— THE RISE OF THE BIOLOGISTS: 
THE CHANGING FACE OF THE 
BIOINFORMATICS INDUSTRY

Leo Lafferty-Whyte, Ph.D. 
Proposition Manager, Discovery & Translational Science, Thomson Reuters

As in-depth biological knowledge is an increasing factor for research success, a shortage of bioinformatic skills presents an exciting opportunity for research biologists.
AbbVie’s recent FDA accelerated approval of Venclexta™ (venetoclax) for patients with chronic lymphocytic leukemia (CLL) was the result of 20 years of work by “a dedicated group of scientists who really understood the biology and how to use the tools,” according to Michael Severino, AbbVie’s chief scientific officer, in an interview with biopharmaceutical news source BioWorld Today.

As the saying goes: “The only thing constant in life is change.” Never is that more true than for the biological research and development field. The fast-paced rate of research makes the industry an exciting place to work, but also a challenging environment in which to keep in-depth biological knowledge up to date. AbbVie attributed such biological knowledge as critical to its recent success. Recent publications by both AstraZeneca and Pfizer also highlighted the importance of such knowledge through direct correlations with clinical success and understanding of disease and target biology. 

We’ve all seen the graphs showing the exponential growth of the volume of data that has been produced over recent years. According to IBM, we are producing 2.5 quintillion bytes of data each day, with 90% of all data in existence being produced in only the past two years. The challenges of producing and storing such volumes of data alone are significant, but the actual analysis and interpretation of the data to facilitate such knowledge presents another hurdle.

To complicate things more, not all data is the same. Since the data being produced is not all being made on the same machines with the same conditions, nor measuring the same parameters, the challenge of integrating the data presents a further hurdle to the ability to make decisions using this growing resource. The difficulty in data integration was a hot topic at the recent Bio-IT World Conference & Expo in Boston, where Tim Miller (VP of Informatics and Analytics Strategy, IP & Science, Thomson Reuters) presented the challenges and a few thoughts on how to move past them.

Learn more about how IP & Science is approaching data integration in Dr. Miller’s recent article. So there’s a lot of data being produced. Integrating it into a format where it can be analyzed for meaning and then used to make research decisions is a challenge. As per Dr. Miller’s article, we are beginning to find methods that allow us to do this efficiently. Implementing such data integration techniques takes knowledge and skill in various processes such as ontology design, management and curation. As a result, skilled data scientists are becoming a much sought-after and valuable resource.

Bernard Marr discusses a rather interesting solution that is naturally evolving to overcome the shortage of skilled data scientists in his recent Forbes Tech article. Large companies, such as Sears, are empowering other staff to perform data analytics previously reserved for data scientists. The rise of these “citizen data scientists” is empowered by the invention of appropriate tools and software that automate complex integration and analytics and present them back to the user in such a manner that it is immediately interpretable. The concept isn’t to replace the data scientists
entirely but instead to release their highly sought-after skill set to be applied to the more complex and custom analytics that are required above the day-to-day big data analysis.

Data scientists are not the only skill set in short demand that is presenting a challenge in the face of data deluge. In a recent Nature commentary, Jeffrey Chang from the University of Texas Health Science Center observed, “biological data are accumulating faster than people’s capacity to analyze them ... there are not enough bioinformaticians.” While Chang also observes that there was surprisingly little application of a standardized workflow in the projects coming through his department, he also notes that only 5% of his time was spent on “pure bioinformatics – that is, developing new algorithms that merit their own publications.” Such lack of time for innovation in our bioinformatics industry presents a threat to our ability to catch up on the backlog of biological data awaiting analysis by inventing new analytical methods. It also presents a bigger threat to keeping data analytical techniques on par with the changing data generation technologies.

Taking the lead from the scarcity of data scientists, one solution is to empower the biologist to perform some of the analytics themselves. Biological researchers typically generate the data and, where available, work with their bioinformatics teams to have the data processed and analyzed. The results are then returned to them so that they can then interpret the biological meaning and therefore impact of the experiment. With anywhere between two and 100+ biologists for every bioinformatician (in SME/academia and large pharma, respectively) even processing the data and getting the results from the teams can lead to a lag in project progression. If biologists are empowered to perform their own data analytics and understand the potential biological impact, then the bioinformaticians could be freed up to support the more complex second pass and pure bioinformatics processes that the industry needs to keep the wheels of progress and profit moving smoothly.

Much like the citizen data scientists, such citizen bioinformaticians would require the right toolkit to guide them through the process. Tools would need to automatically analyze the data with meaningful algorithms, display results in a manner that was interpretable without being “black box”, and most importantly, be easily reproducible without needing to write a series of small novels capturing the settings and processes that were required to produce it in the first place.

In IP & Science, we have been working with both biologists and bioinformaticians for many years and this rising trend has concerned us. Our bioinformatician customers were increasingly overloaded and our biologists increasingly frustrated with long wait times. Indeed, many biologists have begun to learn to process and analyze their own omics data in an effort to bypass the bottleneck. This kind of approach – while to be applauded for proactivity on one hand – also presents significant risk if the incorrect algorithmic approach is applied and decisions are made on incorrect assumptions. Our solution is Key Pathway Advisor™, the first in a new suite of small applications.
Key Pathway Advisor is specifically designed with biologists in mind. The workflows apply state-of-the-art bioinformatics algorithms such as synergy pathway enrichments and causal reasoning while explaining the results in a manner that a degree in bioinformatics is not required. The focus of the app is to help the user identify the potential biological impact and potential root causes for their expression and gene variant results. The tool also automatically aligns the results with current putative biomarker and drug pipeline knowledge for their disease of interest. Thus producing an interactive report that can be discussed with colleagues (including bioinformaticians) to decide next steps and increase project efficiency; saving everyone time, money and hopefully a few wrinkles.

For an example analysis utilizing the Key Pathway Advisor, please read “Molecular Profiles of Tumor Infiltrating T-lymphocytes in Breast Cancer Patients,” which was presented at this year’s MMTC and Bio-IT World events.

To learn more about Thomson Reuters IP & Science solutions and our Key Pathway Advisor (KPA) offering, please see MetaCore™ and Key Pathway Advisor: Data-mining and pathway analysis.
REFERENCES

1 Cook, David; Brown, Dearg; Alexander, Robert; March, Ruth; Morgan, Paul; Satterthwaite, Gemma and Pangalos, Menelas N.; 2014; Lessons learned from the fate of AstraZeneca's drug pipeline: a five dimensional framework; Nature Reviews Drug Discovery 13; 419-431


3 http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
(Data correct as of 25th April 2016)

4 Marr, B. How The Citizen Data Scientist Will Democratize Big Data. Forbes: Tech, April 1, 2016


6 Thomson Reuters independent customer & market research – Data not published.
