Regular scientific paper reading as an approach to improve the outcome of bioinformatics teaching to graduate students of medical majors

DOI: 10.46981/sfjhv3n3-006

Received in: April 29th, 2022
Accepted in: June 30th, 2022

Xiangning Zhang
PhD in tumor biology
Institution: Department of Pathophysiology, School of Basic Medicine - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: zhangxiangning@gdmu.edu.cn

Jing Huang
MB in clinical science, MMed candidate
Institution: Department of Pathophysiology, School of Basic Medicine - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: 1473876648@qq.com

Longji Wu
MMed in pathophysiology, PhD candidate in cell biology
Institution: Department of Pathophysiology, School of Basic Medicine - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: wulongji@163.com

Zhi-Gang Huang
PhD in biostatistics and epidemiology,
Institution: Department of Epidemiology - School of Public Health - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: hzg71@126.com

Zhiwei He
PhD in pathophysiology of cancer
Institution: Department of Pathophysiology - School of Basic Medicine - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: zhiweihe668@yahoo.com

Hongbing Yu
PhD in stem cell biology
Institution: Chinese American Tumor Institute - Guangdong Provincial Key Laboratory of Medical Molecular Diagnostics - Guangdong Medical University
Address: Dongguan, Guangdong, 523808, People's Republic of China
E-mail: yhb1207@163.com

ABSTRACT
Since the elucidation of the double helix structure of DNA molecule and its role of coding for genetic information, rapid progress of biomedical science including the sequencing of macromolecules in the living system has been witnessed. The technique has allowed the study field to evolve from single genes to genome in species. With demand of deposit and analysis of data obtained, a new science bioinformatics has been formed by integration of computer science and biology. Bioinformatics has found wide application in research of different disciplinary in medical science. To facilitate the grasp of its basic theory and practical application by the postgraduate students at the medical school, regular reading of relevant publications has been proposed as an approach to improve the quality of teaching, and its outcome has been evaluated by comparing the scores earned from the identical test papers. The students acquired the designed intervention had higher score in average statistically with $p<0.01$.

Keywords: postgraduate, basic medicine, bioinformatics, genomic sequence.

1 INTRODUCTION

The rapid progress in biological science has been characterized with the renovation of sequencing of macromolecules, notably after the discovery of double helix structure of DNA and its coding potential of genetic information. The study of the genetic materials has evolved from single genes to whole genomic DNA in species. From the end of 20th century, the initiation of genomic nucleic acids to the complement of the draft of human genome has been witnessed. It has been warranted by the development in genomics not only the technique for sequencing of large amount of nucleic acids with large, but also the efficient approach to process and deposit of the data obtained. The relevant theory has also been widely applied in the field of biomedicine.

By integration of computer science, statistics, and biology, a new science, bioinformatics has been formed in the last years of 20th century. It supports the storage, retrieval and analysis of the laboratory data on structure, sequence and functions of biological study. Bioinformatics is about the molecular world with modeling or data "mining", obtaining knowledge from the fields of biochemistry, molecular biology, molecular evolution, and biophysics. At the early time bioinformatics was synonymous with analysis and management of DNA and protein sequence data, but now it has rapidly expanded with the advancement in full genome sequencing, functional genomics, pharmacogenomics, proteomics, metabolomics research, and biological pathway modeling [1].

Bioinformatics has provided great insights into medical research and practice, and public health. The course taught at medical school would greatly contribute to the design of thesis work as well as the future career of the graduate students of all disciplinary. The teaching strategy for bioinformatics has been proposed [2]. The objectives of the proposed skeletal set of modules addresses basic aspects of broad coverage such as molecular biology, molecular evolution, phylogenetic reconstruction, structural, functional and comparative genomics, chemical and biological kinetics, statistical mechanics of

biological macromolecules, biochemistry, thermodynamics of biological processes, computer science, personalized medicine and biomedical computing, statistics.

In view of the rapid development of bioinformatics, it is a huge task for the students to grasp the complicated concepts included in the class coaching. Regular distribution of literature in relevant scope has been proposed as a solution to facilitate the command of the basic theory as well as its practical application. The present study aims at discussing the outcome of approach to improve the quality of the teaching of bioinformatics.

2 METHODS

2.1 STUDY DESIGN

The course of bioinformatics for graduate students, mainly pursuing their MSc in medical science was started in 2013; the course was lectured by the first author together with at least two faculty members from the fields of cancer research and computational drug design. The involved faculty had long experience in the laboratory research and computational modeling for therapeutic target design with familiarity of latest advance of biomedical science, notably human genome sequencing, functional genomics as related to diseases.

Table-1 number of students entering the study

Year attended/acad. majors	Basic medicine	Medical technology
2020	14	39
2021	30	53

2.2 MANIPULATION

The students entering the study were those majored in basic medicine with specific field biochemistry and molecular biology, pathology, pathophysiology, physiology and pathogen etiology), and medical technology. One of such advantages of the mixed online and on spot teaching was the efficient distribution of reading materials, mainly scientific papers. A major improvement was that since the beginning of the coaching for the class of year 2021, some literature was distributed to the students entering the course of bioinformatics, in keeping up with the progress of teaching. The outcome of homework quality as well as the scores earned at the final tests was comparable between two groups of students, of year 2021 and 2020. Final exams are given at the end of each academic term.

Several papers were distributed within weeks of instruction of the class. The first one recommended to the students dealt with differentially regulated genes in a malignant tumor, nasopharyngeal carcinoma (NPC), which is endemic with southern China and Southeast Asia [3]. A number of students had proposed thesis theme on cancers, some focusing the type of tumor with high

incidence in the local region, NPC. The paper described identification of pathogenic genes in NPC screened with the microarray dataset; data from NPC and control samples were downloaded from the Gene Expression Omnibus (GEO) to analysis.

Up to 424 differentially expressed genes (DEGs), were selected from the dataset. DEGs were mainly enriched in extracellular matrix organization, cilium organization, PI3K-Akt signaling pathway, collagen-containing extracellular matrix, and extracellular matrix-receptor interaction, among others. Besides, the effect of genes CDK1, CDC45, RSPH4A, and ZMYND10 on survival of NPC was validated in GEPIA database. The data revealed novel aberrantly expressed genes and pathways in NPC by bioinformatics analysis, providing novel insights for the molecular mechanisms on genesis and progression of NPC. The identified genes even included the gene that we were working on in the lab, BLU on chromosomal 3p21 region coding for a tumor suppressor ZMYND10. The results described herein were insightful not only to the students entering the course.

The second one dealt with the comparison of primary structure, i.e. amino acid sequence and linear motif composition of a protein family of apoptotic regulators, Bcl-2 [4]. The family is composed of proteins with different motif composition and opposing effects on apoptosis, i.e. anti-apoptotic and pro-apoptotic. They are also highly conserved among species due to evolution. All of these can vary across cell or stress types, or developmental stages, and this can cause the delineation of the roles of BCL-2 family members. The distributed homework contains the questions regarding prediction of the function in apoptosis of individual molecules based on their linear structure. The paper discusses our current knowledge of anti-apoptotic BCL-2 family isoforms. With significant improvements in the potential for splicing therapies, it is important that we begin to understand the distinctions of the BCL-2 family, not limited to just the mechanisms of apoptosis control, but in their roles outside of apoptosis. Two additional papers were presented online, and discussion was initiated thereafter [5, 6].

At the end of the term for class 2020, and in the middle of the term for class 21, a test was given based on the identification of a co-transporter protein, NTCP (Sodium taurocholate cotransporting polypeptide) as the viral receptor of Hepatitis B Virus (HBV) in 2012, with interaction with HBV large envelope protein [7]. Questions were to be answered after reading the paper and analyzing the data obtained by bioinformatics tools from PubMed, Genbank, uniprot. The questions raised were included:

1. To locate the chromosomal mapping of the gene from online database together with its accession no. and its sequence in FASTA format;

2. To specify the mutant amino acid residue to alter the host susceptibility of the host to HBV and the clinical outcome of hepatitis B;

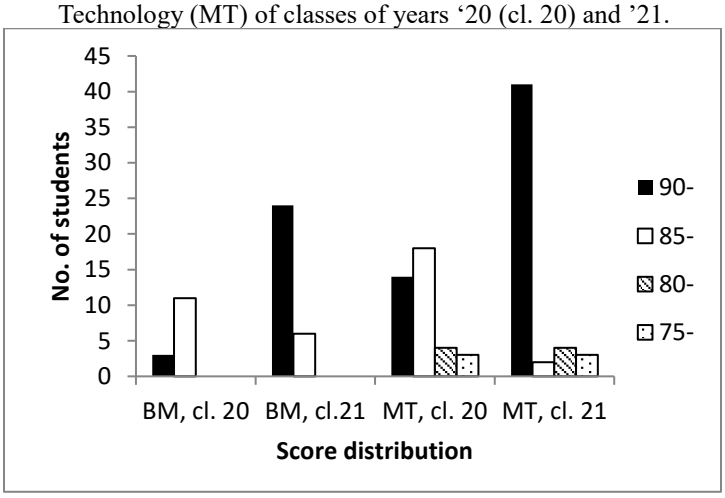
3. To analyze genes or proteins involved in thesis of the student using above-referenced methods, or any gene or its coding product that is interest to you.

The results were evaluated in the identical standard and scored; they were compared in terms of score range distribution and averaged level of the score presented as means ± standard deviation (SD), calculated by SPSS 19.0 software (SPSS Inc. Chicago, Illinois, USA) was indicated in Figure-1 and Table -2.

3 RESULTS

An intervention of recommendation and distribution of literature on algorithm, sequence analysis, functional annotation of the protein molecules was done during the teaching. In fact, during epidemic of COVID-19, a mixed teaching strategy has been introduced at our university. A specially designed software was mounted to include outline of each chapter of the textbook, teaching materials in form of powerpoint and voiced life video together with teacher-student interaction platform to allow chat and text comments. Homework and relevant reading materials are also distributed on this platform. After the students received the literature, they read them and answer the accompanied question in written. The completed homework was returned as e-mail attachment. The students also marked on the homework paper stating that permission was granted to the instructing teacher to use the data within the paper for published analysis purpose. The returned papers were scored in the same standard as the final test of the previous year. The results were processed statistically.

Figure-1 Distribution of scores earned in the same test by students with majors of Basic Medicine (BM) and Medical Technology (MT) of classes of years '20 (cl. 20) and '21.



The scores earned by the two groups of the students, taking the course in consequent years, 2020 and 2012 were compared. The scores of the year 2020 class were obtained the recorded final exam results, and those of year 2021 class were the registered data from their homework. The distribution of the score earned was displayed in Fig.1. As a collegiate regulation, the contents of the final tests are not permitted to repeat between different years. The content of this final test was distributed as home work

for the students entering the course the following years. Comparison of the scores was performed. The conclusion reached would be insightful in further improvement of the teaching in coming years. The final test in 2020 was used as homework for the class of 2021, and the scores earned were compared. As revealed by the current study, the results have been much improved, partly due to the intervention with regular reading.

Table-2 Comparison the average scores earned by students before (class '20) and after (class '21) the described intervention. Mean values were calculated by averaging the scores of each group.

Year attended/majors	Basic medicine.	Med. technology
Class '20	86.7857142±4.353222881	87.6±5.06878095
Class '21	92.37931034±2.922064602	92.2±2.545678786
P values '21 vs '21	0.001044335	3.95075E-05

The scores for comparison were presented in Table-2, in form of mean standard deviation (SD). It has been suggested that the overall levels of the scores for year 2021 classes of two majors have been elevated, and the range of variation of the value has been reduced, as reflected by the smaller SD figures.

4 DISCUSSION

A mixed online and on spot teaching has been introduced to our university, like all schools in China due to the epidemic of COVID-19 in the academic term starting in the first half of 2020. It possessed a challenge to the traditional teaching method, and also provided some opportunity, because the online teaching facilitates the communications between the students and faculty after the class, among which, distribution of literature through internet meets the demand of spread information.

The sequencing technology has rapidly progressed with the advancement of molecular and cell biology, genetics, and ecology. Since the final years of 20th century, the genome of viruses and low free-living organisms has been sequenced. The sequenced species included a protokaryotype bacterium *Haemophilus influenzae* with 1.8 Mb in 1995 [8]，and the baker yeast, *Saccharomyces cerevisiae* in 1996 [9, 10]. The first human chromosome was sequenced in 1999: Chromosome 22 49 Mb, 673 genes [11]. The HUGO (human genome project) focusing on whole genome of human was initiated; the genome with size of 3 billion was sequenced, leading to 2001: draft sequence of the human genome was completed by NIH based public consortium and Celera Genomics in 2001 [12]. The hallmarks in genome sequencing were mentioned in classroom coaching, and our recommended literature provided information well complementary to the lecturing.

The progression of HUGO has prompted the renovation of computer technique, to deposit, analyze and annotate the data obtained from the sequencing. As an integrated science, bioinformatics has been emerged by the combining lab methods of biological science and computational technology.

Meanwhile, the hotspot of biomedical research has evolved from large scale sequencing and annotation of the biological functions, as well as dissecting the linear and spatial structure of the genomic coding products forming the basis of various biological functions. With expansion of the online public database, the professional skills in preparing sequence and structure related lab data submitted to such resource becomes part of daily work in research. Understanding the routine of preparing such materials becomes essential to biomedical researchers. Our lab work deals with a tumor suppressor BLU mapped on chromosomal 3p21 region; the fragment containing the gene is frequently lost in a variety of human tumors. In fact, BLU was initially identified as a lost fragment termed beta-catenin in lung cancer. It has been reported that BLU is inactivated in esophageal and nasopharyngeal cancers [13, 14], and hepatoma [15] due to hypermethylation on its promoter region. We have recently shown that the coding product of BLU, a zinc finger protein ZMYND10 exerts tumor suppression through induction of a repressive histone mark and hence downregulation of cyclins to arrest cell cycle progression [16]. In 2015, we submitted an annotation on BLU/ZMYND10 in terms of its linear sequence, motif composition, and pathogenic role in human cancers in a manner of loss-of-function [17].

Among different reasons, for example, better general scholarly qualification achieved with years, it was reasoned that the improvement teaching approach by regular recommendation of literature in relevant field so as to stimulate the thinking of students contributed to the elevation of the average score earned. In the survey attached to homework paper, students also commented on performance of the classing teaching; this formed the encouragement and motivation for constant improvement of our coaching quality.

It becomes obvious that an adequate command of bioinformatics facilitates the thesis work as well as the future professional career of the medical graduate students in different disciplinary. Improved teaching method would contribute to a better outcome of the coaching.

ACKNOWLEDGMENT

The work described in the present manuscript was included in the project of Southern Medical University- Foundation for Advancement of International Medical Education & Research (FAIMER) training program that XZ registered in September 2021. The work is also in part supported by a research grant of Medical Research Fund from Guangdong Provincial Health Commission (2018A256 to XZ). Dr. Lin Lv, the project advisor is highly acknowledged for her stimulating discussing for the study design and Dr. Xiaoqing Zhang for sharing her valuable knowledge in bioinformatics.

REFERENCES

Dimitrov, R.A. & Gouliamova, D.E. (2009). Bioinformatics Education: Perspectives and Challenges, Biotechnology & Biotechnological Equipment, 23:sup1, 40-42, DOI: 10.1080/13102818.2009.10818360.

Boguski, M.S., & McIntosh, M.W. (2003). Biomedical informatics for proteomics. Nature 2003; 422(6928):233-237. doi: 10.1038/nature01515.

Fang, L., Shi, L., Wang, W., Chen, Q., & Rao X. (2021). Identifying key genes and small molecule compounds for nasopharyngeal carcinoma by various bioinformatic analysis. Medicine (Baltimore). 100(37), e27257. doi: 10.1097/MD.0000000000027257.

Warren, C.F.A., Wong-Brown, M.W., & Bowden, N.A. (2019). BCL-2 family isoforms in apoptosis and cancer. Cell Death Dis. 2019; 10(3):177. doi: 10.1038/s41419-019-1407-6

Mendiratta, G., Ke, E., Aziz, M., Liarakos, D., Tong, M. & Stites, E.C. (2021). Cancer gene mutation frequencies for the U.S. population. Nat Commun. 12, 5961 doi：10.1038/s41467-021-26213-y.

Liu, X., Gan, J., Du, S., Zhu, C.,…, Cai Q. (2021). Proteomic profiling identifies Kaposi's sarcoma-associated herpesvirus(KSHV)-encoded LANA SIM-associated proteins in hypoxia. mSystems. 6:e01109-21.https://doi.org/10.1128/mSystems.01109-21.

Yan, H.，Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., … Li W.. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. Elife. 1:e00049. doi: 10.7554/eLife.00049.

Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes W.S., Borodovsky M., …Koonin EV. (1996). Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. Curr. Biol. 6 (3), 279-291. doi: 10.1016/s0960-9822(02)00478-5.

Bussey, H., Kaback, D.B., Zhong, W., Vo, D.T., Clark, M.W., Fortin, N., Hall J, …Barton AB. (1995). The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 92 (9), 3809-3813. doi: 10.1073/pnas.92.9.3809.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., … Oliver SG. (1996). Life with 6000 genes. Science 274 (5287), 546. doi: 10.1126/science.274.5287.546.

Dunham, I. (2005). Chromosome 22: Sequencing. John Wiley & Sons, Ltd.

Nurk. S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, X.V., Mikheenko, A., … Phillippy AM. (2022). The complete sequence of a human genome. Science. 376(6588); 44-53. DOI: 10.1126/science.abj6987. doi: 10.1126/science.abj6987.

Qiu, G.H., Tan, L.K., Loh, K.S., Lim, C.Y., Srivastava, G., Tsai, S.T., …Tao Q. (2004). The candidate tumor suppressor gene BLU, located at the commonly deleted region 3p21.3, is an E2F-regulated, stress-responsive gene and inactivated by both epigenetic and genetic mechanisms in nasopharyngeal carcinoma. Oncogene. 23: 4793–4806. doi: 10.1038/sj.onc.1207632.

Yi Lo, P.H., Chung Leung, A.C., Xiong, W., Law, S., Duh, F.M., Lerman, M.I., …, Lung, M.L. (2006). Expression of candidate chromosome 3p21.3 tumor suppressor genes and down-regulation of BLU in some esophageal squamous cell carcinomas. Cancer Lett. 234:184–192. doi: 10.1016/j.canlet.2005.03.036.

Tischoff, I., Markwarth, A., Witzigmann, H., Uhlmann, D., Hauss, J., Mirmohammadsadegh, A., … Tannapfel A. (2005). Allele loss and epigenetic inactivation of 3p21.3 in malignant liver tumors. Int J Cancer. 115： 684 – 689. doi: 10.1002/ijc.20944.

Wu, L.J., Zhang, X.N., Wang, J., Kong, X., Zheng, B., & He, Z. (2021). ZMYND10 downregulates cyclins B1 and D1 to arrest cell cycle by trimethylating lysine 9 on histone 3. Life Res. 4(4): 32. doi:10.53388/life2021-0727-132

Zhang, X., Lerman, M.I., & He, Z. (2015). ZMYND10 (zinc finger, MYND-type containing 10). Atlas Genet Cytogenet Oncol Haematol. 19: 319-323.