Emergence of *in-silico* biology in desperate spheres of life sciences in Pakistan

Muhammad Naveed¹*, Mariyam Amin², Shahneela Asif², Ayesha Javed Iqbal², Fazeeka Amjad², Haleema Sadia³ and Mushtaq Ahmed Saleem¹

1. Department of Biotechnology, Faculty of life sciences, University of Central Punjab, Lahore-Pakistan
2. Department Biochemistry and Biotechnology, University of Gujrat, Sialkot sub campus, Sialkot-Pakistan
3. Department of Animal Sciences, Quaid-e-Azam University, Islamabad-Pakistan

*Corresponding author’s email: naveed.quadian@gmail.com ; dr.naveed@ucp.edu.pk

**Citation**

**Abstract**
Bioinformatics, a multidisciplinary field works in collaboration with biological sciences for analysis, prediction and storage of biological data in form of databases. In innovative world of scientific technology, Pakistan is playing its crucial role in development of new era of e-science. Main objective of this review is to highlight bioinformatics applications and its significant role in progress of e-science in Pakistan. In Pakistan, numerous applications such as gene target approaches in agriculture, food industries and 3-dimensional structure analysis of protein have been revealed, where every aspect of nucleic acids processing can be observed. With these advancements, disease diagnosis has now become easy, providing a platform for drug development and enormous jobs for talented ones to make a history in bioinformatics. Similarly, numerous organizations are also taking part in Next Generation Sequencing (NGS) technology and different nationwide companies. However, bioinformatics also faces challenges in data collection and web designing. But introduction of innovative researches has led to formation of highly applicable e-technology of recent era. Hence, this approach has taken primacy to diagnose diseases as cancer. It makes a tremendous distinction in e-science where computer-aided drug design (CADD) gives a ray of hope for many diseased people. Moreover, comprehensive study of human genetics also reveals invention of specified drugs and various biomarkers to discriminate among every single genetic disease. Regardless of various hurdles, scientific transformations in very field related to bioinformatics will help us to take Pakistan beyond skylines and to stand with high tech world association.

**Keywords**: Computational tools; Disease management; E-science in Pakistan; *In-Silico* approaches; Life sciences; NGS technology

**Introduction**
Bioinformatics came into existence to organize all biological raw data with help of different statistical and mathematical rules [1]. Here bioinformatics provides a major advantage over experimental data where information may be lost due to inconvenient methods. So, bioinformatics serves as a hierarchy of biological data where databases have been helping mankind for over a century. Being a part of recent trends, Pakistan is trying to clarify
with outer world, utilizing bioinformatics dug-up genetic irregularities i.e. SNP related disorders, albinism, Down syndrome, epilepsy and to conquest other challenges [2, 3].

In Pakistan, a number of applications have been performed making e-science an emerging trend of recent era. Major applications include agricultural industry promoting frontline for food industry as well as genetically modified crops have been proven beneficial to overcome food deficiency in Pakistan. These modifications are carried out by different gene targeting approaches, having impact on genomics, proteomics and transcriptomics. Bioinformatics plays an advanced role in determination of all key factors involved in foodborne genetic errors. Each and every aspect of a gene can be monitored easily, thereby providing information about structural and functional properties of individual protein [4].

In different aspects similar to genetic errors, personalized drugs provided a gateway towards disease diagnosis where each activity of gene is monitored at molecular level and in case of any abnormality, gene therapy is used to treat that abnormality [5, 6]. Bioinformatics provides an efficient source for locating these types of abnormalities. Only small dimension in Pakistan is working on this aspect but near future possibilities declare the value of bioinformatics in different research projects. Being in the list of developing countries, increased pollution is one of the major problems in Pakistan and is causing harm to all biological species. To avoid this problem, bioinformatics plays a crucial role, as a toolkit for the environmental analysis such as bioremediation that brings a positive impact on the environment [7].

Bioinformatics; Birth of new scientific era

Bioinformatics employs the computational science, algorithm, and statistics to interpret and analyse scattered biological data. Data related to biological molecules is arranged in a static manner for better understanding and localization of information. For the sake of interpretation, biological data is
aligned, identified and analysed by utilizing different tools which are easily available on different databases like GenBank and SWISSPROT [1, 2]. With passage of time, experimental approaches to identify and store biological data steps are forwarded to an advanced era from experimental to statistics analysis. As, on April 2001, a database, the GenBank had information of 11,546,000 entities whereas, SwissProt database of protein confined to 95,320 entities [11] that rises after every 15 months. Moreover, after whole genome publication of the *H. influenzae* genome, about 300 different organisms were identified having range of 450 to 100,000 genes afterwards. Considering gene expression data from myriad of different associated projects, acts as a way to initiate gigantic quantity forming evidence diversity [12].

Bioinformatics opens up a platform for researchers to preserve data as their heritage. The journey of bioinformatics which was started from development of ARPANET database in 19th century, did not confine at any step but has continued its progress till 2017 (Figure 1) where IEEE International Conference on Bioinformatics and Biomedicine (BIBM) gave an explained data on genes. Until now, various databases of different sizes with versatile information have come into existence [13].

![Figure 1. Bioinformatics; journey from 19th to 21st century with formulation of advanced databases](image)

**Bioinformatics; relation with other sciences**

With the moon of 21st century, a lot of development in field of bioinformatics took place and it made glue for modern science holding all disciplines together. Bioinformatics is often confused with Bio computation or Compactional Biology. Although it is quite comparable to computational biology but on the other side it is distinct from it. The first one uses biological data to design various algorithms and models for better understanding of biological systems. While the latter one uses different computational and statistical tools to understand and analyse biological data in a comprehended way. In the field of life sciences, Bioinformatics provide a variety of tools and programmes to the researchers in order to manage their biological data. Most of these tools are available online and provide free access to
user. Being an interdisciplinary field, the tools for in silico analysis are used to study interactions of biomolecules, phylogenetic analysis, visualization of physiochemical properties of proteins and in studying the structure of genes. For example, in the field of biology, it imparts its role in gathering different disciplines as biochemistry to study protein interactions and their structure or as biophysics to find out their functions and effect on biological system. It also involves improvement and implementation of novel informatics in biological sciences. With publication of Human Genome Project (HGP), more challenges have been opened in field of functional proteomics, helping pharmaceutical approaches and as a result specific drug targeting has come under influence [6].

Similarly, mathematics and physics interact with bioinformatics to implement different applications on scattered biological data. The influence of mathematics can be wider in some cases, due to rapid processing and without any extensive transformation steps that usually happen in chemical researches and in permitted pharmaceutical practices [3]. Moreover, Computer science is the root of bioinformatics where data mining of biological system, their informatics management and most importantly web designing for specific data is possible.

Bioinformatics; Urge of era in Pakistan

Globally, bioinformatics is an emerging field due to its unique and diverse features hence its importance cannot be neglected in Pakistan. Naturally, Pakistan is rich with agricultural, farm and dairy sources due to its geographical location and environment. Pakistan needs bioinformatics to study genetic abnormalities including Single Nucleotide Polymorphisms (SNPs) which are responsible for genetic disorders like Diabetes, Alzheimer’s disease, Down syndrome, albinism and epilepsy. In Pakistan, the pharmaceutical industry is widely spread thus, Bioinformatics helps out to design and develop new drug candidates and techniques to identify and eradicate novel diseases. In this prospect, Pakistan aims high to uphold its standards in agricultural and pharmaceutical sectors particularly with aid of bioinformatics [1, 2].

Bioinformatics; a new E-science and its applications

Genotypic Analysis

Genotypic analysis is a key to explore the genetic sequences at molecular level. It also assesses the rare genes and related disorders. 1000’s of genomic projects are in progress in this regard [14]. Entire genome is difficult to analyse, so NGS is employed which gives short fragments of the targeted DNA. Then, these fragments are analysed and sequenced using high-throughput computational analysis to determine the normal sequences or mutations especially SNPs [15]. The selection of tools depends upon the type of analysis going to be carried out by researcher. Tools that are used to analyse the NGS products include ngsPopGen, ngsStat, ngsFST, ngsCovar and ngs2DSFS. ngsPopGen is used for analysis of allelic frequencies and ngsStat for analysis of segregating sites and heterozygosity. While ngsFST, ngsCovar and ngs2DSFS are implemented for analysis of SNP variations, genotypic uncertainties and for comparison of genetic alteration between two populations respectively. Similarly Gene Quiz is used to search for different patterns in the available genomic data and Gene View for the graphical visualization and analysis [15]. Single nucleotide polymorphisms (SNPs) are termed as most recurrent genetic variations. Owing to their high prevalence, they are referred as significant genetic markers having a considerable role in pharmacogenomics and agriculture. SNPs are computationally analysed by executing several disparate bioinformatics tools linked to different features of protein (Table 1). Pakistan is also emerging in the world of automation and is portraying an immense headway towards the bioinformatics analysis in genomics [18]. In Pakistan, some bioinformatics expertise made in
Various protein studies have scrutinized in Pakistan to predict cancer progression along with viral and bacterial diseases. A virus named *Avian adenovirus-4* (AAV-4) is the cause of a poultry disease renowned as Angara Disease (AD). Bioinformatics plays a role in analysing three-dimensional protein encoded by the hexon gene of AAV-4 virus by implementing ExPASy server (www.expasy.org/tool) while the epitopes are detected by ABCpred server. This data has extensive potential to be used for vaccine preparation [24]. KEGG and CD-HIT employed for analysing metabolic pathways and enzyme structure respectively. InterProScan and SVMProt are in practice for the characterization of hypothetical proteins [25]. On the other hand, protein LPAR6 is responsible for a hair loss disease called hypotrichosis, is analysed by various bioinformatics approaches e.g. Ensemble Genome Browser (http://www.ensembl.org) for determination of primary structure and MODELLER 9V8 tool for 3D analysis [26, 27].

**Figure 2. Bioinformatics tools involved in genotyping analysis of different diseases**
Bioinformatics and transcriptomic analysis

Understanding of transcriptome provides a snapshot of whole genome. Targeting of isolated RNA has gained importance in various diagnostic and analytical techniques for exploration of interactions between various molecules. Moreover, targeted RNA sequences are also identified to synthesize random RNA transcript for visualization of various therapeutic and enzymatic applications [7]. Earlier Microarray technologies were used for RNA analysis but focus has now been shifted towards sequencing technique, the most important of which is RNA based sequencing used for transcriptome that provides information regarding novel gene transcripts, alternative splicing mechanism and fusion transcripts [28].

In past years, NGS technology has been employed for complete genomic analysis of 69 years old Pakistani individual, in Karachi by using HiSeq2000 Genome Analyzer while SOAP-linger software (http://soap.genomics.org.cn/soap1/) was used for sequencing and subsequent alignment of data. This sequencing information provides an insight towards the genetic diversity in the South Asian region and knowledge about human ancestry and evolution [29].

Similarly, the transcriptomic analysis of S. incanum (a source of bioactive phenols) has been performed by generating Expressed Sequence Tags (ESTs) using RNA-Sequence. Such analysis can be served as an initiative point in marker development for breeding purpose, exploring different splicing patterns and understanding post-transcriptional modifications [38].

Recently, the exome sequencing has been performed in 121 large consanguineous Pakistani families suffering from an autosomal recessive intellectual disability (ARID), reveals heterogeneous DNA variants in a single gene and also determines the transcriptomic profile of novel candidate genes. The results of such studies have expanded the former existing repertoire of genes involved in ARID and visualizing the mechanisms behind this disorder and transcriptomic maps of ID.

Bioinformatics and health care development

The enormous amount of available proteomics and genomics data provides us an opportunity to develop novel therapeutic approaches and research methodologies for disease prevention and treatment. However, this goal can’t be achieved without effective data management and integrating it with clinical and medical records. The electronic health record (EHR) and clinical data processing techniques will assist to establish an association between patient care, disease progression and public health [39]. Improvements in healthcare sector require data storage, computing infrastructure, effective data management and manipulation techniques to critically analyse biological data (Figure 3) [42, 43].

For effective data management, different tools are applied in different data processing sections [40, 41].

In Pakistan, health care department is trying to overcome hurdles by developing resources and various bioinformatics tools for biological data storage, data mining, data processing and in-silico functional prediction. Pre-natal diagnosis of β-thalassemia with aid of Bioinformatics tools has been made that eventually prevent emergence and spread of disease in Pakistan [8]. Combined study of bioinformatics along with population genetics can be made in Pakistan to visualize various aspects of genetic disorders such as Diabetes and Haemophilia. To accomplish this goal, there is a need to invest in development of infrastructure for genomics and proteomics research and well-equipped bioinformatics laboratories.

A scientific approach for disease analysis

With NGS technologies, many diseases have been characterized, utilizing bioinformatics tools and constituting databases for specific diseases [30]. RegressHaplo is used to characterize and
analyse viral evolution during HIV and SIV infections [31]. In same context, Los Alamos (www.hiv.lanl.gov) is a database for HIV which contains all subtyping, sequencing data related to HIV [32]. Bioinformatics is used to explore common genetic bases between T2D and AD. For instance, role of an hypothetical protein PCNXL2 involved in tumor suppression was analysed based on its functional and in-silico study [33]. Bioinformatics came into play as HaploReg (http://compbio.mit.edu/HaploReg) that analyse SNPs. Software DAVID provided information about commonly shared pathways. This analysis has potential therapeutic implications for treatment of these diseases [34]. For AD treatment, the role of miR-142-5p is testified using bioinformatics as Gene Expression Omnibus provides RNA sequences, whereas miRbase (microRNA database) (http://www.mirbase.org/) and Target Scan provide genomic coordinates of miRNA database [35].

Pathway Mapper provides an easy way to study gene, its interactions and alterations frequencies of genes in cancerous pathways [36]. For cancer treatment, drug combinations are one of the most effective ways to fight back for drug resistance, [37] while bioinformatics also exposes non-tested pairs of therapeutics. In Pakistan, many researches have been carried out for disease treatment and analysis which are summarized in (Figure 2) [38]. In benefit to researchers and developing laboratories, bioinformatics is playing its role i.e. Biomedical informatics lab has been established in PIEAS Islamabad, Pakistan with aim to develop sense of bioinformatics among young scientists. It also aims to solve biological and medical problems by means of computer (http://faculty.pieas.edu.pk/fayyaz/bmi.html). Moreover, another bioinformatics-based laboratory has been established in UET, Lahore named as Bioinformatics Research Lab (BRL) with four complete projects and many more to come in near future (http://www.kics.edu.pk/labs/about/brl).

Figure 3. Various Bioinformatics tools and technologies of data manipulation used for efficient analysis of Biological data
Computational approaches towards personalized medicine

Personalized medicine or customized medicine is based on individual genetic makeup. With the advent of Human Genome Project, personalized medicine has become an ardent topic in medicinal chemistry [39]. Genome-based analysis has been permitted by bioinformatics tools such as genomics and proteomics sequence databases (Table 2) [40]. Brick analysis based on proteomics and proteinchip, set as a remarkable tool in developing personalized medicine. Pharmacoproteomics based on genotyping paved a way to identify single variation having correlation of genotype and phenotype and variations during post-translational modifications [41].

Molecular diagnosis here plays an astonishing role in identifying drug response in presence of mutation such as SNPs. It is carried out by analysing efficacy of drug to its receptor, cluster analysis and protein structure forecasting [39, 42]. Human genome analysis of Pakistani individuals has paved a new way towards personalized medicine, giving crucial information about an individual’s genetic makeup and prevents him from observing adverse effect of any unrelated drug due to variations using pharmacogenomics profiling. Novel genes having effect of pharmaco-genes can also be identified by pharmacogenomics profiling, helping in identification of new treatment strategies for manifold diseases such as hypertension and neurological disorder etc. [43].

Bioinformatics: A new window for agricultural industry

The most crucial sector to be studied is agriculture as most of the developing countries totally rely on it. Advancement in the field of molecular biology paved a way to develop achievements in agriculture sector [44]. Bioinformatics, an e-science plays a remarkable role to inspect and collect all data from molecular biology in collaboration with statistical techniques. Sequencing analysis of several species (Triticum, Oryza sativa, Saccharum officinarum, Gossypium, Helianthus, Brassica, Capsicum annuum L., Allium cepa, Solanum tuberosum) find new features by using different in-silico tools [45]. Moreover, agricultural science is aided by NGS, a formidable tool to study, analyse and identify agriculture genomics. It helps in recognition of diversity, allelic variation, and phenotypic diversity in crops or livestock. The study of plant genomics enables the developers to introduce such species of plant which have enhanced herbicides, pesticides, drought resistance and to improve nutritional quality of various crops.

NGS also plays a valuable role in identification of molecular changes and gene cloning during traits selection. Bioinformatics tools, (Table 3) have been proved helpful to get and apply genetic information to infer structure, function and role of a particular gene on plant or livestock growth [46]. At current, the population of Pakistan is about 180 million and till 2035, it is expected to reach at 240 million. Pakistan not only consumes agricultural products as a source of livelihood but 20.9% of GDP also depends on it. Every year huge quantity of food is spoiled, and cultivation area is reduced due to increased population. The way to overcome these hurdles throughout the world and like so in Pakistan, is to use Genetically Modified Crops. Because of controversies and poor management, the only GM crop approved in Pakistan is Bt cotton which is mainly grown in southern Punjab. Different labs in Pakistan as Centre of Excellence in Molecular Biology (CEMB), Pakistan Agricultural Research Council (PARC) and National Institute for Biotechnology and Genetic Engineering (NIBGE) are working on such safe GM crops. There is still a long journey need to cover and to finds novel solutions to overcome such problems and so on for pathogen identification and microbe profiling to increase soil fertility [47].
Bioinformatics; a tool kit in environmental study

Environmental pollutants disturb environment and ecological balance in a threatening way. The processes of industrialization, technical advancements and consumption of natural resources have produced many new harmful pollutants. Naturally, microbes are involved in cleaning environment by degradation of the organic pollutant via process of bioremediation. Use of bioinformatics regarding this aspect demands data integration from different resources such as reactivity and chemical structures of organic pollutant, comparative genomics, study of proteins/enzymes (their structure and functions) and microbiology. At present, there is lack of bioinformatics tools for dealing with bioremediation but trials are on the way to interrupt genomic and proteomic data obtained from databases to assist environmental cleaning [7]. The University of Minnesota Biocatalysts/Biodegradation Database (UMBBD) is the prominent source to organize and store a large amount of data related to biodegradation and provides information related to almost every aspect of biodegradation including metabolic degradation pathways. Just like UMBBD, database named as MetaRouter allows for consulting database and provides an integrated framework to collect, store and analyse information on organic compounds, their reactivity, enzymes structure and function and microbial community [48]. In Pakistan, there is pathetically low investment of bioinformatics in environmental sector due to less number of researchers and research organizations. Pakistan being a developing country is also trying to catch up with other countries in developing bioinformatics tools, data handling, and data mining processes. There is need to train the researchers in bioinformatics and providing the financial support. So, they can make efforts to work on in silico gene designing for bioremediation and so on.

Bioinformatics: A new frontier to food research

Food is a basic necessity for maintaining cellular machinery of body. Per report of world food and agriculture organization claims that food supply should be 70% till 2050 in order to meet demand of growing population which paved serious problem regarding food safety. NGS opens a new window to investigate bacterial strains for food safety and veritable qualitative solutions of foods [49]. Whole genome sequencing also predicts sequencing information of food borne intoxicant caused by different pathogens [50]. Food allergy is mostly caused by legumes, seeds and nuts, posing a threatening problem to its consumers. Naturally, IgE are specific antibodies, produced in body to cope allergens of similar structure. To eradicate these allergens, Bioinformatics tools such as SDAP, AllerMatch, and FARRP plays an astonishing role in identification and sequencing of food allergens. Bioactive peptides in food are also detected by computational tools such as TreMBL, ExPASy, SwissProt (http://www.mrc-lmb.cam.ac.uk/genomes/madam/pres/swiss2.htm) and UniProtKB (http://www.ebi.ac.uk/uniprot) is used to analyze primary structure of peptide. Other in-silico tools (Figure 4) like PoPS (http://pops.csse.monash.edu.au/) and BIOPEP (http://www.uwm.edu.pl/biochemia/index.php/en/biopep) is also used for proteolysis and discovering new bioactive peptide [51]. According to research, 39.19 % people are sensitive to food allergens in Pakistan [52] and their molecular similarities are easily identified through allergen databases [58]. In future, food safety databases will also act as a new frontier to prevent food from spoilage through microbial identification.
Bioinformatics in gene therapy

Gene therapy involves the replacement of inactivated or mutated gene. For treatment purposes, computational tools are routinely used for characterization of genes, structural determination and physiochemical properties of their products. For gene therapy, the mutated gene sequence and normal gene sequence has to be characterized along with the vector genome information [53]. For this purpose, several tools are employed in gene therapy (Figure 5) [54-57]. In Pakistan as many technologies, tools and treatment are not so much progressed, gene therapy also has not employed so much for disease treatment.

Research institutes working on bioinformatics in the premises of Pakistan

To improve technical knowledge of researchers, there must be an educational platform to address queries related to life and science. To overcome issues related to bioinformatics and to educate researchers, first initiative was taken by Capital University of Science and Technology (CUST) previously known as Muhammad Ali Jinnah University) in 2002 as an undergraduate degree. Later this idea was extended by the COMSATS Institute of Information Technology (CIIT) in 2006. After that, several universities started their undergraduate and postgraduate degree programs to educate and promote awareness of people regarding bioinformatics. In this regard, HEC took a great initiative by introducing scholarships in bioinformatics with the aim to boost up and develop research projects with increased economic values. At present time, almost 20 different universities offering undergraduate and postgraduate degree programs including 5 HEC top-ranked universities [58].

There are almost 98 on-going research programs throughout Pakistan in different research institutes of Pakistan, including almost 46 research programs in HEC top-ranked universities (Table 4). In the province of Punjab, there are 35 on-going projects in different universities. So as 10 on-going projects in Islamabad, 32 in Sindh, 18 in Khyber Pakhtunkhwa, 2 in
Baluchistan and 1 in Azad Jammu Kashmir [6]. On national level, different institutes in Pakistan are organizing workshops and seminars that emphasizes on applications of bioinformatics. In 2011, a workshop was organized by National Institute of Biotechnology and Genetic Engineering (NIBGE) with purpose to provide keen knowledge about applications of bioinformatics and its role in other scientific fields (http://www.nibge.org/data/Brochure.pdf). Further, another workshop was held in NIBGE in 2015 about role of bioinformatics in other experimental techniques including transcriptomics, proteomics and genomics analysis (http://www.nibge.org/data/Brochure_fin.pdf).

Figure 5. Bioinformatics tools for different characterization of genes for gene therapy

Bioinformatics and laboratories in Pakistan
Talking about medical aspects, Pakistan has made certain developments in clinical and diagnostic field such as Shaukat Khanum diagnostic centre and HEJ, Karachi. These clinical laboratories are famous in conducting almost every type of diagnostic tests by means of advanced methodologies. Bioinformatics as an emerging field is
known to be helpful in these diagnostic centres also. Computer based solution of every possible problem is now available, requiring small sample size and cost effective. HEJ Research institute of chemistry in Karachi, Pakistan is taking steps to develop bioinformatics in Pakistan by arranging different internships, seminars and trainings on molecular and drug research in relation to bioinformatics (http://iccs.edu/).

**Bioinformatics business in Pakistan**

Government Policies are essential in the promotion of any field in the country as it derives new possibilities. The bioinformatics is one of the emerging fields in Pakistan having integration with IT. In developed countries, the government provides advanced ways of research and necessary funds for new automated machinery. To face challenges of modern bioinformatics, the government of Pakistan should invest largely in facilitating funded research, advancements in lab equipment and exploring modern technologies. The national centre of biotechnology was created in 1988 with purpose to create an information system of biological and molecular data. Due to this biological information system, the projects like R&D are managed and supported. Biotechnology information system provides support in conducting workshops; preparing dummy data for practices and assists the research communities and biologists to communicate information properly. Pakistan is also an active member of biotechnology information system and in Pakistan various institutes are working on different perspectives of biotechnology. Pakistan Biotechnology Information centre (PABIC) was established at Latif Ibrahim Jamal National Science information centre and in Karachi, the institute named as Sir Syed University of Engineering and Technology arranged. The main purpose is to create awareness, promotion and discussion on scope of bioinformatics. These institutes also aim to focus on the computer technology to manage the biological information in digital form. It merges other fields to understand, process and manage biological data. The technology is flexible and can be applied in drug discovery, new herbicide- resistant crop combination and gene, protein and cell-based research. By establishing research institutes, universities, laboratories and companies bioinformatics can be promoted in Pakistan [8].

**Role of bioinformatics in industries**

In any country, industries have a significant role in providing platform for working and developing tools to create links with biological data and to provide the correct sequence. Pharmaceutical industries offer great jobs in bioinformatics for drug development and functioning [59]. Pharmaceutical sciences and study of functional genomics involve human tissue analysis, pluripotent stem cells and variation in non-coding regions. Computational methods are used to explore databases to examine drug response. Molecular pharmacology helps to understand molecular and cellular mechanisms of biological processes and also involves practices in drug action. The interest is shared with expert programmers to share human disease process. An important focus is to provide skills and make student experts in the clinical and biological research to make drug movement easy and safe [60]. Pharmacogenomics (PSPG) is recognized internationally in pharmacogenomics field, function genomics and precision medicine. Different research works on diabetes, heart disease, gene sequencing, RNA sequencing cancer and HIV are undertaken by pharmacogenomics and Human Genome Project (HGP) [61]. Research tool designing is accompanied by using 3D tools for RNA structural measurement while sequence analysis is involved in generating expression of tertiary protein. Computing methods, important tools for data capturing are involved in arrangement and resolution of biological data while database formation serves as an important factor for recording
and managing information to be served in digital form [62].

**Bioinformatics jobs in Pakistan**

Fields such as computing biological information, sequence managing and handling of biological data need computer experts thus, make job area vast. Companies need computer experts that can manage and sequence the biological data in a statistical way. Vast jobs and applications for database developers have been emerging in the business sector of Pakistan [63]. Laboratories can offer various work fields such as genome sequence and molecular pathway analysis and protein functioning to researchers, where they can work properly for benefits of mankind. Also, linkage of lab with technologies e.g. collection of biological data from different labs and then by using information technology and softwares are applied for better understanding of sequencing, functioning and pathways understanding [64]. Industries that create platform for bioinformatics, are working on a commercial level, including many pharmaceutical companies in Pakistan [IIUI, Department of Bioinformatics & Biotechnology].

**Investment in biotechnology**

The pharmaceutical industries in Pakistan operate on all levels. Industries on smaller and medium scale have to face challenges to survive in market. To fulfil needs of competitive productions and revenue, they also require knowledge about bioinformatics. IT firms found bioinformatics sector beneficial to invest and also invested their resources in the bioinformatics sector to make it 100% digital. Now, industrial sector fulfil their bioinformatics needs by making collaborations with these bioinformatics firms, making it beneficial for both sides [IIUI, Department of Bioinformatics & Biotechnology].

**Investment in R&D**

In Pakistan, 80% of medicines used have a foreign patent with extra amount of royalties. And a large amount of revenue is transferred from country decreasing the national profits. Investments can be made in bioinformatics research and development sector, discovering and developing modern medicines inside the country. This could ensure development of local pharmaceutical industry and secure most of the profits inside the country. This would decreases cost of production, thus decreasing the price of medicine in the market [65].

**Managing clinical and laboratory information**

Many gene sequences cannot be stored or record on paper because gene sequences are too complicated and large in structure so they would be easily subjected to errors. While storing in digital form, tools should be properly spaced and have pre-defined instructions written by professionals with zero chance of error [Thermo Fisher Scientific Inc]. Medical history of a patient is preliminary and most important for a physician while treating with a fatal disease which can’t be completely dragged through conventional ways, so there are softwares which track the patient’s daily routine and vitals etc to keep full record of his treatment history and physical changes after regular intervals [66].

In a clinical research system, software are used to check whether research is being conducted according to set pattern and plans. These softwares also store data regarding research participants, their contacts, credentials residence [67]. Computational biology is related to data structures and algorithms with computer modelling of biological systems. It helps to understand complex biological systems by designing predictive multi-scale models that help in discovery of new biomarkers for disease. Within networks, information at different levels integrates and organizes these networks with help of communication e.g. ISB [68]. The aim of this discipline is to make open-source software that can share information between different pathway models.
The clinical applications of bioinformatics can be observed in the immediate, short, and long term. A comprehensive list of human gene products such as insulin and growth hormones may offer new drugs and gene therapy for single gene diseases. The most common disorders of hypochromic microcytic anaemia (HMA) are Thalassemia traits (TTs) and iron deficiency anaemia (IDA). The aim of present study is to differentiate TTs from IDA through evaluating discrimination formulas and provides wide-ranging data of haemoglobin disorders which are dominant in Pakistan. There are 12 published discrimination formulas, among these 12 formulas, 6 formulas—MI, EF, G&K, RDWI, R, and HHI—were the most efficient to discriminate TTs from IDA. The random forest (RF) decision-tree approach was used to improve cut off values [75].

In preclinical stages of research and development, MS-based targeted phospho-proteomics was used and believed that final translation of this technology will expose new accesses in clinical setting. Phospho-proteomics, as an integrative approach with genomics and other omics data, may have a future hand in addressing the tasks of prostate cancer diagnosis and drug development by identifying active pathways. The technology would also cover the way for broadenfield of pharmaco-omics to logically select and modify a patient’s drug therapy for different diseases that have low mutation burden [76].

Major challenges in bioinformatics
A key task for bioinformatics is to find out useful information from a tremendous diversity of available biological data and correspond such information with biological knowledge [69]. However, it proves to be highly challenging in many cases: (1) increased data size typically handled with traditional computational approaches (2) information rich data is heterogeneous in nature, corrupted and incomplete, as well as deceptive outliers and (3) biological systems due to compliancy evaluation, sacking, robustness and exposure are exceedingly complex [70]. Elucidation of biological data demands finding of intricate relationships hidden in data. Due to these challenges, the precision of augury from a database is often not satisfactory [71]. It is clear that there is much room for further enhancement and development claiming novel theoretical structure and computational techniques. Other than technical challenges, human factors (cognitive and perceptual abilities of users) are also of great importance [72]. These challenges in bioinformatics have given rise to a broad spectrum of studies from computer sciences. Almost all available computer science techniques have petitions in bioinformatics [73]. Few applications of computational and statistical methods in bioinformatics e.g. dynamic programming, neural networks, hypothesis testing and clustering etc. have great importance. Software is easy to design rather than gadgets like cars and buildings. So, it is generally anticipated that future would meet the development of extensively huge and complicated software systems [9]. Software Engineering is closely associated to combinatorial issues along with new kind of challenges with superintend complexity.

Bioinformatics challenges; faced by biologists
Collins and colleagues from the US National Human Genome Research Institute wrote about future of genome research and proved computation biology as a valuable resource. At a bioinformatics conference in 2002, EBI’s Ewan Birney, MIT’s Chirs Burge, and Glaxo Smith K lines Jim Fickett proposed a precise and predictive model of transcription, initiation and termination to predict where and when transcription will occur in genome [74]. The precise and predictive model of RNA splicing/ alternating splicing provides splicing pattern of any primary transcript in any tissue. Vast understanding of effectual gene ontologies provides systematic ways to discover functions of any gene or protein [75]. Challenges in bioinformatics are
difficult to handle, i.e. they require even more time to integrate effects from other world groups to be solved. These challenges will be solved with help of vast amounts of data collected in the laboratories, analysed manually or semi-naturally to produce accurate data that are termed as bench markers for hypothesis testing. Every challenge has input and output and together they form a knowledge network to understand biological processes [76].

**Advancements of bioinformatics in Pakistan in different fields**

Bioinformatics is widely practiced within the pharmaceutical industry in the development of health-care products, and also in agricultural and environmental protection. Developments in pharmaceutical industry require recent tools and methods for discovery of drug targets, drug design and for the identification of new disease markers to improve early diagnosis and development of new therapeutic strategies. Pakistan is aiming to raise its research standards in biomedical sectors and in other fields by exploiting bioinformatics approaches in many life-science domains.

**Pharmaceutical areas**

Bioinformatics plays important role in all features of drug discovery, drug designing and drug calculation. This is because of bioinformatics tools to predict, analyse and help in clarification in clinical and preclinical findings. Increase in pressure to generate more and more drugs within limited life span with low risk has resulted in remarkable interest in bioinformatics. Computer-aided drug design (CADD) is a new separate field where bioinformatics provides a huge support to overcome the cost and time context in various ways. It also provides a wide range of databases and softwares, used for drug designing [77].

**Database management**

CLIP-seq, RIP-seq, ChIRP-Seq or Shape-Seq are RNA based methods which are recently invented and is a central topic in molecular biology to investigate RNA based regulation. RNA Bioinformatics provides a comprehensive set of analysis tools and work flows for data analysis. Incorporation of these tools in the Galaxy framework allows easy access to RNA workbench. The current developments in this regard are RNA-centric annotation efforts, sequence/structure integrated motif finding, extending RBP peak callers to new protocols, or detecting RNA modifications. Finally, an important task for the future will be designing and development of novel analysis tools and integration in workbench to provide accommodations in technological progress [10].

**Table 1. Different bioinformatics tools for genotyping of Single nucleotide polymorphisms (SNPs)**

<table>
<thead>
<tr>
<th>Bioinformatics Tools</th>
<th>Features</th>
<th>Website</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PinSnps</td>
<td>Prediction of SNPs having effective protein interactions</td>
<td><a href="http://fraternalilab.kcl.ac.uk/PinSnps/">http://fraternalilab.kcl.ac.uk/PinSnps/</a></td>
<td>[78, 79]</td>
</tr>
<tr>
<td>MuD</td>
<td>Predicting non-neutral and neutral amino acid sequences</td>
<td><a href="http://mud.tau.ac.i">http://mud.tau.ac.i</a></td>
<td>[80]</td>
</tr>
<tr>
<td>Meta-SNP</td>
<td>Genotyping of non-synonymous SNPs</td>
<td><a href="http://snps.biofold.org/meta-snp">http://snps.biofold.org/meta-snp</a>.</td>
<td>[81]</td>
</tr>
<tr>
<td>PROVEAN</td>
<td>Analysing functional substitutions of amino acids</td>
<td><a href="http://provean.jcvi.org">http://provean.jcvi.org</a></td>
<td>[82]</td>
</tr>
<tr>
<td>SIFT 4.0.4</td>
<td>Genotyping of missense SNPs</td>
<td><a href="http://sift.jcvi.org">http://sift.jcvi.org</a></td>
<td></td>
</tr>
<tr>
<td>PolyPhen-2</td>
<td>Genotyping of missense SNPs</td>
<td><a href="http://genetics.bwh.harvard.edu/pph2">http://genetics.bwh.harvard.edu/pph2</a></td>
<td>[83]</td>
</tr>
</tbody>
</table>
Table 2. Different Databases based on variations in each individual

<table>
<thead>
<tr>
<th>Database</th>
<th>Website</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGMD</td>
<td><a href="http://www.hgmd.cf.ac.uk/">http://www.hgmd.cf.ac.uk/</a></td>
<td>Disease-related gene lesions</td>
</tr>
<tr>
<td>OMIM</td>
<td><a href="http://www.omim.org/">http://www.omim.org/</a></td>
<td>Human genes and genetic disorders</td>
</tr>
<tr>
<td>VnD</td>
<td><a href="http://vnd.kobic.re.kr/">http://vnd.kobic.re.kr/</a></td>
<td>Variation and drugs</td>
</tr>
<tr>
<td>PinSnps</td>
<td><a href="http://fraternalilab.kcl.ac.uk/PinSnps/">http://fraternalilab.kcl.ac.uk/PinSnps/</a></td>
<td>Protein-protein interaction networks</td>
</tr>
<tr>
<td>HUMA</td>
<td><a href="https://huma.rubi.ru.ac.za">https://huma.rubi.ru.ac.za</a></td>
<td>Comprehensive biological database including variation</td>
</tr>
<tr>
<td>DGVa</td>
<td><a href="http://www.ebi.ac.uk/dgva">http://www.ebi.ac.uk/dgva</a></td>
<td>Structural variation</td>
</tr>
<tr>
<td>NHGRI-EBI catalog</td>
<td><a href="http://www.ebi.ac.uk/gwas/home">http://www.ebi.ac.uk/gwas/home</a></td>
<td>Manually curated database of published genome-wide association studies</td>
</tr>
<tr>
<td>SNPs3D</td>
<td><a href="http://www.snp3d.org/">http://www.snp3d.org/</a></td>
<td>Functional effects of no synonymous SNPs</td>
</tr>
<tr>
<td>UniProt</td>
<td><a href="http://www.uniprot.org/">http://www.uniprot.org/</a></td>
<td>Protein database including no synonymous SNPs</td>
</tr>
<tr>
<td>EVA</td>
<td><a href="http://www.ebi.ac.uk/eva/">http://www.ebi.ac.uk/eva/</a></td>
<td>Public variation archive</td>
</tr>
<tr>
<td>COSMIC</td>
<td><a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a></td>
<td>Cancer-associated mutations</td>
</tr>
</tbody>
</table>

Table 3. Bioinformatics tools for next generation sequencing of genome and transcriptome

<table>
<thead>
<tr>
<th>Features</th>
<th>Program</th>
<th>Website</th>
<th>Repeat-relevant parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP detection</td>
<td>GATK</td>
<td><a href="http://www.broadinstitute.org/gsa/wiki/index.php/Downloading_the_GATK">http://www.broadinstitute.org/gsa/wiki/index.php/Downloading_the_GATK</a></td>
<td>None available or none required</td>
</tr>
<tr>
<td></td>
<td>Sniper</td>
<td><a href="http://kim.bio.upenn.edu/software/snip">http://kim.bio.upenn.edu/software/snip</a> er.shtml</td>
<td>Mapping parameters</td>
</tr>
<tr>
<td></td>
<td>Var Scan</td>
<td><a href="http://varsan.sourceforge.net">http://varsan.sourceforge.net</a></td>
<td>None available or none required</td>
</tr>
<tr>
<td>Short-read alignment</td>
<td>BFAST</td>
<td><a href="http://bfast.sourceforge.net">http://bfast.sourceforge.net</a></td>
<td>Reports all locations by default</td>
</tr>
<tr>
<td>De novo assembly</td>
<td>CABOG</td>
<td><a href="http://wgs-assembler.sf.net">http://wgs-assembler.sf.net</a></td>
<td>Reassemble misclassified non-unique unitigs: doToggle = 1</td>
</tr>
<tr>
<td>Spliced read alignment</td>
<td>GSNAP</td>
<td><a href="http://share.gene.com/gmap">http://share.gene.com/gmap</a></td>
<td>------</td>
</tr>
<tr>
<td>Reference-guided transcript assembly</td>
<td>Cufflinks</td>
<td><a href="http://cufflinks.cbcb.umd.edu">http://cufflinks.cbcb.umd.edu</a></td>
<td>Improve repeat read mapping estimate: --ulti-read-correct</td>
</tr>
<tr>
<td>De novo transcript assembly</td>
<td>Multiple-k</td>
<td><a href="http://www.surgetgroba.ch/downloads">http://www.surgetgroba.ch/downloads</a></td>
<td>None required or none available</td>
</tr>
</tbody>
</table>
Table 4. On-going bioinformatics research projects in HEC top-ranked universities [6, 58]

<table>
<thead>
<tr>
<th>Research objective</th>
<th>Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping and screening copy number variation in Pakistani individuals</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Descriptor based modelling for the identification of anti-inflammatory agents in Pakistan</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>2D-QSAR analysis of anti-inflammatory and anticancer compounds</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Genome wise association studies using microarray chip data</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Structure/ligand based virtual screening for identification of novel leads for Human and Plant diseases</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Epitope prediction of different pathogenic viruses in Pakistani diseases</td>
<td>Centre of Excellence in Molecular Biology, University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Next generation multi scale cancer systems biology modelling platforms</td>
<td>Biomedical informatics research laboratory, Lahore University of Management Sciences, Lahore</td>
</tr>
<tr>
<td>Brain modelling with brain computer interface for wheel-chair navigation</td>
<td>Al-Khawarizmi institute of computer science, University of Engineering &amp; Technology, Lahore</td>
</tr>
<tr>
<td>Serodiagnostic tests for infectious diseases of the poultry</td>
<td>University of Karachi, Karachi</td>
</tr>
<tr>
<td>Rapid diagnosis and immune-prophylaxis of acute viral diseases in Canines</td>
<td>University of Veterinary and Animal Sciences, Lahore</td>
</tr>
<tr>
<td>Identification, characterization of drug resistant <em>E. coli</em> and its treatment: A perspective of novel antibacterial agents</td>
<td>University of the Punjab, Lahore</td>
</tr>
<tr>
<td>Isolation, purification and characterization of protease inhibitors from plants</td>
<td>Centre for Advance Studies in Biotechnology, University of Sindh</td>
</tr>
<tr>
<td>Production and characterization of engineered proteins containing non-natural amino acids; potential tools for specific drug targeting</td>
<td>University of Karachi</td>
</tr>
<tr>
<td>Development of new antiviral and anticancer steroids by combinatorial synthesis and high-throughput biological screening—a new approach to drug discovery</td>
<td>University of Karachi</td>
</tr>
<tr>
<td>Peptidyl antibiotics: structure-function relationship of the antimicrobial peptides and bacteriocins/antibiotics from indigenous Enterococci</td>
<td>University of Karachi</td>
</tr>
<tr>
<td>Photo dynamic Diagnosis System: A new method for the early recognition bladder tumour</td>
<td>Liaqat University of Medical and Health Sciences, Jamshoro, Sindh</td>
</tr>
<tr>
<td>Studies on the effect of Vitamin C supplementation on lead levels of blood, hair and urine in adults</td>
<td>The Aga Khan University, Karachi</td>
</tr>
<tr>
<td>Determination of essential and trace elements in biological samples of human subjects with various physiological disorders</td>
<td>University of Sindh, Jamshoro</td>
</tr>
<tr>
<td>Atherosclerosis—an inflammatory disease: role for Cyclooxygenase-2</td>
<td>The Aga Khan University</td>
</tr>
<tr>
<td>Phenolic allelochemical of sunflower as natural herbicide for weed management in wheat</td>
<td>University of the Punjab, Lahore</td>
</tr>
</tbody>
</table>
Conclusion
Our main focus is on current applications of bioinformatics in Pakistan, emerging an era of e-technology development and taking the concept of science towards a new horizon. PABIC is contributing fundamental part in digital data and research routes, creating an opportunity for growth of many novel fields in Pakistan. System architecture and designing of biological data are involved in making bioinformatics promising and inclusive platform for numerous arenas. Talking about Pakistan, a nature rich country, it plays a vibrant role in the chief sectors including pharmaceutical and agricultural zones. There are number of relinquished institutes proposing bioinformatics in their systematic curriculum. In Pakistan, it is the prerequisite of time for young scientists to get acquaint with the benefits and opportunity of bioinformatics in this era, providing new ones with the opportunity to create their own world of web-tech and play experimentation with it. At this time, it is obligatory need to team up with global biotic oriented intelligence to advocate standards. Pakistan in this regard plays its role in reshaping genomic and proteomic areas, where life is at eye site to observe and modify. Future towards new vista containing e-fashioned world in Pakistan where there are chances of introduction of Computer aided drugs along with the implementation of phospho-proteomics. In near century, there will be fewer miseries and more technologies and Pakistan will lead to high e-tech biological routes leaving digital biologic designs for its heirs.

Authors’ contributions
Conceived and designed the experiments: H Sadia & M Naveed Analysed the data: MA Saleem, M Naveed, M Amin & F Amjad, contributed reagents/ materials/ analysis tools: AJ Iqbal & M Naveed, Wrote the paper: H Sadia, M Amin & S Asif.

References
4. Naveed M et al. (2018). Annotation and curation of hypothetical proteins:

| Biochemical characterization of in vitro salt tolerant cell lines and regenerated plants of potato and subsequent establishment under ex-vitro | University of the Punjab, Lahore |
| Generation of salt resistant tomato by using HAL II genes and molecular analysis of transgenic | Quaid-i-Azam University |
| Identification of biochemical indicators for salt tolerance and their genetic basis in potential oilseed crops, canola and sunflower | University of Agriculture, Faisalabad |
| Reconstruction of chromosomal inheritance in pedigree of Pakistani wheat cultivars | Kohat University of Science & Technology |
| Identification, tissue-specific expression and immunolocalization of stress proteins (dehydrins) under high temperature stress | University of Agriculture, Faisalabad |
| Vanadium complexes as Insulin Mimetic Agents: Coordination chemistry, characterization and biological studies of vanadium (IV) and vanadium (V) complexes of various ligands containing carboxylate group | Quaid-i-Azam University, Islamabad |


62. Rao VS *et al.* (2008). Recent developments in life sciences research:
Role of bioinformatics. Afr J Biotech 7(5).