

Tweeting Biology for Better Data and Better Connections

Yue Zhang*

Division of Genetics and Development, Toronto Western Research Institute, Toronto Western Hospital, University Health Network, Canada

*Corresponding author: Yue Zhang, PhD, Division of Genetics and Development, the Toronto Western Research Institute, Toronto Western Hospital, University Health Network, 60 Leonard Avenue, Toronto, Ontario, Canada, Tel: +1-416-603-5800 Ext 4797; E-mail: zy101@yahoo.com, zhanglee2006@gmail.com

Received date: Nov 24, 2014, Accepted date: Nov 24, 2014, Publication date: Nov 27, 2014

Copyright: © 2014 Zhang Y. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The challenge lying ahead is how to exact information efficiently during data mining and computation modeling for our body but systematic network intrinsically having both biological elements and social features for data -sharing is lacking. Revealed by a critical analysis of published information from the principal investigators (PIs) in one typical institute, an unexpected trend, the shrink of “monopoly” status of authority PI and an increase of democratic one, inspires us to propose that re-screenings of impact genes crossing institute PIs’ foci may set up “nodes” contributing to biological and social networks. By sharing, this could maximize the potential of some valuable reagents, incentivize institute to address ethical and intellectual property (IP) issues, help detect irreproducibility, and accelerate novel discoveries.

Introduction

Assemblies of biological systems at different levels may be created by using genome expression profiling, interacting signalling components, regulatory networks and more [1]. These need model systems’ self-organisation maps, information processing and robustness to explore why and how individual cells [2], tissues [3], organs (e.g. brain initiatives in China [4], EU and US) and bodies respond to signals. One could leap from cells to whole-body simulations to predict behaviour and create interventions [1–3]. Since computation biology needs actual spatiotemporal data to test and model, leading institutions [1,5] need to construct “hubs” and principal investigators (PIs) worldwide may collect the data.

Thorough understanding of the body needs multidimensional knowledge or cubics biology. Cubics is an approach borrowed from graph theory proposed as a model for high-quality networkable data/databases. Like Cubist art, cubics biology takes multiple views of a subject simultaneously, including views across time. The strands of omics biology – genomics, proteomics and metabolomics –could be linked in cubics biology [1,5] (Figure 1). Omics research is seldom accessible for small labs. However, labs with highly reputed PIs could share reagents, ideas and protocols with better reproducibility. Internal networks inside labs (including mentor–student relationships) create mini-cubics biology (Figure 1). Omics centres generate “big data” but are limited by a rarity or lack of “nodes” contributing to biological and social external networks; and external mini-cubics networks are also insufficient or even lacking [4].

Toward Institute-level Cubics Biology (ICB)

As a step towards new networks, we analyse the publication patterns of the University Health Network (UHN), specifically the Toronto Western Research Institute (TWRI) and the Ontario Cancer Institute (OCI), with departments covering rheumatology, cancer biology and

neurobiology. We wonder whether any changes in these patterns may be applied for accelerating discoveries. Our starting point is the impact genes highlighted by PIs, which influence multiple systems simultaneously (e.g., mTOR [6,7]).

The pattern for UHN (TWRI plus OCI) PIs’ recent publications is outlined below:

1. Roughly half of the 1744 articles published in 2013 by 520 UHN PIs from relate to primary research using impact gene-centered data [8].
2. On average, one PI may share co-authorship, ideas, analysis and reagents with a few UHN PIs. Typically, collaborations combine effort by people with different expertise. These efforts may create landmark discoveries. Classic publication highlights positive results and significant changes over valuable but seemingly small or negative ones. Authors are somehow encouraged to explore the data that best demonstrate their conclusions.
3. Historically, recent publication patterns for impact genes such as p53, PTEN and E2F1, ~30–50% of articles on one impact gene have contributions from only one or a few PIs; a variable number of PIs account for the rest. Greater contributor diversity is happening nowadays more often, with a trend towards shrinking of the “monopoly” status of authority PIs and towards “democracy”. Most of those articles are based on simple unbiased screening platforms, e.g. miRNA screens. The data from Omics platform may facilitate the buildup of impact for PIs.
4. Most UHN PIs seem to focus on ~380 impact genes. PIs may have one gene that is “their” gene, one gene that is shared frequently with other PIs (~6–12% of PIs) (e.g. Ras/MAPk, PTEN/PI3K/mTOR) and two that are shared with a handful other PIs (~1–6% PIs) (IDH, EZH2, TNE, etc.).

Re-screening Impact Genes for ICB

The increased diversity of PIs authorship for single impact gene may indicate that the field is advancing. As the greatest influence on diseases appears to be genetic and we wish to maximize the value of landmark genetic discoveries cost-effectively (e.g. simply exploring the common mechanisms in different settings), we propose the re-screening of impact gene sets focused by all institute PIs by using qPCR from different PIs’ published cDNAs. This seldom takes place due to funding contract requirements, mutual professional respect and intellectual property (IP). However, it is essential to have a database of screens that focus on the impact genes highlighted by individual PIs from a single institute.

By running a customized 384-well PCR array for the 380 impact genes focussed on by the UHN PIs, new information would appear and be integrated into an ICB database. The raw data may be run on a

Gene Expression Dynamics Inspector program as a surrogate for whole tissues, cells or organs [9]. This may become a valuable reference once personalized medicine arrives. Since the human genome has 99.99% similarity, when we compile the data for longitudinal studies and simulations, any expression level – overexpression, down-regulation, up-regulation, slight changes or seemingly no change – is equally valued. Gene expression profiles (GEPs) should include dynamic, heterochronic, tissue-specific, spatiotemporal expression chaotic, and other features. A critical impact gene may be overexpressed in the setting of one PI but repressed in another; moreover, PIs focussing on a diseased state could see opposite results to those focussing on healthy people. Arguably, a limited number of specimens may have a limited value in any experiment; however, the real data will be reliable because they are likely to make sense eventually in that pooled data may reveal further significance. As a whole, data may accrue from all system levels (cells, tissue, organs, etc.) and across time “crowdsourcing” from PIs worldwide. Additionally, one patient may visit different hospitals at different times, leaving his/her data in different settings alongside different treatments, acting as a “golden” biological reference to help researchers alongside the data of their patients to join the dots in a cubics biology matrix (CBM) (Figure 1A).

Feasibility and Significance of ICB

Feasibility depends on the features of the data. It is unlikely that a single GEP will immediately alert every UHN PI; however, since the gene has been highlighted, individual PIs would be likely to pay more attention to it. Some PIs may increase the number of impact genes they know about 50- to 100-fold compared to their former focus just a few impact genes. These GEPs would be high-quality since the original sample cDNAs have been published. With double-checking, expanded re-screens of the impact genes will filter irreproducible work, as more attention would be given to a single publication (Figure 1B).

Economic Feasibility

Sharing-based re-screens could place unexpected burdens on institutional budgets. However, ~800 molecular biology papers in 2013 by PIs in the UHN contain molecular gene expression data. Any high-quality cDNAs published could be run on a 384-well qPCR array. PCR analysis costs around \$100.00 per run, so combined tests would cost \$80,000.00 per year. This could be generally affordable, given that total external funding was \$197M for OCI and TWRI in 2013 [8] and all qPCR re-screens would therefore cost around 0.04% of annual external funding (Figure 1B).

Ethical and IP Issues

Regulatory bodies should be involved to resolve ethical and IP issues. The expanded re-screening of impact genes may be done just before/after a fresh publication and add a new “node” to the global CBM. PIs would judge each paper’s quality to minimise any irreproducibility.

Expansion and Development of ICB

Any data from omics platforms could be extracted to highlight the impact genes of interest to individual PIs, stimulating them to give these genes more attention by “tweeting” key messages, preferably into ICB databases. Re-screening data from archived cDNAs from past

publications will be added. Any discoveries could be published as additions in cases where PIs agree; possibly, this may create more efficient networks than traditional conferences and seminars.

Examples and Significance of ICB

When discoveries are reported, researchers may wonder if a common mechanism could exist in their own research and if starting from it to discover something new. Many PIs forget this question quickly, leading to missed chances. Some of these findings may be revealed, but they often lag behind similar findings in other settings. As one example, New Zealand Black (NZB) mice develop a lupus-like syndrome. One UHN lab showed that an increased proportion of NZB T1 B-cells increases Bcl-2 expression in vivo, and proved that up-regulation of Bcl-2 follows contact with self-Ag in vivo. Thus, aberrant cell signaling results in the survival of autoreactive B-cells, predisposing NZB mice to autoimmunity [10]. A decade later, another group performed cell sorting analysis in B-cells from lupus patients, revealing that lupus B-cell subsets showed decreased PTEN expression compared with B-cells from healthy controls, and PTEN expression had an inverse correlation with disease activity [11]. If an ICB had included PTEN in its re-screens 10 years ago, the latter finding might have been obtained earlier.

Another instance is the hypothesis that cancer triggers scleroderma in patients with antibodies to RPC1. Three JHU PIs in separate departments (rheumatology, immunology and cancer biology) made associations between scleroderma and an immunologic response to cancer [12]. This made their distinct directions regarding autoimmune diseases and cancer converge into a ground-breaking discovery.

As another example, research on mTOR/autophagy in osteoarthritis qPCR array screen 84 genes covering mTOR, BCL-2, PTEN, MMP13, TGF- β , and more [13] applied on fibrosis, osteosarcoma, stomach adenocarcinoma, etc. Rescreening of 380 impact genes may have been more influential.

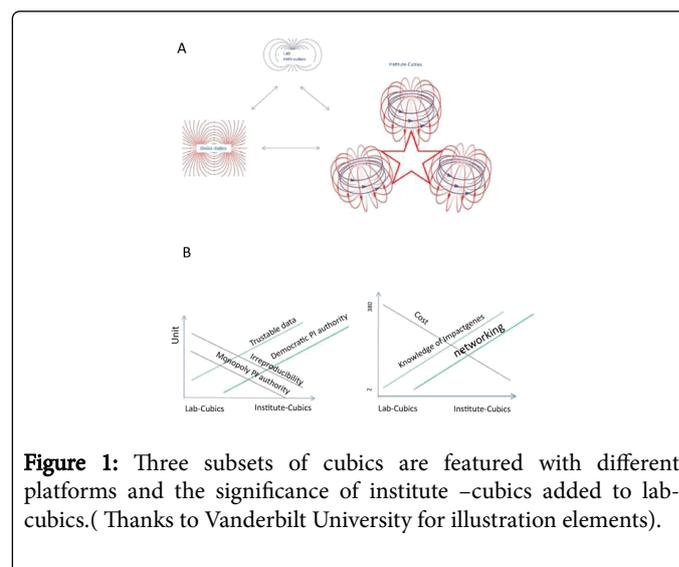


Figure 1: Three subsets of cubics are featured with different platforms and the significance of institute -cubics added to lab-cubics. (Thanks to Vanderbilt University for illustration elements).

Discussion

Most research provides separate pieces of the biology puzzle. Unbiased re-screens of impact genes, alongside data from other omics platforms, could pave the way for combining many views into one

system. Once screens are shared widely by tweeting concise expression pattern (e.g. ideally via concise messages suitably communicated with internet or mobile devices, PIs would become more aware of each other's results and be more likely to network. Research might be conducted successfully through intermediate and weak ties [14]. ICB research networking could help individual researchers to reach targets of interest in five to seven steps [14]. Once ICB networks exist, a full CBM could form via local, national, international collaborations and/or "crowd-sourcing", where small labs derive ideally benefits from this matrix. Its actual success also depends on individual institution incentives and novel paradigm thinking in regulatory systems. Re-screening is only the first step toward ICB, with more methods to go. Data from ICB would improve computer models and offer a real-world test of *in silico* systems. One may argue that current research pattern seems work well. However, the traditional thinking and cost-effectiveness of least favorable way covered would stay within society's least threshold in willingness-to-bear unless more viewpoints may be debated publicly.

References

1. Weinberg R (2010) Point: Hypotheses first. *Nature* 464: 678.
2. Roberts E, Magis A, Ortiz JO, Baumeister W, Luthey-Schulten Z (2011) Noise contributions in an inducible genetic switch: a whole-cell simulation study. *PLoS Comput Biol* 7: e1002010.
3. Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, et al. (2014) A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 345: 668-673.
4. Cyranoski D (2014) Chinese science gets mass transformation. *Nature* 513: 468-469.
5. Golub T (2010) Counterpoint: Data first. *Nature* 464: 679.
6. Vellai T, Takacs-Vellai K, Zhang Y, Kovacs AL, Orosz L, et al. (2003) Genetics: influence of TOR kinase on lifespan in *C. elegans*. *Nature* 426: 620.
7. Johnson SC, Rabinovitch PS, Kaerberlein M (2013) mTOR is a key modulator of ageing and age-related disease. *Nature* 493: 338-345.
8. <http://www.uhnresearch.ca/news/reports/RR2013final.pdf>
9. Bar-Yam Y, Harmon D, de Bivort B (2009) Systems biology. Attractors and democratic dynamics. *Science* 323: 1016-1017.
10. Roy V, Chang NH, Cai Y, Bonventi G, Wither J. Aberrant (2005) IgM signaling promotes survival of transitional T1 B cells and prevents tolerance induction in lupus-prone. New Zealand black mice. *J Immunol* 175: 7363-7371.
11. Wu XN, Ye YX, Niu JW, Li Y, Li X, et al. (2014) Defective PTEN regulation contributes to B cell hyperresponsiveness in systemic lupus erythematosus. *Sci Transl Med* 6: 246ra99.
12. Joseph CG, Darrah E, Shah AA, Skora AD, Casciola-Rosen LA, et al. (2014) Association of the autoimmune disease scleroderma with an immunologic response to cancer. *Science* 343: 152-157.
13. Zhang Y, Vasheghani F, Li YH, Blati M, Simeone K, et al. (2014) Cartilage-specific deletion of mTOR upregulates autophagy and protects mice from osteoarthritis. *Ann Rheum Dis* .
14. Dodds PS, Muhamad R, Watts DJ (2003) An experimental study of search in global social networks. *Science* 301: 827-829.