毒性学

Percellomeトキシコゲノミクスの進捗

Progress in percellome toxicogenomics

Percellomeトキシコゲノミクスプロジェクトとは

2006年に本誌の当欄にて，毒性学の高精度解析手法として開始した“Percellomeトキシコゲノミクスプロジェクト”を紹介させていただいた[1]．当毒性部の基本姿勢は変わらず，さまざまな物質が身体に取り込まれた際に生じる可能性のある毒性(有害性)を予測し，それらの使用に際しての被害を未然に防ぐのが毒性学の役割であるとの考えに立脚し，身のまわりにあり，体のなかに入ってくるすべての“もの”について，どのような場合に(胎児・新生児・小児など，吸い込む・飲み込むなど)，どのくらいの量で，どのような症状が現れるか(急性毒性，発癌を含む慢性毒性，遅発性毒性など)について研究を継続している．

具体的には実験動物の診断所見をヒトに外挿すべく実施しているが，従来法では種差や個体差は“安全係数”により量的な安全マージンをとることで勘案されてきた．しかし，サリドマイド奇形に代表されるように，これには科学的な限界があり，“毒性学の近代化”が必要である．医薬品の場合はヒトで治験を行える場合があるが，それも胎児や新生児には実施困難であり，一般的な物質の毒性を検討することを考えると現状では動物実験は不可避である．そこで，著者らはヒトの身代りとしての実験動物(遺伝子改変動物の活用を含む)を対象とした，Percellomeトキシコゲノミクス研究を開始した次第である．

これは生体というブラックボックスの中身を遺伝子発現ネットワークの面から解明することにより，生体反応メカニズムに基づいた分子毒性学を構築することを目的としている．その際，毒性を見落とさない“網羅性”を確保する必要性から，全遺伝子のトランスクリプトーム情報のなかから生物学的に有意と判断される反応ネットワークを網羅的に抽出するアプローチをとっている．複数の実験から得られる大量のデータを蓄積し横断的な解析を加えることが必須であることから，マイクロアレイデータの標準化と互換性確保のために“細胞1個当りのmRNAコピー数”を得るPercellome法[2]を開発し，プロジェクトを軌道に乗せたところまでを前回の記事でご紹介した．

最近の展開

その後の数年間に，100種類超(医薬品，一般化学物質，食品関連物質を含む)の化学物質によるマウス肝の初期応答データを含む，延べ3.5億遺伝子情報からなるPercellomeデータベースを得た．これは，基本的に投与後の時間，曝露用量，遺伝子発現量の3軸からなる三次元曲面データにより構成される(図1)．解析には，この三次元曲面の特徴抽出という独創的な方法をとり，解析ソフトウェア群(相崎健一ら)は独自開発である．また，動物実験レベルからのシステム管理により，高精細かつ高再現性を実現している．

得られたデータの例としては，アリル炭化水素受容体(AhR)に結合するダイオキシン(2，3，7，8-TCDD)が比較的少数のAhR直下の遺伝子の発現を2時間目に誘導し，4，8，24と時間が経過する

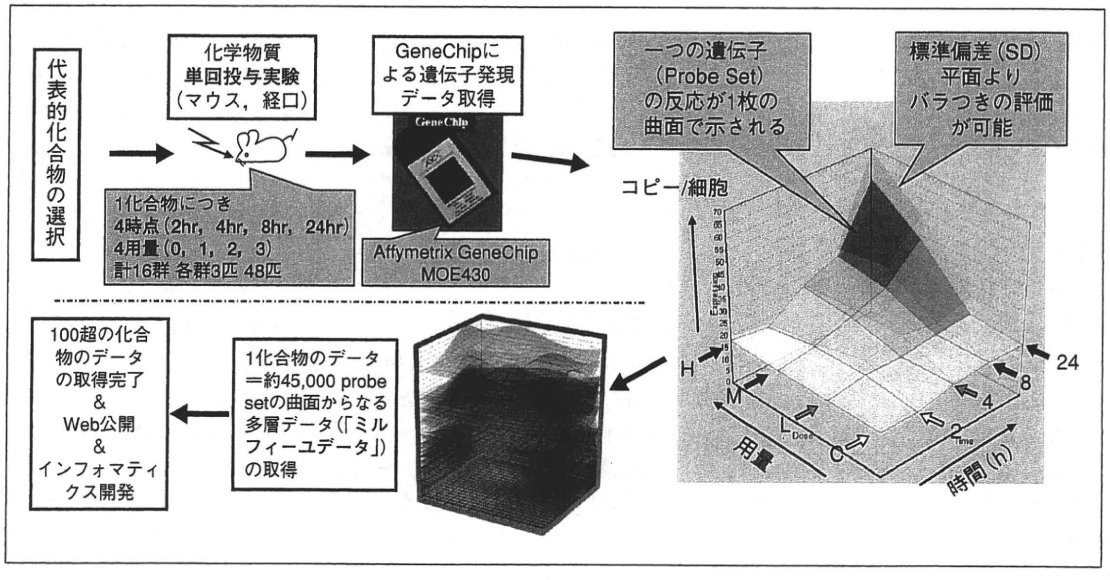


図1　Percellomeデータベースの概要

につれ数を増す状況が確認された．ダイオキシンの体内半減期が25時間であるにもかかわらず，2時間目のみの一過性発現のパターンをとるもの，持続的に発現が増加するものなどが観測されている．シックハウス症候群の指針値程度の，ごく低濃度域での吸入毒性トキシコゲノミクスも実施しており，ごく低濃度のホルマリン(0.1 ppm 付近)で肺の複数の遺伝子発現が明確に誘導されることをみている．サリドマイドは近年，癌治療薬として使用されていることから，複数の臓器における初期誘導を観測したところ，肺の2時間目に用量相関性をもって発現誘導のピークを示す遺伝子に，Cdkn1a (P21)が認められた．類似の発現パターンを示す初期応答遺伝子には，Fas，Foxo3a，Gata2など50あまりがあり，酸化的ストレスが誘発されることが推測された．実際，癌患者にサリドマイドが間質性肺炎を誘発する報告が増加しており，ヒトで確認された形となっている．

また，Percellomeトキシコゲノミクスを発生毒性へも適用している．妊娠マウスにサリドマイドを投与し胎児で発現変動が認められた遺伝子のなかに，マウス胚の肢部形成に重要な分子が見出され(その遺伝子をノックアウトしたマウス胚にアザラシ肢症に類似の奇形が生じる)，サリドマイド奇形の標的分子検索の糸口が示唆された．さらに，胎生期～幼若期の発達中の脳に対する神経シグナル撹乱が脳構造や神経回路の形成に影響を及ぼし，成熟後に行動異常などの脳高次機能の障害として顕在化することを見出している．これについては，妊娠マウスへ神経伝達物質類似物質を投与し，生まれたマウスに誘発される遅発性中枢毒性と海馬の遺伝子発現異常の関連解析から標的ネットワークが示唆されつつある．

このほかにも投与した化学物質に関して，いままで報告のないあらたな遺伝子発現変動現象を多数見出し，そのいくつかには特定の毒性との連鎖を示唆する分子生物学的情報がみつかっていることから，それらを順次報告および一般公開する準備を進めている(http://www.nihs.go.jp/tox/TTG_Archive.htm；現在更新中．2010年度中再開予定)．

プロジェクトの今後

さらに，マイクロアレイのクロスハイブリダイゼーションを修正するアルゴリズムの開発を終え(特許出願準備中)，その実装準備中である(NTTデータおよび日本テラデータとの委託共同研究)．また，遺伝子ネットワークと毒性の動的な因果関係を導き出すインフォマティクスの構築研究やPercellomeデータの統合的提示方法の開発にも本格的に取り組んでおり(ソニーコンピュータサイエンス研究所との共同研究)，段階的に皆様にご披露できる予定である(厚生労働科学研究費補助金，環境研究総合推進費などによる)．

1) 菅野　純：毒性の高精細解析に向けてのトキシコゲノミクス．医学のあゆみ，**218**：1035-1036，2006.
2) Kanno, J. et al.："Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. *BMC Genomics*, **7**：64, 2006.

菅野　純／Jun KANNO
国立医薬品食品衛生研究所
安全性生物試験研究センター毒性部

循環器内科学

心筋トロポニンの高感度測定の有用性

Clinical utility of high-sensitivity cardiac troponin assay

従来の心筋トロポニン測定は検出感度が低いため，急性冠症候群の診療以外で用いられることはまれであった．最近，検出感度が5倍以上改善された高感度測定が臨床の場に登場した．この高感度測定は，従来測定では検出不可能であった小さな心筋障害を診断できる．そのため，超急性期の心筋梗塞診断の精度[1,2]や慢性心不全における予後予測の精度[3]を高めることが示されている．さらに，外来診療や検診・人間ドック分野へのあらたな展開も期待される．

急性冠症候群の診療

トロポニンが上昇している不安定狭心症は，突然死や急性心筋梗塞発症の危険度が高い．このトロポニンの上昇は，破砕したプラークや血栓が引き起した末梢の微小塞栓による微小心筋障害を反映している．そのため，2000年に公表されたヨーロッパ心臓病学会/アメリカ心臓病学会(ESC/ACC)の心筋梗塞の再定義[4]は，トロポニンが上昇している不安定狭心症を急性心筋梗塞に包括した．さらに，ヨーロッパ心臓病学会/アメリカ心臓病学会/アメリカ心臓協会/世界心臓協会(ESC/ACC/AHA/WHF)の共同タスクフォースは，2007年に急性心筋梗塞の診断基準の再改定[5]を公表した．新しい診断基準では，トロポニンの心筋梗塞診断における基準値を健常人の99thパーセンタイル値より大と定めた．一般に，測定値の相対的なばらつき(変動係数．coefficient of variation：CV)が小さいほど測定値の精度は高い．共同タスクフォースは試薬の精度にも言及しており，健常人の99thパーセンタイル値における変動係数が10%以下である試薬を用いることを推奨した．従来の試薬はこの条件を

APPLICATIONS NOTE

Vol. 26 no. 10 2010, pages 1381–1383
doi:10.1093/bioinformatics/btq143

Systems biology

Advance Access publication April 5, 2010

Payao: a community platform for SBML pathway model curation

Yukiko Matsuoka[1,2], Samik Ghosh[1], Norihiro Kikuchi[3] and Hiroaki Kitano[1,4,5,*]

[1]The Systems Biology Institute, Tokyo, [2]JST ERATO Kawaoka Infection-induced Host-response Network Project, Tokyo, [3]Mitsui Knowledge Industry Co. Ltd, Tokyo, [4]Okinawa Institute of Science and Technology, Okinawa and [5]Sony Computer Science Laboratories, Tokyo, Japan

Associate Editor: Jonathan Wren

ABSTRACT

Summary: Payao is a community-based, collaborative web service platform for gene-regulatory and biochemical pathway model curation. The system combines Web 2.0 technologies and online model visualization functions to enable a collaborative community to annotate and curate biological models. Payao reads the models in Systems Biology Markup Language format, displays them with CellDesigner, a process diagram editor, which complies with the Systems Biology Graphical Notation, and provides an interface for model enrichment (adding tags and comments to the models) for the access-controlled community members.

Availability and implementation: Freely available for model curation service at http://www.payaologue.org. Web site implemented in Seaser Framework 2.0 with S2Flex2, MySQL 5.0 and Tomcat 5.5, with all major browsers supported.

Contact: kitano@sbi.jp

Received on January 21, 2010; revised on March 12, 2010; accepted on March 30, 2010

1 INTRODUCTION

Creating an extensive model of gene-regulatory and biochemical networks with the latest data is a painstaking task. Curation is essential to creating an accurate model. Yet as science and technology advances rapidly, once curated models soon become out-of-date and need to be revised constantly. Many pathways and networks are now available online via pathway databases, such as Reactome, BioModels.net, Panther Pathways and many pathway editors are available (Bauer-Mehren *et al.*, 2009). What is needed is a framework to facilitate tracking and update mechanism for modelers and researchers in the community to contribute to the collaborative model building and curation process.

WikiPathways (Pico *et al.*, 2008) is an effort for such a collaborative platform in the Wiki style. While the Wiki system has its strength in collaborative editing and version tracking, it does not provide access control or explicit community tagging mechanisms. In a community-driven model enrichment environment, it is effective to differentiate privileges to special interest group (SIG) members for curation activities—commenting on existing tags, adding tags to models, annotating individual component inside a model and validating the annotations. In view of the complexity of biological pathways and the expertise of biologists in different areas, a community platform for biology requires an exquisite balance of federated resource sharing and quality control of information by a SIG of experts in the particular pathway or process. An access control privilege system allows the community to share and disseminate the knowledge, while enabling a dedicated SIG to maintain high-quality, curated information.

To provide such a curation framework, we have developed a system called 'Payao'. The system is named after a fish aggregating device, an artificial floating raft where fish congregate and popular in Okinawa/Philippine area. Payao aims to become a biological knowledge aggregating system, which enable a community to work on the same models simultaneously, insert tags as pop-up balloon to the parts of the model, exchange comments, record the discussions and eventually update the models accurately and concurrently.

The current workflow for pathway curation has two phases working in a cyclical manner, as shown in Figure 1: pathway editing using biological pathway editors (CellDesigner) and community-driven pathway enrichment and knowledge sharing. Payao serves for enrichment phase of the curation. Payao is a web-based platform, providing an interface for adding tags and comments to the components (such as Species, Reactions and specified area) of the model, as well as community management functionality. The information on the users and tag data is stored in a relational database (RDBMS) on the server. Payao adopts community standards, accepting Systems Biology Markup Language (SBML; Hucka *et al.*, 2003) format models and displays them in Systems Biology Graphical Notation (Le Novere *et al.*, 2009) compliant CellDesigner (Funahashi *et al.*, 2008) graphical notation. Curation data on Payao can be easily reintegrated into the original model via CellDesigner.

Fig. 1. Workflow of community pathway curation.

*To whom correspondence should be addressed.

2 PLATFORM

Payao consists of the server application, client user interface and database. The server application has been implemented in Java on Seaser Framework 2.0 with S2Flex2. Tomcat 5.5 was used as the servlet container to build the web application. The client user interface has been implemented in ActionScript on Flex framework 2.0.1, which allows us to build a Rich Internet Application to visualize SBML models. The server communicates with the client via the Action Message Format (AMF3) protocol on S2Flex2, which enables us to translate between Java Objects in Seaser and Action Script Object in Flex. The server can handle CellDesigner models including SBML models and the visual information using CellDesigner API ver.4.0. It parses SBML files sent from client to create CellDesigner models, and provides the information as CellDesigner Plugin classes. The client application receives the model information and draws the model. MySQL 5.0 is used to store information on user, model and tags/comments in the database.

3 FEATURES

As Payao accepts pathway models stored in SBML format and uses CellDesigner APIs for visualization, the most suitable SBML editor for Payao is CellDesigner. In SBML format, models can capture details of biochemical process descriptions, not only protein–protein interactions. Adopting SBML format enables the models to be easily used as the base of computational data analysis or simulation of dynamic behaviors. The Payao platform enriches the model curation process by providing a host of features for user management, tagging and model updates [detailed are available for reference in (Payao User guide, 2009)].

3.1 Community management

Forming a community is an important step for curation. Different expertise groups can contribute variety of information to the model. As web-based Payao can be accessible from all physical locations, it enables experts across the world to communicate in a collaborative curation effort.

Community is formed around a pathway model. It is the model owner who sets access control over the registered model. In the Payao system, access controls can be set by specifying the privileges to individuals as well as to user categories, such as guest, login user and model user (who are invited to access the model by the model owner). This enables a user to stage the curation process, initiate the curation within a small group (e.g. SIG) and then switch the access control of the model for public viewing.

3.2 Model management

The model owner registers and manages the SBML model. Upon registration, the model owner specifies the basic model information including, thumbnail image, references and copyright. The owner sets the access privilege to the user in three levels (browsing, adding tags, adding comments) by user categories or by individual users. The registered models can be sorted by Register Date and by Popularity. Popularity is measured by the activity level (number of tags and comments) and ranked in the list. All the registered models are listed with the thumbnails in the top screen in the right panel (Fig. 2). Registered models are stored in the database.

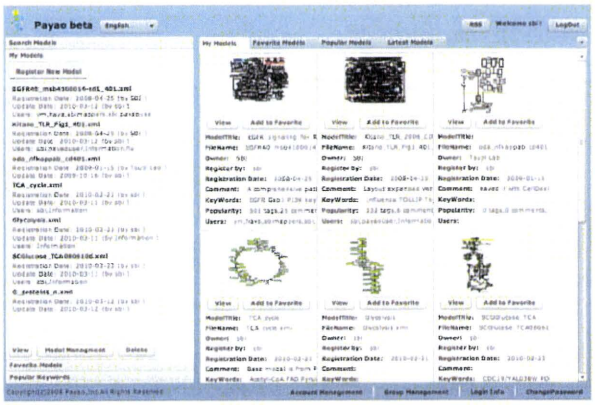


Fig. 2. Payao Model Browser screen—browse the registered models in the list format (left) and the thumbnails with summary data and statistics (right).

3.3 Community tagging and commenting

The tagging on the visually represented pathways is a characteristic of Payao, which makes the curators easy to grasp the nature of the pathway while discussing on the specific component of the pathways. Like Google Maps, tags are displayed in a bubble form attached to the items (Species, Reactions or any specified area), and click to expand and display the content of the information in the tag. Tags can specifically be keywords, links, PubMed IDs as well as free text, as shown in Figure 3. A TagSet groups a set of tags and can be color coded for ease of viewing. User-defined TagSets also allow access control features (browsing, editing and editing tags) to be set by a user, in the same way as for model access privilege settings. Thus, a user can set permissions to a 'My Tagset', which hides comments and tags from the community members. Inside the tag, comments can be added in the free text format. While tags anchor the points for annotation in the model, comments function serves as the discussion space.

3.4 Model update: tag data export and import

Data in the tags can be exported from the Payao system, evaluated outside the system and integrated into the Payao system or to the SBML model file. All the information, including the model file (CellDesigner format), Tags/Comments and model image (png/jpg/pdf format) in a ZIP file format are exported in a batch. Information stored in Tags is exported as a table in .csv format, which can be edited and imported back to the system. The tag information can also be imported into the base CellDesigner-format model as 'notes' using a CellDesigner plugin (import notes) and stored in SBML 'notes' annotation. Once the base model has been updated using CellDesigner, the model owner can reregister the model onto Payao for further curation.

4 FUTURE DEVELOPMENT

Payao will facilitate the evolution of CellDesigner from a simple tag-based pathway curation tool to a more versatile, comprehensive platform. The next version of CellDesigner will adopt support for Minimum Information Required in the Annotation of Models (MIRIAMs; Le Novère *et al.*, 2005) standard for annotation and

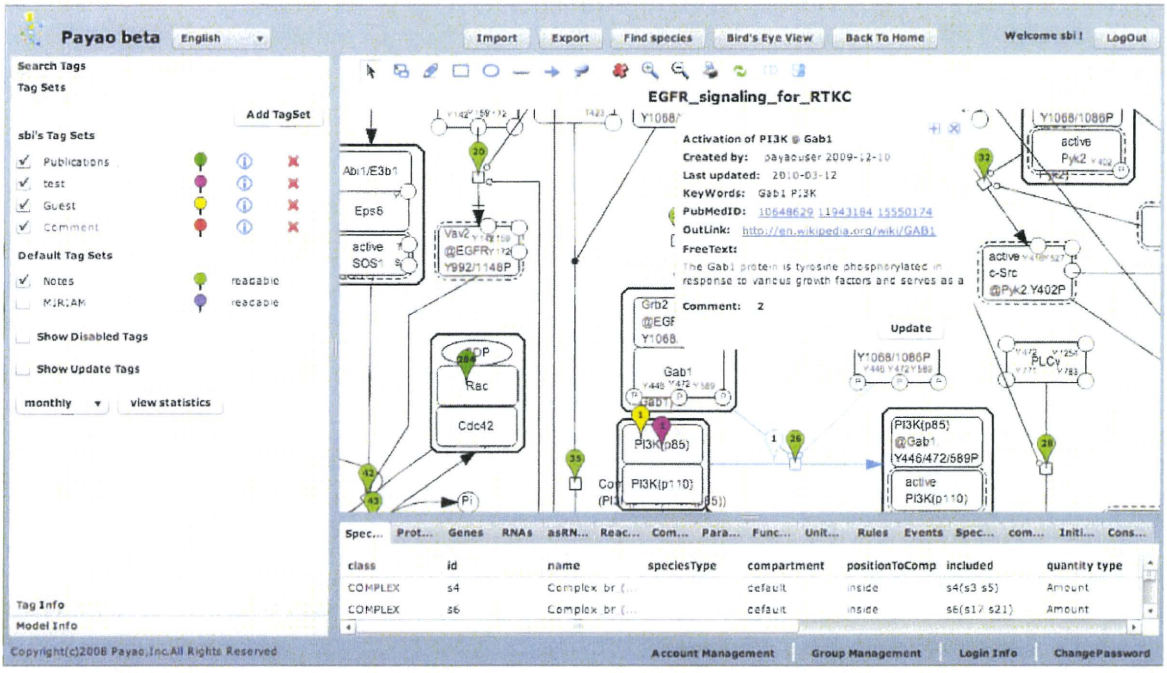


Fig. 3. Payao Model Viewer screenshot—model view panel and tag example. In the left panel, tagsets are listed to identify by specified color. In the right panel, under the tagset, assigned model users can add/edit tags associated with the nodes/reactions. Other users can add comments to the tag for providing further information or raising an issue against the tag contents.

allow MIRIAM annotations to be viewed on Payao. Modules to support import of different pathway exchange formats (BioPax), visual mapping of experimental data on pathway components in Payao, would be developed in future versions. Future enhancements also include integration of Payao with biological text mining techniques to facilitate literature-driven knowledge enrichment of existing pathway models. Tracking model updates using an RSS (Really Simple Syndication) feed will assist curators as well as public viewers. Statistics for measuring contribution of individuals to the model curation process as well as community interest (measured by count of tags and comments on a model) would be integrated into Payao in future releases. In the long run, we envisage this system to be a platform for aggregation, dissemination and community exchange for biological pathway models.

Funding: Okinawa Institute of Science and Technology (http://www.oist.jp); Establishment of a Human Genome Network Project (MEXT); Japan Science and Technology Agency/Biotechnology and Biological Sciences Research Council Strategic International Corporation Program; The Systems Biology Institute.

Conflict of Interest: none declared.

REFERENCES

Bauer-Mehren,A. *et al.* (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol. Syst. Biol.*, **5**, 290.

Funahashi,A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Pro. IEEE*, **96**, 1254–1265.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Le Novère,N. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.

Le Novère, N. *et al.* (2009) Systems biology graphical notation. *Nat. Biotech.*, **8**, 735–741.

Pico,A. *et al.* (2008) WikiPathways: pathway editing for the people. *PLoS Biol.*, **6**, e184.

Payao User Guide. (2009) Available at http://celldesigner.org/payao/payaopreview.html (last accessed date April 1, 2010)

frontiers in
PHYSIOLOGY

OPINION ARTICLE
published: 07 May 2010
doi: 10.3389/fphys.2010.00003

Grand challenges in systems physiology

Hiroaki Kitano[1,2,3]*

[1] The Systems Biology Institute, Tokyo, Japan
[2] Sony Computer Science Laboratories, Inc. Tokyo, Japan
[3] Okinawa Institute of Science and Technology, Okinawa, Japan
*Correspondence: kitano@sbi.jp

Systems physiology is an integrated discipline. It combines experimental, computational, and theoretical studies to advance our understanding of the physiology of human and other living creatures. In other words, systems physiology is systems biology with a physiology (i.e., functionally)-centered view. Understanding the principle behind the system is one of the fundamental challenges in systems physiology and systems biology. One can not make the use of sophisticated computational models or arrays of biological data to deepen our understanding of biological function without in-depth insights into the systems as a whole. For example, robustness and its trade-offs have been proposed as a fundamental principles (Kitano, 2004, 2007). This view, although still speculative, provides a framework for the conceptualization of data and observed phenomena. Identification of a series of such principles and their relationships can enrich our understanding of biological systems. The beauty of a good theory is that it reshapes our view of the world, so that the same data and phenomena may be re-interpreted in the light of the introduced concepts. Such transformation of our conceptualization often leads to true advances in science.

While such theoretical and explorative research are expected, it is also important to consolidate various efforts to achieve high impact objectives; these efforts are often referred as "Grand Challenges." Defining grand challenges provide an effective approach that both illuminates unresolved issues and helps focus research effort on these problems and thereby advances the state-of-the-art in systems physiology. It is most effective when used for engineering-oriented projects where progress can be made by the effective coordination of research and development programs along with a series of technological innovations, rather than merely waiting for serendipitous explorations. While basic scientific explorations are still indispensable and much needed in this field, it is also true that coordinated efforts on relatively well-defined missions can dramatically change the way we do science and apply it to medical practice. In this article, I attempt to define a series of grand challenges that are interlinked and designed to accomplish the ultimate goal of creating an integrated understanding and platform for human healthcare services, biomedical research, and drug discovery.

The grand challenge is to create highly accurate and broad coverage computational model of organisms that are backed up by well-controlled high precision experimental data. In practice, the true challenge is not only to build such a model, but also to establish a system of technologies that enable us to build these models cost-effectively, because these models must match genetic and epigenetic diversity. With this technology, both virtual human and virtual mouse models shall be developed. In addition, models of specific cell lines shall be developed. This set of models shall be consistent with a set of cells and organisms used for drug discovery and biomedical research. The reality of the drug discovery pipeline is that it uses cell lines and animal models before moving into clinical trial. Thus, it is important that not only human models, but also mouse and cell line models are developed with an equal level of quality. Accomplishment of this grand challenge will enable us to use computational models and associated experimental verification systems to understand disease mechanisms, and to predict drug efficacy, side effects, and therapeutic strategy outcomes. At a workshop held in Tokyo in February 2008, a group of researchers agreed to initiate a project to create a "virtual human" in next 30 years (Jones, 2008). They also announced the Tokyo Declaration that reads in part as follows: "Recent advances in Systems Biology indicate that the time is now ripe to initiate a grand challenge project to create over the next 30 years a comprehensive, molecules-based, multi-scale, computational model of the human ('the virtual human'), capable of simulating and predicting, with a reasonable degree of accuracy, the consequences of most of the perturbations that are relevant to healthcare."[1]

Although creation of a virtual human (a comprehensive computational model of human being) has been the subject of much discussion in variety of conferences and workshops, the real implications and difficulties with the model need to be re-addressed. There is no doubt that simulation, if properly used, can be a powerful tool for scientific and engineering research. Modern aircrafts cannot be developed without help of computational fluid dynamics (CFD). CFD is one of the most successful computational approaches used in the engineering design process.

There are three major reasons why CFD is now widely accepted. First, the Navier-Storkes equation has been well established to provide a computational basis for fluid dynamics with reasonable accuracy. While there are yet unresolved issues on how to compute tabular flows accurately, the Navier-Storkes equation provides an acceptable practical solution for most needs. Second, many CFD results are compared and calibrated against wind-tunnel experiments that are highly controlled and extensively monitored. Due to the existence of the wind-tunnel, CFD models can be improved for their accuracy and reliability of predictions. In wind tunnels, air flow speed, temperature, and other parameters can be adjusted within a very small error margin, for example within 0.01% error margin. Third, decades of effort have been spent on improving CFD and related fluid dynamics research. Thus, the current status of CFD is a result of decades of effort.

For computer simulation and analysis in biology to parallel the success of CFD, it must establish a fundamental computing paradigm comparable to the Navier-Storkes equation, to create a wind-tunnel equivalent for biological

[1] http://www.systems-biology.org/~myukiko/FC-SB2008/doku.php?id=workshop:statement

experiments, and to maintain a constant focus on these problems for decades. Of course, the biological system is much more heterogeneous and complex than fluids, but a set of basic equations must be established so that fundamental principles behind the computations point in the right direction. It is essential that both interaction networks and the physical structures are modeled together so that the resulting model provides an improved reality, particularly for high-resolution modeling of complex mammalian cells. Second, highly controlled and high-precision experimental systems that will serve as the "wind-tunnel" in biology are essential. Microfluidics and other emerging technologies may provide us with experimental paradigms that have remarkably high precision (Balagaddé et al., 2005).

Even if technologies can be developed, their full potential can not be reached unless they are used properly. There are, at least, two issues that must be carefully examined in order to make the best use of a computational simulation. First, the purpose of simulation has to be well defined, and model has to be constructed to maximize the purpose of the simulation. This affects the choice of modeling technique, levels of abstractions, the scope of modeling, and parameters to be varied. Second, the simulation needs to be well placed in the context of the whole analysis procedure. In most cases, simulation is not the only methods of analysis. Thus, the part of analysis that uses numerical simulation and the other parts that use non-simulation methods must be well coordinated in order to maximize overall success of the analysis activity. An example from racing car design illustrates these issues. CFD is extensively used in Formula-1 car design in order to obtain optimal aerodynamics (Ziemelis and Wenz, 2004); that is, a higher downward force coupled with a lower drag. Particular interest has been placed on effects of various aerodynamics components such as front wings, rear wings, and ground effects. However, the complicated interference between front wings, suspension members, wheels, and break air intake ducts must also be investigated. Combustion in engine is the other issue where simulation studies are often used, but simulated separately from CFD model. This exemplifies the practice of proper focus and abstraction. Thus, one can infer from this example that attempts to create a computational model of the human being without defining the model's use cases and the expected insights to be gained from the model would be economically inefficient and unlikely to be successful.

It should also be noted that CFD is not the only tool used for aerodynamics design. F-1 racing cars are initially designed using CFD (*in silico*), then further investigated using wind tunnel (*in physico*), followed by actual run at the test course (*in vitro*) before being deployed in actual races (*in vivo*). CFD in this case is used for initial search of candidate designs that are subject of further investigation and modification based upon results obtained from wind tunnel testing. This sequence of computational design followed by physical testing (experimentation) is the key for success in engineering design. It is highly likely the same would be true for biology. If so, a series of corresponding experimental platforms and methodologies may need to be developed to make the best use of the results obtained from the computational modeling approach.

Looking at the modeling platform, there are three major issues: scaling, sharing, and merging of biological models. The scaling problem, in turn, has three aspects: problem scaling, layer scaling, and scope scaling. Problem scaling means that the approach or computing framework enable models to get larger and larger to cover a substantial part of the organism. Developing a large-scale model is beyond the scope and capability of a single laboratory, and, in fact, may not even be possible within a national framework. It is critically important to establish an international collaborative framework to provide the infrastructure necessary to supports these activities in order to develop large scale models. This issue is directly related to the issue of sharing and merging of models. This requires installation of platform that fosters a global initiative. For example, there must be a mechanism by which the multiple models developed by different research groups can be combined into a single consistent model. Of course, an underlying assumption is that the models can be shared, requires well-informed communication within the community and the establishment of standards for models as seen in SBML (Hucka et al., 2003)[2] and SBGN (Le Novere et al., 2009)[3].

Layer scaling means that the model can incorporate multiple layers of description from the sub-cellular, cellular, and tissue level to the whole organism and assembly of organisms. This is a non-trivial issue because each layer may have different modalities of operation and a suitable way to represent these layers into models in a consistent and integrated manner is yet to be understood.

Finally, scope scalability can be defined as the capability of modeling approach to allow for an integrated treatment of both interactions between the layers and the physical structures (Kitano, 2006). While many models often used in systems biology focus on molecular interactions and gene regulatory networks, they often neglect the important structures and dynamics of intracellular and intercellular systems as well as the whole body. This is especially the case for physiological studies. For example, models that combine cytoskeleton dynamics, hence cell deformation and movements, with molecular and genetic interactions are at best rare, but more typically totally lacking.

It must be emphasized that one must first clearly establish what scientific questions are to be answered by using computational approach before the model of the biology system is developed. While this criterion has been already stated, it is so critical for the success but all too frequently forgotten during the course of model development that I shall repeat the point again. Mere attempts to create computational models that behave like an actual cell and organisms does not in themselves constitute a good scientific practice. It must be remembered what simulation and modeling represent is an abstraction of the actual phenomena. Without first carefully framing the scientific questions, a proper determination of the right level of abstraction and the scope of the model to be created is not possible. This is also the case in CFD. CFD as used in racing car design has a clear and an explicit optimization goal, namely to maximize the downward force while minimizing drag. The problem for biological simulations is that the information to be discovered by the simulation is much more complex than needed for racing car design. However, the questions must be a well-defined one in order to make the best

[2]http://www.sbml.org/
[3]http://www.sbgn.org/

use of computational machinery. With the right question and framing of problem, the model can become the starting point for a broad range of applications.

The discussion as presented so far outlines a new set of problems and challenges that has not been common in traditional biomedical sciences. This fact may have implications on how scientific communications, including journals, need to be organized and directed. For example, models and other resources that are gaining more importance have not been properly credited. There needs to be mechanisms by which proper credit can be assigned to the large groups of international experts who contribute to the incremental improvement of existing models and other knowledge resources. The proper consolidation of knowledge is as equally important for scientific advancement as are novel discoveries because the simple assembly of isolated knowledge, regardless of originality of discovery itself, does not enable us to achieve the grand challenge.

In order to resolve this issue, this journal attempts to provide a forum for the publication of modeling and mapping studies, results that have often been difficult to publish. The development of precision models, molecular interactions maps, and other knowledge-intensive resources are critical for the advancement of systems physiology and systems biology. In the past, the value of submissions describing these results have not been fully appreciated due to the assumption that these studies fail to provide novel insights despite the fact that important insight is often gained by efficient use of these models and maps. Just like an engineer who can design and build a great car without being a great driver, those who develop precision models and maps can provide functional insights that others can experimentally confirm or refute. In a similar fashion, being a great driver does not means that one can design and build a great car. Thus, someone who can gain insights from the models and maps may not be the one who can actually develop the resources used in the model. Model and map construction are engineering and infrastructure work that requires specific skills and dedication that provide resources essential for promoting systems physiology and such contributions need to be properly credited. A major goal of this journal is to establish an innovative forum of scientific exchange in the new area of web-based scientific activity.

A grand challenge for systems physiology entails this exciting objective. We need a series of innovations, discoveries, collaborative efforts, and dedications to accomplish it. The impact will be massive.

REFERENCES

Balagadde, F. K., You, L., Hansen, C. L., Arnold, F. H., and Quake, S. R. (2005). Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science* 309, 137–140.

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A. A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J. L., Kremling, A., Kummer, U., Le Novere, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., Nielsen, P. F., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., Spence, H. D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., and Wang, J. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531.

Jones, D. (2008). All Systems Go. *Nat. Rev. Drug Discov.* 7, 128–129.

Kitano, H. (2004). Biological robustness. *Nat. Rev. Genet.* 5, 826–837.

Kitano, H. (2006). Computational cellular dynamics: a network–physics integral. *Nat. Rev. Mol. Cell Biol.* 7, 163.

Kitano, H. (2007). Towards a theory of biological robustness. *Mol. Syst. Biol.* 3, 137.

Le Novere, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., and Kitano, H. (2009). The systems biology graphical notation. *Nat. Biotechnol.* 27, 735–741.

Ziemelis, K., and Wenz, C. (2004). Formula 1 racing: Science in the FAST LANE. *Nature* 431, 736–738.

Received: 28 March 2010; accepted: 31 March 2010; published online: 07 May 2010.

Citation: (2010) Kitano H. Grand challenges in systems physiology. Front. Physio. 1:3. doi: 10.3389/fphys.2010.00003

This article was submitted to Frontiers in Systems Physiology, a specialty of Frontiers in Physiology.

Molecular Systems Biology 6; Article number 384; doi:10.1038/msb.2010.40
Citation: *Molecular Systems Biology* 6:384

www.molecularsystemsbiology.com

molecular systems biology

PERSPECTIVE

Violations of robustness trade-offs

Hiroaki Kitano[1,2,3*]

[1] The Systems Biology Institute, Minato, Tokyo, Japan,
[2] Sony Computer Science Laboratories, Inc., Shinagawa, Tokyo, Japan and
[3] Okinawa Institute of Science and Technology, Kunigami, Okinawa, Japan
* Corresponding author. The Systems Biology Institute, Falcon Building 5F, 5-6-9 Shirokanedai, Minato, Tokyo 108-0071, Japan. Tel.: +81 3 5789 2181; Fax: +81 3 5789 2182; E-mail: kitano@sbi.jp

Received 21.10.09; accepted 12.5.10

Biological robustness is a principle that may shed light on system-level characteristics of biological systems. One intriguing aspect of the concept of biological robustness is the possible existence of intrinsic trade-offs among robustness, fragility, performance, and so on. At the same time, whether such trade-offs hold regardless of the situation or hold only under specific conditions warrants careful investigation. In this paper, we reassess this concept and argue that biological robustness may hold only when a system is sufficiently optimized and that it may not be conserved when there is room for optimization in its design. Several testable predictions and implications for cell culture experiments are presented.
Molecular Systems Biology **6**: 384; published online 22 June 2010; doi:10.1038/msb.2010.40
Subject Categories: metabolic and regulatory networks
Keywords: evolution; portfolio selection; robustness; suboptimality; trade-offs

Introduction

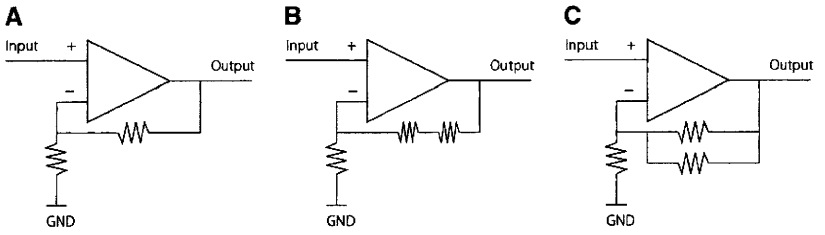
It has been claimed that trade-offs exist between robustness, fragility, performance, and resource demands in biological and engineering systems (Csete and Doyle, 2002; Kitano, 2004, 2007). Determination of the conditions in which this conjecture would hold is of great interest for systems theory in biology. For example, systems that are optimized to be robust against certain perturbations are often extremely fragile against unexpected perturbations. This trade-off is also known as the 'robust yet fragile' nature of highly optimized systems (Csete and Doyle, 2002). Principles such as the Bode integral formula (Bode, 1945) and the summation theorem in metabolic control analysis (Fell, 1997) underscore this trade-off in certain conditions. Although such theorems provide a basis for understanding robustness trade-offs, their applications are limited to specific aspects of the system. The Bode integral formula applies specifically to conservation of the sensitivities of negative feedback circuits on a frequency axis, and the summation theorem assumes linearization for minor perturbations. In addition, real systems are likely to exhibit more complex robustness–fragility trade-offs because of the involvement of component failures and other aspects not taken into account for these theorems (see Supplementary information). Such trade-offs may thus hold only when design and implementation are sufficiently optimized. This means that the system can be made more robust without undermining other features (see Box 1). It should be noted that the system can be optimized by redesign and reimplementation of engineering systems. In biological systems, an evolutionary selection is required for such design optimizations. Although qualitative observations exist for this trade-off, quantitative experimental verification of this trade-off has not been conducted.

In contrast, the trade-off between robustness and performance is more tractable, and several experimental and computational reports discussing such a trade-off have been published (Ibarra *et al*, 2002; Stelling *et al*, 2002; Fischer and Sauer, 2005; Andersson, 2006). In short, the trade-off dictates that high-performance systems are often more fragile than systems with suboptimal performance. Interestingly, there are studies reporting suboptimal metabolism performance in *Bacillus subtilis* and *Escherichia coli* (Stelling *et al*, 2002; Fischer and Sauer, 2005). If the trade-off holds, metabolic performance has to be kept suboptimal to ensure a certain level of robustness against environmental perturbations.

As such studies observe cultured microorganisms and cells, changes in performance and robustness can be attributed to either of the two scenarios: emergence and rebalancing. The emergence scenario assumes that random mutation gives rise to a genetic subtype that fits the perturbed environment better and that this subtype quickly proliferates in culture. The rebalancing scenario assumes that a specific mutant strain that fits the environment better may already exist in a heterogeneous population even before perturbations are imposed, and that this strain proliferates faster than other strains under the perturbed environment.

It is important to clearly define robustness and adaptation through evolutionary selection. Here, 'robustness' means an individual organism's capability of tolerating external and internal perturbations, such as environmental fluctuations, the addition of drugs, and mutations. Robustness–performance trade-off means that, when two individuals are compared, one is found to be more robust than the other but is outperformed by the other; thus, no individual can be more robust and at the same time exhibit higher performance than others. In general, organisms can be 'optimized' or 'adapted' to a certain condition by evolutionary selection; thus, they can be more robust against perturbations implicated in the

Box 1 Design suboptimality and robustness–fragility trade-offs



The robustness–fragility trade-off is one of the most widely known trade-offs for biological and engineering systems. A simple toy example using electric circuits is presented here to show that the trade-off holds only when system design and implementation are sufficiently optimized. In other words, a circuit can be made more robust without scarifying other features.Take a simple electronics example: Assume a simple feedback amplifier in which a feedback loop consists of resistors (Design **A**). There is clearly a robustness–fragility trade-off due to the feedback loop, as depicted in the Bode theorem (Bode, 1945). Without affecting the trade-off due to the feedback, the actual implementation of the amplifier can be changed to have two serially connected resistors in its feedback loop (Design **B**). Failure of one of these resisters may cause dysfunction in the feedback loop and undermine the system stability. Alternatively, the use of resistors connected in parallel would reduce such a failure risk (Design **C**). The parallel implementation is more robust against component failure than both a single and a serial connection configuration. Assuming that a probability of failure of each resister is P, Design B is susceptive to component failure, as a probability of failure of the feedback loop simply doubles ($2P$). Design A reduces such a risk to half of that of Design B because it only uses one resister; hence, probability of failure is P. Design C improves robustness against component failure on feedback loops because of parallel construction. Now it has only P^2, whereas resource demand is equivalent to Design B. For example, a design change from Design B to Design C improves the robustness of the system against component failures without increasing fragility elsewhere, undermining performance, or requiring major additional resources. Thus, robustness is improved without substantial trade-offs. Change from Design B to A actually reduces resource demands slightly and improves robustness. Change from Design A to C improves robustness with minimum increase in resource demands. However, attempts to totally eliminate component failure, not only for a feedback loop but for every aspect of the system, would require multiple redundancies for every aspect of the system, which would require major resources. How design affects the system vulnerability to component failure is a complex issue, and biological cases need to be further explored. It will be further complicated when feedback loops are involved (see Supplementary information).

selection pressure or can have higher performance than preselected individuals. Thus, if the robustness–performance trade-off holds, descendents of organisms can be more robust than their ancestors when they are adapted for perturbations imposed during evolutionary selection, but they may be outperformed by their ancestors or by other individuals adapted for other conditions in which performance is favored. By the same token, the descendents of organisms can outperform their ancestors when selection pressure favors high-performing individuals, but may be less robust than their ancestors and other individuals evolved under conditions that favor more robust individuals.

In this paper, we examine the idea that such trade-offs appear only when the system is sufficiently optimized, and thus may not be observed when systems are yet to be fully optimized. This implies that there should be cases in which descendents of organisms can be more robust and perform as well as or better than their ancestors, which is not possible if robustness–performance trade-off holds universally.

A primer on portfolio selection

In this article, the portfolio selection concept used in modern investment theory is introduced to explain robustness–performance trade-offs. Portfolio selection is an idea to combine various investment options to minimize risk while attaining the desired return on investment (Markowitz, 1991). In the modern portfolio theory used in investment practice, it is well understood that there is trade-off between risk (uncertainty of the return shown by s.d.) and performance (expected return). High-yield financial products generally have higher risk, and modest-yield products have lower risk. Risk in this context refers to the s.d. of the asset price. Performance is measured by the expected percentage of return. Any investment item (asset) can be mapped on a yield–risk space.

As investors generally invest in multiple financial assets with different expected yields and risks, the question is how to find the optimal mix of assets with a desirable yield and acceptable risk. The concept of efficient frontier needs to be introduced here. The efficient frontier is a set of points that represent an optimal combination of assets (mostly securities in a financial context) that maximizes the return for any given level of s.d. Any point not on the efficient frontier represents a portfolio that is inferior to a portfolio on the efficient frontier, either because it generates less return at the same level of risk or is exposed to higher risk at the same expected level of return. In Figure 1A, Portfolio X is inferior to both Portfolios Y and Z. Portfolio Y has a higher expected return than X at the same level of risk, and Z has lower risk than X with the same expected return. Portfolio X can be reorganized to reach the efficient frontier. Thus, theoretically, any portfolio not on the efficient frontier can improve its yield without increasing risk, or reduce risk without undermining the expected yield. However, on the efficient frontier, any change in yield affects risk and *vice versa*. Trade-off between risk and yield takes place on an optimal portfolio that is on the efficient frontier. A similar trade-off concept is also investigated in the context of multiobjective optimization as the Pareto efficiency, originally proposed by Pareto (1935). For a Pareto-efficient solution, no individual parameter can be improved without undermining another parameter. A set of Pareto-efficient solutions constitute a Pareto surface, also called a Pareto frontier.

An indifference curve projects valuation criteria on the yield–risk space. It has graded utility levels depending on

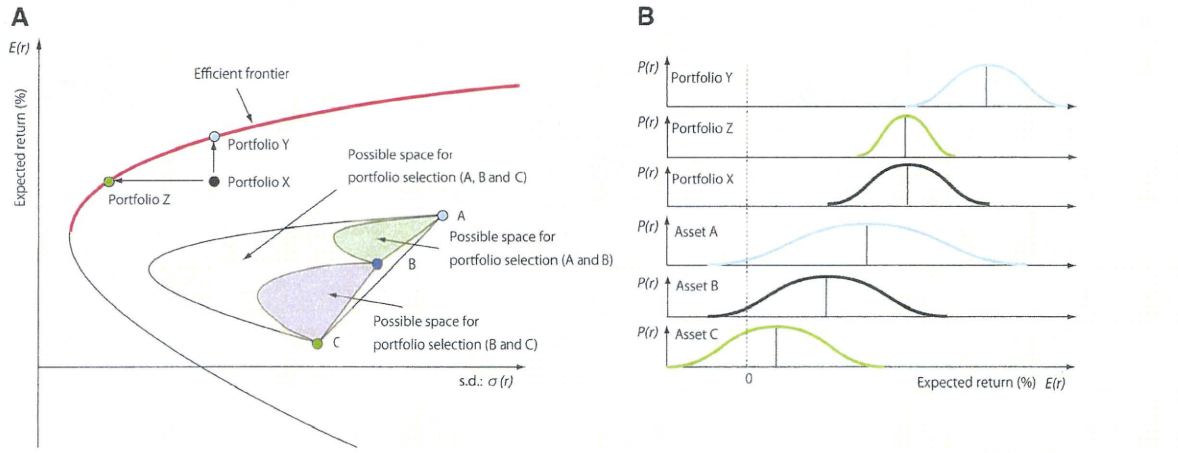


Figure 1 Basic concepts of portfolio selection. (**A**) Any asset can be mapped on the $E(r)$–$\sigma(r)$, or yield–risk, space. Combining A and B creates a possible space of yield and risk depending on the mixture and covariance of the two assets against fluctuation. Maximum risk reduction is achieved when the prices of the two assets change in opposite directions because this offsets fluctuation. Increasing the number of assets involved generally reduces risk. The efficient frontier is achieved by optimally combining all available assets. Availability of a larger number of assets with different yield–risk characteristics improves the overall portfolio, analogous to an increase in the degree of design of freedom in highly optimized tolerance (Carlson and Doyle, 1999; Reynolds *et al*, 2002). In actual investment planning, investment with a fixed return asset is considered to form a capital market line, but this is not considered in biological applications because there is no zero-risk-fixed-yield genotype. (**B**) The probability distribution of expected return is shown for each asset and portfolio in (A) to visually illustrate their relationships.

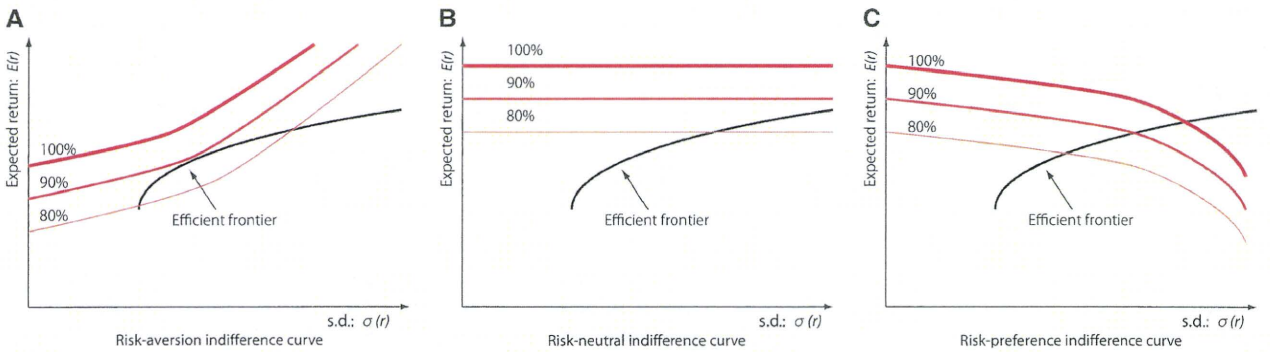


Figure 2 Indifference curves. Three indifference curves are shown: (**A**) risk-aversion, (**B**) risk-neutral, and (**C**) risk-preference curves. Percentile numbers associated with each line indicate the level of utility, hence the level of satisfaction or optimality in the given context.

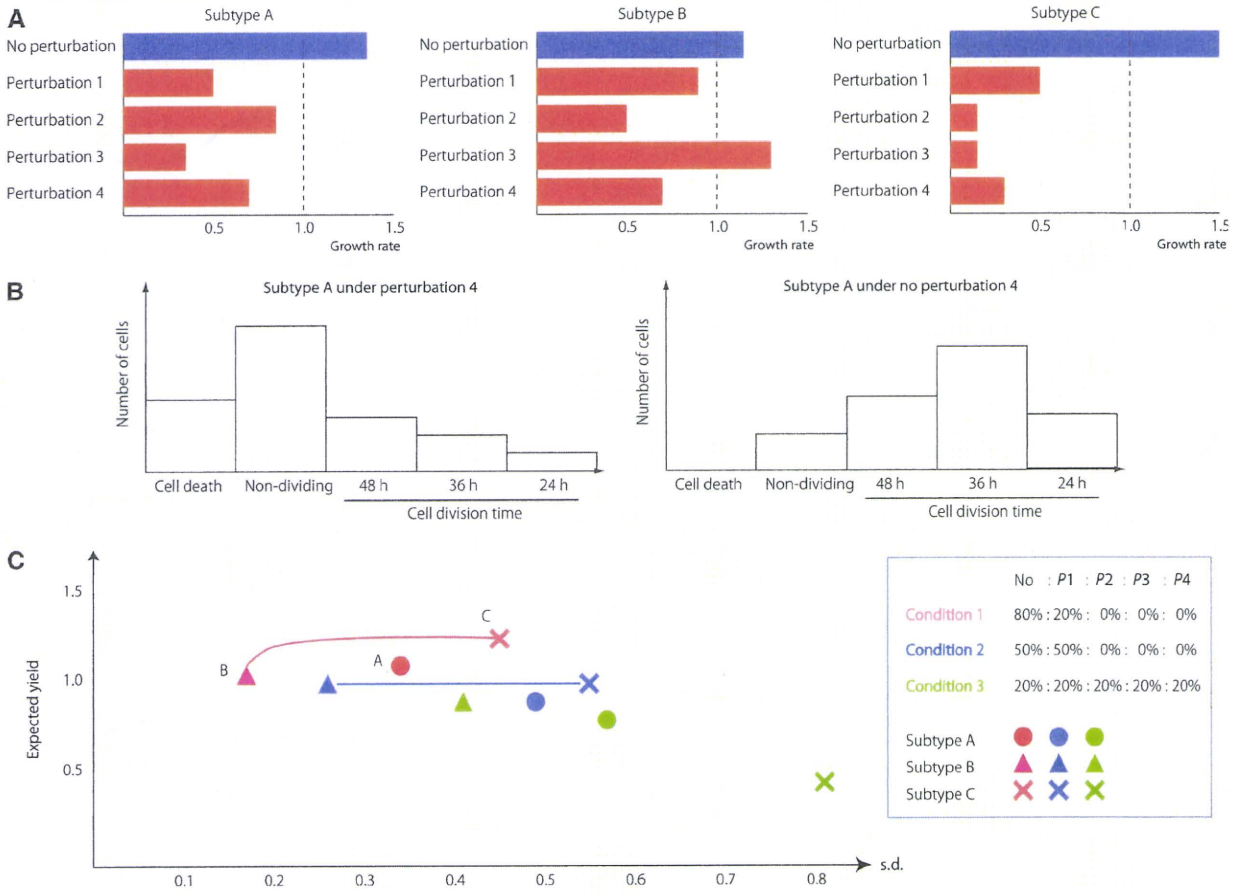
whether the desired portfolio is selected on the basis of risk preference. A risk-aversion indifference curve represents a portfolio chosen to maximize the expected return but avoids risk (Figure 2A). A risk-neutral indifference curve is used when only the expected return is considered (Figure 2B). A risk-preference indifference curve is used when higher risk is preferred for an equal expected return (Figure 2C). Obviously, the risk-preference indifference curve would be an odd choice for an investment situation. Thus, the risk-aversion indifference curve is used in general.

Genetic portfolio selection: translating investment theory into biology

Portfolio selection, which seems remote from biology, can be applied to understand the evolution of microorganisms and cells in specific conditions. Thus, it may help us understand robustness–performance trade-offs.

Each organism or cell can be mapped onto the yield–risk space. A position of the yield–risk space that characterizes the biological entity X can be called 'a projected position of X.' Yield is an expected performance, such as reproduction speed or biomass production rate. Risk (equivalent to fragility) is the degree to which a growth rate or biomass production rate is affected by perturbations. In general, it can be represented by s.d. and calculated by assuming possible perturbations, their probabilities of occurrence, and expected yield under each perturbation (Box 2). These indexes can be measured by tracing behaviors of individual cells and their biomass production or lineage for reproductive efficiency under various conditions. Alternatively, they can be measured by the growth of the population under various perturbations in which the population can be kept monoclonal. In this case, the distribution of projected positions of a certain cell or organism

Box 2 Mapping cell culture onto yield–risk space



Suppose the yield is the growth of an organism. Expected yield and its s.d. can be determined through a series of perturbation experiments. For example, the growth rate of each subtype may be measured under a normal culture condition (no perturbation or reference) and with different perturbations (**A**). Selection is imposed at an individual level. Histograms (**B**) indicate the number of cells that result in cell death, nondividing cells after a certain period of time, and dividing cells with different intervals for subtype A under different conditions. Expected yields in (A) are the results of such responses of individual cells for each condition.Subtypes can be mapped on the yield–risk space as shown in (**C**). Under the condition that 80% of the time there is no perturbation and 20% of the time perturbation 1 occurs, each subtype can be mapped on the yield–risk space as shown by red symbols. In this case, a combination of subtypes B and C reduces risk without major sacrifice in yield. Because of the covariance of B and C that has a very small positive value (0.068), mitigation of risk over the risk of subtype B is limited. The projected position of each subtype in the yield–risk space will be different if assumed perturbations and their probability of occurrence are different. Under the assumption of no perturbation and perturbation 1 conditions, each with a 50% probability, the projected position of each subtype in the yield–risk space is shown by blue symbols, which are located differently from the red symbols that assume no perturbation for 80% probability. The expected yields of subtypes B and C are now equal; thus, a combination of B and C only reduces risk, yield is not reduced. Yet another assumption of perturbations, in which no perturbation and perturbations 1, 2, 3, and 4 each occur with 20% probability, would result in each subtype being located as shown by the green symbols. In this case, subtypes A and C suffer seriously from perturbations; thus, the population will be dominated by subtype B, which has robust yet low-yield characteristics.

for its wild type and various mutants on the yield–risk space is contained by the efficient frontier (Figure 3). Changes in the distribution of projected positions in the yield–risk space for randomly sampled cells will test the conjecture that the robustness–performance trade-off holds only at the efficient frontier.

Next, we consider the cases in which such a trade-off holds in a population of microorganisms and cells. Analysis at the population level is biologically important because cell cultures that have a substantial level of heterogeneity are often used in biological experiments. In addition, certain tumors are known to be composed of heterogeneous cancer cells. Furthermore, populations of organisms and cells are used to measure growth rate and how organisms respond to environmental changes in the context of the study of general biology and in drug screening.

Growth rate (yield or performance) is generally measured by the size of a colony, by numbers of cells, or by other means that reflect the number of cells in the population. Risk is an s.d. of growth rate under various possible perturbations. Experimentally, it can be measured by repeated perturbation experiments. In a heterogeneous cell culture, the projected position of the cell culture in the yield–risk space is determined by the population composition. This is illustrated in Figure 4. Initially, the cell culture is mainly composed of subtypes A and B, with a negligible amount of C (Culture 1 in Figure 4A). Subtype C has a better fit with the culture condition and has

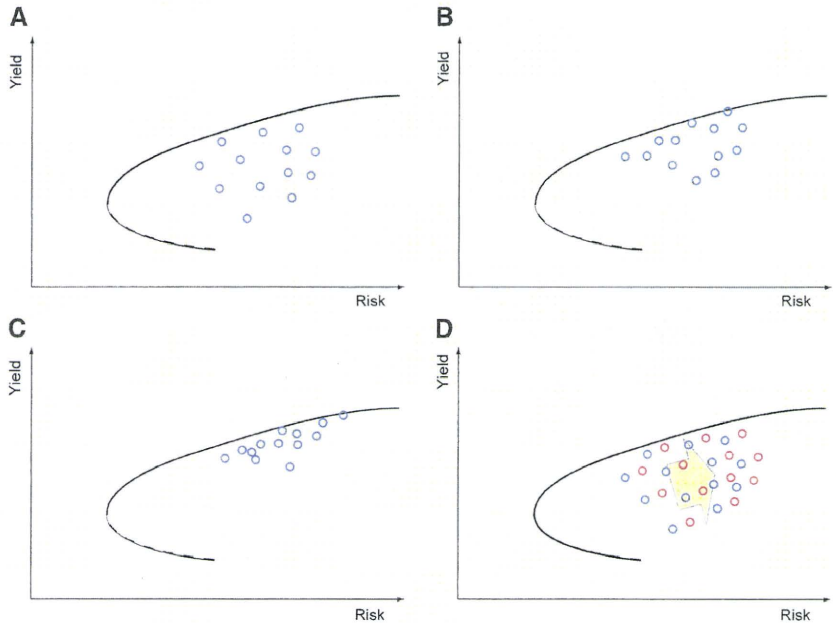


Figure 3 Distribution of randomly sampled cells on yield–risk space. A heterogeneous culture of cells is composed of a mixture of wild type and various mutant cells. (**A**) At the initial stage of culturing, random sampling of cells and mapping onto the yield–risk space may result in a broad distribution for the yield and risk of each cell. (**B**) Culturing this population under the stationary condition for multiple passages may result in evolution of the culture toward a high-yield genotype. Each circle represents randomly sampled cells at this time point, and not the same cell sampled in (A), because multiple passages have occurred. (**C**) Further passages under the stationary condition result in a distribution strongly favoring high-yield individuals. If there is a trade-off between robustness and performance, hence an inverse of risk and yield holds, distribution will be constrained within a certain envelope represented by the efficient frontier. Artificial evolution experiments with random sampling for yield–risk space mapping test the conjecture of robustness–performance trade-off at the efficient frontier. (**D**) If the robustness–performance trade-off holds at any time, even without being at the efficient frontier, the center of gravity of all randomly sampled points after passages will simply shift toward the upper right. This contrasts to the case that the trade-off holds only at the efficient frontier. This difference can be experimentally verified.

higher yield. Thus, the proportion of subtype C increases over that of subtypes A and B (Culture 2 in Figure 4A). Random mutation gives rise to subtype D. It has higher reproductive potential under this specific culture condition, and thus it quickly proliferates in the population (Culture 3 in Figure 4A). However, when extra perturbations are imposed on the culture, fast-growing but less-robust subtypes (subtype D) may substantially decrease in their proliferation speed or the number of cells. At the same time, low-yield but more-robust subtypes may continue to grow at a similar rate. These population changes result in a composition that better fits a condition with a higher degree of perturbations. In this case, the projected position of the culture on the yield–risk space map may move left to that of Culture 4 in Figure 4A. In contrast, the fast-growing subtype may establish its dominance when the environment reaches a more stable condition that is ideal for the fast but less-robust subtype (Culture 5 in Figure 4A). Both emergence and rebalancing scenarios are included, but for the sake of explanation, only the wild type and its mutational variants are used as subtypes. However, cells with different epigenetic modifications can be considered as subtypes if these modifications affect the yield–risk characteristics of the cell.

Although translation of portfolio theory for biology is shown to be possible, some clear and essential differences have to be made explicit and given a new interpretation that is consistent with biology. First, portfolio selection assumes that there are investors and fund managers making decisions regarding the composition of assets selected for the portfolio. This is clearly not the case in biology. The population composition changes because of the relative growth rate of each cell subtype that is the aggregated effect of individual cell reproduction cycles. Individual cells and organisms will be the subject of selection. Second, in portfolio selection theory, investors decide on the indifference curve to be used on the basis of their appetite for risk taking. In the biological translation, indifference curves are only a reflection of the level of perturbation imposed on organisms and cells (Figure 4B). When organisms and cells are cultured in a highly stationary environment, the use of the risk-neutral curve may best predict their possible evolutionary paths. Risk-aversion curves represent the situation in which perturbations are imposed on organisms and cells.

Performance suboptimality and robustness trade-offs

Studies report that suboptimal metabolism performance exists in microorganisms such as *B. subtilis* and *E. coli* (Stelling *et al*, 2002; Fischer and Sauer, 2005). Fischer and Sauer (2005) argue that several regulatory mutants that have improved biomass production efficiency were 'almost exclusively regulators of not-yet-activated adaptive responses, suggesting that *B. subtilis* invests valuable resources in anticipation of changing environmental conditions at the expense of optimal growth'. As almost all mutations to enhance biomass production

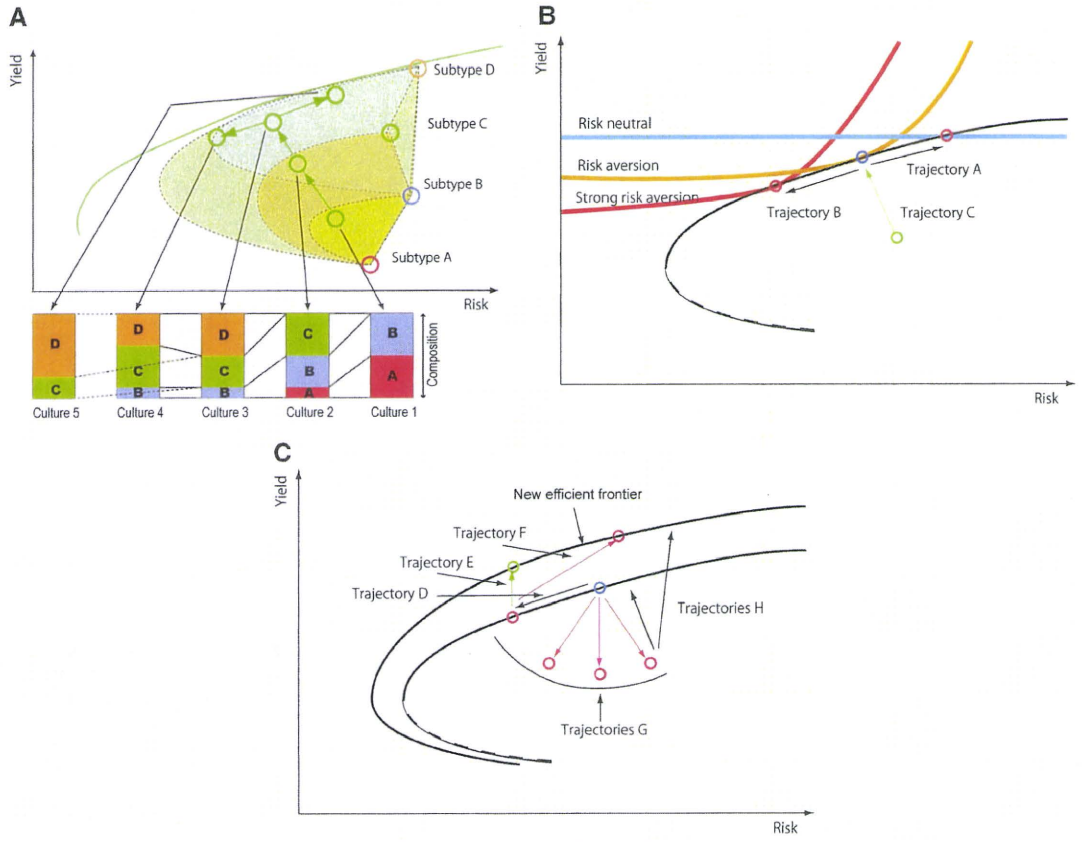


Figure 4 (**A**) In biological application of portfolio selection, each individual is mapped on the yield–risk space. A culture of cells may contain multiple different subtypes with mutations and epigenetic modifications. Assume a culture of cells (or organisms) composed of subtypes A, B, C, and D. At the initial stage, A and B may be dominant and C and D may be negligible (Culture 1). However, subtype C better fits the environment and grows faster than A and B (Culture 2). Subtype D starts to grow faster than the others and changes the subtype composition of the culture (Culture 3). At this stage, the composition of the culture may be sufficiently optimized for the given culture condition. Suppose the culture condition is changed now to have greater perturbations. Subtype D may not be able to tolerate it and will decrease the rate of proliferation and may even reduce in number, and subtype C may grow faster than the other subtypes (Culture 4). Alternatively, subtype D may continue to grow faster than other subtypes if the environment becomes even more stable. (**B**) A population of cells may evolve toward the efficient frontier. Under the risk-aversion indifference curve, the population arrives at the blue circle on the efficient frontier. The risk-aversion curve represents cases in which higher-level perturbations are imposed on the culture compared with a risk-neutral case. Under the stable condition in which selection pressures other than growth speed are not significant, the risk-neutral indifference curve is likely to be applied. The population follows Trajectory A and maximizes its growth rate at the cost of robustness. Imposing a higher level of perturbation may result in transition of the state through Trajectory B. (**C**) Cost-free resistance may be a result of taking Trajectory E or F to a new efficient frontier. There may be cases in which the population moves back to suboptimal regions (Trajectories G). Chemotherapy for cancer may shift the point inside the efficient frontier with different end points because of heterogeneous subpopulations. However, tumor cells may again evolve to gain proliferation potential despite the presence of anticancer drugs (Trajectories H). Tumor cells that undergo this transition may be too optimized for this specific therapeutic intervention, which implies possible efficacy when therapeutic regimens are switched. This may explain the collateral sensitivity of drug resistance tumor cells (Skipper *et al*, 1972).

involve genes related to adaptations to environmental changes, it is likely that the projected position of *B. subtilis* is on the efficient frontier and optimized for the risk-aversion indifference curve (orange line in Figure 4B), because any change to increase biomass would be associated with an increased risk under environmental fluctuations (Trajectory A in Figure 4B). Thus, although Fischer and Sauer argue that *B. subtilis* exhibits suboptimal metabolism, it can be considered optimal for the risk-aversion indifference curve.

Culturing microorganisms or cells under a chemo-static laboratory setup may enable the creation of an almost perturbation-free environment, unless explicitly imposed. In such a case, a risk-neutral indifference curve can be applied to determine the optimal portfolio (blue line in Figure 4B). For example, a population of *E. coli* evolved in the laboratory may quickly arrive at one of the projected positions on the efficient frontier and move along the trajectory of maximum biomass production through evolution (Ibarra *et al*, 2002) (Trajectory A in Figure 4B). The line of optimality in the phenotype phase plane as shown in Ibarra *et al* (2002) may correspond to the efficient frontier under stationary condition. In fact, their Figure 4 shows that the yield of bacteria evolves toward the line of optimality that is consistent with Figure 3 (A–C) in this paper. qIt should be noted that the line of optimality does not reflect the risk factor against perturbations. It only illustrates yield optimization. Therefore, it can be speculated that the *E. coli* that achieved maximum biomass production is more fragile against environmental change than those that have not evolved.

Interestingly, there are studies reporting that some pathogens evolve to overcome decreased growth fitness. As a result,

they can become drug resistant and maintain a competitive growth rate (Andersson, 2006). Although there is controversy with regard to this concept of 'cost-free evolution,' if it is correct, this compensatory evolution is thought to adapt to drugs that result in individuals that are more robust against drug insults, without increasing fragility elsewhere or undermining system performance measured by growth rate. In general, pathogens that have acquired drug resistance are known to have less growth fitness than pathogens without drug resistance. This can be seen as a trajectory along the efficient frontier, but toward the lower-left direction in Figure 4B (Trajectory B) and 4C (Trajectory D). Because of the existence of a drug, pathogens that are able to proliferate under drug exposure grow faster than others. Such pathogens may arise because of random mutations that better fit the drug-exposed environment. As a result, the population evolves to be optimal under the strong risk-averse indifference curve. This implies that trade-offs exist between increased drug resistance and a competitive growth rate against nonresistant pathogens. Again, 'robustness' is used as a general term to define the organism's tolerance against perturbations, including environmental fluctuations, the addition of drugs, and mutations. Although acquisition of drug resistance does not affect an organism's capability to cope with non-drug perturbations, overall robustness was considered to be increased because of added tolerance to the drug. Although the existence of 'cost-free evolution' seems to breach the concepts of robustness–performance trade-off, it is consistent because this is a process of moving toward a new efficient frontier (hence, the population is evolving), and robustness–performance trade-off can be observed only on the efficient frontier. The efficient frontier has changed because of the presence of a drug that was not a factor in determining the original efficient frontier (Trajectory E or F in Figure 4C).

Predictions and implications

The application of the concept of portfolio selection to biological systems results in testable predictions. First, it is predicted that the growth rate or biomass production of organisms and cells can be improved through evolution without increasing fragility until the efficient frontier is reached. A trade-off emerges only at the efficient frontier. This can be tested by random sampling of cells in artificial evolution experiments under a stationary culture condition, as shown in Figure 3. By the same token, the growth rate or biomass production of a population of organisms and cells can be improved without increasing their fragility against perturbations. Only when the population is on the efficient frontier do changes in yield and risk affect each other, hence a trade-off emerges. Therefore, robustness (or fragility) is not always conserved; it is conserved only when the system is on the efficient frontier.

Second, once the population or individual cells are on the efficient frontier and a stable culture condition persists, their position in the yield–risk space may move along the efficient frontier to the right for higher yield at the cost of robustness. This is because of the higher growth rate of a subtype over others that better fit the condition. Whether the population is on a trajectory toward the efficient frontier or moving along the efficient frontier can be distinguished by looking at the types of mutations and upregulation of genes that occur during such transitions. If the projected position of the population of cells on the yield–risk space moves toward the efficient frontier from a suboptimal portfolio space (Trajectory C in Figure 4B), mutations and gene up- or downregulations can be observed in broader functional classes of genes. In contrast, if the projected position of the population of cells on the yield–risk space moves along the efficient frontier to the right (Trajectory A in Figure 4B), mutations that generate high-risk high-yield phenotypes and downregulation of genes that accounts for perturbations may be observed. Prediction can be tested by sampling populations of cell and individual cells for sequencing and expression profile measurements to identify distribution of genes that are affected. In addition, different genes may be upregulated in a culture condition in which multiple perturbations are constantly imposed, because this would push the population to a lower-yield projected position. Genes that are accountable for environmental perturbations will be upregulated and genes that attain a higher yield may be downregulated.

If these conjectures hold and are proven to have wider applicability, there will be several implications for how we handle cell culture experiments. Cells cultured for multiple generations may have the problem of being optimized for a culture-specific condition and higher growth rate rather than for robustness against broader perturbations. Consider the drug-screening process. Drugs are initially screened using cell cultures. When a cancer cell line is used, for example, various drug candidates are applied for various cell lines. Cells in the culture are those that best fit the specific culture condition, which does not necessarily represent an *in vivo* environment for tumor cells. The most successful drug candidate may then be the one that undermines the growth of cells that are optimized for this specific condition. As *in vivo* cancer cells may be optimized for surviving under various perturbations, but may not for growth rate, a serious discrepancy exists between cells used for screening and actual cancer cells. Such a discrepancy may be mitigated if a culture condition can be set to impose various perturbations mimicking the cancer cells to which the body may be exposed. Thus, a deeper understanding of the type of perturbations that tumor cells in the body may be exposed to may make it possible to develop a multiple perturbation culture system that may improve the screening process.

By the same token, induced stem cells that are screened for therapeutic purposes may entail a similar problem. Cells with undermined robustness may be selected in favor of efficient reprogramming and upregulation of induction and differentiation markers, rather than cells that maintain robustness against broader perturbations. Currently, multiple generations are required for induction of pluripotent stem cells and elimination of epigenetic traces that are reminiscent of original cells (Masaki *et al*, 2007). During this process, which often requires multiple passages, new epigenetic modifications that are introduced by specific culture conditions are inevitable (Rubin, 1994; Meissner *et al*, 2008). It remains to be seen whether characteristics coselected for

high-yield subtypes entail any unwanted characteristics for therapeutic use.

With the introduction of the portfolio selection concept, observed breaches of trade-offs and enigma of performance suboptimality can be explained. Further studies and verifications are expected to lead to solid theories for biological systems and their applications to medical research.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (http://www.nature.com/msb).

Conflict of interest

The author declares that he has no conflict of interest.

References

Andersson DI (2006) The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol* **9:** 461–465

Bode H (1945) *Network Analysis and Feedback Amplifier Design.* Melbourne, FL: Krieger

Carlson JM, Doyle J (1999) Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **60** (2 Pt A): 1412–1427

Csete ME, Doyle JC (2002) Reverse engineering of biological complexity. *Science* **295:** 1664–1669

Fell D (1997) *Understanding the Control of Metabolism.* London: Portland Press

Fischer E, Sauer U (2005) Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. *Nat Genet* **37:** 636–640

Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420:** 186–189

Kitano H (2004) Biological robustness. *Nat Rev Genet* **5:** 826–837

Kitano H (2007) Towards a theory of biological robustness. *Mol Syst Biol* **3:** 137

Markowitz H (1991) *Portfolio Selection: Efficient Diversification of Investments.* Hoboken, NJ: John Wiley & Sons

Masaki H, Ishikawa T, Takahashi S, Okumura M, Sakai N, Haga M, Kominami K, Migita H, McDonald F, Shimada F, Sakurada K (2007) Heterogeneity of pluripotent marker gene expression in colonies generated in human iPS cell induction culture. *Stem Cell Res* **1:** 105–115

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454:** 766–770

Pareto V (1935) *The Mind and Society (English version).* New York: Dover

Reynolds D, Carlson JM, Doyle J (2002) Design degrees of freedom and mechanisms for complexity. *Phys Rev E Stat Nonlin Soft Matter Phys* **66** (1 Pt 2): 016108

Rubin H (1994) Cellular epigenetics: control of the size, shape, and spatial distribution of transformed foci by interactions between the transformed and nontransformed cells. *Proc Natl Acad Sci USA* **91:** 1039–1043

Skipper HE, Hutchison DJ, Schabel Jr FM, Schmidt LH, Goldin A, Brockman RW, Venditti JM, Wodinsky I (1972) A quick reference chart on cross resistance between anticancer patients. *Cancer Chemother Rep* **56:** 493–498

Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420:** 190–193

Large-Scale Analysis of Network Bistability for Human Cancers

Tetsuya Shiraishi[1]*, Shinako Matsuyama[2], Hiroaki Kitano[1,3,4]

1 Sony Computer Science Laboratories, Shinagawa-ku, Tokyo, Japan, **2** Sony Corporation, Shinagawa-ku, Tokyo, Japan, **3** The Systems Biology Institute, Shinjuku-ku, Tokyo, Japan, **4** Okinawa Institute of Science and Technology, Kunigami, Okinawa, Japan

Abstract

Protein–protein interaction and gene regulatory networks are likely to be locked in a state corresponding to a disease by the behavior of one or more bistable circuits exhibiting switch-like behavior. Sets of genes could be over-expressed or repressed when anomalies due to disease appear, and the circuits responsible for this over- or under-expression might persist for as long as the disease state continues. This paper shows how a large-scale analysis of network bistability for various human cancers can identify genes that can potentially serve as drug targets or diagnosis biomarkers.

Citation: Shiraishi T, Matsuyama S, Kitano H (2010) Large-Scale Analysis of Network Bistability for Human Cancers. PLoS Comput Biol 6(7): e1000851. doi:10.1371/journal.pcbi.1000851

Editor: Nathan D. Price, University of Illinois at Urbana-Champaign, United States of America

Received May 14, 2009; **Accepted** June 3, 2010; **Published** July 8, 2010

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Tetsuya.Shiraishi@jp.sony.com

Introduction

Understanding diseases within the context of biological networks is one of the major challenges in systems biology. Diseases often persist and resist therapeutic intervention. The persistence of a disease in a system must be reflected in the ability of the system's networks to maintain the state underlying the disease. In other words, networks are "locked-in" to disease states and maintain their stability. Thus, it is important to understand how such multi-stable states are achieved within the context of network topology and to understand the dynamics of these states. A network robust against a range of perturbations can maintain a healthy state but can also, when affected by a disease, transition to a new steady state that is often also robust against perturbations, making the disease state persistent. A series of disease progressions may be the result of a sequence of state transitions in the network dynamics (Fig. 1A). Bistable circuits may drive such transitions and are thus critical in enabling the initiation and progression of diseases to be understood (Fig. 1 B).

Complex networks exhibiting such multi-stability must have a set of bi-stable or multi-stable circuits consisting of proteins and genes. The identification of circuits that exhibit bi- or multi-stability within large protein-interaction and gene-regulation networks would provide information useful for understanding the mechanism(s) of network bistability. Furthermore, circuits exhibiting bistability can be potential drug targets or biomarkers for classifying disease states.

Network dynamics are regulated by the structure of the network and the flow of information through feedforward and feedback loops. Mutual activation or mutual inhibition configurations can maintain the flow of biological information between two molecules and act as network memories or switches. Furthermore, an activation-inhibition configuration, in which one molecule stimulates the other while the latter inhibits the former, generates dynamics with periodicity like that seen in circadian rhythms and cell cycles [1]. The stability and characteristics of Boolean networks comprising these configurations were studied in detail by Kauffman et al. [2]. In the study reported here, we focused on mutual inhibition, which is thought to be involved in the stable deviations of a system observed during the progression of tissue from a normal to a diseased state.

There are several important network motifs for system configurations [3–6] in protein-protein networks. One of them, a toggle switch that converts a continuous input signal into a discontinuous ON or OFF response, plays a fundamental role in information processing and decision making. Among the naturally occurring toggle switches that have been reported are the lambda phage lysis–lysogeny switch [7–9], switches in the lactose operon repressor system [10–12], the mitogen-activated protein kinase (MAPK) cascade [13–20], the Sonic hedgehog network in stem-cell differentiation [21], cell-cycle regulatory circuits [22–24], and the rapid lateral propagation of receptor tyrosine kinase activation [25]. Genetically engineered toggle switches have been constructed experimentally in Escherichia coli [26,27] and in mammalian cells [28].

A robust toggle switch behaves as a signal memory unit by using a hysteresis mechanism [29]. Once in the ON state, a toggle switch remains in the ON state even if the stimulus concentration falls below the threshold level [11,13,23,24,30,31]. A molecular network's persistence in a disease state might be due to the hysteresis of toggle switches.

To identify circuits exhibiting bi- and multi-stability, we topologically analyzed activation and inhibition in proteins on a large scale by using various databases containing expression array data for various diseases. We compared the progression stages of these diseases with those of control samples by using data for healthy individuals taken from available databases, and we identified sets of switch circuits possibly responsible for maintaining the persistent disease states by using network topologies to analyze that data.

Author Summary

Since most disease states exhibit a certain level of resilience against therapeutic interventions, each disease state can be considered to be homeostatic to some extent. There must be one or more mechanisms that cause the gene-regulatory network to maintain a certain state, and one such mechanism is a bistable switch. In this work, bistable switch networks were constructed and their ON(upregulated)/OFF(downregulated) states were compared between human cancers and healthy control samples. Changes in the ON/OFF state with the progression of cancer were demonstrated. A series of genes that might serve as a drug target or diagnosis biomarker was identified. The approach presented here should provide useful insights into the states of biological networks, which may lead to the discovery of novel drug targets and therapeutic interventions.

Results

Extraction of bistable toggle switches

There are theoretically many system configurations that can lead to bistability [18,32–35]. We focused on bistable toggle switches (BTSs) with double-negative feedback. Such switches can be constructed from any two genes that mutually repress their expression. We considered three types of network motifs that can exhibit bistable behavior (Fig. 2).

1. Type-1 BTS: A type-1 BTS uses a basic motif that has been identified in E. coli [26] and has mutually inhibitory interaction and positive autoregulators. In a circuit with a double-negative feedback loop, proteins A and protein B inhibit or repress each other. Positive autoregulation is a type of feedback in which proteins directly activate the transcription of their own genes. Under the right circumstances, there could be a stable steady state in which A is "ON" and B is "OFF" or B is "ON" and A is "OFF." This bistability is maintained through positive autoregulation.
2. Type-2 BTS: Only a small number of transcription factors with a positive autoregulation ability have been reported. From the viewpoint of dynamic properties, positive autoregulation has the same functional meaning that a positive feedback loop (double-positive feedback or double-negative feedback) does [36]. We thus defined two mutually inhibitory nodes with a positive feedback loop between them as a type-2 BTS.
3. Type-3 BTS: A theoretical study of modeling genetic switches with positive feedback loops [37] revealed that mutual inhibition is maintained even if a molecule that signals information intervenes between the molecules constituting a switch. We defined two nodes that inhibit each other through other genes (mediators) as a type-3 BTS. Although it is theoretically possible that a positive feedback loop can be formed even if the intervening molecules are identical, in the present study we excluded this possibility.

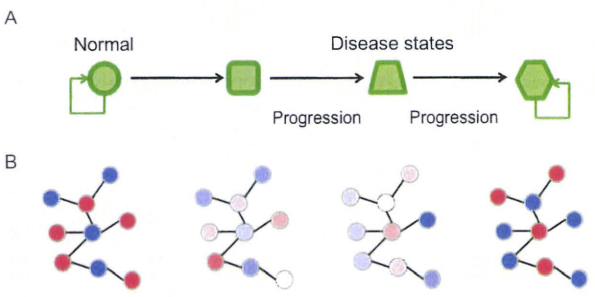


Figure 1. State transitions in network dynamics and disease progression. A: A network in a healthy state is robust against a range of perturbations, so it can continue to maintain a healthy state. With the onset of a disease, however, the network transitions to a new steady state that is also often robust against perturbations, making the disease state persistent. B: These state transitions might be driven by bistable switch networks. The nodes represent genes and the edges between them represent the pairing of bistable toggle switches. Red and blue nodes correspond to ON (upregulated) and OFF (downregulated) states.
doi:10.1371/journal.pcbi.1000851.g001

It is possible that double-negative feedback can be a bistable toggle switch when both nodes have positive feedback loops. Two BTSs can share their mutual inhibition configurations as positive feedback loops and can form network configurations.

Next, bistable toggle switches defined above was extracted from large-scale databases (ResNet 3.0, Ariadne Genomics Inc.) containing data for interaction networks. We detected 6585 pairs of bistable toggle switches, and these switch nodes formed a large network. Four-hundred and forty-two genes are involved in these BTS pairs, and the hubs of switch nodes in the network are clearly visible because of their high degree of connectivity (Fig. 3). A complete list of the BTS pairs is provided in Protocol S1, and a Cytoscape session file is provided in Protocol S2. It should be noted that this network was constructed using text mining and that the molecular details of each interaction were not verified. It is nevertheless a reasonable starting point, and whether or not a listed BTS actually exhibits bistability can be further examined using microarray data.

Tests using mRNA microarray data

ArrayExpress microarray data were used to further examine the states of the BTS pairs. It is obvious that a BTS has four possible states: "ON/ON," "ON/OFF," "OFF/ON," and "OFF/OFF." Mathematical analysis of bistability for the chosen parameter condition demonstrated that the probability of "ON/OFF" and "OFF/ON" states is high, that of "ON/ON" is low, and that of "OFF/OFF" is extremely low [38]. This is the reason we focused on the BTSs that demonstrated "ON/OFF" or "OFF/ON" states.

The ArrayExpress experimental categories and the mean number of corresponding BTS pairs with a significant ON/OFF change are shown in Fig. 4. In the set of 6585 candidate BTSs the number of pairs with a significant ON/OFF change ranged from 0 to 1927 (mean = 298.6), while in a set of 6585 randomly selected gene pairs the number of pairs with a significant ON/OFF change ranged from 0 to 273 (mean = 72.1).

The switching of a molecule's function to the ON state generally means the molecule's intrinsic function related to intracellular molecular systems has become stronger, whereas switching to the OFF state means it has become weaker. The ON state of a molecule is produced not only by an increase in the absolute amount of that molecule but also by actions such as activation due to phosphorylation-induced transformation of the molecule's three-dimensional structure or to translocation of the molecule to an location where it can carry out its function properly.

In these studies using mRNA expression data from microarrays, the toggling of a BTS pair was defined as an instance in which a sample's mRNA level for one of that pair's molecules increased (relative to a control) and the mRNA level for the other of that pair's molecule decreased (relative to the same control).

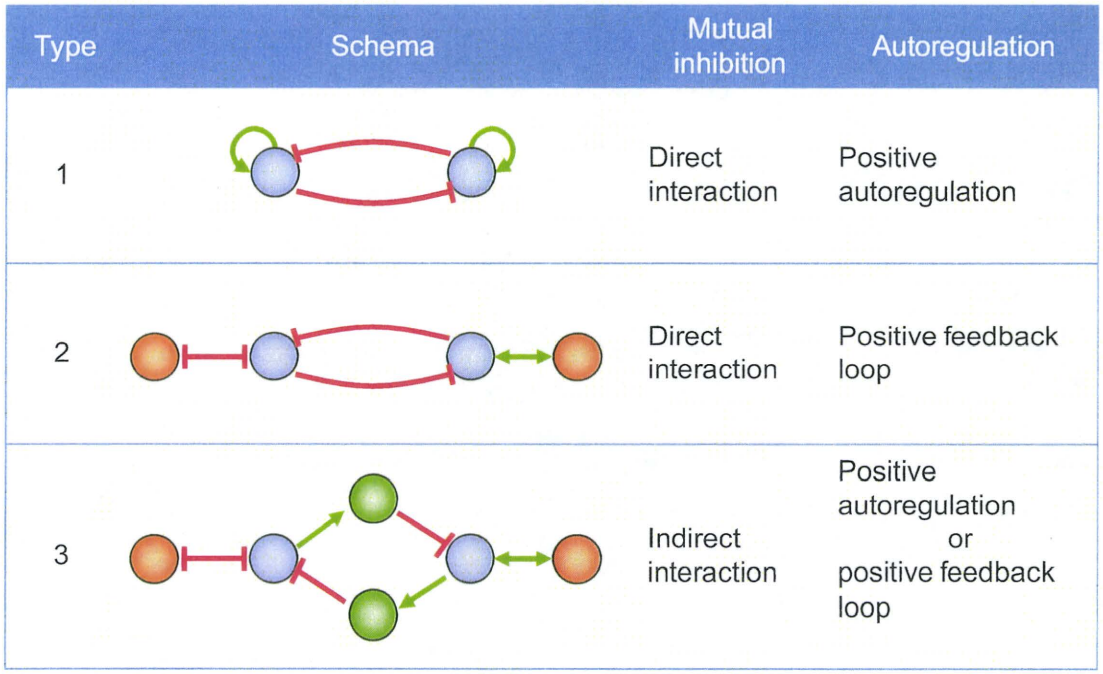


Figure 2. Motifs of bistable toggle switches. A type-1 bistable toggle switch (BTS) contains two genes with positive autoregulation. Each gene mutually inhibits the other's expression. The two genes in the type-2 BTS also suppress each other's expression. Each gene has double positive or negative feedback with the other gene, so the same function as a type-1 BTS may be exhibited. A type-3 BTS was constructed on the basis of a theoretical study on the modeling of genetic switches with positive feedback loops. The blue, green, and orange nodes respectively correspond to switch genes, mediators, and genes constituting a feedback loop. Positive (upregulated) interactions are indicated by green lines and negative (downregulated) interactions are indicated by red lines.
doi:10.1371/journal.pcbi.1000851.g002

A notable finding is that when mRNA levels were compared between induced pluripotent stem (iPS) cells and donor controls, more than 1000 BTS pairs demonstrated significant changes in the ON/OFF states. The high frequency of these changes in iPS cells is reasonable in that an iPS cell is in an undifferentiated state committed to differentiation to a particular lineage, in which many BTSs might be involved [39]. iPS cells have been generated from mouse and human somatic cells by using retroviruses or lentiviruses to introduce Oct3/4 and Sox2 with either Klf4 and c-Myc or Nanog and Lin28 [40]. These factors have been reported to result in bistability when they combine with other factors and form mutual-activation and mutual-inhibition motifs [41–43].

Lung cancer

Lung cancer is the leading cause of cancer-related deaths [44], and tobacco smoking is the strongest etiological factor associated with lung cancer. Prior studies have demonstrated that smoking creates a field of molecular injury throughout the airway epithelium exposed to cigarette smoke [45].

Figure 5A depicts the toggling of BTS ON/OFF states inferred from time-dependent data (ArrayExpress ID: E-GEOD-10700 and E-GEOD-10718) for the mRNA expression in normal human bronchial epithelial cells exposed to cigarette smoke for 24 hours. Toggling began at 2 hours (Fig. 5B) and was observed most frequently at 4 hours (Fig. 5C). SOCS3 (suppressor of cytokine signaling 3) was observed early, while BTSs related to HMOX1 (heme oxygenase 1), CSF2 (colony stimulating factor 2), and SPP1 (secreted phosphoprotein 1) were observed throughout the 24-h period.

SOCS3 inhibits cytokine signaling via the JAK(Janus kinase)/STAT(signal transducers and activators of transcription) pathway. Recent research has demonstrated that the activation of SOCS3 in the lung occurs during the acute inflammatory response [46]. Frequent hypermethylation in the CpG islands of the functional SOCS3 promoter has been found in lung-cancer tissue samples to correlate with its transcription silencing [47]. The OFF states of EGF (epidermal growth factor) and MAPK8 (mitogen-activated protein kinase 8) were linked to the ON states of CSF2 and HMOX1, which became the main players at four or more hours of exposure. CSF2 and HMOX1 were connected through several genes in the OFF state, including IL13 (interleukin 13), IFNG (interferon gamma), and FN1 (fibronectin 1), which are related to inflammatory responses and wound healing.

Figure 6 illustrates the state of BTS toggling for a comparison of mRNA expression (ArrayExpress ID: E-GEOD-10072) in non-small cell lung carcinoma (NSCLC) patients with a history of smoking (Fig. 6A) along with those currently smoking (Fig. 6B) with mRNA expression seen in normal lung tissue. The bold black frames surround molecules that are also in the BTS molecules whose toggling is shown in Fig. 5A.

ON/OFF patterns of FN1-SPP1 (Fig. 6A) and IGF1-SPP1 (Fig. 6B) were observed in the data gathered in experiments exposing normal human bronchial epithelial cells to cigarette smoke. SPP1 is a secreted integrin-binding glycoprotein that is overexpressed in various tumors and has been reported to be involved in tumorigenesis and metastasis. High expression of SPP1 is a significantly unfavorable prognostic factor for the survival of patients with NSCLC [48].

In addition, although some EDN1(endothelin-1)-related BTS pairs and SHC1(Src homology 2 domain containing transforming protein)-related BTS pairs are shared in lung cancer tissue in current and former smokers, a considerable number of differing

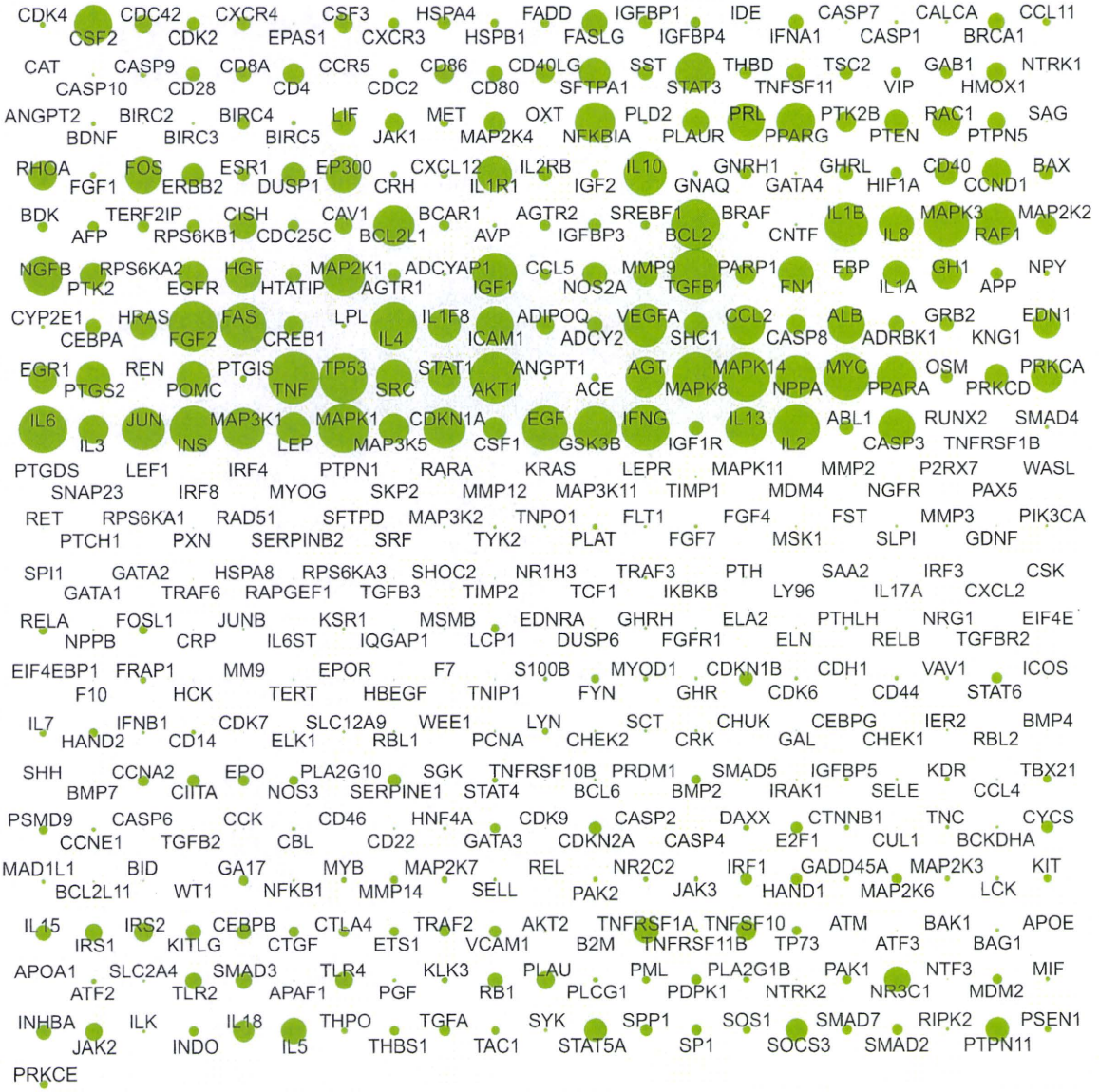


Figure 3. Cytoscape visualization of network composed of bistable toggle switch pairs. Four-hundred and forty-two genes are involved in 6585 bistable toggle switch pairs. Nodes are shown in sizes proportional to their connectivity, making the hubs of switch nodes clearly visible. The Cytoscape session file for this network is available in Protocol S2.
doi:10.1371/journal.pcbi.1000851.g003

patterns are evident. This suggests that the mechanisms for carcinogenesis differ depending on the lengths of time that current and former smokers have smoked. EDN1, which is a hypoxia-inducible angiogenic growth factor for surrounding epithelial and endothelial cells, plays an important role in cancer-stromal interactions and tumor progression, and its expression is related to poor prognosis in NSCLC [49].

Small molecules that can put these BTS pairs into normal ON/OFF states might be useful in preventing the progression of lung cancer in both current and former smokers.

Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is a primary cancer that originates in hepatocytes and typically follows cirrhosis or chronic-hepatitis virus infections [50], and the most significant risk factors for HCC are chronic infections with either hepatitis B virus or hepatitis C virus (HCV).

Figure 7 is a BTS toggling graph in which mRNA expression data (ArrayExpress ID: E-GEOD-6764, [51]) for tissues from patients with HCV-induced dysplasia and HCC are compared with mRNA expression data for normal liver tissue. The molecules surrounded by bold lines are BTSs for which toggling was observed when comparing dysplastic liver tissue (cirrhotic tissue and dysplastic nodules), a precursor of liver cancer, with normal liver tissue. The two tissue types share many BTSs associated with PTGS2 (prostaglandin-endoperoxide synthase 2; COX-2) and IL1B (interleukin 1, beta). It has been demonstrated that the expression pattern of PTGS2, a key enzyme of the