Metabolically re-modeling the drug pipeline
Matthew A Oberhardt1,2,3, Keren Yizhak1,2,3 and Eytan Ruppin1,2

Costs for drug development have soared, exposing a clear need for new R&D strategies. Systems biology has meanwhile emerged as an attractive vehicle for integrating omics data and other post-genomic technologies into the drug pipeline. One of the emerging areas of computational systems biology is constraint-based modeling (CBM), which uses genome-scale metabolic models (GSMMs) as platforms for integrating and interpreting diverse omics datasets. Here we review current uses of GSMMs in drug discovery, focusing on prediction of novel drug targets and promising lead compounds. We then expand our discussion to prediction of toxicity and selectivity of potential drug targets. We discuss successes as well as limitations of GSMMs in these areas. Finally, we suggest new ways in which GSMMs may contribute to drug discovery, offering our vision of how GSMMs may re-model the drug pipeline in years to come.

Addresses
1 School of Computer Sciences, Tel Aviv University, Tel Aviv 69978, Israel
2 The Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel
3 These authors contributed equally to this study.

Introduction
Despite huge advances in biology in the last few decades, the resources pumped into drug development become ever larger and yet pipelines are not producing more drugs [1–3]. Industry analysts have blamed this in part on stagnation in R&D due to risk aversion in Big Pharma [4], and yet with strains of bacteria emerging that are resistant to nearly every existing antibiotic [5] and many of the most devastating diseases in the developed and developing world still lacking cures, new drugs are desperately needed. A major pitfall of traditional approaches (such as the ‘one drug, one target’ paradigm) is their inability to account for the full complexity of biological systems [6]. With many systems biology techniques now being developed specifically to navigate this complexity, a synthesis between the traditional drug pipeline and systems methods may soon become not an R&D luxury, but an industry necessity [4].

Systems biology addresses biological questions not from the perspective of individual protein or gene actors, but rather taking into account full interacting cellular networks. Modern omics technologies, as well as huge advances in computational capabilities, have made systems biology both appropriate and crucial for fully utilizing the resources for biological discovery that are available today. One of the most influential paradigms within computational systems biology is constraint-based modeling (CBM), which focuses on building and analyzing genome-scale metabolic models (GSMMs). CBM imposes context-specific constraints on a space of possible metabolic behaviors and allows prediction of numerous metabolic phenotypes, including growth rates, nutrient uptake rates, gene essentiality and more (Figure 1) [7]. The CBM field has been rapidly expanding, and today dozens of manually curated GSMMs are available in addition to over a hundred bacterial GSMMs built through an automated system called SEED [8]. In 2007, two genome-scale models of human metabolism were published [9,10], followed by a new version that has been published very recently [Recon2 [11] ]. While these models are generic and thus not specific to any mature cell-type or tissue-type, they have successfully served as a basis for context-specific models of human tissues [12–22]. GSMMs have been used in areas including metabolic flux analysis [23,24], studies of network evolution [25], and metabolic engineering [26,27]. These uses and many others have been thoroughly reviewed elsewhere [28,29].

The increasing availability of high-quality GSMMs presents new opportunities in drug development and design both for combating pathogens and for treating human disorders, which together form the focus of this review.

Determining the best drug targets
Traditional drug development left the determination of targets for a second step following the discovery of lead compounds with potential therapeutic efficacy (i.e. phenotypic screens), whereas the genomics era brought a surge of interest in the reverse ‘target-based’ approach, that is, determining good potential gene or protein targets first, and then screening for suitable small molecules to hit them [30]. Despite great interest in what seemed a more rational approach to drug development, target based screens have under-performed phenotypic screening in drug development [30,31]. One of the reasons cited for this failure is not target based screening itself, but the
2 New technologies

Inputs and outputs to GSMMs. A genome-scale metabolic model (GSMM) integrates information on compounds present in the growth medium, substrate uptake rates (if known), and information on the internal state of a cell (such as gene knockouts or large-scale omics datasets) to predict the steady state fluxes and activities of all reconstructed pathways in metabolism. Outputs of a GSMM analysis typically include growth yields (or growth rates if substrate uptake data is available), gene essentiality information, and prediction of wild-type or perturbed flux distributions across the whole metabolic network.

need to integrate new technologies in order to predict better targets [32]. There has been a major focus in the GSMM community on using models to predict drug targets, as GSMMs can directly predict the effects of perturbations such as gene knockouts on growth and thus lend themselves naturally to this task. Therefore, we focus here on the ways GSMMs have been used to predict novel drug targets. Drugs serve two major purposes: either to kill or to functionally alter a target tissue or cell. Here, we review GSMM methods for both of these aims.

Drugs to kill

Selective killing of certain target cells while sparing healthy human cells is the fundamental goal of many drugs. This goal is common, for example, in targeting both bacterial pathogens as well as cancer. About a third of current antibiotics target metabolic genes (derived from DrugBank: [33]), and metabolism comprises a large potential pool for more targets. Because of their basic structure and capabilities, GSMMs are a natural choice for rational drug target development.

GSMMs have been recognized for quite a while as vehicles to predict lethal metabolic drug targets in pathogens [34]. However, despite many proofs-of-principle such as accurate prediction of essential genes using flux balance analysis (FBA) (e.g. [35]) and ample predictions of possible targets in over a decade of use, these efforts have led to few promising new drug leads, and to our knowledge, no new drugs. Part of the reason for this is that many key factors essential in a good drug target have been out of the scope of GSMMs — for example, the ‘drugability’ of a target, which can involve its cellular localization, its three-dimensional structure, and its potential binding with known classes of inhibitors, among other factors [32,36].

More recently, merging druggability information with GSMM-based essentiality predictions has shown tremendous potential. A recent study aimed to repurpose already FDA-approved drugs to kill the deadly parasite, Leishmania major, combining GSMM predictions of gene essentiality and synthetic lethality (i.e. lethality of two genes in combination) along with ‘drugability’ scores and known
Drug target related predictions using GSMMs. A map of metabolism is shown, with descriptions of some of the highest confidence drug targets (or lead molecules) that have been identified in GSMM studies. Human (blue, solid border) and bacterial (orange with dashed border) targets are both shown.

chemical-protein interactions [37*]. This led to successful in vitro killing of *L. major* by Halofantrine, an antimalarial agent, as well as synergistic killing by four repurposed drug pairings (Figure 2). Such repurposing efforts, especially when aimed at neglected diseases (of which *L. major* infection is one), may have enormous impact.

Recent literature includes several other mergers of GSMM predictions with structural or druggability data in order to discover new drugs. In a study of *Escherichia coli* and *Staphylococcus aureus*, predicted enzyme targets in the FASII pathway were virtually screened for binding against a library of small molecules. Out of 41 predicted inhibitors, eight proved strongly and nine weakly active in extracted enzyme assays [38]. Another study of the same two organisms also used GSMMs alongside protein-molecule binding predictions to identify drugs leads that target the histidine biosynthesis pathway, which showed marked in vitro efficacy [39]. Some recent efforts have also brought the focus from genes to metabolites, analyzing 'choke points' (i.e., reactions uniquely consuming or producing a metabolite) [40,41] or metabolite essentiality [42,43]. Metabolite-centric methods have the advantage that small molecules can be sought based directly on their similarity to the essential metabolite, and will often have the potential to hit multiple enzymatic targets, a desirable drug attribute [6]. This line of reasoning led to identification of chorismate as a prime drug target for four prominent pathogens [44]. Chorismate is an important biochemical intermediate in plants and microorganisms and a precursor of metabolites such as aromatic amino acids, salicylic acid, and more. Importantly, there is no significant homology between genes that encode for reactions that utilize chorismate and the human genome, suggesting that blocking this compound might affect the pathogen and not the human host. A later study from the same group screened live cells of *Vibrio vulnificus* with 352 analogs of five promising target metabolites,
as identified through metabolite-centric GSMM methods. This screen yielded a small molecule inhibitor with bacteriocidal properties in the range of currently used antibiotics [45**].

These studies demonstrate how combining GSMM-based target predictions with information on druggability and enzyme/metabolite structural analysis can lead to novel drug candidates. These pipelines are new, but as they mature we may begin to see their effects in the drug industry. So far, attempts to use GSMMs to develop lethal drugs have focused mostly on microbial pathogens, but there is increased focus and interest in turning GSMM methods toward cancer. The power of GSMM in this arena was recently shown in a study of the role of fumarase, a TCA-cycle enzyme whose loss in the germ line can cause leiomyomatosis and renal carcinoma [46]. In the study, a kidney cancer GSMM was used to predict what metabolic mechanisms these cells might use to survive without fumarase. A novel pathway was predicted and then confirmed in vitro in immortalized kidney cancer cells. This in turn led to promising new drug targets in the haem oxygenase pathway that selectively killed cancer cells that were fumarase negative while sparing fumarase positive cells (Figure 2), a major step forward in selective targeting of this type of cancer.

Drugs to alter
While most drugs targeting microbes aim to kill the target cells, the majority of human-cell targeted drugs obviously aim not to kill the cells but rather to alter their function in a therapeutic manner. For example, 16 of the 20 most prescribed therapeutic drug classes in the US in 2011 aim to alter the functions of human cells (derived from: [47]). GSMMs are attractive tools for developing new drugs in this area, as they provide a large-scale, high confidence vehicle for predicting genome-wide effects of a perturbation. Hundreds of past studies have explored the effects of drugs or metabolic alterations (such as gene knockouts or changes in environmental conditions) on microbes, providing proofs of principle for many types of uses of GSMMs in this arena (see, e.g. [48]). In an outstanding example, a GSMM of E. coli was recently used to build a range of ensemble models that account for production of reactive oxygen species (ROS) [49]. Gene deletions were then predicted that increase endogenous production of ROS, and knockout of these genes in vitro was shown to increase susceptibility of E. coli to oxidants and to antibiotics. This study shows the power of GSMMs in developing adjuvant therapies to microbial pathogens, and more broadly, shows how prediction of metabolic alterations can suggest new targets for drugs.

Until very recently, usage of GSMMs to develop drugs that alter cells’ state has been limited because of the lack of a GSMM for human cells. However, with the recent introduction of human metabolic reconstructions [19] and a recent update [11*], methods to explore drug-related alterations in human metabolism are emerging. Generally speaking, many of the techniques mentioned in the To Kill section can also be applied to altering metabolism, but rather than inhibiting production of cellular biomass (and hence killing a cell), the goal is to inhibit the production of a metabolite known to have systemic impact on disease (e.g. cholesterol). Conversely, our lab recently developed a new GSMM method to address the problem of restoring diminished functions to cells. The method, which identifies gene targets most likely to globally transform a diseased metabolic state toward a healthy one, identified novel lifespan extending genes, which were then successfully experimentally validated in yeast. In a promising follow-up, the method identified the eicosanoids pathway, a known age related pathway, as a high-priority target for age-reversal in mammals (Yizhak et al., personal communication).

Despite these few examples, this area of drug development using GSMMs is still relatively underdeveloped. Potential future applications include developing adjuvant therapies (as mentioned above), reducing the metastatic potential of cancer, and even predicting dietary interventions that may augment or replace a certain drug [50].

Using GSMMs to predict drug toxicity and selectivity
Despite extensive advancements in drug discovery, drug toxicity and selectivity remain major drivers of failures and cost during the drug development process. These failures result from unexpected negative systems-level effects of hitting the appropriate target, as well as off-targeting in both primary and secondary tissues. Early pre-clinical elimination or modification of troublesome lead compounds is hence an important task that may significantly bolster drug industry R&D efficiency. GSMMs fit naturally into this scheme, as they are well suited to predicting system-wide effects of metabolic perturbations. Importantly, these effects can be evaluated at multiple levels that include both diseased and healthy tissues, and thus can give valuable insight into selective targeting.

The prediction of ligand binding sites using protein structures is a crucial step toward identifying drug side-effects. This type of structural analysis can identify as-yet uncharacterized drug targets that may participate directly in drug response. While ligand binding predictions go beyond GSMM-based methods, the integration of both approaches is complementary and hence of much value. Such a combined approach was recently applied [20] to investigate the hypertensive side effect of the cholesteryl ester transfer protein inhibitor torcetrapib in the context of human renal function. Focusing on 41 predicted off-targets of torcetrapib, the authors used a renal metabolic model to show that the inhibition of PTGIS precludes the
secretion of PG12. This metabolite has been linked to blood pressure and is thus expected to have a hyperten-
sion effect. As mentioned earlier, GSMMs have also been used alongside ligand binding models in the prediction of novel antimicrobial compounds [38,39], an approach that shows marked promise.

Many drugs work to inhibit a certain enzyme or function of the cell, but some drugs are rather designed to stimulate or trigger a response. Clearly, inhibitory and stimulatory responses should be modeled differently. Our lab has recently developed CBM-based method termed EDGE that is designed to address this challenge (Wagner et al., personal communication). Given a species metabolic model, EDGE works to systematically predict whether the over-activation of a metabolic enzyme is expected to be toxic to the organism or not. As an initial proof of concept, EDGE was shown to successfully detect genes whose expression is toxic in microorganisms, both in rich and in poor media. EDGE therefore may help predict unintended drug side effects in future studies in humans.

The capacity of a drug to produce the desired effect only at a selected target cell or tissue is another major chal-
lenge in drug design. This problem is encountered at various levels: first, antibiotics that are designed to kill only a certain spectrum of bacteria (bacteria versus bacteria); second, antibiotics that should spare the host human cells (bacteria versus human); and third, drugs that must target diseased cells while keeping healthy tissues in our body intact (human versus human). The recent development of GSMMs, both for hundreds of bacterial species and for numerous human tissues, provides for the first time an opportunity to address these issues. GSMMs have been used to target specific bacterial spectra [44], and reducing likely toxicity by filtering out gene or metabolite targets present in humans is standard practice in predicting novel antibacterial drug targets [45*]. Furthermore, GSMMs have been used to study host–pathogen interactions, resulting in a joint Mycobacterium tuberculosis — alveolar macrophage model [51] that displayed higher accuracy than the non-joined model in predicting gene essentiality. Although the tuberculosis study did not focus on drug development, the growth of M. tuberculosis is both so sluggish and so interconnected with its macrophage host that such models may be required in order to boost both effectiveness and selectivity of future drugs.

The issue of using GSMMs to develop selective drug targets has already been initially addressed in the context of cancer. An attempt to accomplish this goal was taken by Folger et al. [15] who built a metabolic model of cancer metabolism and identified drug targets that can potentially affect cell proliferation. In parallel, the effect of these targets on the production of ATP in a generic non-proliferating human cell was examined, resulting in a set of predictions that is highly enriched with FDA-approved metabolic anticancer drugs (Figure 2). As mentioned previously, a GSMM of the kidney cancer was used to determine new drug targets in the haem pathway that are selective for certain cancer cells, relying on utilizing synthetic lethality [46]. Finally, Facchetti et al. [52] developed a novel method for identifying drug targets by inhibiting a predefined objective function such as cell growth in one cell type while preserving the metabolism of another. This analysis correctly identified known che-
motherapeutic agents, lending a strong proof-of-principle to its use in novel drug discovery. Since these studies have been published, multiple GSMMs of human tissues have been reconstructed. Utilizing them, we can therefore expect a more thorough analysis of drug selectivity in a broad span of human metabolic disorders in the near future.

Future directions

GSMMs lend themselves naturally to the early stages of drug development, most notably the determination of new targets for target-based screens. However, as direct drivers of innovation and as scaffolds for interpretation of complex large-scale datasets, the potential of GSMMs in drug development is yet largely untapped. We put some current and projected future uses in Figure 3.

One of the great strengths of GSMMs is their ability to integrate high throughput data to, for example, model complex lifestyles of pathogens in vivo or in situ. As such, GSMMs have been used to explore the progression of a chronic cystic fibrosis lung infection [53], lifecycle stages of the parasites causing Chagas and Malaria [54,55], and features of other complex diseases [51,56]. Large efforts such as the ENCODE project [57] have recently revealed many new areas of biology, such as genome-wide human DNA methylations, copy number variations, and SNPs [58], which may also soon be integrated with GSMMs following along the lines of recent multi-system modeling efforts [59,60*]. GSMMs provide a solid framework for contextualizing such data, and may be integrated into future drug development pipelines in ways analogous to how physiologically based pharmacokinetic (PBPK) models already have been [61].

Additionally, GSMMs may be important for developing entirely new classes of drugs, such as probiotic treatments in which engineered bacteria outcompete pathogens in the human gut. Such ideas are not so outlandish considering, for example, the significantly higher cure rates of Clostridium difficile infections via fecal transplant (a probiotic treatment) than via antibiotics [62]. GSMMs are natural platforms for use in developing such probiotics, as they can integrate with 16s rRNA or metagenomic data to produce predictive community models [63,64], and can also aid in engineering of specific probiotic strains [65].
6 New technologies

As mentioned before, GSMMs can also be used to suggest new ways to repurpose existing drugs [37,66]. Drug repurposing sits far later in the drug development pipeline than target discovery, the prediction of toxicity, and most of the other applications previously discussed (see Figure 3), and thus represents a break from the emphasis on the very early stages of drug development.

Many other portions of drug development have not even seen a preliminary GSMM study. This is especially striking in phenotypic screening, as despite clear potential, GSMMs have never (to our knowledge) been used in either developing phenotypic screens or in interpreting the mechanisms of lead compounds from them, even in proof of principle. Concerns that arise later in the lifetimes of drugs, such as the common occurrence of resistance in bacteria and cancer, also have not yet been addressed using GSMMs, despite the demonstrated usefulness of GSMMs in examining important resistance factors such as gene promiscuity [67] and cellular evolution [68,69]. Finally, GSMM methods subtle enough to predict differential phenotypes of a tissue across many humans, a requirement for personalized medicine, are only just now being developed (Yizhak et al., personal communication). Such methods will also help in developing drugs that target certain tissues over others. GSMMs are obviously not magic bullets, but once they are noticed and adopted by the pharmaceutical industry, they will undoubtedly add value and insight that would be difficult to gain otherwise.

Acknowledgements
KY is partially supported by a fellowship from the Edmond J. Safra Bioinformatics Center at Tel-Aviv University and is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship. MO gratefully acknowledges support from the Whitaker Foundation and from the Dan David Foundation. ER’s research is supported by grants from the Israeli Science Foundation (ISF), the Microme project (EU FP7), and by the Israeli Centers of Research Excellance, Gene Regulation in Complex Human Disease Center (4/11).

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:
• of special interest
○ of outstanding interest


Metabolically re-modeling the drug pipeline

Oberhardt, Vizhak and Ruppin


This article presents Recon 2, a community-driven consensus GSSM, which is the most comprehensive, model-ready representation of human metabolism to date. Recon 2 is shown to outperform its predecessor Recon 1 in predicting metabolite biomarkers, and may serve as a scaffold for integrating diverse data types in various human diseases.


Chavali et al. present a computational pipeline for repurposing already FDA-approved drugs toward killing of the neglected tropical disease parasite, L. major. Genes predicted to be essential to L. major either alone or in combination are compared to target genes of existing drugs, and likely targets are then filtered based on several criteria such as druggability and toxicity. This led to successful in vitro killing of L. major by Halofantrine, an antimalarial agent, as well as synergistic killing by four repurposed drug pairings.


8 New technologies


This study presents a full pipeline for novel drug target development using GSSMs along with structural data. Briefly, a GSSM for V. vulnificus is developed and used to predict essential metabolites, that is, those whose removal fully obstruct biomass production. Next, the essential metabolites (and related enzymes) are filtered to remove those present in human metabolism. Finally, analogues of these metabolites were screened against live cells, yielding a high-priority lead compound with promising anti-bacterial properties.


This article presents a metabolic model of the kidney that emulates hereditary leiomysarcoma and renal-cell cancer (HLRCC), both of which are conditions caused by a germline mutation in the enzyme furamate hydratase (FH). The authors use the GSSM to identify an enzyme in the haem oxygenase pathway that kills FH-deficient cells while sparing healthy cells. This target was validated experimentally in vitro in immortalized kidney cancer cells.


This study presents an ensemble modeling approach that integrates ROS production into a GSSM of E. coli, following the observation that ROS production enhances killing by antibiotics. The model is then used to predict gene deletions that increase ROS production within cells. The deletions are then shown in vitro to increase susceptibility of E. coli to oxidants and antibiotics. This study represents an exemplary use of GSSMs to develop new adjuvant therapeutics to assist current antibiotics.


In this article, the authors present the first ever whole-cell computational model of a cell. The model captures the life cycle of the human pathogen Mycoblasma gerlmentum, including every known gene function as well as all known molecular components and interactions. The model was validated against a broad range of data, and provides insights into many cellular behaviors, including in vivo rates of protein-DNA association and an inverse relationship between the durations of DNA replication initiation and replication.


