Proceedings of the 2021 International Conference on Public Art and
Human Development (ICPAHD 2021)

What Contribution Does Bioinformatic Analysis Perform in the Advancement of Conventional Biochemical Research

Ziran Yin[†]

McMaster University
Hamilton Ontario Canada
[†] Corresponding author: yinz33@mcmaster.ca

ABSTRACT

The precision and expense of biological research have grown in inverse proportion during decades of ongoing bioinformatics technology development, with steadily increasing accuracy and gradually decreasing spending. In the meantime, the demand for data analysis has skyrocketed for questions that have remained unanswered since the relatively mature biological sector. During the 1990s and 2000s, considerable advances in sequencing technology, combined with lower costs, resulted in an exponential increase in data analysis demand. The emergence of 'Big Data' has posed new data mining and administration issues, necessitating the incorporation of greater computer science knowledge. Traditional experimental approaches have started to demonstrate some flaws, such as inefficiency and inaccuracy. Combining earlier experimental methodologies with modern bioinformatics data analysis approaches could be a great place to meet today's more demanding data analysis demands. However, numerous opportunities for boosting efficacy using bioinformatic data analytic applications are still being summarised. This review will summarize the current application of the latest bioinformatics technology in biological research disciplines such as biomedicine, medicine, and public health. By evaluating significant biology data with current-updating software, it could better comprehend how gene sequencing, medicine, and diagnostic screening method exploration, quick coping, and update in public health study might be better achieved.

Keywords: *Bioinformatics technology, DNA/RNA sequencing, Cancer screening, Novel disease detection*

1. INTRODUCTION

Bioinformatics began more than 50 years ago, at a time when desktop computers were still a theory and DNA could not yet be sequenced. The use of computational approaches to protein sequence analysis created the groundwork for bioinformatics in the early 1960s. DNA analysis emerged later due to parallel[l] advances in molecular biology methods, which made DNA manipulation and sequencing easier, and computer science, which saw the rise of increasingly miniaturized and powerful computers and novel software better suited to handle bioinformatics tasks1. The information for terrestrial life's hereditary and metabolic features is eventually encoded in the order of nucleic acids in polynucleotide chains. As a result, the biological study requires the capacity to quantify or infer such sequences. The data explosion in the 1960s aided the development of bioinformatic abilities[2]. The emergence of 'Big Data' has posed new issues regarding data mining and management, necessitating greater computer science knowledge.

Table 1. Huge spending on gene sequencing in 2011[3]

		Whole genome sequencing				**RNA-Seq**		
		2011 cost	**output**	**2011 time**		**2011 cost**	**output**	**2011 time**
Sample collection and experimental design	from blood samples (easy to collect) to brain tissue (hard to collect)	~$100 onwards		from a few hours to several days	same as for whole genome sequencing			
Sequencing	library preparation + running the sequencer (whole dual flow cell)	~$6500 = ~$500 + ~$6000	~380 M reads/lane; 1 individual: ~1140 M total reads (~3 lanes for a 30 × coverage); ~250 Gb (intermediate files)	~11-12 day	library preparation + running the sequencer (whole dual flow cell)	~$3300 = ~$300 + ~$3000	~380 M reads/lane	~12-14 day
	Data storage, low-level processing							
	Alignment (transfer* and storing raw data + mapping)	~$40 = ~$33 + ~$7	300 Gb (BAM file)	~1/2 day *** (including transferring 250 Gb FASTQ ~7.5 hrs)	Alignment (transfer* and storing raw data + mapping)	~$5 = ~$3 + ~$2	~30 Gb (BAM); ~22 Gb (MRF)	< 2 hrs ***
	(data transfer and storage for 10 days)*; **	~$40		~8.5 hrs	(data transfer and storage for 10 days)*; **	< $4		< 1 hr
Data reduction and management	***High-level summaries*** ***							
	SNP calling (compute + transfer out)	< $5 = ~$4 + ~$0.60	< 1 Gb	~3 hrs	Gene and exon expression quantification	< $1	< 1 M	< 1 hr (1 CPU)
	Indel calling (compute + transfer out)	< $35 = ~$32 + ~$0.60	< 1 Gb	~1 day	Isoform quantification	~$6	< 1 M	~4 h
	SV calling (compute + transfer out)	< $35 = ~$32 + ~$0.60	< 1 Gb	~1 day				
Downstream analyses		> $100 K	~310 Gb	months		> $100 K	~30 Gb	months
Total of sequencing, data management and reduction		**~$6500**	**~310 Gb**	**~15 days**		**~3500**	**~30 Gb**	**~12-14 days**

Over the last fifty years, a large number of researchers have worked to develop techniques and technologies that will make this task easier, such as sequencing DNA and RNA molecules, moving from sequencing short oligonucleotides to millions of bases, and moving from attempting to deduce the coding sequence of a single gene to rapid and widely available whole-genome sequencing[4]. The "next-generation" technologies of cyclic-array, hybridization-based, nanopore, and single-molecule sequencing have progressed from labour-intensive slab gel electrophoresis to automated multiCE systems employing fluorescence labelling with multispectral imaging[5] and computational sequencing methods like Hifi or Rsubreads. Biological Big Data had and continues to have substantial implications on the predictive capacity and repeatability of bioinformatics results in a range of domains, especially when combined with an ever-increasing number of bioinformatics tools1, including epidemiology in exploring medicine and new disease screening[6] and public health in detecting viral genomes[7]. This review will focus on various relevant fields, including biology, medicine, and public health. It will be demonstrated the irreplaceable importance of massive amounts of data analysis to the modern biological and medical fields and the methodologies and research that bioinformatics research could offer support and promotion in approaching heavy use of data analytics in those particular fields.

2. ENHANCE IN DNA/RNA SEQUENCING

Bioinformatics approaches can assist in analyzing large amounts of data using code or software for data sets that would be time-consuming and hard to evaluate using standard experimental methods.

2.1. Defects in the traditional sequencing method

A DNA sequence length of 10-25 kb, which could not be sequenced effectively until the 1950s, can now be read with accuracies better than 99.5 percent utilising the newest bioinformatic sequencing technology[8]. Traditionally, researchers have concentrated their efforts on sequencing the most abundant populations of relatively pure RNA species, such as microbial ribosomal or transfer RNA, or the genomes of single-stranded RNA bacteriophages[4]. Those simplest sequences satisfy the first-generation sequence constraint in that they are not only easily bulk-produced in culture, but they are also not complicated by a complementary strand and are frequently much shorter than eukaryotic DNA molecules[9]. Fred Sanger and colleagues implemented a similar technique based on detecting radiolabelled partial-digestion fragments after two-dimensional fractionation in 1967, which allowed researchers to steadily add to the growing pool of ribosomal and transfer RNA sequences by filling in the ends with radioactive dideoxy nucleotides[10], deriving sequence by feeding each nucleotide one at a time and measuring incorporation[11]. However, because of the impure material, the outcomes of this basic sequencing techniques might easily be muddled. My past year's microbial resistance study's sanger sequencing picture displayed mixed signals initially. Base determination was still limited to small segments of DNA, and it still required a significant amount of analytical chemistry and fractionation techniques4. Parallelising bead-based emPCR and flowcell-binding sequencing were designed during second-generation DNA sequencing[12,13]. These techniques have substantially reduced the cost and complexity of DNA sequencing. In recent years, the Illumina sequencing platform has been credited with making the most significant contribution to the second generation of DNA sequencers[14]. These precise short reads are ideal for calling single-nucleotide variations (SNVs) and small insertions and deletions (indels). However, they are less effective for de novo assembly, haplotype phasing, and structural variant detection, requiring information over larger sequence stretches[15]. While sequencing capabilities between 2004 and 2010 doubled every five months[16] and relatively low accuracy in long-red sequencing[17], emphasizing merging conventional sequencing-method development with advancements in computer technology.

2.2. With the use of bioinformatics and computational approaches

Multiple pass circular consensus sequencing of prolonged (up to 25 kb) individual molecules generates very accurate long sequencing reads, increasing the value of noisy long-read sequencing (HiFi reads). This approach may provide long, high-fidelity readings with an average length of 13.5 kilobases (kb) that are highly accurate[15] (99.8%). HiFi sequencing uses circular consensus sequencing (CCS), which yields reliable reads from noisy individual subreads by constructing a consensus sequence from successive runs of a single template molecule[18]. Quality evaluation of CCS reads, tiny variant detection using CCS reads (50bps), and enhancing small variant detection with haplotype phasing are tasks that modern data analysis can accomplish that prior manual sequencing couldn't. Read-to-read error correction, a computationally costly technique that could only be done by computational data analysis, can yield high consensus accuracy[19]. The average read accuracy for sequencing the human HG002 genome to 28-fold coverage with an average read length of 13.5 kb was 99.8%, reaching the accuracy of short reads for minor variation identification while accessing more of the genome, especially in medically significant genes. CCS readings may also discover structural variants and assemble them from scratch, with equal contingency and significantly greater concordance than noisy long reads[15]. Many branches of science might benefit from this high-accuracy long-read sequencing, including chromosomal-level genome assembly, metagenome analysis, comprehensive variate identification, full-length transcripts and isoforms sequencing, and target sequencing of variants in specific regions[8]. Rsubread is a computational sequencing approach for aligning and quantifying RNA sequencing reads simpler, quicker, cheaper, and more accurate[20]. Subread is a Bioconductor package that implements current high-performance but computationally intensive RNA-seq read alignment and read counting algorithms as R algorithms, resulting in a matrix of read counts that could be inspected afterwards. It incorporates read mapping and quantification in a R package to reduce the computational cost of the most computationally expensive elements of an RNA-seq study, with mapping and counting both contributing for a significant portion of the total cost.

3. DISPROVE THE SEEMINGLY PROMISING CANCER SCREENING METHOD

To optimize the survival rate of suspected cancer patients with previously restricted data analysis tools, the major approaches to identify cancer include radiological examination, cytology of the sputum, radio-isotope lung scanning, and mediastinoscopy. Low-dose CT became popular as an alternative to standard-dose CT since it is less aggressive and produces less cell or DNA damage[21]. According to the findings of the National Lung Screening Trial, both policy and clinical decision-making concerning LDCT screening must take into account the potential advantages of screening (lower lung cancer mortality) as well as the potential hazards[22]. With a limited screening dosage, screening cannot yield the most accurate findings. Traditional screening procedures have a history of false-positive results, overdiagnosis, and unnecessarily intrusive testing. With modern bioinformatics applications, the software might collect a vast quantity of data from patients for further summarization and analysis. Based on efficient and reliable data analysis, common diagnostic challenges might be addressed. Physicians will have to decide in 2011 whether or not to start conducting low-dose computed tomography (LDCT) of the chest to screen for lung cancer in patients who have smoked in the past. The full effectiveness of routine cancer screening has yet to be shown because the advantage of low-dose frequent cancer screening was biochemically promising. The puzzle was knotty before bioinformatics analysis approaches were utilized. The LDCT group had a 20% decrease in lung cancer-specific mortality, according to the published NLST. In 2009, The PLCO (Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial) reported seven years of follow-up of over 76 000 randomly assigned participants to demonstrate the efficacy of early cancer screening, which would have been nearly impossible without the use of a bioinformatic analytical tool due to the large sample size. Despite the significant reduction in mortality found in the NLST trial cohort when utilizing LDCT, it is premature to suggest lung cancer screening in general practice. The lessons from prostate and breast cancer screening should warn us that the predicted reductions in fatalities are not as easy to accomplish as previously thought. Furthermore, the risks of false-positive chest computed tomography findings are extremely high based on current data. When the NLST approach is applied in non-specialty care settings and among the population at highest risk, namely those with smoking-related comorbid conditions, the morbidity and even mortality associated with invasive diagnostic testing and surgical resection due to false- and true-positive findings on computed tomography is likely to increase[23]. More accurate analysis models and a larger data sample should be researched in order to uncover improved cancer screening for more real-life situations.

4. NOVEL DISEASE DETECTION METHOD EXPLORATION

Bioinformatics also contributes to the medical area by supporting the development of novel diagnostic screening methods. Diseases were often diagnosed in the late twentieth and early twenty-first centuries based on well-understood biological mechanisms, making fresh diagnostic exploration difficult. For example, in a 1985 research by L Kovács, it was proposed to make a diabetes diagnosis by measuring plasma and urine osmolality since it may provide critical information on changes in water balance in diabetic patients[24] (diabetes insipidus and antidiuretic hormone excess). It was difficult for researchers to establish which biological metabolites for the patients to detect the illness without insufficient theoretical backing. The bioinformatic analysis could assist in finding the target metabolizing biomarker for such a condition in a reasonably quick and straightforward manner due to advancements in software and data processing tools.

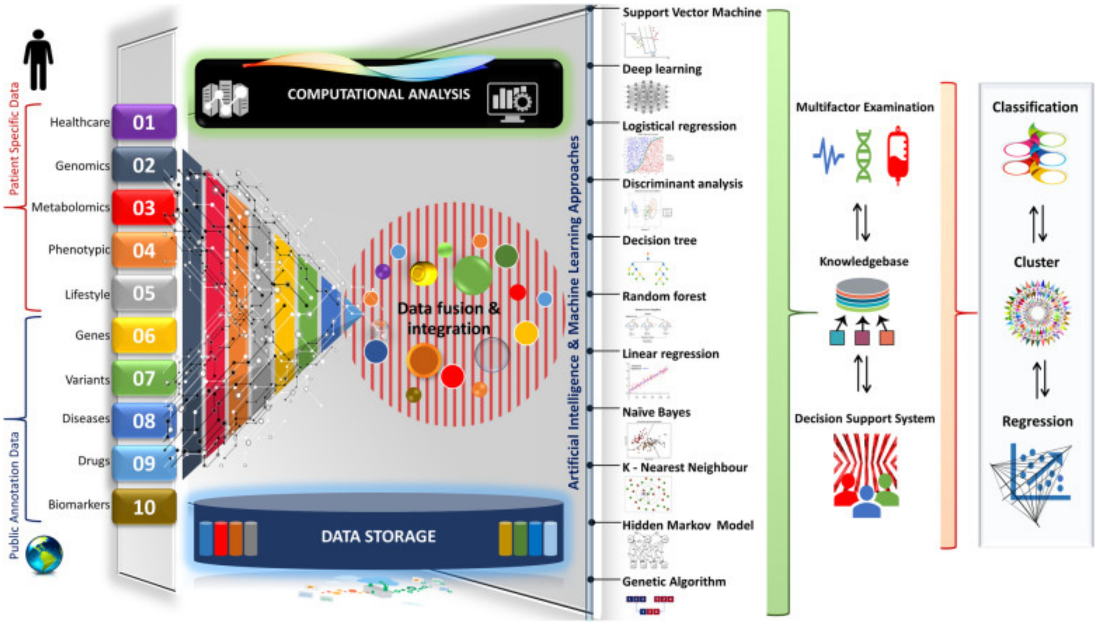

Figure 1 Modeling of patient-specific healthcare, genomics, metabolomics, phenotypic, and lifestyle data, as well as publically accessible annotated data such as genes, variations, illnesses, medicines, and biomarkers. AI and machine learning techniques are used in the analysis[25]

Instead of focusing on every aspect of a disease, researchers only need to uncover the data correlation between metabolite concentration and diagnostic status, together with a rudimentary comprehension of the disease's underlying principles. As an alternative to comprehending the extensive and complicated biological context, bioinformatic studies applied data analysis to prove the relationship between metabolites and illnesses. For example, a study published in 2020 by Musher stated that circulating tumour DNA was a promising indicator for screening recurrent colorectal cancer, with a series of significant improvements in saving patients' lives without knowing the biological reasoning for the emergence of higher ctDNA concentrations. By detecting circulating tumour DNA (ctDNA) concentration in patients' plasma samples, researchers were able to identify recurrences of early-stage colorectal cancer, which is an aggressive, likely widespread, and incurable malignancy. Researchers also assessed sensitivity for recurrence with well-performance confidence intervals and specificity of diagnosis using computational data analysis[26].

Furthermore, a panel of urine biomarkers was being developed to detect early stages of pancreatic cancer without knowing the precise raising the concentration of chosen waste biomarkers. Five hundred ninety urine samples were collected and analyzed for the study. The PancRISK score algorithm was introduced to establish that the metabolizing waste LYVE1, REG1A, REG1B, and TFF1 had a high chance of being used as a potential screening indicator for early stages of pancreatic cancer[27] (AUC = 0.992 (95 percent CI 0.983-1.000).

5. CONTRIBUTION IN RELIEVING AND CONTROLLING OF PANDEMIC

The bioinformatic analysis makes a significant contribution to the public health service since it can swiftly evaluate vast volumes of detailed data and assist in the search for links, patterns, or trends. These advantages might significantly impact the progress of a pandemic, allowing researchers to create epidemic preventive strategies and test the efficacy of new treatments more rapidly. When SARS-COVID epidemics occurred in China in 2003, the bioinformatics application and public health data analysis methodologies were still in the early stages of development, making it impossible to evaluate the infection data at the time. Only a few studies or papers have looked into this virus in-depth to find an efficient technique to combat the infection's spread. The study that associated infection data interpretation is rudimentary, estimating that, on average, a single infection case can infect three secondary cases[28]. The pandemic's capacity to be controlled was hampered by a lack of knowledge, which increased mortality.

The COVID-19 outbreaks occurred in 2019, however, with several bioinformatics data analyses being employed in public health research. Sufficient data gathering and analysis resulted in detailed pandemic responses promptly. For example, in research conducted by Kuma Diriba, immunocompromised pregnant women were shown to have a faster worsening clinical course and a higher risk of injury to both the mother and the fetus when infected with the COVID-19 virus[29]. SARS-CoV-2 has been compared to SARS-CoV and influenza pandemics in transmissibility, hospitalization, and fatality rates[30]. Cleo Anastassopoulou's research examined the correlation between human genetic variables and vulnerability to SARS-CoV-2 infection and the severity of COVID-19 illness[31]. Megan O'Driscoll's survey suggests that age-specific mortality and immunity patterns of SARS-CoV-2 had a consistent pattern[32]. The current knowledge on the evolutionary and structural aspects of the COVID virus is summarised concisely to understand its mutational pattern and likely participation in the ongoing pandemic[33].

All these studies contributed significantly and showed how bioinformatics analysis may have saved the lives of millions of people during the outbreak.

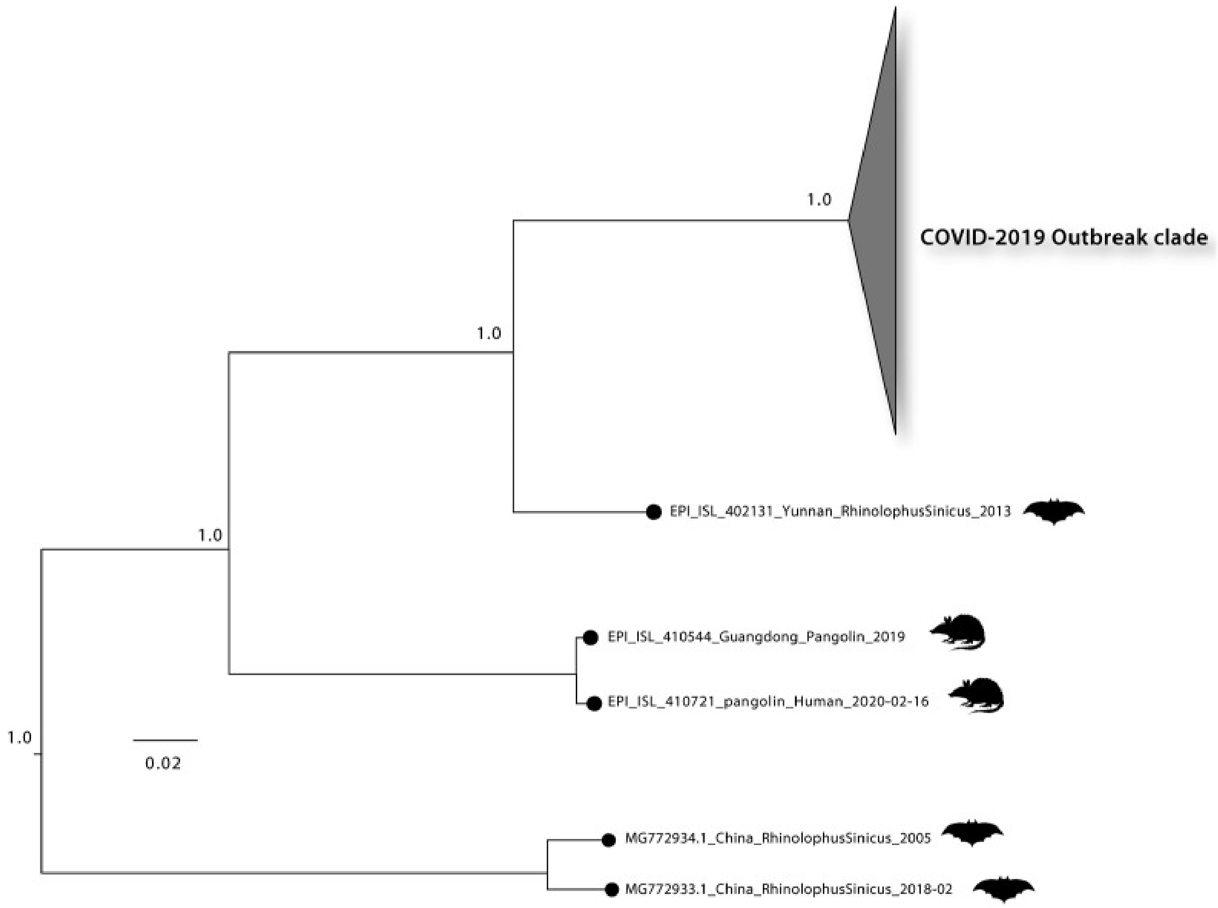
1.0
COVID-2019 Outbreak clade
1.0
EPI_ISL_402131_Yunnan_RhinolophusSinicus_2013
1.0
EPI_ISL_410544_Guangdong_Pangolin_2019
EPI_ISL_410721_pangolin_Human_2020-02-16
1.0
0.02
MG772934.1_China_RhinolophusSinicus_2005
MG772933.1_China_RhinolophusSinicus_2018-02

Figure 2 Maximum likelihood phylogeny was estimated with n = 143 full genome sequences from the current SARS-CoV-2 pandemic (2019–2020) plus n = 3 closely related bats[33]

6. CONCLUSION

With its unbeatable value in data processing and interpretation, bioinformatics research contributes to a wide range of disciplines of study, including biology, medicine, and public health. Bioinformatic tools will be developed in the future to solve traditional methods' deficiency, with fresh data sets and the ongoing development of new technologies allowing for the extension of biomedical, medical, and public health investigations, bringing personalized medicine. Despite its near-optimal success, the bioinformatic analysis still has its deficiency. Medical education and medical scholarship rely heavily on data collecting and analysis. Despite the technological ability to collect more data and medical education's reliance on data, it is becoming more difficult for medical schools and educational scholars to collect and use data, particularly in terms of regulations, security concerns, and the growing reluctance of learners and others to participate in data collection activities. The unstable link between these fields and bioinformatics and the lack of analytical precision has all contributed to the developing problem. As a result, a system-wide corrective is becoming increasingly necessary: a shift in practice that makes data usage more practical and productive while preserving high professional standards. Greater clarity of usage of data; greater clarity on what constitutes "good" data; changes to how data is collected; better strategic stewardship of existing data; and deliberate and strategic attention to "data readiness" in support of medical education and medical education scholarship are five core areas that can contribute to a system-wide correction. In the face of primarily regulatory obstacles, these solutions are essentially practical and conceptual improvements. Medical educators must also connect with emerging practice areas such as learning analytics and examine the altering social contract for data use in medical education.

REFERENCES

[1] Gauthier, J.; Vincent, A. T.; Charette, S. J.; Derome, N. A Brief History of Bioinformatics. Brief Bioinform 2019, 20 (6), 1981–1996. https://doi.org/10.1093/bib/bby063.

[2] Akalin, P. K. Introduction to Bioinformatics. Mol Nutr Food Res 2006, 50 (7), 610–619. https://doi.org/10.1002/mnfr.200500273.

[3) Sboner, A.; Mu, X. J.; Greenbaum, D.; Auerbach, R. K.; Gerstein, M. B. The Real Cost of Sequencing: Higher than You Think! Genome Biol 2011, 12 (8), 125. https://doi.org/10.1186/gb-2011-12-8-125.

[4] Heather, J. M.; Chain, B. The Sequence of Sequencers: The History of Sequencing DNA. Genomics 2016, 107 (1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.

[5] Karger, B. L.; Guttman, A. DNA Sequencing by Capillary Electrophoresis. Electrophoresis 2009, 30 (Suppl 1), S196–S202. https://doi.org/10.1002/elps.200900218.

[6] Seow, W. J.; Shu, X.-O.; Nicholson, J. K.; Holmes, E.; Walker, D. I.; Hu, W.; Cai, Q.; Gao, Y.-T.; Xiang, Y.-B.; Moore, S. C.; Bassig, B. A.; Wong, J. Y. Y.; Zhang, J.; Ji, B.-T.; Boulangé, C. L.; Kaluarachchi, M.; Wijeyesekera, A.; Zheng, W.; Elliott, P.; Rothman, N.; Lan, Q. Association of Untargeted Urinary Metabolomics and Lung Cancer Risk Among Never-Smoking Women in China. JAMA Netw Open 2019, 2 (9), e1911970. https://doi.org/10.1001/jamanetworkopen.2019.11970.

[7] Escher, F.; Pietsch, H.; Aleshcheva, G.; Bock, T.; Baumeier, C.; Elsaesser, A.; Wenzel, P.; Hamm, C.; Westenfeld, R.; Schultheiss, M.; Gross, U.; Morawietz, L.; Schultheiss, H.-P. Detection of Viral SARS-CoV-2 Genomes and Histopathological Changes in Endomyocardial Biopsies. ESC Heart Fail 2020, 7 (5), 2440–2447. https://doi.org/10.1002/ehf2.12805.

[8] Hon, T.; Mars, K.; Young, G.; Tsai, Y.-C.; Karalius, J. W.; Landolin, J. M.; Maurer, N.; Kudrna, D.; Hardigan, M. A.; Steiner, C. C.; Knapp, S. J.; Ware, D.; Shapiro, B.; Peluso, P.; Rank, D. R. Highly Accurate Long-Read HiFi Sequencing Data for Five Complex Genomes. Sci Data 2020, 7, 399. https://doi.org/10.1038/s41597-020-00743-4.

[9] Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA Sequencing at 40: Past, Present and Future. Nature 2017, 550 (7676), 345–353. https://doi.org/10.1038/nature24286.

[10] Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. Proc Natl Acad Sci U S A 1977, 74 (12), 5463–5467.

[11] Brownlee, G. G.; Sanger, F. Nucleotide Sequences from the Low Molecular Weight Ribosomal RNA of Escherichia Coli. J Mol Biol 1967, 23 (3), 337–353. https://doi.org/10.1016/s0022-2836(67)80109-8.

[12] Shendure, J.; Ji, H. Next-Generation DNA Sequencing. Nat Biotechnol 2008, 26 (10), 1135–1145. https://doi.org/10.1038/nbt1486.

[13] Fedurco, M.; Romieu, A.; Williams, S.; Lawrence, I.; Turcatti, G. BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies. Nucleic Acids Res 2006, 34 (3), e22. https://doi.org/10.1093/nar/gnj023.

[14] Greenleaf, W. J.; Sidow, A. The Future of Sequencing: Convergence of Intelligent Design and Market Darwinism. Genome Biol 2014, 15 (3), 303. https://doi.org/10.1186/gb4168.

[15] Wenger, A. M.; Peluso, P.; Rowell, W. J.; Chang, P.-C.; Hall, R. J.; Concepcion, G. T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N. D.; Töpfer, A.; Alonge, M.; Mahmoud, M.; Qian, Y.; Chin, C.-S.; Phillippy, A. M.; Schatz, M. C.; Myers, G.; DePristo, M. A.; Ruan, J.; Marschall, T.; Sedlazeck, F. J.; Zook, J. M.; Li, H.; Koren, S.; Carroll, A.; Rank, D. R.; Hunkapiller, M. W. Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome. Nat Biotechnol 2019, 37 (10), 1155–1162. https://doi.org/10.1038/s41587-019-0217-9.

[16] Stein, L. D. The Case for Cloud Computing in Genome Informatics. Genome Biol 2010, 11 (5), 207. https://doi.org/10.1186/gb-2010-11-5-207.

[17] Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; Dewinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-Time DNA Sequencing from Single Polymerase Molecules. Science 2009, 323 (5910), 133–138. https://doi.org/10.1126/science.1162986.

[18] Travers, K. J.; Chin, C.-S.; Rank, D. R.; Eid, J. S.; Turner, S. W. A Flexible and Efficient Template Format for Circular Consensus Sequencing and SNP Detection. Nucleic Acids Res 2010, 38 (15), e159. https://doi.org/10.1093/nar/gkq543.

[19] Jain, M.; Koren, S.; Miga, K. H.; Quick, J.; Rand, A. C.; Sasani, T. A.; Tyson, J. R.; Beggs, A. D.; Dilthey, A. T.; Fiddes, I. T.; Malla, S.; Marriott, H.; Nieto, T.; O'Grady, J.; Olsen, H. E.; Pedersen, B. S.; Rhie, A.; Richardson, H.; Quinlan, A. R.; Snutch, T. P.; Tee, L.; Paten, B.; Phillippy, A. M.; Simpson, J. T.; Loman, N. J.; Loose, M. Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long

Reads. Nat Biotechnol 2018, 36 (4), 338–345. https://doi.org/10.1038/nbt.4060.

[20] Liao, Y.; Smyth, G. K.; Shi, W. The R Package Rsubread Is Easier, Faster, Cheaper and Better for Alignment and Quantification of RNA Sequencing Reads. Nucleic Acids Res 2019, 47 (8), e47. https://doi.org/10.1093/nar/gkz114.

[21] Gholizadeh-Ansari, M.; Alirezaie, J.; Babyn, P. Low-Dose CT Denoising Using Edge Detection Layer and Perceptual Loss. Annu Int Conf IEEE Eng Med Biol Soc 2019, 2019, 6247–6250. https://doi.org/10.1109/EMBC.2019.8857940.

[22] Tanoue, L. T.; Tanner, N. T.; Gould, M. K.; Silvestri, G. A. Lung Cancer Screening. Am J Respir Crit Care Med 2015, 191 (1), 19–33. https://doi.org/10.1164/rccm.201410-1777CI.

[23] Silvestri, G. A. Screening for Lung Cancer: It Works, but Does It Really Work? Ann Intern Med 2011, 155 (8), 537–539. https://doi.org/10.7326/0003-4819-155-8-201110180-00364.

[24] Kovács, L.; Némethova, V.; Gucalová, Y.; Lehotská, V.; Cintala, J.; Michajlovskij, N.; Michalicková, J.; Lichardus, B. Simple Diagnosis of Diabetes Insipidus and Antidiuretic Hormone Excess. Exp Clin Endocrinol 1985, 85 (2), 228–234. https://doi.org/10.1055/s-0029-1210441.

[25] Ahmed, Z. Practicing Precision Medicine with Intelligently Integrative Clinical and Multi-Omics Data Analysis. Hum Genomics 2020, 14, 35. https://doi.org/10.1186/s40246-020-00287-z.

[26] Musher, B. L.; Melson, J. E.; Amato, G.; Chan, D.; Hill, M.; Khan, I.; Kochuparambil, S. T.; Lyons, S. E.; Orsini, J.; Pedersen, S. K.; Robb, B.; Saltzman, J.; Silinsky, J.; Gaur, S.; Tuck, M. K.; LaPointe, L. C.; Young, G. P. Evaluation of Circulating Tumor DNA for Methylated BCAT1 and IKZF1 to Detect Recurrence of Stage II/Stage III Colorectal Cancer (CRC). Cancer Epidemiol Biomarkers Prev 2020, 29 (12), 2702–2709. https://doi.org/10.1158/1055-9965.EPI-20-0574.

[27] Debernardi, S.; O'Brien, H.; Algahmdi, A. S.; Malats, N.; Stewart, G. D.; Plješa-Ercegovac, M.; Costello, E.; Greenhalf, W.; Saad, A.; Roberts, R.; Ney, A.; Pereira, S. P.; Kocher, H. M.; Duffy, S.; Blyuss, O.; Crnogorac-Jurcevic, T. A Combination of Urinary Biomarker Panel and PancRISK Score for Earlier Detection of Pancreatic Cancer: A Case-Control Study. PLoS Med 2020, 17 (12), e1003489. https://doi.org/10.1371/journal.pmed.1003489.

[28] Lipsitch, M.; Cohen, T.; Cooper, B.; Robins, J. M.; Ma, S.; James, L.; Gopalakrishna, G.; Chew, S. K.; Tan, C. C.; Samore, M. H.; Fisman, D.; Murray, M. Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. Science 2003, 300 (5627), 1966–1970. https://doi.org/10.1126/science.1086616.

[29] Diriba, K.; Awulachew, E.; Getu, E. The Effect of Coronavirus Infection (SARS-CoV-2, MERS-CoV, and SARS-CoV) during Pregnancy and the Possibility of Vertical Maternal-Fetal Transmission: A Systematic Review and Meta-Analysis. Eur J Med Res 2020, 25 (1), 39. https://doi.org/10.1186/s40001-020-00439-w.

[30] Petersen, E.; Koopmans, M.; Go, U.; Hamer, D. H.; Petrosillo, N.; Castelli, F.; Storgaard, M.; Al Khalili, S.; Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and Influenza Pandemics. Lancet Infect Dis 2020, 20 (9), e238–e244. https://doi.org/10.1016/S1473-3099(20)30484-9.

[31] Anastassopoulou, C.; Gkizarioti, Z.; Patrinos, G. P.; Tsakris, A. Human Genetic Factors Associated with Susceptibility to SARS-CoV-2 Infection and COVID-19 Disease Severity. Hum Genomics 2020, 14 (1), 40. https://doi.org/10.1186/s40246-020-00290-4.

[32] O'Driscoll, M.; Ribeiro Dos Santos, G.; Wang, L.; Cummings, D. A. T.; Azman, A. S.; Paireau, J.; Fontanet, A.; Cauchemez, S.; Salje, H. Age-Specific Mortality and Immunity Patterns of SARS-CoV-2. Nature 2021, 590 (7844), 140–145. https://doi.org/10.1038/s41586-020-2918-0.

[33] Giovanetti, M.; Benedetti, F.; Campisi, G.; Ciccozzi, A.; Fabris, S.; Ceccarelli, G.; Tambone, V.; Caruso, A.; Angeletti, S.; Zella, D.; Ciccozzi, M. Evolution Patterns of SARS-CoV-2: Snapshot on Its Genome Variants. Biochem Biophys Res Commun 2021, 538, 88–91. https://doi.org/10.1016/j.bbrc.2020.10.102.