INGOT: Towards Network-driven *In Silico* Combination Therapy

Sourav S Bhowmick, Huey-Eng Chua, Jie Zheng
School of Computer Engineering, Nanyang Technological University, Singapore
Email: assourav|chua0530|ZhengJie@ntu.edu.sg

***Abstract*—*Combination therapy*, where several drugs interact with multiple targets, holds tremendous promise for effective clinical outcomes in the management of chronic, complex diseases such as cancer. In this paper, we take a step towards this grand goal by laying out the vision of a novel *in silico*, data-driven *combination therapy* framework called INGOT for complex network diseases. Given the genomic and proteomic profiles of a patient population, it automatically predicts "optimal" set of *synergistic* drug combinations and corresponding dosages, which can potentially achieve the *therapeutic goal* while minimizing any *off-target* effects. Towards this goal, we present the architecture of INGOT and discuss various non-traditional design challenges and innovative features. Specifically, in INGOT, a disease-related *probabilistic signaling network* (PSN) is constructed by integrating publicly-available disease-specific signaling networks with expression data. Next, topology and dynamics of the PSN, which can be noisy and incomplete, are analyzed as a whole using probabilistic network analytics techniques to identify promising *target combinations* with desirable properties (*e.g.,* synergistic in nature, good efficacy and minimum off-target effect) to regulate the activities of key disease-related molecular players. Finally, *optimal* candidate drug combinations to modulate these targets are predicted by integrating and analyzing drug information (*e.g.,* DrugBank) with the target nodes. Successful realization of this framework can result in an effective platform for *in silico* screening of drug combinations in a rational way, by aiding early discovery of suitable combination therapy and guiding the design of further *in vitro* and *in vivo* experiments.**

I. Introduction

Cells use sophisticated communication between proteins in order to perform a variety of functions such as growth, survival, proliferation and development. As signaling proteins rarely operate in isolation through linear pathways, cell signaling can be viewed as a large and complex network. Specifically, the network view emerges due to 'cross-talks' between different signaling pathways. Such a network contains numerous features such as feedback and feedforward loops [2], which render it virtually impossible to manually comprehend how signals are integrated in these pathways. Understanding signal flow in the network is paramount as alterations of cellular signaling events, such as those that arise by gene mutations or epigenetic changes, can result in various diseases. For example, alterations to the genes that encode key signaling proteins, such as `RAS` and `PI3K`, are commonly observed in many types of cancers. In fact, although we have medically prevailed over numerous diseases that have plagued humanity throughout the ages, we are still not able to provide effective cures for diseases of greater complexity such as diabetes, cancer, heart disease, and Alzheimer's disease.

Therapeutic drug discovery that can target these altered signaling pathways to restore the physiological state of a disease network to normalcy has long been dominated by the "one-target one-drug" paradigm (identify a single chemical entity that binds to a single target) in the past decades. However, most complex diseased states are *polygenic* and are characterized by a combination of interacting genes and their products instead of a single gene. Although some degree of efficacy is possible using such a paradigm, it is well-known that such a paradigm did not yield an increasing number of successful drugs as expected [1] because multiple targets have to be involved in disease control due to redundancy and multi-functionality of biological processes. For example, cell proliferation leverages the combined control of multiple growth factor receptor pathways, and genetic experiments reveal that inhibition of any single receptor is only partially effective in blocking growth. Furthermore, clinical treatment of this paradigm may give rise to unexpected off-target effects due to cross-talks between pathways in the disease network.

To address the limitations of single-target-based drugs, increasing attention has been diverted to *combination therapy* by targeting multiple molecules simultaneously in a disease-related signaling network [24]. Specifically, in this therapy, instead of a single compound interacting with a single target, a concerted pharmacological intervention of several compounds interacting with multiple targets is made. Such a strategy has the potential to yield better benefits compared to a single molecule (mono-therapy) for complex diseases due to its ability to use lesser dosage to achieve efficacy, reduction of the frequency at which acquired drug resistance arises by combining drugs with minimal cross-resistance, and enhancing potency (amount required to produce an effect of given intensity) by leveraging *additivity* or *synergism* in the biochemical activities of multiple drugs. Examples of such a strategy can be found in the combination therapy of AIDS, cancer, and hypercholesterolaemia [24]. Even for diseases that are caused entirely by disruption of a single pathway, combination therapy might still offer benefits over monotherapy by virtue of spreading out the side effects to sub-toxic levels, while concentrating the desired effects on the target pathway. However, not all combination therapies produce better effects than monotherapies. For instance, in a study of combinations of analgesic drugs, some combinations (*e.g.*, `aspirin` and `pentazocine`) were beneficial, while others (*e.g.*, `acetaminophen` and `pentazocine`) were detrimental [22]. Hence, it is important to formulate strategies to develop good drug combinations which maximize the overall *therapeutic effect* while minimizing the *off-target effects*.

The traditional approach for combination therapy is gener-

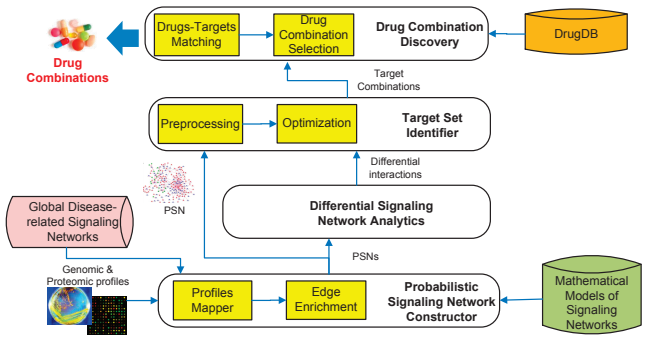

Fig. 1. System architecture of of INGOT.

ally based on designing combinations based on the clinical experience of doctors, knowledge of biological mechanisms, and practical constraints in the design of clinical trials [32]. Most therapies were initially developed as effective single agents and only later combined clinically. A common assumption in this case is that only drugs that are effective individually should be used as part of a drug combination. Because the effect of drugs depends on the dose, several doses need to be studied for the drug combinations, and the number of possible combinations can grow exponentially. For instance, a cancer chemotherapy regimen typically consists of six or more drugs from more than 100 anticancer drugs. However, investigating all six combinations out of 100 (including partial combinations) at three different doses to determine which combination is effective generates 8.9×10^{11} possibilities. As clinical trials are very expensive, exhaustive study of all possible combinations clearly becomes intractable. Although high-throughput screening technology allows the testing of pairs of drugs over a range of doses, combinatorial explosion still prevents exhaustive measurement of combinations of more than two drugs [37]. Hence, alternative procedures are needed to enable the rapid search for superior drug combinations targeting disease-related networks.

With the fast accumulation of experimental data from high-throughput screening and omic data measurements along with growing availability of disease networks, data-driven tools that can facilitate early detection of efficacious and nontoxic drug combinations *in silico* can serve as a powerful discovery and pre-screening platform when coupled with other complementary technologies such as high-throughput screening. Several groups have recently undertaken a data-driven *model-based approach* [3], [5], [35] where biological measurements (either obtained by the authors or derived from the literature) are used to build explicit models of the target biological network for the optimization of drug combinations using simulations. For example, Yang et al. [35] modeled the arachidonic acid metabolic network, related to inflammation, with a 2-step simulated annealing approach to optimize (1) drug candidates and (2) levels of intervention for multiple targets. On the other hand, advocates of *model-free* approach [6], [17], [34] have described biological search algorithms, where the search is not conducted *in silico* but directly using biological measurements from *in vitro* or *in vivo* systems. These methods can be stochastic, for example using Monte Carlo or evolutionary algorithms, or non-stochastic. For instance, Calzolari et al. [6] used non-stochastic algorithms derived from one used in digital decoding to perform biological searches for drug combinations, both *in vivo* (using fruit flies) and *in vitro* (using cell lines).

The model-based and model-free approaches, however, suffer from two key limitations. First, they ignore the off-target effects of the combinations. Consequently, these approaches may yield combinations satisfying the user-desired therapeutic effect, but with excessive off-target effects, rendering the combinations useless due to their toxicity. Note that in [35] a user needs to specify *apriori* specific side effects (as input to the algorithm) in terms of the ratio of concentration of two relevant nodes. Due to the complexity of biological networks, such a strategy is often impractical as it is highly unlikely for a user to know all *system-wide* side effects ahead of time. Second, although the target activity affects the combination effects, it is chosen randomly without considering *synergistic*[1] combinations. A judicious selection process that is "synergism-aware" can provide us an opportunity not only to select superior drug combinations but also improve efficiency of the overall process by ignoring non-synergistic combinations.

In this paper, we take a step towards addressing the aforementioned limitations by presenting a vision of a novel generic framework called INGOT (**I**n Silico **N**etwork-enabled Dru**G** C**O**mbination **T**herapy) for *in silico* combination therapy. Figure 1 shows the system architecture of INGOT. Given publicly available expression profiles data of patients for a specific disease (*e.g.,* MGH-Sanger Database (http://www.cancerrxgene.org/), Broad Institute Database (http://www.broadinstitute.org/ccle/home), dbDEPC), first, the *Probabilistic Signaling Network Constructor Module* constructs a *probabilistic signaling network* (PSN) by integrating publicly-available global human disease signaling network with these expression profile data. Specifically, since diseases like cancer implicate many signaling pathways, this module integrates these pathways into a large network. The *Differential Signaling Network Analytics* Module identifies important *differential* interactions in the PSN under the disease condition in comparison to the interactions under normal condition. Next, the topology and dynamics of the PSN, which can be noisy and uncertain, are analyzed by the *Target Set Identifier Module* in its entirety using novel network analytics techniques to identify "optimal" collection of potential molecules that can be targeted together (*target combinations*) by drugs. In the final step, the *Drug Combination Discovery Module* leverages publicly-available drug databases (*e.g.,* DrugBank [33]) to automatically identify *effective* drug combinations for these target. We elaborate on these components in the subsequent sections.

II. PSN Construction

In this section, we discuss how a disease-related *probabilistic signaling network* (PSN) can be constructed from genomic and proteomic expression data of patients. For ease of exposition, we use cancer patients as a running example to illustrate the PSN.

Genes that, when mutated or silenced, result in tumorigenesis often lead to the aberrant activation of certain downstream signaling nodes resulting in dysregulated growth, survival

[1] A joint action of two drugs in such a manner that one supplements or enhances the action of the other to produce an effect greater than that which may be obtained with either one of the drugs in equivalent quantity or produce effects that could not be obtained with any safe quantity of either drug, or both.

and/or differentiation. Hence, the architecture of cancer signaling network[2] is important for understanding the regions at which a genetic defect is involved in cancer and determining biological targets for diagnostic and therapeutics. Specifically, it presents a global picture of the mechanisms affecting cancer cell signaling and tumor progression. In INGOT we take the following two-step approach to create such a *probabilistic* signaling network.

Mapping Expression Profiles to Disease-related Signaling Network. First, we identify key molecular players (*e.g., mutated* and *methylated* genes) and their expression levels from the publicly available expression data of patients with a specific type of cancer. Second, given the set of molecular players and their expression values, we superpose them in a publicly-available global cancer-related signaling network (*e.g.,* [11]) to generate subnetworks containing nodes that are implicated in cancer as well as their neighborhood connections. Specifically, the expression data can be used to enrich the subnetworks by encoding the activity of relevant genes (proteins). For example, these expression data may enrich the p53 region of the network containing tumor suppressor proteins such as p53, Rb, BRCA1, BRCA2, etc. Note that due to the complexity of the cancer causing mechanism, the PSN tends to be large in size consisting of several subnetworks (*e.g.,* p53, MAPK-PI3K, RAS, TGF-β subnetworks) implicated in cancer.

It is worth noting that there are several recent efforts to identify relevant subnetworks in protein-protein interaction (PPI) networks that are differentially expressed in different types of cancer by leveraging high-throughput genomic and proteomic data [10], [27]. However, we construct a signaling network instead of a PPI network as the latter is insufficient to support development of superior data-driven techniques for drug combination discovery for the following reasons. First, edges in PPI networks are undirected; there is neither flow of information nor mass between nodes - an edge simply indicates that two proteins bind. Hence, they cannot be used to provide insights into the dynamics of the interacting molecular players as well as models of signal transduction, key prerequisites for inferring potential drug targets [3] (see Section IV). Second, PPIS have high false-positive rate, in the sense that although these proteins can truly physically bind they may never do so inside cells, because of different localization, or because they are never simultaneously expressed.

Enrichment of Edges with Reaction Dynamics. Signaling networks can be modeled at different levels of details ranging from detailed mathematical models to graphical representations. The PSN constructed in the preceding step only details topological structure of the network and roles of various molecular players in a disease state. It does not contain any mathematical models that can provide insights into the dynamics and functions of various signal transduction pathways in the PSN. It is increasingly acknowledged that existence of such models can greatly facilitate predictive analysis of combination therapy [3]. Hence, in this step we enrich the edges (reactions) of the PSN with mathematical models for quantitatively describing the types of reactions.

[2]A signaling network is represented as a hypergraph, in which nodes represent proteins. The edges are directed and represent *activation* (positive edge) or *inhibition* action (negative edge). Note that for some edges the types may be unknown.

Several mathematical models based on ordinary differential equations (ODE) have been formulated and their parameters optimized to fit experimental observations [3], [20]. However, formulating such detailed models is a difficult problem requiring a huge amount of experimental data, which are not commonly available, certainly not at a PSN-wide scale. Hence in INGOT, we need to devise automated techniques for extracting the mathematical models from heterogeneous information sources such as curated signaling network databases (*e.g.,* BioModel [26]) and biomedical literature.

It is worth mentioning that the enriched PSN resulting from the aforementioned steps is inherently *probabilistic* in nature due to the following key reasons. First, the PSN is incomplete and some edges may not be correct. Therefore, connectivity patterns in the PSN cannot be always interpreted as well-defined wiring schemes. Second, it is overly optimistic to assume the availability of mathematical models of *all* reactions in the PSN from heterogeneous information sources or experimental studies. Realistically, the ODES of several reactions are unavailable due to lack of experimental data. Hence only a subset of subnetworks in the PSN will be enriched with mathematical models. Even for those edges which are enriched with ODES, they may be impaired because of incomplete knowledge about the concentrations and kinetics of signaling intermediates. Additionally, while some subnetworks enriched with mathematical models may not be noticeably affected by the specific choice of *initial conditions* (*e.g.,* EGFR network [3]), it may be critical for subnetworks with complicated topologies as they may exhibit very complex responses such as ultrasensitivity, bistability and oscillations. Hence, any subsequent PSN processing techniques need to be probabilistic in nature to tackle these aforementioned challenges.

III. Differential Signaling Network Analytics

In this section, we present the functionality of the *Differential Signaling Network Analytics* Module. Notice that the above construction of a disease-related PSN by high-throughput mapping of expression profiles on a global disease-related signaling network can only represent a static "snapshot" of the network under a disease condition. It cannot distinguish between changes in network state and changes in network wiring in the disease PSN compared to the healthy PSN (*e.g.,* cancer vs normal). That is, it does not provide an answer to the following question: What interactions are most affected by the disease-state? Note that answers to this question shall pave way for identifying important *differential* interactions in the PSN which are a reflection of cellular processes that are differentially important under the disease condition. Specifically, some interactions may appear or disappear in the disease state, intensity of some interactions may alleviate or aggravate when in disease state compared to healthy condition, and others may remain strong irrespective of the state. Consequently, system-wide changes in signaling networks in different conditions will guide the subsequent targets identification as it shed light on which components need to be regulated in order to restore normalcy to the disease-related PSN.

Recently, there is a growing interest in studying the system-wide responses of interaction networks following environment or condition change [18]. One representative method for mapping the genetic interaction responses following environmental

changes is the dE-MAP approach [4]. In this method, two static gene interaction networks for each condition are first obtained using the *epistatic miniarray profile* (E-MAP) approach [29] which constructs a quantitative genetic interaction landscape of *S. cerevisiae*. Using these two static E-MAP networks, a *differential network* is then constructed that maps the interaction differences between the two static networks. However, analysis of such differential networks is largely done manually (*i.e.,* functional roles of different components of the network is done manually). Consequently, it does not scale to large networks such as cancer-related interaction networks.

We construct a *differential* PSN from the genetic interaction differences between cancer cells and normal cells. Note that the PSN of a normal cell can be constructed by following the similar technique as described in Section II. Then, large-scale novel analytics techniques need to be built on top of it to identify differentially important components under the disease condition. Results of this analysis will be fed to the *Target Combination Identification Module* to determine which molecular players/components need to be regulated.

IV. Target Combination Identification

Given a PSN and a *therapeutic goal* (*e.g.,* 50% inhibition of phosphorylated ERK), the goal of this component is to find suitable (top-k) sets of drug targets and the required *target activities* (type and extent of perturbations) for these targets. The intuition behind this step is to find target sets which can be modulated according to the required target activities by drugs[3] to restore normalcy to the biological network. Intuitively, the molecular targets can be identified by comparing the result of their perturbation in the disease-PSN with the PSN modeling the normal physiological state (Sections II and III).

The aforementioned problem is not only biologically challenging but also computationally. Complexity of the PSN (numerous potential drug targets) and potentially a wide range of target activities for each target make brute-force search to test all sets of potential target combinations infeasible since the number of testable combinations increases exponentially with the number of variables associated with the PSN. Furthermore, any target combination identification technique must generate target sets that are *synergistic* in nature and minimizes any *off-target effects* on the network. The *off-target effects* refer to systems level implications of targeting particular sets of nodes (*e.g.,* effect of activations of Akt and MEK on the PSN). Importantly, such off-target effects may have a deleterious impact on the physiological state of a patient. *Synergistic* target combinations require smaller target activities (drug dosage) to achieve the same therapeutic effect. This in turn, minimizes the potential off-target effects. Last but not the least, these challenges are further exacerbated by the noisiness and incompleteness of the PSN. Specifically, the therapeutic effect of a set of targets can be measured as the normalized change in area under the concentration-time series curves (temporal evolution of the ODEs) of relevant nodes before and after the target combination is perturbed. However, in order to quantify this accurately, the associated mathematical models need to be available, which may not be the case for several regions in the PSN as discussed earlier. Consequently, *probabilistic* target combinations identification approach that can operate in noisy and incomplete environment needs to be developed.

The *target combination identification problem* (TCIP) essentially consists of two subgoals: (a) determine the targets in the combination and (b) determine the type of action (*i.e.,* activation or inhibition) and the extent of the action on the targets. Both these subgoals must ensure that it achieves desired therapeutic effect while minimizing off-target effects. Hence, the problem can be modeled as optimization of a *constraint satisfaction problem* (CSP) which is NP-hard [13]. Due to the complexity of this problem, it is desirable to seek for an approximate solution.

Notably, there are early efforts in the systems biology community to address this problem in the context of well-studied signaling networks [9], [35]. However, these techniques cannot be easily adopted in the context of PSN as they assume that the underlying signaling network is complete and the mathematical model to describe it is completely available. Furthermore, [35] is not only "synergism-unaware" but also does not automatically consider off-target effects. Lastly, these techniques are computationally expensive as they are designed for a small subset of cellular reactions. That is, they do not scale to networks containing thousands of nodes (*e.g.,* PSN).

In order to tackle the issue of incompleteness and noisiness of mathematical models in a PSN, an intriguing possibility is to explore whether topology of the PSN themselves may provide valuable information in assessing targets and their combinations. That is, can we facilitate identification of target combinations by leveraging topological features without completely relying on the availability of complete set of mathematical models? This is more so as recent studies have strengthen the hypothesis that network topology is an essential feature in the emergent system function of the protein when it is perturbed [16]. For instance, *bridging nodes* (nodes with high *bridging centrality*) have been suggested as potential drug targets, although modulation of the bridging targets themselves may still be indirect [16]. An initial network analysis of the current drug targets of approved drugs indicated that drug targets are commonly highly connected but not essential nodes [25], [36]. Additionally, studies have shown that drug resistance is typically a result of the existence of redundant pathways in the network [31]. Hence, topological knowledge of whether a node is located in a redundant pathway in the PSN may increase the efficacy of designing target combination.

Approach. Broadly speaking, the TCIP can be tackled in two key steps. In the *preprocessing* step, we first compute the individual target activities required to achieve the therapeutic goal for *each* node by leveraging on the ODEs whenever available. The purpose of performing this step is to facilitate selection of only synergistic target activities for the next step to identify candidate target combinations. Techniques such as Monte Carlo Simulated Annealing (MCSA) can be leveraged here where the target activity is allowed to vary over a specific range during annealing. Observe that the target activities for some nodes need to be computed probabilistically as some of the edges may not be associated with any ODEs. Next, a *prioritization rank* for each node is computed by taking into account its topological properties in the PSN as well as its dynamic behavior. Specifically, properties such as bridging centrality,

[3] *Target activity*, quantified as the ratio of concentration of a drug to its dissociation constant, models the phenomenon of a drug at particular dosage hitting its target resulting in a particular response of the network.

location in redundant pathways, *target downstream effect*[4] [8], and *profile shape similarity*[5] [8] can be exploited to rank them based on their sensitivity to the therapeutic goal (nodes with high probabilities as drug targets are ranked higher). In the *optimization* step, the candidate target combinations can be first generated by applying a set of heuristics to reduce the exponential search space. For instance, targets that have higher prioritization ranks are preferentially selected as candidates (the ranks can be converted to selection probabilities by using a *rank-based fitness function*). Additionally, target activities of a combination are selected within a synergistic range. *Loewe additivity theory* [12] can be used to determine synergism by replacing drug dosages in its *combination index* with target activities. Next, the therapeutic effect and the off-target effects are computed for each candidate combination generated from the previous step by *perturbing* the PSN network. The perturbation needs to be probabilistically simulated leveraging on the ODEs whenever available. Specifically, a combination is accepted if it achieves the desired therapeutic goal and results in potentially fewer off-target effects. The selection is terminated once k "good" target combinations are identified.

V. Drug Combination Discovery

Given the target combination set, the objective of this step is to develop *in silico* techniques for discovering and characterizing "good" drug combinations for these targets. This problem is challenging in at least two fronts. First, drug promiscuity and drug-drug interaction effects make predicting efficacy and safety of drug combination difficult since off-targets of the drugs may be implicated in toxicity and interaction amongst drugs may result in changes in drug metabolism or uptake. Hence, relevant drug-related information needs to be carefully garnered and leveraged along with characteristics of the targets in the PSN to discover effective drug combinations. Second, as remarked earlier, there is an exponential number of possible combinations of drugs available for a disease treatment, making any brute-force approach prohibitively expensive. Hence, any *in silico* solution must not only be able to use drug-related information and incorporate semantic reasoning to map appropriate drug combinations (synergistic, potentiative, additive, antagonistic, reductive) to the target combinations by analyzing their effects, but also must be efficient and scalable in pruning exponential search space.

Approach. *Pharmacodynamics* (mechanism of action of a drug) and *pharmacokinetics* (fate of a drug in the body) properties of a drug play important roles in determining not only suitability of a drug for a target but also in determining appropriate drug combinations. However, recent *in silico* approaches [30], [35] for identifying drug combinations focus on analysis of the signalling networks via perturbation of the network parameters, ignoring drug-related information such as *drug-drug interaction* or *drug-target information* (*e.g.,* inhibitor potency of drugs, which molecules cause metabolism of the drugs, etc). In contrast, in INGOT we propose a framework that leverages such drug-related information for superior drug combinations prediction. Specifically, the *drug combination discovery* problem comprises of two key subgoals: (a) *identification* of drugs relevant to targets by analyzing characteristics of drugs and targets and (b) *selection* of "good" drug combinations for the target combinations.

Identification of drug-target relationship involves finding mapping between targets and *known* or *new* drugs. Specifically, for finding known drugs for targets, we may leverage on publicly-available drug-related databases (*e.g.,* DrugBank [33]). Intuitively, targets that do not have any matching drugs in the databases are first pruned and then various pharmacodynamics and pharmacokinetics properties (such as effect on targets, mechanism of action, absorption rate, kinetic constants, etc.) of selected drugs from these databases can be exploited to find suitable drugs. The latter case of identifying *new* drug-target relationship can be tackled by leveraging on approaches based on sequence similarity, 2D and 3D structural similarity [15], [19], [21], [23], side-effect similarity [7], text mining-based [14], or hybrid strategies that combine different drug-drug and target-target similarity measures [28].

The second subgoal to select "good" drug combinations for the target combinations from the selected drugs needs to address the following two important questions. First, if multiple drugs are mapped to a target, then which one do we choose? Second, how do we assess *in silico* if a particular drug combination is a good choice? We create a *combination rules* database that semi-automatically extracts various drug combination characteristics and rules from the literature by leveraging text mining techniques. For instance, since we would like to form drug combinations that have high efficacy and low toxicity, a possible strategy would be to choose drugs having higher potency. As drug information such as K_i and IC_{50} (inhibitor potency) are typically used to infer potency, rules such as smaller K_i and IC_{50} imply higher potency can be stored in the rule database. Next, this rulebase can be exploited by a *reasoner* to identify potential drug combinations. Specifically, the reasoner can use the rules to produce a list of logical statements which can be used to automatically determine drug combinations. Observe that such reasoning framework reduces the solution space considerably. Then the required dosages of these drugs (*i.e.,* dosage of drug required to produce the drug effect) can be estimated using global optimization techniques such as simulated annealing (similar to the TCIP). In the final step, we can evaluate the effect of the drug combinations *in silico* by incorporating kinetic parameters of the drugs (*e.g.,* dissociation rate, degradation rate) into the ODEs involving the targets and simulating the model at the system level for various drug inhibitory concentrations.

VI. Conclusions

The value of combination therapy, in general, has been suggested by numerous experimental studies, and its practicality has been demonstrated by the ubiquity of multicomponent drugs in the treatment of diseases like cancer, AIDS, etc. In this paper, we laid out the vision of an *in silico* combination therapy framework called INGOT, to improve the design and development of multicomponent drugs for patients suffering from complex diseases that implicate multiple molecular pathways, by aiding early discovery of optimal drug combinations and guiding the design of further *in vitro* and *in vivo* combination therapy experiments. Although increasing efforts are

[4] It assesses the effect of perturbing a target on the entire signalling network and can be computed as the sum of the product of the probability of perturbing a downstream node and the likelihood that the downstream node would cause an off-target effect.

[5] Identifies most relevant upstream regulators by assessing the similarity of the concentration-time series profiles of a target and its upstream regulators. Activators tend to have similar profiles to the target while inhibitors have inversely-similar profiles to the target.

under way to develop experimentally verified models of cell signaling, INGOT rejects the view that network-enabled *in silico* combination therapy is useful only when 'complete' mathematical models of cells or tissues are available. Specifically, a salient feature of INGOT is that it does not await the availability of complete models of signaling networks relevant to human disease that realistically capture, in mathematical form, actual cellular signaling events. Typically these mathematical models encompass only a small subset of cellular reactions in relatively simple biological settings. Hence, INGOT assumes that merging of these small models in a disease-related signaling network is inherently noisy and many regions of the network are starved of such mathematical models. Consequently, the proposed framework aims to leverage on probabilistic techniques towards predicting optimal drug combinations by accepting the noisiness and incompleteness of the PSN. We believe that whatever the limitations of current signaling networks, they will almost certainly be better guides for combination therapy than the prevailing practice based on clinical experiences.

Acknowledgment

The authors were supported by Singapore MOE AcRF Tier-1 Grant RGC 1/13.

References

[1] Food and Drug Administration. Innovation and Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products. *FDA White Paper* (http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.pdf), 2004.

[2] U. Alon. Network Motifs: Theory and Experimental Approaches. *Nat Rev Genet*, 8, 2007.

[3] R. P. Araujo, E. F. Petricoin, and L. A. Liotta. A Mathematical Model of Combination Therapy Using the EGFR Signaling Network. *BioSystems*, 80, 2005.

[4] S. Bandyopadhyay, M. Mehta, et al. Rewiring of Genetic Networks in Response to DNA Damage. *Science*, 330, 2010.

[5] D. Calzolari, G. Paternostro, et al. Selective Control of the Apoptosis Signaling Network in Heterogeneous Cell Populations. *PLoS ONE*, 2(6), 2007.

[6] D. Calzolari, S. Bruschi, et al. Search Algorithms as a Framework for the Optimization of Drug Combinations. *PLoS Comput Biol*, 4(12), 2008.

[7] M. Campillos, et al. Drug Target Identification Using Side-effect Similarity. *Science*, 321, 2008.

[8] H.-E Chua, S. S. Bhowmick, et al. PANI:A Novel Algorithm for Fast Discovery of Putative Target Nodes in Signaling Networks, In *ACM BCB*, 2011.

[9] H.-E Chua, S. S. Bhowmick, et al. STEROID: In Silico Heuristic Target Combination Identification for Disease-Related Signaling Networks. In *ACM BCB*, Oct 2012.

[10] H.-Y. Chuang et al. Network-based Classification of Breast Cancer Metastasis. *Molecular Systems Biology*, 3:140, 2007.

[11] Q. Cui, Y. Ma, et al. A Map of Human Cancer Signaling. *Molecular Systems Biology*, 3:152, 2007.

[12] J. B. Fitzgerald, et al. Systems Biology and Combination Therapy in the Quest for Clinical Efficacy. *Nat Chem Biol*, 9, 2006.

[13] E. C. Freuder, et al. Systematic versus Stochastic Constraint Satisfaction. *In IJCAI*, 1995.

[14] R. Frijters, et al. Literature Mining for the Discovery of Hidden Connections Between Drugs, Genes, and Diseases. *PLoS Comput Biol*, 6(9), 2010.

[15] P. Hawkins, et al. Conformer Generation with Omega: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, 50(4), 2010.

[16] W. C. Hwang, A. Zhang, and M. Ramanathan. Identification of Information Flow-Modulating Drug Targets: A Novel Bridging Paradigm for Drug Discovery. *Clin Pharmacol Ther*, vol. 84, 2008.

[17] S. Iadevaia, Y. Lu, et al. Identification of Optimal Drug Combinations Targeting Cellular Networks: Integrating Phospho-Proteomics and Computational Network Analysis. *Cancer Research*, 70(17), 2010.

[18] T. Ideker, N. J. Krogan. Differential Network Biology. *Molecular Systems Biology*, 8:565, 2012.

[19] M. Keiser, et al. Relating Protein Pharmacology by Ligand Chemistry. *Nature Biotechnology*, 25(2), 2007.

[20] B. N. Kholodenko. Cell-signaling Dynamics in Time and Space. *Nat. Rev. Mol. Cell. Bio.*, 7, 2006.

[21] M. Kuhn, et al. Stitch: Interaction Networks of Chemicals and Proteins. *Nucleic Acids Research*, 36:1, 2008.

[22] F. J. López-Muñoz, et al. Analgesic Effects of Combinations Containing Opioid Drugs with Either Aspirin or Acetaminophen in the Rat. *Drug Develop Res*, 29(4), Wiley Subscription Services, Inc., 1993.

[23] E. Lounkine, et al. Large-scale Prediction and Testing of Drug Activity on Side-effect Targets. *Nature*, 486(7403), 2012.

[24] M. S. Mitchell. Combinations of Anticancer Drugs and Immunotherapy. *Cancer Immunology, Immunotherapy,* 52, 2003.

[25] J. C. Nacher and J. M. Schwartz. A Global View of Drug-therapy Interactions. *BMC Pharmacol.*, 8(5), 2008.

[26] N. Le Novre et. al. BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems. *Nucleic Acids Research*, 34:D689-91, 2006.

[27] R. K. Nibbe, M. Koyuturk, and M. R. Chance. An Integrative-omics Approach to Identify Functional Sub-Networks in Human Colorectal Cancer. *PLOS Comp. Bio.*, 6:11, January 2010.

[28] L. Perlman, et al. Combining Drug and Gene Similarity Measures for Drug-target Elucidation. *Journal of Computational Biology*, 18(2), 2011.

[29] M. Schuldiner, S. R. Collins, et al. Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. *Cell*, 123, 2005.

[30] M. Schulz, B. Bakker, and E. Klipp. TIde: A Software for the Systematic Scanning of Drug Targets in Kinetic Network Models. *BMC Bioinformatics*, 10, 2009.

[31] S.V. Sharma and J. Settleman. Oncogene Addiction: Setting the Stage for Molecularly Targeted Cancer Therapy. *Genes & Development*, 21:3214-3231, 2007.

[32] R. A. Weinberg. The Biology of Cancer. *New York: Garland Science*, 2007.

[33] D. S. Wishart, C. Knox, A. C. Guo, et al. DrugBank: A Knowledge-base for Drugs, Drug Actions, and Drug Targets. *Nucleic Acids Research*, vol. 36, 2008.

[34] P. K. Wong, F. Yu, et al. Closed-loop Control of Cellular Functions using Combinatory Drugs Guided by a Stochastic Search Algorithm. *Proc Natl Acad Sci U S A*, 105(13), 2008.

[35] K. Yang, H. Bai, et al. Finding Multiple Target Optimal Intervention in Disease-related Molecular Network. *Mol Syst Biol*, vol. 4, 2008.

[36] M. A. Yildirim, K. I. Goh, et al. Drug-target Network. *Nature Biotech*, 25, 2007.

[37] G. R. Zimmermann, J. Lehar, and C.T. Keith. Multi-target Therapeutics: when the Whole is Greater than the Sum of the Parts. *Drug Discov Today*, 12(1-2), 2007.