Reverse Translation

Using Computational Modeling to Enhance Translational Research

Daniel Gratz, Thomas J. Hund, Michael J. Falvo, Loren E. Wold

Animal models provide powerful tools for examining human disease; however, translation of findings from these models to human patients is often challenging. To this end, we discuss modern tools to support the process of selecting and validating animal models with relevance to humans. We draw from data mining and computational modeling approaches to examine how large data sets may be leveraged to identify suitable models with the greatest translational potential.

Rodent models for human disease are indispensable tools for modern biomedical research, particularly because of the high degree of control they afford in examination of complex pathophysiological phenomena. There is no doubt within the scientific community that animal models are both useful and necessary; however, the value of any animal model depends on its ability to replicate the human condition. Curing disease in a mouse is fascinating but short on merit without application to the understanding of human disease. Thus, there is great need for a robust process to translate a human disease into an appropriate rodent model, termed reverse translation, to distinguish it from the more familiar act of moving findings from animal models into human patients (translation). Studies on the quality of rodent models have reported disparate results depending on the specific physiological system of study. For example, studies on the acute inflammatory response to blunt trauma, burn, and endotoxin have found a poor correlation between the response in mouse and that in human, leading to the conclusion that murine models have limited translational potential.[1] This weak association may be due, in part, to a highly variable response in mice that is not observed in humans, underscoring the low success rate in translating mouse findings on acute inflammatory response to human. However, rodents fare better as a model in chronic inflammation, where there is a much better correlation between the response in rodents and humans. In fact, an examination of the signaling pathways activated in chronic inflammation revealed that 7 of the top 10 pathways were found to be common in rodents and humans.[2] Similar concerns apply not only to studies involving rodents but to large animal models, as well. For example, collateral artery formation is different in the canine versus pig or human.[3] This anatomical difference contributes to less severe infarct formation and more complete recovery after coronary ligation in the canine compared with porcine models. At the same time, the ability of the canine to readily grow collateral arteries makes it an interesting model to study the arteriogenesis response to atherosclerosis with relevance to the human. These examples highlight the importance of a reverse translation process where the human condition is mapped to an appropriate animal model before rigorous analysis. To that end, it is illuminating to consider state-of-the-field with respect to experimental and computational methods for animal model validation and reverse translation. In particular, advanced computational and mathematical methods have evolved to the point that they have great potential to revolutionize the reverse translation (and subsequent translation) pipeline.

Fundamental benefits of exercise have been further strengthened with the use of reverse translation. For example, studies have shown that aerobic fitness was associated with hippocampal volume and memory task performance in older adults.[4] These clinical observations, through reverse translation, were confirmed in an animal model showing that exercise (eg, running) induced striking effects on neural hippocampal progenitor cells.[5] However, the use of reverse translation was simply to validate in this example, and these associations could be strengthened with the use of data mining/computational modeling approaches herein described.

With the invention of full gene expression assays, it is now possible to obtain a much more complete description of the attributes of an animal model. Numerous studies have used this technique to retrospectively evaluate the quality of mouse models already used for translational research and thus comment on the continued potential of using mice for further studies.[1,2] Such studies have gathered gene expression profiles from a cohort of human patients (typically >100) and compared them to a healthy cohort. They then gathered similar data from the mice and searched for genes in the mice and humans that were similarly altered in the affected patient population. These studies would potentially be even more helpful if incorporated into the reverse translation process to identify a priori the most appropriate animal model (rather than for validation after the fact). Even if such a large patient population cannot be tested, a nonsignificant but trending result could give the researchers added confidence that their mouse model

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Dorothy M. Davis Heart and Lung Research Institute (D.G., T.J.H., L.E.W.) and Department of Internal Medicine (T.J.H.), The Ohio State University Wexner Medical Center, Columbus; Department of Biomedical Engineering, College of Engineering (D.G., T.J.H.), College of Nursing (L.E.W.), and Department of Physiology and Cell Biology (L.E.W.), The Ohio State University, Columbus; War Related Illness and Injury Study Center, Department of Veterans Affairs, New Jersey Health Care System, East Orange (M.J.F.); and Rutgers Biomedical and Health Sciences, New Jersey Medical School, Newark (M.J.F.).

Correspondence to Loren E. Wold, PhD, The Ohio State University, 603 Davis Heart and Lung Research Institute, 473 W 12th Ave, Columbus, OH 43210. E-mail loren.wold@osumc.edu

(*Circ Res*. 2018;122:1496-1498.
DOI: 10.1161/CIRCRESAHA.118.313003.)

***Circulation Research* is available at http://circres.ahajournals.org**
DOI: 10.1161/CIRCRESAHA.118.313003

has potential. Additionally, the gene expression profiles provide a list of interesting genes for use in further study using, for example, methods from data science.

Association rule mining is one such approach that may be exploited to extract meaningful information from large data sets. This machine-learning technique takes sets of elements (called transactions) and creates rules from them. For example, if the elements were individual items from a grocery store and a transaction is the group of items purchased together, association rule mining can extract a rule such as: if a person at a grocery store buys bread, they also buy jam and peanut butter 80% of the time and bread, jam, and peanut butter are bought in 10% of the transactions. In the case of gene profiles, individual genes would be considered the elements, and the gene profile of a particular individual would be the transaction. Rules would then take the form: when gene A is upregulated, we often see that gene B is also upregulated. Numerous tools have been developed to use association rule mining for determining the involvement of both known and novel pathways. One such tool is GenMiner,[6] which uses gene expression data, as well as annotated information, to define association rules with predictive potential. Annotations can be anything from genes listed together in a paper to genes known to be linked in a pathway. By using multiple levels of information, GenMiner enables researchers to easily synthesize previous work into their current projects. In reverse translation, rules found in humans could be compared with the rules found in rodents (or other species). The overlap of associations and pathways found would point to similarities and differences in the models that may be used to select the most appropriate model and direct future study. For example, when examining the relationship between exercise and hippocampal volume, rules linking varied gene expression in humans and mice could be compared, to search for interesting pathways shared between species. Association rule mining thus provides both directions of future study and increased confidence in the validity of an animal model.

Another important criterion for selection of an animal model is its similarity to the human disease phenotype. With large amounts of patient data becoming increasingly available to researchers, it is imperative to consider how such data may be leveraged to construct the best animal models possible. This is especially important when no genotypic data are available and phenotypic data alone are used to validate a model. One possible way to use large data sets is by applying data mining techniques to extract a larger number of characteristics of the disease to incorporate into the animal models. Clustering algorithms are one such type of technique. A clustering algorithm takes features (eg, height and weight) from many patients and uses them to find similarities to form groups of patients with similar attributes (ie, clusters), which may then be compared with potential animal models. Such an approach may be useful, for example, in our effort to better understand long-term cardiovascular consequences of moderate duration, high-intensity particulate matter exposures[7] for US military veterans.[8,9] Through modeling approaches, we have attempted to extrapolate the exposure paradigms in humans to mouse exposures, ensuring that exposure concentrations are normalized for species. Clustering also allows for examination of subsets within the population, allowing us to consider the differences among the veterans when evaluating our animal model. Effective modeling allows clinically relevant therapeutic development in mice that would not otherwise be translatable.

Beyond biostatistical and data science methods, computational simulations provide a valuable tool to help with the reverse translation process. Although computational models cannot replace traditional experiments, they may be used to enhance investigation into complex phenomenon and suggest new experiments in a closed-loop approach. Computational simulations enable researchers to test hypotheses and explore the full parameter space of a biological system in a way that is difficult/impossible experimentally, which may help increase a researcher's understanding of the phenomena before the reverse translation process. For example, before considering reverse translation, it may be useful to use computational models to explore the phenomena of interest across temporal and spatial scales. Using a model, a researcher may begin by constructing a cellular model based on observed data (eg, cell electrophysiology data). This cellular model may then be incorporated into a model of the tissue featuring unique spatial and even temporal scales,[10] which will be informed by unique data (eg, magnetic resonance imaging data) and governed by distinct mathematics/biophysics. The components of the fitted cell model can be transferred into existing animal cell models and fit and scaled to develop a model of disease in animals. Examination of the physiological aspects of the computational models can then be used to inform selection of an in vivo animal model (by comparing computational models of rodent versus large animal, for example). Experiments may then be used to inform further development of the computational model. Using such a closed-loop, iterative approach, computer simulation allows for tight coupling of animal and human models, enhancing and expediting future research.

We have discussed multiple examples where computational algorithms and models may facilitate the endeavor of high-quality reverse translation (and further future translation). Full gene assays provide a wealth of data that may be used for reverse translation, as well as mined for valuable insights. Furthermore, there are an increasing number of large data sets on patients that may be leveraged to identify valid animal models through other data mining techniques such as clustering. Ideally, both genotypic and phenotypic data would be used in the model selection; however, as it is sometimes not possible to gather genotypic data, successful validation can be performed with the phenotype alone. Finally, computational simulations provide a valuable resource for reducing the gap between animal and human models accelerating the reverse translation/translation loop. A further application of these techniques could also be to aid in the choice of species for a model, by enabling a more detailed comparison. As the quantity and quality of data increases, it will become increasingly imperative that researchers push the boundaries of data mining, biostatistical and computational techniques to maximize the potential for impactful translational studies.

Sources of Funding

We are supported by the National Institutes of Health grants HL135096, HL134824, and HL114893 to T.J. Hund and AG057046, HL139348, NR012618, and ES019923 to L.E. Wold.

Disclosures

None.

References

1. Seok J, Warren HS, Cuenca AG, et al; Inflammation and Host Response to Injury, Large Scale Collaborative Research Program. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci USA*. 2013;110:3507–3512. doi: 10.1073/pnas.1222878110.
2. Sweeney TE, Lofgren S, Khatri P, Rogers AJ. Gene expression analysis to assess the relevance of rodent models to human lung injury. *Am J Respir Cell Mol Biol*. 2017;57:184–192. doi: 10.1165/rcmb.2016-0395OC.
3. Helisch A, Schaper W. Arteriogenesis: the development and growth of collateral arteries. *Microcirculation*. 2003;10:83–97. doi: 10.1038/sj.mn.7800173.
4. Erickson KI, Prakash RS, Voss MW, Chaddock L, Hu L, Morris KS, White SM, Wójcicki TR, Mcauley E, Kramer AF. Aerobic fitness is associated with hippocampal volume in elderly humans NIH Public Access. *Hippocampus*. 2009;19:1030–1039. doi:10.1002/hipo.20547.
5. Farioli-Vecchioli S, Mattera A, Micheli L, Ceccarelli M, Leonardi L, Saraulli D, Costanzi M, Cestari V, Rouault JP, Tirone F. Running rescues defective adult neurogenesis by shortening the length of the cell cycle of neural stem and progenitor cells. *Stem Cells*. 2014;32:1968–1982. doi: 10.1002/stem.1679.
6. Martinez R, Pasquier N, Pasquier C. GenMiner: mining non-redundant association rules from integrated gene expression data and annotations. *Bioinformatics*. 2008;24:2643–2644. doi: 10.1093/bioinformatics/btn490.
7. Engelbrecht JP, McDonald EV, Gillies JA, Jayanty RK, Casuccio G, Gertler AW. Characterizing mineral dusts and other aerosols from the Middle East–Part 2: grab samples and re-suspensions. *Inhal Toxicol*. 2009;21:327–336. doi: 10.1080/08958370802464299.
8. Zuo L, Youtz DJ, Wold LE. Particulate matter exposure exacerbates high glucose-induced cardiomyocyte dysfunction through ROS generation. *PLoS One*. 2011;6:e23116. doi:10.1371/journal.pone.0023116.
9. Falvo MJ, Osinubi OY, Sotolongo AM, Helmer DA. Airborne hazards exposure and respiratory health of Iraq and Afghanistan veterans. *Epidemiol Rev*. 2015;37:116–130. doi: 10.1093/epirev/mxu009.
10. Moreno JD, Zhu ZI, Yang PC, Bankston JR, Jeng MT, Kang C, Wang L, Bayer JD, Christini DJ, Trayanova NA, Ripplinger CM, Kass RS, Clancy CE. A computational model to predict the effects of class I anti-arrhythmic drugs on ventricular rhythms. *Sci Transl Med*. 2011;3:98ra83. doi: 10.1126/scitranslmed.3002588.

KEY WORDS: atherosclerosis ■ data mining ■ endotoxin ■ inflammation ■ mice

Circulation Research
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Reverse Translation: Using Computational Modeling to Enhance Translational Research

Daniel Gratz, Thomas J. Hund, Michael J. Falvo and Loren E. Wold

Circ Res. 2018;122:1496-1498
doi: 10.1161/CIRCRESAHA.118.313003
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7330. Online ISSN: 1524-4571