

Fallot and Pentalogy of Cantrell [30,31]. It is also significant in the context of our research program, because *Aldh1a2* is an essential gene for heart development [32,33], and its presence in the cardiac fibroblasts strongly reinforces our recent findings of a cardiogenic gene expression signature in the normal adult cardiac fibroblasts [13]. We therefore used qPCR to confirm that *Aldh1a2* is indeed strongly up-regulated in heart fibroblasts, but not in tail fibroblasts or the whole heart (Figure 3B, and Additional file 1: Text S1).

In addition to our customised TF layout algorithm, we also provide implementations of popular open-source layout algorithms from the NodeXL API, including Fruchterman-Reingold [34], Harel-Koren [35], and Sugiyama [36]. For the Gata4-Tbx20 case study, the customised VISIONET layout method was substantially faster than other layout methods (Figure 3C), while still providing a human-readable display (Figure 3A).

Discussion

Transcription factors often act in concert to co-regulate genes, and therefore the overlap of multiple TF networks carries important biological implications [7,8,37]. VISIONET is designed to be a simple but powerful visualisation tool for biologists to study these overlapping networks. Our tool brings together a number of key features that enable expertise-driven discovery, including allowing biologist users to interact directly with the TF network and determine the components of the network that are highly relevant to the biological question. The overlapping TF networks can be supplied, or derived, by using straightforward algorithms, and filtering based on microarray data can subsequently reduce the complexity of the network.

VISIONET was developed in-house and tailored to our biological question of cardiac fibroblasts characterisation,

but is nevertheless broadly suitable for expertise-driven biological discovery. In a VISIONET graph, the network topology represents the transcription factor networks constructed from ChIP-Seq datasets, and the node colours represent the transcriptomic profile obtained from microarray experiments. The transcriptomic profile can then be used as a filter to limit the number of nodes visible to the biologist users, allowing the identification of important genes based on human expertise. Although microarray values may not be easily inferred solely from the node colours, users can interactively set the threshold for VISIONET to only display nodes above (or below) a certain value. By limiting the number of visible genes in the network, researchers can then apply their expertise to identify strongly relevant or unexpected transcripts, as we demonstrated with *Aldh1a2* (Figure 3A-B) and *Hand2* [3].

Being custom-made, VISIONET has features not yet addressed by other popular general-purpose visualization platforms, including CellDesigner and Cytoscape. There exists a myriad of visualisation platforms that provides a partial list of features provided by VISIONET (Figure 1A), but none have provided the streamlined user experience tailored to the specific task of expertise-driven discovery, as provided by VISIONET. This has been reinforced by our experience of uncovering the common target of Tbx20 and Gata4 in the adult cardiac fibroblasts (Figure 3A-B), where the discovery of *Aldh1a2* would not have been achieved using other visualisation platforms (Figure 2).

The majority of Tbx20 targets (58%) are also Gata4 targets (Figure 2B), consistent with other studies showing that Gata4 and Tbx20 interact to regulate heart structure and function in development [14,27]. The identification of the *Aldh1a2* gene as a target of Tbx20 based on ChIP-Seq data agrees with our previous findings that cardiac fibroblasts in Tbx20 conditional knockout mice displayed reduced

Aldh1a2 activity [13]. Furthermore, *Aldh1a2* is an established direct downstream target of *Gata4* [38,39], as confirmed by our Tbx20-*Gata4* case study using VISIONET.

Biological network visualisation has been an active area of research, and VISIONET is also designed to continuously improve its limitations and to adapt to the changing technological landscape. In our *Gata4*-Tbx20 case study, we have applied VISIONET for two overlapping TF networks (Figure 3A-B) due to the nature of the experiment, but the customised layout algorithm can also accommodate any larger number (>2) of TF overlapping networks. We currently implemented VISIONET using the NodeXL API for the node filtering support and web accessibility. Since the popular platforms CellDesigner [40] and Cytoscape are widely used, future work will also develop VISIONET as a plug-in for CellDesigner and/or Cytoscape. Also, as VISIONET depends on input CHIP-Seq and microarray data, it also inherits the technological pitfalls of these transcriptomic technologies. The peak list of CHIP-Seq contains uncertainties in the TF target gene information, and our microarray values may not indicate accurately the *in vivo* gene expression level [41].

Conclusions

VISIONET is an in-house and streamlined tool for the specific tasks of expertise-driven gene discovery, with an implementation of features that are not all concurrently available in the popular comprehensive analysis platforms. In concert with other computational tools under development by the systems biology community, VISIONET bridges the gap between complex dataset and biologist users for a better understanding of biological systems.

Availability and requirements

Project name: VISIONET

Project home page: <http://VISIONET.erc.monash.edu.au>

Source code, desktop companion software and tutorials: <http://VISIONET.erc.monash.edu.au>

Operating system(s): Platform independent

Programming language: C#, ASP.NET

Other requirements (desktop application only): Microsoft Excel 2007 or later.

License: GNU GPL

Any restrictions to use by non-academics: none

Additional file

Additional file 1: Supplementary text and table.

Abbreviations

VISIONET: Visualisation of overlapping transcription factor network; TF: Transcription factor; CHIP-Seq: Chromatin immunoprecipitation sequencing; qPCR: quantitative real-time polymerase chain reaction.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

HTN and SEB designed the research. HTN developed the algorithm, implemented the web server, and performed data analyses. MBF prepared the animal samples and MWC performed the qPCR experiments. NAR and HK provided expertise in cardiac biology, and visualisation and layout, respectively. HTN prepared the manuscript. All authors edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Jerico Revote for the web server support, and Prof. Falk Schreiber for very helpful critique and comments.

Funding

ARC Discovery Grant DP130104792 to SEB and HK, and an NH&MRC Australia Fellowship to NR. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

Author details

¹Systems Biology Institute (SBI) Australia, Monash University, Clayton, VIC 3800, Australia. ²Australian Regenerative Medicine Institute, Monash University, Clayton, VIC 3800, Australia. ³National Heart and Lung Institute, Imperial College London, London W12 0NN, UK. ⁴Sony Computer Science Laboratories, Inc., Higashigotanda, Shinagawa, Tokyo, Japan. ⁵Okinawa Institute of Science and Technology, Onna, Onna-son, Kunigami, Okinawa, Japan.

Received: 9 January 2015 Accepted: 20 April 2015

Published online: 01 May 2015

References

- Ghosh S, Matsuoka Y, Asai Y, Hsin KY, Kitano H. Software for systems biology: from tools to integrated platforms. *Nat Rev Gen.* 2011;12(12):821–32.
- Hurley D, Araki H, Tamada Y, Dunmore B, Sanders D, Humphreys S, et al. Gene network inference and visualization tools for biologists: application to new human transcriptome datasets. *Nucleic Acids Res.* 2012;40(6):2377–98.
- Nim HT, Boyd SE, Rosenthal NA. Systems Approaches in Integrative Cardiac Biology: Illustrations from Cardiac Heterocellular Signalling Studies. *Progr Biophys Mol Biol.* 2014. doi:10.1016/j.pbiomolbio.2014.10.111.1006.
- Alworth SV, Watanabe H, Lee JS. Teachable, high-content analytics for live-cell, phase contrast movies. *J Biomol Screen.* 2010;15:968–77.
- Kikuta J, Wada Y, Kowada T, Wang Z, Sun-Wada GH, Nishiyama I, et al. Dynamic visualization of RANKL and Th17-mediated osteoclast function. *J Clin Invest.* 2013;123:866–73.
- Biggin MD. Animal transcription networks as highly connected, quantitative continua. *Dev Cell.* 2011;21:611–26.
- He A, Kong SW, Ma Q, Pu WT. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A.* 2011;108(14):5632–7.
- Schlesinger J, Schueler M, Grunert M, Fischer JJ, Zhang Q, Krueger T, et al. The cardiac transcription network modulated by *Gata4*, *Mef2a*, *Nkx2.5*, *Srf*, histone modifications, and microRNAs. *PLoS Genet.* 2011;7(2):e1001313.
- Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. *Nat Methods.* 2010;7(3 Suppl):S56–68.
- Palsson B, Zengler K. The challenges of integrating multi-omic data sets. *Nat Chem Biol.* 2010;6(11):787–9.
- von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, van Wijk JJ, Fekete JD, et al. Visual analysis of large graphs: state-of-the-art and future research challenges. *Comput Graph Forum.* 2011;30(6):1719–49.
- Lusis AJ, Weiss JN. Cardiovascular networks: systems-based approaches to cardiovascular disease. *Circulation.* 2010;121(1):157–70.
- Furtado MB, Costa MW, Pranoto EA, Salimova E, Pinto AR, Lam NT, et al. Cardiac genes expressed in cardiac fibroblasts contribute to heart development and repair. *Circ Res.* 2014;114(9):1422–34.
- Stennard FA, Costa MW, Elliott DA, Rankin S, Haast SJ, Lai D, et al. Cardiac T-box factor Tbx20 directly interacts with *Nkx2-5*, *GATA4*, and *GATA5* in

- regulation of gene expression in the developing heart. *Dev Biol.* 2003;262:206–24.
15. Sakabe NJ, Aneas I, Shen T, Shokri L, Park SY, Bulyk ML, et al. Dual transcriptional activator and repressor roles of TBX20 regulate adult cardiac structure and function. *Hum Mol Genet.* 2012;21:2194–204.
 16. Duester G. Retinoic acid synthesis and signaling during early organogenesis. *Cell.* 2008;134:921–31.
 17. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H. CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc IEEE.* 2008;96(8):1254–65.
 18. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H. Payao: a community platform for SBML pathway model curation. *Bioinformatics.* 2010;26(10):1381–3.
 19. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* 2010;27:431–2.
 20. Hu Z, Chang YC, Wang Y, Huang CL, Liu Y, Tian F, et al. VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.* 2013;41:W225–31.
 21. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol.* 2008;6:e184.
 22. Hansen D, Smith MA, Shneiderman B. Analyzing Social Media Networks with NodeXL: Insights from a Connected World. 1st ed. Burlington, Massachusetts: Morgan Kaufmann; 2010.
 23. Brandes U, Eiglsperger M, Herman I, Himsolt M, Marshall MS. GraphML progress report structural layer proposal. In: Mutzel P, Jünger M, Leipert S, editors. *Graph drawing*, vol. 2265. Berlin Heidelberg: Springer; 2002. p. 501–12.
 24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
 25. Takeuchi JK, Mileikovskykaia M, Koshiba-Takeuchi K, Heidt AB, Mori AD, Arruda EP, et al. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. *Deve (Cambridge England).* 2005;132:2463–74.
 26. Oka T, Mailliet M, Watt AJ, Schwartz RJ, Aronow BJ, Duncan SA, et al. Cardiac-specific deletion of Gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circ Res.* 2006;98:837–45.
 27. Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, et al. Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Dev (Cambridge England).* 2005;132:2451–62.
 28. Shen T, Aneas I, Sakabe N, Dirschinger RJ, Wang G, Smemo S, et al. Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function. *J Clin Invest.* 2011;121:4640–54.
 29. Leda M, Fu JD, Delgado Olguin P, Vedantham V, Hayashi Y, Bruneau BG, et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. In: *Cel*, vol. 142. United States: Elsevier Inc; 2010. p. 375–86.
 30. Steiner MB, Vengoechea J, Collins 2nd RT. Duplication of the ALDH1A2 gene in association with pentalogy of Cantrell: a case report. *J Med Case Rep.* 2014;7:287.
 31. Pavan M, Ruiz VF, Silva FA, Sobreira TJ, Cravo RM, Vasconcelos M, et al. ALDH1A2 (RALDH2) genetic variation in human congenital heart disease. *BMC Med Gen.* 2009;10:113.
 32. Kikuchi K, Holdway JE, Major RJ, Blum N, Dahn RD, Begemann G, et al. Retinoic acid production by endocardium and epicardium is an injury response essential for zebrafish heart regeneration. *Dev Cell.* 2011;20:397–404.
 33. Castillo HA, Cravo RM, Azambuja AP, Simoes Costa MS, Sura Trueba S, Gonzalez J, et al. Insights into the organization of dorsal spinal cord pathways from an evolutionarily conserved raldh2 intronic enhancer. *Dev (Cambridge England).* 2010;137:507–18.
 34. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exp.* 1991;21(11):1129–64.
 35. Harel D, Koren Y. A Fast Multi-scale Method for Drawing Large Graphs. In: *Proceedings of the 8th International Symposium on Graph Drawing*. 729397: Springer-Verlag 2001: 183–196.
 36. Sugiyama K, Tagawa S, Toda M. Methods for visual understanding of hierarchical system structures. *Syst Man Cybernet, IEEE Transac.* 1981;11(2):109–25.
 37. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell.* 2010;140(5):744–52.
 38. Xuan S, Borok MJ, Decker KJ, Battle MA, Duncan SA, Hale MA, et al. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J Clin Invest.* 2012;122:3516–28.
 39. Liang D, Jia W, Li J, Li K, Zhao Q. Retinoic acid signaling plays a restrictive role in zebrafish primitive myelopoiesis. *PLoS One.* 2012;7:e30865.
 40. Funahashi A, Morohashi M, Kitano H, Tanimura N. Cell Designer: a process diagram editor for gene-regulatory and biochemical networks. *Drug Discov Today BIOSILICO.* 2003;1(5):159–62.
 41. Furtado MB, Nim HT, Gould JA, Costa MW, Rosenthal NA, Boyd SE. Microarray profiling to analyse adult cardiac fibroblast identity. *Genomics Data.* 2014;2:345–50.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



システム・トキシコロジーの展開

北野 宏明 博士 (特定非営利活動法人 システム・バイオロジー研究機構 代表)

毒性学 (toxicology) は、我々の健康を守り、薬の安全性を担保し、安全な化粧品や加工食品を世に送り出すために必要不可欠な研究分野である。しかし、その対象範囲があまりに広範なことで未解明の毒性出現機序が数多く残されているために、ともすると後追いの展開を余儀なくされてきた。しかし近年、企業のリスク管理の観点もあって、毒性を事前に予測することへの関心が急速に高まってきている。今回から、システムバイオロジーの応用領域の一つである毒性学を取り上げ、どのようにこの問題にアプローチするべきか議論していきたい。

最近我々の研究機構 (SBI) に共同研究などのリクエストが急速に増えているのが毒性分野である。創薬分野では、候補化合物の副作用の予測が重要な課題である。臨床試験が進んだ段階や上市後に重篤な副作用が見つかる大きな経営リスクとなるので、少しでもそれを事前に回避したいのである。また、化粧品などの分野では、動物実験ができない (EU では禁止されている) という問題があり、毒性の計算予測への期待は特に大きい。これらの状況から、システム・トキシコロジー (Systems Toxicology) は我々の中核研究分野となっている。

創薬などで候補化合物の有効性を予測する場合、対象となる細胞や組織を想定し、細胞の増殖抑制などのエンドポイント (評価項目) を設定する。ここで一番大切なのは、設定されたエンドポイントを満たすかどうかであって、少なくとも最初の段階ではそれ以外の要因は考慮しない。しかし、毒性予測はこれとは対極である。

つまり、遺伝的にもエピジェネティックにおいても不均質な集団を対象として、極めて多様な各臓器とその細胞種を調べ、どれかに大きな問題が発生すれば、その物質は大きな毒性リスクを持つ、とするのである。これを実行・実現するには、極めて広範かつ網羅的なデータや知識ベースなどが整備されている必要がある。

我々はいままで、何社もの民間企業や国立医薬品食品衛生研究所 (NIHS)、米国食品医薬品局 (FDA) などと共同研究を進めてきた。これまでの経験から、毒性分野は公共性の極めて高い分野であると感じている。とともに、多くの課題が残された分野であることを知った。例えば、極めて広範で質の高いデータベース、モデル、実証系などが必要であるが、これは個別の企業努力で整備する規模をはるかに超えている。また毒性を事前に予測するには、毒性発現の機序を理解することが重要であるが、これには多くの研究者によるシステムティックな対応が必要になる。



北野 宏明 博士

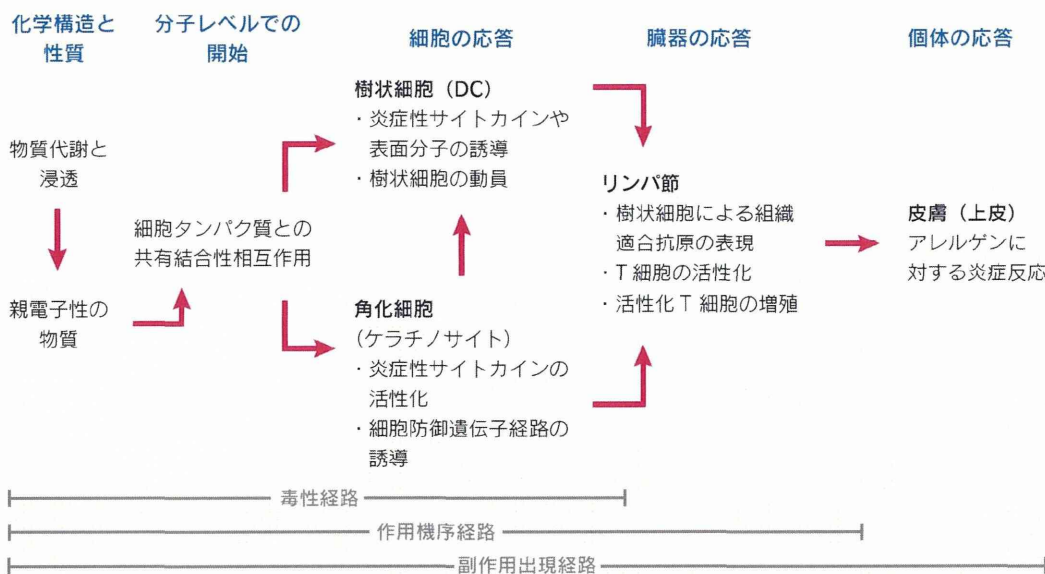


図 1. OECD-AOP の枠組みの一例 ((OECD, 2012) より)

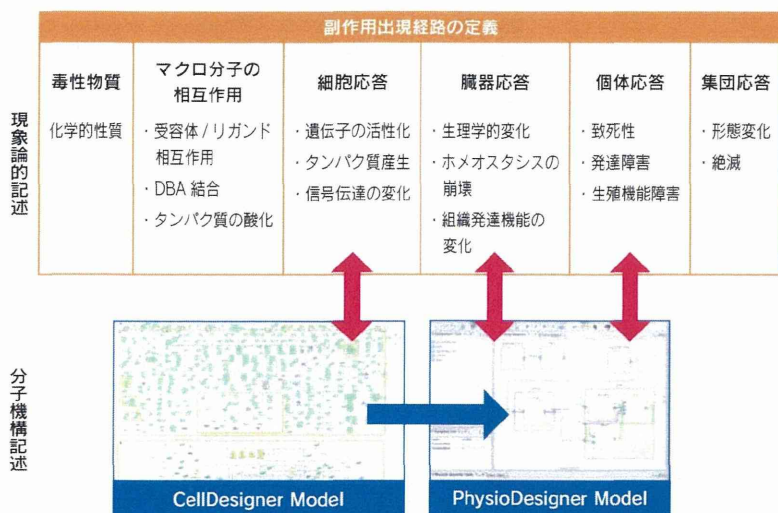


図 2. 現象論記述の OECD-AOP と分子機構に立脚するモデルの統合を行なう

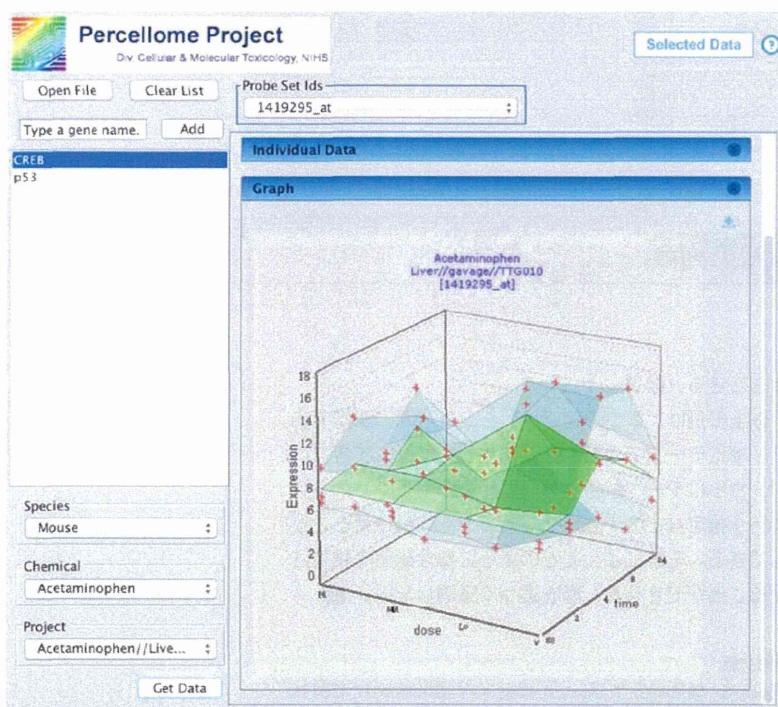


図 3. Garuda Application として実装された PerCellome のスクリーンショット。ここでは、マウスに対して Acetaminophen を暴露した場合の肝臓における CREB の発現変化のデータを可視化している。

さらに、生理現象として毒性が顕在化する機序についても、十分には解明されていると言えない。特に、低用量長期反復暴露で引き起こされる問題に関しては、どのようにその機序を解明し、予測を行なうかは大きな課題である。これら多くの問題を解決するために、まずはしっかりした研究基盤を構築する必要があると考えた。

OECD-AOP との連動

そこで、我々は Systems Toxicology Platform を、既に紹介した Garuda Platform (Ghosh, Matsuoka, Asai, Hsin, & Kitano, 2011) 上に開発することとした。そして、一般化学物質の毒性予測に関しては、OECD が国際標準化を目指して進めている AOP (Adverse Outcome Pathway: 副作用出現経路) との連動性を重視して、その枠組みとの一貫性と相互補完性を検討した。そこで判明したことは、OECD-AOP では、毒性物質から始まり分子相互作用、細胞応答、臓器応答、個体応答、集団応答の階層が定義されている (図 1)。しかし、その具体的記述をみると、各々の階層での記述は、関連論文における記載を文章のまま列挙する形式である ((OECD, 2012)などを参照のこと)。つまり、その背後にある分子機構などを一貫した体系として記述し、さらにシミュレーションや統計解析などへと橋渡しするという枠組みにはなっていない。これでは、既知の毒性に関する百科事典的記述とは異なるが、毒性予測を行なうことはできない。

そこで我々は、OECD-AOP の記述的な部分に対して、その背後にある分子機構やネットワーク動態などの部分を計算予測やシミュレーション、統計解析が可能な形式で記述し、OECD-AOP と統合する方針を定めた (図 2)。

この場合、分子機構を記述するパスウェイ (反応経路) などは CellDesigner を用いてモデル化し、それらを統合した臓器レベル、個体レベルのモデルは PhysioDesigner を利用することを前提とする。これらのモデリング・ツールに対し、一連の解析ツール、シミュレーション・ツール群が Garuda Platform 上で連動することで、包括的な毒性解析プラットフォームとなると考えている。

Garuda Platform 上での毒性関連ツール群

まず、Garuda Platform 上に既に提供されている毒性関係のツール群を紹介しよう。まず、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の提供する Percellome データベースがある。Percellome

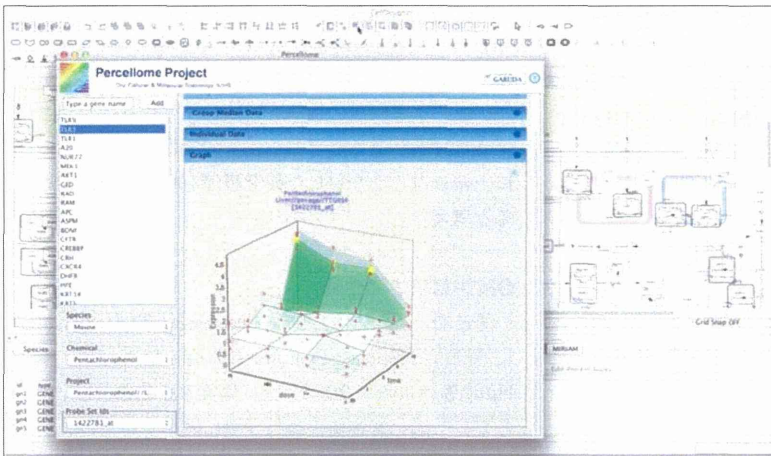


図 4. Garuda Platform 上での、CellDesigner と PerCellome の連携。背景に、CellDesigner による TLR Pathway map が表示されて、PerCellome の左側には、TLR Pathway に関わる遺伝子のリストが送られている。

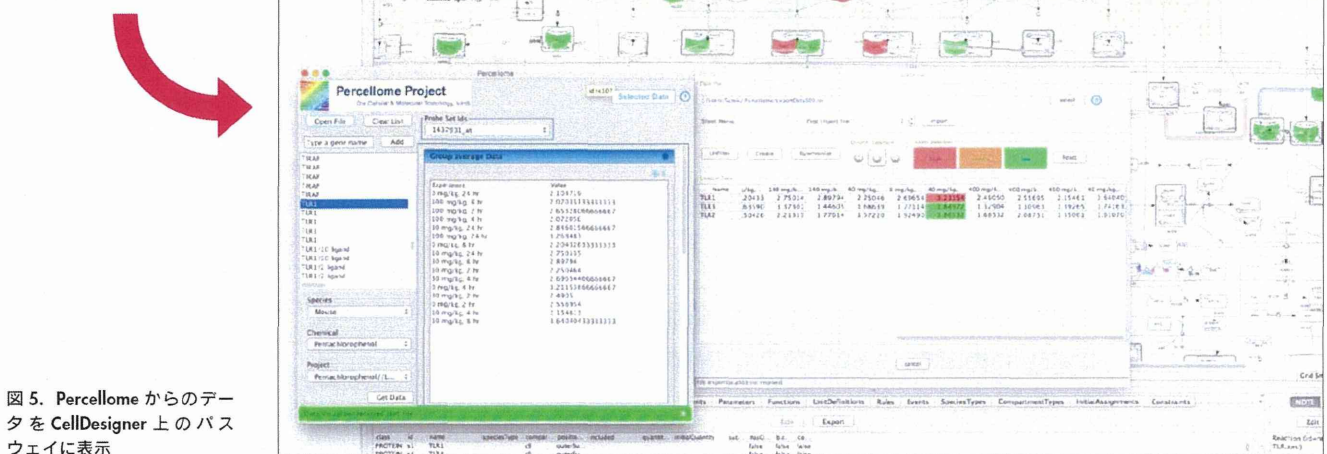


図 5. PerCellome からのデータを CellDesigner 上のパスウェイに表示

は、マウスに化学物質を暴露させた場合の遺伝子発現の変化を高精度で測定する手法である (Kanno et al., 2006)。Garuda Platform 上で、単純に PerCellome を呼び出し、そこにあるデータの一部を表示させたのが図 3 である。

Garuda Platform では、他のツールとの連携が可能である。例えば、TLR シグナリングネットワークに関わる遺伝子群に対して、特定の化学物質がどのような影響を与えるか調べたいとしよう。この場合、まず CellDesigner で、TLR Pathway map をオープンする。その上で、そこにあるタンパク質のリストを抽出し、それを PerCellome へと送る。するとそのリストが PerCellome の遺伝子リストに現れるので、その中から調べたい遺伝子をクリックすると結果が表示される (図 4)。こんな便利な使い方ができる。

さらに、PerCellome から取得したデータを、CellDesigner 上の TLR パスウェイにマッピングするこ

ともできる (図 5)。

以上の例は、2 つのアプリケーション間の連携であり、単にデータアクセスを効率化しただけにすぎないが、さらにデータ解析、分子間相互作用予測、ネットワーク推定などのアプリケーションを連動させることもできる。それによってどのように毒性研究を推進させることができるか、順を追って議論したい。 ■

参考文献

Ghosh, S., Matsucka, Y., Asai, Y., Hsin, K. Y., & Kitano, H. (2011). Software for systems biology: from tools to integrated platforms. *Nat Rev Genet*, 12(12), 821-832. doi: 10.1038/nrg3096

Kanno, J., Aisaki, K., Igarashi, K., Nakatsu, N., Ono, A., Kodama, Y., & Nagao, T. (2006). "Per cell" normalization method for mRNA measurement by quantitative PCR and microarrays. *BMC Genomics*, 7, 64. doi: 10.1186/1471-2164-7-64

OECD. (2012). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins, Part 1: Scientific Evidence. In OECD Environmental Directorate (Ed.), *Series on Testing and Assessment*, No. 168. Paris: OECD.

システム・トキシコロジーの展開（第2回）

北野 宏明 博士（特定非営利活動法人 システム・バイオロジー研究機構 代表）

前回は、薬はもちろん化粧品や加工食品などの安全性を担保し、我々の健康を守る上で重要な役割を果たしている毒性学について紹介し、それがシステム・バイオロジーの中でいかに展開されつつあるか、その全体像を紹介した。今回は一歩踏み込んで、実際に解析する場合を取り上げ、どのような考え方とどんな仕組みで進められているかを紹介する。

薬効予測と毒性予測を比較した場合、システム・トキシコロジーやイン・シリコ解析においては、基盤となっている計算上の要素技術には共通するものが多いものの、全体の解析フローは、大きく異なっている。

薬効予測の場合、ある化合物が、定義されたエンドポイント（評価項目）を達成するかどうかを予測するが、この際に、この化合物と、ターゲットとなりうる生体分子群との相互作用の予測、これらが複数のパスウェイ（反応経路）に与える影響、その結果としてフェノタイプ（表現型）に与える影響を追っていくことになる。つまり、化合物→生体分子群→パスウェイ→フェノタイプという順方向に（つまり生体内での反応に沿った方向に）、フォワード予測を行なうことになる（図1）。

一方、毒性予測の場合、生体内のどこかで、副作用の可能性となる何らかの現象が引き起こされるなら、要注意リストに取り上げて、集中的に毒性リスクの解析をする必要がある。しかし、ここでフォワード予測を採用するのは現実的ではない。すべての予測されるフェノタイプに対して詳細解析をすることになってしまい、解析コストの観点からも、現実的ではないからだ。実際には、経験上、最初に確認しておくべき副作用群というものがあるので、イン・シリコ解析ではまず、候補化合物が、これらの副作用を引き起こす可能性があるかどうか、調べる。

この最初のステップのところ、フォワード解析とは逆方向のバックワード解析を導入する（図2）。つまり、まず副作用フェノタイプ群を特定し、そこから上流方向へとさかのぼっていくのである。具体的に言うと、一連のまず確認すべき副作用群の各々に対して、それに関連する（因果関係があると思われる）遺伝子群と転写因子群を同定する。これによって、遺伝子発現データなどいろいろなデータを使った解析が可能になる。なお、遺伝子発現や転写因子群の代わりに、メタボロームデータ（代謝物質の網羅的データ）を利用するのも非常に有効な手段である。

遺伝子発現や転写因子群を使った場合、この段階で得られるのは、各々の副作用に関連する蓋然性の高い

遺伝子と転写因子のリストとなる。このリストには、蓋然性の尺度として、信頼度の値が各遺伝子に付与されている。

最終的なリスク解析結果の求め方

次の段階では、候補化合物がこれらの遺伝子群や転写因子群に影響を与えるかどうかを解析する。この段階では、フォワード解析に近い手法が用いられる。具体的には、毒性をチェックしようとしている化学物質が、一連の転写因子の上流域にあるパスウェイ群上のどの分子に対して相互作用を持つのかを解析し、その影響に関してパスウェイ解析を行なって、リスク評価する（図3）。

一連の生体内分子に対して候補化合物が相互作用を持つかどうか、純粋にイン・シリコ解析で進めるとすると、このあとは、網羅的なドッキングシミュレーションなどを利用する。我々が開発しているドッキングシミュレーション・システムは、「複数のドッキングシミュレーションアルゴリズムを並列に実行し、事前に行なった機械学習の結果から、各々のドッキングペアに対して最も信頼度の高い推定を行なうと思われるアルゴリズムの結果を軸に、最終的な推定結果 ▶



北野 宏明 博士

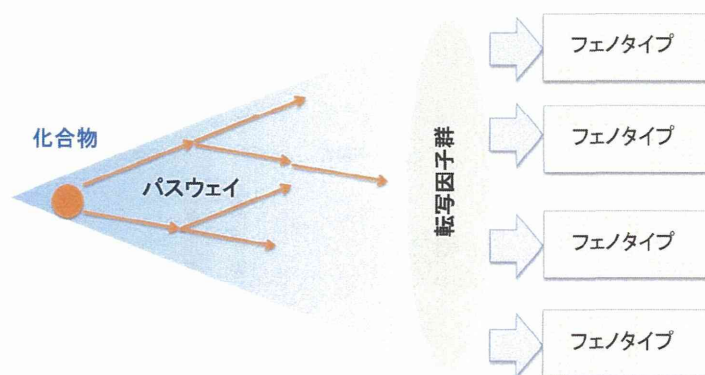


図1. フォワード解析（この場合のフェノタイプ（表現型）とは、細胞増殖、細胞死、特定タンパク質の生産量などのことで、調べる化合物がそれらを変化させるかどうか、予測する。）

を得る」という方法になっている。それを、候補化合物とパスウェイ全体の分子に対して網羅的に実行し、推定を行なう。その結果、パスウェイ上の各々の分子には、候補化合物との相互作用強度と信頼度が表示される(図4)。このあと、総合的な影響の推定が行なわれる。

この方式の1つのポイントは、遺伝子発現・転写因子を境に、解析の方向を分けているということであり、フェノタイプが出現するかどうかを判定する中間表現として、遺伝子・転写因子の動態を利用しているということである。もちろん、サイトカインやプロテオーム、メタボロームをこれに加えることもできるし、原理的には、それが望ましい。現実には、今現在で最も網羅的かつ容易にデータが蓄積可能なのが遺伝子発現レベルなので、それを利用しているということである。

この一連のプロセスにおいて、副作用リスクがあると判断された候補化合物は、その段階で候補から排除されるか、実験でリスクを確認する作業に入る。

このスクリーニングで副作用リスクが認められない候補化合物群は、次の解析ステージに進むが、この最

初のステージで分かることは、以下の点である。すなわち、「イン・シリコ解析によれば、候補化合物単独では、事前に設定された一連の副作用を引き起こす可能性が低いと推定される」ということである。これを「単独スクリーニング」と呼ぶ。この段階では、事前に設定された以外の副作用や複数の化合物の相互作用による副作用などのチェックは、まだなされていない。また、個別の遺伝的背景に起因する副作用については、解析データやモデルにその遺伝的違いを組み込んでいくかいないか、つまり設定条件に依存する。

複数スクリーニング

次のプロセスは、1つは、想定される副作用を大きくして、同じプロセスをくり返してみることである。そして2つ目は、複数の化合物の組み合わせの効果を推定することである。

単一の化合物では毒性が顕在化しないが、複数の化合物に曝露されることで、毒性が引き起こされることはありうる。そんな場合に対応したイン・シリコ解析がこの推定で、「複合スクリーニング・プロセス」と

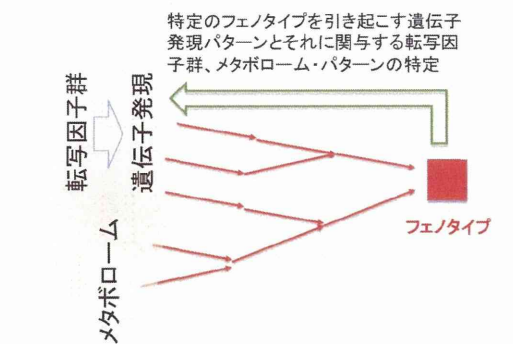


図2. バックワード解析

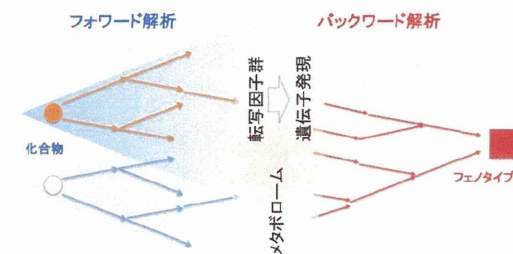


図3. フォワード解析とバックワード解析の融合

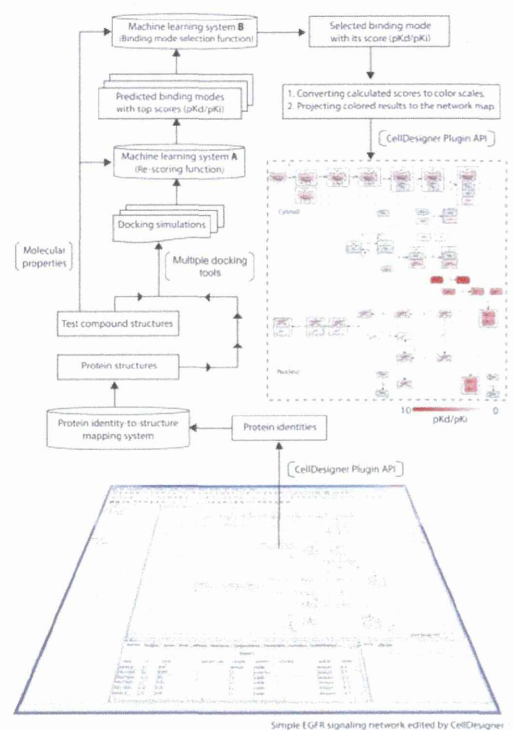


図4. 網羅的高精度ドッキングシミュレーション
Hsin, K.Y., Ghosh, S., and Kitano, H., PLoS One, Dec. 31, 2013; DOI: 10.1371/journal.pone.0083922

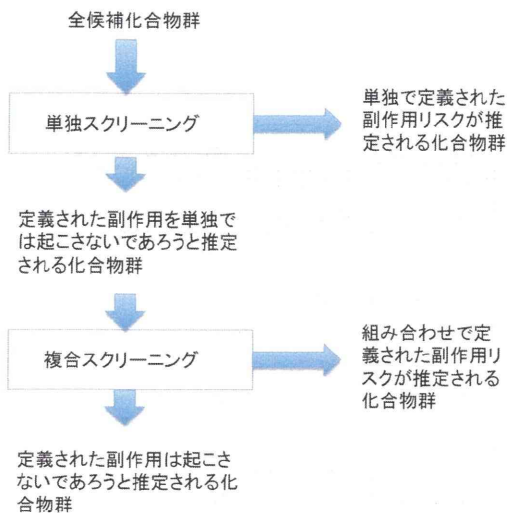


図5. 多段階スクリーニング・プロセス

呼ぶ。この解析の対象となる化合物は、単独では毒性が推定されていないものであり、他の化合物との併用で毒性が生じる可能性を推定する。得られる情報は、「避けるべき組み合わせのリスト」ということになる。こうした分析の結果、組み合わせで使用しても毒性が生じないと推定される化合物群のリストが生き残る。

全体の流れは、図5にあるような多段階スクリーニング・プロセスとなる。

イン・シリコ解析の意義

我々の多段階スクリーニング・プロセスは、前回述べた OECD-AOP と親和性が高く、密な連動が可能である (OECD-AOP は、OECD が国際標準化を進めている副作用出現経路 AOP のこと)。図6に示すように、OECD-AOP の「臓器応答、個体応答、集団応答」から検討を必要とするフェノタイプ群を選択し、毒性物質のリストからスクリーニング対象となる物質のリストが生成される。バックワード解析から導出される遺伝子発現やメタボロームの特徴的変動は、OECD-AOP の「細胞応答」から「臓器応答」の一部までの基本データと対応する。また、フォワード解析の部分は、おもに「マクロ分子の相互作用」の部分に相当する。

この解析手法はすべて、このコラムで何度も紹介している Garuda Platform 上に実装され、製薬企業などへの解析サービスなどに実際に利用されている。

創薬という視点からは、各々の解析において対象となる候補化合物やその組み合わせは、あまり大きな

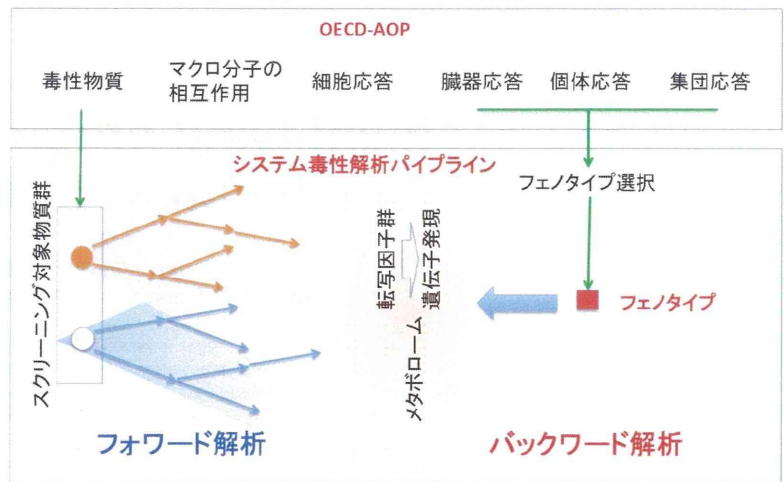


図6. OECD-AOP と多段階スクリーニング・プロセスの関係

い。しかし、非常に多くの物質が同時に利用されるケースもある。それが例えば漢方薬や化粧品などの場合で、同時に生体内に導入される物質の種類は非常に大きな数になる。この場合、実験的に検証するには膨大なコストがかかる。そこで、計算的手法への期待が高まる。現実の問題として、コストなどの理由から、複数物質による相乗作用 (シナジー) の影響については、十分には調べられていない。その意味でも、イン・シリコ解析プロセスが稼働していることは、きわめて大きな意義があると思っている。

イン・シリコ解析でも、組み合わせの数は、コストを引き上げてしまう重要な要因である。例えば 100 個の化合物群から 2 種類の化合物を取り出してチェックするだけでも、約 5000 組のペアが必要になる。

さらに、各々の組み合わせにおいて、用量の違いによって複合的な影響の出る様子が変わってくる。各々の組み合わせに対して、4 点の用量でスクリーニングするとして、 $4 \times 4 = 16$ 回の推定が必要になる。つまり、合わせて約 8 万回の推定プロセスを実行する必要がある。現段階で、このプロセス結果が完全に自動的に出てくるといった状況にはない。我々は、この推定プロセスの自動化レベルを向上させ、さらに推定精度を高める研究に取り組んでいる。

注意すべきは、これらのプロセスには、時間軸の議論が入っていないことだ。実際には、短期的な毒性と同時に、長期暴露による毒性の問題も存在する。これらの問題に対する解析方法の開発も必要である。□



Identifying problematic drugs based on the characteristics of their targets

Tiago J. S. Lopes^{1,2,3*}, Jason E. Shoemaker^{1,3}, Yukiko Matsuoka^{1,4},
Yoshihiro Kawaoka^{1,2,3} and Hiroaki Kitano^{1,4,5,6,7*}

¹ Japan Science and Technology Agency ERATO Kawaoka Infection-Induced Host Responses Project, Minato-ku, Japan, ² Department of Pathobiological Sciences, School of Veterinary Medicine, Influenza Research Institute, University of Wisconsin-Madison, Madison, WI, USA, ³ Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo, Japan, ⁴ The Systems Biology Institute, Tokyo, Japan, ⁵ Sony Computer Science Laboratories, Inc., Tokyo, Japan, ⁶ Integrated Open Systems Unit, Okinawa Institute of Science and Technology, Okinawa, Japan, ⁷ Laboratory for Disease Systems Modeling, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

OPEN ACCESS

Edited by:

Thomas J. Anastasio,
University of Illinois at
Urbana-Champaign, USA

Reviewed by:

Giovanna Cenini,
Universität Bonn, Germany
Ghanshyam Upadhyay,
City College of New York, USA

*Correspondence:

Tiago J. S. Lopes,
Influenza Research Institute, University
of Wisconsin, 575 Science Dr.,
Madison, WI 53711, USA
dasilvalopes@wisc.edu;
Hiroaki Kitano,
The Systems Biology Institute, Falcon
Building 5F, 5-6-9 Shirokanedai,
Tokyo 108-0071, Japan
kitano@sbi.jp

Specialty section:

This article was submitted to
Experimental Pharmacology and Drug
Discovery,
a section of the journal
Frontiers in Pharmacology

Received: 18 June 2015

Accepted: 17 August 2015

Published: 01 September 2015

Citation:

Lopes TJS, Shoemaker JE, Matsuoka
Y, Kawaoka Y and Kitano H (2015)
Identifying problematic drugs based
on the characteristics of their targets.
Front. Pharmacol. 6:186.
doi: 10.3389/fphar.2015.00186

Identifying promising compounds during the early stages of drug development is a major challenge for both academia and the pharmaceutical industry. The difficulties are even more pronounced when we consider multi-target pharmacology, where the compounds often target more than one protein, or multiple compounds are used together. Here, we address this problem by using machine learning and network analysis to process sequence and interaction data from human proteins to identify promising compounds. We used this strategy to identify properties that make certain proteins more likely to cause harmful effects when targeted; such proteins usually have domains commonly found throughout the human proteome. Additionally, since currently marketed drugs hit multiple targets simultaneously, we combined the information from individual proteins to devise a score that quantifies the likelihood of a compound being harmful to humans. This approach enabled us to distinguish between approved and problematic drugs with an accuracy of 60–70%. Moreover, our approach can be applied as soon as candidate drugs are available, as demonstrated with predictions for more than 5000 experimental drugs. These resources are available at <http://sourceforge.net/projects/psin/>.

Keywords: multi-target drugs, drug safety, target validation, machine learning, protein networks, supervised learning

Introduction

New compounds are traditionally discovered by using large biological screening techniques to identify substances that cause the desired effects. While this approach has been effective for years and produced the drugs used today, technological advances are shifting the drug discovery process toward a more rational approach, with computational drug-design and pathway analysis playing major roles. With the costs of compound design dramatically increasing and most of these funds being spent on drugs that never make it to market (Munos, 2009; Scannell et al., 2012), there is a clear need for new technologies to develop more specific, less toxic compounds.

Recently, *in silico* analyses have been successfully applied throughout the drug discovery pipeline. Examples include methods to help understand the changes caused by candidate compounds in protein interaction networks (Csermely et al., 2005; Yildirim et al., 2007), and algorithms to develop specific ligands that inhibit the activity of pathogen proteins (Fleishman et al., 2011; Whitehead et al., 2012). In addition, computational analyses in key studies have revealed