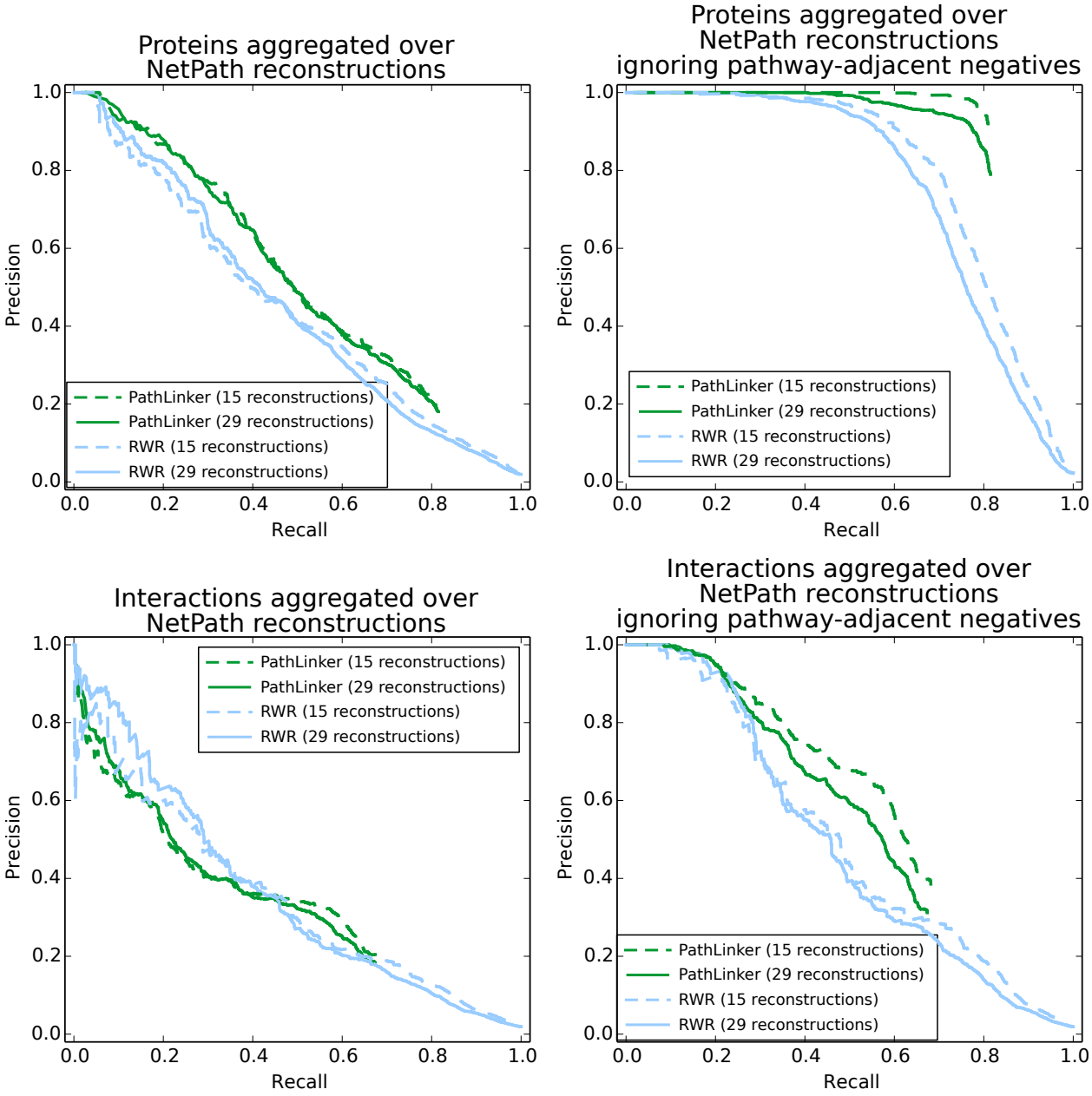


Figure S6: Precision and recall of the proteins and interactions in the 29 NetPath pathways that contained at least one receptor and at least one TR (but may not have three paths from receptors to TRs). Dashed lines denote the precision and recall aggregated over 15 NetPath pathways.

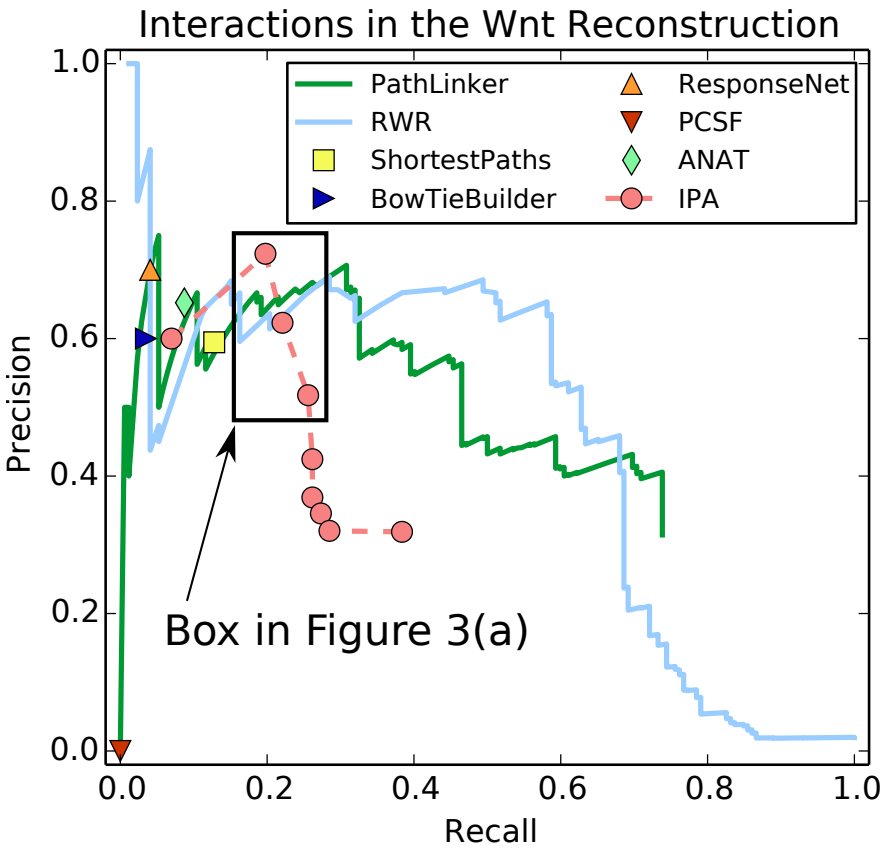


Figure S7: Precision and recall of the interactions in the Wnt pathway reconstructions computed by PathLinker and other algorithms.

Table S6: Precision and Recall of Networks from Figure 3(a).

					Ignoring Interactions in KEGG	
Method	Threshold	#Edges	Precision	Recall	Precision	Recall
PathLinker	top 247 paths	193	0.648	0.203	0.745	0.203
RWR	flux $\geq 8.26 \times 10^{-4}$	208	0.614	0.203	0.680	0.203
IPA	$n_{max} = 10$	213	0.723	0.198	0.739	0.198

5 PathLinker's Reconstruction of the Wnt Signaling Pathway

Here, we discuss PathLinker's reconstruction of the Wnt signaling pathway (Figure 3(b)).

Differences between NetPath and KEGG. In the canonical branch of Wnt signaling, β-catenin activity is controlled by the destruction complex. The PathLinker network included the core constituents of the β-catenin destruction complex (AXIN1, APC and GSK3β), as well as the accessory proteins Dishevelled 1, 2, and 3 (DVL1, DVL2, DVL3) [22]. While proteins in the Fzd and Dvl families are present in NetPath, the interactions among them are captured better in the KEGG database.

The KEGG database documents the Ca^{2+} branch of Wnt signaling, which occurs in a β-catenin-independent fashion [23]. Even though the NetPath database does not include this branch, PathLinker's reconstruction (Figure 3(b)) included paths from Frizzled receptors to phospholipase C proteins (PLCB1, PLCB2, PLCB3, PLCB4) and protein kinase C (PRKCA). In the presence of Wnt, Frizzled receptors activate phospholipase C proteins, resulting in increased intracellular concentrations of Ca^{2+}, the production of diacylglycerol, and the subsequent activation of protein kinase C [24]. However, the reconstruction did not include the Ca^{2+}-sensitive protein phosphatase calcineurin PPP3CC, CHP1, CHP2) family of proteins or their activation of the NFAT family of transcriptional regulators [24].

Proteins not in NetPath or KEGG. The PathLinker network included 16 proteins not previously known to be in the NetPath or KEGG representations of the Wnt pathway (Figure 3(b)). Ten of these proteins (MAPK1, MAPK3, EGFR, NOTCH1, SMAD2, SMAD3, SMAD7, PIK3CA, PIK3R1, and SRC) have been shown to crosstalk with the Wnt signaling pathway. Through a feedback loop, MAPK1 and MAPK3 phosphorylate GSK3β and activates the Wnt signaling pathway, thereby stabilizing β-catenin and activating Raf-1, which in turn activates MAPK1 and MAPK3 [25]. WNT1 and WNT5 have been shown to transactivate EGFR in mammary epithelial cells [26]. Through its interaction with the NOTCH1 intracellular domain, Dishevelled links the Wnt and Notch signaling pathways [27]. The SMAD proteins (SMAD2, SMAD3, and SMAD7), β-catenin, and LEF form a transcriptional complex in the nucleus [28]. Finally, PIK3CA and PIK3R1 are members of the PI3K/Akt signaling pathways. Though this pathway and the Wnt pathway share a key protein (GSK3β), the extent of crosstalk between the two pathways has been disputed [29, 30]. The SRC kinase catalyzes several signal transduction pathways, and is known to phosphorylate β-catenin [31].

Two G-protein coupled receptors in the PathLinker reconstruction (GNAQ and GNAO1) have been shown to be involved in β-catenin signaling in Drosophila and murine models, respectively [32, 33]. Two other proteins identified by PathLinker, UBA52 and RPS27A, both encode for ubiquitin. The reconstruction may have included them because ubiquitination is a common post-translational protein modification. A third protein, FLNA, is a cytoskeletal scaffold for other membrane-bound proteins [34]. It is unknown if FLNA specifically scaffolds Wnt/β-catenin signaling proteins.

CFTR was the highest ranked of all proteins not previously known to be in Wnt pathway

in the NetPath or KEGG databases. PathLinker indicated that CFTR acted as a signal transducer from Ryk, a receptor tyrosine kinase involved in Wnt signaling and organismal development [35–38], to Dab2, a known negative regulator of β-catenin signaling [39, 40]. As Wnt signaling is associated with several types of cellular differentiation and specification, the closing of membrane channels to facilitate morphological changes is biologically relevant [41].

6 A*-augmented Yen's Algorithm

We briefly describe Yen's algorithm using Dijkstra's algorithm as a subroutine, and then explain how we achieve a considerable speedup in practice by augmenting Yen's with the A* heuristic.

Yen's Algorithm. Given a directed graph $G = (V, E)$ with n vertices, m edges and two vertices s and t in V, Yen's algorithm finds the k shortest loopless paths from s to t in $O(kn(m + n \log n))$ time using Dijkstra's algorithm as a shortest path subroutine [1].

Let the ith shortest s-t path in G be π_i and let the jth vertex in that path be $\pi_{i,j}$. Yen's algorithm operates on the principle that each new shortest path π_i can be generated from some previous shortest path $\pi_{i'}$, $i' < i$, by assuming that π_i deviates from $\pi_{i'}$ after some vertex $\pi_{i',j'}$. Yen's algorithm computes this path by executing a shortest path search from $\pi_{i',j'}$ to t on a graph G', which is constructed by removing from G all the vertices in $\{\pi_{i',1}, \pi_{i',2}, \ldots, \pi_{i',j'-1}\}$ in addition to any outgoing edges from $\pi_{i',j'}$, which are in a previously found path. This construction guarantees that the path found in G' represents a new, loopless s-t path. Over all possible deviation vertices $\pi_{i',j'}$, this process results in $O(kn)$ calls to Dijkstra's algorithm to compute shortest paths. Thus, these calls yield the stated $O(kn(m + n \log n))$ time complexity for Yen's algorithm.

Integrating with A*. The running time of PathLinker is dominated by the use of Yen's algorithm to calculate the k shortest loopless paths in a network. We improve the performance of Yen's algorithm in practice with a simple modification: rather than using Dijkstra's algorithm as the shortest path subroutine, we use the A* algorithm. Given a heuristic function $h : v \to \mathbb{R}$ for $v \in V$ that is an estimate of the shortest path distance from v to t, A* is a "best first search" algorithm that computes an optimal solution to the shortest path problem while attempting to search a much smaller subset of the graph than Dijkstra's algorithm [2].

Let $d_G(v)$ be the distance from v to t in graph G. The heuristic h is *admissible* if and only if $h(v) \leq d_G(v)$, for all $v \in V$. The tighter the lower bound, the better A* will perform. If the heuristic satisfies the additional property that $h(u) - h(v) \leq w(u, v)$, for all $u, v \in V$, where $w(u, v) > 0$ is the weight of the edge from u to v, it is said to be *monotone*. Given an admissible, monotone heuristic function, A* is guaranteed to return the shortest paths from s to all nodes in G. The A* heuristic used by PathLinker is the distance from the target in the original graph, i.e., $h(v) = d_G(v)$. Each call to the shortest path subroutine in Yen's algorithm will be on some subgraph $G' \subseteq G$. Since all edge weights are non-negative, the distance of a vertex v to t in the original graph G is a lower bound for the distance of v to t in all subgraphs G'. Since $d_G(v) \leq d_{G'}(v)$, h is admissible. Furthermore, h is monotone.

Dijkstra's algorithm keys the priority queue for exploring nodes by $c(v)$, the shortest path length to v from s considering only nodes that have been explored so far. We implement A* as a modification of Dijkstra's algorithm, where we key the priority queue by $c(v) + h(v)$, rather than just by $c(v)$.

While this optimization does not affect the asymptotic running time for Yen's algorithm, it yields considerable speed ups in practice, running 11 to 41 times faster than the traditional implementation of Yen's algorithm on the pathways (Supplementary Figure S8). This improvement facilitated the computation of the top 20,000 paths in the interactome.

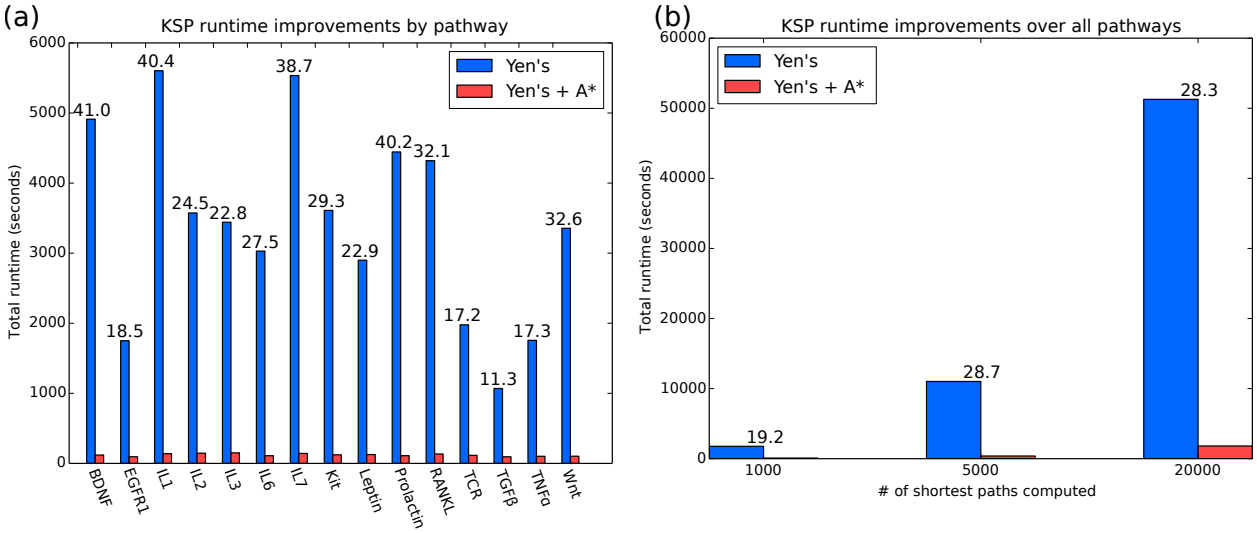


Figure S8: Comparison of the running time of A*-augmented Yen's algorithm (red bars) to a standard implementation of Yen's algorithm (blue bars). (a) The running time for each NetPath pathway ($k = 20,000$). The number above each blue bar is the speed up afforded by the improved algorithm for the corresponding pathway. (b) The total running time over all NetPath pathways for $k = 1,000$, $k = 5,000$, and $k = 20,000$.

7 Experimental Methods

Efficacy of siRNA silencing. Cells were routinely passaged and cultured as described in Clark et al. [68] in DMEM containing 10% fetal bovine serum and 1% penicillin/streptomycin at 37 °C in the presence of 5% CO_2. Invitrogen silencer select validated siRNAs (Dab2: s3896, Ryk: s12390, CFTR: s2945) were dissolved in 500 µL of provided water resulting in a final concentration of 10 mM and stored in 30 µL aliquots at -20 °C. Efficacy of siRNA silencing was determined by western blot in a dose dependent manner. Approximately 200,000 HEK293 were plated in 24 well plates and allowed to adhere for 24 h in 1 mL complete media. Cells were washed twice with room temperature (22 °C) dPBS and incubated with 900 µL of complete media. A siRNA-RNAiMax solution was prepared as described by the manufacturer. Briefly, 3 µL of RNAiMax and 0-4 µL of respective siRNA (10 mM stock concentration) were added to separate tubes of 50 µL of DMEM and allowed to incubate for 15 min. Solutions were subsequently pooled, mixed by gentle pipetting, and allowed to

incubate at room temperature for 30 min. Complexed siRNA solution (100 µL) was added to each well and incubated for 48 h. Cells were washed twice with room temp dPBS and harvested in 100 µL NP-40 buffer containing protease inhibitor and flash frozen in liquid nitrogen. 25 µL of 5x Loading dye was added to each sample, mixed and heated to 80 °C for 5 min.

Western blots. Processed samples were run via SDS-PAGE on a 7.5% polyacrylamide gel of 1.5 mM thickness. Samples were transferred for 1.5 h at a constant 300 mA onto hybond-C extra membranes. Membranes were kept submerged in 20 mL of PBS-T (Sigma P3813 + 0.1% Tween20) + 3% BSA for 1 h at room temperature. Appropriate primary antibody was spiked in (Table S7) and membranes were stored overnight at 4 °C on an orbital shaker. Membranes were washed thrice with 20 mL of PBS-T for 10 min while shaking. Membranes were probed with appropriate secondary antibody (Table S7) for 1 h at room temperature while shaking. Membranes were washed twice with 20 mL of PBS-T for 10 min shaking and stored in 20 mL PBS for no more than 10 min. Membranes were exposed to 8 mL of chemiluminescence substrate (SuperSignal^{TM} West Pico Chemiluminescent Substrate 34080) for 5 min in the dark and subsequently imaged in a Chem-Doc XRS+ workstation using Image Lab Software. Images were recorded over 10 min every 10 s.

Table S7: Antibodies used and paired for this study. Abbreviation: Horseradish Peroxidase (HRP).

Antigen	**Antibody**	**Source**	**Dilution**	**Vendor**	**Catalog #**
CFTR	Primary	Rabbit Polyclonal IgG	1:2,000	Santa Cruz	sc-10747
	Secondary	Goat Anti-Rabbit Polyclonal IgG-HRP	1:10,000	GE Healthcare	RPN4301
Dab2	Primary	Rabbit Polyclonal IgG	1:2,000	Santa Cruz	sc-13982
	Secondary	Goat Anti-Rabbit Polyclonal IgG-HRP	1:10,000	GE Healthcare	RPN4301
Ryk	Primary	Rabbit Polyclonal IgG	1:2,000	abcam	ab135670
	Secondary	Goat Anti-Rabbit Polyclonal IgG-HRP	1:10,000	GE Healthcare	RPN4301
β-catenin	Primary	Mouse Monoclonal IgG1	1:10,000	Santa Cruz	sc-7963
	Secondary	Goat Anti-Mouse Polyclonal IgG-HRP	1:10,000	R&D Systems	HAF007
GAPDH	Primary	Goat Polyclonal IgG	1:20,000	R&D Systems	AF5718
	Secondary	Rabbit Anti-Goat Polyclonal IgG HRP	1:10,000	R&D Systems	HAF017

Transient overexpression of Wnt proteins in siRNA silenced background. Cells were silenced via lipofection as described above. The Wnt plasmid library (addgene Kit # 1000000022) [20], specifically secreted Wnt proteins lacking any engineered epitopes, were utilized for the study. Approximately 24 h post RNAiMAx transfection, cells were washed twice with room temp dPBS and incubated with 900 µL of complete media. Lipofectamine LTX- plasmid solution was prepared as described by the manufacturer. Briefly, 4 µL of Lipofectamine was added to a tube containing 50 µL of DMEM. 1 µL of plus solution and 100 ng of a given secreted Wnt, pM50 Super 8x TOPFlash (7 sequencing TCF/LEF promoter binding sites fused to firefly luciferase) [21], and constitutive expression of Renilla luciferase plasmid (pGL4.74[hRluc/TK], promega E6921) was added to a separate tube containing 50 µL of DMEM. Tubes were allowed to incubate for 15 min and the solutions were subsequently pooled, mixed by gentle pipetting, and allowed to incubate at room temp for 30 min.

The LTX -plasmid solution (100 μL) was added to each well and incubated for 30-36 h prior to the luciferase reporter assay or determination of β-catenin levels via western blot.

Luciferase reporter assay. The dual glow luciferase reporter assay was conducted as described by the manufacturer (Promega #E2940). Briefly, treated cells were washed once with room temp dPBS, and incubated in 100 μL of dual glow buffer for 5 min at 37 °C. Luminescence was determined via integration of 1000 ms top reading using a SpectraMax M5. Subsequently, 100 μL of freshly prepared Stop and Glow buffer was added to each well and incubated for 5 min at 37 °C. Renillia control luminescence was determined via integration of 1000 ms top reading using a SpectraMax M5. Normalized luminescence was determined by dividing the dual glow luminescence (firefly luciferase activity) by the Stop and glow luminescence (Renilla luciferase activity).

Co-immunoprecipitation. HEK293 cells (10^6 cells) were transfected with sWnt and control plasmids and incubated as previously described for approximately 48 h. Cells were gently washed 2x with room temp dPBS and re-suspended in 1 mL of Extraction Buffer in the presence of protease inhibitors (Roche #11697498001) and incubated on ice for 15 min. Antibody coupling to Dynabeads (M-270 Epoxy) resin and co-immunoprecipitation via magnetic separation was followed as described by the manufacturer (Invitrogen #14321D). 1.5 mg of antibody (150 μL) was transferred to a fresh tube and washed with 900 μL of Extraction Buffer using magnetic separation. Cell lysate was added to washed beads and incubated for 45 min at 4 μC on a vertical rotator. Magnetic beads were washed three times with 200 μL of Extraction Buffer. Beads were incubated with 200 μL of Last Wash Buffer for 5 min at room temp on a vertical rotator. Beads were transferred to a clean tube and re-suspended in 60 μL of Elution Buffer using magnetic separation. Samples (10 μL) were run on a 7.5% SDS-PAGE gel and probed with appropriate antibody pairs (Supplementary Table S7).

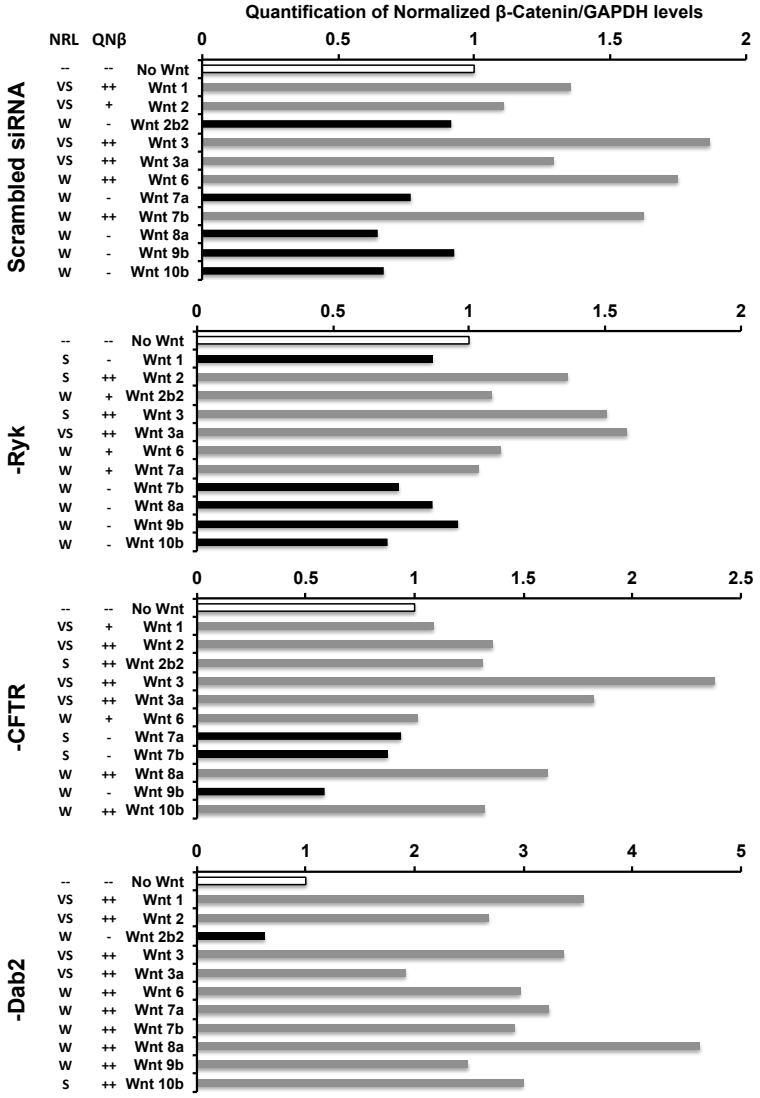


Figure S9: Quantification of western blot band Intensity. We applied the Bio-Rad Image Lab software on SCN files obtained from the Bio-rad ChemiDoc-XRS+ system. We normalized Band Intensity by GAPDH intensity and then against control No Wnt samples. We performed qualifications and comparisons only for samples on a given scan. Black or gray colors for individual bar graphs signify a normalized intensity less than or greater than the No Wnt control, respectively. QNβ: Qualification of Normalized β-catenin intensity, "++": $\geq$1.3-fold, "+": 1.3-fold>x$\geq$1-fold, "-": <1-fold. We compared QNβ values to qualifications of the normalized relative luminescence (NRL), "VS": very strong ($\geq$30-fold),"S": strong (30-fold>X$\geq$15-fold), "W": weak (<15-fold).

References

[1] Yen JY. Finding the K Shortest Loopless Paths in a Network. Management Science. 1971;17(11):712–716.

[2] Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics. 1968;4(2):100–107.

[3] Haveliwala TH. Topic-Sensitive PageRank. In: Proceedings of the 11th International Conference on World Wide Web; 2002. p. 517–526.

[4] Yosef N, Zalckvar E, Rubinstein AD, Homilius M, Atias N, Vardi L, et al. ANAT: A Tool for Constructing and Analyzing Functional Protein Networks. Science Signaling. 2011;4(196):pl1+.

[5] Tuncbag N, Braunstein A, Pagnani A, Huang SS, Chayes J, Borgs C, et al. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting Steiner forest problem. Journal of Computational Biology. 2013 Feb;20(2):124–136.

[6] Yeger-Lotem E, Riva L, Su LJJ, Gitler AD, Cashikar AG, King OD, et al. Bridging high-throughput genetic and transcriptional data reveals cellular responses to alpha-synuclein toxicity. Nature Genetics. 2009 March;41(3):316–323.

[7] Ingenuity Pathway Analysis (IPA). IPA Network Generation Algorithm; 2005. http://www.ingenuity.com/wp-content/themes/ingenuity-qiagen/pdf/ipa/IPA-netgen-algorithm-whitepaper.pdf.

[8] Supper J, Spangenberg L, Planatscher H, Drager A, Schroder A, Zell A. BowTieBuilder: modeling signal transduction pathways. BMC Syst Biol. 2009;3:67.

[9] Aranda B, Blankenburg H, Kerrien S, Brinkman FS, Ceol A, Chautard E, et al. PSICQUIC and PSISCORE: accessing and scoring molecular interactions. Nat Methods. 2011 Jul;8(7):528–529.

[10] Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, et al. NetPath: a public resource of curated signal transduction pathways. Genome Biology. 2010;11(1):R3.

[11] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research. 2012 Jan;40(Database issue):D109–114.

[12] Paz A, Brownstein Z, Ber Y, Bialik S, David E, Sagir D, et al. SPIKE: a database of highly curated human signaling pathways. Nucleic Acids Research. 2011 Jan;39(suppl 1):D793–D799.

[13] Almen M, Nordstrom K, Fredriksson R, Schioth H. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biology. 2009;7(1):50+.

[14] Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. Cell. 2010 Mar;140(5):744–752.

[15] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009 Apr;10(4):252–263.

[16] Tastan O, Qi Y, Carbonell JG, Klein-Seetharaman J. Prediction of interactions between HIV-1 and human proteins by information integration. Pac Symp Biocomput. 2009;p. 516–527.

[17] Qi Y, Bar-Joseph Z, Klein-Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins. 2006 May;63(3):490–500.

[18] Sun MG, Sikora M, Costanzo M, Boone C, Kim PM. Network evolution: rewiring and signatures of conservation in signaling. PLoS Comput Biol. 2012;8(3):e1002411.

[19] Park Y, Marcotte EM. Revisiting the negative example sampling problem for predicting protein-protein interactions. Bioinformatics. 2011 Nov;27(21):3024–3028.

[20] Najdi R, Proffitt K, Sprowl S, Kaur S, Yu J, Covey TM, et al. A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. Differentiation. 2012 Sep;84(2):203–213.

[21] Veeman MT, Slusarski DC, Kaykas A, Louie SH, Moon RT. Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. Curr Biol. 2003 Apr;13(8):680–685.

[22] Clevers H. Wnt/β-catenin signaling in development and disease. Cell. 2006;127(3):469–480.

[23] Niehrs C. The complex world of WNT receptor signalling. Nat Rev Mol Cell Biol. 2012 Dec;13(12):767–779.

[24] Wang Hy, Malbon CC. Wnt signaling, Ca2+, and cyclic GMP: visualizing Frizzled functions. Science. 2003;300(5625):1529–1530.

[25] Jin C, Samuelson L, Cui CB, Sun Y, Gerber DA. MAPK/ERK and Wnt/β-Catenin pathways are synergistically involved in proliferation of Sca-1 positive hepatic progenitor cells. Biochem Biophys Res Commun. 2011 Jun;409(4):803–807.

[26] Civenni G, Holbro T, Hynes NE. Wnt1 and Wnt5a induce cyclin D1 expression through ErbB1 transactivation in HC11 mammary epithelial cells. EMBO Rep. 2003 Feb;4(2):166–171.

[27] Collu GM, Hidalgo-Sastre A, Acar A, Bayston L, Gildea C, Leverentz MK, et al. Dishevelled limits Notch signalling through inhibition of CSL. Development. 2012 Dec;139(23):4405–4415.

[28] Guo X, Wang XF. Signaling cross-talk between TGF-beta/BMP and other pathways. Cell Res. 2009 Jan;19(1):71–88.

[29] Voskas D, Ling LS, Woodgett JR. Does GSK-3 provide a shortcut for PI3K activation of Wnt signalling? F1000 Biol Rep. 2010;2:82.

[30] Ng SS, Mahmoudi T, Danenberg E, Bejaoui I, de Lau W, Korswagen HC, et al. Phosphatidylinositol 3-kinase signaling does not activate the wnt cascade. J Biol Chem. 2009 Dec;284(51):35308–35313.

[31] Piedra J, Martinez D, Castano J, Miravet S, Dunach M, de Herreros AG. Regulation of beta-catenin structure and activity by tyrosine phosphorylation. J Biol Chem. 2001 Jun;276(23):20436–20443.

[32] Liu X, Rubin JS, Kimmel AR. Rapid, Wnt-induced changes in GSK3beta associations that regulate beta-catenin stabilization are mediated by Galpha proteins. Curr Biol. 2005 Nov;15(22):1989–1997.

[33] Bikkavilli RK, Feigin ME, Malbon CC. G alpha o mediates WNT-JNK signaling through dishevelled 1 and 3, RhoA family members, and MEKK 1 and 4 in mammalian cells. J Cell Sci. 2008 Jan;121(Pt 2):234–245.

[34] Robertson SP, Twigg SR, Sutherland-Smith AJ, Biancalana V, Gorlin RJ, Horn D, et al. Localized mutations in the gene encoding the cytoskeletal protein filamin A cause diverse malformations in humans. Nat Genet. 2003 Apr;33(4):487–491.

[35] Lu W, Yamamoto V, Ortega B, Baltimore D. Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth. Cell. 2004 Oct;119(1):97–108.

[36] Yoshikawa S, McKinnon RD, Kokel M, Thomas JB. Wnt-mediated axon guidance via the Drosophila Derailed receptor. Nature. 2003 Apr;422(6932):583–588.

[37] Keeble TR, Halford MM, Seaman C, Kee N, Macheda M, Anderson RB, et al. The Wnt receptor Ryk is required for Wnt5a-mediated axon guidance on the contralateral side of the corpus callosum. J Neurosci. 2006 May;26(21):5840–5848.

[38] Bovolenta P, Rodriguez J, Esteve P. Frizzled/RYK mediated signalling in axon guidance. Development. 2006 Nov;133(22):4399–4408.

[39] Jiang Y, He X, Howe PH. Disabled-2 (Dab2) inhibits Wnt/β-catenin signalling by binding LRP6 and promoting its internalization through clathrin. EMBO J. 2012 May;31(10):2336–2349.

[40] Jiang Y, Luo W, Howe PH. Dab2 stabilizes Axin and attenuates Wnt/beta-catenin signaling by preventing protein phosphatase 1 (PP1)-Axin interactions. Oncogene. 2009 Aug;28(33):2999–3007.

[41] Chen Y, Rice W, Gu Z, Li J, Huang J, Brenner MB, et al. Aquaporin 2 promotes cell migration and epithelial morphogenesis. J Am Soc Nephrol. 2012 Sep;23(9):1506–1517.