Scientific
Research

The Impact of Online Networks and Big Data in Life Sciences

Ruchita Gujarathi[1,2,3,4], Fabricio F. Costa[1,2,3,5*]

[1]Genomic Enterprise, Chicago, USA
[2]DataGenno Interactive Research Ltda, Campos dos Goytacazes, Rio de Janeiro, Brazil
[3]1871: DataGenno Interactive Research, Chicago, USA
[4]Department of Bioengineering, University of Illinois at Chicago, Chicago, USA
[5]Cancer Biology and Epigenomics Program, Ann and Robert Lurie Children's Hospital of Chicago Research Center
and Department of Pediatrics, Northwestern University's Feinberg School of Medicine, Chicago, USA
Email: *fcosta@luriechildrens.org, fcosta@datagenno.com, fcosta@genomicenteprise.com

ABSTRACT

Advances in Information Technology (IT) have enhanced our ability to gather, collect and analyze information from individuals and specific groups of people online. The emergence of online networks has facilitated connections between individuals by leveraging data exchange in a variety of fields. Online networking in life sciences transforms data collection into actionable information that will improve individual and population health, deliver effective therapies and, consequently, reduce the cost of healthcare. These novel tools might also have a direct impact in personalized medicine programs, since the adoption of new products by health care professionals in life sciences and peer-to-peer learning could be improved using social networks and big data analytics. However, one of the main concerns of information exchange online is data privacy. In this article, we will review how online networks and big data analytics are impacting the life sciences sector.

KEYWORDS

1. Introduction

In the early 1650s, when digital technology (computers/internet/online networks) did not exist, public places such as coffeehouses were considered as the best way to socialize [1]. People would stay in coffeehouses not only to drink coffee but also to read, discuss the latest pamphlets and news-talk, rumors, gossip and to check for new mail [1]. It was the social-networking site of its day [1]. Communication has become easier with the industrial revolution and the advent of digital technology. Online networks nowadays allow people to communicate in more efficient ways than in primeval times. According to the state of the news media, in 2012, in United States alone, 44% of people use smart phones, 18% use tablets

and 77% use laptop or desktop in their day-to-day life [2]. With the improving technology, online networking has increased to a level where sometimes it is accused as the enemy of productivity.

Coffee is still a traditional drink but, eventually, social networking sites such as Facebook, Twitter, Skype, LinkedIn, etc. have diminished the importance of the coffeehouses for socializing purposes. Such sites are established to maintain and cultivate social relationship, close friendship to casual acquaintances online [3]. There is no geographical limitation for people to socialize due to wireless communication. It is now extremely interactive, fast and less time consuming to socialize than it was in the 17th century. Not just socializing, but to share and backup data on the internet have also progressed with the advancement of different technologies.

*Corresponding author.

The health field also progressed with advances in technology. During an early era, patient's information was recorded on paper charts and stored at one place. An Electronic Medical Record (EMR) is a digital version of a paper chart that contains all of a patient's medical history from one practice and can be accessed at any location in real-time [4]. It can be also called Electronic Health Record (EHR) when it goes beyond just clinical data with a broader view of the patient's care. In addition, an EMR contains the standard medical and clinical data gathered in one provider's office, while EHRs are designed to contain and share information from all the providers involved in a patient's care [5]. The number of hospitals with a basic electronic health record system in place jumped to 44% in 2012, which is a 17 percentage point increase from 2011 and has tripled since 2010 [6].

In the last five years, Social Networking Sites (SNS) have also grown from a niche to a mass online activity, in which tens of millions of internet users are engaged, both in their leisure time, and at work [3]. Social media has opened new opportunities to collaborate with scientific and healthcare provider communities on research and product development for the life sciences industry. Even with a lack of regulatory clarity as the largest barrier, a trend towards social networks and the use of social media for clinical studies by pharmaceutical companies, information exchange between physicians and scientists and other applications in life sciences are starting to become reality [7]. Increasing growth in the amount of biomedical information, also called Big Data, provides an even greater opportunity to improve patient care with the help of future personalized medical programs [8]. It is now clear that the combination of an individual's genetics information with clinical data will facilitate personalized medicine applied to life sciences in clinical studies and for the development of new drugs and therapies [8].

In this article, we will discuss the current state and future developments occurring in the field of online networking applied to the life sciences sector. Moreover, an overview of emerging health technologies, network approaches to diseases, data sharing and implications of personalized medicine programs in the life sciences sector will be described.

2. A New Open Source Model?

Big data, a new information technology concept, is designed to store ample amounts of information gathered using different sources and utilize heavy server processing in order to capture, discover and analyze the data in a secure manner [8]. Big data in biomedicine is coming from two ends; the genomics-driven end (genotyping, gene expression, and now next-generation sequencing data or NGS); and the payer-provider end (electronic medical records, pharmacy prescription information, insurance records, etc.) [9] [see also **Figure 1**]. Traditional tools do not provide the complete advantage, which big data analytics can provide these days. It has created a bridge between life sciences and the medical industry to facilitate more interactions between patients, doctors and pharmaceutical companies. Online networking and web-based tools are becoming very common and considered the best and fastest way of communication since there is real-time feedback. Additionally, it is important to note that sharing information reduces cost and saves time. It also improves patient-physician relations as they interact directly without any mediator.

Advanced technology has increased the availability of mobile apps, which allows patients to access their information easily and quickly [**Figure 1**]. Social networking is emerging as a new standard as patients expect faster response and organizations strive to reduce their costs. On the other hand, if the information is blindly shared without acknowledging rules and regulations it could be leaked. Reliable sources play an important role when collecting data in the web. Privacy issues have been the prime disadvantage for the use of the internet and online data storage.

Several decades ago, medical information was stored in registers, research panels and static databases. Accessing such information was time consuming, expensive and geographically limited. However, the arrival of EMRs with the digital technology era, dynamic databases and processing technologies are fundamentally changing the face of healthcare [10,11]. Such health systems are rapid, less time consuming and can be accessed anywhere. Clinicians choosing products that allow data sharing in order to participate in the Medicare and Medicaid Incentive Programs should not be held captives to EHRs that reduce their efficiency and strangle innovation [10,11]. One successful example in the improvement of patient care is the use of EHRs by the Memphis information-sharing network with a Health Information Exchange (HIE) system [12].

Health-related social media sites, such as Patients LikeMe, Sermo, Mediguard, Spine Connect, Ozmosis, Inspire and others that enable people to share information with others who face similar health challenges, are gaining popularity [7]. However, there is no "universal" platform that provides a connection between both basic and translational research and the development of platforms for clinical and scientific data storage are warranted. For example, Datagenno (http://www.datagenno.com/), a new platform, is a database with storage capacity that could be used by healthcare professionals to access patient information in real-time [13]. Data Genno could also bridge the gap between healthcare professionals and
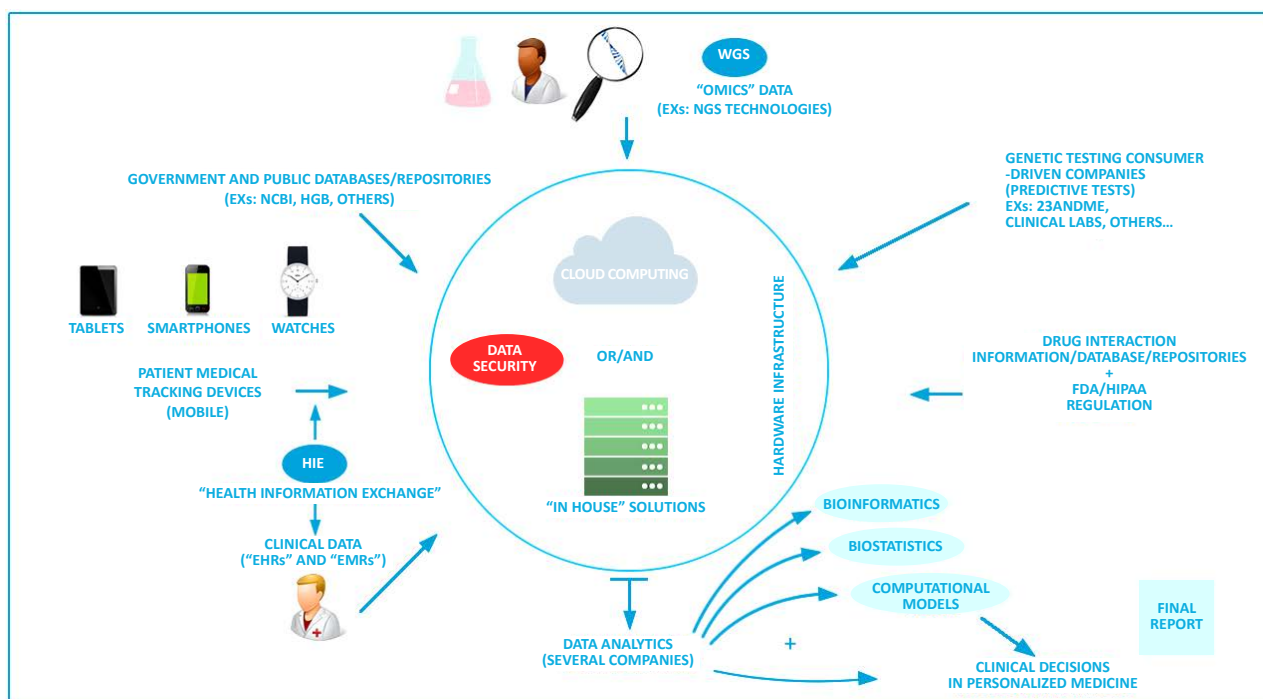
Figure 1. This figure depicts how digital technologies such as the internet, online networks, mobile devices, tracking devices, etc can facilitate the collection of different sources of information in life sciences and medicine. With advances in Information Technology (IT), data obtained from clinical (left lower side), research (center up), different mobile/tracking devices (left middle side) and information from public institutions (left upper side) and private companies/databases (right upper side) have been accumulating in "in house" servers and in the cloud (center). All this "big data" can now be stored and shared using online tools and networking. This new concept enables data and models to be shared by health professionals and patient communities in an open-source manner. The collected data is secured by encryption/de-identification and analyzed using different tools such as bioinformatics, biostatistics and computational predictive models to facilitate personalized medicine decisions. Clinical reports for translational medicine are generated and could be used in the clinic by health professionals in order to improve patient care. Abbreviations: EMR: Electronic Medical Record, EHR: Electronic Health Record, HGB: Human Genome Browser; HIE: Health Information Exchange, WGS: Whole Genome Sequencing, FDA: Food and Drug Administration, HIPPA: Health Insurance Portability and Accountability Act, NCBI: National Center for Biotechnology Information. Figure design by Eduardo B. F. Junior.

scientists, combining clinical, genetic and genomic information from patients since it provides clinical data along with all genetic information available from patients [13,14]. We believe that new solutions with more data security to share health information will become available in the years to come and patients will become more empowered by their clinical data.

3. Network Approaches to Diseases

The development of different Computer Aid Diagnostic (CAD) tools and the mining of medical record systems has increased remarkably since 2002, and it is expected that the emergence of large EHR-based databases will only stimulate this activity further [15]. Information exchange between doctors, patients and pharmaceutical companies could reduce medical errors, health discrepancy, inefficiency, inappropriate care and duplicative care, which would save the cost and time of patients, doctors and insurance companies.

Effective implementation of online networking solutions with the advent of big data has the potential to not only increase the predictive values of the clinical information but also to decrease the amount of time spent on analyzing the information on diseases. It also provides reliable sources for the information requested. Social networks present an unprecedented opportunity for engaging and interacting with patients, prescribers, and targeted health communities [16]. Thus, approaching social media as a cross-brand, cross-functional program will help organizations take advantage of existing process infrastructure [16].

Social networking could also help the industry gain insights into how people use and react to pharmaceuticals and medical devices; and how they benefit from them. Such information could be applied into the product development process as a "lean" process to test new products [17]. This new channel could also help the health care industry build brand equity, educate the scientific

community and consumers, increase patient compliance with prescribed treatments, and ultimately assemble information to use it in improving the quality and effectiveness of healthcare [16]. For Example, ResearchGate is a social network for scientists in a variety of fields to exchange information and collaborate that already has more than 3 million members with a profile webpage, sharing information about specific fields of their interest and also creating groups of discussion. ResearchGate has been called the Facebook of Science [7]. Several other solutions are becoming available in the web as networking tools to approach science and medicine.

In 2010, the pharmaceutical and biotechnology industries spent more than $100 billion and generated only 21 successful US Food and Drug Administration (FDA) approved New Chemical Entities (NCEs) [18]. Drug development takes too long, and it is too costly with unpredictable outcomes. Hence, pharmaceutical companies have begun to invest in Public Private Partnerships (PPPs) to overcome the current stagnant situation. PPP's objectives tend to focus on the creation of new tools and methods to support faster and more cost effective drug development, rather than pharmaceutical drug discovery and development itself [18].

Starting with investments in basic science, academic institutions excel in research for new discoveries and are now working, in a variety of ways, to ensure that their discoveries translate to better patient health. Patient driven research groups are continually innovating and have tenaciously pioneered creative new models of collaboration with academia, industry, and government agencies [19]. Almost all of the top pharmaceutical companies, biotechnology firms and medical device manufacturers have some social network presence; some are partnering with third-party social networks such as PatientsLikeMe and Sermo to communicate and collaborate with external stakeholders [20]. These collaborations will empower companies to facilitate drug discovery, enable personalized medicine programs to be publicized among people and give patients enough basic knowledge to make their own healthcare decisions.

4. Data Sharing in Life Sciences

In the life sciences industry, social networks are already improving applications such as market research, physician relations, clinical trials, and medical education. According to Deloitte's research, 65% of surveyed life science professionals said that their company is using or plans to use social networks in some capacity combined with the right application [21]. People use online social networks when they want information from people in similar circumstances. Establishing the right kind of chemistry can be as simple as allowing physicians to

speak with other physicians and patients to speak with other patients. For example, in response to physician demands, Pfizer and Sermo developed a service called AskRx, which allows physicians to request scientific information from medically trained Pfizer personnel [21].

Online social networking applications have become the fastest growing mechanism to exchange personal and professional information [22]. Although computers and the internet have become faster, we have a lack of computational infrastructure that is needed to securely generate, maintain, transfer, and analyze large-scale information in biomedicine and to integrate *omics* data (which includes all information generated by new DNA sequencing technologies such as Whole Genome Sequencing, transcriptome sequencing and others) with other data sets, such as clinical information from patients and mobile data collected through the web from different sources [8] [see also **Figure 1**]. Medical professionals who use social networking sites have unique responsibilities, since information posting and sharing could violate patient privacy. Life sciences companies should also consider that communication on social networks is highly public, which allows small issues to escalate into public incidents [21]. The basis of every patient-physician interaction is trust, because shared personal health information is kept confidential and only used exclusively for the patient's benefit. Thus, the main barrier for the adoption of online networks in health care and medicine will be the privacy and security of the data [14].

Privacy in healthcare consists of anonymity and confidentiality. It is a major concern for patients, providers, regulators, and legislators [10]. Millions of records, including personal information, are transferred through the web and different mobile devices [**Figure 1**]. Private medical information disclosure can create a negative impact on future job assignments and career development. Healthcare vendors should adapt modern technology to provide the best secrecy to their customers.

Many patients engage in clinical trials where if they come to an end without the result they wanted, they are easily motivated to share the data they contributed to the trial [23]. As discussed, concerns about privacy are always an issue when sharing clinical data. In order to overcome the privacy issue, cloud computing and the use of "in house" computer infrastructure has been established. However, in order to comply to HIPAA requirements and to a "Protected Health Information" (PHI), cloud solutions will need constant auditing and "security" officers to protect the information. In that regard, suitable solutions are in practice that balances privacy, data utility and system usability [see **Figure 1**]. Could computing is used when in-house storage systems are not capable of storing accumulating data (for example, in the case of genetics and WGS data). It also facilitates data storage

and transfer. Cloud computing is ideal for projects that require periodic computational bursts, rapid prototyping, or fast turnaround time [24].

Another challenge is to transfer data from one location to another since it mainly done mailing external drives containing encrypted data. Security and privacy of the data from individuals is also a concern before and during data transfer. Possible solutions to these issues include the use of better security systems with advanced encryption and de-identification algorithms, like the ones used by banks in the financial sector to secure privacy of their clients [8]. The future of Big Data in life sciences is full of insecurities and challenges, but changes in several sectors are occurring, especially sectors that deal with security of online information, in order to address these issues [8].

5. Implications for Personalized Medicine

Personalized medicine is a rapidly advancing practice that uses an individual's unique molecular characteristics to make decisions regarding the diagnosis, treatment and prevention of diseases. It is a possibility that a certain medicine will work for everyone else but may not work for one specific person. It is already known that each person has a unique genome and gene structure. In 2003, after more than a decade of research, the US Department of Energy and the National Institutes of Health studied the order and the locations of the 3 billion units of DNA that makes human genome under Human Genome Project (HGP) [25]. After this milestone, it is believed that personalized genomics will be able to help individuals in health decisions and change their lives depending on the results obtained from predictive genetic tests. Personalized genomics will also help in the understanding of human variations and their impact in common diseases, thus facilitating the development of more specific drugs [26].

With advances in technology, we will see a shift in population-based healthcare to personalized medicine that includes targeted diagnostics and treatment based on each person's history [26]. Great challenges, including the use of more specific drugs with fewer side effects and less toxicity lie ahead for personalized medicine to be applied in the clinic [26]. The main advantage of the use online networks in personalized medicine programs is for patients, who do not have to wait for years for a clinical trial, to monitor and participate in real-world natural experiments.

Personalized medicine could improve both the effectiveness and efficiency of health management at the individual level [14]. Risk factors and genetic susceptibilities can be predicted earlier to detect the issues before using genetic testing and Whole-Genome Sequencing (WGS)

technology. WGS is a laboratory process that determines the complete DNA sequence of an individual [27]. This may persuade a patient to make changes in his diet and/or lifestyle. Early screening can resolve issues predicted before they appear and can save a person's life. It could also help to reduce more serious conditions by taking preventive medicine decisions. Personalized medicine will help health professionals to figure out faster which medicines works for the individuals and which ones do not instead of utilizing universal medicine with blockbuster drugs. In addition, patient and doctor relationship can be more interactive due to better diagnosis and more effective treatments.

Although personalized medicine is in its early stages, we believe that a big shift in targeted therapies based on an individual's history will occur. The knowledge of the genetic basis of human disease is ushering a new trend in drug development that is focused on targeted therapies. Great challenges, including the use of more specific drugs with fewer side effects and less toxicity lie ahead for personalized medicine to be applied to clinical medicine [28]. However, the integration of these new online and digital solutions such as online networking and big data analytics with personalized medicine programs will greatly impact in health decisions.

6. Conclusions and Future Directions

Online Networks and Big Data analytics are growing fields that will greatly impact healthcare. Life science companies will need to adapt to the digital technology era in order to develop more dynamic and cost-efficient drugs and therapies. Online networking will not replace healthcare professionals but it will aid them in gaining and delivering high quality care. Online networking in life sciences using big data will improve individual and population health, deliver effective therapies and reduce the cost of care. It will also engage physicians to help improve the adoption of new products and companies, thus facilitating peer-to-peer learning. Social networking already assists companies to execute market research cheaper and faster, providing information to speed up clinical trials. Personalized Medicine programs will also be impacted by these new technologies. However, privacy issues have to be taken into account when dealing with health information. In conclusion, we foresee a bright future if suitable use of online networks and big data analytics are implemented in life sciences in order to improve patient care while greatly reducing costs.

Acknowledgements

REFERENCES

[1] T. Standage, "Social Networking in the 1600s," *New York Times*, 2013.
http://www.nytimes.com/2013/06/23/opinion/sunday/social-networking-in-the-1600s.html?pagewanted=all&_r=0

[2] A. Mitchell, T. Rosenstiel and L. Christian, "Mobile Devices and News Consumption: Some Good Signs for Journalism|State of the Media," *The State of the News Media* 2013.
http://stateofthemedia.org/2012/mobile-devices-and-news-consumption-some-good-signs-for-journalism/

[3] R. Cachia, "Social Computing : Study on the Use and Impact of Online Social Networking," 2008.
http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=1884

[4] S. Terry and P. Terry, "Power to the People: Participant Ownership of Clinical Trial Data," *Science Translational Medicine*, Vol. 3, No. 69, 2011, pp. 1-4.
http://dx.doi.org/10.1126/scitranslmed.3001857

[5] "Definition and Benefits of Electronic Medical Records (EMR)|Providers & Professionals|HealthIT.gov," *HealthIT.gov*. The Official Site for Health IT Information, 2013.
http://www.healthit.gov/providers-professionals/electronic-medical-records-emr

[6] Reuters, "Hospitals Speed Transition to Electronic Medical Records, Study Says," *Breaking News and Opinion on the Huffington Post*, 2013.
http://www.huffingtonpost.com/2013/07/08/hospitals-digital-electronic-medical-records-transition_n_3563071.html

[7] M. P. Coutinho and F. F. Costa, "Social Media in Science and Medicine," *G.I.T. Laboratory Journal*, Vol. 7, 2011, pp. 10-12.

[8] F. F. Costa, "Big Data in Biomedicine," *Drug Discovery Today*, 2013, in Press.
http://dx.doi.org/10.1016/j.drudis.2013.10.012

[9] K. Miller, "Big Data Analytics in Biomedical Research," *Biomedical Computation Review*, 2012, pp. 14-21.
http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Big+data+analytics+in+Biomedical+Reseach#2

[10] B. Malin, K. El Emam and C. O'Keefe, "Biomedical Data Privacy: Problems, Perspectives, and Recent Advances," *Journal of the American Medical Informatics Association*, Vol. 20, No. 1, 2013, pp. 2-7.
http://dx.doi.org/10.1136/amiajnl-2012-001509

[11] K. Mandl and I. Kohane, "Escaping the EHR Trap—The Future of Health IT," *New England Journal of Medicine*, Vol. 366, 2012, pp. 2240-2242.
http://dx.doi.org/10.1056/NEJMp1203102

[12] M. E. Frisse, K. B. Johnson, H. Nian, C. L. Davison, C. S. Gadd, *et al.*, "The Financial Impact of Health Information Exchange on Emergency Department Care," *Journal of the American Medical Informatics Association*, Vol. 10, 2011, pp. 1-6.

[13] F. F. Costa, L. S. Foly and M. P. Coutinho, "DataGenno: Building a New Tool to Bridge Molecular and Clinical Genetics," *The Application of Clinical Genetics*, Vol. 4, 2011, pp. 45-54.
http://dx.doi.org/10.2147/TACG.S17597

[14] F. F. Costa, "Social Networks, Web-Based Tools and Diseases: Implications for Biomedical Research," *Drug Discovery Today*, Vol. 18, No. 5-6, 2013, pp. 272-281.
http://dx.doi.org/10.1016/j.drudis.2012.10.006

[15] W. T. Kerr, E. P. Lau and G. E. Owens, "The Future of Medical Diagnostics: Large Digitized Databases," *Yale Journal of Biology and Medicine*, Vol. 85, No. 3, 2012, pp. 363-377.

[16] Pricewaterhouse Coopers, "Joining the Conversation: Life Science Industry Ventures into Social Media Despite Regulatory Uncertainty," *Knowledge Line*, 2011.

[17] S. Blank, "Why the Lean Start-Up Changes Everything," *Harvard Business Review*, 2013.
http://hbr.org/2013/05/why-the-lean-start-up-changes-everything

[18] T. C. Norman, C. Bountra, A. M. Edwards, K. R. Yamamoto and S. H. Friend, "Leveraging Crowd Sourcing to Facilitate the Discovery of New Medicines," *Science Translational Medicine*, Vol. 3, No. 88, 2011, p. 88.
http://dx.doi.org/10.1126/scitranslmed.3002678

[19] M. Anderson, "No Time to Waste," *Science Translational Medicine*, Vol. 5, 2013.

[20] P. Keckley and M. Hoffmann, "Social Networks in Health Care: Communication, Collaboration and Insights," *Deloitte*, 2010.
http://www.healthcarevisions.snapmonkey.net/f/2010_Deliotte_Social_Networks.pdf

[21] T. Hisey, C. Franck and M. Hoffmann, "To Friend or Not ? New Insights about Social Networks in the Life Sciences Industry," *Deloitte*, 2010.
http://www.deloitte.com/assets/Dcom-Mexico/Local%20Assets/Documents/mx(es-mx)Social%20Networking%20in%20Life%20Sciences_2010.pdf

[22] L. Thompson, *et al.*, "Protected Health Information on Social Networking Sites: Ethical and Legal Considerations," *Journal of Medical Internet Research*, Vol. 13, No. 1, 2011, p. e8.
http://dx.doi.org/10.2196/jmir.1590

[23] S. Terry and P. Terry, "Power to the People: Participant Ownership of Clinical Trial Data," *Science Translational Medicine*, Vol. 3, No. 69, 2011, pp. 1-4.
http://dx.doi.org/10.1126/scitranslmed.3001857

[24] V. A. Fusaro, P. Patil, E. Gafni, D. P. Wall and P. J. Tonellato, "Biomedical Cloud Computing with Amazon Web Services," *PLoS Computational Biology*, Vol. 7, No. 8, 2011, Article ID: e1002147.
http://dx.doi.org/10.1371/journal.pcbi.1002147

[25] "Personalized Medicine," *US News & World Report*, 2011.
http://health.usnews.com/health-conditions/cancer/person
alized-medicine

[26] F. F. Costa, "Genomics, Epigenomics and Personalized Medicine," *BioForum Europe*, 2009, pp. 2-4.
http://genomicenterprise.com/yahoo_site_admin/assets/do
cs/Costa_BFE1109-1.203215611.pdf

[27] ACMG Board of Directors, "Points to Consider in the Clinical Application of Genomic Sequencing," *Genetics in Medicine*, Vol. 14, No. 8, 2012, pp. 759-761.
http://dx.doi.org/10.1038/gim.2012.74

[28] M. Najafzadeh, L. D. Lynd, J. C. Davis, S. Bryan, A. Anis, M. Marra and C. A. Marra, "Barriers to Integrating Personalized Medicine into Clinical Practice: A Best-Worst Scaling Choice Experiment," *Genetics in Medicine*, Vol. 14, No. 5, 2012, pp. 520-526.
http://dx.doi.org/10.1038/gim.2011.26

Abbreviations

CAD: Computer Aid Diagnostics; EHR: Electronic Health Record; EMR: Electronic Medical Record; HGP: Human Genome Project; HIE: Health Information Exchange; IT: Information Technology; NCEs: New Chemical Entities; NGS: Next-Generation Sequencing; PHI: Protected Health Information; PPPs: Public Private Partnerships; SNS: Social Networking Sites; WGS: Whole Genome Sequencing.