Network-based approaches for extending the Wnt signalling pathway and identifying context-specific sub-networks

Sudipto Saha

Center for Proteomics and Bioinformatics,
Case Western Reserve University,
Cleveland, Ohio 44106, USA
Email: sudipto.saha@case.edu

Theodore Roman

Lane Center for Computational Biology,
Carnegie Mellon University,
Pittsburgh, Pennsylvania 15213, USA
Email: troman@andrew.cmu.edu

Alex Galante

School of Medicine,
Case Western Reserve University,
Cleveland, Ohio 44106, USA
Email: alex.galante@case.edu

Mehmet Koyutürk

Department of Electrical Engineering and Computer Science,
Case Western Reserve University,
Cleveland, Ohio 44106, USA
and
Center for Proteomics and Bioinformatics,
Case Western Reserve University,
Cleveland, Ohio 44106, USA
Email: mxk331@case.edu

Rob M. Ewing*

Center for Proteomics and Bioinformatics,
Case Western Reserve University,
Cleveland, Ohio 44106, USA
Email: rob.ewing@case.edu
*Corresponding author

Abstract: Wnt signalling is a critically important signalling pathway regulating embryogenesis and differentiation, and is broadly conserved amongst multicellular animals. In addition, dysregulation of Wnt signalling contributes to the pathogenesis of many human cancers, in particular colorectal cancer. Core members of the Wnt signalling pathway are quite well defined, although it has become apparent that a much broader network of interacting proteins regulates Wnt signalling activity. The goal of this paper is first to identify novel members of the Wnt regulatory network; and second, to identify sub-networks of the larger Wnt signalling network that are active in different biological contexts. We address these two questions using complementary computational approaches and show how these approaches may identify potentially novel Wnt signalling proteins as well as defining Wnt sub-networks active in different stages of colorectal cancer.

Keywords: Wnt signalling network; eigenvector centrality; random walk; sub-network; simulated annealing.

Reference to this paper should be made as follows: Saha, S., Roman, T., Galante, A., Koyutürk, M. and Ewing, R.M. (2012) 'Network-based approaches for extending the Wnt signalling pathway and identifying context-specific sub-networks', *Int. J. Computational Biology and Drug Design*, Vol. 5, Nos. 3/4, pp.185–205.

Biographical notes: Sudipto Saha is currently a Postdoctoral Scholar in the Center for Proteomics and Bioinformatics at Case Western Reserve University. He received his PhD in Bioinformatics from Jawaharlal Nehru University, New Delhi, India in 2007. His research interests include studying protein–protein interaction networks and systems biology approaches.

Theodore Roman received his BS in Computer Science and BA in Mathematics from Case Western Reserve University, Cleveland, Ohio, in 2011. Currently he is a Doctoral Student at the Joint Carnegie Mellon University–University of Pittsburgh program in Computational Biology, sponsored by the Lane Center for Computational Biology at Carnegie Mellon University in Pittsburgh, Pennsylvania.

Alex Galante received his BA in Biology from Case Western Reserve University, Cleveland, Ohio in 2011. Currently he is a Medical Student at Case Western Reserve University, School of Medicine.

Mehmet Koyutürk is currently T&D Schroeder Assistant Professor of Computer Science and Engineering at Case Western Reserve University in Cleveland, Ohio. He received his PhD in Computer Science from Purdue University, West Lafayette, Indiana in 2006. His research focuses on development of computational algorithms development for data analysis in systems biology.

Rob Ewing is currently an Assistant Professor in the Center for Proteomics and Bioinformatics at Case Western Reserve University in Cleveland, Ohio. He received his PhD in Molecular Genetics from University of Oxford, UK. His research interest is in mapping protein interaction networks using computational and experimental approaches, and understanding the effects of perturbing protein interaction networks through mutation and signalling.

This paper is a revised and expanded version of a paper entitled 'Systematic discovery of condition specific Wnt signalling sub-networks' presented at '2nd Workshop on Integrative Data Analysis in Systems Biology (IDASB), in Conjunction with 2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)', 12–15 November 2011, Atlanta, Georgia, USA.

1 Introduction

The Wnt signalling pathway is conserved throughout metazoans and involved in fundamental biological processes such as embryogenesis and in human developmental disorders and cancer. Several decades of Wnt research have led to the identification of many components of the Wnt signalling pathway, and enabled the definition of canonical (Wnt/β-catenin) and non-canonical Wnt pathways (Wnt/Calcium and the Wnt/Jun N-terminal Kinase (JNK) pathways) (KEGG Pathway: hsa04310; see http://www.genome.jp/dbget-bin/www_bget?pathway+hsa04310). Different biological roles have been associated with these different branches of Wnt signalling. For example, canonical Wnt signalling determines cell fate, whereas the non-canonical, β-catenin independent pathways are involved in the regulation of cell polarity. Nineteen Wnt family ligands have been identified in mammals and all are secreted proteins with glycolipid modifications. Some of these ligands activate specific Wnt receptors and function in specific branches of Wnt signalling, whereas for others it is less clear (Katoh, 2002). In addition, canonical and non-canonical Wnt pathways have different transcriptional targets. Well-studied canonical targets include MYC, CCND1, FGF20, WISP1, JAG1, DKK1 and GCG; while non-canonical signalling activates JNK, ROCK, PKC, MAP3K7, NFAT and associated signalling cascades (Katoh and Katoh, 2007).

It is currently challenging to dissect this important pathway for the following reasons. First, many components of Wnt signalling pathways are multi-functional proteins (for example β-catenin), that are essential not only for Wnt signalling but also for other intercellular signalling networks (Chilov et al., 2010). Second, many pathways, for example Hedgehog, FGF, Notch, BMP, ERK and P13K, overlap or crosstalk with the Wnt signalling pathway in embryogenesis and carcinogenesis (Moreno, 2010). In addition, the canonical and non-canonical branches of the Wnt signalling pathways are themselves highly interconnected, and cross-regulate each other. Third, the Wnt signalling cascade is dependent on biological state, for example the presence or absence of specific Wnt receptors. Dependencies between ligands and receptors determine whether specific branches of Wnt signalling will be activated. For example, WNT5A can activate Wnt/calcium and the Wnt/JNK in cancer as well as Wnt/β-catenin pathway in the presence of FZ4 and LRP-5 receptors (McDonald and Silver, 2009). Fifth, different concentrations of Wnt ligands can elicit different intracellular responses. For example, low concentrations of WNT3A trigger Wnt/calcium signalling, while high concentrations of WNT3A activate Wnt/β-catenin signalling (Nalesso et al., 2011). In addition, since Wnt signalling is integral to different biological processes and pathologies, it is important to understand how the pathway is modulated under different biological conditions. In summary, rather than thinking of Wnt as a fixed linear pathway, it is a network of interconnected molecular events that is modulated under different biological conditions. Our goal in this paper is to identify novel Wnt pathway regulators and finally develop a platform for defining how the Wnt signalling network changes according to biological state.

Several studies have successfully used computer models of biological networks to identify key players in these processes (Colland et al., 2004; Niida et al., 2004; Major et al., 2007; Major et al., 2008). The paradigm of these types of studies is to use the *in silico* models to predict important in vivo proteins/processes, and then test those

predictions using various functional assays. Xenopus larval axis duplication is known to be induced by Wnt signalling, and is now a standard in detecting Wnt functional activity (Moon and Kimelman, 1998; Tamai et al., 2004). TOP-flash reporters have also been successfully used to confirm the presence of Wnt signalling by activating a constructed luciferase reporter gene at TCF (the major Wnt transcription factor) binding sites along the genome (DasGupta et al., 2005).

Different computational approaches have been applied to understanding the Wnt signalling network. On the one hand, mathematical models have been developed to understand the dynamics of the pathway, focusing on the core canonical β-catenin destruction complex (Lee et al., 2003; Wawra et al., 2007). On the other hand, broader views of Wnt signalling network have been undertaken to identify new regulators of Wnt signalling. These approaches have used integrative data analysis of genetic screens, Protein–Protein Interactions (PPI) and post-translational modifications of Wnt signalling components. Major et al (2008) mapped physical and functional Wnt pathway components by merging the small interfering RNA (siRNA) screen with the Wnt/β-catenin protein interaction network. Kestler and Kühl (2008) reported a graphical representation of Wnt network and evolved the concept that Wnt proteins activate a signalling network instead of an individual pathway.

Here, we first construct a network of PPI with context to Wnt signalling pathway, and attempt to identify novel members of the Wnt signalling pathway by scoring the relevance of proteins with respect to their connectivity to the known Wnt signalling members using local and global network approaches. We hypothesise that the highest scoring proteins will be qualitatively more relevant to Wnt signalling, and our aim is identify novel Wnt signalling proteins whose functional significance can be confirmed using the standard functional assays, such as Xenopus larvae axis duplication or TOP-flash reporters (Moon and Kimelman, 1998; Tamai et al., 2004; DasGupta et al., 2005). Second, to better understand how the Wnt signalling network is activated under different biological states, we identify activated Wnt sub-networks under different conditions. By analysing an interaction network seeded from the Wnt pathway, we are able to discovery signalling circuits without being limited to a specific pathway boundary or set of genes with similar expression profiles. This approach allows us to address important questions of similarities or conservation of active sub-networks patterns across Wnt perturbed healthy versus disease cell lines and tissues. We used a simulated annealing approach in overlaying significant changes in gene expression on a Wnt-focused protein interaction network. Similar approaches have been applied to studying signalling circuits in yeast (Ideker et al., 2002) to classification of breast cancer samples (Chuang et al., 2007) and previously by us to discriminate different stages of colon cancer (Nibbe et al., 2009).

The remainder of this paper is organised as follows. Section 2 focuses on the data sets, parameters and methods used in our study. We then used local and global network analyses to identify novel interactors to Wnt pathway using the hypergeometric test and eigenvector centrality, respectively. Next, we integrate gene-expression data sets with a Wnt-focused interaction network and apply a simulated annealing approach to identify Wnt sub-networks associated with different biological conditions. We identify similarities between the sub-networks using the Jaccard index and explore the underlying biology associated with the commonalities between tissues, cells and disease states. In Section 4, we summarise our key findings, applications to biomedicine, highlighting the pros and cons of our integrated approach.

2 Materials and methods

2.1 Data sets

Two principal data sets were used in this study. First, we compiled two PPI networks, one focused on the known Wnt-signalling pathways (Wnt-specific network) and other including a comprehensive set of interactions among human proteins (global network). Although sub-network analysis techniques are applicable to large-scale PPI networks, we used a Wnt-focused PPI network in addition to the comprehensive PPI network in our analyses because we are primarily interested in sub-networks related to Wnt signalling.

The Wnt-specific network was constructed as follows. A core set of Wnt pathway proteins was extracted from the Wnt proteins listed in four pathway databases: KEGG (Kanehisa et al., 2012), BioCarta (see http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways), Protein Lounge (http://www.proteinlounge.com/Default.aspx) and Cancer cell map (see http://cancer.cellmap.org/cellmap/about.do) (Table 1). Proteins which were listed as Wnt signalling members in at least three of these databases were included in our 'Core' set, and represent the most well-annotated members of the Wnt pathway. Our network was expanded from this core to include all proteins experimentally shown to interact with the core. Proteins which interacted with the core and that were in at least one of all four databases were labelled as 'Union' proteins. All other proteins that were experimentally shown to interact with the core, but were not annotated members of the pathway were labelled as 'Others'. The distribution of proteins in the network was as follows: 31 core proteins, 74 union proteins and 511 other proteins. The 31 core proteins were superimposed on version 9 of the Human Protein Reference Database (HPRD) (Keshava Prasad et al., 2009) and a network was extracted by including all proteins no more than one hop from the core Wnt proteins, resulting in a network of 363 proteins and 2072 PPIs. The global PPI network was constructed by including all human PPI from the HPRD for the global analysis network (8959 proteins and 33,528 pairwise PPIs among these proteins).

Table 1 List of Wnt core proteins

Gene name	*Gene name*	*Gene name*	*Gene name*	*Gene name*
APC	DKK1	FZD4	GSK3B	RAC1
AXIN1	DKK2	FZD5	LEF1	WNT1
BTRC	DVL1	FZD6	LRP5	WNT2
CCND1	FRAT1	FZD7	LRP6	WNT3
CTNNB1	FZD1	FZD8	MAP3K7	WNT4
DAAM1	FZD2	FZD9	NLK	WNT5A

Second, a data set of Wnt-perturbed human gene-expression measurements was compiled. Seven human gene-expression data sets were selected from the Gene Expression Omnibus (GEO), based upon their experimental annotations (activation of the Wnt signalling pathway). The seven studies were selected as representative of the diverse range of cells and tissue types with activated Wnt signalling. Although this is a relatively small subset of the possible Wnt-related gene-expression studies in GEO, we used this

subset to define our analysis process and as a proof-of-principle. Table 2 lists the GEO studies with brief annotations. GEO data sets were processed, analysed and annotated using custom R code and the GEOquery module (Sean and Meltzer, 2007). In general, since the GEO studies represent studies using different microarray platforms, we relied on submitter-provided intensity and/or ratio values. Log_2 ratios of intensities and significance values (Student's *t*-test) were computed for all microarray features. A core set of genes represented on all microarray platforms represented in the seven GEO studies was identified, providing a data-matrix of 8044 genes by seven studies. In addition, the smaller network size allows us to use the computationally intensive simulated annealing approach to identify Wnt sub-networks. Finally, to further annotate the proteins represented in the network, we used the cancer-related gene list from the Cancer Gene Census (see http://www.sanger.ac.uk/genetics/CGP/Census/) (457 cancer genes and 19 colorectal cancer genes as of September 2011).

Table 2 GEO series (GSE) datasets used in this study

GEO accession	*Tissue or sample type*	*Number samples*	*Study abbreviation*	*Reference*
GSE10972	Tumour vs. normal colon tissue	48	CTT	Jiang et al. (2008)
GSE14107	Wnt3A treated vs. control PC9 (lung adenocarcinoma) cells	16	LCA	Nguyen et al. (2009)
GSE1473	CTNNB1 mutant vs. normal HEK293T (embryonic kidney) cells	8	EKC	Chamorro et al. (2005)
GSE16186	Wnt3A treated vs. control mesenchymal stem cells	24	MSC	Qiu et al. (2010)
GSE17385	CTNNB1 knock-out vs. normal multiple myeloma cells	6	MMC	Dutta-Simmons et al. (2009)
GSE6120	CTNNB1 mutant vs. normal Wilms tumour tissue	39	WTT	Li et al. (2004)
GSE8671	Colorectal adenoma vs. normal tissue	64	CCA	Sabates-Bellver et al. (2007)

2.2 Local network analysis

The 'local' analysis considers the direct interactions with Wnt core members. In order to test our hypothesis that proteins that exhibit significant connectivity to the core are likely to be involved in Wnt signalling, we assessed the statistical significance of the number of interactions between each protein in the network and the proteins in the core set. For this analysis, we considered a network constructed from human, mouse, rat and Drosophila proteins and their orthologs. We used each protein's degree (the total number of interactions), each protein's degree with respect to the Wnt core (number of interactions with core proteins), the total number of interactions of all core proteins in the

network, and the total number of binary interactions in the network. Using these statistics, the hypergeometric test was used to assess the significance of the degree of each protein with respect to the Wnt core as follows:

$$P(X \geq i) = \sum_{k=i}^{n} \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \tag{1}$$

Here 'X' is a random variable representing the number of connections from a protein to the core, 'i' is the observed value of X, 'n' is the degree of the same protein, 'm' is the total number of interactions of the core in the entire network and 'N' is the total number of interactions in the network. Namely, $P(X \geq i)$ is the probability that the protein would have at least i interactions with the core if all interactions in the network were drawn at random while preserving the connectivity of the core to the rest of the network., i.e. the smaller the p-value is, the more significant is the protein's connectivity to Wnt core. This test allowed us to rank proteins in the network based on the significance of their connectivity to the core Wnt proteins. Note that correction of multiple hypothesis testing was not necessary here since the p-values were only used to compare the proteins against each other.

2.3 Global network analysis: eigenvector centrality/random walk

In addition to the direct connections to the 'core' proteins, we expected that proteins that are involved in Wnt signalling would also have increased indirect connections to the core in a global analysis. To test this hypothesis, we quantified the level of (direct or indirect) connectivity to core proteins by generalising the concept of eigenvector centrality. Local analysis takes into account only the direct interactions with the core to assess connectivity, while eigenvector centrality based scoring considers the indirect interactions as well. In a more general setting, eigenvector centrality is a measure of each node's influence in a network. It has been used in a diverse range of applications to assess the centrality of each node in a network, including Google's page rank algorithm for ranking the importance of web pages (Brin and Page, 1998) and assessment of the relationship between essentiality and network topology in biological networks (Zotenko et al., 2008).

While eigenvector centrality is a measure of the general influence of a node in the network, the objective here is to assess the influence of each protein in the network on the Wnt core proteins. For this reason, we use a generalised version of eigenvector centrality, by using a 'random walk with restarts' model. This model simulates a random walk across the PPI network, where the walk makes frequent restarts at the Wnt core proteins and the score of each protein is computed as the probability that the random walk will be at that protein at infinity. In this model, the frequent restarts introduce a bias to the scores of proteins that interact with the core and this bias is subsequently propagated across the network by the random walk. Consequently, the resulting score represents the connectivity of each protein to the Wnt core for a global network perspective. This method has been shown to be successful in predicting novel functions for proteins (Nabieva et al., 2005) and prioritising candidate disease genes (Chen et al., 2009.

Let $G := (V, E)$ denote the global PPI network where V denotes the set of proteins and E denotes the set of interactions. Let A denote the corresponding adjacency matrix, i.e. $A = (a_{i,j})$ where $a_{i,j} = 1$ if protein i interacts with protein j, and $a_{i,j} = 0$ otherwise. We first normalise this matrix by dividing each column of A to the 1-norm of that column and obtain stochastic matrix W, where the entries in each column of W sum up to 1. Furthermore, we define restart vector r by setting $r_i = 1/k$ if protein i is in the Wnt core, $r_i = 0$ otherwise. Here k denotes the number of proteins in the Wnt core, so the 1-norm of r is equal to 1 as well. Subsequently, we define the 'random walk-based score' vector x with the following mutually reinforcing relationship:

$$x = cr + (1-c)Wx. \tag{2}$$

Here $0 \le c \le 1$ denotes the restart probability, i.e. it is a parameter that adjusts the balance between connectivity to Wnt core and the overall network topology. In the context of disease gene prioritisation, the effect of c was shown to be minimal as long as it is not very close to 0 or 1 (Erten et al., 2011). In our experiments, we use $c = 0.5$ to allow the Wnt core and global network topology to equally contribute to the score of each protein. In practice, the vector x can be computed iteratively by initialising x to a vector of $1/|V|s$ and executing the operation defined by equation (2) to recompute a new x until x does not change any more.

Recently, it was shown that random walk-based scores are heavily influenced by the network connectivity of individual proteins and proteins with a high number of known interactions

$$Ax = \lambda x;\ x_i = \frac{1}{\lambda}\sum_{j=1}^{n} a_{ij}x_j;\ i = 1,\ldots,n \tag{3}$$

are favoured by the random walk model (Erten et al., 2011). Since our purpose here is to identify new components of the Wnt signalling network, which may be relatively less studied and hence may have a lower number of known interactions, we correct the random walk-based scores using the standard eigenvector centrality scores. To achieve this, we define the global Wnt-connectivity score of a protein as the log-ratio of the likelihood that the protein will be visited by random walk that makes frequent restarts at the Wnt core to the likelihood that the protein will be visited by a random walk that does not make restarts. Namely, x_i denotes the random walk-based score of protein i that is computed by setting $c = 0.5$ and y_i denotes that computed by setting $c = 0$, the global Wnt-connectivity score s_i of protein i is computed as

$$S_i = \log(X_i/Y_i) \tag{4}$$

and these s_i scores are used to rank to proteins according to their global connectivity to the Wnt core. Here x_i denotes the likelihood that protein i will be visited by a random walk that makes frequent restarts (at every other step, on an average) at Wnt core proteins, whereas y_i denotes the likelihood that protein i will be visited by a random walk that does not make restarts. Hence, x_i measures proximity to Wnt core proteins, while y_i measures proximity to all proteins in the network. Consequently, s_i provides a measure of the proximity to Wnt core as compared to all other proteins in the network. Correction with the centrality score will increase the rank of a protein if the protein is loosely

connected but all its connections are in close proximity of Wnt core proteins. Similarly, if a protein is in close proximity of many Wnt core proteins, but if it is also heavily connected to many other proteins in the network, then the correction with centrality score will decrease the rank of this protein since its proximity to Wnt core is not specific.

2.4 Identifying Wnt signalling sub-networks

To identify sub-networks in the integrated PPI/gene-expression network, we used the jActiveModule (Cytoscape plugin) software (Cline et al., 2007) in simulated annealing mode as follows. The *p*-values in the gene-expression data were converted to *z*-scores, with smaller p-values corresponding to larger z-scores. The aggregate *z*-score for a sub-network *A* with *K* genes is computed by summing up the z_i over all the genes in the sub-network:

$$Z_A = \frac{1}{\sqrt{k_{i\in A}}} \sum z_i \tag{5}$$

High-scoring sub-networks in the PPI network were searched using simulated annealing (Kirkpatrick et al., 1983; Ideker et al., 2002). In this algorithm, each node is associated with an active/inactive state. The parameters for quenching were set with start temperature 1.0 and end temperature 0.01. The overlap threshold to identifying sub-networks was set at 0.5 and with iteration of 10^6. The highest scoring sub-networks of *z*-score greater than of 3.0 were selected as active network in our study.

2.5 Additional methods

To identify similarities in sub-networks identified with each gene-expression data set, we used the Jaccard index as follows (Steinbach et al., 2005).

$$\text{Jaccard Index}\left(\text{A}, \text{B}\right) = \left(\frac{A \cap B}{A \cup B}\right) = \frac{n_i}{\left(n_a + n_b + n_i\right)} \tag{6}$$

where n_a and n_b are the number of proteins in two sub-networks A and B, respectively, and n_i is the number of proteins in common between A and B. Spotfire software (TIBCO) was used for computation of gene-expression correlations (Pearson's correlation) and hierarchical clustering.

3 Results

3.1 Local analysis of Wnt networks

For each protein in the network, a single *p*-value representing the significance of that protein's connections to the core was obtained from the hypergeometric test. It was observed that the significance of connectivity to the core was ordered as follows: core > union > other. The top ten ranked 'other' proteins (*p*-value < .01) not annotated as Wnt signalling in the four databases used for network construction are listed in Table 3. Interestingly, although they are not in the Wnt pathway databases, they have all been the

focus of separate experimental studies linking them to Wnt signalling in the literature (Briggs et al., 2002; Kioussi et al., 2002; Hering and Sheng, 2002; Ren et al., 2002; Oishi et al., 2003; Hsieh et al., 2003; Lu et al., 2004; Edlund et al., 2005; Weiske et al., 2007; Tanegashima et al., 2008). The complete table of proteins in HPRD and their connectivity to the Wnt pathway are listed in Supplementary Table S1.[1]

Table 3 Top ten proteins based on local analysis not previously annotated as Wnt pathway members in the selected public pathway databases

Gene Symbol	*Total Degree*	*Degree-to-core*	*Percent-to-core*	*p-value*
ROR2	4	3	0.75	5.86E-06
ARHGJ	4	3	0.75	5.86E-06
MESD2	2	2	1	1.30E-04
SMAD7	56	5	0.089286	4.50E-04
RYK	6	2	0.333333	0.001891
PITX2	7	2	0.285714	0.002627
MUC1	26	3	0.115385	0.003161
IQGAP1	35	3	0.085714	0.007375
FHIT	12	2	0.166667	0.00795
DLG4	113	5	0.044248	0.009749

3.2 Global analysis of Wnt networks

The random walk-based score for each protein was calculated, representing the total percentage of time spent at each node for a given restart parameter. To capture both high-degree and low-degree nodes, we used a restart parameter of .5 (raw score) as well as a restart parameter of 0. From there we calculated a 'corrected score' as the log (base 2) of the raw score divided by the non-restart centrality score. We then ranked the proteins both according to their raw random walk scores and corrected random walk scores. To ensure the score most benefited a protein, we considered the better rank, whether raw or corrected, as the final rank of that protein. Because percentage and log work on different scales, the scores were normalised by percentile. Then, each group was validated by random set generation, with significance determined by random testing. Figure 1 illustrates the ROC curve assessing the success of the global analysis in identifying the connectivity of union proteins to the core proteins. The significance of the union set in terms of its global connectivity to the Wnt core is quite pronounced, suggesting that the proposed ranking scheme indeed ranks the proteins that are functionally associated with Wnt signalling higher. The top ten ranking proteins that are newly identified using global analysis are listed in Table 4.The complete table of protein in global analysis and their connectivity to the Wnt pathway are listed in Supplementary Table S2.[1] We also compared the results from local analysis and global analysis, and identified ROR2 and RYK as proteins occurring within the top ten most ranks of both analyses. Both of these proteins, though not present in the pathway databases have been shown to function in Wnt signalling (Lu et al., 2004; Li et al., 2008).

Figure 1 The performance of global network analysis in scoring the global connectivity of 'Union' proteins (i.e. those occurring in at least one Wnt pathway database) to the 'Core' proteins (i.e. those occurring in at least 3 of the 4 databases Wnt pathway databases). All proteins in the network were ranked according to their global Wnt connectivity scores. Subsequently, the fraction of Union proteins (sensitivity) that rank in the top x% (1-specificity) according to global connectivity score was computed and plotted for each value of x. The red curve shows 'Union' proteins, while the error bars show the distribution of the respective value for random sets of proteins with cardinality equal to that of the 'Union' set (see online version for colours)

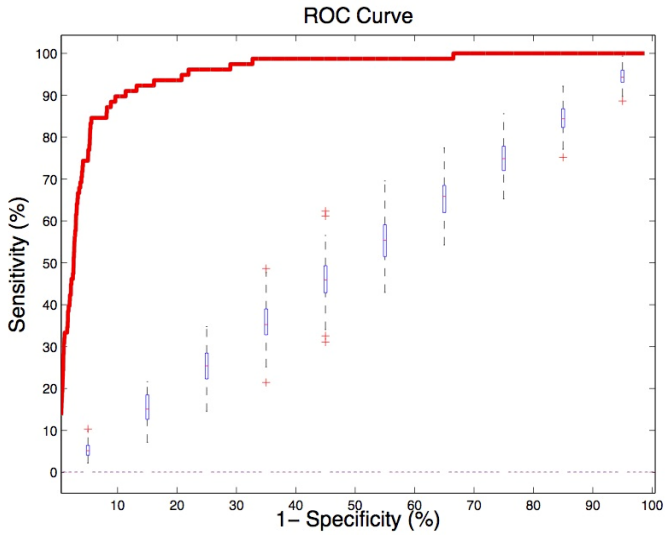


Table 4 Top ten proteins based on global analysis not previously annotated as Wnt pathway members in the selected public pathway databases

Gene Symbol	*Total Degree*	*Degree-to-core*	*Centrality score*	*Likelihood score*
DKK2	1	1	4.20E-005	8.92
SFRP1	4	4	8.40E-005	7.13
ROR2	4	3	8.40E-005	6.82
KREMEN1	1	1	4.20E-005	6.81
MESDC2	1	1	4.20E-005	6.86
CCND1	43	2	2.80E-004	5.92
RYK	5	2	9.40E-005	5.62
CXXC4	1	1	4.20E-005	5.57
CCDC88C	1	1	4.20E-005	5.57
SFRP2	5	2	9.40E-005	5.56

We next analysed functional and pathway groups for significantly connected proteins that were not already included in the core or union sets. We first ranked the list of proteins from the global analysis by likelihood, and then analysed enriched functional groups (Ingenuity Pathways Tool). We first partitioned the set of connected proteins into those with log likelihood score > 3 (set 1) (excluding all core and union proteins) and those with log likelihood <3 but >1 (set 2). The top ten most enriched pathways and functional

groups are shown in Figures 2–5. For the proteins in set 1, there is very significant occurrence of proteins annotated as functioning in Wnt signalling, Cancer and pluripotency. The occurrence of Wnt signalling as the top most significant process/pathway for this set indicates that our computational approach is working, but also points to the discrepancies between pathway databases, and Gene Ontology annotations, in terms of which proteins are considered as functioning in the Wnt signalling pathway. In set 2, Wnt signalling is not as significant, but there are many other significant processes that relate to Wnt signalling such as cellular growth and proliferation and tissue development and pathways such as Rac and Rho signalling.

Figure 2 Top ten canonical signalling pathway of set of proteins with log likelihood score greater than 3 and outside of core and union sets (see online version for colours)

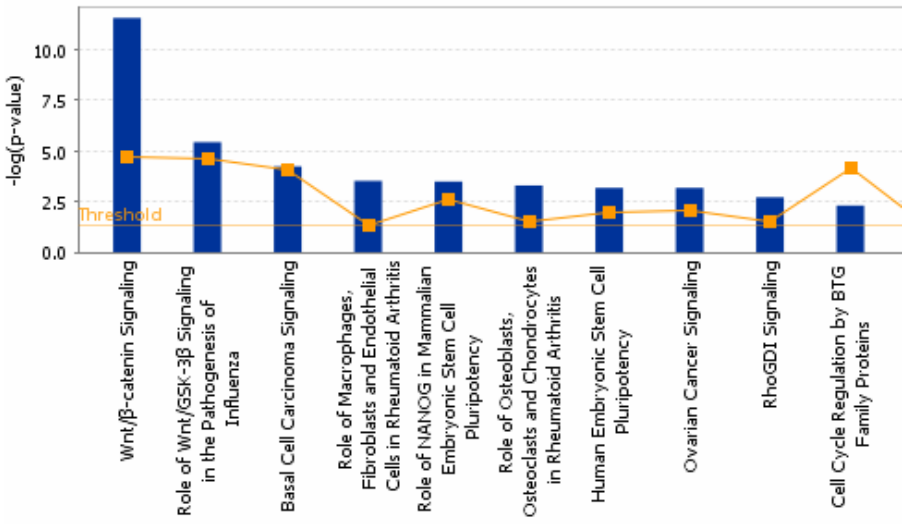


Figure 3 Top ten functional category of set of proteins with log likelihood score greater than 3 and outside of core and union sets (see online version for colours)

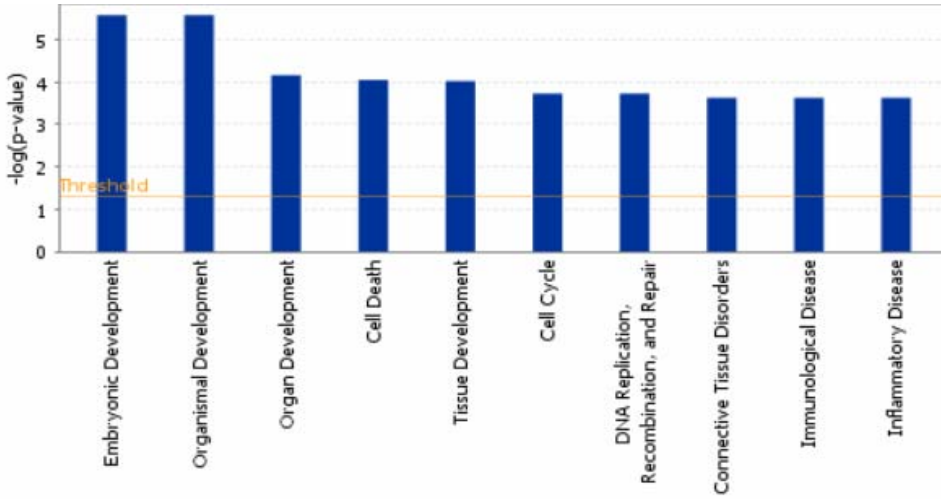


Figure 4 Top ten canonical signalling pathway of set of proteins with log likelihood score less than 3 but greater than 1 and outside of core and union sets (see online version for colours)

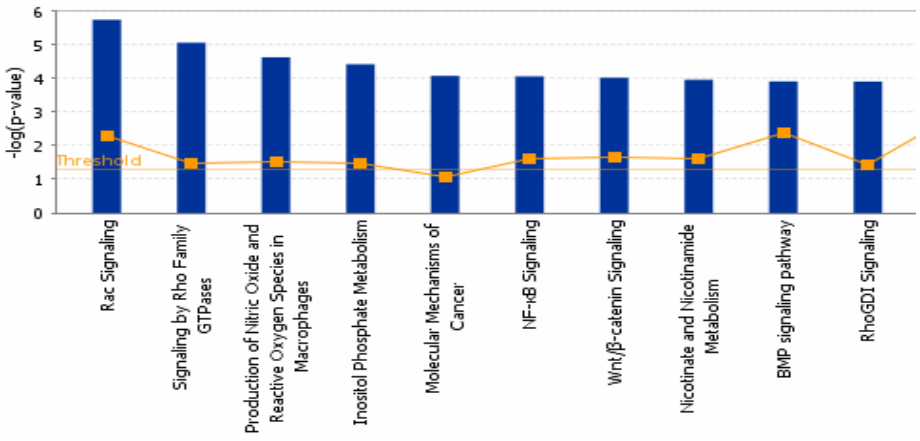
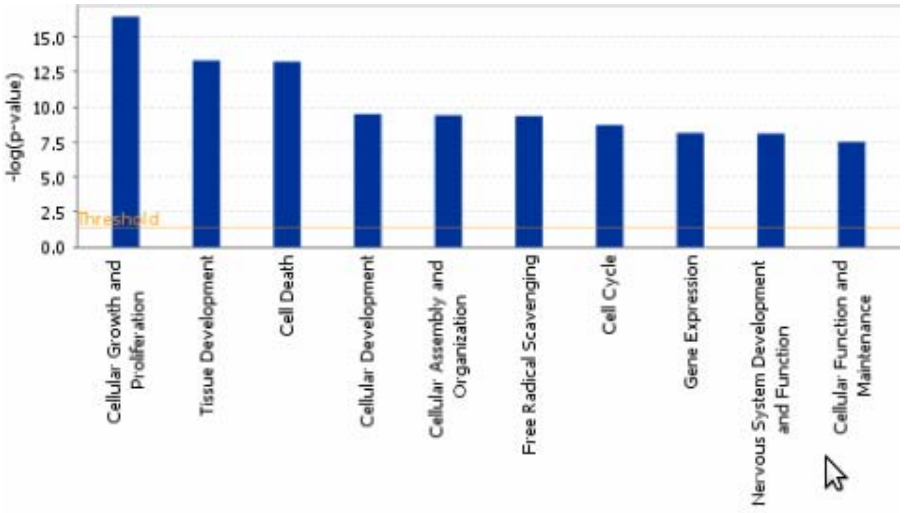


Figure 5 Top functional category of set of proteins with log likelihood score less than 3 but greater than 1 and outside of core and union sets (see online version for colours)



3.3 Tissue specific Wnt sub-networks

Next, rather than focusing on identifying network nodes that are significantly connected to the Wnt pathway, we aim to identify sub-networks that are activated under selected Wnt perturbations. The Wnt perturbations are represented by the set of gene-expression studies, selected from GEO. For each gene-expression study representing activation of the Wnt signalling pathway (Table 2), we searched for active sub-networks in the Wnt-focused protein interaction network (2072 protein interactions, 363 proteins). Simulated annealing was used to identify sub-networks with the following parameters: (number of

iteration (N) = 10^6; T_{start} = 1; T_{end} = 0.01; Number of modules = 5; d_{min} = 100). Top scoring sub-networks (size of sub-network, z-score and number of annotated cancer genes from the Cancer Gene Census) for each gene-expression study are shown in Table 5. We identified ten top scoring sub-networks using a threshold of z-score > 3. Where multiple significant (z-score > 3) sub-networks were identified for a gene-expression study, the sub-network with greatest representation of Wnt core components was selected for further analysis (sub-network 2 in GSE1473 (EKC Sn 2), sub-network 1 in GSE16186 (MSC Sn 1) and sub-network 2 in GSE17385 (MMC Sn 2). Each high-scoring sub-network consists of 31.25–46.87% of core Wnt components and ~3% cancer annotated genes. For each pair of gene-expression studies, we computed the pairwise similarity between high-scoring sub-networks as in Figure 6. Several pairs of similar sub-networks were observed. Notably, high-scoring sub-networks in lung cancer adenocarcinoma LCA Sn 1 (GSE14107) and colorectal cancer adenoma CCA Sn 1 (GSE8671) were the highest scoring pair (Figure 6). SIF format files of all sub-networks are provided in the supplementary data.[1]

Figure 6 Heat map of sub-network similarities based on Jaccard index. Grey scale colours indicate high (black) to low (white) similarity

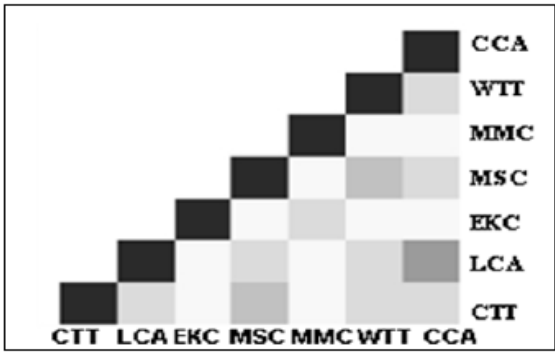


Table 5 Summary of active sub-networks perturbed in Wnt signalling network

GEO Accession	*Sub-network*	*Number nodes*	*Z-score*	*Wnt Core targets**	*Cancer genes***
GSE10972	CTT Sn 1	52	7.6	10 (31.25%)	14 (3.06%)
GSE14107	LCA Sn 1	90	7.7	14 (43.75%)	19 (4.15%)
GSE1473	EKC Sn 1	46	6.2	6 (18.75%)	11 (2.4%)
GSE1473	EKC Sn 2	64	4.7	11 (34.37%)	10 (2.18%)
GSE16186	MSC Sn 1	55	6.9	15 (46.87%)	12 (2.62%)
GSE16186	MSC Sn 2	72	4.2	3 (9.37%)	13 (2.84%)
GSE17385	MMC Sn 1	42	6.7	2 (6.25%)	8 (1.7%)
GSE17385	MMC Sn 2	60	3.8	10 (31.25%)	6 (1.31%)
GSE6120	WTT Sn 1	66	7.4	14 (43.75%)	15 (3.28%)
GSE8671	CCA Sn 1	89	8.4	13 (40.62%)	15 (3.28%)

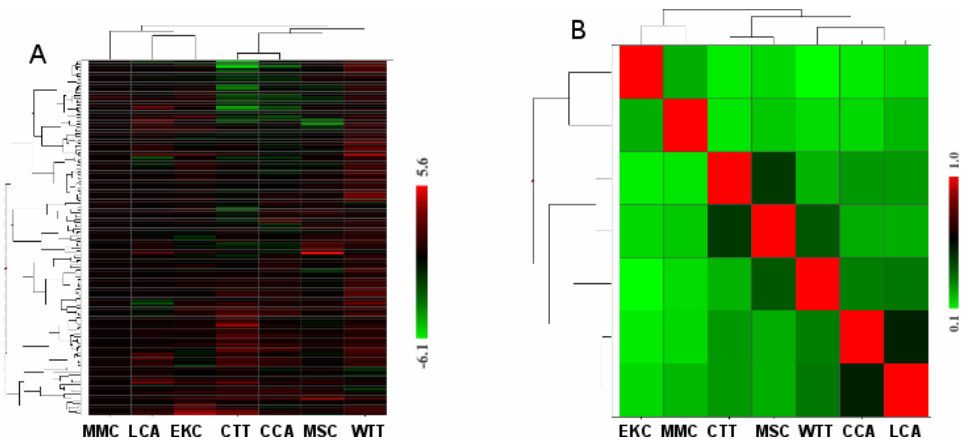
Notes: * Wnt core targets (total 32 proteins); ** Annotated cancer genes (from a total of 457 genes).

3.4 Classifying samples by analysis of gene-expression profiles and sub-networks

We hypothesised that comparisons of sub-networks obtained from our analysis might reveal commonalities between studies not revealed by more conventional means of classifying gene-expression data sets, such as classification by clustering of gene-expression profiles (Sorlie et al., 2003). We first clustered the complete gene-expression data set (8044 genes × 7 studies) or the set of 363 genes represented in the PPI (363 genes × 7 studies) using hierarchical clustering. Figure 6A shows hierarchical clustering of the set of 363 genes (clustering of the complete data set is provided in Figure 2). The patterns of similarities between studies were similar regardless of whether the full set of 8044 gene-expression profiles or the set of 363 gene-expression profiles was used or clustering. Notably, and as expected, the tissue of origin is a principal driver of the groupings observed in the hierarchical clustering of the gene-expression data alone. For example, the Colon Tumour Tissue (CTT) study is most similar to the colorectal adenoma (CAA) study.

We next used the Jaccard Index matrix of similarities (Figure 6) to hierarchically cluster and reorder the studies according to the similarity of their highest scoring sub-networks (Figure 7B). The sub-network-based clustering generated quite different patterns of similarity between studies, grouping the Colon Cancer Adenoma (CCA) study with the lung cancer adenocarcinoma (LCA) study. Thus, we observed that grouping studies according to similarity of their activated Wnt sub-networks rather than according to similarity of global gene-expression profiles gave different results. We hypothesised that similarity according to Wnt-focused sub-networks might reveal similarities between studies that are not identified by gene-expression clustering, and may represent Wnt sub-networks with important functional characteristics.

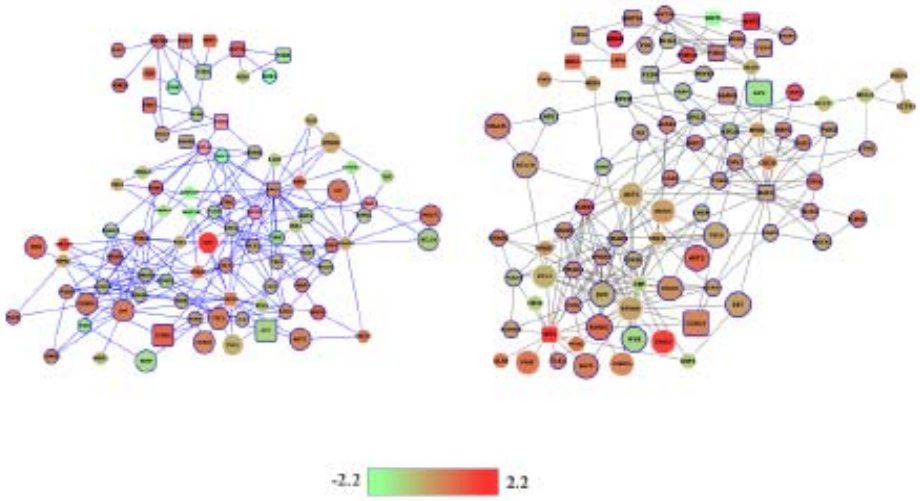
Figure 7 Comparison of clustering using (A) gene expression data alone (log_2 fold-change) or (B) sub-networks from integrated gene-expression and protein interaction networks (Jaccard index) (see online version for colours)



3.5 Comparison of adenoma sub-networks

To explore in more depth similar sub-networks observed between studies, we analysed the sub-networks identified in the lung cancer adenocarcinoma (LCA) and CCA studies. The pair of high-scoring sub-networks from these two studies was found to have the greatest similarity according to the Jaccard index (Figure 6). Sixty-three genes were in common between the lung adenocarcinoma and colorectal adenoma sub-networks, with an additional ~30 genes unique to each sub-network (a statistically significant overlap, p-value = $3.66e^{-28}$ using Fisher's exact test). Gene-expression profiles corresponding to this core set of 63 genes are provided in Supplementary Table S1.[1] In both the Wnt signalling perturbed sub-networks, we observed ~41% of Wnt core components and ~3% annotated cancer genes. Interestingly, both of these networks encapsulate proteins involved at different levels of the Wnt signalling pathway (from Wnt ligands such as WNT2, WNT3, WNT3A and WNT5A), to receptors such as FZD1 and FZD7 to transcription factors such as MYC, CCND1 and CREB1 (Figure 8). We observed both known Wnt canonical components (MYC, DKK1, CCND1, WNT2 and WNT3A) and non-canonical components (CREB1 and WNT5A) in these active sub-networks, further supporting the notion of the Wnt signalling network, without separation of canonical and non-canonical pathways. There are also several genes not usually linked directly to the Wnt signalling pathway such as DLG1, DLG2, DLG3, RUNX1 and EP300). We observed in both the sub-networks several cancer genes such as AKT2, CCND1, RUNX1 and MALT1 that were induced while others such as APC and BCL10 were repressed. In addition, there are specific-cancer genes represented in each sub-network such as MYC in the colorectal adenoma and CDX2 in the lung adenocarcinoma that might represent tissue- or organ-specific components.

Figure 8 Comparison of adenoma sub-networks: (A) An active sub-network of colorectal cancer adenomas (CCA); (B) An active sub-network of lung cancer adenocarcinoma (LCA). Wnt core genes are shown as rectangles, and circles represent other signalling circuit genes; cancer genes are shown in large shape nodes. Induced and repressed genes are shown as red and green colour nodes respectively ($\log_2$ fold-change). Nodes in common between 'CCA' and 'LCA' sub-networks are shown with blue borders (see online version for colours)



4 Discussion

In this study, we present two approaches to (a) extend the Wnt signalling network and (b) identify context-specific Wnt sub-networks that can be compared, and potentially used to classify Wnt perturbations.

Our motivation for attempting to extend the Wnt signalling pathway is that the underlying cellular network that governs Wnt signalling is most probably considerably more extensive than current representations of core Wnt pathway components suggest. This is evidenced, for example, by the multiple RNAi screens that have been performed, that typically identify significant numbers of genes whose expression is able to alter Wnt activity in genetic screens (e.g. Major et al., 2008). Although false positives may contribute to the numbers of genes identified in these screens, it still seems likely that a broad cellular network governs overall activity of the pathway. Our approach, using either local or global connectivity of proteins to core Wnt pathway members provides a computational screen in which novel proteins may be identified, and future work integrating this computational approach with the results from genetic/RNAi screens may also be fruitful. Our analysis also takes into consideration differences between pathway databases as to the composition of the Wnt signalling pathway. Although there exists a core set of proteins that are present in all four databases, many more proteins 'union' set feature in one or two pathway databases. These discrepancies are due to the subjectivity in defining the Wnt pathway (or any other biological pathway). Network-based approaches, as described here, will be required in future definitions of what the core or extended composition of signalling pathways are.

The sub-networks are scored by integrating a Wnt-focused PPI network and multiple gene-expression studies. We identify multiple sub-networks that incorporate known and unknown components of the Wnt signalling pathway as well as linking genes whose expression may be induced or repressed. Using a focused Wnt PPI network instead of a large-scale PPI network enables us to use the computationally intensive simulated annealing approach rather than a greedy approach and to identify smaller sub-networks specific to Wnt signalling.

Identification of similarities between gene-expression studies based upon shared sub-networks may reveal commonalities between studies that are not revealed by analysis of gene-expression data alone. Thus, classification of studies using sub-networks grouped two adenoma ('CCA' and 'LCA') studies together, while the gene-expression clustering grouped studies based upon tissue type. A possible explanation for the common sub-network found for the colorectal adenoma ('CCA') and lung adenocarcinoma ('LCA') studies is that they both represent early stages of tumorigenesis and may have common mediators of chemotactic invasion and colony outgrowth-dependent Wnt signalling (Sabates-Bellver et al., 2007; Nguyen et al., 2009). Wnt ligands such as WNT5a and WNT2 were observed in both sub-networks and it has been reported that WNT5A high expression increased motility and invasion in cancer progression (Da Forno et al., 2008). In both the adenoma sub-networks, cancer genes such as AKT2, CCND1, RUNX1 and MALT1 were induced. The role of RUNX1 in disease prognosis is not yet well reported. However, RUNX3 has been shown as an initiator of colonic carcinogenesis by linking the Wnt oncogenic and TGF-beta tumour suppressive pathways (Subramaniam et al., 2009). Identifying sub-networks important during tumorigenesis is becoming increasingly important, since networks themselves, rather than individual proteins are now seen as

potential drug targets (Brehme et al., 2009). Future analysis of cancer-associated or signalling pathway associated sub-networks may identify sub-networks that can be targeted therapeutically.

Although, our study and the work of many others using similar approaches shows promise, we recognise several current drawbacks. For example, to understand the dynamics of PPI networks, we require accurate measures of protein abundance and post-translational state. Although gene-expression measurements are convenient, and widely available, in many instances they may not reflect the expression of corresponding proteins. Second, tissue-type or cell-type specific PPI data are not available on a large scale, hence the use of a reference database in our study. There may be many interactions occurring in cell- or tissue-specific manner that are not represented, and conversely many interactions in the reference database may not occur in the tissue or biological state of interest. In addition, PPI data from large-scale studies are of variable quality, and a future direction would be to weight the interactions in the network accordingly. In addition, we recognise that this study utilises only a small number of diverse gene-expression studies. Using a larger set of gene-expression studies may well help better define the important sub-networks.

Acknowledgements

TR and AG were funded by NSF grant number DUE-0634612 through the RIBMS program at CWRU and support from SOURCE via the Case Alumni Association (CAA). RME acknowledges funds from the Center for Proteomics and Bioinformatics and Cleveland Foundation used in part to fund this study.

References

Brehme, M., Hantschel, O., Colinge, J., Kaupe, I., Planyavsky, M., Köcher, T., Mechtler, K., Bennett, K.L. and Superti-Furga, G. (2009) 'Charting the molecular network of the drug target Bcr-Abl', *Proceeding of the National Academy of Science USA*, Vol. 106, No. 18, pp.7414–1719.

Briggs, M.W., Li, Z. and Sacks, D.B. (2002) 'IQGAP1-mediated stimulation of transcriptional co-activation by beta-catenin is modulated by calmodulin', *Journal of Biological Chemistry*, Vol. 277, No. 9, pp.7453–7465.

Brin, S. and Page, L. (1998) 'The anatomy of a large-scale hypertextual web search engine', *Computer Networks and ISDN Systems*, Vol. 30, pp.107–117.

Chamorro, M.N., Schwartz, D.R., Vonica, A., Brivanlou, A.H., Cho, K.R. and Varmus, H.E. (2005) 'FGF-20 and DKK1 are transcriptional targets of beta-catenin and FGF-20 is implicated in cancer and development', *EMBO Journal*, Vol. 24, No. 1, pp.73–84.

Chen, J., Aronow, B.J. and Jegga, A.G. (2009) 'Disease candidate gene identification and prioritization using protein interaction networks', *BMC Bioinformatics*, Vol. 10, p.73.

Chilov, D., Sinjushina, N., Saarimäki-Vire, J., Taketo, M.M. and Partanen, J. (2010) 'Beta-catenin regulates intercellular signalling networks and cell-type specific transcription in the developing mouse midbrain-rhombomere 1 region', *PLoS One*, Vol. 5, No. 6, e10881p.

Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. and Ideker, T. (2007) 'Network-based classification of breast cancer metastasis', *Molecular Systems Biology*, Vol. 3, 140p.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T. and Bader, G.D. (2007) 'Integration of biological networks and gene expression data using cytoscape', *Nature Protocols*, Vol. 2, No. 10, pp.2366–2382.

Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P. and Gauthier, J.M. (2004) 'Functional proteomics mapping of a human signaling pathway', *Genome Research*, Vol. 14, no. 7, pp.1324–1332.

Da Forno, P.D., Pringle, J.H., Hutchinson, P., Osborn, J., Huang, Q., Potter, L., Hancox, R.A., Fletcher, A. and Saldanha, G.S. (2008) 'WNT5A expression increases during melanoma progression and correlates with outcome', *Clinical Cancer Research*, Vol. 14, No. 18, pp.5825–5832.

DasGupta, R., Kaykas, A., Moon, R.T. and Perrimon, N. (2005) 'Functional genomic analysis of the Wnt-wingless signaling pathway', *Science*, Vol. 308, No. 5723, pp.826–833.

Dutta-Simmons, J., Zhang, Y., Gorgun, G., Gatt, M., Mani, M., Hideshima, T., Takada, K., Carlson, N.E., Carrasco, D.E., Tai, Y.T., Raje, N., Letai, A.G., Anderson, K.C. and Carrasco, D.R. (2009) 'Aurora kinase A is a target of Wnt/beta-catenin involved in multiple myeloma disease progression', *Blood*, Vol. 114, No. 13, pp.2699–26708.

Edlund, S., Lee, S.Y., Grimsby, S., Zhang, S., Aspenström, P., Heldin, C.H., and Landström, M. (2005) 'Interaction between Smad7 and beta-catenin: importance for transforming growth factor beta-induced apoptosis', *Molecular Cell Biology*, Vol. 25, No. 4, pp.1475–1488.

Erten, S., Bebek, G., Ewing, R.M. and Koyutürk, M. (2011) 'DADA: degree-aware algorithms for network-based disease gene prioritization', *BioData Mining*, Vol. 4, 19p.

Hering, H. and Sheng, M. (2002) 'Direct interaction of Frizzled-1, -2, -4, and -7 with PDZ domains of PSD-95', *FEBS Letters*, Vol. 521, No. 1, pp.185–189.

Hsieh, J.C., Lee, L., Zhang, L., Wefer, S., Brown, K., DeRossi, C., Wines, M.E., Rosenquist, T. and Holdener, B.C. (2003) 'Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity', *Cell*, Vol. 112, No. 3, pp.355–367.

Ideker, T., Ozier, O., Schwikowski, B. and Siegel, A.F. (2002) 'Discovering regulatory and signalling circuits in molecular interaction networks', *Bioinformatics*, Vol. 18, Suppl. 1, pp.S233–S2340.

Jiang, X., Tan, J., Li, J., Kivimäe, S., Yang, X., Zhuang, L., Lee, P.L., Chan, M.T., Stanton, L.W., Liu, E.T., Cheyette, B.N. and Yu, Q. (2008) 'DACT3 is an epigenetic regulator of Wnt/beta-catenin signaling in colorectal cancer and is a therapeutic target of histone modifications', *Cancer Cell*, Vol. 13, No. 6, pp.529–541.

Katoh, M. (2002) 'WNT and FGF gene clusters', *International Journal of Oncology*, Vol. 21, No. 6, pp.1269–1273.

Katoh, M. and Katoh, M. (2007) 'Comparative integromics on non-canonical WNT or planar cell polarity signaling molecules: transcriptional mechanism of PTK7 in colorectal cancer and that of SEMA6A in undifferentiated ES cells', *International Journal of Molecular Medicine*, Vol. 20, No. 3, pp.405–409.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. and Tanabe, M. (2012) 'KEGG for integration and interpretation of large-scale molecular data sets', *Nucleic Acids Research*, 40(Database issue), pp.D109–D114.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R. and Pandey, A. (2009) 'Human Protein reference database–2009 update', *Nucleic Acids Research*, Vol. 37 (Database issue), pp.D767–D772.

Kestler, H.A. and Kühl, M. (2008) 'From individual WNT pathways towards a WNT signalling network', Philosophical Transactions of the Royal Society B: Biological Sciences, Vol. 363, No. 1495, pp.1333–1347.

Kioussi, C., Briata, P., Baek, S.H., Rose, D.W., Hamblet, N.S., Herman, T., Ohgi, K.A., Lin, C., Gleiberman, A., Wang, J., Brault, V., Ruiz-Lozano, P., Nguyen, H.D., Kemler, R., Glass, C.K., Wynshaw-Boris, A. and Rosenfeld, M.G. (2002) 'Identification of a Wnt/Dvl/beta-Catenin --> Pitx2 pathway mediating cell-type-specific proliferation during development', *Cell*, Vol. 111, No. 5, pp.673–685.

Kirkpatrick, S., Gelatt Jr, C.D. and Vecchi, M.P. (1983) 'Optimization by simulated annealing', *Science*, Vol. 220, No. 4598, pp.671–680.

Lee, E., Salic, A., Krüger, R., Heinrich, R. and Kirschner, M.W. (2003) 'The roles of APC and Axin derived from experimental and theoretical analysis of the WNT pathway', *PLoS Biology*, Vol. 1, No. 1, pp.E10.

Li, C.M., Kim, C.E., Margolin, A.A., Guo, M., Zhu, J., Mason, J.M., Hensle, T.W., Murty, V.V., Grundy, P.E., Fearon, E.R., D'Agati, V., Licht, J.D., Tycko, B. (2004) 'CTNNB1 mutations and overexpression of Wnt/beta-catenin target genes in WT1-mutant Wilms' tumors', *American Journal of Pathology*, Vol. 165, No. 6, pp.1943–1953.

Li, C., Chen, H., Hu, L., Xing, Y., Sasaki, T., Villosis, M.F., Li, J., Nishita, M., Minami, Y. and Minoo, P. (2008) 'Ror2 modulates the canonical Wnt signaling in lung epithelial cells through cooperation with Fzd2', *BMC Molecular Biology*, Vol. 9, 11p.

Lu, W., Yamamoto, V., Ortega, B. and Baltimore, D. (2004) 'Mammalian Ryk is a Wnt coreceptor required for stimulation of neurite outgrowth', *Cell*, Vol. 119, No. 1, pp.97–108.

Major, M.B., Camp, N.D., Berndt, J.D., Yi, X., Goldenberg, S.J., Hubbert, C., Biechele, T.L., Gingras, A.C., Zheng, N., MacCoss, M.J., Angers, S. and Moon, R.T. (2007) 'Wilms tumor suppressor WTX negatively regulates Wnt/ß-catenin signaling', *Science*, Vol. 316, No. 5827, pp.1043–1046.

Major, M.B., Roberts, B.S., Berndt, J.D., Marine, S., Anastas, J., Chung, N., Ferrer, M., Yi, X.H., Stoick-Cooper, C.L., von Haller, P.D., Kategaya, L., Chien, A., Angers, S., MacCoss, M., Cleary, M.A., Arthur, W.T. and Moon, R.T. (2008) 'New regulators of Wnt/β-catenin signaling revealed by integrative molecular screening', *Science Signaling*, Vol. 1, No. 45, ra12p.

McDonald, S.L. and Silver, A. (2009) 'The opposing roles of WNT-5a in cancer', *British Journal of Cancer*, Vol. 101, No. 2, pp.209–214.

Moon, R.T. and Kimelman, D. (1998) 'From cortical rotation to organizer gene expression: toward a molecular explanation of axis specification in Xenopus', *BioEssays*, Vol. 20, pp.536–545.

Moreno, C.S. (2010) 'The sex-determining region Y-box 4 and homeobox C6 transcriptional networks in prostate cancer progression: crosstalk with the WNT, notch, and PI3K pathways', *American Journal of Pathology*, Vol. 176, No. 2, pp.518–527.

Nabieva, E., Jim, K., Agarwal, A., Chazelle, B. and Singh, M. (2005) 'Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps', *Bioinformatics*, Vol. 21, Suppl. 1, pp.i302-i310.

Nalesso, G., Sherwood, J., Bertrand, J., Pap, T., Ramachandran, M., De Bari, C., Pitzalis, C. and Dell'accio, F. (2011) 'WNT-3A modulates articular chondrocyte phenotype by activating both canonical and noncanonical pathways', *Journal of Cell Biology*, Vol. 193, No. 3, pp.551–564.

Nguyen, D.X., Chiang, A.C., Zhang, X.H., Kim, J.Y., Kris, M.G., Ladanyi, M., Gerald, W.L. and Massagué, J. (2009) 'WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis', *Cell*, Vol. 138, No. 1, pp.51–62.

Nibbe, R.K., Markowitz, S., Myeroff, L., Ewing, R. and Chance, M.R. (2009) 'Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer', *Molecular & Cell Proteomics*, Vol. 8, No. 4, pp.827–845.

Niida, A., Hiroko, T., Kasai, M., Furukawa, Y., Nakamura, Y., Suzuki, Y., Sugano, S. and Akiyama, T. (2004) 'DKK1, a negative regulator of Wnt signaling, is a target of the beta-catenin/TCF pathway', *Oncogene*, Vol. 23, No. 52, pp.8520–5826.

Oishi, I., Suzuki, H., Onishi, N., Takada, R., Kani, S., Ohkawara, B., Koshida, I., Suzuki, K., Yamada, G., Schwabe, G.C., Mundlos, S., Shibuya, H., Takada, S. and Minami, Y. (2003) 'The receptor tyrosine kinase Ror2 is involved in non-canonical Wnt5a/JNK signaling pathway', *Genes to Cells*, Vol. 8, pp.645–654.

Qiu, W., Hu, Y., Andersen, T.E., Jafari, A., Li, N., Chen, W. and Kassem, M. (2010) 'Tumor necrosis factor receptor superfamily member 19 (TNFRSF19) regulates differentiation fate of human mesenchymal (stromal) stem cells through canonical Wnt signaling and C/EBP', *Journal of Biological Chemistry*, Vol. 285, No. 19, pp.14438–14349.

Ren, J., Li, Y. and Kufe, D. (2002) 'Protein kinase C delta regulates function of the DF3/MUC1 carcinoma antigen in beta-catenin signaling', *Journal of Biological Chemistry*, Vol. 227, No. 20, pp.17616–17622.

Sabates-Bellver, J., Van der Flier, L.G., de Palo, M., Cattaneo, E., Maake, C., Rehrauer, H., Laczko, E., Kurowski, M.A., Bujnicki, J.M., Menigatti, M., Luz, J., Ranalli, T.V., Gomes, V., Pastorelli, A., Faggiani, R., Anti, M., Jiricny, J., Clevers, H. and Marra, G. (2007) 'Transcriptome profile of human colorectal adenomas', *Molecular Cancer Research*, Vol. 5, No. 12, pp.1263–1275.

Sean, D. and Meltzer, P.S. (2007) 'GEOquery: a bridge between the gene expression omnibus (GEO) and bioconductor', *Bioinformatics*, Vol. 23, No. 14, pp.1846–1847.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.L. and Botstein, D. (2003) 'Repeated observation of breast tumor subtypes in independent gene expression data sets', *Proceedings of the National Academy Science USA*, Vol. 100, No. 14, pp.8418–8423.

Steinbach, M., Tan P-N., Xiong, H. and Kumar, V. (2005) 'Objective measures for association pattern analysis', *Introduction to Data Mining*, Addison-Wesley, Boston, pp.1–21.

Subramaniam, M.M., Chan, J.Y., Yeoh, K.G., Quek, T., Ito, K. and Salto-Tellez, M. (2009) 'Molecular pathology of RUNX3 in human carcinogenesis', *Biochimica et Biophysica Acta*, Vol. 1796, No. 2, pp.315–331.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z. and He, X. (2004) 'A mechanism for Wnt coreceptor activation', *Molecular Cell*, Vol. 13, No. 1, pp.149–156.

Tanegashima, K., Zhao, H. and Dawid, I.B. (2008) 'WGEF activates Rho in the Wnt-PCP pathway and controls convergent extension in Xenopus gastrulation', *The EMBO Journal*, Vol. 27, No. 4, pp.606–617.

Wawra, C., Kühl, M. and Kestler, H.A. (2007) 'Extended analyses of the WNT/beta-catenin pathway: robustness and oscillatory behaviour', *FEBS Letters*, Vol. 581, No. 21, pp.4043–4048.

Weiske, J., Albring, K.F. and Huber, O. (2007) 'The tumor suppressor Fhit acts as a repressor of B-catenin transcriptional activity', *PNAS*, Vol. 104, No. 51, pp.20344–20349.

Zotenko, E., Mestre, J., O'Leary, D.P. and Przytycka, T.M. (2008) 'Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality', *PLoS Computational Biology*, Vol. 4, No. 8, e1000140p.

Note

1 This information is available by contacting authors.