Article

The role of protein interaction domains in the human cancer network

Shady S. Ibrahim[1], Maha AR. Eldeeb[1], Mona AH. Rady[1], Karim M. Abdel Hady[1], Mohamed S. Lotfy[1], Noha S. Farag[1], Stephan Verleysdonk[2] and Christoph P. Bagowski[3]

[1]German University Cairo, Faculty of Pharmacy and Biotechnology, New Cairo City, Egypt

[2]Interfaculty Institute for Biochemistry, University of Tübingen, Hoppe-Seyler-Str. 4, 72076 Tübingen, Germany

[3]Centogene GmbH, Department of New Technologies, 18057 Rostock, Germany

E-mail: christoph.bagowski@centogene.de

Received 18 March 2011; *Accepted 21 April 2011*; *Published online 14 May 2011*

IAEES

Abstract

Protein-protein interaction networks provide a global picture of cellular function and biological processes. Proteins interact largely through specific domains which constitute the main building blocks of an interaction network. Perturbed or dysfunctional protein interactions are linked to many diseases, including cancer.

In this study we describe the major pathways and connections within the human cancer network by a novel approach in which we overlay the human cancer network with all protein interaction domain (PID) superfamilies. Based on 38,777 experimentally derived interactions, we constructed a cancer network with 8 different levels and identified all major protein hubs within this cancer interactome. Only one percent of the cancer genes constitute over 50 percent of all interactions within the network.

In addition, we mapped 56 PID superfamilies onto the cancer network, and discovered that over 10% of protein interaction domains are overrepresented within the cancer interactome when compared to the normal protein network. We present here a comprehensive list of all PIDs in the cancer network, identify the most important hubs within it and discover several individual genes which had previously not been linked to cancer. These proteins constitute excellent targets for the development of novel cancer therapeutics. Our results further hint to a partial molecular commonality between cancer and neurodegenerative diseases such as Alzheimer's and Huntington's.

Keywords protein domain; interaction network; PDZ domain; systems biology; cancer; tumour; superfamily; metastasis; interactome; p53; signaling; protein binding; neurodegenerative disease; Alzheimer; Huntington; NMDAR.

1 Introduction

Cancer is a leading cause of death worldwide. The disease accounted for 7.4 million deaths (around 13% of all deaths worldwide) in 2007 and its associated mortality rate is expected to increase even further according to the World Health Organization.

Cancer is a heterogenous disease that can affect any part of the body. One of its defining feature is the rapid proliferation of abnormal cells that grow beyond their usual boundaries, invade adjoining parts of the body, and subsequently spread through blood and lymphatic vessels to form metastases in other organs which can lead to secondary tumours.

Malignant tumours and neoplasms arise from one single cell in a multistage process, which typically involves progression from a pre-cancerous lesion to a malignant tumour. The changes leading to cancer are the result of the interaction between a person's genetic predisposition and external factors. In 2000, Hanahan and Weinberg defined six hallmarks of cancer (Hanahan and Weinberg, 2000): 1) self-sufficiency in growth signals, 2) insensitivity to antigrowth signals, 3) evasion of programmed cell death (apoptosis), 4) limitless replicative potential, 5) sustained angiogenesis and 6) tissue invasion and metastasis. Now, ten years later, it is becoming increasingly clear that tumour progression is even more complex and requires additional hallmarks, such as a shift in cellular metabolism (Tennant et al., 2009) as well as factors from the tumour microenvironment and from the tumour stroma interaction (Pietras and Ostman, 2010).

All of these processes are ultimately linked and executed by the proteome and the road to malignancy is regulated by changes in protein activity, cellular protein composition and/or changes in protein interactions and in the protein network.

The wealth of information gathered on cancer in general and on cancer related genes, including oncogenes and tumour suppressor genes, has thus far not transformed into a clear understanding of the molecular mechanisms and the exact cause underlying the formation of malignant tumours. A major challenge is to understand how pathological phenotypes arise from changes in the protein network or alterations in proteome connectivity in this complex multigene disease. In this study we compile and characterise the cellular cancer network and analyse its network of protein-protein interactions as one of the major driving forces of tumourigenesis.

Within a complex network of interactions, a single molecule can affect a wide range of other cellular components (Chautard et al., 2009). Central to protein-protein networks or interactomes are modular protein interaction domains (PID), which organize them by promoting protein-protein binding. Within the network, adaptor proteins composed exclusively of PIDs mediate molecular interactions and link signaling proteins such as activated cell-surface receptors to downstream effectors thereby directing the flow of information (Pawson, 2007).

Protein-protein interaction (PPI) databases have become a major resource for investigating cellular protein networks and signaling pathways. A number of publicly available repositories for human PPIs are currently available and each of them has its own unique pattern and depth of annotations.

According to the integration provided by the meta-database APID, the six major primary databases (BIND, BioGRID, DIP, HPRD, IntAct, and MINT), construct the largest known interactome with a total of 80,032 interactions. The largest primary database that includes around 50% of the human PPIs (De Las Rivas and Fontanillo, 2010) is the Human Protein Reference Database (HPRD). It is a resource for experimentally derived information about the human proteome including protein–protein interactions, post-translational modifications (PTMs) and tissue expression data (Mathivanan et al., 2006; Keshava Prasad et al., 2009).

The human interactome is defined as the entirety of interactions occuring in a human cell.

The PPI data is sub-classified as binary (direct interactions between two proteins) or complex interactions based on topology and the number of participants.

In the present study we used the HPRD as the largest available list of experimentally derived human interactions to generate the human cancer interactome and performed a comparative analysis of the role of protein interaction domains in the human interactome versus the cancer network.

2 Material and Methods

2.1 Constructing and visualizing the human interactome

The HPRD (8th release) list of human binary protein-protein interactions in tab delimited format was obtained from the HPRD website. This list was loaded into the cytoscape software (version 2.6.3) to visualize the human interactome network.

2.2 Generating the human cancer network

A comprehensive list of cancer genes was obtained from the Wellcome Trust Sanger Institute website (http://www.sanger.ac.uk/). The cancer list contained 401 cancer genes, the Entrez ID and HUGO gene symbols of these were converted into the HPRD ID list format by using the converter at the Biomedical Information Research Center (BIRC) of the National Institute of Advanced Industrial Science and Technology (AIST) (http://biodb.jp/ids). After the conversion, 360 cancer gene IDs were found back in the human interactome, and their appropriate HPRD data were used to construct cancer networks on 8 different levels (examples of visualized networks are shown in Fig. 1). The first level cancer network was created from the 360 cancer nodes. Only the direct interactions between the cancer genes were included in this network. The next level cancer network included the interactions of the cancer genes with their 1st neighbours, but excluded neighbour-neighbour interactions. The third network was set to include the cancer nodes as well as their 1st and 2nd neighbours. The edges in this network represent the interactions between the cancer genes themselves, between them and their 1st or 2nd neighbours and between 1st and 2nd neighbours, but not the interactions among 2nd neighbours . The other levels of the network were constructed in the same way, and the maximum network size was obtained at the level of the 8th neighbours. Due to the large size of the higher level cancer networks and their similarity to the general interactome, further analysis was restricted to the level of the 3rd neighbour and below. Table 1 shows the connectivity of all constructed networks, and of the entire interactome.

2.3 Investigation of Protein interaction domains (PID's)

Lists of all genes encoding the respective 56 PIDs from the Superfamily database (http://supfam.cs.bris.ac.uk/SUPERFAMILY/) (Gough et al., 2001)were prepared, annotated and batch converted using the Biomart online converter (www.biomart.org). Lists of domain encoding genes were converted to the HPRD ID format (http://biodb.jp/ids) in order to facilitate their investigation within the interactome. Proteins without matching HPRD IDs were not loaded into the network (Table 2). For each domain, the list of genes was loaded into the previously designed cancer networks as well as into the total human interactome network. Data from different networks including the number of nodes, edges, average interactions, as well as theoretical number of nodes and edges were calculated and the networks were analysed using the Cytoscape Network Analyser plugin (release 2.6.1) to calculate the average number of neighbours and the average node degree.

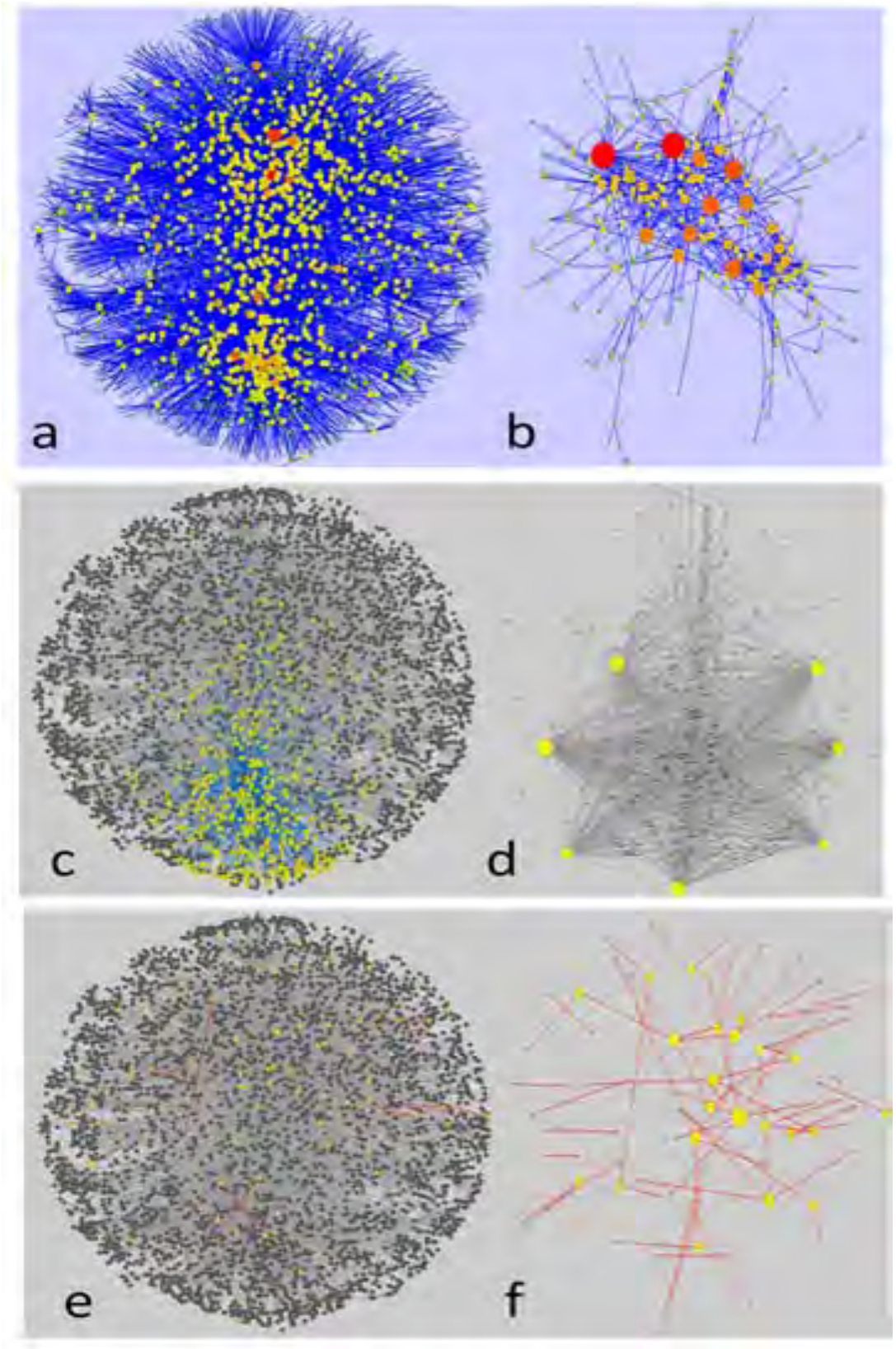

Fig. 1 Shown are different levels of the cancer network: (a) 3,068 nodes connected with 6,572 edges form the cancer network that include the cancer genes connected to their first neighbours. (b) The 360 cancer genes (nodes) connected by 695 edges, which forms the smallest network and includes only direct interactions and self-loops. Examples of Protein Interaction Domain (PID) networks are shown for the SH2 domain (in c and d) and for the WD40 domain (e and f). All figures were generated using Cytoscape.

Table 1 Shown are the numbers of nodes and edges found in the different interaction networks we constructed. The network representing the 360 cancer genes connected to their 1st neighbours was used in the whole study and is referred to herein as the cancer network.

Network	Nodes (% of Interactome)	Edges (% of Interactome)
Human Interactome	**9630 (100%)**	**38 777 (100%)**
Cancer nodes only	360 (3.73%)	695 (1.79%)
Cancer+ 1st neighbours	**3068 (31.85%)**	**6572 (16.95%)**
Cancer +2nd neighbours	7904 (82.07%)	31783 (81.96%)
Cancer +3rd neighbours	9077 (94.25%)	38001 (98%)
Cancer +4th neighbours	9202 (95.55%)	38382 (98.98%)
Cancer +5th neighbours	9219 (95.73%)	38424 (99.09%)
Cancer +6th neighbours	9224 (95.78%)	38431 (99.11%)
Cancer +7th neighbours	9226 (95.8%)	38433 (99.11%)
Cancer +8th neighbours (max.)	9227 (95.81%)	38435 (99.12%)

Table 2 List of protein interaction domains (PID) and the number of proteins containing at least one of the corresponding domains (in brackets the number of proteins found back in the interactome after conversion to the respective HPRD IDs using Biomart)

14-3-3	7 (7)	ENTH	27 (19)	PX	51 (32)
ADF	17 (16)	F-BOX	47 (26)	RGS	38 (31)
ANK	251 (115)	FERM	44 (27)	RING	309 (165)
ARM	293 (185)	FF	4 (3)	SAM	109 (74)
BAR	39 (25)	FHA	52 (40)	SH2	107 (99)
BEACH	8 (3)	FH2	22 (11)	SH3	205 (164)
BH	13 (12)	FYVE	108 (64)	SPRY	237 (115)
BIR	9 (8)	GAT	8 (8)	TIR	21 (18)
BRCT	23 (20)	GRIP	1 (1)	TPR	169 (100)
BROMO	89 (30)	GYP	3 (3)	TRAF	24 (19)
BTB-POZ	183 (87)	HECT	27 (21)	TSNARE	15 (11)
C1	58 (46)	LIM	123 (105)	TUBBY	5 (2)
C2	153 (100)	MH1	8 (8)	TUDOR	52 (23)
CH	74 (50)	NZF	4 (3)	UBA	35 (26)
CHROMO	24 (18)	PAS	33 (23)	UEV	52 (39)
DEATH	81 (65)	PB1	13 (12)	VHL	2 (2)
DEP	199 (131)	PDZ	149 (123)	WD40	256 (139)
DH	70 (45)	PH	377 (255)	WW	42 (37)
EF-hand	236 (145)	Polo-box	6 (4)		

3 Results and Discussion

3.1 Comparative analysis of protein-protein interaction networks

Recent advances in proteomics and bioinformatics have made large datasets of experimentally derived protein-protein interactions available and expanded the wealth of information on protein networks. The complete map of protein interactions that can occur in a living organism is called the interactome (Cusick et al., 2005).

Recently, interactome mapping has become an important focus of biological and biomedical research (De Las Rivas and Fontanillo, 2010). In order to compare a large global protein interaction network to one restricted to cancer-related genes, we generated two different interaction networks with different levels, using the HPRD and a list of cancer genes from the Sanger Institute in Cambridge (www.sanger.ac.uk), respectively. A network representing 360 cancer genes from this list connected to their 1st neighbours was used in the whole study and is referred to herein as the cancer network.

We created our networks in cytoscape, an open source bioinformatics software platform, which enables comparative analysis and visualization of molecular interaction networks.(www.cytoscape.org) (Cline, Smoot et al., 2007). The human interactome we constructed included 9,630 nodes and 38,777 edges (and is from now on simply called the "interactome"). The human cancer networks produced from those genes listed in the HPRD that are also featured in the list of cancer genes available at the Sanger Institute website contains from 360 (0th level) to 9077 (3rd level) nodes and from 695 (0th level) to 38 382 (3rd level) edges (Fig. 1 and Table 1) The number of total interactions (edges) in the interactome is 5.9 times higher than the one in the 1st level cancer network (Table 1). Similarly, 3.1 times more nodes are observed in the entire interactome than in the cancer network (Table 1). Almost a two-fold difference is observed in the ratio of nodes and edges, indicating that less nodes are necessary for the same number of intercations within the cancer network compared to the general interactome indicating higher connectivity in the cancer network than in the interactome. Recently, it was shown that cancer genes exhibit a network topology that is different from that of proteins not documented as being mutated in cancer with cancer proteins showing an increase in the number of proteins they interact with (Jonsson and Bates, 2006) and signaling domains were found more often in intermodular hub proteins associated with oncogenesis (Taylor et al., 2009). These findings are in accordance with our results on network topology.

3.2 Investigation of the role of protein interaction domains (PID) in the interactome and in the cancer network

Currently, 1,777 domain superfamilies are recognized in the latest release of the database Structural Classification of Proteins (SCOP) and of these less than 5% are protein interaction domains (Andreeva et al., 2008). Despite this relative low abundance in the proteome and amongst the domain families, these PID's naturally make major contributions to protein-protein interactions within protein networks. In order to asses each PID's overall contribution to the interactome and to delineate its particular role in wiring the protein-protein network, we generated a comprehensive list of all currently described PIDs within the 56 superfamilies defined in the SUPERFAMILY database (Gough et al., 2001) (Table 2). Genes coding for proteins with two different PIDs will appear in both respective superfamilies. For each of these PID superfamilies, a list of proteins containing the domain was generated from the database. The lists were then batch converted using Biomart converter and Ensemble ID's, gene names and UnigeneID's were extracted and finally converted to HPRD ID's. Subsequently the HPRD IDs were used to construct the networks and analyse the different contribution of the PID's to the entire interactome and the cancer network. Additionally, by loading the list of each PID containing protein in HPRD ID and attaching these proteins to their 1st neighbours we could create

independent PID networks. The extracted PID networks were used to calculate the average number of interactions of the PIDs in the interactome by dividing the number of edges linked to all proteins containing this PID by the total number of proteins containing this PID. The average number of interactions for the mean of all 56 PID was determined to be 15.27 interactions/protein. We plotted the average interactions of individual PIDs against this mean which highlights (from a quantitative standpoint focusing solely on binding via the PIDs) the most important PIDs in the human interactome. It is important to note that other mechanisms of binding than via PIDs can contribute to the total protein interactions, these non-PID dependent interactions are obviously not accounted for in this analysis. Table 3 shows the individual fold changes compared to the mean for all PIDs. A total number of 6 PIDs showed a >1.5 fold elevated number of average interactions with two PIDs showing exceptionally high elevation (14-3-3 and MH1 domains with around 6.5 times the average of interactions of other PIDs per gene) (Table 3). It is noteworthy that the MH1 domain is also a well known protein/DNA binding domain and it is possible that DNA binding complexes contribute to these above average interactions of MH1 domains. Conversely, 26 PIDs showed less average interactions (again with a cut off of < 1.5 fold) (Table 3).

Table 3 Shown is the number of average interactions for each PID divided by the number of average interactions of all proteins in the interactome (=15.27).

Table (2): Protein interaction domains representation in the Interactome, Cut-off =1.5 and 0.66							
PID	Folds of AI	PID	Folds of AI	PID	Folds of AI	PID	Folds of AI
14-3-3	6.781	WW	1.012	VHL	0.687	FYVE	0.492
MH1	6.571	HECT	0.973	PH	0.669	BEACH	0.480
FF	2.008	FERM	0.962	BIR	0.655	TIR	0.454
SH2	1.971	TSNARE	0.952	RGS	0.651	SPRY	0.429
FHA	1.786	C2	0.949	BAR	0.623	ADF	0.429
TRAF	1.726	PAS	0.905	DH	0.615	NZF	0.370
BROMO	1.460	BRCT	0.855	RING	0.612	TPR	0.368
C1	1.375	PDZ	0.839	EF-hand	0.605	WD40	0.368
SH3	1.239	UBA	0.824	ARM	0.599	PX	0.366
BH	1.146	DEATH	0.804	UEV	0.586	TUBBY	0.360
FH2	1.143	CHROMO	0.793	DEP	0.539	BTB-POZ	2.902
LIM	1.047	GAT	0.753	ANK	0.523	F-BOX	0.344
PB1	1.026	ENTH	0.730	TUDOR	0.506	GRIP	0.262
Polo-box	1.015	CH	0.704	SAM	0.498	GYP	0.262

In a subsequent step we analysed the contribution of PIDs to the cancer network and further compared the numbers of interactions directly to our data obtained for the entire interactome. The cancer network includes 360 cancer-related genes (these were genes with known HPRD IDs derived from the Sanger list of a total of 401 cancer-related genes) and their 1st neighbours. The 1st level cancer network consists of 3068 nodes and 6572 edges. Similar to our approach for the interactome, the average number of interactions per PID was calculated in the cancer network. In the 1st level cancer network, the average number of interactions determined for the 56 PIDs was 6.7. Only 3 domains, the BEACH, GRIP and GYP domains, were completely

absent from the cancer network a (did not have any nodes and were not correlated to either cancer genes directly or to their 1st neighbours). Threfore these domains probably play no role for the cancer interactions. In the cancer network, the FH2, BROMO and MH1 domains had the highest average interactions with 6.8, 4.7 and 4.6 times the average number of interactions of other PIDs, respectively (Table 3). The FH2 domain occurs in proteins regulating rearrangements of the actin cytoskeleton, which is a crucial process in cytokinesis and therefore of obvious relevance for the development and maintenance of cancer. Furthermore, f-actin rearrangements are of paramount importance in cell motility, which in turn is required for tumor tissue invasion and metastasis (Castrillon and Wasserman, 1994). BROMO domains mediate the recognition of acetylated histones and promote an open, transcriptionally active chromatin structure. Since tumor progression requires high transcription rates, the overrepresentation of BROMO domains in the cancer interaction network developped here is an expected result. Conversely, the CHROMO domain promotes the recognition of methylated chromatin and the formation of a closed, transcriptionally silent chromatin structure. Since this state is incompatible with tumor progression, an underrepresentation of the CHROMO domain is likewise expected. Other significant quantitative and qualitative differences were found between the interactome and the cancer network (see Fig. 3A and Table 4). In our comparison of the interactome to the cancer network, we

Table 4 Shows the number of average interactions of each PID protein divided by the number of average interactions of all proteins in the cancer network (=6.7). The Cut-offs were 1.5 fold (above green line) and 0.67 fold (below red line).

Protein interaction domain representation in the cancer network							
PID	**Fold AI**	**PID**	**Fold AI**	**PID**	**Fold AI**	**PID**	**Fold AI**
FH2	6.87	**BTB-POZ**	1	**DEP**	0.5	**CHROMO**	0.35
BROMO	4.71	**DH**	0.91	**SAM**	0.49	**PX**	0.33
MH1	4.61	**LIM**	0.91	**PDZ**	0.45	**RGS**	0.33
VHL	2.99	**ARM**	0.86	**Polo-box**	0.45	**WW**	0.26
FHA	2.3	**PH**	0.79	**TUBBY**	0.45	**TPR**	0.23
BRCT	1.96	**C2**	0.77	**TRAF**	0.43	**WD40**	0.23
SH2	1.8	**EF-hand**	0.76	**BAR**	0.42	**HECT**	0.22
14-3-3	1.69	**ANK**	0.71	**TUDOR**	0.41	**TIR**	0.2
BH	1.64	**DEATH**	0.7	**FF**	0.37	**ADF**	0.15
PAS	1.37	**BIR**	0.7	**UBA**	0.37	**NZF**	0.15
SH3	1.13	**F-BOX**	0.69	**CH**	0.36	**TSNARE**	0.14
FERM	1.09	**PB1**	0.57	**GAT**	0.36		
C1	1.07	**FYVE**	0.54	**UEV**	0.36		
RING	1.06	**ENTH**	0.53	**SPRY**	0.35		

divided the PIDs in four groups: 1) PIDs with a high number of average interactions in both normal (interactome) and the cancer network (12 PIDs); 2) PIDs with low average interactions in both normal and the cancer network (21 PIDs); 3) PIDs with low average interactions in the normal but high average interactions in the cancer network (6 PIDs) and 4) PIDs with high average interactions in normal but low interactions in the cancer network (10 PIDs). In total, 18 PIDs in (groups 1 and 3) have a higher number of interactions in the cancer network, while 31 PIDs (groups 2 and 4) have a lower number of interactions in the cancer network (Fig. 3 and Table 4). Our data shows (Fig.2) that there is a clear difference in the amount of interactions in the

interactome and the cancer network and the PIDs with the highest average interactions differ significantly between both networks.

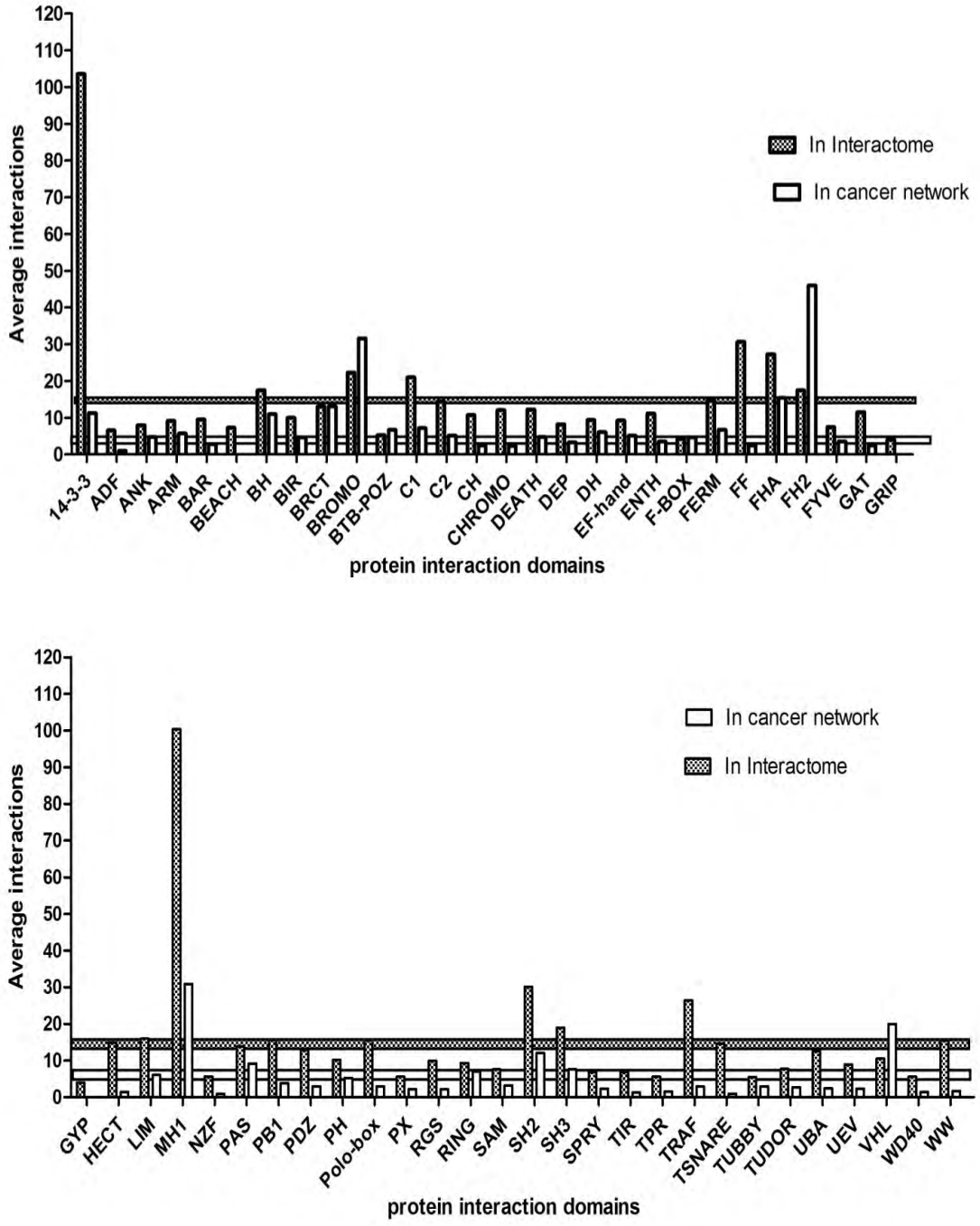

Fig. 2 Shown in alphabetical order are the average interactions for the total 56 PIDs in the interactome (grey bars) and the cancer network (white bars). The line in both graphs depicts the average number of interactions per PID per protein (15.71).

A.

B.

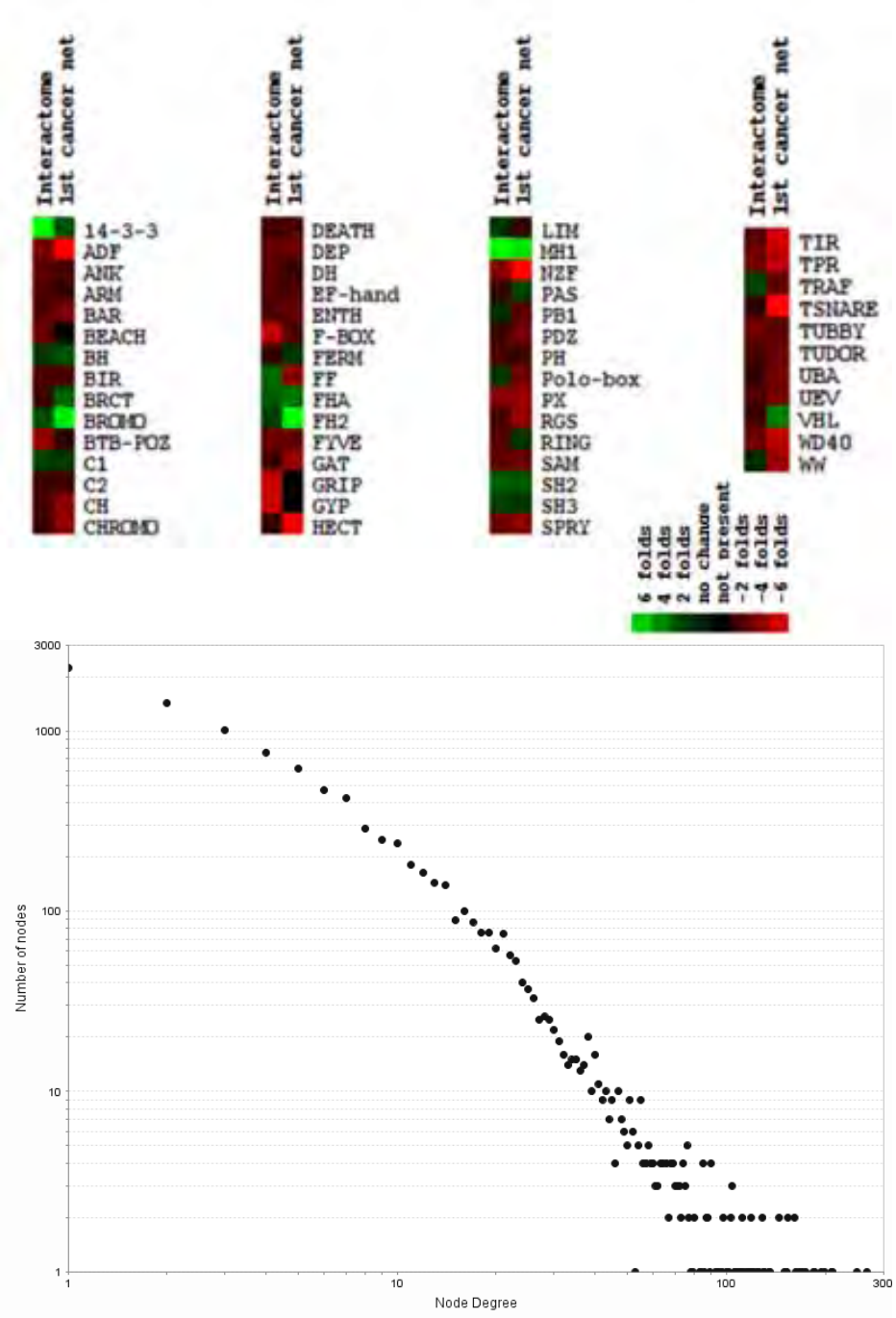

Fig. 3 (A) A heat map for the average interactions of PIDs in the ineractome and the cancer network shows the fold differences compared to the respective mean of all PIDs. Only 3 PIDs were not present in the cancer network (black) (B)The node degree distribution in the constructed interactome indicates a scale-free network.

3.3 Analysis of network connectivity and identification of major cancer hubs

In a network, a hub is a node with high connectivity. In accordance, in our interaction networks a hub is a protein with many interactions including self-binding (self-loops). To a certain extent it can be assumed that these hubs proteins play important roles in the interaction network. However, there is a balance between qualitative and quantitative contribution and their effects on respective functionality. Our plot of the node

degree distribution for the interactome shows that the network is scale-free (Fig. 3B). By further investigating the hub sizes in both the interactome and the cancer network we identified the most connected proteins for both networks, which are thus involved in the majority of interactions. We determined all hub sizes and focused our analysis on the hubs of more than 5 to above 50 interactions and used different databases and a literature search to determine their known or unknown relation to cancer. In the cancer network we found that hubs with a node degree > 50 all represent very well-characterised proteins linked to cancer, including some major oncogenes (e.g. Raf) and tumour suppressor genes (p53, Rb) (Fig.4 and Supplementary Material). By applying a filter to the interactome that maps the nodes according to their node degree, we found that nearly 10% of the nodes (942 out of 9,630) are connected directly through around 70% of the interactions (27,734 out of 38,777). However when we applied the same filter to the cancer network, we found that only 3% of the nodes (87 out of 3,068) are connected to 70% of the interactions (4,583 out of 6,572). These 87 proteins represent the core interactors in the cancer network and 82 of them are found in the list of cancer genes maintained by the Sanger Institute. Not surprisingly, all 87 genes are well known for their relation to cancer. We assembled a comprehensive list of proteins at different node degrees (a summary is represented in (Table 5) and the complete list can be found in Supplementary Material. All genes (with node degrees of 5 or above) were searched for in the major databases GeneCards, Novoseek, OMIM and in the literature. The power of our approach to uncover yet unknown relationships of certain proteins to cancer is highlighted by the identification of the NMDA receptor type 1 among the interactors with node degrees between 5 and 10. While the gene of this protein had previously appeared to be completely unrelated to cancer, North et al. (North et al., 2010) very recently reported its expression in breast cancer cell lines Mcf-7 and SKBR3. Furthermore, the

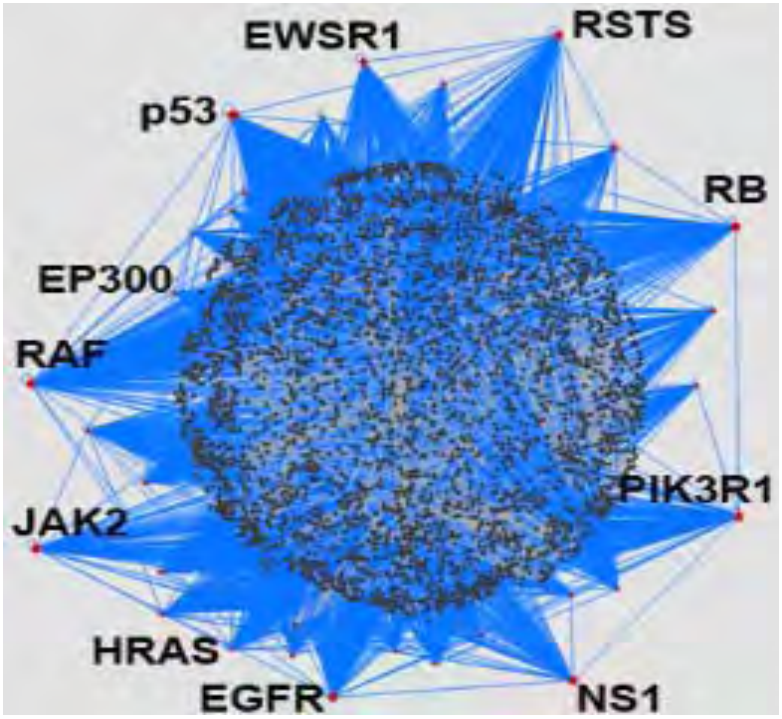

Fig. 4 Shown are the largest hubs in the cancer network (in red) connected with their direct interactors (1st neighbours) (in blue). The node degree is mapped to node size. Major oncogenes (e.g. RAF, Harvey RAS and other oncogene products and tumor suppressors are identified including the EGFR, a receptor tyrosine kinase which signals via RAF-RAS to the MAPKinases in order to promote cell proliferation. EGFR is a major target for cancer therapies today.

presence of the NMDA receptor was shown to be necessary for the growth of human breast cancer xenografts in mice (North et al., 2010). That topological features of PPI networks and protein domain compositions are good predictors for cancer genes had recently been demonstrated (Li et al., 2009). We report here further the discovery of 5 proteins with node degrees between 5-10 which have not been linked to cancer before (TUB, TEC tyrosine kinase, TCF14, Presenilin1 and Huntingtin). In Figure 5 the direct interactions of all 5 genes are depicted. Two of these, Presenilin and Huntingtin, are important genes involved in the neurodegenerative diseases Alzheimer's and Huntington's, respectively. This suggests a partial underlying commonality in the wiring of these rather different diseases and cancer. The significance of these findings, especially in regard to functionality of these proteins requires further investigation.

Table 5 Quantification of different hub sizes and the number of cancer genes present in different categories. 5 ND refers to hubs with 5-9 edges, 10 ND to 10-14, 15 ND to 15-19, 20 to 20-24, 25 to 25-49 and 50 to 50 and above edges.

Hubs present at different node degrees (ND)						
Number of Node Degree	5ND	10ND	15ND	20ND	25ND	50ND
Number of Hubs	506	251	163	113	77	59
Hubs not present in the cancer gene list of the Sanger institute	180	34	11	8	3	1

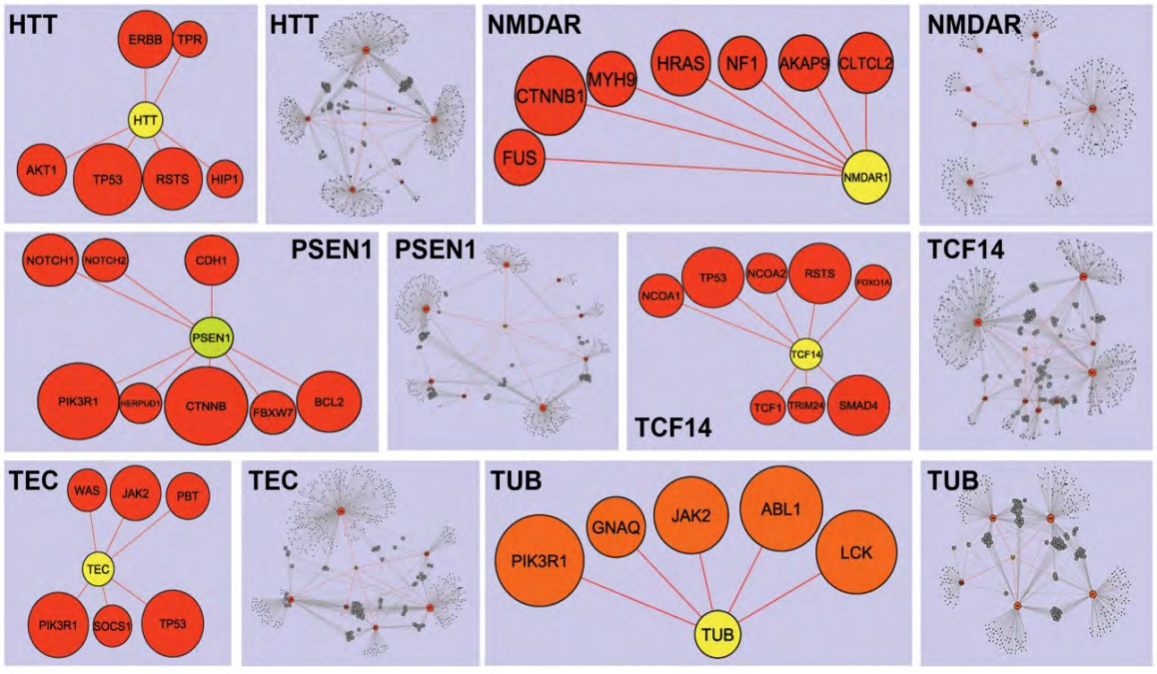

Fig. 5 Shown are the direct interactions of the 5 genes which we discovered here, which have more than 5 edges in the cancer network but have so far not been linked to cancer. In addition, we included the interactions of the NMDAR, a gene which we identified with above 5 interactions in our cancer network which just recently was linked to breast cancer by a single publication (North et al., 2010).

Acknowledgement

We would like to thank Rob de Jong, Wouter I. Bruins and Wouter II. Hollander for initial help with the collection of PIDs.

References

Andreeva A, Howorth D, Chandonia JM, et al. 2008. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Research, 36(Database issue): D419-25

Castrillon DH, Wasserman SA. 1994. diaphanous is required for cytokinesis in Drosophila and shares domains of similarity with the products of the limb deformity gene. Development, 120(12): 3367--3377

Chautard E, Thierry-Mieg N and Ricard-Blum S 2009. Interaction networks: from protein functions to drug discovery. A review. Pathologie Biologie (Paris), 57(4): 324-33

Cline MS, Smoot M, Cerami E, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. Nature Protocols, 2(10): 2366-2382

Cusick ME, Klitgord N, Vidal M, et al. 2005. Interactome: gateway into systems biology. Human Molecular Genetics, 14(Supppl.2): R171-181

De Las Rivas J, Fontanillo C. 2010. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. PLoS Computational Biology, 6(6): e1000807

Gough J, Karplus K, Hughey R, et al. 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of Molecular Biology, 313(4): 903-919

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell, 100(1): 57-70

Jonsson PF, Bates PA. 2006. Global topological features of cancer proteins in the human interactome. Bioinformatics, 22(18): 2291-2297

Keshava Prasad TS, Goel R, Kandasamy K, et al. 2009. Human Protein Reference Database--2009 update. Nucleic Acids Research, 37(Database issue): D767-772

Li L, Zhang K, Lee J, et al. 2009. Discovering cancer genes by integrating network and functional properties. BMC Medical Genomics, 2: 61

Mathivanan S, Periaswamy B, Gandhi TK, et al. 2006. An evaluation of human protein-protein interaction data in the public domain. BMC Bioinformatics, 7(Suppl 5): S19

North WG, Gao G, Memoli VA, et al. 2010. Breast cancer expresses functional NMDA receptors. Breast Cancer Research and Treatment, 122(2): 307-314

Pawson T. 2007. Dynamic control of signaling by modular adaptor proteins. Current Opinion in Cell Biology, 19(2): 112-116

Pietras K, Ostman A. 2010. Hallmarks of cancer: interactions with the tumor stroma. Experimental Cell Research, 316(8): 1324-1331

Taylor IW, Linding R, Warde-Farley D, et al. 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature Biotechnology, 27(2): 199-204

Tennant DA, Duran RV, Boulahbel H, et al. 2009. Metabolic transformation in cancer. Carcinogenesis, 30(8): 1269-1280