Artificial Intelligence in Clinical Genomics and Healthcare

[1]Alim Al Ayub Ahmed (alim@jju.edu.cn)
[1]School of Accounting, Jiujiang University, Jiujiang, Jiangxi, CHINA

[2]Praveen Kumar Donepudi
[2]Enterprise Architect, Information Technology, UST-Global, Inc., Ohio, USA

[3]ABM Asadullah
[3]Kulliyyah of Economics and Management Sciences, International Islamic University Malaysia (IIUM), MALAYSIA

Abstract: Genomics creates large databases for the discovery, study and production of new therapeutics worldwide. It would not be impossible to imagine that 3 billion base pairs comprising the humanoid genetic makeup may now be studied to find genetic differences within the population by artificial intelligence. Large pharmaceutical firms such as Astra Zeneca are aiming to research up to 2 million genomes by 2026 and review vast quantities of patient data points from their clinical drug trials. AI will be used in genomics for multiple omics experiments, such as transcriptomics, as we introduce more instruments. AI is increasingly being used by healthcare firms in accordance with HEOR (Health Economics Outcome Research), i.e. In order to help classify possible clinically important genes, AI is used to combine data produced from genomic studies with analysis from science literature. Machine learning today plays an integral role in the development of the genomics industry. In this paper, we set out to explore the uses of genomics machine learning to help market leaders consider existing and evolving developments in the field. We have discussed history terms and distilled perspectives from various study. Current applications of machine learning in gene technology boost up future applications of genomics machine learning.

Keywords: Machine learning, genomics, artificial intelligence, healthcare, gene technology

1. Introduction

Genomics is an interdisciplinary biology field which focuses on the study of genome structure ,function, mapping, and editing. A genome is a full collection of an organism's DNA; all of the genes are included. We can split genomics into several subsets i.e. genomics of control, genomics of structure, and genomics of function. Nearly every industry has been affected by artificial intelligence and machine learning. No exception is healthcare. Innovations have long been embraced by the industry, and now a rising number of researchers are turning their attention towards advancements in artificial intelligence. Genomics is one of these fields. In the evolution of this area, machine learning plays an increasingly important part. Researchers can examine the increasing amount of genomic imagery data by connecting deep learning with computer vision techniques. Machine learning models are able to solve tasks in computer vision, such as semantic segmentation, recognition of images and withdrawing of images (Rahman et al., 2020). It is possible to examine a vast volume of genomics-related text that can be found in publicly accessible research papers by integrating machine learning with natural processing techniques. Researchers may solve problems such as relationship extraction, retrieval of information, or identification of named individuals in this way. Due to the enormous amount of study carried out in this area at the moment, certain systems are ideally appropriate for working with natural language processing activities (Donepudi et al., 2020).

2. AI and Genomics Background and Insights Up front

DNA sequencing empowers specialists to peruse the hereditary arrangement which administers the exercises of every single living being. As the route from DNA to RNA to protein, the essential science doctrine is summed up to give history. DNA comprises of the basic matches A sets with T and C sets with G, in light of 4 crucial units (A, C, G and T) called nucleotides. A sum of 23 sets in people are split into chromosomes.

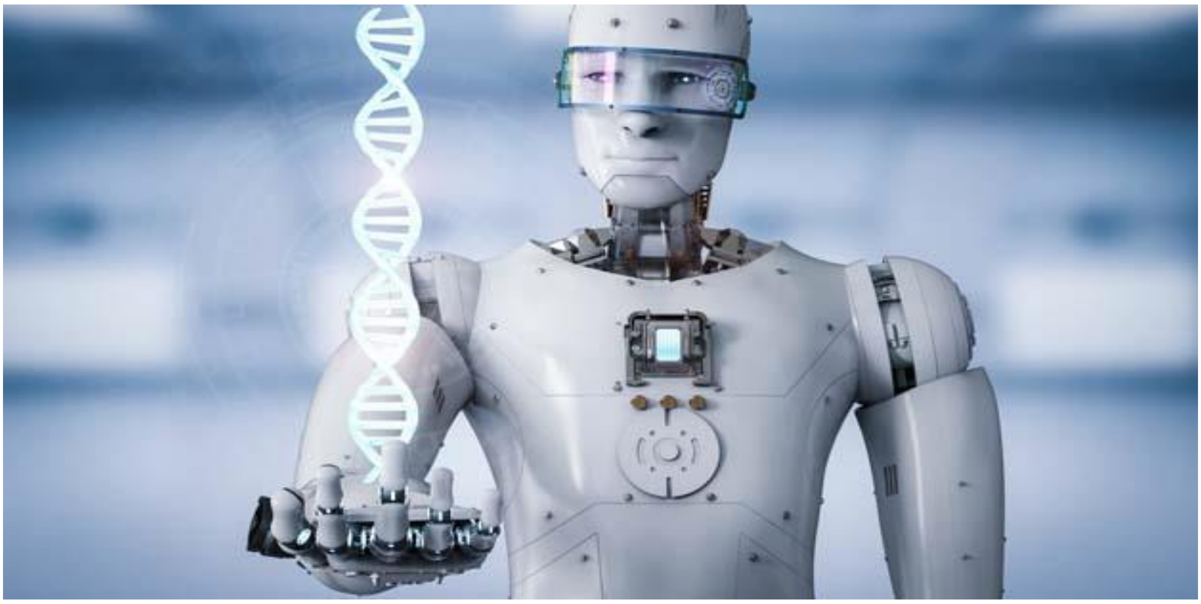

Figure 1: AI is next big player in genomics

Chromosomes are more coordinated into fragments of DNA called genes which produce or encode proteins. The genome is named the quantity of qualities that a creature conveys. Individuals have around 20,000 chromosomes and base sets of around 3,000 billion. Zero in on genomics examination and business, protein is encoded by only 2 percent of the human genome. It is a basic area. Accuracy is firmly connected with genomics. The region of Precision Medicine (otherwise called tweaked medication) is a way to deal with medical care that joins science, perspectives and the network with the goal of applying a patient or populace explicit clinical mediation, instead of a one-size-fits-all methodology, with a market size assessed to hit $87 billion by 2023. For instance, a man who needs a blood bonding will be matched to a benefactor who has a similar blood classification rather than a haphazardly picked contributor to limit the probability of entanglements. Actually, there are two major obstacles to greater precision medicine implementation: high prices and infrastructure restrictions. Many researchers are applying machine learning approaches to solve the large volume of patient data that must be gathered and processed, and to help reduce costs.

Fortunately, the expense of decoding a genome tends to decline year-over-year for researchers and genomics firms, even after a massive relative decrease in costs between 2007 and 2012.

3. Applications of AI and Machine Learning in Genomics

New machine learning technologies in the field of genomics have an effect on how genetic testing is done, how specialists offer clinical administrations to make genomics more open to people who are keen on studying how their heritage can affect their wellbeing.

Sequencing of Gene

The method of deciding the nucleic acid series, the order of nucleotides in DNA, is said to be DNA sequencing. It requires any procedure or technology used to establish the sequence of the four bases adenine, cytosine, guanine, and thymine. Entire Genome Sequencing (WGS) has arisen as a field of interest in clinical diagnostics. Cutting edge Sequencing has arisen as a trendy expression that includes innovative DNA sequencing strategies, helping researchers to grouping.

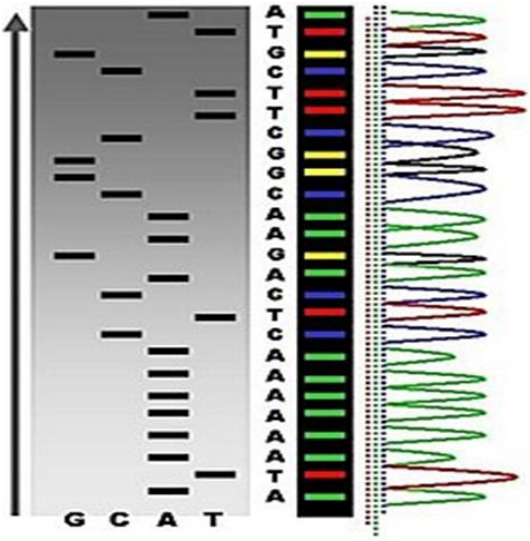

Figure 2: Genome sequencing

To help researchers interpret genetic variance, organisations such as Deep Genomics use machine learning. Specifically, algorithms are built based on patterns found in broad genetic data sets that are then converted into computer models to help clients understand how critical cellular processes are influenced by genetic variation. The metabolism, DNA repair, and development of cells are examples of cellular processes. Disruption of these pathways' natural functioning will theoretically induce diseases such as cancer.

The Toronto-based startup, which was founded in 2014, has raised a combined $3.7 million in seed financing from three U.S. venture capital companies. In fact, the supporters of Deep Genomics reportedly advised the business to continue to expand in Toronto instead of going to Silicon Valley. The decision will reflect the recent allocation of $125 million (Canadian dollars) by the Canadian government to the Pan-Canadian Strategy for Artificial Intelligence. As of April 2017, seven publications concerning its science have been cited by Deep Genomics, most of which forecast or suggest possible genetic variants. Relevant findings of this study, though, are yet to be published within the sense of diseases or possible therapies

Editing of Gene

The technique in which the slight and precise changes are made at cellular levels is called gene editing. The instrument which is responsible for editing of genome is said to be CRISPR. It does the editing in quick way with least expenses. The investigators should choose a suitable goal sequence so as to apply CRISPR. This can be a daunting system involving multiple decisions and unforeseeable effects. Machine learning provides the potential to greatly minimise the time ,expense and effort taken to define a reasonable sequence of goals.

Figure 3: CRISPR gene editing

Desktop Genetics, based in London, is a tech firm where AI and CRISPR intersect. Formed in 2012, 7 investors, representing a combination of accelerators, venture capital companies, and biotech business and DNA sequencing veteran Illumina, have raised $5.8 million in overall equity investment.

Two key results from a recent study are stated by the company i.e. an increased volume of training data increases the precision of an algorithm in its ability to predict CRISPR behavior, and when applied to a particular animal, the accuracy of the model declines, such as humans vs. mice. None of these results was especially shocking, and Desktop Genetics recognizes that extensive analysis would be required to continue refining processes and push the limits of how CRISPRR can affect machine learning.

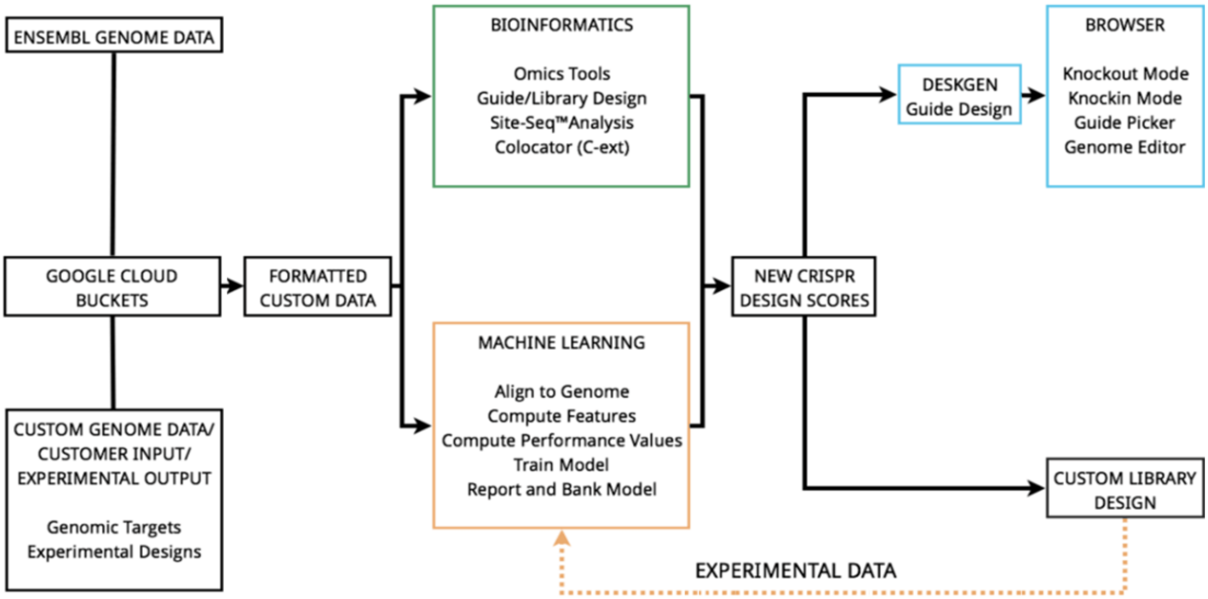

Figure 4: This leads new CRISPR designs which can then be tested in the lab, generating FASTQ data which once again feeds back into the workflow

Pharmacogenomics

The process is about how genes control the reaction of a person to drugs. The relatively recent discipline integrates pharmacology (drug science) and genomics (genetic testing and its functions) to produce reliable and safe medicines and dosages customized for the genetic make-up of an individual. There is evidence of studies concerning machine learning, though the field is still very recent. In February 2017, for instance, what is considered the first study to apply machine learning models to evaluate a safe dose of Tacrolimus in patients with renal transplantation was released. In order to avoid acute rejection of new organs, tacrolimus is usually given to patients after strong organ transplantation.

Screening Methods for Newborns

Over the next decade, experts expect that infant genetic screening will become common practice. Information acquired upon entering the world would be effectively fused into the EHR of individuals, and ladies during breastfeeding would approach non-obtrusive screening capacities for genuine problems, for example, Down Syndrome. AI was presented by the infant Screening Center at the National Taiwan University Hospital to improve the exactness of its online infant metabolic deformity screening framework.. Genetic screening of newborns is now an increasingly popular procedure. Diseases like Down syndrome during birth can be diagnosed by this non-invasive genetic screening. Based on available evidence, artificial intelligence can forecast results and the risks associated with curing genetic diseases.

For agriculture

An new field of concern and hope within the agricultural sphere is the potential for genomics to help boost soil quality and crop yield. Illumina has lent funding to California-based entrepreneurs through its Illumina Accelerator. The start-up is described as a combination of genomics and machine learning to create diagnostic tools for crop disease prediction and prevention.

The business is now known as Trace Genomics and seems to have turned its focus more to soil health. It may allow farmers to better predict and maximize yields if genetic data can be used to predict crop yield or health (and the resulting effects on soil). The global increases in crop yields that have resulted from past genetic alterations could also increase those developments used on a scale.

4. AI in Clinical Genomics

Emulating human insight is the reason for AI algorithm (Donepudi, 2017). However, when moved toward utilizing conventional numerical techniques, AI usage in medical genomics like to target undertakings that are wasteful to execute with human knowledge and powerless against blunder. A large number of the above strategies have been changed to determine the various advances associated with clinical genomic research, including variant calling, genome explanation, variation marking, and correspondence from aggregate to-genotype, and perhaps they may likewise be utilized for forecasts of genotype-to-aggregate at long last. Here, we recognize the vital classifications of issues examined in clinical genomics by AI.

Variant calling

The clinical understanding of genomes is powerless against the recognition, including serious explicitness of individual hereditary varieties inside the large numbers populating every genome. Precise mistakes related with the nuances of test preparing, sequencing innovation, grouping foundation, and the regularly unforeseen impact of science, for example, physical mosaicism are defenceless to typical variation calling apparatuses (Li, 2014). To determine these issues, a blend of measurable methodologies with hand-created qualities, for example, strand-predisposition or populace level conditions was utilized, bringing about high exactness yet slanted mistakes (DePristo et al., 2011). AI calculations may get these inclinations from a solitary genome with a perceived highest quality level of reference variant calls and produce unrivalled variant calls. A CNN-put together variant caller legitimately prepared with respect to peruse arrangements with no serious information on genomics or sequencing stages, deep Variant has as of late been appeared to outperform the benchmark. (Poplin et al., 2018). It is expected that the expanded exactness is because of the capacity of CNNs to perceive dynamic conditions in sequencing information. In addition, ongoing discoveries show that profound learning can reform straightforward calling (and, as a result, variation distinguishing proof) for nanopore-based sequencing advancements that have customarily attempted to contend with demonstrated sequencing innovation because of the blunder inclined existence of prior base-calling algorithms. (Wick et al., 2019).

Genome explanation and variant order

The investigation of humanoid genome results after variant calling depends on the distinguishing proof by past information on specific hereditary variations and the suspicion of the impact on hereditary variation practical genomic components. Man-made intelligence calculations can energize the utilization of earlier information by giving phenotype to-genotype planning. Here, the same number of the AI calculations used to anticipate the presence of a utilitarian component from essential DNA grouping information are likewise used to foresee the impact of a hereditary minor departure from such useful components, both genome explanation and variation arrangement are set up.

Classification of coding variants

A few of strategies have been generated to group the nonsynonymous variations (Tang & Thomas, 2016). Meta indicators (models that loop and aggregate the expectations generated by a few different indicators) that are focused on deep learning have been coupled with either of these techniques., when incorporated utilizing relapse or other AI draws near, beat both their individual prescient segments and the mix of those prescient parts (Kircher et al., 2014). For example, in an AI algorithm, the combined annotation based depletion techniques (CADD) join various prescient qualities to foresee the perniciousness of hereditary variations. Utilizing similar arrangement of info includes as CADD yet joined in a profound neural organization, a profound learning-based expansion of CADD, called DANN, exhibited better execution (Quang et al., 2015). This specialized expansion of CADD shows that profound learning can be an excellent technique for combining established characteristics that forecast deleteriousness.

Classification of non-coding variants

An open challenge in human genomics is the computational detection and prediction of noncoding pathogenic variation (Chatterjee & Ahituv, 2017). Latest results suggest that our ability to interpret non-coding genetic variation would be significantly enhanced by AI algorithms. At least 10 percent of unusual pathogenic genetic mutation is responsible for splicing defects in genes, however, because of the complexity of enhancers, silencer, isolators and other combinatorial and long-range DNA interactions, which influence the splicing of the genes,, they can be difficult to classify (Soemedi et al., 2017).

Phenotype-to-genotype mapping

A person genetic makeup possesses different hereditary varieties, autonomous of individual wellbeing status, which are either recently distinguished as pathogenic or expected to be pathogenic (Telenti et al., 2016). Along these lines, the recognition of pathogenic fluctuations plus the assurance of the communication between the aggregate of the ailing creature and those anticipated to happen from every up-and-comer pathogenic variation are likewise important for sub-atomic analysis of the illness. Computer based intelligence algorithms, particularly via the withdrawal of more significant level symptomatic rules which are installed in clinical pictures and EHRs, may enormously improve the mapping of phenotype to genotype.

Genotype-to-phenotype prediction

Ultimately, genetics' therapeutic aim is to include future disease risk diagnoses and projections. Relatively easy statistical approaches to the prediction of polygenic risk allow risk stratification for some common complex diseases, both personally and clinically useful (Torkamani, et al., 2018). A few analyses have endeavored to gnomically show unpredictable humanoid attributes utilizing AI algorithms, anyway most extreme of those recorded to date in the writing are probably going to be over fit in light of the fact that they supposedly depict extensively more variety in qualities than ought to be practical based on heritability gauges. One use of AI to genomic stature expectation has had the option to give sensibly solid estimates inside anticipated cutoff points, demonstrating that computational methods can be improved utilizing AI-based. The genuine advantage of AI-based ways to deal with genotype-to-phenotype expectation, be that as it may, is probably going to originate from consolidating various types of wellbeing information and danger factors into powerful infection hazard indicators (Lello et al., 2018).

Image to genetic diagnosis

There are 4526 disorders and 2142 mutations associated with these anomalies (Köhler et al., 2019). A dysmorphologist much of the time orders these abnormalities separately and sums up them into an expert assessment. Clinical determination would then be able to educate particular quality sequencing and aggregate educated examination regarding more explicit hereditary information. Clinical finding and atomic analysis given by individuals frequently cover, yet they don't coordinate impeccably in view of the phenotypic varieties between hereditarily various conditions. Profound Gestalt, a calculation for the CNN exploration of the facial picture, is unmistakably more than human dysmorphologists in this position and is fittingly explicit to separate atomic diagnoses mapped to the similar Noonan syndrome (Gurovich et al., 2019). PEDIA, a genome analysis method integrating Deep Gestalt, was able to use phenotypic characteristics derived from facial images when paired with genomic data to reliably priorities candidate pathogenic variants across 679 individuals for 105 distinct monogenic disorders (Hsieh et al., 2019). Deployment of Deep Gestalt as a face-scanning app has the power to both democratize and revolutionize genetic syndrome recognition (Dolgin, 2019).

Hereditary syndromes found by facial assessment can be effectively checked by DNA screening, however, now and again of malignancy, there is generally lacking material for substantial change testing. Notwithstanding, it is essential to perceive the hereditary establishments of a tumor while getting ready treatment. Once more, AI will conquer the hole between aggregates created from photos and their conceivable hereditary starting point.. A survival CNN was able to obtain an understanding of the histological features of brain tumors, linked to the survival position, which is a CNN hybrid, with Cox's proportional risk-driven findings (a form of predictive survival analysis). This algorithm was not enough intended to expressly anticipate genomic deviations. The investigation of the CNN ideas to make endurance expectations distinguished new histological attributes that are significant for choosing visualization. These findings, including the expressions of those within a genetically overlying phenotypical syndrome, demonstrate that photos of the historical tumor which specifically forecast genomic aberrations underlying the tumor. More broadly, machine vision applications focused on AI seem to be able to predict genetic aberrations that are likely to exist in the genome of a person which are based on the composite phenotypes encoded in suitable medical photos (Mobadersany et al., 2018).

5. Conclusion

Machine learning in genomics already has an effect on several touch points, including how genetic testing is carried out, how physicians deliver medical care and genomics accessibility to people interested in learning more about how their heredity can influence their health. Smart business is an attempt to introduce AI to help speed up the journey from bench-to-bedside and make precision medicine more commonplace (readers will want to explore our recent article on the applications of machine learning in medicine and pharma) with a deeper interest in this topic (Donepudi, 2018). Such activities can also be beneficial for organizations capable of offering tangible and viable solutions to the problems facing precision medicine. Although there is much hope, it is still an difficult task to contend for precision medicine with many physicians searching for more clarification on therapeutic value and insurance providers not treating it as a need. Therefore, education and concise examples of the usefulness and importance of this technology would have to supplement the data interpretation capabilities available by machine learning. Pharmacogenomics is a core field of emerging machine learning technologies of genomics, but this is only one instance and there are diverse possible future applications. With restricted results evidence, however, time will tell which fields stand to reap the greatest value from investing in AI. As we believe that this will be an active arena with further machine learning applications in the near future, we will continue to closely track the area of genomics.

6. References

1. Chatterjee S., & Ahituv N. (2017). Gene Regulatory Elements, Major Drivers of Human Disease. *Annual Review of Genomics and Human Genetics, 18*, 45-63. https://doi.org/10.1146/annurev-genom-091416-035537

2. DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., Daly M.J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806
3. Dolgin, E. (2019). AI face-scanning app spots signs of rare genetic disorders. *Nature*. https://doi.org/10.1038/d41586-019-00027-x
4. Donepudi, P. K. (2017). Machine Learning and Artificial Intelligence in Banking. Engineering International, 5(2), 83-86. https://doi.org/10.18034/ei.v5i2.490
5. Donepudi, P. K. (2018). Application of Artificial Intelligence in Automation Industry. Asian Journal of Applied Science and Engineering, 7(1), 7–20. http://doi.org/10.5281/zenodo.4146232
6. Donepudi, P. K., Ahmed, A. A. A., & Saha, S. (2020). Emerging Market Economy (EME) and Artificial Intelligence (AI): Consequences for the Future of Jobs. PalArch's Journal of Archaeology of Egypt / Egyptology, 17(6), 5562 - 5574. Retrieved from http://palarch.nl/index.php/jae/article/view/1829
7. Gurovich, Y., Hanani, Y., Bar, O. *et al.* (2019). Identifying facial phenotypes of genetic disorders using deep learning. *Nature Medicine*, *25*, 60-64. https://doi.org/10.1038/s41591-018-0279-0
8. Hsieh, T. C., Mensah, M. A., Pantel, J. T. *et al.* (2019). PEDIA: prioritization of exome data by image analysis. *Genetics in Medicine*, *21***,** 2807–2814 https://doi.org/10.1038/s41436-019-0566-2
9. Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, *46*(3), 310-315. https://doi.org/10.1038/ng.2892
10. Köhler, S., Carmody, L., Vasilevsky, N., et al. (2019). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.*, *47*(D1), D1018-D1027. https://doi.org/10.1093/nar/gky1105
11. Lello, L., Avery, S. G., Tellier, L., Vazquez, A. I., de Los Campos, G., Hsu, S. D. H. (2018). Accurate Genomic Prediction of Human Height. *Genetics*, *210*(2), 477-497. https://doi.org/10.1534/genetics.118.301267
12. Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples, Bioinformatics, *30*(20), 2843–2851, https://doi.org/10.1093/bioinformatics/btu356
13. Mobadersany, P., Yousefi S., Amgad, M., Gutman, D. A., Barnholtz-Sloan, J. S., Vega, J. E. V., Brat, D. J., Cooper, L. A. D. (2018). Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(13), E2970-E2979. https://doi.org/10.1073/pnas.1717139115
14. Poplin R., Chang P.C., Alexander D., Schwartz S., Colthurst T., Ku A., Newburger D., Dijamco J., Nguyen N., Afshar P.T., Gross S.S., Dorfman L., McLean C.Y., DePristo M.A. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology. 36(10), 983-987. https://doi.org/10.1038/nbt.4235
15. Quang, D., Chen, Y., Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, *31*(5), 761-763. https://doi.org/10.1093/bioinformatics/btu703
16. Rahman, M. M., Chowdhury, M. R. H. K., Islam, M. A., Tohfa, M. U., Kader, M. A. L., Ahmed, A. A. A., & Donepudi, P. K. (2020). Relationship between Socio-Demographic Characteristics and Job Satisfaction: Evidence from Private Bank Employees. American Journal of Trade and Policy,7(2), 65-72. https://doi.org/10.18034/ajtp.v7i2.492
17. Soemedi R., Cygan K.J., Rhine C.L., Wang J., Bulacan C., Yang J., Bayrak-Toydemir P., McDonald J., Fairbrother W.G. (2017). Pathogenic variants that alter protein code often disrupt splicing. Nature Genetics, 49(6), 848-855. https://doi.org/10.1038/ng.3837
18. Tang, H., & Thomas, P. D. (2016). Tools for Predicting the Functional Impact of Nonsynonymous Genetic Variation. *Genetics*, *203*(2), 635-647. https://doi.org/10.1534/genetics.116.190033

19. Telenti A., Pierce L.C., Biggs W.H., di Iulio J., Wong E.H., Fabani M.M., Kirkness E.F., Moustafa A., Shah N., Xie C., Brewerton S.C., Bulsara N., Garner C., Metzker G., Sandoval E., Perkins B.A., Och F.J., Turpaz Y., Venter J.C. (2016). Deep sequencing of 10,000 human genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(42), 11901-11906. https://doi.org/10.1073/pnas.1613365113
20. Torkamani A., Wineinger N. E., Topol E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews, Genetics*, *19*(9), 581-590. https://doi.org/10.1038/s41576-018-0018-x
21. Wick, R.R., Judd, L.M. & Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology* *20*, 129. https://doi.org/10.1186/s13059-019-1727-y