Vol 2:54-60

JEMI-PEARLS

Leveraging Cloud Computing for Genome Analysis to Provide Patient-Specific HIV Treatments and Insight Into HIV Susceptibility Patterns

Derek Zhang
Department of Microbiology and Immunology, University of British Columbia

BACKGROUND INFORMATION

Developed over 40 years ago, the introduction of Sanger sequencing by Frederick Sanger revolutionized the field of biology. The Sanger method quickly became the most widely used sequencing method for over three decades, with sequencing capacities hitting 150,000 bases per 24 hours per machine [1]. However, with the advent of "second generation sequencers", our ability to generate sequences exploded exponentially, with the latest Illumina HiSeq platform hitting throughputs of 600 gigabases per 24 hours [1]. It was soon apparent that our ability to generate sequences has far surpassed our capabilities to process them. Taking a look at benchmarks of commonly used DNA sequence aligners on a standardized single-processor machine, the popular Burrows-Wheeler Aligner (BWA) aligned 1,000,000 sequences of 100 base pairs in length to a human reference sequence in 97 minutes [2]. A more accurate program known as GMAP subsequently took over 2887 minutes, or just over 2 days, to perform the same alignment to a higher degree of accuracy [2]. However, the sequences aligned in this test represents only 0.016% of the 24-hour throughput of an Illumina HiSeq platform. Not only does this benchmark reveal problems of speed with sequence analysis, the costs of maintaining high-powered servers is also an issue. In order for precision medicine and personalized treatments for HIV to truly become viable for a large portion of the population, the problems associated with large scale sequence analysis, namely cost and speed, must be solved.

Instead of attempting to solve these problems by hiring more IT staff at bioinformatics labs and building more server infrastructure from scratch, leveraging well-established cloud computing platforms is becoming more attractive as a possible solution. Instead

Received: Feb./17 **Accepted:** Feb./17 **Published:** July/17

Citation Zhang, D. 2017. Leveraging Cloud Computing for Genome Analysis to Provide Patient-Specific HIV Treatments and Insight Into HIV Susceptibility Patterns. JEMI-PEARLS. 2:54-60.

of purchasing servers outright, cloud computing platforms allow any user to rent pre-partitioned computing resources at a set rate, billed hourly. Users can rent as many instances as needed with as much processing power or storage as desired. All of the instances are interconnected via high-speed links, making them ideal for cluster computing and workloads that are easily distributed among many workers, such as sequence analysis [3]. Large-scale data centers deployed across the world by tech leaders in the industry such as Amazon and Google also ensures a practically limitless number of servers and instances available to be utilized and a high level of uptime and redundancy to guard against data loss and corruption. This large scale also means that the instance rates are affordable and is significantly cheaper than running an in-house server. In a study conducted by Langmead *et al.* in 2009, researchers successfully re-aligned a human genome and simultaneously performed single-nucleotide polymorphism detection using 320 Amazon EC2 instances [4]. The endeavor took only 2.5 hours and cost approximately $113 US Dollars. Looking at historical pricing, the same task could be done today for less than $47 due to the ever-lowering prices of cloud computing [5, 6, 7]. Due to improvements in processor technology since 2009, the time taken by the test would also be significantly shorter today. This strongly suggests that the idea of utilizing cloud computing to overcome the problems associated with accessible personalized treatments of HIV is both financially and technically sound.

> "In order for precision medicine and personalized treatments for HIV to truly become viable for a large portion of the population, the problems associated with large scale sequence analysis, namely cost and speed, must be solved"

RESEARCH QUESTIONS

According to data collected in 2014, over 75,000 Canadians are infected with HIV and around 2,500 new individuals are infected annually [8]. The global HIV rates are much worse with over 36.7 million individuals infected worldwide and over 1.1 million deaths due to HIV annually, with the majority of cases originating in Sub-Saharan Africa [9]. Not only is HIV severely affecting the lives of millions, the economic damage due to HIV epidemics is significantly hindering the growth of developing nations [10]. Unfortunately, due to the high mutation rates of HIV, effective vaccines and universal treatments have proven to be difficult to develop. However, if we could collect sequence data on both the HIV type currently circulating within the patient as well as the patient's own DNA makeup, perhaps we could prescribe personalized treatment regimens for each individual based on the type of HIV identified. Though personalized HIV treatments are not novel, current solutions could be made more efficient and cost-effective by utilizing cheap and abundant cloud computing resources. The first question I wish to investigate is "*Can we leverage cloud computing and high throughput sequencing to quickly and cheaply recommend an effective drug regimen for specific HIV patients?*"

By sequencing and analyzing a large number of HIV patient and viral genomes, I also propose that we establish a centralized database where participating clinics and nations can contribute sequence data associated with HIV patients. Because the storage and analysis of such an enormous amount of data is not feasible with local servers, we must once again turn to the computing power and storage capabilities afforded to us by cloud computing. Through big data collection and analysis, I also hope to answer my second research question: "*Can the anonymous collection, analysis, and storage of global HIV patient genomes using cloud/distributed computing definitively reveal all genetic markers associated with "elite controllers"?*". As of right now, our knowledge of such genetic markers is limited to certain HLA types, certain mitochondrial DNA haplotypes, CCR5 deletions, and certain protein mutations [11, 12, 13]. Not only is this list far from exhaustive, our understanding of the mutation's phenotypic consequences is hazy at best. I believe that with cloud computing, we can greatly expand this list of genetic markers and help us understand the mutations associated with elite controllers.

On a similar vein, it's commonly believed that HIV exhibits cooperative mutations where one particular mutation enables the emergence of another seemingly unrelated mutation [14]. Due to HIV's high mutation

rate, it would be extremely advantageous if we could use a map of the possible cooperative mutations within the HIV genome to stay "one step ahead" of HIV mutations in a particular patient. I propose that we also leverage the data within our centralized database and cloud computing to attempt to create such a mapping. The final research question I wish to answer is "*Can we use machine learning and mathematical models facilitated by cloud computing to predict individual patient HIV mutations before they occur?*"

PROJECT NARRATIVE

Can we leverage cloud computing and high throughput sequencing to quickly and cheaply recommend an effective drug regimen for specific HIV patients?

In order to obtain the amount of data needed to answer my first research question, we must first undertake the monumental task of building an efficient and scalable centralized database of patient and HIV virus genomes. As mentioned before, affordable cloud computing solutions could offer significant cost and practicality improvements over existing infrastructure. There are two main avenues of obtaining cloud servers for this purpose. The main method is to rent cloud computing instances from tech giants such as Amazon, Google, and Microsoft. Looking at the price of storage, Amazon S3 currently charges $0.023 USD per gigabyte per month. Archival storage featuring slower access speeds lowers this cost down to $0.004 per GB per month [15]. Assuming *all* of the data associated with whole genome sequencing and analysis is around 310 GB [16], *everything* regarding a patient's genome, including a 30x genome sequence coverage, SNP/Indel calling, and alignment data can be stored for only $7.13 per month. Better yet, once the analysis is performed, this data could be moved to archival storage and kept for only $1.24 per month. Additional bloat such as redundant sequencing data can be further pruned to reduce storage costs down to mere pennies per patient.

While renting cloud instances from private companies is one way to build our platform, we could also partner with government-sponsored cloud computing initiatives. One such example is Compute Canada. With 27 data centers, 2 petaflops of compute power, 20 petabytes of storage capacity, and a strong focus on advancing Canada's research computing platform, Compute Canada seems to be an ideal government partner for this project [17].

Once a centralized database and computing platform is established for the purpose of HIV personalized medicine, I would invite participating countries, clinics, and bioinformatics labs to begin submitting raw sequence data of both HIV patient genomes and their viral genomes with the consent of the patient. Though the patient's own genetic makeup has little impact on HIV treatment effectiveness, this data is still crucial to obtain, especially when attempting to answer the other questions outlined in this proposal. Ideally, this data would be processed by a team of recruited bioinformaticians with experience in the area of cloud computing. Due to our ability to spin up instances and quickly obtain access to our database directly on the cloud, data analysis can be performed by our experts without any local hardware. The HIV genome data will be scanned for SNPs using a robust variant calling tool such as Samtools. These variants will then be compared against known mutations and any drug resistances that they confer to the virus. Any unknown variants would be documented for further investigation by future studies.

When enough data is collected for our experts to confidently profile the majority of possible drug resistance genetic markers in the HIV genome, we would open up the database for query. Clinicians should be able to take a sample of the patient's HIV genome, upload it to our database, and our cloud analysis platform should be able to quickly and accurately recommend a tailor-made drug regimen for the patient at hand.

Because the technologies we have in place already allows us to perform this type of analysis (albeit with inferior costs and speeds), the only potential roadblock is integrating cloud technology into our existing bioinformatics pipelines. As cloud computing instances are not significantly different from on-site servers, I do not expect this to be a big hurdle. In fact, I expect that my proposed steps of establishing a centralized database, soliciting patient and viral genome samples from participating bodies, analyzing the genome samples by recruited bioinformaticians, and establishing a global querying database for quick HIV drug resistance profiling will not be technically challenging. To address my first proposed research question, I do expect that we will be able to leverage cloud computing to recommend effective drug regimens for individual patients.

If the platform mentioned is realized, it could mean huge quality of life improvements for existing HIV patients. The CDC estimated in 2010 that the lifetime cost of treating HIV is around $379,000 USD and nearly 30% of infected individuals are uninsured [18]. Not only is this a huge amount of money for any individual, it's simply out of the question for residents in developing

nations. With my project, I hope to alleviate some of the economic burdens by ensuring that HIV patients are only prescribed treatments that are known to be effective for their particular viral type. While this initiative will not decrease the costs of the actual drugs, I hope that the scientific knowledge gained during this endeavor will spark improvements in HIV medication and treatment methods.

So far, the analysis done on the data we hope to collect focused on the HIV genome itself. As we know, host genetic factors play a huge role in HIV infectivity and virulence. We can also leverage the host genome data in our database to answer important questions related to host factors.

Can the anonymous collection, analysis, and storage of global HIV patient genomes using cloud/distributed computing definitively reveal all genetic markers associated with "elite controllers"?

Despite the progressive and crippling nature of HIV, there is roughly a 5-15% of HIV positive individuals who do not appear to have compromised CD4 cell counts despite being infected by HIV [19]. These individuals are known as long-term nonprogressors, or "elite controllers". While some of these cases are well characterized and are explained by genetic factors such as CCR5 deletions or overexpression of A3G [13], an apolipoprotein which doubles as a reverse-transcriptase inhibitor, the causes of the other cases remain a mystery.

Because I hope to establish a database with potentially millions of HIV positive patient genomes as their associated viral genomes, I wish to begin answering my second research question by also collecting information on patient viral loads, estimated length of infection, and CD4 counts along with sequencing information. Since phenotypes are rarely binary, I wish to explore not only the genetic markers which give rise to elite controllers, but also the extent of the protection each mutation confers to the host. By using cloud computing instances and once again hiring bioinformatics professionals, I hope to assemble a near exhaustive list of genetic markers which can affect host susceptibility to HIV. This process can once again be facilitated by SNP discovery tools such as Samtools. Any SNPs discovered can be cross referenced with some of the many available human gene mutation databases such as the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff, or ClinVar by the NCBI. Any previously unknown mutations can once again be documented and researched in follow-up studies. Once the polymorphisms are called and recorded, they will be paired up with the corresponding patient's HIV viral load, CD4 count, and normalized based on infection duration. Ideally, we would be able to extract correlations between certain mutations and possible elite controller statuses. Using this information, we can build a database of host gene mutations that could affect host susceptibility to HIV, both positively and negatively.

While this research question is more technically challenging than the first due to our limited understanding of many genotypic mutations, I believe that our knowledge today can provide a very solid foundation for the establishment of such a database. Though I do not expect us to be able to construct an exhaustive list of genetic markers with associations to elite controllers as indicated in my second research question, we should be able to apply mathematical modeling to our collected data to obtain a respectable list that can only be improved as more data is collected.

The impacts of establishing such a database might not be immediately felt in our current stage of technological development. We could provide genetic profiling services to individuals to inform them of their HIV susceptibility, but this information may only be of interest to high risk individuals such as those belonging to MSM or IDU populations. A more exciting use-case for this database relies on the development and maturation of gene editing techniques such as the promising CRISPR/Cas9 system. Once we identify putative genetic markers which confer HIV resistance with minimal adverse phenotypic effects, we could potentially eliminate the risk of HIV for any individual by directly introducing these mutations into their genome.

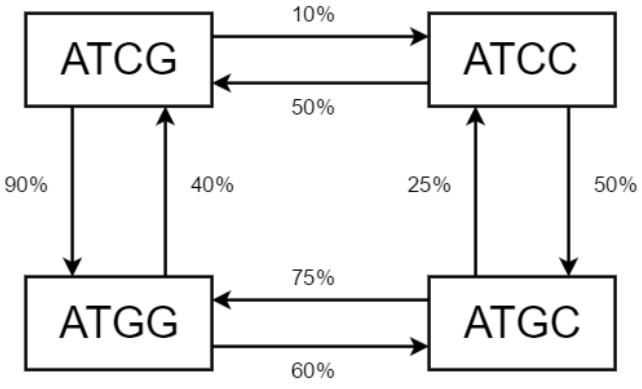
For this particular research question, we have only applied mathematical modeling to host genetic factors. If we expand this analysis to include HIV genomes as well, we could potentially uncover even more pieces of the HIV puzzle.

Can we use machine learning and mathematical models facilitated by cloud computing to predict individual patient HIV mutations before they occur?

While, on the surface, viral mutations appear to come and go at random, many viruses exhibit cooperative mutations. Explained previously, cooperative mutations have been the center of many studies regarding Influenza A [20], but studies into HIV regarding this topic have been sparse. Because one of HIV's hallmarks is its fast mutation rate, I believe that it's important to study and characterize HIV mutation patterns.

In order to create a mapping of the cooperative mutations associated with HIV and answer our third and final research question, we must once again leverage affordable computing services from cloud providers. By tapping into the data in our centralized database, we can simply retrieve the SNP data already present from our steps in the first research question. Because finding pathways between mutations and identifying possible cooperative mutations is not as simple as identifying variants in the genomes, we must apply mathematical modeling on our variant data. One way of identifying possible pathways is to use a well-established probability model known as Markov chains. The underlying principle of Markov chains in relation to biological sequences is that all unique mutated states can be represented as a discrete state in a Markov chain (Fig. 1). A mutation which causes a sequence to change is modeled as a transition in the Markov chain state [21]. The probability of this mutation taking place is simply the probability of the corresponding Markov chain state transition. These probabilities can be mathematically estimated using "training data", which in this case, would be our genome database. Intuitively, the larger the training data set, the more accurate the Markov chain transition probabilities would be. Once our Markov chain is complete, we can once again open it up for query. Given a particular patient's HIV genome, our Markov chain could help us predict the possible and most likely mutations that the virus will undergo. Another method that is becoming more and more popular is utilizing neural networks and machine learning to identify these mutation patterns [22]. These methods rely on similar concepts and also require training data, but are significantly more computationally intensive and are not as well field-tested.

As the methods and techniques required by this proposal are well understood, I expect that we will definitely be able to construct a mathematical model for HIV mutation prediction. However, I am concerned about the accuracy of the resulting model. Depending on the size of our training data, and the methods used to generate our model, I am expecting extensive tuning and refining before the output can be generalized and applied to the average HIV patient. While I do not expect us to be able to predict HIV mutation patterns to a clinically significant degree during the first iteration, I do believe that gradual improvements will eventually produce a robust mathematical model that will be able to make such predictions.

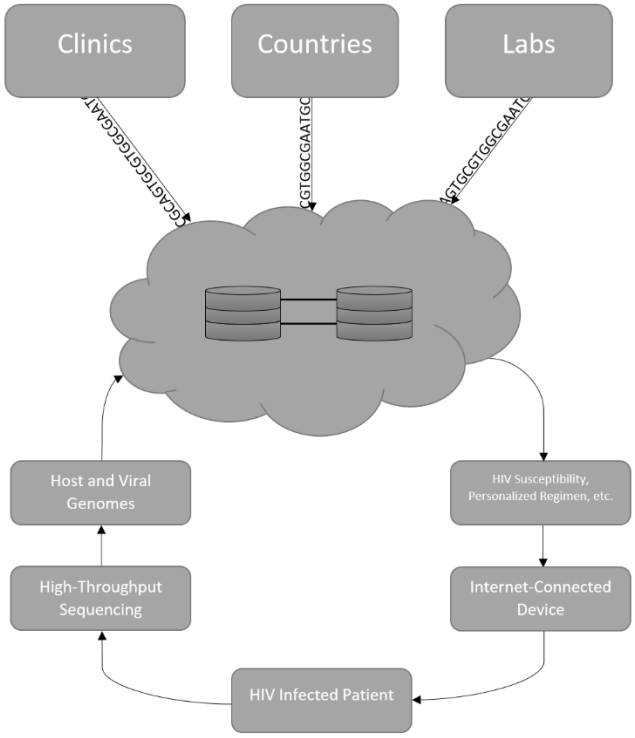


If we are eventually successful at creating such a prediction model, we could enable physicians to prescribe treatments against not only the circulating virus within the patient, but also the expected viral type in the coming months. This could potentially increase patient quality of life by helping physicians choose treatments that are likely to be effective for a longer period of time. Since regimen adherence by HIV patients could be improved [23], a simplified and stable treatment plan with minimal changes along the way could help patients stay on track. Learning more about the possible mutations in the HIV genome and the mutation pathways that the virus can take can also help us in our goal of creating an HIV vaccine. Since HIV is essentially a life-long disease requiring indefinite treatment, any knowledge that could help us prevent infections would prove invaluable to the field of medicine.

If we are able to answer yes to any of the research questions I have proposed, it could mean sweeping changes to the field of HIV treatment and bioinformatics pipelines across the industry. The field of precision medicine has long been dismissed as a pipe dream with arguments of barriers such as unrealistic costs. However, as cloud computing matures as a platform, it could quickly become a solution to the many challenges associated with precision medicine and personalized treatments of viral infectious diseases.

SUMMARY AND CONCLUSION

The proposal that I have outlined lays down the framework for a high-tech, scalable, and accessible database for HIV patient and viral genomes (summarized in Fig. 2). The two biggest challenges facing precision medicine today, cost and speed, are big roadblocks preventing groundbreaking research on not only HIV. If HIV can serve as a proof of concept for breaking these roadblocks via cloud computing, this opens the door for the establishment of similar databases for other viral infectious diseases and many other initiatives related to personalized medicine.



The ultimate goal of my research proposal is realized

FIG. 2 Visualization of the proposed cloud platforms. This figure gives a general overview of the cloud infrastructure and how it will be used. HIV host and viral genomes will be uploaded to the cloud via participating countries, clinics, and bioinformatics labs to generate training data for our cloud database. Once robust models for personalized treatments, HIV susceptibility prediction, and mutation prediction are established, it can be opened for query by clinicians to aid in the treatment process of real HIV patients.

if a patient is able to step into a clinic, perform a DNA sequence, and obtain information on their HIV susceptibility, viral mutation patterns, and recommended treatments, all for less than $10. This can prevent unnecessary spending on ineffective treatments that the patient is likely to become resistant to in the near future. For patients residing in rural areas with inconvenient access to a clinic, remote access to test results via any internet connected device is also a possibility due to cloud computing. Since data processing is already done on the cloud, retrieval of test results through the internet is only a matter of establishing an intuitive interface for the end-user. Other exciting opportunities enabled by big data collection and massively parallel data analysis include the development of highly sensitive and specific tests for host HIV susceptibility and eventual vaccine development. As mentioned before, once the platform for HIV matures, a similar approach can be taken to establish databases for other viral diseases as well. Due to the prevalence of co-infections of viral diseases, large scale data collection on a number of viruses and cross-analysis between the viral databases could also reveal missing links and shed light on any discrepancies discovered when looking at only one virus.

However, my research proposal does present numerous challenges, both technical and bioethical. As with any large-scale data collection initiative, the identity of patients must be protected at all costs. Not only is genomic data extremely personal, HIV is still heavily stigmatized across many cultures. An accidental leak of patient identities and their HIV statuses could prove disastrous for future big data collection initiatives in the field of precision medicine. On the same vein, database breaches and hacks must be prevented. Methods such as encryption using community-audited encryption algorithms that are free of government backdoors such as AES-256 is a viable approach [24]. Data loss due to corruption must also be considered and prevented using methods provided by cloud computing companies such as offshore backups in multiple data centers across the world. The storage and analysis of patient genomic data across multiple jurisdictions might also raise legal concerns in certain countries. Some more trivial challenges include retraining of bioinformatics staff, restructuring of bioinformatics pipelines, and the inability to access the platform in areas of poor internet connectivity. However, these concerns will hopefully be resolved during the course of the research as part of the project.

If my research proposal is to be carried out and realized, every one of these concerns must be addressed and efficiently managed. However, if the challenges are overcome, it could spark a flurry of advancements in the field of precision medicine. The success of this project will serve as a prototype for future PM initiatives and could lead to fast and affordable personalized treatments for the masses in the future.

ACKNOWLEDGEMENTS

I would like to thank Dr. François Jean for providing a huge amount of guidance and resources for my research topic. He contributed greatly to the development of my topic and was a fantastic mentor throughout this process. I would also like to thank my MICB 406 discussion group members Michael Li, Daniel Chen, and Manish Hosane. Their insights, suggestions, and eagerness to help improve my proposal were invaluable.

REFERENCES

1. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016 Jan;107(1):1-8.
2. Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. *Bioinformatics*. 2012 Oct;28(24):3169-3177.
3. Microsoft. What is cloud computing? A beginner's guide [Internet]. Microsoft; 2017 [updated date unknown; cited 2017 Jan 19]. Available from: https://azure.microsoft.com/en-ca/overview/what-is-cloud-computing/
4. Langmead B, Schatz MC, Lin J, Pop M, Salzberg SL. Searching for SNPs with cloud computing. *Genome Biol* [Internet]. 2009 Nov [cited 2017 Feb 19];10(11):R134. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-11-r134 DOI: 10.1186/gb-2009-10-11-r134
5. Mor Y. Analyzing AWS EC2 Price Drops Over the Past 5 Years [Internet]. Cloudyn.com; 2015 [updated 2015 Nov 15; cited 2017 Feb 19]. Available from: https://www.cloudyn.com/blog/analyzing-aws-ec2-price-drops-over-the-past-5-years/
6. Barr J. Amazon S3 Price Reduction [Internet]. Amazon.com; 2012 [updated 2012 Feb 6; cited 2017 Feb 19]. Available from https://aws.amazon.com/blogs/aws/amazon-s3-price-reduction/
7. Barr J. AWS Storage Update – S3 & Glacier Price Reductions [Internet]. Amazon.com; 2016 [updated 2016 Nov 21; cited 2017 Feb 19]. Available from: https://aws.amazon.com/blogs/aws/aws-storage-update-s3-glacier-price-reductions/
8. Public Health Agency of Canada. Summary: Estimates of HIV Incidence, Prevalence and Proportion Undiagnosed in Canada, 2014 [Internet]. Public Health Agency of Canada; 2015 [updated 2015 Nov; cited 2017 Feb 19]. Available from: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/summary-estimates-hiv-incidence-prevalence-proportion-undiagnosed-canada-2014.html
9. WHO. HIV/AIDS [Internet]. World Health Organization; 2015 [updated date unknown; cited 2017 Feb 19]. Available from: http://www.who.int/gho/hiv/en/
10. Dixon S, McDonald S, Roberts J. The impact of HIV and AIDS on Africa's economic development. *BMJ*. 2002 Jan;324(7331):232-234.
11. Leipzig J. Mitochondrial DNA Haplogroups influence AIDS Progression. *AIDS*. 2008 Nov;22(18):2429-2439.
12. Kindberg E. A nonsense mutation (428G-->A) in the fucosyltransferase FUT2 gene affects the progression of HIV-1 infection. *AIDS*. 2006 Mar;20(5):685-689.
13. Pasquale M, Kourteva Y, Allos T, D'Aquila R. Lower HIV Provirus Levels Are Associated with More APOBEC3G Protein in Blood Resting Memory CD4+ T Lymphocytes of Controllers In Vivo. *PLoS One* [Internet]. 2013 Oct [cited 2017 Feb 19];8(10):e76002. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0076002 DOI: 10.1371/journal.pone.0076002
14. Foulkes-Murzycki JE, Rosi C, Yilmaz NK, Shafer RW, Schiffer CA. Cooperative Effects of Drug-Resistance Mutations in the Flap Region of HIV-1 Protease. *ACS Chem Biol*. 2013 Mar;8(3):513-518.
15. Amazon AWS. Cloud Storage Pricing – Amazon Simple Storage Service (S3) [Internet]. Amazon.com; 2017 [updated date unknown; cited 2017 Feb 20]. Available from: https://aws.amazon.com/s3/pricing/
16. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think!. *Genome Biol* [Internet]. 2011 Aug [cited 2017 Feb 20];12(125):1. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2011-12-8-125 DOI: 10.1186/gb-2011-12-8-125
17. Compute Canada. Renewing Canada's Advanced Research Computing Platform [Internet]. Compute Canada; 2017 [updated date unknown; cited 2017 Feb 20]. Available from: https://www.computecanada.ca/renewing-canadas-advanced-research-computing-platform/
18. Wenger A. The Cost of Treating HIV: One Man's Monthly Medical Bill [Internet]. San Francisco: Healthline; 2014 [updated 2014 Dec 16; cited 2017 Feb 20]. Available from: http://www.healthline.com/health/hiv-aids/monthly-cost-treating-hiv#1
19. Broder M. New Research on HIV Elite Controllers [Internet]. New York: MedPage Today; 2015 [updated 2015 Mar 23; cited 2017 Feb 20]; Available from: http://www.medpagetoday.com/resource-center/HIV/Elite-Controllers/a/50624
20. Chen H, Zhou X, Zheng J, Kwoh C. Rules of co-occurring mutations characterize the antigenic evolution of human influenza A/H3N2, A/H1N1 and B viruses. *BMC Med Genomics* [Internet]. 2016 Dec [cited 2017 Feb 20]. Available from: https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-016-0230-5 DOI: 10.1186/s12920-016-0230-5
21. Djordjevic IB. Markov Chain-Like Quantum Biological Modeling of Mutations, Aging, and Evolution. *Life (Basel)*. 2015 Sep;5(3):1518-1538.
22. Vural S, Wang X, Guda C. Classification of breast cancer patients using somatic mutation profiles and machine learning approaches. *BMC Syst Biol*. 2016 Aug;10(3):62.
23. Simoni JM, Kurth AE, Pearson CR, Pantalone DW, Merrill JO, Frick PA. Self-report measures of antiretroviral therapy adherence: A review with recommendations for HIV research and clinical management. *AIDS Behav*. 2006;10(3):227-245.
24. National Institute of Standards and Technology. Announcing the ADVANCED ENCRYPTION STANDARD (AES). Gaithersburg, Maryland: National Institute of Standards and Technology; 2001 Nov 26. 51 p. Report No.: 197.