

Review Article

Translational Biomedical Informatics in the Cloud: Present and Future

Jiajia Chen,^{1,2} Fuliang Qian,¹ Wenyang Yan,¹ and Bairong Shen¹

¹ Center for Systems Biology, Soochow University, Suzhou 215006, China

² School of Chemistry and Biological Engineering, Suzhou University of Science and Technology, Suzhou 215011, China

Correspondence should be addressed to Bairong Shen; bairong.shen@suda.edu.cn

Received 8 December 2012; Accepted 17 February 2013

Academic Editor: Ming Ouyang

Copyright © 2013 Jiajia Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Next generation sequencing and other high-throughput experimental techniques of recent decades have driven the exponential growth in publicly available molecular and clinical data. This information explosion has prepared the ground for the development of translational bioinformatics. The scale and dimensionality of data, however, pose obvious challenges in data mining, storage, and integration. In this paper we demonstrated the utility and promise of cloud computing for tackling the big data problems. We also outline our vision that cloud computing could be an enabling tool to facilitate translational bioinformatics research.

1. Introduction

The rate of accumulation of biomolecular data is increasing astonishingly. This information explosion is being driven by the development of low-cost, high-throughput experimental technologies in genomics, proteomics, and molecular imaging, amongst others. Success in the life sciences will depend on our ability to rationally interpret these large-scale, high-dimensional data sets into clinically understandable and useful information, which in turn requires us to adopt advances in informatics. Translational informatics, given the available data resources, is now evolving as a promising methodology that can drive the translation of laboratory data at the bench to health gains at the bedside. This “translation” involves correlating genotype with phenotype, which often requires dealing with information at all structural levels ranging from molecules and cells to tissues and organs, individuals to populations.

2. Translational Bioinformatics: Imperative to Collaborate

According to the scale of investigation, the fields of translational informatics can be roughly classified into four sub-disciplines [1]: (1) bioinformatics (molecules and cells); (2) imaging informatics (tissues and organs); (3) clinical

informatics (individuals); and (4) public health informatics (populations). Each of the subfields is directed at a particular level of research scale. Table 1 outlines the spectrum of translational bioinformatics activities. The four subfields of translational bioinformatics are compared along several dimensions, including (1) areas of research purpose; (2) data types; (3) informatics tools to support practice.

Bioinformatics traditionally concerns applying computational approaches to the analysis of massive data from genomics, proteomics, metabolomics, and the other “-omic” subfields. Such research might help better comprehend the intricate biological details at molecular and cellular levels. Imaging informatics is focused on what happens at the level of tissues and organs. The essential informatics techniques to extract and manage the biological knowledge from images are summarized in Table 1. At the individual level, clinical bioinformatics is oriented to provide the technical infrastructure to understand clinical risk factors and pathophysiological mechanisms. As for public health informatics, the stratified population of patients is at the center of interest. Such research relies on informatics solutions to study shared risk factors for disease on a population level.

At each of these levels, large amounts of experimental data are being generated. To fully understand a disease phenomenon, however, it is important to gather data at various levels and analyze them in an integrated fashion. While the four

TABLE 1: Spectrum of translational bioinformatics activities.

Subfields	Research purpose	Data types	Informatics tools
Bioinformatics	Sequencing	Sequence information	Pattern recognition [45–47]
	Structure analysis	Microarray	Data mining [48–50]
	Expression analysis	Mass spectrum	Machine learning [47, 51, 52]
	Phylogenetic analysis	SNP	Visualization [53–55]
	Structure modeling	Haplotypes	Automatic annotation [49, 55–58]
Imaging informatics	Image feature identification	DICOM	Content-based image retrieval [59, 60] Natural language processing [61, 62]
	Image segmentation	JPEG	
	Image reconstruction	TIFF	
	Image annotation	PNG	
	Image indexing	GIF	
	Image visualization	BMP	
Clinical informatics	Clinical decision support	Clinical laboratory results	Probabilistic decision-making [63]
	Clinical information access	Physical examination	Expert reasoning system [64–66]
	Electronic patient record system	Symptoms and signs	Assessment and validation vocabularies [67, 68]
	Disease reclassification	Patient history Prescriptions	Text-parsing tools [69]
Public health informatics	Tracking of infectious diseases	PHCDM-based health information	Access control [70, 71]
	Assessment clinical interventions		Information security technology [37, 72]
	Monitoring disease risk factors		Semantic and syntactic standards [73] Structured data-collection techniques [74, 75]

areas of research differ in their scientific foundations, they nevertheless share a core set of informatics methodologies, such as data acquisition systems, controlled vocabularies, knowledge representation, simulation and modeling, information retrieval, and signal and image processing, which provide a basis for their intersection.

3. Crisis Looms for Multidisciplinary Collaboration

The current push for personalized disease treatment is encouraging bioinformatics to seamlessly integrate data acquired from multiple levels of investigation, from molecular scale to organisms and tissues and further to individuals and populations. To achieve this goal, multidisciplinary collaboration between the fundamental aspects of translational informatics (e.g., bioinformatics, imaging informatics, clinical informatics, and public health informatics) has become essential.

However, the large scale and high dimensionality of data have posed obvious challenges in data mining, storage, and integration. Traditionally, basic research, clinical research, and public health are seen as different worlds based on distinct or incompatible principles. Data transfer, access control, and model building rank are among the most pressing challenges.

4. Cloud Computing to the Rescue

Recent studies and commentaries [2–6] have proposed cloud computing as a solution that addresses many of the limitations mentioned above. Cloud computing is a relatively recent invention. It refers to a flexible and scalable internet infrastructure where processing and storage capacity are

dynamically provisioned. The basic idea of cloud computing is to divide a large task into subtasks, which can then be executed on a number of parallel processors. A key technology with the cloud is the virtual machine (VM) that can be prepackaged with all software needed for a particular analysis.

Large utility-computing services have been emerging in the commercial sector, for example, the Amazon Elastic Compute Cloud (EC2) (<http://aws.amazon.com/ec2/>) [7], and noncommercial public cloud computing platforms also exist to support research, such as the IBM/Google Cloud Computing University Initiative [8] launched by Google and IBM.

Cloud computing infrastructures offer a new way of working. It features a special parallel programming model (e.g., MapReduce [9] designed by Google) to efficiently scale computation to many thousands of commodity machines. These commodity machines form a cluster that can be rented over the internet. Applications in the cloud have also benefited from hadoop (<http://hadoop.apache.org/>) [10], an open-source implementation of MapReduce. Since it is easy to fine tune and highly portable, Hadoop, together with MapReduce, has been widely used for large-scale distributed data analysis in both academy and industry.

Cloud computing infrastructures offer a highly flexible and economical means of working. Cloud computing provides scalable, flexible access to larger computer processing power and storage and avoids the fixed cost of capital investments in local computing infrastructures, computing maintenance, and personnel. The end users are essentially renting capacity on their demand [11].

Cloud computing allows the sharing of data in real-time collaboration with other users. It addresses one of the challenges related to transferring and sharing data. Researchers

TABLE 2: Application of cloud computing in bioinformatics research.

Software	Website	Description	Reference
ArrayExpressHTS	http://www.ebi.ac.uk/Tools/rwiki/	RNA-Seq data processing and quality assessment	[76]
Bioscope	http://www.lifescopelcloud.com/	Reference-based read mapping	[77]
Cloud-MAQ	http://sourceforge.net/projects/cloud-maq/	Read mapping and assembly	[78]
CloudAligner	http://sourceforge.net/projects/cloudaligner/	Read mapping	[79]
CloudBurst	http://cloudburst-bio.sourceforge.net/	Reference-based read mapping	[6]
CloudCoffee	http://www.tcoffee.org/homepage.html	Multiple sequence alignment	[20]
Contrail	http://contrail-bio.sourceforge.net/	De novo read assembly	[80]
Crossbow	http://bowtie-bio.sourceforge.net/crossbow/	Read mapping and SNP calling	[4]
FX	http://fx.gmi.ac.kr/	RNA-Seq data analysis for gene expression levels and genomic variant calling	[81]
GeneSifter	http://www.geospiza.com/Products/AnalysisEdition.shtml	Customer-oriented NGS data analysis services	[82]
Myrna	http://bowtie-bio.sf.net/myrna/	Differential expression analysis for RNA-Seq data	[83]
PeakRanger	http://www.modencode.org/software/ranger/	Peak caller for ChIP-Seq data	[84]
Roundup	http://rodeo.med.harvard.edu/tools/roundup/	Optimized computation for comparative genomics	[19]
SeqMapReduce	http://www.seqmapreduce.org/	Read mapping	[85]
YunBe	http://tinyurl.com/yunbedownload/	Gene set analysis for biomarker identification	[86]

can store their data in the cloud with high availability. For example, Amazon web services provide free access to many useful data sets, for example, the Ensembl [12] and 1000 Genomes data [13]. In addition, the users can have thousands of on-demand powerful computers ready to run their analysis. To this end, cloud computing has the potential to facilitate large-scale efforts in translational data integration and analysis.

5. Translational Bioinformatics Research in the Cloud

There is considerable enthusiasm in the bioinformatics community to deploy open-source applications in the cloud. Various services provided by cloud-computing vendors are described below.

5.1. Cloud-Based Application in Bioinformatics. Numerous of studies have reported the successful application cloud computing in bioinformatics research. Most of these cloud computing applications deal with high-throughput sequence data analysis. CloudBLAST [14] is the first cloud-based implementation to solve sequence analysis problems. Other projects have since been launched on the cloud. Some initiatives have utilized preconfigured software on cloud systems to support large-scale sequence processing. Some tools are available for sequence alignment, short read mapping, SNP identification, genome annotation, and RNA differential expression analysis, amongst others (Table 2). Efforts in

comparative genomics [15–20] and proteomics [21] have also incorporated the cloud to expedite their data processing.

5.2. Cloud-Based Application in Imaging Informatics. The volumes of high-resolution and dynamic imaging data can be estimated to reach petabytes, which indicates that the image reconstruction and analysis is computationally demanding. Cloud-computing is an obvious potential contributor to this end. Image clouds would enable multinational sharing of imaging data, as well as advanced analysis of imaging data away from its place of origin.

Many studies have shown the utility of MapReduce for solving large-scale medical imaging problems in a cloud computing environment. For example, Meng et al. [22] developed an ultrafast and scalable image reconstruction technique for 4D cone-beam CT using MapReduce in a cloud computing environment. Avila-Garcia et al. [23] proposed a cloud computing-based framework for colorectal cancer imaging analysis and research for clinical use. Silva et al. [24] implemented a set of DICOM routers interconnected through a public cloud infrastructure to support medical image exchange among institutions.

Imaging clouds is also making unprecedentedly large-scale imaging research feasible. For example, Euro-Bioimaging [25], a pan-European research infrastructure project aims to deploy a distributed biological and biomedical imaging infrastructure in Europe in a coordinated and harmonized manner. It is expected to offer platforms for storing, remotely accessing, and post-processing imaging data on a large scale.

5.3. Cloud-Based Application in Clinical Informatics. A major challenge for clinical bioinformatics pertains to the accommodation of the range of heterogeneous data into a single, queryable database for clinical or research purposes. Electronic health record (EHR), an integrated clinical information storage systems, has recently emerged and stimulated increased research interest. EHR is a record in digital format that is capable of organizing clinical data by phenotypic categories. An ideal EHR provides complete personal health and medical summary by integrating personal medical information from different sources. The inclusion of genetic imaging and population-based information in EHR has the potential to provide patients with valuable risk assessment based on their genetic profile and family history and to carve a niche for personalized cancer management.

The potential benefits of cloud computing facilitating EHR sharing and EHR integration have been realized. With cloud computing, EHR service could store data into cloud servers. In this way the resources could be flexibly utilized and the operation cost can be reduced. It is envisioned that through the internet or portable media, cloud computing can reduce electronic health record startup expenses, such as hardware, software, networking, personnel, and licensing fees and therefore will promise an explosion in the storage of personal health information online [26–29].

Many previous studies proposed different cloud-based frameworks in an attempt to improve EHR. Among them, Chen et al. [30] proposed a new patient health record access control scheme under cloud computing environments which allows accurate access to patient health record with security and is suitable for enormous multiusers. Chen et al. [31] proposed an EHR sharing and integration system in healthcare clouds. Doukas et al. [32] presented the implementation of a mobile system that enables electronic healthcare data storage, update and retrieval using cloud computing. Rolim et al. [33] proposed a cloud-based solution to automate processes for patients' vital data collection via a network of sensors connected to legacy medical devices and to deliver the data to a medical center's cloud for storage, processing, and distribution. The system provides users with real-time data accessibility labor work to collect, input, and analyze the information. Rao et al. [34] introduced a pervasive cloud-based healthcare application called Dhatri, which leveraged the power of cloud computing and mobile communications technologies to enable physicians to access real-time patient health information from remote areas.

Besides academic researches described above, multiple commercial vendors are competing on this relatively new market. Many world-class commercial companies have heavily invested in the cloud, offering personal medical records services, such as Microsoft's HealthVault [35], which is currently the largest commercial personal health report platform.

5.4. Cloud-Based Application in Public Health Informatics. Public health informatics heavily relies on the data exchange between public health departments and clinical providers. However, public health's information technology systems lack

the capabilities to accept the types of data proposed for exchange. Data silos across organizations and programs will present a set of challenges. With cloud services, however, public health applications, software systems, and services would be made available to health departments, therefore facilitating the exchange of specified types of data between different organizations. In addition, through remote hosting and shared computing resources, public health departments could overcome the problem of funding constraints and insufficient infrastructure for public health systems.

6. Concerns and Challenges for the Biomedical Cloud

Cloud computing offers new possibilities for biomedical research, as data can now be easily accessed and shared. Despite the potential gains achieved, there are also several important issues to be addressed before the cloud computing can become more popular. The most significant concerns pertain to information security and data transfer bottlenecks.

6.1. Information Security and Privacy. Lately, many healthcare organizations are looking to move data and applications to a cloud environment. While this offers flexibility and easy access to computational resources, it also introduces security and privacy concerns, which are particularly evident in fields such as clinical informatics and public health informatics. Highly specialized data, such as clinical data from human studies, have exceptional security needs. Hosting such data on publicly accessible servers may increase the risk of security breaches. There are additional privacy concerns relating to personal information. Therefore these data must be posted according to privacy and security rules, such as the Health Insurance Portability and Accountability Act (HIPAA) [36]. For a biomedical cloud to be viable, a secure protection scheme will be necessary to protect the sensitive information of the medical record. For example, sensitive data will have to be encrypted before entering the cloud. Also, only authorized users are allowed to place and acquire sensitive security metadata in the cloud. More advanced encryption measures as well as access control schemes need to be deployed under cloud computing environments.

So far, some research efforts have been made to build security and privacy architectures for biomedical cloud computing [37, 38]. Main cloud service providers (e.g., Amazon, Microsoft, and Google) have also made commitments to develop best practices to protect data security and privacy.

6.2. Data Transfer Bottlenecks. Another major obstacle to moving to the cloud is the time and cost of data transfer. Biomedical research institutions may need to frequently export or import large volumes of data (on the order of terabytes and soon to be petabytes) to and from the cloud. Given the size of the data set, one may find that there is a data transfer bottleneck. Networking bandwidth limitation causes delays in data transfer and incurs high bandwidth costs from service providers. Bandwidth costs might be low for smaller internet-based applications that are not

data intensive. However, as applications continue to become more data intensive, these costs can quickly add up, making data transfer costs an important issue. For applications that require substantial data movement on a regular basis, cloud computing currently does not make economic sense.

7. Future Developments and Applications

As discussed above, the future of translational medical bioinformatics will depend on integration of diverse data types of patient characteristics. It is therefore crucial to develop an open, data-sharing environment. We suggest that future initiatives should include (1) development of standards to facilitate informational exchange, (2) integration of databases to allow cross-referencing of multilevel data.

7.1. The Need for Standardized Data Formats. Data exchange across the subfields of translational bioinformatics is often difficult because the data come from heterogeneous informatics platforms and are stored in different formats (e.g., numerical values, free text, and graphical and imaging material). The high dimensionality of potential data types mandates standards to represent data in a uniform manner. To work toward this goal, integrated medical/biological terminologies and ontologies have to be adopted, together with advanced semantic-based models and natural language processing (NLP) techniques to objectively describe medical and biomolecular findings.

Numerous attempts have been made in developing standards for data integration in specialized domains. For example, minimum information about microarray experiment (MIAME) [39] is a standard developed to represent and exchange microarray data. In the field of imaging informatics, existing standards include the foundational model of anatomy clinical community, and digital imaging and communications in medicine (DICOM) [40]. Health level 7 (HL7), clinical data standards interchange consortium (CDISC), systematized nomenclature of medicine (SNOMED) and the international statistical classification of diseases and related health problems (ICD-10) represent the standard for clinical community.

These community-specific standards alone, however, are not sufficient to enable intercommunity data sharing. In this regard, the development of integrated standards will be essential. While it is unlikely to develop a single standard to cover all domains, semantic mapping between terminologies seems more practical. Several pioneering medical informatics projects are underway to define such intercommunity standard. For example, the ACGT project [41] launched by the European Union developed a set of methodological approaches as well as tools and services for semantic integration of distributed multilevel databases.

7.2. The Need for Unified Databases. Currently different layers of biomedical data are stored within databases that are highly distributed, and often not interoperable. Even the databases that hold large data sets are often specialized and fragmented, obstructing the path to information sharing. We

need database integration to allow cross-referencing of multilevel data for research or clinical purposes. Opportunities to develop integrated storage systems are increasing as a result of participatory initiatives. Funded by the US National Institutes of Health (NIH), many local platforms in the biomedical informatics space have been established to support data sharing, including informatics for integrating biology and the bedside (i2b2) [42], cancer biomedical informatics grid (caBIG) [43], and biomedical informatics research network (BIRN) [44].

NIH-funded i2b2 Center developed an open-source scalable informatics framework that integrates clinical research data from medical record and genomic data from basic science research. This platform helps better understand the genetic bases of complex diseases. To date, i2b2 has been deployed by over 70 sites internationally. caBIG aims to provide open source standards for data exchange and interoperability in cancer research. At the heart of the caBIG approach is a grid middleware infrastructure called caGrid. caGrid is a service-oriented platform that provides the tools for organizations to integrate data silos, securely share data, and compose analysis pipelines. caBIG enjoys widespread adoption throughout the cancer community. BIRN is an initiative funded by NIH to provide infrastructure, software tools, strategies, and advisory services for sharing biomedical research across disparate groups. These efforts contributed to the transfer and integration of distributed, heterogeneous and multilevel data across the major realms of translational bioinformatics.

8. Conclusion

Biomedical cloud, given the proper architecture, could integrate all the petabytes of available biomedical informatics data in one place and process them on a continuous basis. In this way, we would continuously observe the connections between genotypic profiles and phenotypic data. We can envision that the cloud-supported translational bioinformatics endeavors will promote faster breakthroughs in the diagnosis, prognosis, and treatment of human disease.

Conflict of Interests

The authors declare that there is no conflict of interests.

Authors' Contribution

J. Chen and F. Qian contributed equally to this work.

Acknowledgments

The authors gratefully acknowledge financial support from the National Natural Science Foundation of China Grants (91230117, 31170795, and 91029703), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20113201110015), International S&T Cooperation Program of Suzhou (SH201120), and the National High Technology

Research and Development Program of China (863 program, Grant no. 2012AA02A601).

References

- [1] K. A. Kuhn, A. Knoll, H. W. Mewes et al., "Informatics and medicine—from molecules to populations," *Methods of Information in Medicine*, vol. 47, no. 4, pp. 283–295, 2008.
- [2] L. D. Stein, "The case for cloud computing in genome informatics," *Genome Biology*, vol. 11, no. 5, article 207, 2010.
- [3] M. Baker, "Next-generation sequencing: adjusting to data overload," *Nature Methods*, vol. 7, no. 7, pp. 495–499, 2010.
- [4] B. Langmead, M. C. Schatz, J. Lin, M. Pop, and S. L. Salzberg, "Searching for SNPs with cloud computing," *Genome Biology*, vol. 10, no. 11, article R134, 2009.
- [5] M. C. Schatz, B. Langmead, and S. L. Salzberg, "Cloud computing and the DNA data race," *Nature Biotechnology*, vol. 28, no. 7, pp. 691–693, 2010.
- [6] M. C. Schatz, "CloudBurst: highly sensitive read mapping with MapReduce," *Bioinformatics*, vol. 25, no. 11, pp. 1363–1369, 2009.
- [7] *Amazon Elastic Compute Cloud (Amazon EC2)*, <http://aws.amazon.com/ec2/>.
- [8] *IBM/Google Cloud Computing University Initiative*, <http://www.ibm.com/ibm/ideasfromibm/us/google/index.shtml>.
- [9] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [10] *Hadoop*, <http://hadoop.apache.org/>.
- [11] A. Rosenthal, P. Mork, M. H. Li, J. Stanford, D. Koester, and P. Reynolds, "Cloud computing: a new business paradigm for biomedical information sharing," *Journal of Biomedical Informatics*, vol. 43, no. 2, pp. 342–353, 2010.
- [12] P. Flicek, M. R. Amode, D. Barrell et al., "Ensembl 2012," *Nucleic Acids Research*, vol. 40, Database issue, pp. D84–D90, 2012.
- [13] *1000 Genomes Project*, <http://www.1000genomes.org>.
- [14] A. Matsunaga, M. Tsugawa, and J. Fortes, "CloudBLAST: combining MapReduce and virtualization on distributed resources for bioinformatics applications," in *Proceedings of the 4th IEEE International Conference on eScience (eScience '08)*, pp. 222–229, December 2008.
- [15] G. Sudha Sadasivam and G. Baktavatchalam, "A novel approach to multiple sequence alignment using hadoop data grids," *International Journal of Bioinformatics Research and Applications*, vol. 6, no. 5, pp. 472–483, 2010.
- [16] D. P. Wall, P. Kudtarkar, V. A. Fusaro, R. Pivovarov, P. Patil, and P. J. Tonellato, "Cloud computing for comparative genomics," *BMC Bioinformatics*, vol. 11, article 259, 2010.
- [17] I. Kim, J. Y. Jung, T. F. Deluca, T. H. Nelson, and D. P. Wall, "Cloud computing for comparative genomics with windows azure platform," *Evolutionary Bioinformatics Online*, vol. 8, pp. 527–534, 2012.
- [18] G. Zhao, D. Bu, C. Liu et al., "CloudLCA: finding the lowest common ancestor in metagenome analysis using cloud computing," *Protein Cell*, vol. 3, no. 2, pp. 148–152, 2012.
- [19] P. Kudtarkar, T. F. Deluca, V. A. Fusaro, P. J. Tonellato, and D. P. Wall, "Cost-effective cloud computing: a case study using the comparative genomics tool, roundup," *Evolutionary Bioinformatics Online*, vol. 6, pp. 197–203, 2011.
- [20] P. di Tommaso, M. Orobitch, F. Guirado, F. Cores, T. Espinosa, and C. Notredame, "Cloud-Coffee: implementation of a parallel consistency-based multiple alignment algorithm in the T-coffee package and its benchmarking on the Amazon Elastic-Cloud," *Bioinformatics*, vol. 26, no. 15, pp. 1903–1904, 2010.
- [21] B. D. Halligan, J. F. Geiger, A. K. Vallejos, A. S. Greene, and S. N. Twigger, "Low cost, scalable proteomics data analysis using Amazon's cloud computing services and open source search algorithms," *Journal of Proteome Research*, vol. 8, no. 6, pp. 3148–3153, 2009.
- [22] B. Meng, G. Pratz, and L. Xing, "Ultrafast and scalable cone-beam CT reconstruction using MapReduce in a cloud computing environment," *Medical Physics*, vol. 38, no. 12, pp. 6603–6609, 2011.
- [23] M. S. Avila-Garcia, A. E. Trefethen, M. Brady, F. Gleeson, and D. Goodman, "Lowering the barriers to cancer imaging," in *Proceedings of the 4th IEEE International Conference on eScience (eScience '08)*, pp. 63–70, Indianapolis, Ind, USA, December 2008.
- [24] L.A. Silva, C. Costa, and J. L. Oliveira, "DICOM relay over the cloud," *International Journal of Computer Assisted Radiology and Surgery*, 2012.
- [25] *Euro-Bioimaging*, <http://www.eurobioimaging.eu/>.
- [26] E. J. Schweitzer, "Reconciliation of the cloud computing model with US federal electronic health record regulations," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 161–165, 2011.
- [27] G. Fernandez-Cardenas, I. de la Torre-Diez, M. Lopez-Coronado, and J. J. Rodrigues, "Analysis of cloud-based solutions on EHRs systems in different scenarios," *Journal of Medical Systems*, vol. 36, no. 6, pp. 3777–3782, 2012.
- [28] J. Haughton, "Look up: the right EHR may be in the cloud. Major advantages include interoperability and flexibility," *Health Management Technology*, vol. 32, no. 2, p. 52, 2011.
- [29] J. Kabachinski, "What's the forecast for cloud computing in healthcare?" *Biomedical Instrumentation and Technology*, vol. 45, no. 2, pp. 146–150, 2011.
- [30] T. S. Chen, C. H. Liu, T. L. Chen, C. S. Chen, J. G. Bau, and T. C. Lin, "Secure Dynamic access control scheme of PHR in cloud computing," *Journal of Medical Systems*, vol. 36, no. 6, pp. 4005–4020, 2012.
- [31] Y. Y. Chen, J. C. Lu, and J. K. Jan, "A secure EHR system based on hybrid clouds," *Journal of Medical Systems*, vol. 36, no. 5, pp. 3375–3384, 2012.
- [32] C. Doukas, T. Pliakas, and I. Maglogiannis, "Mobile healthcare information management utilizing Cloud Computing and Android OS," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2010, pp. 1037–1040, 2010.
- [33] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, and G. S. Salvador, "A cloud computing solution for patient's data collection in health care institutions," in *Proceedings of the 2nd International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED '10)*, pp. 95–99, New York, NY, USA, February 2010.
- [34] G. S. V. R. K. Rao, K. Sundararaman, and J. Parthasarathi, "Dhatri—a pervasive cloud initiative for primary healthcare services," in *Proceedings of the 14th International Conference on Intelligence in Next Generation Networks (ICIN '10)*, Berlin, Germany, October 2010.
- [35] *Microsoft Health Vault*, <http://www.microsoft.com/en-us/health-vault/>.

- [36] E. Hansen, "HIPAA (Health Insurance Portability and Accountability Act) rules: federal and state enforcement," *Medical Interface*, vol. 10, no. 8, pp. 96–102, 1997.
- [37] M. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," in *Advances in Cryptology-EUROCRYPT 2004*, pp. 1–19, 2004.
- [38] V. Danilatu and S. Ioannidis, "Security and privacy architectures for biomedical cloud computing," in *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB '10)*, November 2010.
- [39] A. Brazma, P. Hingamp, J. Quackenbush et al., "Minimum information about a microarray experiment (MIAME)—toward standards for microarray data," *Nature Genetics*, vol. 29, no. 4, pp. 365–371, 2001.
- [40] H. Blume, "DICOM (Digital Imaging and Communications in Medicine) state of the nation. Are you afraid of data compression?" *Administrative Radiology Journal*, vol. 15, no. 11, pp. 36–40, 1996.
- [41] M. Tsiknakis, D. Kafetzopoulos, G. Potamias, A. Analyti, K. Marias, and A. Manganas, "Building a European biomedical grid on cancer: the ACGT Integrated Project," *Studies in Health Technology and Informatics*, vol. 120, pp. 247–258, 2006.
- [42] I. S. Kohane, S. E. Churchill, and S. N. Murphy, "A translational engine at the national scale: informatics for integrating biology and the bedside," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 181–185, 2011.
- [43] K. K. Kakazu, L. W. Cheung, and W. Lynne, "The Cancer Biomedical Informatics Grid (caBIG): pioneering an expansive network of information and tools for collaborative cancer research," *Hawaii Medical Journal*, vol. 63, no. 9, pp. 273–275, 2004.
- [44] J. R. MacFall, W. D. Taylor, D. E. Rex et al., "Lobar distribution of lesion volumes in late-life depression: the Biomedical Informatics Research Network (BIRN)," *Neuropsychopharmacology*, vol. 31, no. 7, pp. 1500–1507, 2006.
- [45] P. Baldi and A. D. Long, "A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes," *Bioinformatics*, vol. 17, no. 6, pp. 509–519, 2001.
- [46] W. Pan, "A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments," *Bioinformatics*, vol. 18, no. 4, pp. 546–554, 2002.
- [47] P. Tamayo, D. Slonim, J. Mesirov et al., "Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2907–2912, 1999.
- [48] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.
- [49] S. Götz, J. M. García-Gómez, J. Terol et al., "High-throughput functional annotation and data mining with the Blast2GO suite," *Nucleic Acids Research*, vol. 36, no. 10, pp. 3420–3435, 2008.
- [50] K. G. Becker, D. A. Hosack, G. Dennis et al., "PubMatrix: a tool for multiplex literature mining," *BMC Bioinformatics*, vol. 4, article 61, 2003.
- [51] J. Quackenbush, "Computational analysis of microarray data," *Nature Reviews Genetics*, vol. 2, no. 6, pp. 418–427, 2001.
- [52] F. M. Selaru, Y. Xu, J. Yin et al., "Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions," *Gastroenterology*, vol. 122, no. 3, pp. 606–613, 2002.
- [53] A. J. Saldanha, "Java Treeview—extensible visualization of microarray data," *Bioinformatics*, vol. 20, no. 17, pp. 3246–3248, 2004.
- [54] K. Rutherford, J. Parkhill, J. Crook et al., "Artemis: sequence visualization and annotation," *Bioinformatics*, vol. 16, no. 10, pp. 944–945, 2000.
- [55] A. A. Porollo, R. Adamczak, and J. Meller, "POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins," *Bioinformatics*, vol. 20, no. 15, pp. 2460–2462, 2004.
- [56] W. Fleischmann, S. Möller, A. Gateau, and R. Apweiler, "A novel method for automatic functional annotation of proteins," *Bioinformatics*, vol. 15, no. 3, pp. 228–233, 1999.
- [57] L. B. Koski, M. W. Gray, B. F. Lang, and G. Burger, "AutoFACT: an automatic functional annotation and classification tool," *BMC Bioinformatics*, vol. 6, article 151, 2005.
- [58] F. Meyer, D. Paarmann, M. D'Souza et al., "The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics*, vol. 9, article 386, 2008.
- [59] J. R. Smith and S. F. Chang, "VisualSEEK: a fully automated content-based image query system," in *Proceedings of the 4th ACM international conference on Multimedia*, pp. 87–98, November 1996.
- [60] D. M. Squire, W. Müller, H. Müller, and T. Pun, "Content-based query of image databases: inspirations from text retrieval," *Pattern Recognition Letters*, vol. 21, no. 13-14, pp. 1193–1198, 2000.
- [61] L. Ohno-Machado, V. Bafna, A. A. Boxwala et al., "iDASH: integrating data for analysis, anonymization, and sharing," *Journal of the American Medical Informatics Association*, vol. 19, no. 2, pp. 196–201, 2012.
- [62] C. Friedman, "A broad-coverage natural language processing system," *Proceedings/AMIA. Annual Symposium. AMIA Symposium*, pp. 270–274, 2000.
- [63] J. M. Beck, W. J. Ma, R. Kiani et al., "Probabilistic population codes for Bayesian decision making," *Neuron*, vol. 60, no. 6, pp. 1142–1152, 2008.
- [64] E. S. O'Neill, N. M. Dluhy, and E. Chin, "Modelling novice clinical reasoning for a computerized decision support system," *Journal of Advanced Nursing*, vol. 49, no. 1, pp. 68–77, 2005.
- [65] H. Lindgren, "Decision support system supporting clinical reasoning process—an evaluation study in dementia care," *Studies in Health Technology and Informatics*, vol. 136, pp. 315–320, 2008.
- [66] J. Diaz Guzman and A. Gomez de la Camara, "The diagnostic process in Neurology: from clinical reasoning to assessment of diagnostic tests," *Neurologia*, vol. 18, supplement 2, pp. 1–2, 2003.
- [67] A. Subramanian, B. Westra, S. Matney et al., "Integrating the nursing management minimum data set into the logical observation identifier names and codes system," *Proceedings/AMIA. Annual Symposium. AMIA Symposium*, p. 1148, 2008.
- [68] A. W. Forrey, C. J. McDonald, G. Demoor et al., "Logical Observation Identifier Names and Codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results," *Clinical Chemistry*, vol. 42, no. 1, pp. 81–90, 1996.
- [69] C. Friedman, "Semantic text parsing for patient records," *Medical Informatics*, vol. 8, pp. 423–448, 2005.

- [70] G. Eysenbach and A. R. Jadad, "Evidence-based patient choice and consumer health informatics in the Internet age," *Journal of Medical Internet Research*, vol. 3, no. 2, p. E19, 2001.
- [71] K. D. Mandl, P. Szolovits, and I. S. Kohane, "Public standards and patients' control: how to keep electronic medical records accessible but private," *British Medical Journal*, vol. 322, no. 7281, pp. 283–287, 2001.
- [72] S. A. Buckovich, H. E. Rippen, and M. J. Rozen, "Driving toward guiding principles: a goal for privacy, confidentiality, and security of health information," *Journal of the American Medical Informatics Association*, vol. 6, no. 2, pp. 122–133, 1999.
- [73] D. M. Lopez and B. Blobel, "Enhanced semantic interoperability by profiling health informatics standards," *Methods of Information in Medicine*, vol. 48, no. 2, pp. 170–177, 2009.
- [74] S. Baxter, A. Killoran, M. P. Kelly, and E. Goyder, "Synthesizing diverse evidence: the use of primary qualitative data analysis methods and logic models in public health reviews," *Public Health*, vol. 124, no. 2, pp. 99–106, 2010.
- [75] G. Q. Hu, K. Q. Rao, and Z. Q. Sun, "Identification of a detailed function list for public health emergency management using three qualitative methods," *Chinese Medical Journal*, vol. 120, no. 21, pp. 1908–1913, 2007.
- [76] A. Goncalves, A. Tikhonov, A. Brazma, and M. Kapushesky, "A pipeline for RNA-seq data processing and quality assessment," *Bioinformatics*, vol. 27, no. 6, pp. 867–869, 2011.
- [77] M. H. Rahimi, *Bioscope: Actuated Sensor Network for Biological Science*, University of Southern California, Los Angeles, Calif, USA, 2005.
- [78] A. K. Talukder, S. Gandham, H. A. Prahald, and N. P. Bhattacharyya, "Cloud-MAQ: the cloud-enabled scalable whole genome reference Assembly application," in *Proceedings of the 7th IEEE and IFIP International Conference on Wireless and Optical Communications Networks (WOCN '10)*, Colombo, Sri Lanka, September 2010.
- [79] T. Nguyen, W. Shi, and D. Ruden, "CloudAligner: a fast and full-featured MapReduce based tool for sequence mapping," *BMC Research Notes*, vol. 4, article 171, 2011.
- [80] M. C. Schatz, A. L. Delcher, and S. L. Salzberg, "Assembly of large genomes using second-generation sequencing," *Genome Research*, vol. 20, no. 9, pp. 1165–1173, 2010.
- [81] D. Hong, A. Rhie, S. S. Park et al., "FX: an RNA-Seq analysis tool on the cloud," *Bioinformatics*, vol. 28, no. 5, pp. 721–723, 2012.
- [82] C. Sansom, "Up in a cloud?" *Nature Biotechnology*, vol. 28, no. 1, pp. 13–15, 2010.
- [83] B. Langmead, K. D. Hansen, and J. T. Leek, "Cloud-scale RNA-sequencing differential expression analysis with Myrna," *Genome Biology*, vol. 11, no. 8, article R83, 2010.
- [84] X. Feng, R. Grossman, and L. Stein, "PeakRanger: a cloud-enabled peak caller for ChIP-seq data," *BMC Bioinformatics*, vol. 12, article 139, 2011.
- [85] Y. Li and S. Zhong, "SeqMapReduce: software and web service for accelerating sequence mapping," *Critical Assessment of Massive Data Analysis (CAMDA)*, vol. 2009, 2009.
- [86] L. Zhang, S. Gu, Y. Liu, B. Wang, and F. Azuaje, "Gene set analysis in the cloud," *Bioinformatics*, vol. 28, no. 2, pp. 294–295, 2012.