Leading Edge
Commentary

Cell

From Scientific Discovery to Cures: Bright Stars within a Galaxy

R. Sanders Williams,[1,2,*] Samad Lotia,[1] Alisha K. Holloway,[1,2] and Alexander R. Pico[1,2]
[1]Gladstone Institutes, San Francisco, CA 94158, USA
[2]University of California, San Francisco, San Francisco, CA 94143, USA
*Correspondence: rs.williams@gladstone.ucsf.edu
http://dx.doi.org/10.1016/j.cell.2015.09.007

We propose that data mining and network analysis utilizing public databases can identify and quantify relationships between scientific discoveries and major advances in medicine (cures). Further development of such approaches could help to increase public understanding and governmental support for life science research and could enhance decision making in the quest for cures.

Governments and philanthropists provide financial support for life sciences primarily with the expectation that research will lead to cures—defined broadly here as measures to prevent, eradicate, or ameliorate serious diseases. However, public understanding of how scientific discoveries actually result in cures is limited, and research to elucidate principles of biological processes may appear to non-scientists as esoteric and irrelevant to public expectations. Recent examples of important cures are evident, but public support for biomedical research as reflected by federal funding for the U.S. National Institutes of Health has eroded over the past decade (FASEB 2015), indicating the absence of a strong electoral consensus that the life science enterprise is meeting public expectations. Why is public support for life science research wavering at a time when the pace of discovery is strong and scientists see expanding opportunity, and can actions to increase public understanding of how new cures are developed lead to more sustained and predictable funding of life science?

We propose that data mining and network analytics (Nicholson 2006; Nishikawa and Motter, 2011) applied to what we call "cure network informatics" could help to increase public appreciation of the societal value of life science discoveries. Thoughtful metrics emerging from this concept perhaps can be developed and molded into forms embraced broadly among life scientists and by those providing their funding and can be used to guide decision making in ways that would accelerate progress toward cures. Here, we describe a step in this direction by means of an analytical model and topology-based algorithms that quantify relationships between scientific discoveries and cures.

We established and automated data collection and network analysis protocols utilizing publicly accessible databases, including www.fda.gov, www.clinicaltrials.gov, www.pubmed.gov, and www.webofknowledge.com. In a pilot study, we considered the recently successful applications for regulatory approval of two new drugs: ipilimumab in oncology and ivacaftor for cystic fibrosis. These medical advances are sufficiently novel and important to be reasonably characterized as "cures" (*vide supra*). Ipilimumab is the first successful entry into the new and burgeoning field of immuno-oncology (Sharma and Allison, 2015) by which sustained clinical remissions are being induced in patients with previously intractable cancers by releasing immune effector cells from checkpoint inhibition. Ivacaftor corrects the structure of a specific loss-of-function mutation in the cystic fibrosis transmembrane conductance regulator and is the first targeted therapy of this heritable disease. Beginning with the references cited in clinical trials and information provided to the U.S. Federal Drug Administration (FDA) for regulatory approval of these drugs (FDA, 2011; FDA, 2012), we extracted two consecutive rounds of retrospective citations and constructed network models of articles, authors, and institutions contributing to the network. Assumptions underlying this approach are: (1) that the authors of FDA applications and clinical trials will appropriately cite publications reporting new knowledge critical to the development of a new drug candidate and (2) that further retrospective rounds of citations will identify previous discoveries that were most important in establishing the base of knowledge that enabled the successful drug development program.

We learned that the nature of a cure discovery citation network is complex and fundamentally collaborative with respect to the number of different scientists and institutions making contributions to a cure. For example, the citation network leading to ipilimumab includes 7,067 different scientists who listed 5,666 different institutional and departmental affiliations and includes discoveries spanning 104 years of research (Figure 1A). Results for ivacaftor are similar: 2,857 different scientists from 2,516 different institutional and departmental affiliations, with discoveries spanning 59 years of research (Figure 1B).

We next characterized individual scientists within each citation network by two metrics. Propagated in-degree rank (PIR) is based on the number and citation count of articles that a given author published within the citation network and is a measure of *influence* within this selective set of publications. Ratio of basic rankings (RBR) is based on how *selectively* a given author published within the cure discovery citation network relative to background networks of topically related publications similar in size, scope, and structure. This ratio helps to normalize their overall publication output.

By applying the metrics of PIR and RBR to the entire cure discovery citation

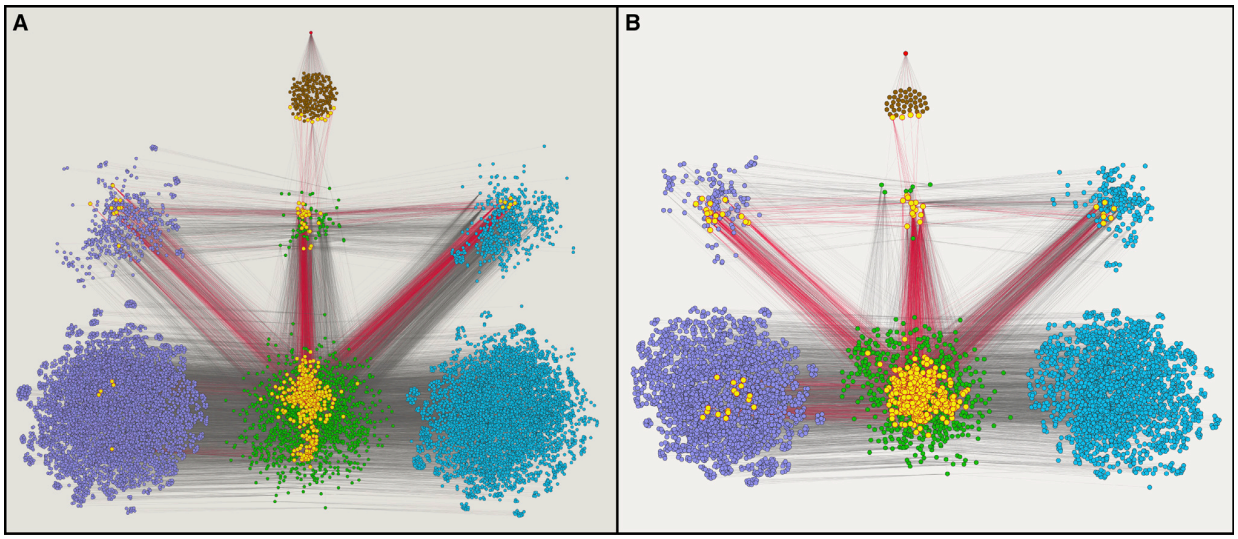

Figure 1. Cure Networks: The Constellation of Publications, Scientists, and Institutions Contributing to Drug Discovery
The red dot at the apex of the cluster is the drug ipimilumab (A) or ivacaftor (B). Relevant clinical trials and the FDA application are illustrated in brown. Publications cited in the clinical trials and FDA applications are shown in green. Likewise, papers cited by those publications are also shown in green. Authors of the papers are shown in purple, and institutional affiliations listed on the papers are shown in blue. The most influential contributors to the network as assessed by PIR and RBR (see Table S1), their articles, and their institutions are highlighted in yellow with red connecting lines.

network, the most influential and selective contributors to these massive networks emerge. Thus, in the case of ipilimumab, 15 scientists and 7 institutions associated with 433 articles spanning 46 years are characterized as elite performers (Figure 1A and Table S1). Elite performers within the ivacaftor network exhibiting similar properties as defined by the same metrics include 33 scientists and 7 institutions associated with 355 articles spanning 47 years (Figure 1B and Table S1). These elite performer subnetworks are integral to their overall citation networks, serving as hubs for 31% of the ipilimumab network and 49% of the ivacaftor network.

These data quantify how the knowledge base on which important advances in medicine ("cures") depend includes contributions from a large and diverse set of individual scientists working in many locales. This insight should be instructive for policy makers by suggesting that future cures will depend on broadly based public support of life sciences. Narrowly targeted funding initiatives may well have value but are unlikely in isolation to generate the breadth of new knowledge required to lay the foundation for future cures.

We call on the scientific community to embrace and advance the concept of cure network informatics so as to develop advanced and sophisticated analytical tools to increase understanding of how scientific discoveries lead to cures, including predictive metrics that may guide decision making with respect to work in progress. All of the code necessary to reproduce and extend this initial effort is freely available and open source (Lotia and Pico, 2015). This network informatics approach can be applied to any "cure" with a cited publication trail. Curators of publically available databases could play important roles in these efforts by considering cure network informatics in the design of database architecture and embedded tools. It will be important to identify trends that hold across all cures and ones that are specific to certain types of cures. It will also be useful to identify features of hubs within cure networks that are essential to the flow of knowledge required to create a cure.

A need for better metrics for assessing performance and for decision support within the life sciences is widely acknowledged by leaders and commentators in biomedicine (Sarli and Carpenter, 2014; University of Gothenburg, 2013). Metrics that are readily understandable by non-scientists, grounded in outcomes that the general public values highly (cures), and faithful to what scientists know to be the richly intersecting and often unpredictable nature of scientific discovery should be more useful for influencing policy makers than currently available alternatives. Further development of new and useful tools for cure network informatics should contribute to increased public trust in, and support for, the life science enterprise.

SUPPLEMENTAL INFORMATION

Supplemental Information includes one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.09.007.

ACKNOWLEDGMENTS

We are grateful for financial support of this project from the Schwab Foundation, Bruce and Martha Atwater account. We also thank Eleanor Prezant for project management and communications relating to this research.

REFERENCES

FASEB (2015). http://www.faseb.org/Portals/2/PDFs/opa/2015/2.10.15%20NIH%20Funding%20Cuts%202-pager.pdf.

FDA (2011). http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/125377Orig1s000TOC.cfm.

FDA (2012). http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203188s000TOC.cfm.

Lotia, S., and Pico, A.P. (2015). https://github.com/gladstone-institutes/bibliometrics/tree/1.0.1.

Nicholson, S. (2006). Inf. Process. Manage. *42*, 785–804.

Nishikawa, T., and Motter, A.E. (2011). Sci. Rep. *1*, 151.

Sarli, C., and Carpenter, C.R. (2014). Mo. Med. *111*, 399–403. http://digitalcommons.wustl.edu/cgi/viewcontent.cgi?article=1048&context=becker_pubs.

Sharma, P., and Allison, J.P. (2015). Cell *161*, 205–214.

University of Gothenburg. (2013). ScienceDaily. http://www.sciencedaily.com/releases/2013/10/131014094212.htm.