Drug-target network of taxanes revealed by data mining

Shao-Jun Chen*, Wei-Cai and Zhao-Hui Li
Department of Traditional Chinese Medicine, Zhejiang Pharmaceutical College, Ningbo, China

Abstract: Taxanes, mainly group paclitaxel and docetaxel, are amongst the most promising anticancer agents that are widely used for a variety of tumor types. It is a great challenge to gain a quick overview of the molecular mechanisms of taxanes, owning to the massive amounts of data have been produced. Network pharmacology will be a powerful tool to uncover the drug-targets network of taxanes. In this study, drug-targets network of paclitaxel and docetaxel were constructed via STITCH by database mining, and its topological parameters and important nodes were analyzed. All will provide a systematic understanding for molecular mechanisms of pacltaxel and docetaxel in a quick and visual way.

Keywords: Data mining, Docetaxel, Network pharmacology, Paclitaxel, Taxanes.

INTRODUCTION

Taxanes are amongst the most promising antitumor agents available at hand today, of increasing importance given that cancer is currently one of the world's major public health problems which need to be dealt urgently for the benefit of affected patients (Fauzee, 2011). In brief, the taxanes mainly group paclitaxel (Taxol) and Docetaxel (Taxotere) as well as taxanes homologs (Fauzee, 2011). Paclitaxel, identified in 1971, was originally isolated from the bark of the yew tree (*Taxusbrevifolia*) (Yared *et al*., 2012). Later, due to the initial limited availability of paclitaxel, a new taxane drug, docetaxel was discovered by a French group (Yared *et al.*, 2012, de Weger VA *et al.*, 2014).

Nowadays, paclitaxel and docetaxel are the effective drug for a broad range of human cancer, especially breast, ovarian, lung, prostate, gastric and head and neck cancer (Geney *et al*., 2005, Yared *et al*., 2012, de Weger VA *et al*., 2014). Their antitumor activity is primarily by stabilization of the microtubule dynamics, disruption of the cell cycle and thereby causing cancer cell death (Yared *et a*l., 2012, de Weger *et al*., 2014). During the past decades, these unique hydrophobic mitotic inhibitors have been thoroughly investigated through numerous experimental and clinical trials, which have brought over tens of thousands of publications in the Pubmed database (Fauzee, 2011). So, it is a big challenge to gain a systematic overview of the mechanisms from such massive amounts of data.

Network pharmacology, first proposed by Hopkins in 2007, is a new approach to drug design that encompasses systems biology, network analysis, connectivity, redundancy and pleiotropy (Hopkins, 2007 and 2008). It offers a good way to take many factors into consideration in drug discovery, including improving clinical efficacy, understanding side effects, evaluating toxicity and so on (Hopkins, 2007 and 2008). Therefore, a drug-target network approach of integrating information on a drug, gene, protein, cell and a disease can be a good solution for more efficient drug discovery, especially as more biological, chemical and interaction data are produced (Lee *et al*., 2009). Now, it is considered that systematic effects of a drug-target network will be eventually take into account in the drug design, which could pave the way to more specific drugs for diseases (Yildirim *et al.*, 2007). So, network pharmacology will be a powerful approach to understand the molecular mechanisms of paclitaxel and docetaxel.

The purpose of this study is to construct the drug-target network of paclitaxel and docetaxel by database mining. Next, the topological parameters of these networks were analyzed by network-related tools. We proposed to gain a quick and comprehensive understanding of molecular pharmacological mechanisms of taxanes.

METHODS AND MATERIALS

Network constructed by STITCH 4.0 database

Firstly, paclitaxel and docetaxel were entered into the STITCH 4.0 (accessed at http://stitch.embl.de) search panel respectively. The high confidence of the required confidence (score) was set to 0.700. The maximum number of interactions was set to 500. The other parameters were kept as default values. Then, the network files were downloaded.

Network parameters analyzed by network analyzer

The files obtained in the previous step were imported into Cytoscape soft, then into Network Analyzer and treated as undirected. The other parameters were set to default values. Then, a set of topological parameters (clustering coefficient, closeness centrality, between ness centrality, and topological coefficient etc) were calculated.

Top nods identified by STITCH

A confidence score, assigned by STITCH database, was

**Corresponding author:* e-mail: chenshaojun@hotmail.com

used to assess each protein-protein, chemical-protein, or chemical-chemical interaction (Kuhn *et al* 2008, 2010, 2012 and 2014) Paclitaxel and docetaxel were entered into the STITCH search panel individually, and the parameter settings were as describe above.

RESULTS

Network constructed via database mining

Networks can be manually constructed with the user assigning positions and node description characteristics, or could be formed by query public databases to assign further node descriptions (Pastrello *et al.*, 2013). As described in the methods section, we set a threshold of high confidence in the network to separate relevant interactions, and constructed the drug-target networks of taxanes querying from STITCH database on November 18, 2014. With this constraint, the paclitaxel network consists of 157 nodes and 752 edges, and the docetaxel has 76 nodes and 344 edges (fig. 1 and table 1).

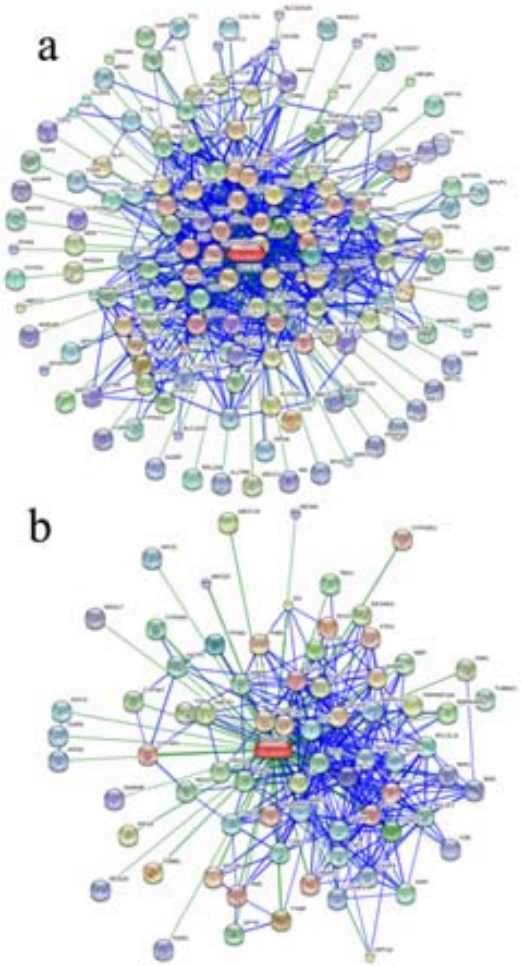

Fig. 1:.Network around paclitaxel and docetaxel. (a) paclitaxel; (b) docetaxel. Chemical is represented as pill-shaped nodes, while proteins are shown as spheres (Kuhn et al., 2008). Nodes that are associated to each other are linked by an edge.

Parameters analyzed by Network Analyzer

As listed in table 1 and shown in fig. 2, a group of topological parameters are provided by NetworkAnalyzer, including the degree distributions, average clustering coefficients, topological coefficients, and characteristic path lengths, network diameters, average number of neighbors, network density and so on.

Top nodes identified by STITICH

According confidence scores, EGFR, CYP2C8 and ABCB1, is the top 3 nodes in paclitaxel network, as well as CYP3A4, TYMP and ABCB1 in docetaxel network (table 2). And the two networks share many same nodes, such as EGFR, ABCB1, AKT1, BIRC5 and CAP3A4 (table 2). The top 10 nodes also formed a sub-network (fig. 3).

DISCUSSION

Target discovery is the most critical step in the modern drug development (Yang *et al.*, 2009). A drug that interact with multiple targets can form an interaction network, which may offer higher efficacy and limit drawbacks of a single-target drug (Anighoro *et al.*, 2014).With the explosion of biological data in the rapid development of omics era, network pharmacology will be the next paradigm in drug discovery (Hopkins, 2008, Yang *et al.*, 2009). In addition, many methods were created for its studies. For example, STITCH (search tool for interaction of chemicals) is a database for protein-chemical interactions (Kuhn *et al* 2008, 2010, 2012 and 2014). It contains 390 000 chemicals and 3.6 million proteins from 1133 organisms (Kuhn *et al.*, 2014). Recently, the database was used to search for the potential D-amino acid oxidase inhibitors (Zhao *et al.*, 2014). And in a systematic experimental research, the drug-target interactions from three kinase inhibitor studies were compared with the data in the ChEMBL and STITCH databases (Tang *et al.*, 2014). Another example, Cytoscape, a general-purpose and open-source software, can integrate, visualize and analyze biological networks (Shannon *et al*., 2003, Saito *et al.*, 2012). The Network Analyzer plugin, installed in Cytoscape by default, calculates and displays a set of topological parameters of biological networks such as the distribution of node degrees (Assenov *et al.*, 2008, Saito *et al.*, 2012). And it was fully utilized in a protocol for network topology analysis (Doncheva *et al.*, 2012). Therefore, these tools can be useful to understanding the drug-target networks of taxanes.

There, we constructed drug-target network of paclitaxel and docetaxel by query public STITCH database. A typical feature of scale-free networks is that most proteins connect to each with only a few interactions, while a few participate in dozens (Barabasi *et al.*, 2004). So, fig. 1 showed the drug-target networks of paclitaxel and docetaxel are also scale-free.

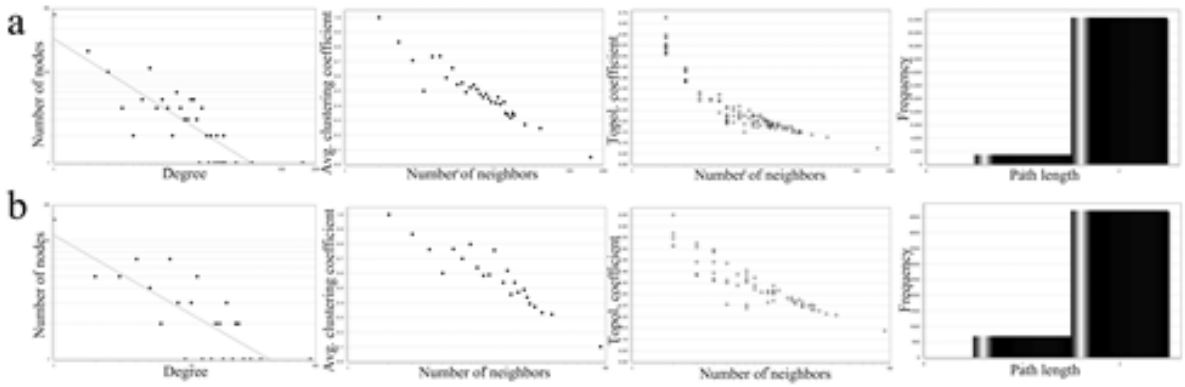

Fig. 2: The topological parameters of taxanes networks. (a) paclitaxel; (b) docetaxel.

Table 1: Topological parameters of the paclitaxel and docetaxel networks.

Parameters	Paclitaxel	Docetaxel
Number of nodes	157	76
Number of edges	752	344
Characteristic path length	1.937	1.873
Network heterogeneity	1.539	1.142
Network centralization	0.938	0.876
Avg. number of neighbors	9.580	9.053
Network density	0.061	0.121
Clustering coefficient	0.458	0.525
Network diameter	2	2

Table 2: The 10 top-ranked nodes in paclitaxeland docetaxel network.

Ranking	Paclitaxel	Docetaxel
1	EGFR	CYP3A4
2	CYP2C8	TYMP
3	ABCB1	ABCB1
4	AKT1	CASP3
5	CASP3	CYP3A5
6	BIRC5	CYP1B1
7	CYP3A4	BIRC5
8	CDK1	AKT1
9	BCL2L1	ABCG2
10	PARP1	EGFR

Network topology is a description of the arrangement of elements (nodes, interactions)in a network (Saito *et al.*, 2012). Degree (or connectivity), one of the most important feature of a network, denotes how many the nodes connect with others (Barabasi *et al.*, 2004). The degree distribution in fig. 2 illustrates its scale-free topology, which is consistent with the fig. 1. Another topological parameter, path length means the distance in networks, which tell us how long the expected distance between interconnecting nodes (Barabasi *et al.*, 2004, Chen, 2013). The average number of neighbors refers to the average connectivity of a node in the network (Chen, 2013). And network centralization indicates the concentration of communication in the network (Patterson *et al.*, 2013). As listed in table1, the paclitaxel network has the longer characteristic path length, the more network heterogeneity, the more network centralization, and the higher average number of neighbors. The average clustering coefficient characterizes the overall tendency of nodes to form clusters or groups, and network density shows how densely the network is populated with edges (Barabasi *et al.*, 2004, Chen, 2013). As shown in table 1, the docetaxel network has the more network density and the higher clustering coefficient. These computational analysis and interactive visualization will help us to gain insights into taxanes' biological processes (Doncheva *et al.*, 2012).

Hubs, nodes with a high degree, refer to some important proteins that can interact with a large number of nodes in a biological network (Doncheva NT, 2012, Patil A *et al*, 2010). They play the important roles in the network's

function and stability (Doncheva NT, 2012, Patil A *et al*, 2010). As shown in table 2, the top 10 nodes were identified (table 2), and formed a sub-network (fig. 3). The top three nodes in paclitaxel network corresponds to EGFR, CYP2C8 and ABCB1, whereas CYP3A4, TYMP and ABCB1 in docetaxel network.

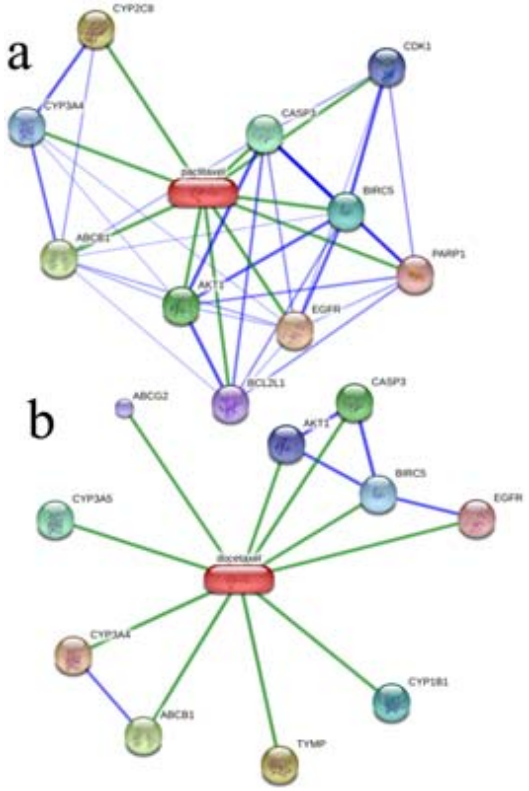

Fig. 3: Sub-network created by top 10-nodes. (a) paclitaxel; (b) docetaxel. Oblong: chemicals. Colorful spheres: target protein; Protein-protein interactions are shown in blue line, chemical-protein interactions are shown by green lines (Kuhn *et al*., 2008, Kuhn *et al*., 2010, Kuhn *et al*., 2012, Kuhn *et al*., 2014).

Cytochrome enzyme P450 is one of the most important nodes in the two networks (table 2). And, a review summarized that paclitaxel is primarily metabolized by the CYP2C8 and CYP3A, while docetaxel is only by CYP3A4 (de Weger *et al.*, 2014). It consists with the result of table 2. In addition, the human ABCB1 protein, a plasma membrane-bound glycoprotein, is involved in multidrug resistance and affects drug pharmacokinetics during the treatment of various diseases (Kimura *et al.*, 2007, Wolf *et al.*, 2011). Tulsyan *et al* reported that CYP and ABCB1 gene polymorphisms have an important influence on the efficacy of taxanes in breast cancer treatments (Tulsyan *et al.*, 2014). The epidermal growth factor receptor (EGFR), one of four homologous transmembrane proteins, regulates cell differentiation, proliferation and survival through binding its ligands, playing a vital role in the development of cancer (Jorissen *et al.*, 2003, Mendelsohn J, 2001). Paclitaxel can transiently transactivate EGFR, modulate the cell survival pathway, and induce cancer cell resistance, indicating a combination treatment may be better in the ovarian cancer management (Qiu *et al*., 2005). Thymidine phosphorylase (TYMP) can stimulate cancer growth, or enhance the antitumoral efficacy of chemotherapeutic agent, playing a dual role in tumor development and treatment(Bronckaers *et al.*, 2009). Puglisi reported that combination of low doses of capecitabine with weekly docetaxel is active against metastatic breast cancer, and TYMP expression may be a promising marker for medical benefit (Puglisi F *et al.*, 2008). Therefore, these hubs can serve as a biological marker in the paclitaxel and docetaxel network.

CONCLUSION

With the rising flood of biomedical data in an omics era, data mining approaches that discover previously unknown information from different written resources will play an increasingly significant part in targets selecting, interaction network constructing and cellular mechanisms understanding in the first stage of drug development (Yang *et al.*, 2009). Network visualization can be a powerful approach for data interpretation and analysis (Pastrello *et al.*, 2013). In this article, we constructed drug-target network of paclitaxel and docetaxel by data mining from STITCH 4.0 database. Then, we analyzed the network parameters by computational tool. Although the deficiencies in this article may include the false positive and data missing, the paper may provide a quick understanding the targets of taxanes, which may help a systematic understanding of their molecular mechanisms in a visual manner.

ACKNOWLEDGEMENTS

This work was supported by Zhejiang Provincial Natural Science Foundation of China (LY15H280009) and Zhejiang Pharmaceutical College (ZPCSR2015018).

REFERENCES

Anighoro A, Bajorath J and Rastelli G (2014). Polypharmacology: Challenges and opportunities in drug discovery. *J. Med. Chem.*, **57**(19): 7874-7887.

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T and Albrecht M (2008). Computing topological parameters of biological networks. Bioinformatics, **24**(2): 282-284.

Barabasi AL and Oltvai ZN (2004). Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**(2): 101-113.

Bronckaers A, Gago F, Balzarini J and Liekens S (2009). The dual role of thymidine phosphorylase in cancer development and chemotherapy. *Med. Res. Rev.*, **29**(6): 903-953.

Chen X (2013). Target network analysis of adiponectin, a multifaceted adipokine. *J. Cell. Biochem.*, **114**(5):

1145-1152.

de Weger VA, Beijnen JH and Schellens JH (2014). Cellular and clinical pharmacology of the taxanes docetaxel and paclitaxel a review. *Anticancer Drugs*, **25**(5): 488-494.

Doncheva NT, Assenov Y, Domingues FS and Albrecht M (2012). Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.*, **7**(4): 670-685.

Fauzee NJ (2011). Taxanes: Promising anti-cancer drugs. *Asian Pac. J. Cancer. Prev.*, **12**(4): 837-851.

Geney R, Chen J and Ojima I (2005). Recent advances in the new generation taxane anticancer agents. *Med. Chem.*, **1**(2): 125-139.

Hopkins AL (2007). Network pharmacology. *Nat. Biotechnol.*, **25**(10): 1110-1111.

Hopkins AL (2008). Network pharmacology: The next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**(11): 682-690.

Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW and Burgess AW (2003). Epidermal growth factor receptor: Mechanisms of activation and signalling. *Exp. Cell. Res.*, **284**(1): 31-53.

Kimura Y, Morita SY, Matsuo M and Ueda K (2007). Mechanism of multidrug recognition by MDR1/ABCB1. *Cancer Sci.*, **98**(9): 1303-1310.

KuhnM, SzklarczykD, FranceschiniA, CampillosM, von Mering C, Jensen LJ, Beyer A and Bork P (2010). STITCH 2: An interaction network database for small molecules and proteins. *Nucleic Acids. Res.*, **3**8 (Database issue), D552-556.

KuhnM, SzklarczykD, Franceschini A, von Mering C, Jensen LJ and Bork P (2012). STITCH 3: Zooming in on protein-chemical interactions. *Nucleic Acids. Res.*, **40** (Database issue): D876-880.

Kuhn M, Szklarczyk D, Pletscher-Frankild S, Blicher TH, von Mering C, Jensen L J and Bork P (2014). STITCH 4: Integration of protein-chemical interactions with user data. *Nucleic Acids Res.*, **42** (Database issue), D401-407.

KuhnM, von MeringC, Campillos M, Jensen LJ and Bork P(2008). STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, **36** (Database issue), D684-688.

Lee S, Park K and Kim D (2009). Building a drug-target network and its applications. *Expert Opin Drug Discov.*, **4**(11): 1177-1189.

Mendelsohn J (2001). The epidermal growth factor receptor as a target for cancer therapy. *Endocr. Relat. Cancer*, **8**(1): 3-9.

Pastrello C, Otasek D, Fortney K, Agapito G, Cannataro M, Shirdel E and Jurisica I (2013). Visual data mining of biological networks: One size does not fit all. *PLoS Comput. Biol.*, **9**(1): e1002833.

Patil A, Kinoshita K and Nakamura H (2010). Hub promiscuity in protein-protein interaction networks. *Int. J. Mol. Sci.*, **11**(4): 1930-1943.

PattersonPD, Pfeiffer AJ, Weaver MD, Krackhardt D, Arnold RM, YealyDM and Lave JR (2013). Network analysis of team communication in a busy emergency department. *BMC Health Serv. Res.*, **13**: 109.

Puglisi F, Cardellino GG, Crivellari D, Di Loreto C, Magri MD, Minisini AM, Mansutti M, Andreetta C, Russo S, Lombardi D, Perin T, Damante G and Veronesi A (2008). Thymidine phosphorylase expression is associated with time to progression in patients receiving low-dose, docetaxel-modulated capecitabine for metastatic breast cancer. *Ann. Oncol.*, **19**(9): 1541-1546.

Qiu L, Di W, Jiang Q, Scheffler E, Derby S, Yang J, Kouttab N, Wanebo H, Yan B and Wan Y (2005). Targeted inhibition of transient activation of the EGFR-mediated cell survival pathway enhances paclitaxel-induced ovarian cancer cell death. *Int. J. Oncol.*, **27**(5): 1441-1448.

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD and Ideker T (2012). A travel guide to Cytoscape plugins. *Nat. Methods*, **9**(11): 1069-1076.

ShannonP, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**(11): 2498-2504.

Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K and Aittokallio T (2014). Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis. *J. Chem. Inf Model.*, **54**(3): 735-743.

TulsyanS, Chaturvedi P, SinghAK, AgarwalG, Lal P, Agrawal S, Mittal RD and Mittal B (2014). Assessment of clinical outcomes in breast cancer patients treated with taxanes: Multi-analytical approach. *Gene*, **543**(1): 69-75.

Wolf SJ, Bachtiar M, Wang J, Sim TS, Chong SS and Lee CG (2011). An update on ABCB1 pharmacogenetics: Insights from a 3D model into the location and evolutionary conservation of residues coreesponding to SNPs associated with drug pharmcokinetics. *Pharmcogenomics J.*, **11**(5): 315-325.

Yang Y, Adelstein SJ and Kassis AI (2009). Target discovery from data mining approaches. *Drug Discov. Today*, **14**(3-4): 147-154.

Yared JA and Tkaczuk KH (2012). Update on taxane development: New analogs and new formulations. *Drug Des. Devel. Ther.*, **6**: 371-384.

Yildirim MA, Goh KI, Cusick ME, Barabasi AL and Vidal M (2007). Drug-target network. *Nat. Biotechnol.*, **25**(10): 1119-1126.

ZhaoM, Chang HT, ZhouQ, Zeng T, Shih CS, Liu ZP, Chen L and Wei DQ (2014). Predicting protein-ligand interactions based on chemical preference features with its application to new D-amino acid oxidase inhibitor discovery. *Curr. Pharm. Des.*, **20**(32): 5202-5211.