What is translational bioinformatics?

CrossMark

In 2001, the Human Genome Project successfully completed the first full human genome sequence (Lander et al., 2001; Venter et al., 2001). With the advent of next generation sequencing (NGS) platforms, what took over 3 billion dollars and more than 10 years to complete can now be done for as little as $1000 and in only a few days. The application of NGS has enabled for the interrogation of the human genome in ways not previously considered. In recent years, large "big-data" projects have utilized this technology to interrogate the human genome in finite detail and provide valuable insights into the molecular mechanisms underlying diseases. The 1000 genome project (Genomes Project C et al., 2010, 2012) has revealed the diverse genetic variation across different population groups. The ENCODE project (Consortium EP, 2012) identified and characterized functional elements throughout the genome. Similarly, The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) has yielded valuable clues into the molecular biology of a variety of cancers. Collectively, these studies and others have given new clues to the genome that can ultimately be leveraged into clinical applications.

While these large projects have not only taught us a lot of the nature of the human genome, they have also generated new methods to effectively and efficiently mine the vast amount of data generated. With this technology, the bioinformatics field has flourished and developed new and complex ways to effectively and efficiently mine the data. Over time, bioinformaticians have improved our ability to map the sequence reads to the genome, and correlate the results to disease phenotypes, and in some cases translate the results to a clinical setting (Chong et al., 2015; Knoppers et al., 2015; Shen et al., 2015). Yet, our ability to establish a better, more practical and accelerated path from discovery to improved patient care lags. Certainly in the future, these, and future, technologies will transform our approach to medicine and change the ways in which we diagnose and treat diseases. But, in order for this to be achieved new and specific tools to translate the genome to the clinic are needed. A new field, 'translational bioinformatics' is the domain to bridge this gap.

The American Medical Informatics Association defines translational bioinformatics as the development of storage, analytic and interpretive methods to optimize the transformation of increasingly voluminous genomic (proteomic, transcriptomic, metabolomic, epigenomic, enviromic, interactomic, pharmacogenomic, phenomics) and biomedical data into proactive, predictive, preventive and participatory health (https://www.amia.org/applications-informatics/translational-bioinformatics). And in the past ten to fifteen years, bioinformaticians have largely devoted themselves to exactly these tasks: data analysis, library cataloging, database management, distribution specialization and software engineering. While these tasks have well served certain translational needs, other new needs have evolved and the field continues to adapt to these needs, as well as define others. Currently, the transformation of huge volumes of complex data into clinically useful knowledge requires the convergence molecular bioinformatics, biostatistics, statistical genetics and clinical informatics into computational methods to establish a better/more practical, and accelerated, path from discovery to improving patient care, and important discoveries for practical use. In other words, achieving these goals involves interdisciplinary collaboration, and one that is essential to advancing translational genomics, similar to those used in large "big data" research projects.

The importance of translational bioinformatics may be best understood in the things it is teaching us, things not previously knowable. For example, it is identifying flawed science, improving estimates of relative pathogenicity of human genetic variants, inferring new insights about underlying genetic mechanisms of disease, and identifying promising new drug indications based on curating large volumes of scientific literature. While, sequencing an exome for a clinical diagnosis can be a routine task, the interpretation of the data to make an actual diagnosis or treatment plan is much more complex. Out of the many thousands of variants identified, many of them will have to be evaluated for their clinical utility. At times, for perhaps a simple Mendelian disorder this may be as simple, as only a single variant will need to be identified and considered. But for more complex diseases (e.g. cancers, diabetes, or neurodegenerative diseases) multiple variants will need to be identified. It is only by asking the correct questions about the patient and the disease, along with employing the right computational tools that correct answers can be achieved. In some cases appropriate questions and tools already exist and can be applied, while in other cases, they still need to be developed. Thus, the translational bioinformatics field is poised to address these issues and help generate the right answers.

Determining the right question to ask is now a collaborative effort involving non-conventional sources. In the process, unaddressed areas or unchallenged assumptions can be identified and so researched. Bioniformaticians can, and for that matter do, create new tools to answer the questions, pioneer and validate their use, and report findings demonstrating the tool's ability to provide *good* answers. Thus, the field has moved from mere doing to also questioning and therefore to being integrally involved in research design. In this way innovative bioinformatics methods are revolutionizing translational science. Crowd based discovery, new methods for using genome/transcriptomes to recommend cancer treatments and more trained systems, such as machine learning, are but some examples. Furthermore, it is our increasing ability to directly investigate that biological responses in real time is enabling the field to understand the complexity of disease processes.

Thus, with the right questions and the right tools to answer them, translational bioinformatics can translate DNA sequences into new discoveries, novel diagnostics and fundamental causes of disease as targets

for new therapeutics. The scope of application, though, goes beyond the biomedical sciences to health care delivery, health economy and health policy. In this way, translational bioinformatics is crucial to advancing translational genomics (omics).

In this special issue "What is Translational Bioinformatics?" we present on overview papers that address some of the key questions in the field. Four articles address the need of improved tools and methods to effectively transform genomic data into translatable results. In 'The problem with big data in translational medicine', Jordan presents an overview of the evolution of bioinformatics and look to the future of translational medicine, while discussing current hindrances to applying with big data techniques. In 'Empowered Genome Community: Leveraging a bioinformatics platform as a citizen-scientist collaboration tool', Wendelsdorf et al., discuss the potential research benefits of a citizen to researcher to citizen tool, highlighting how QIAGEN's ingenuity platform and its ability to integrate clinical with the genomic data to optimally identify correlative variants can expedite research. Glazer et al., in 'Atoms, Bits, and Cells' shows how cloud computing can be leveraged to increase the ability and speed at which genomic data can be analyzed. Tobias et al., in 'Developing educational iPhone, Android and Windows smartphone cross-platform apps to facilitate understanding of clinical genomics terminology', present a simple and easy to use smartphone app geared towards clinicians to aid in the interpretation of genomic data. Finally, in 'OncDRS: An Integrative Clinical and Genomic Platform for Enabling Translational Research and Precision Medicine', Methew, et al., describe their OncDRS platform which can integrate clinical and genomic data to enhance translational medicine and research. 'Genome Interpretation: clinical correlation is recommended' by Segal discusses different ways to clinically correlate genomic information. In 'Interdisciplinary training to build an informatics workforce for Precision Medicine', Williams et al., present a model for a collaborative approach to the training of future bioinformaticians and clinicians to better implement precision medicine. Two papers by Karikari et al., 'Developing expertise in bioinformatics for biomedical research in Africa' and Widening participation would be key in enhancing bioinformatics and genomics research in Africa', discuss the need for translational bioinformatics in Africa, and specifically how the field can enable knowledge generation by under-represented populations and in particular Africa. They show this both in terms of the development of the infrastructure of genome centers and the need to increase the education of bioinformatic researchers. Isaacson Barash, et al., in TranSMART Foundation Datathon 1.0: The Cross Neurodegenerative Diseases Challenge, present the success of a datathon in neurodegenerative diseases; across Alzheimer's disease, Parkinson's disease and others and the preliminary new findings it generated. Together, these articles display a window into the various types of research questions and clinical needs important to the translational bioinformatics field, which can ultimately help bridge the gap between genomics and an effective clinical utility.

References

Chong, J.X., et al., 2015. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. Am. J. Hum. Genet. 97 (2), 199–215.

Consortium EP, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489 (7414), 57–74.

Genomes Project C, et al., 2010. A map of human genome variation from population-scale sequencing. Nature 467 (7319), 1061–1073.

Genomes Project C, et al., 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491 (7422), 56–65.

Knoppers, B.M., Zawati, M.H., Senecal, K., 2015. Return of genetic testing results in the era of whole-genome sequencing. Nat. Rev. Genet. 16 (9), 553–559.

Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409 (6822), 860–921.

Shen, T., Pajaro-Van de Stadt, S.H., Yeat, N.C., Lin, J.C., 2015. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Front. Genet. 6, 215.

Venter, J.C., et al., 2001. The sequence of the human genome. Science 291 (5507), 1304–1351.

Eric R. Londin
Computational Medicine Center, Department of Pathology, Anatomy and Cellular Biology, Sydney Kimmel College of Medicine at Thomas Jefferson University, United States
Corresponding author.
E-mail address: eric.londin@jefferson.edu.

Carol Isaacson Barash
Helix Health Advisors, United States