

# Systems multiple molecule drug design with less side-effects via drug data mining and genome-wide data identification: Drug design specifications approach

Bor-Sen Chen

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan

---

## RESEARCH

---

Please cite this paper as: Chen BS. Systems multiple molecule drug design with less side-effects via drug data mining and genome-wide data identification: Drug design specifications approach. AMJ 2018;11(6):361–369. <https://doi.org/10.21767/AMJ.2018.3439>

---

### Corresponding Author:

Bor-Sen Chen  
Department of Electrical Engineering  
National Tsing Hua University, Kuang Fu Road, No.101,  
Hsinchu, Taiwan  
Email: [bschen@ee.nthu.edu.tw](mailto:bschen@ee.nthu.edu.tw)

---

## ABSTRACT

---

### Background

Drugs fail in the clinic for two main reasons; one is that they do not work and another is that they are not safe. As such, two of the most important steps in developing new drugs should be drug targets identification and side-effect validation.

### Aims

The identification of drug targets and their restoration of cellular dysfunctions to normal cellular functions with less side-effects are considered as drug design specifications of systems medicine discovery. Since the effect on the normal expression of house-keeping genes and proteins is also considered as a restriction on drug design, the proposed multi-molecules drug strategy might be helpful for systems drug design with less-side effects.

### Methods

By systems biology method, genetic and epigenetic networks (GENs) are constructed to identify network

biomarker for drug targets of diseases by genome-wide high throughput data. An integration of computational network-based approach for multiple drug targets with drug data mining is also proposed for systems drug discovery with more precise medicine and less side-effects. Finally, some systematic drug design specifications for drug design are proposed to restore to the normal functions of multiple drug targets with less side-effects.

### Results

A systematic method is introduced to find multiple drug targets based on pathogenic mechanism investigated by network identification through genome-wide high-throughput data. Then a multi-molecule drug design strategy is also proposed to select a set of multi-molecule drugs with less side-effects via drug data mining method.

### Conclusion

Systematic engineering design methods seem applicable to systems drug discovery and design.

### Key Words

Systems drug design, drug targets identification, side-effect validation, multiple drug targets, multiple molecule drugs, pathogenic mechanism, drug data mining, drug design specifications

---

### What this study adds:

#### 1. What is known about this subject?

Drug side effects are unavoidable for the traditional single target and multiple drugs could be with less side-effects.

#### 2. What new information is offered in this study?

A drug design specifications approach is introduced for multiple drug design with less side effect via drug data mining method.

### 3. What are the implications for research, policy, or practice?

The design specifications in system engineering designs could be employed in system medical design to restore cellular functions with less side-effects.

---

#### Background

For almost a century, drug discovery was driven by the quest for magic bullets, which primarily followed the idea of “one drug- one target- one disease”<sup>1,2</sup> However, this concept is far from biological reality and even the most successful rationally designed drugs show a quite promiscuous binding behaviour, i.e., drugs rarely bind specifically to a single target and this fact challenges the concept of a magic bullet. Despite considerable progress in genome- and proteome-based high-throughput screening method and rational drug design, the number of successful single-target drugs did not increase appreciably during the past decade. Several highly efficient drugs, steroidal anti-inflammatory drug (NSAIDs), salicylate, metformin or gleevec, affect many targets simultaneously.<sup>1</sup> Furthermore, combinatory therapy, which represents another form of multiple molecule drugs, is used increasingly to treat many types of diseases, such as AIDS, cancer and atherosclerosis.<sup>2</sup> Thus, the use of multiple molecules is apparently an evolutionary success story of systems medicine design. Further, traditional medical treatments often use multiple components extracts of natural products and network model suggests that partial inhibition of a surprisingly small number of targets can be more efficient than the complete inhibition of a single target. Recently, more analyses of drugs and drug-target networks have indeed shown a rich pattern of interactions among drugs and their targets.<sup>2,3</sup> Obviously, drug side effects are unavoidable for the traditional single target – drug paradigm of magic bullets since drugs acting on a single target seem to be the exception.<sup>4,5</sup>

In general, drug side effects are very complex phenomenological observations in the treatment of human diseases. They have been attributed to a number of molecular scenarios, including the interaction with primary or additional targets, downstream pathway perturbations, kinetic and dosage effects, drug-drug interference, insufficient metabolization, effect of active metabolites, and aggregation or irreversible target binding of the drug.<sup>4</sup> Of these, direct interaction with proteins seems to be of the most important scenarios. Usually, these unexpected activities derived from off-targets are called drug side-effects, which are unwanted and harmful to patient.<sup>5</sup>

Further, wrong selection of molecular target implies the lack of expected efficacy, which is the most important cause of failure in clinical trials.<sup>5,6</sup> The lack of efficacy problem is even prominent when dealing with complex and multifactorial diseases. Consequently, the selection of right targets requires the complete understanding of entire molecular interconnected system where molecular targets play very specific roles on large and precise molecular machinery. Therefore we need many novel computational methods for drug target identification, including molecular networks to better represent the biological system to intervene in clinical trials.<sup>7</sup> Moreover, network-based representation enables the integration of multiple sources of information such as protein-protein interaction (PPI), target druggability assessment, gene-disease association, compound-protein interaction or protein side effects association that eventually resembles the reality in which the decision of molecular target is based on multiple distinct factors.<sup>8</sup> Frequently, the selected molecular target plays multiple cellular functions. Hence, the inhibition or activation of the drug target can lead to severe side effects that do not compensate the positive or negative ones.<sup>9</sup> A less harmful alternative consists in the specific regulation of molecular interaction that is associated to the treating diseases. Therefore, application of cellular network biology can help identify new PPIs and gene regulations amenable to be disrupted by small molecule treatment. Unfortunately, targeting PPIs or gene regulations is a very challenging task<sup>10</sup> and the progress is still limited to certain classes of PPIs or gene regulations.<sup>11</sup> Further, drug-target interactions are discussed in more detail from the perspective of cellular networking.<sup>10,12-14</sup>

Based on the above analyses, an integration of computational network-based approaches for multiple drug targets with drug data mining for multiple molecule drugs could not only reduce the time and expenses of pre-clinical stages, but also could lead to more precise medicines with less side-effects that will eventually be translated into lower attrition rates due to multiple drug targets and multiple molecule drugs. From the flowchart of systems drug discovery in Figure 1, the first step is to construct a candidate genetic and epigenetic network (GEN) for diseases by big database mining (if an infectious disease is considered, interspecies candidate cross-talk genetic and epigenetic network should be constructed). Since there are large amount of false positives in the GEN, the network identification method needs to be used to prune these false positives to obtain the real GENs of every stage of disease by parameter estimation scheme and system order detection method via genome-wide high throughput data of

different stages of diseases (i.e., microarray data, NGS data or protein PCR data<sup>15</sup>). Since GENs of diseases are still very complex, it is not easy to find drug targets for drug discovery from them. A principal network projection (PNP) method is employed to extract core GENs from GENs based on the principal component analysis (PCA) of GENs in different stages of diseases.<sup>16-19</sup> Then we project the core GENs to KEGG pathways (i.e., denotation based on KEGG pathways) to obtain the core pathways from which we could investigate the pathogenic mechanism based the changes of core pathways of different stages of disease. We could identify some significant biomarkers for multiple drug targets based on pathogenic mechanism. In order to efficiently design multiple molecule drugs with less side-effects to restore the cellular dysfunctions of these multiple drug targets to their normal cellular functions, those highly activated drug targets must be inhibited, those highly repressed drug targets must be activated, and those genes or proteins with housekeeping or without differential expression should be less influenced. These conditions and limitations on the drug design could be considered as drug design specifications of system multiple molecule drug design with less side-effects such as design specifications for systematic design to achieve multi-purposes in the system engineering design.<sup>20</sup> Connectivity Map (CMAP) and DGldb databases provide a large amount of compounds (drug molecules), among which some existing compounds have been approved by FDA and some potential compounds have not been approved by FDA yet. By drug data mining method and DGldb databases, we could find some compounds (drug molecules) to most meet these drug design specifications as the multiple molecule drugs for therapeutic treatment with less side-effects. Finally, some design examples of multiple molecule drugs with less side-effects are introduced to illustrate the proposed systematic design specifications scheme by restoring to the normal genetic expressions of these drug targets without influencing the expressions of housekeeping genes and proteins via drug data mining.

## Method

### Systems multiple drug design with less side effects: Drug design specification approach

From the systems biology perspective,<sup>15</sup> a disease is caused by a perturbation of genetic and epigenetic network (GEN). Obviously, the first thing of systems medicine is how to design multiple molecule drugs to remove the perturbation of GEN to restore to their normal cellular functions of perturbed GEN in disease. In general, the perturbed GEN is very complex, it is not easy to restore its whole cellular functions. In this situation, we could restore cellular dysfunctions to normal cellular functions of some significant

drug targets selected from network biomarkers based on the pathogenic molecule mechanism of a disease. In order to investigate pathogenetic molecule mechanism, we need to construct the GENs of different stages of disease by big data mining and system identification method via genome-wide high throughput data.<sup>13-16</sup> The systematic design precedence of multiple molecule drugs in Figure 1 will be discussed in detail in the following.

### Multiple drug targets identification via systems biology method

Network models suggest that partial inhibition and activation of a small number of drug targets can be more efficient than the complete inhibition of a single target because cellular networks are robust and could prevent major changes due to one target.<sup>21</sup> The combinational therapies of multiple molecules drugs could lead us to suggest that the systematic drug design strategies should be directed against multiple drug targets, and that this novel drug-design paradigm might often result in the development of more efficient molecules than the current favoured single-target drugs.<sup>22,23</sup> In general, the final effect of partial, but multiple, drug actions might often surpass that of a complete drug action at a single target. Therefore, the future success of this novel drug-design paradigm will depend not only on system model to identify the correct multiple drug targets, to find their multiple fitting, low-affinity drug molecules but also on more- efficiency *in vivo* testing.<sup>24,25</sup> A comparison of various strategies suggests that multiple but partial attacks on carefully selected targets are almost inevitably more efficient than the knockout of a single target though an equal and well selection.<sup>21,26-28</sup> A plausible explanation for this higher efficiency might be that even partial but multi-target attacks could block an increased number of individual interactions or network links in GENs. The reason underlying the attack efficiency of multiple target is easier to understand from the systematic point of view.<sup>29-31</sup> In general multiple target attacks are much better because they affect the GEN at more sites if they are distributed in the entire network.<sup>21,23,24</sup> In order to gain much better multiple target attacks, a structure-based and ligand-based drug design has been further discussed in.<sup>24,25,32</sup>

Based on the above analysis, the first step of the systems multiple drug design is to find a network biomarker as multiple drug targets. To identify a network biomarker for multiple drug targets, we need to investigate pathogenic mechanism by comparing the progression molecular mechanisms between connective stages of a disease. Before that, we need to identify GENs and further extract core

pathways by projecting core GENs to KEGG pathways. The steps for constructing GENs, core GENs and core pathways for pathogenic mechanisms of different stages of a disease by big data mining, system modelling, system identification, and principal network projection via high throughput data are shown in Figure 1. The procedure of constructing core GENs can be divided to the following four steps: (1) we employed big data mining and pre-processing of gene/miRNA/lncRNA expression data and DNA methylation data of patients (in infectious disease host/pathogen two-sided data are necessary). (2) we then constructed candidate GENs by using candidate PPI network and candidate gene/ miRNA/lncRNA regulatory networks (in infectious diseases, candidate interspecies cross-talk GEN is necessary); (3) we identified real GEN (or real interspecies GEN) of each stage of a disease by using the genome-wide high throughput data of the disease; (4) we then applied the principal network projection (PNP) method to extract the core nodes of GENs such as core proteins, genes, miRNAs, and lncRNAs to construct core GENs in different stages of a disease.<sup>16-19</sup>

By comparing the core GENs between two connective stages of a disease and projecting them to KEGG pathways to get core signalling pathways of the disease, we then could extract differential core signalling pathways between different stages to get insight into pathogenic mechanism of the disease. Based on pathogenic mechanism, we could select multiple biomarkers for multiple drug targets to design multiple molecule drugs by applying drug database mining to CMAP and DGIdb databases for therapeutic treatment of the disease.<sup>30,31</sup> CMAP provides the levels of 14,825 genes under 6,100 different conditions containing 1,327 different compounds (i.e., drugs) and different concentrations of these compounds.<sup>33</sup> The correlation coefficients between the gene expression levels and the concentrations of compounds denote the relationship between compounds and genes. If the correlation coefficient is greater than zero or some positive value, the gene is said to be up-regulated by applying the compound. If the correlation coefficient is less than zero or some negative value, the gene is said to be down-regulated by applying the compound. After applying correlation coefficient between drug concentration and mRNA activity in microarray data of CMAP, we then rank different compounds (drugs) based on the successful drug target cases and the number of unaffected housekeeping genes, which cannot be affected by the drug. How to design multiple molecule drugs for a disease with fewer side effects will be discussed in the following.

### **Multiple molecule drug design with less side-effects based on design specifications**

In the section, systems multiple molecule drug design for a disease with less side-effects will be discussed based on design specifications (i) the restoration of cellular dysfunctions to normal cellular functions of those selected multiple drug targets (ii) less side-effects on housekeeping and non-differential expression genes or proteins in the above section. The proposed drug design methods will be divided into two classes: one for chronic or aging-related diseases like cancer and another for infectious diseases.

In chronic diseases, we need to select multiple molecules to inhibit and activate the corresponding drug targets to restore their cellular dysfunctions to normal functions with less side effects on housekeeping cellular functions.<sup>26-28</sup> For infectious diseases, we need to select multiple molecules to destroy cellular functions of core toxic proteins of pathogen and restore the host cellular dysfunctions due to pathogenic toxicity to normal cellular functions without side-effects on host cells simultaneously. Therefore there are some different drug design methodologies in chronic diseases and infectious diseases. They will be discussed in detail as follows:

#### *a. Multiple molecule drug design with less side-effects in chronic or aging-related diseases*

In the chronic or aging-related diseases like diabetes, Alzheimer's disease or cancer,<sup>22,23</sup> we could extract core GENs of different stages of these chronic or aging-related disease and project these core GENs to KEGG pathways to investigate core pathways for investigating the pathogenic mechanism of this disease as shown in Figure 1. Then, based on the pathogenic mechanism, we could identify a set of multiple biomarkers (network biomarker) for drug targets, among which the higher expressions of some drug targets must be inhibited and some lower expressions of drug targets must be activated when compared with the expressions of neighbouring normal cells. Further, the expressions of housekeeping genes or proteins should not be influenced by the proposed multiple molecule drugs. The above three drug design specifications are introduced as the design objectives of systems multiple molecule drug design with less side-effects via system modelling, drug data mining and genome-wide high-throughput data identification.

For example, based on systems biology method via system identification and NGS data in Figure 1, we found the dysfunctions of three signalling pathways, i.e., SOP, TNF and ER signalling pathways play important roles in the early and late stages of bladder cancer cells.<sup>17</sup> In the early bladder

cancer, we found the following genes in bladder cancer cells are highly expressed than those of the neighbouring normal cells in the above three pathways<sup>17</sup>

*ARM1, COP5, PSMD8, SUMO2, CALR, PDIA3, DNAJB11, HSPA5, RPN1, CUL1, HSP90B1, KPNA2, PSMD12, ECT 2, TK1, TUBA1, HN1 and ENO1* (1)

The following genes in the early bladder cancer cells are lower expressed than those of neighbouring normal cells in the above three pathways

*UBC, JUN, RARRES3, and FOS* (2)

The following genes are nondifferentially expressed between the early bladder cancer cells and neighbouring normal cells

*BAG6, HUWE1, PAAF1, PSMD10, FAF 2, PCYT1A, and PAMD10* (3)

Based on the above analyses on the early bladder cancer, the dysfunctional expression genes in (1) and (2) can be considered as multiple drug targets and the genes with normal expressions in (3) could not be influenced by the multiple molecule drugs (i.e., the side-effects to be avoided). Therefore for the therapeutic treatment of the early bladder cancer, we need to select an adequate set of molecule drugs by drug data mining from databases CMAP and DGIdb to inhibit the highly expressed genes in (1) and activate the highly suppressed genes in (2) to restore to their normal expressions for proper cellular functions without influencing the expressions of non-differential genes (i.e., without side-effects) in (3). These design objectives and constraints could be considered as the drug design specifications of multiple drug design of early bladder cancer. CMAP contains the genome-wide microarray data in response to 1327 drugs in five cell lines, while DGIdb comprises a drug-gene interaction database. According to the above drug design specifications for multiple molecule drug design strategy, a set of potential multiple molecule drugs fulvestrant, estradiol, yohimbine and gefitinib are selected as a multiple drug combination from the proposed drug data mining method for treating the early stage of bladder cancer.<sup>14</sup> In the proposed combined multiple molecule drugs for the therapeutic treatment of the early stage bladder cancer, estradiol, gefitinib and fulvestrant are existing molecular drugs approved by FDA, and yohimbine is a new potential drug molecule, which has not been approved by FDA yet.

For the later stage bladder cancer, we found the following

genes are highly expressed than the neighbouring normal cells<sup>14</sup>

*ADRM1, COP5, PSMD8, SUMO2, RNF126, CALR, PDIA3, DNAJB11, HSPA5, RPN1, HSP90 AA1, HSPA1B, METTL23, RARRES3, KPNA2, PSMD12, ECT, JUM, TK1, TUBA1C, HN1 and EN 01* (4)

The following genes are suppressed in the later bladder cancer cells than the neighbouring normal cells<sup>14</sup>

*BCL3, FOS, UBC and GTF2A1* (5)

and the following genes have non-differential expression in the later bladder cancer cells than the neighboring normal cells.<sup>14</sup>

*BAG6, HUWE1, PAAF1, PSMD10, FAF 2, PCYT1A, and PSMD10* (6)

Then by drug data mining through CMAP database, based on the proposed drug design specifications for systems drug discovery of the later bladder cancer, a set of multiple molecule drugs comprising of gefitinib, estradiol, chlorpromazine and Ly294002 is selected to inhibit the highly expressed genes in (4) and to activate the highly repressed genes in (5) to restore the dysfunctions to normal functions of these three signaling pathways without influencing the expression of genes in (6) (i.e., to decrease the side-effects). In the proposed multiple molecule drugs for the therapeutic treatment of the later stage bladder cancer, gefitinib, estradiol and chlorpromazine have been approved by FDA, and Ly294002 is still a new drug molecule, which has not been proven by FDA yet.

#### *b. Multiple Molecule Drug Design with Less Side-Effects in Infectious Diseases*

*Clostridium difficile* (*C. difficile*) is the leading cause of nosocomial antibiotic-associated diarrhea, and the major etiologic agent of Pseudomembranous colitis. In severe cases, *C. difficile* infection (CDI) can cause toxic megacolon, intestinal perforation, and death. The intestinal epithelium is the first tissue encountered in the adhesion and colonization of *C. difficile*, and serves as a physical defense barrier against infection. Based on systems biology method via host/pathogen two-sided microarray data, system identification, and principal network projection method in Figure 1, we could obtain core cross-talk GENS of early and late CDI, which could be projected to KEGG pathways to find the core cross-talk pathways for infection and defense

mechanism of host and pathogen in the early and late CDI. From the infectious and defensive mechanism investigated by two core cross-talk pathways in the early and late stage of CDI, we found the following two cell wall proteins

*CD2787 and CD0237* (7)

play an important role in cell adhesion and pathogen defence mechanism; the following crucial proteins<sup>13</sup>

*CD1214, CD2629, and CD2643* (8)

are employed by *C. difficile* for sporulation.

Therefore, by drug data mining based on the proposed drug design specifications, we propose a potential multiple molecule drug, which contains E64, IgY, REP3123, camptothecin and apigenin for the therapeutic treatment of CDI owing to their abilities to inhibit the above targets proteins in (7) and (8) and to maintain the homeostasis of host dysfunctional proteins.<sup>16</sup> The combinations of camptothecin and apigenin can up-regulate the expression of dysfunctional proteins and down-regulate the inflammation- and apoptosis-related protein of host cells. This multiple-molecule drug could induce efficient prevention and elimination of *C. difficile* and remedial effects to restore gene expression homeostasis of host cells. The cysteine protease inhibitors E64 and CD0237-specific IgY can inhibit the activities of CD2787 and CD0237 in (7), thus interfering with cell adhesion and cell surface protein maturation.<sup>34</sup> In addition, the inhibition of CD2787 and CD0237 in (7) will limit the toxin production and the formation of biofilm, reducing not only the probability of cell adhesion but also the cytotoxicity of *C. difficile*. REP3123 can repress the spore formation and toxin production of *C. difficile*. The repressed toxin production could limit pathogenesis progression and the inhibition of sporulation could prevent spore-mediated re-infection.<sup>16</sup> Furthermore, the combination of human molecule drugs (camptothecin and epigenin) can promote the expression of dysfunctional proteins (RHOA, CDC42, RAC1, HSP90B1, HSPA5, and HSP90B2P) and repress inflammation-related proteins (NFkB1, REL, and IL-8) against the severe pathogenic effects induced by *C. difficile*. They could also provide potential antibiotic activity based on recent studies. The results in Li et al.<sup>16</sup> also suggest that CD2356, CD0171 and CD0179 could participate in the defence mechanisms of *C. difficile* against oxidative stress. The cooperation among these proteins provides a well-designed protection against human-produced ROS. The inhibition of these antioxidative proteins facilitates the host eliminating ability against pathogens,

and the scattered ROS can induce the rapid necrosis of pathogen cells. Therefore, CD2356, CD0171, CD1064 and CD0129 are recommended as potential drug targets for further drug design.<sup>16</sup>

## Discussion

In practical engineering designs,<sup>20</sup> we are beforehand with the design specifications including the desired targets, specific limitations and performance indices. For example, control engineers design a controller for a missile to attack an aircraft. First, engineers need to know the system model of the missile, which must be identified from flight experimental data. In general, the system model of missile is a very complex dynamic system of nonlinear and partial differential equations. For the convenience of control design, the complex dynamic model must be reduced to a simple significant model by model reduction method via principal component analysis (PCA) method.<sup>12</sup> In order to attack an aircraft, a robust controller must be designed for the missile to robustly track the desired target under some design specifications to tolerate uncertainties, modelling errors, external disturbance and some physical constraints. Based on the systematic design methodology of meeting the prescribed design specifications, the achievement in control system engineering designs is outstanding.<sup>20</sup> Even the biological systems are more complex than the engineering systems, the systematic design methodology may be useful for systems medicine design. Based on this systematic design methodology, we use big data mining method to construct candidate genetic and epigenetic network (GEN) and to prune the false positives of candidate GEP to obtain GENs of a disease by its genome-wide high-throughput data. Because the biological systems are more complex than the physical systems in engineering fields and cannot be modelled by conventional physical laws. The identification of GENs in different stages of a disease by genome-wide high throughput data is like the system identification for the missile model by experimental flight data. Complex physical systems (especially for nonlinear partial differential missile system) are not easily designed and should be truncated to more simple significant models so that a robust simple controller needs to be designed. Similarly, the GENs are still too complex to analyse for the pathogenic mechanism to investigate a set of multiple drug targets for multiple molecule drug design. Therefore the principal network projection (PNP) method needs to be employed to extract core GENs for investigating pathogenic mechanism of a disease to identify its multiple drug targets. Finally, a robust control based on simple significant system model is designed to achieve a desired target tracking under several design constraints and requirements (i.e., design

specifications). Similarly, the proposed systems drug design method needs some drug design specifications to achieve some desired therapeutic targets with less side-effects, i.e., to activate some targets, which are highly repressed in disease, to inhibit some targets, which are highly activated in disease, and do not influence some housekeeping genes or proteins. These prescribed therapeutic goals, constraints and limitations, which could restore cellular dysfunctions to normal cellular functions of drug targets with less side-effects, are called design specifications of systems medicine design.

Therefore, based on systematic design methodology in engineering, we proposed a systems drug design methodology in Figure 1. First, we construct the real GENs by big database mining method, system modelling and high throughput data identification. Then we extract the core GENs by principal network projection (PNP) not only to investigate pathogenic mechanism of a disease but also to identify the significant drug targets for multiple molecule drug design. The activation of these highly repressed targets, the inhibition of those highly activated targets and no influence on non-differential and housekeeping genes are considered as the design specifications of systems medicine design. Therefore, we employ systems biology and drug data mining method to select a set of multiple molecules to meet the above design specifications as potential multiple molecule drugs for *in vivo* testing of the treatment of a disease.

## Conclusion

In order to make drugs more workable and safe in the therapeutic process, a systems drug design method is proposed for multiple molecule drugs with less side-effects. Based on systematic control design methodology with robust control design purpose, to efficiently achieve desired therapeutic targets under several drug design specifications through a simple network model, simple core GENs are constructed through big database mining, system modelling and principal network projection method through genome-wide high throughput data. We then project core GENs to KEGG pathways to investigate the pathogenic mechanism for the identification of the network biomarker from which a set of multiple drug targets could be selected for the therapeutic treatment of the disease. The activation of highly repressed targets and the inhibition of highly activated targets to restore their cellular dysfunctions to normal cellular functions with less side-effects on the housekeeping and non-differential expression genes or proteins can be considered as the drug design specifications as the conventional systematic engineering designs. Then

we employ drug data mining method to select adequate multiple drug molecules from drug databases mining to meet the above drug design specifications as potential multiple molecule drug with less side-effects for the disease. Two design examples for chronic diseases and infectious diseases are also given to illustrate the proposed systems multiple molecule drug design method with less side-effects.

---

## References

1. Campillos M, Kuhn M, Gavin AC, et al. Drug target identification using side-effect similarity. *Science*. 2008;321(5886):263–266.
2. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One*. 2011;6(12):e28025.
3. Duran-Frigola M, Aloy P. Recycling side-effects into clinical markers for drug repositioning. *Genome Med*. 2012;4(3):1–4.
4. Hughes JP, Rees S, Kalindjian SB, et al. Principles of early drug discovery. *Br J Pharmacol*. 2011;162:1239–1249.
5. Sliwoski G, Kothiwale S, Meiler J, et al. Computational methods in drug discovery. *Pharmacol Rev*. 2014;66:334–395.
6. Hutchinson L, Kirk R. High drug attrition rates- where are we going wrong? *Nat Rev Clin Oncol*. 2011;8:189–190.
7. Csermely P, Korcsmacs T, Kiss HJM, et al. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther*. 2013;138:133–408.
8. Lindsay MA. Target discovery. *Nat Rev Drug Discov*. 2003;2:831–838.
9. Wang X, Thijssen B, Yu H. Target essentiality and certainty characterize drug side effects, *PLoS Comput Biol*. 2013;9:e1003119.
10. Arkin MR, Tang Y, Wells JA. Small molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol*. 2014;21:1102–1114.
11. Scott DE, Bayly AR, Abell C, et al. Small molecules, big targets: drug discovery faces the protein- protein interaction challenge. *Nat Rev Drug Discov*. 2016.
12. Wong YH, Chiun CC, Lin CL, et al. A new era for cancer target therapies: Applying systems biology and computer-aided drug design to cancer therapies. *Curr Pharm Biotechnol*. 2016;17(14):1246–1267.
13. Wong YH, Wu CC, Lai HY, et al. Identification of network-based biomarkers of Cardioembolic stroke using a systems biology approach with time series data. *BMC Syst Biol*. 2015;9:S4:1–21.
14. Chu LH, Chen BS. Construction of Cancer- perturbed protein-protein interaction network for discovery of apoptosis drug targets. *BMC Syst Biol*. 2008;2:56.

15. Chen BS, Li CW. Big mechanisms in systems biology: Big data mining, network modeling, and genome-wide data identification. Academic Press. Elsevier, London, UK 2017.
16. Li CW, Su MH, Chen BS. Investigation of the cross-talk mechanism Caco<sup>-2</sup> cells during *clostridium difficile* infection through genetic and epigenetic interspecies networks: Big data mining and genome wide identification. Front Immunol. 2017;8:901.
17. Li CW, Chen BS. Network biomarkers of bladder cancer based on a genome-wide genetic and epigenetic network derived from next generation sequencing data. Dis Markers. 2016;2016:4149608.
18. Wu CC, Chen BS. Key immune events of the Pathomechanisms of early Cardioembolic stroke: multi-database mining and systems biology approach. Int J Mol Sci. 2016;17(3):305.
19. Chen BS, Li CW. Constructing an integrated genetic and epigenetic cellular network for whole cellular mechanism using high-throughput next-generation sequencing data. BMC Syst Biol. 2016;10:1–18.
20. Chen BS, Wu CC. H<sup>∞</sup> robust design and their applications to control, signal processing, communication, systems and synthetic biology. Nova Science Pub Inc., New York, 2016.
21. Csermely P, Agoston V, Ponger S. The efficiency of multi-target drug: The network approach might help drug design, Trends Pharmacol Sci. 2005;26(4):178–182.
22. Chiang JH, Cheng WS, Hood L, et al. An epigenetic biomarker panel for glioblastoma multiform personalized medicine through DNA methylation analysis of human embryonic stem cell-like signature. J Integrative Biology. 2014;18(5):310–323.
23. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: Predictive, preventive, personalized and participatory, New Biochnology. 2012;29(6):613–624.
24. Wong YH, Lin CL, Chen TS, et al. Multiple target drug cocktail design for attacking the core network markers of your cancer using ligand-based and structure-based virtual screening methods. BMC Med Genomics. 2015;8(suppl.4):S4.
25. Wong YH, Chen RH, Chen BS. Core and specific network markers of carcinogenesis from multiple cancer samples. J Theor Biol. 2014;362:17–34.
26. Dickson M, Gagnon JP. Key factor in the rising cost of new drug discovery and development. Nat Rev Drug Discovery. 2004;3:417–429.
27. Gonem M. Predicting drug-target interaction from chemical and genomic kernels using Bayesian matrix factorization. Bioinformatics. 2012;28:2304–2310.
28. Ozturk H, Ozkirimli E, Ozgur A. A comparative study of SMILES-based compound similarity function for drug-target interaction prediction. BMC Bioinformatics. 2016;17:1–11.
29. Wong YH, Wu CC, Lai HY, et al. Identification of network-based biomarkers of Cardioembolic stroke using a systems biology approach with time series data. BMC Syst Biol. 2015;9(S6):54.
30. Li CW, Chang PY, Chen BS. Investigating the mechanism of hepatocellular carcinoma progression by constructing genetic and epigenetic networks using NGS data identification and big database mining method. Oncotarget. 2016;7(48):79453–79473.
31. Li CW, Chen BS. Genetic and epigenetic interspecies networks for cross-talk mechanisms and multi-molecule drug design in human macrophages and dendritic cells during MTB infection. Front Cell Infect Microbiol. 2016;6:124.1–25.
32. Wong YH, Li CW, Chen BS. Evolution of network biomarkers from early to late stage bladder samples. Biomed Res Int. 2014;2014:159078.
33. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene- expression signatures to connect small molecules, genes and diseases. Science. 2006;313(5795):1929–35.
34. Janoir C, Pechine S, Grosdidier C, et al. Cwp84, a surface-associated protein of *clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. J Bacteriol. 2007;189(20):7174–7180.

## ACKNOWLEDGEMENTS

The authors acknowledge the participation of all participants in the study.

## PEER REVIEW

Not commissioned. Externally peer reviewed.

## CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

## FUNDING

None



**Figure 1: The flowchart of systems drug discovery based on design specifications for multiple molecule drugs through big data mining, network identification, multiple drug target selection and drug data mining**

