Preface

The emerging advances of Bioinformatics and the need to improve Healthcare and the Management of Medical Systems has promoted research collaborations among researchers from the field of Bioinformatics and Health Informatics together with administrators, clinicians and data scientists. There is increased need to improve target drug therapy, personalized medicine and the way clinical decisions are made for the welfare of patients and on the management of medical systems. These needs demand Big Data Analytics incorporating the latest computational intelligence and statistical methodologies together with Data Mining and Machine Learning Methodologies. Big Data aspects such as data volume, data velocity, data veracity and data value are considered and examined in terms of usefulness and importance as to which is the most decisive criterion turning data from big to smart as a real time assistance for the improvement of living conditions. The increased need to improve healthcare and the welfare of patients and people in general, requires that we fast improve prognosis, diagnosis and therapies in order to advance personalized medicine and targeted drug/gene therapy (Alyass, et al., 2016; Chen et al., 2013; Tenenbaum, 2016; Greene et al., 2013).

The overall scope and main objective of the book is to expose the reader to the latest developments in Bioinformatics and Health Informatics but to also put emphasis on increasing awareness of all stakeholders of the importance to move from a data management organizational culture to a learning organization culture. We believe that the carefully chosen individual chapters address effective ways of communication and dissemination of the biological relevance of genomic and proteomic discoveries and related specific gene expression to realizing the clinical potential to make possible targeted therapy, personalized medicine and enhancement of human wellness and social cohesion. The recent advances in Bioinformatics and in Healthcare Informatics for next generation medical research are examined with the goal to improve medical practices and the management of Medical Systems both at the national and global level. The book includes chapters on innovation of advanced methods and techniques from medicine bioinformatics/ health informatics as also from computer science, statistics, and information theory which infer the relationships and dynamics among genes and their products. Both the genes and their products could be targets for treatment but also help propose predictive models which will enable the industry to develop an array of algorithms and software and overall accurate and intelligent computational systems for next generation medicine and medical systems.

THE CHALLENGES

The rapid development of biomedical research has given rise to an increasing demand of various computational and mathematical approaches to analyze and integrate the resulting large-scale data with the molecular and bioinformatics basis of clinical science. The new approaches will provide useful therapeutic targets to improve diagnosis, therapies and prognosis of diseases but to also help toward the establishment of better and more efficient next generation medicine and medical systems.

Declines in the cost of generating genomic and proteomic data have made the approaches to DNA sequencing, RNA- sequencing, and high-throughput screening and protein analysis more efficient and effective to analyze data but also have created new challenges in data analysis. Researchers, such as data analysts, specialize in developing computational methods for extracting insights from next-generation sequencing and high-throughput screening data with the goal to find the most efficient and effective ways to analyze data and generate insights of the function of biological systems. The emerging field of Systems Biology with its recent technical breakthroughs in next generation sequencing (NGS) technology with advanced gene and protein bioinformatics analysis, provides formal approaches to modeling and simulating regulatory processes in biological systems and has accelerated the convergence of discovery science with clinical medicine and improvement of medical systems (He et al., 2013; Hirak, et al., 2014; Huang et al., 2013; Lewis et al., 2012; Lin et al., 2015; Luo et al., 2016; Margolis et al., 2014; Ng et al., 2014; Niemenmaa et al., 2012).

One very big challenge in the healthcare system is how to connect it to basic research and its applications together with the technical and structural problems encountered. In addition to powerful computational tools that need to be developed, many practical problems exist as far as the proper training and handling and analysis of big data is concerned and in the development of a common technical language and terms so that meta-data is aligned and clinical data is properly analyzed and shared. There have been a number of international efforts to develop global and local frameworks via global and local associations and alliances which take into consideration ethical and regulatory challenges and address privacy and security issues via secure electronic methods (European Bioinformatics Institute (EBI), 2016; Global Alliance for Genomics and Health (GA4GH) 2016; The Precision Medicine Initiative® (PMI) Cohort Program, 2016; The Phenotype-Genotype Integrator (PheGenI), 2016;The International Medical Informatics Association (IMIA), 2016; United Nations (UN), 2016; World Health Organization (WHO), 2016). Academics, commercial groups, and industry in general must come together and through good team work decide on what is best for an individual's well-being and health and also psychological and social needs. Improved synergies and good team work of talented and well trained people could help improve personalized medicine and possibly precision medicine.

SEARCHING FOR SOLUTIONS

Researchers, clinicians and administrators continue looking for solutions to the many challenges of Big Data Analytics in the field of Health Informatics in relation to its applications and implications. For example, by translating information from cancer genomics into diagnostics and therapeutics will revolutionize cancer treatment and management as they develop next generation sequencing (NGS) and advanced bioinformatics-based non-invasive cancer management systems that may provide a better way to monitor and diagnose cancer recurrence and therapy effectiveness. Identification of genomic and

protein alterations that lead to health problems enhances our understanding and classification of human diseases and accelerates the discovery of new approaches for clinical diagnosis, outcome prediction and risk stratification. Emphasis nowadays is put on the importance of establishing the biological relevance of genomic discoveries and related specific gene expression to realizing the clinical potential to make possible targeted therapy, personalized medicine and effective medical management systems.

RECOMMENDATIONS

At present time, within the field of Bioinformatics, massive amounts of biological information becomes available, such as genome sequences, gene expression data, protein sequences, protein interaction data etc. and therefore, there is high need for more efficient, sensitive, and specific big data analytic technology in Bioinformatics. For example, in biological and biomedical imaging process and analysis, large volumes of data are generated. As a consequence the research community faces many challenges mainly those that have to do with storage, indexing, managing, mining, and visualizing big data. The field of Bioinformatics enhances therefore, the development of databases, algorithms, computational and statistical techniques and tools to solve a variety of practical problems, and abstract biological data by mainly analyzing and correlating genomic and proteomic information (Greene et al., 2013; He et al., 2013; Hirak, et al., 2014; Huang et al., 2013; Lewis et al., 2012; Lin et al., 2015; Luo et al., 2016; Margolis et al., 2014; Ng et al., 2014; Niemenmaa et al., 2012).

Similarly in Health Informatics, healthcare organizations and companies are trying to digitize, store as well as manipulate medical data efficiently and cost effectively. We recommend that more serious efforts should be made to use predictive analytic models and apply risk adjustment methodologies embedded within data analysis platforms and thus allowing the healthcare organizations to make predictions for the cost and help develop strategies on population health management. In addition with Big Data Analytics, healthcare organizations can be helped to perform risk assessments and adjustments so the action plans fall within the budget and at the same time be helped to develop better treatment guidelines, plan care management strategies, measure physician performance, and also work closely with insurance companies (European Bioinformatics Institute (EBI), 2016; Global Alliance for Genomics and Health (GA4GH) 2016; The Precision Medicine Initiative® (PMI) Cohort Program, 2016; The Phenotype-Genotype Integrator (PheGenI), 2016;The International Medical Informatics Association (IMIA), 2016; United Nations (UN), 2016; World Health Organization (WHO), 2016).

The Bioinformatics discipline together with that of Health Informatics should facilitate the development of big data analytic technology as in the case of the integration of genetic test results, patient-specific sequencing, expression profiling, tissue image data and overall clinical data in patients' medical records. This fusion will provide opportunities for personalized medicine, targeted drug research and therapy but also, create new challenges for big data analytics from database design, data mining, data knowledge representation, to data analytics, and clinical decisions.

The target audience of this book would be undergraduate and graduate students, practitioners, researchers, clinicians, and data scientists in the area of Bioinformatics and Health Informatics who would be exposed to the latest findings in the field, and would be helped to explore the intersections between Bioinformatics and Health Informatics as well as the new research areas brought by advancement in big data analytics, data mining, machine learning and statistical learning.

The book can also serve as a guide for professionals, researchers, clinicians, and data scientists on ways of communication and how to share opinions and exchange ideas, so as to facilitate a better fusion of Bioinformatics and Health Informatics, academic and industry research, and the improvement in the quality of people's daily health and life activities. Big data analytics will therefore support and promote such research activities.

The potential benefits to individual and the society can be great if the efforts to transform Big Data into Smart Data and from that to Wise Data, as we propose, are successful. This transformation can be achieved when the motivation of people to participate and the rewards of collaboration are high and are based on mutual respect of all stakeholders. This we think, is not an unachievable goal and all stakeholders should be working toward that direction. The fact that a good number of associations, organizations and world alliances are working toward establishing policies, regulations, and safety measures including risk analysis studies, is a very positive step.

In our mind, it is clear that genomics and health informatics data can be useful in selection of therapies and prevention. We must not forget however, that each one and each society has different value priorities, different understanding of health and quality of life.

In July 2016, an international team published a map of the human brain's cerebral cortex (Glasser et al., 2016). Despite the highest resolution ever achieved in human brain-brain connectivity map (1 cubic millimeter spatial resolution), each voxel contains tens of thousands of neurons but still it is not comparable to the mapped at single-cell resolution in the fruit fly. Neuroscientists however have started sharing and integrating big brain data following a team approach which is not easy. Using multi-modal magnetic resonance images from the Human Connectome Project (HCP) and semi-automated neuroanatomical approach for example, they have made accurate real maps of cortical architecture, function and connectivity (Glasser et al., 2016). Despite the advances in bioinformatics and the revolution that swept the genomics field decades ago, in neurobiology where big data is truly big and a single neuroimaging data measures in terabytes the challenges are many. Brain mappers face richer sets of imaging and electrophysiological data which goes without saying that a concentrated effort is needed to produce accurate real brain maps (Glasser et al., 2016). A culture shift will be required so that scientists not only develop new computational tools but to also share and visualize the resulting data, so that we expand the limits of what is possible. Mapping the Human brain will be indeed one of the greatest future challenges that scientists will be facing which however will greatly advance our understanding of ourselves as a *Home sapiens* species, the wise one.

We, as editors of this book but also as co-authors of one of its chapters we hold the belief that "Making big data smart implies ability to technically connect relevant data to identify patterns. Turning smart data to wise implies the integral consideration of social, cultural and political considerations to ensure the holistic wellbeing of the individual, but also a community that promotes health, wellness and spiritual integrity above anything else".

ORGANIZATION OF THE BOOK

The book is organized into 16 chapters. A brief description of each of the chapters follows:

Chapter 1 reviews the emerging advances of Bioinformatics which have already contributed toward the establishment of better next generation medicine and medical systems by putting emphasis on improvement of prognosis, diagnosis and therapy of diseases including better management of medical

systems. The authors of this chapter explore ways by which the use of Bioinformatics and Smart Data Analysis offer solutions to challenges in the fields of genomics, medicine and Health Informatics. They focus on Smart Data Analysis and ways needed to filter out the noise. The chapter addresses challenges researchers and data analysts are facing in terms of the developed computational methods used to extract insights from NGS and high-throughput screening data. The authors propose the concept "Wise Data" reflecting the distinction between individual health and wellness on the one hand, and social improvement, cohesion and sustainability on the other, leading to more effective medical systems, healthier individuals and more socially cohesive societies.

Chapter 2 is an overview of Bioinformatics in relation to data mining, data visualization, secretome analysis, mass spectrometry, chemical cross-linking reagents, Software Product Line (SPL), protein kinase, microRNAs (miRNAs, clinical bioinformatics and cancer. The author focuses on how bioinformatics helps reveal the wealth of biological information hidden in the large amounts of data and obtain a clearer insight into the fundamental biology of organisms. The author finally points to the need of the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve the formal and practical problems arising from the management and analysis of biological data.

Chapter 3 identifies the existing challenges in the management of computational techniques in protein structure prediction such as comparative modeling, threading and ab initio modelling. In this chapter the authors elaborate on the three phases of modelling (the pre-modelling analysis phase, model construction and post-modelling refinement phases), the programs available for performing each, issues, possible solutions and future research areas.

Chapter 4 examines the idea of personalized medicine system as an evolution of holistic approach of treatment and in more evidence based manner. The authors begin the chapter with an introduction of how body systems work naturally and examine the impact of modern medicine on overall health, followed by a historical background and brief review of literature providing the description that the concept of personalized medicine is not new but a very old ideology which stayed neglected until the development in the field of medical genetics, followed by the role of omics in modern medicine, the comparison of modern medicine and personalized medicine, medical concepts relevant to proteomics in personalized medicine, impact of proteomics in drug development and clinical safety and finally closing the chapter with future prospects and challenges of proteomics in personalized medicine.

Chapter 5 presents an analysis of issues and concerns in managing cloud computing and have examined virtual document storage platforms as providers of amenities with minimal expense in the corporate society. The authors discuss how cloud computing is appeasing to owners of clinics and businesses as it banishes the requirement of planning, provisioning, and allows corporations to advance their filling systems according to service demand. The authors identify the importance that Medical practices urgently need to revolutionize their storage standards, to keep up with the growing population.

Chapter 6 examines how Informatics and Data Analytics supports Exposome-based discovery Part I. The totality of exposures from conception onwards, simultaneously identifying, characterizing and quantifying the exogenous and endogenous exposures and modifiable risk factors that predispose to and predict diseases throughout a person's life span are discusses. The authors discuss how the unravelling of the exposome implies that both environmental exposures and genetic variation can reliably be measured simultaneously. To achieve this, they claim, we need to bring together a comprehensive array of novel technologies, data analysis and modelling tools that support efficient design and execution of exposome studies. This requires an innovative approach bringing together and organizing environmental, socio-economic, exposure, biomarker and health effect data; in addition, this effort includes all the procedures

and computational sequences necessary for applying advanced bioinformatics coupling advanced data mining, biological and exposure modelling so as to ensure that environmental exposure-health associations are studied comprehensively. The authors explore the type of novel tools exposome studies will require to address the complexity of emerging environmental health issues. Critical for success will be the ability to bring together existing geospatial, environmental, health and socioeconomic data, and to collect new high resolution data using innovative environmental micro-sensors, remote sensing or other community and omics/systems biology based approaches to describe the exposome for e.g. endocrine disruption-related syndromes and sex-related changes (menopause), neurodegenerative or respiratory diseases.

Chapter 7 examines how Informatics and Data Analytics supports Exposome-based discovery Part II. The authors outline the connectivity-based methodology and introduce a new exposome-based paradigm for interdisciplinary scientific work in the area of environment and health. It is an approach that builds on the exploration of the interconnections between the co-existence of multiple stressors and the different scales of biological organization that together produce the final adverse health effect. The authors state that their approach marks a clear departure from the conventional paradigm, which seeks to shed light on the identification of singular cause-effect relationships between stressors and health outcomes. It entails creating a new way of combining health-relevant information coming from different disciplines, including (but not limited to) environmental science, epidemiology, toxicology, physiology, molecular biology, biochemistry, mathematics and computer science. According to the connectivity approach, all factors affecting the internal and external exposome are treated as co-variates, rather than just as confounders. The functional integration of these different information classes into a unique framework results in understanding the complex interaction between the genome and exposure to environmental factors that determines physiological response to environmental insults and, ultimately, the onset or exacerbation of adverse health outcomes.

Chapter 8 presents an analysis of issues and concerns in managing the big omics data generated in the recent decade which flooded the internet with transcriptomic, genomics, proteomics and metabolomics data. The authors refer to a number of software, tools, and web-servers that have developed to analyze the big data omics. In this review, the authors integrate the various methods that have been employed over the years to interpret the gene regulatory and metabolic networks. They illustrate random networks, scale-free networks, small world network, bipartite networks and other topological analysis which fits in biological networks. They point to the fact that transcriptome to metabolome network is of interest because of key enzymes identification and regulatory hub genes prediction. This chapter also provides an insight into the understanding of omics technologies, generation of data and impact of in-silico analysis on the scientific community.

Chapter 9 focuses on protein docking as an integral part to structure-based drug design and molecular biology. The authors argue that the recent surge of big data in biology, the demand for personalised medicines, evolving pathogens and increasing lifestyle-associated risks, asks for smart, robust, low-cost and high-throughput drug design. Computer-aided drug design techniques allow rapid screening of ultra-large chemical libraries within minutes. This the authors think is immensely necessary to the drug discovery pipeline, which is presently burdened with high attrition rates, failures, huge capital and time investment. With increasing drug resistance and difficult druggable targets, there is a growing need for novel drug scaffolds which is partly satisfied by fragment based drug design and de novo methods. This chapter discusses various aspects of protein docking and emphasises on its application in drug design.

Chapter 10 focuses on the design process of an effective and efficient dashboard which displays management information for an Electronic Health Record (EHR) in Dutch long-term and chronic healthcare. The authors presents the actual design and realization of a management dashboard for the YBoard 2.0 system, which is a popular solution on the Dutch market. The authors based their design decisions in this investigation on human perception and computer interaction theory, in particular Gestalt theory. The empirical interviews with medical professionals supplemented valuable additional insights into what the users wanted to see most of all in a dashboard in their daily practices. This study successfully shows how effective and efficient dashboard design can benefit from theoretical insights related to human perception and computer interaction such as Gestalt theory, in combination with integrated end user requirements from daily practices.

Chapter 11 analyses and compares how search engines and social media are two different online data sources where search engines can provide health related queries logs and Internet users' discuss their diseases, symptoms, causes, preventions and even suggest treatment by sharing their views, experiences and opinions on social media. This chapter hypothesizes that online data from Google and Twitter can provide vital first-hand healthcare information. An approach is provided for collecting twitter data by exploring contextual information gleaned from Google search queries logs. Furthermore, the authors investigated whether it is possible to use tweets to track, monitor and predict diseases, especially Influenza epidemics. The authors argue that their obtained results show that healthcare institutes and professional uses of social media helps provide up-to date health related information and interact with public. Moreover, the proposed approach is beneficial for extracting useful information regarding disease symptoms, side effects, medications and to track geographical location of epidemics affected area.

Chapter 12 addresses the issue of how the semi-supervised learning has become one of the most interesting fields for research developments in the machine learning domain beyond the scope of supervised learning from data. Medical diagnostic process works mostly in supervised mode, but in reality, we are in the presence of a large amount of unlabeled samples and a small set of labeled examples characterized by thousands of features. This problem is known under the term "the curse of dimensionality". In this chapter, the authors propose, as solution, a new approach in semi-supervised learning that they call Optim Co-forest. The Optim Co-forest algorithm combines the re-sampling data approach with two selection strategies. The first one involves selecting random subset of parameters to construct the ensemble of classifiers following the principle of Co-forest. The second strategy is an extension of the importance measure of Random Forest (RF). Experiments on high dimensional datasets confirm the power of the adopted selection strategies in the scalability of their method.

Chapter 13 focuses on how Medical Diagnosis has been gaining significance in everyday life and the need for good medical analysts. The diseases and their symptoms are highly varying and there is always a need for a continuous update of knowledge needed for the doctors and the associated stakeholders' i.e. medical analyst. The diseases fall into different categories and a small variation of symptoms may leave to different categories of diseases. This is further supplemented by the medical analysts for a continuous treatment process. A small level of error in disease identification introduces overhead in diagnosis and further consequences in treatment.

Chapter 14 explores pathway information for cancer detection which helps to find co-regulated gene groups whose collective expression is strongly associated with cancer development. In this chapter the authors propose a collaborative multi-swarm binary particle swarm optimization (MS-BPSO) based gene selection technique that outperforms to identify the pathway marker genes. The authors have compared their proposed method with various statistical and pathway based gene selection techniques for different

popular cancer datasets as well as a detailed comparative study is illustrated using different meta-heuristic algorithms like binary coded particle swarm optimization (BPSO), binary coded differential evolution (BDE), binary coded artificial bee colony (BABC) and genetic algorithm (GA). Experimental results show that the proposed MS-BPSO based method performs significantly better and the improved multi swarm concept generates a good subset of pathway markers which provides more effective insight to the gene-disease association with high accuracy and reliability.

Chapter 15 identifies the existing challenges and concerns in the management of the diseases and health problems managers of the Unified Health System face and the high costs. The World Health Organization focused on prevention of chronic diseases to prevent millions of premature deaths in the coming years, bringing substantial gains in economic growth by improving the quality of life. Few countries appear to be aimed at prevention, if not note the available knowledge and control of chronic diseases and may represent an unnecessary risk to future generations. The authors support the idea that early diagnosis of these diseases is the first step to successful treatment in any age group. Their objective is to build a model, from the establishment of a Bayesian network, for the early diagnosis of nursing to identify eating disorders bulimia and anorexia nervosa in adolescents, from the characteristics of the DSM-IV and Nursing Diagnoses. The need for greater investment in technology in public health actions aims to increase the knowledge of health professionals, especially nurses, contributing to prevention, decision making and early treatment of problems.

Chapter 16 explores how predicting patterns to extract knowledge can be worth the efforts despite the difficulties. The authors argue that in order to accomplish a task good knowledge of the type of data is needed together with adequate time for analysis and also adapt it to the algorithms and technologies used together with the appropriate methodologies. In this chapter a real case is examined following all this process. In particular, a classification problem is shown as an example of using machine learning methodologies to find out whether a hospital patient should be admitted or not in Cardiology department.

Miltiades D. Lytras
Deree – The American College of Greece, Greece

Paraskevi Papadopoulou
Deree – The American College of Greece, Greece

REFERENCES

Alyass, A., Turcotte, M., & Meyre, D. (2016). From Big Data analysis to personalized medicine for all: Challenges and opportunities. *BMC Medical Genomics*, *8*(33). PMID:26112054

Chen J, Qian F, Yan W. (2013). Translational biomedical informatics in the cloud: present and future. *Biomed Res Int.* 658925.

European Bioinformatics Institute (EBI). (2016). Retrieved November 21, 2016, from http://www.ebi.ac.uk/

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., & Van Essen, D. C. et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. doi:10.1038/nature18933 PMID:27437579

Global Alliance for Genomics and Health (GA4GH). (2016). Retrieved November 21, 2016, from http://www.ebi.ac.uk/about/news/press-releases/ewan-birney-leads-global-alliance-genomics-and-health

Greene, A. C., Giffin, K. A., Greene, C. S. & Moore, J. H. (2015). Adapting bioinformatics curricula for Big Data. *Briefings in Bioinformatics-Bioinformatics Curricula,* 1-8.

He, C., Fan, X., & Li, Y. (2013). Toward ubiquitous healthcare services with a novel efficient cloud platform. *IEEE Transactions on Bio-Medical Engineering*, *60*(1), 230–234. doi:10.1109/TBME.2012.2222404 PMID:23060318

Hirak, K., Hasin, A. A., Nazrul, H., Swarup, R. & Dhruba, K. B. (2014). Big Data Analytics in Bioinformatics: A Machine Learning Perspective. *Journal of Latex Class Files, 13*(9), 1-20.

Huang, H., Tata, S., & Prill, R. J. (2013). BlueSNP: R package for highly scalable genome-wide association studies using Hadoop clusters. *Bioinformatics (Oxford, England)*, *29*(1), 135–136. doi:10.1093/bioinformatics/bts647 PMID:23202745

Lewis, S., Csordas, A., Killcoyne, S., Hermjakob, H., Hoopmann, M. R., Moritz, R. L., & Boyle, J. et al. (2012). Hydra: A scalable proteomic search engine which utilizes the Hadoop distributed computing framework. *BMC Bioinformatics*, *13*(1), 324. doi:10.1186/1471-2105-13-324 PMID:23216909

Lin, W., Dou, W., Zhou, Z., & Liu, C. (2015). A cloud-based framework for home-diagnosis service over big medical data. *Journal of Systems and Software*, *102*, 192–206. doi:10.1016/j.jss.2014.05.068

Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomedical Informatics Insights*, *8*, 1–10. doi:10.4137/BII.S31559 PMID:26843812

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., & Green, E. D. et al. (2014). The National Institutes of Healths Big Data to Knowledge (BD2 K) initiative: Capitalizing on biomedical Big Data. *Journal of the American Medical Informatics Association*, *21*(6), 957–958. doi:10.1136/amiajnl-2014-002974 PMID:25008006

Ng, K., Ghoting, A., Steinhubl, S. R., Stewart, W. F., Malin, B., & Sun, J. (2014). PARAMO: A PARAllel predictive MOdeling platform for healthcare analytic research using electronic health records. *Journal of Biomedical Informatics*, *48*, 160–170. doi:10.1016/j.jbi.2013.12.012 PMID:24370496

Niemenmaa, M., Kallio, A., Schumacher, A., Klemela, P., Korpelainen, E., & Heljanko, K. (2012). Hadoop-BAM: Directly manipulating next generation sequencing data in the cloud. *Bioinformatics (Oxford, England)*, *28*(6), 876–877. doi:10.1093/bioinformatics/bts054 PMID:22302568

Tenenbaum, J. D. (2016). Translational Bioinformatics: Past, Present, and Future. *Genomics, Proteomics & Bioinformatics*, *14*(1), 31–41. doi:10.1016/j.gpb.2016.01.003 PMID:26876718

The International Medical Informatics Association (IMIA). (2016). Retrieved November 21, 2016, from http://imia-medinfo.org/wp/welcome-to-imia-2/

The Phenotype-Genotype Integrator (PheGenI). (2016). Retrieved November 21, 2016, from https://www.ncbi.nlm.nih.gov/gap/phegeni

The Precision Medicine Initiative® (PMI) Cohort Program. (2016). Retrieved November 21, 2016, from https://www.nih.gov/precision-medicine-initiative-cohort-program

United Nations (UN). (2016). *Sustainable Development Goals. 17 Goals to Transform the World.* Retrieved November 21, 2016, from http://www.un.org/sustainabledevelopment/sustainable-development-goals/

World Health Organization (WHO). (2016) Retrieved November 21, 2016, from http://www.who.int/en/