IOSR Journal of Biotechnology and Biochemistry (IOSR-JBB)
ISSN: 2455-264X, Volume 1, Issue 5 (Jul. – Aug. 2015), PP 22-26
www.iosrjournals.org

Bioinformatics in the Clinical Pipeline: Contribution in Genomic Medicine

Amiy Dutt Chaturvedi and P.K Patra
Second phase of Task force Biomedical informatics center" under Bioinformatics center of ICMR
Department of Biochemistry
Pt.J.N.M Medical College Raipur Chhattisgarh India .492001

Abstract: *In this review report we like to focus on the new challenges in methodology of modern biology be used in medical science. Today human health is a primary issue to cure disease, undoubtedly the answer to this is bioinformatics or (In-silco) tools has change the concept of treating patients to understand the need of genomic medicine in use. Those with new modes of action in clinical treatment, is a major health concern in medical science. On global prospective scientific role in constructing new ideas to remediate health care to treat disease exciting in nature is challenging task. So awareness needs to accelerate store clinical datasets for scientific represents to design genomic drugs. This new outline will drive the medical to discover public data and create a cognitive approach to use technology cheaper at cost effective mode.*
Key word: *Bioinformatics (In-silco), clinical datasets, genomic medicine, cognitive*

I. Introduction

Over the past years, molecular focus to analysis the medical driven datasets by genomic, proteomic, DNA Sequencing and gene expression using microarrays has been used to generate in lager amount [1,2,12]. Today, biomedical research over the past decade has been in early stage for development of (In-silco) biology or bioinformatics for publication. So we can identify more disease at molecular level (DNA sequence) and can image biological data in a month than was previously able to asses in years. It is just like a few years, we have gone to identify one complete human genome sequence to dozens, with thousands of immediate horizon. Collection of genomic dataset for biological analysis are resourceable well online today such as National Centers for biomedical Computing (NCBC), Cancer Bioinformatics Grids (caBIC) to diagnosis the disease is expanded by leaps and bounds by high-throughput genome sequencing or large-scale image acquisition. New scientific biomedical informatics approaches are needed to store, mine and analyze this unprecedented.

Evolutionary medicine and informatics research encompasses broad aspects of drug development. (In-silco) computational biology or bioinformatics research in medical science is thought to have the inherent approach to speed up the rate of translation or medicine discovery while reducing the need for expensive R&D work and clinical trials including the development of theoretical models, empirical analysis of large scale datasets method., biomedical software development, and establishment of unique databases that accelerate biological and biomedical discovery [13]. Computational, mathematical, and statistical techniques are used to compile the larger dataset throw bioinformatics to address important questions in basic biomedical research. Modern research ranges from theoretical structure modeling and simulation to develop tools with data scaling to analysis a single critical gene within a population to its full genome sequence comparisons across species as well as also to analysis of individual images capturing functional and expression domains to thousands of images describing genome-wide expression patterns[14].

Walter Gilbert, a renowned scientist, described this shift in biology as follows:

"The new paradigm, now emerging, is that all of the "genes" will be known (in the sense of being resident in databases available electronically), and that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a theoretical conjecture only then turning to experiment to follow or test that hypothesis."

Bioinformatics plays dynamic role for the integration of broad disciplines are biology to understand the complex mechanisms of the cell system in body. Bioinformatics also aids the way in which biomedical investigators use the database information in their testing. The complete process of data collection to analysis the results tests may be categorized under a separate area named "Clinical Informatics". The major advantage with the clinical informatics is the concept of storage of data by Electronic Medical Records Data (EMRD) information system can be easily accessed and shared in comparison to traditional medical records data. EMRD also drastically reduces the possibilities errors due to frustration and other psychological disturbances during the manual data entry record process after collecting the necessary information of patients on paper. It also helps to eliminate the manual task of extracting data from charts or filling out specialized data sheets. The National Center for Biotechnological Information (NCBI) Gene expression Omnibus (GEO) has complied 3848 in

1,40,1476 samples from experiment [8]. Data required for a R&D study can be obtained directly from the stored electronic record, thus making research data collection for analysis, may be used as byproduct to routine clinical record. The clinical record environment can help to assure compliance with a research protocol, pointing out to a clinician when a patient is ethically eligible for a study, or when the protocol for a study calls for a specific health care management plan given the currently available data about that patient. In the near future one can see a situation where the complete genomic sequence information of disease and its mutational change at particular gene on the patient can be accessed from the EMRD. This information can be of any type, ranging from rational drug trial data of Clinical R&D data's of various tests performed on that patient and the outcome of such experiments. Thus challenge in such cases will be to organize and integrate the heterogeneity of the information into a comprehensive, knowledge based database from which an individual can access the necessary portion of the record for any research analysis for social benefit.

Medical science and Bioinformatics

In the late 1950, medical data was implemented clinically and recorded [1,2] have a rapid development in data storage at computational level [4,5]. Bioinformatics has a profound impact in Clinicians (Medical Sciences) research. There are several scientific journal publish biological databases (BD) are helping clinician and clinical researchers to diagnose the disease and develop compact strategies for its therapy [7,8]. Consider a situation where a patient with a genetic form of Hemolytic anemia (Sickle Cell anemia) meets clinicians. The clinicians are not sure with the symptoms of the disease but have the only clue that the patient's family may be suffering from hemolytic anemia. The sickle disease syndromes are caused by a mutation in the β-globin gene that changes the sixth amino acid from glutamic acid to valine HbS ($\alpha_2\beta_2^{6\ Glu\rightarrow Va1}$) polymerizes reversibly when deoxygenated to form a gelatinous network of fibrous polymers that stiffen the RBC membrane, increase viscosity, and cause dehydration due to potassium leakage and calcium influx. These changes also produce the sickle shape surf the web to get the information on the disease by checking out the OMIM (Online Mendelian Inheritance of Man) resources available at http://www.ncbi.nlm.nih.gov/omim/ which provides detailed information on genetic disorders. A focused search for sickle cell anemia would reveal multiple disorders including microvascular vasoocclusion and premature RBC destruction (hemolytic anemia) also provides the information. The mutated sickle gene database literature linked to DNA sequence to link with protein sequence is in references of MEDLINE [4]. Following this MEDLINE literature database, the original research article on different complications cause due to sickle (where the association hemoglobin is discussed) is obtained. Protein sequence detailed information is obtained from the SWISS-PROT database and Protein Information Resource (PIR). The information on the crystal structure can be obtained by following the link to Protein Data Bank (PDB) provided in the SWISSPROT database. Genetic disorder database, GENBANK, the nucleotide sequence of the gene is obtained along with records of gene irregularities. Thus the clinician uses a number of databases to collect recent advance information about the disease which aids him to diagnose and device strategies for future therapy.

Drug Designing and Bioinformatics

In-silico research in Pharmaceutical Industries (drug development) is thought to have the effective approach for trial and error process to speed up the rate of drug designing at molecular level. Modern drug discovery for infectious diseases have a comprehensive strategy for reducing world's biggest killers of children and young adults. So far first genetic disease is yet to have a stepping stone to cure disease [9,10]. According to the WHO reports (Developing and Developed) countries should need to have expensive R&D work and clinical trials to develop cheaper medicine efficient for social benefit to mankind [11]. So drug development is the major problem for recent immune diseases in nature, one way to achieve this is by producing and screening rational drug using modern informatics science, answers for this problem is Bioinformatics by high-throughput screening (HTS) system. Medical data are very much lexicon reliable on scientific drug design (Computational and synthetically) for decision making to convert better resources for suitable drug development. Hospital Data collection manually may have error but now a day's collections of data throw computationally are more reliable and effective plays a new oil in the drug development. This communication technique will induce better understanding in social media. Universal data in scientific community can relate or unrelated in all research disciplines [3]. The role of clinician and pharmacist is never been static to evolve computational datasets. Furthermore, structure base rational drug design process could reduce the time and cost for developing suitable pharmacological molecular agents. Computational biology (In-silico) methods are used to predict 'drug-likeness' which is nothing but the identification and elimination of candidate molecules to Drug-likeness could be predicted by statistically by (genetic algorithm and neural network) based approaches. Just like FDA has recommended only one drug for sickle disease "Hydroxyurea" is used to control the crises happening in the blood against abnormally sickle shaped cell. Hydroxyurea belong to the drug group antimetabolites[16]. That does not have a significant cure for sickle disease and several oncological problems. So doctors recommend

another drug (folic acid; Vitamin) to avoid this side effect. In recent year pharmaceutical industries are working on to identify a suitable drug which could reduce the risk factor cause due to Hydroxyurea. In this reference In-silco biology will be able to answer all the possibilities by using the mutated protein of infected cell by statistical using (genetic algorithm) to allocated design protein to replace the infected protein by reducing the energy level of infection. Technical barriers are there to show and realize the changes happing in designing the selected protein molecules to form drug. Bimolecular interaction (Ligand and Receptor) through computational approach is a tuff task requires a high screening of 0.5 Gflop and 1 Tflop, which is quite expensive in the present day scenario. Realizing the rational problem for storage EMRD recently IBM has announced 500 times more powerful processor world's fastest existing computer and 2 million times faster than the today's fastest desktop PC. This new computer nicknamed "Blue Gene" by IBM researchers will be capable of performing close to one Petaflop (1015 operations per second) [17,18].

Genomic medicine and Bioinformatics

In the Biomedical research data analysis is important to get insight future planning to diagnosis disease. In the scientific analysis statistic data can be relevant to support bioinformatics tools to design genomic medicine. Genomic medicines in the past years have been in the post genomic era. The genomic research momentum across all clinical consortium helps to treat patients at primary level in recent years. But scenario has changed the symptom of diagnosis disease medical record data must be a basic tool in treatment. In this investigation some database list are mentioned helps clinicians to find the new approaches implemented in health care [Table 3].

Next-generation sequencing for clinical diagnostics

In past years sanger DNA sequencing approach are used to diagnosis the gene expression of disease. As the pace advancement in the technology growing day by day has changed the concept of biomedical research. Although in recent decade performance of high-throughput sequencing (NSG) using bioinformatics tools have over lapping the sanger DNA sequencing to do clinical research are needed to expertise in handling the medical record data (MRD) in their treatment level. So as the development of treatment system high-throughput sequencing DNA sequence termed Next generation Sequencing (NSG) is novel approach for genomic analysis at a cost feasible for routine clinical diagnostics in present scenario. In the clinical pipeline contribution to design genomic medicine will accomplish the bioinformatics tools implemented in medical science. Clinical research may investigate the disease mechanism to transcript DNA and RNA sequencing. According to the complex-city of disease treatment throw bioinformatics tools may reduce the cost effective approach of laboratory bench work throw high-throughput sequencing technology at therapeutic level [19]. Next-generation sequencing (NSG) datasets are used to minimize the workflow of laboratories and ethical consideration for human trails. Some latest next-generation sequencing tools have been design by industries in past year are summarized are been used clinical diagnosis [Table 2].

Ethical issues

Biomedical research has been inexorable and intertwined the challenges of ethical task. However, genomic diagnosis treatments by datasets are new task design for clinicians to treat patients in health care. Now few questions that stimulate hindrance in mind that modern technology will be easy to implicate the clinical data without any cognitive risk. So accuracy of datasets in the medical science is important to judge socially to predict the disease at public level. These few ethical points are now a primary level to understand the modern technology to cure diseases [25].

II. Conclusion

Technology and health care are two most collaborated social ethical issues to be discussed in new era of human treatment and diagnosis. As the emergence of genomic medicine (personalized medicine) are closely related to fulfill the need of individual patient treatment. However, cost effective approach of bioinformatics in medical science has been a multidimensional platform to give a meaningful thought at (Inter)-national level. Electronic record data managements system are moderate in future will bring closer to us in diagnosis disease in nature. So biomedical and informatics community need to construct data policy to procure the cognitive risk in public health. Despite of risk, limitations and reduction in treatment cost undoubtedly high-throughput technology will change the thinking of clinical trials. Meanwhile, high-throughput (Bioinformatics) technology has focused on supporting different applications of scientific research to outbreak health care. For this clinicians and informaticians should shake hand in applying informatics tools to diagnosis the public data will discover fundamental cause of therapy. The transfer of technology in both the ways may benefits the clinicians and scientist to develop expertise for solving many unanswered health care problems in environment. As above

discussed study bioinformatics tools in clinical pipeline will make a significant change in recent years to improve human health.

Acknowledgement

The authors express their gratitude to ICMR for sectioning project no. BIC/12/14/2013 "Second phase of Task force Biomedical informatics center" under Dr P.K Patra Prof/Head & P.I, Bioinformatics center of ICMR, Department of Biochemistry, Pt. J.N.M Medical College Raipur (C.G) for their support and kind endeavor.

References

[1]. Barnett GO: The application of computer-based medical record systems in ambulatory practice. N Engl J Med. 1984; 310; 1643–50.
[2]. McDonald CJ, Overhage JM, Tierney WM, et al: The Regenstrief medical record system: a quarter century experience. Int J Med Inform. 1999; 54: 225–53.
[3]. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. New York: Springer-Verlag, 1991.
[4]. Shortiffe EH, Perreault L, Fagan L, Wiederhold G (eds). Medical Informatics: Computer Applications in Health Care (ed 2). New York: Springer Verlag, 2001.
[5]. Weed LL. Medical records that guide and teach: N Engl J Med. 1968; 278:593–9.
[6]. Lindberg DA, Siegel ER, Rapp BA, Wallingford KT, Wilson SR: Use of MEDLINE by physicians for clinical problem solving. JAMA. 1993; 269:3124–9.
[7]. Rosati RA, Wallace GG, Stead EA: The way of the future. Arch Intern Med. 1973; 131:285–8. 27.
[8]. Fries J: Time oriented patient records and a computer data bank. JAMA. 1972 ; 222:1436–542.
[9]. Kohane I. Bioinformatics and clinical informatics: the imperative to collaborate: J Am Med Inform Assoc. 2000; 7:439–43.
[10]. Kulikowski C. The micro-macro spectrum of medical informatics. Challenges: from molecular medicine to transforming health care in a globalizing societ: Methods Inf Med. 2001; 41:20–4.
[11]. Miller P. Opportunities at the intersection of bioinformatics and health informatics J Am Med Inform Assoc: 2000; 7:4318.
[12]. Altman R. The interactions between clinical informatics and bioinformatics. J Am Med Inform Assoc. 2000; 7:439–43.
[13]. Computational methods for the prediction of 'drug-likeness' D.E. Clark and S.D. Pickett. Drug Discovery Today: 2000. 5(2):49-57.
[14]. Bioinformatics in support of molecular medicine. R.B. Altman. Author's site at http://smi-web.stanford.edu/people/altman/
[15]. New advances in Pharmacogenomics. B. Destenaves and F. Thomas. Current Opinion in Chemical Biology: 2000. 4:440-444.
[16]. Proteomics Factories. E. Russo. The Scientist: 2000; 14(3):1.
[17]. Trends in Computational Biology: A Summary based on a RECOMB Plenary lecture. J.C. Wooley. Journal of Computational Biology: 1999; 6(3/4):459-474.
[18]. The Computer Meets Medicine and Biology: Emergence of a Discipline. (Chapter. 1 of the course material). E.H. Shortliffe and M.S. Blois.
[19]. Next Generation DNA Sequencing and the Future of Genomic Medicine, Matthew W. Anderson and Iris Schrijver Genes: 2010; 1, 38-69
[20]. Margulies, M.; Egholm, M.; Altman, W.E.; Attiya, S.; Bader, J.S.; Bemben, L.A.; Berka, J.; Braverman, M.S.; Chen, Y.J.; Chen, Z.; et al: Genome sequencing in microfabricated high-density picolitre reactors. Nature **2005**, 437, 376-380.
[21]. Ronaghi, M.; Karamohamed, S.; Pettersson, B.; Uhlen, M.; Nyren, P. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem: **1996**, 242, 84-89.
[22]. Ronaghi, M.; Uhlen, M.; Nyren: P. A sequencing method based on real-time pyrophosphate. Science 1998, 281, 363, 365.
[23]. Shendure, J.; Porreca, G.J.; Reppas, N.B.; Lin, X.; McCutcheon, J.P.; Rosenbaum, A.M.; Wang, M.D.; Zhang, K.; Mitra, R.D.; Church, G.M: Accurate multiplex polony sequencing of an evolved bacterial genome. Science: **2005**, 309, 1728-1732.
[24]. The Church Laboratory. Polony sequencing protocols http://openwetware.org/wiki/ Church Lab: PoloProt (accessed April 2010).
[25]. Clayton, E.W. Ethical, legal, and social implications of genomic medicine. N. Engl. J. Med.: **2003**; 349, 562-569.

Table1. Data sources collected for clinical R&D used in bioinformatics

Data source	**Data size**	**R&D Bioinformatics**
Raw DNA sequence	8.2 million sequences	Separating coding and non-coding regions
	(9.5 billion bases)	Identification of introns and exons
		Gene product prediction
		Forensic analysis
Protein sequence	300,000 sequences	Sequence comparison algorithms
	(~300 amino acids	Multiple sequence alignments algorithms
	each)	Identification of conserved sequence motifs
Macromolecular	13,000 structures	Secondary, tertiary structure prediction
structure	(~1,000 atomic	3D structural alignment algorithms
	coordinates each)	Protein geometry measurements
		Surface and volume shape calculations
		Intermolecular interactions
		Molecular simulations
		(force-field calculations,
		molecular movements,
		docking predictions)
Genomes DNA/RNA	35, 807complete genomes	Characterisation of repeats
	(1.6 million –	Structural assignments to genes
	3 billion bases each)	Phylogenetic analysis
		Genomic-scale censuses
		(characterisation of protein content, metabolic

		pathways)
		Linkage analysis relating specific genes to diseases
mRNA	1,140 complete genomes	Genome sequence
Gene expression	largest: ~20 time	Correlating expression patterns
	point measurements	Mapping expression data to sequence, structural and
	for ~6,000 genes	biochemical data
Other data		
Literature	20 million citations	Digital libraries for automated bibliographical searches
		Knowledge databases of data from literature
Metabolic pathways		Pathway simulations
Next Genration Sequencing	>99.9% raw base accuracy	Nucleotide for template sequence

Table -2 Next Generation Sequencing (NSG Model) used in Bioinformatics

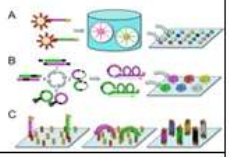
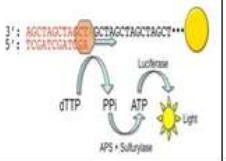
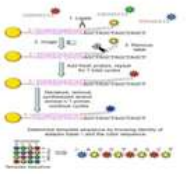
Next Generation Sequencing (NSG Model)	Utility	Application
Roche/454 Life Sciences	DNA library (single stranded DNA or PCR amplicons) containing flanking adaptor sequences which are used to immobilize the DNA library fragments to capture beads.	
4 GS FLX	Pyrosequencing technology to perform the sequencing reaction	
Applied Biosystems/SOLiD	Probe ligation sequencing chemistry	

Table3. Database URLs list

Database	URL
International Nucleotide sequence database collaboration	www.insdc.org
DDBJ (DNA Database of Japan)	www.ddbj.nig.ac.jp
ENA (European Nucleotide Achieves	www.ebi.ac.uk/ena
Gene bank	www.ncbi.nlm.nih.gov/genbank
Nucleotide Database	www.ncbi.nlm.nih.gov/nucleotide
EMBL	www.ebi.ac.uk
NCBI	www.ncbi.nlm.nih.gov
Protein Database	www.rcsb.org
SWISS-PROT	web.expasy.org/docs/swiss-prot
PDBJ (Protein databank of Japan)	www.pdbj.org
PDBTM (Protein databank transmembrane protein)	www.pdbtm.enzim.hu
HGPD (Human Gene and Protein Database)	www.hgpd.lifesciencedb.jp
DisProt (Database of Protein disorder)	www.disprot.org
Genomic Database	
Ensembl	www.asia.ensembl.org/
GOLD (Genome online Database)	www.gold.jgi-psf.org
MBGD (Microbial Genome Database)	www.mbgd.genome.ad.jp/