BIOINFORMATICIANS: THE HIDDEN HEROES OF THE COVID-19 PANDEMIC

As the COVID-19 pandemic continues to evolve, we explore how bioinformatics has paved the path to this point.

There has been a lot of skepticism surrounding how scientific processes were able to move so quickly during the COVID-19 pandemic, bringing treatments, vaccines and prevention solutions to the public sphere in record time. While this was very much a collaborative effort from the pharmaceutical industry, healthcare systems, regulators, academia and government, it was initiated, informed and progressed through the generation and application of data.

To leverage big data, from the first publicly available SARS-CoV-2 genome sequence and computational models to understand the virus and predict mutations, through to screens for potential therapies and vaccines – bioinformaticians have had a crucial role to play every step of the way. These bioinformatic methods have dramatically reduced experimental lab time and enabled the communication of key information between the aforementioned collaborators, and to the public both on a national and worldwide scale.

The pandemic has led to a multitude of learnings about the scientific process and how researchers work. Both had to adapt to a new pace, as well as the vast volumes of data that were produced and required to ensure decisions could be made quickly, research could progress, and public concerns were allayed. There was also a 'work-from-home' call, with many labs closing completely or limiting their physical presence, so it was crucial for researchers and research to adapt to computational methods for work to continue remotely.

It now feels as though the world is beginning to move on from COVID-19 – legal restrictions have been abolished in many countries and populations are now re-living their pre-2020 'normal'. The role of bioinformaticians during the pandemic has been widespread and essential to the feeling of moving on from COVID-19 that so many are now experiencing. We now have a holistic view of the SARS-CoV-2 virus. We're preparing for future mutations and resulting variants with current vaccination programs and ongoing vaccine development.

Let's look at how bioinformatics has enabled scientific progress during the COVID-19 pandemic and the crucial role it's played in our ability to move forward with preparedness, prevention, and control.

SEQUENCING SARS-COV-2

Traditional virology research methods cannot obtain data as urgently and at such a scale as has been necessary during the COVID-19 pandemic. The first SARS-CoV-2 genome sequence was established early and published on GenBank on 10 January 2020 [1] (although James Shore of The Native Antigen Company (Figure 1) could argue it was publicly available prior to this date) [2].

A nucleic acid sequence on its own cannot tell us very much. Bioinformatic analysis is crucial to understanding more about the origins, similarities and evolution of a virus. Sequence alignment and phylogenetic analysis was used to identify that COVID-19 was caused by a coronavirus and the likely origin of SARS-CoV-2 could be a bat [3].

The protein sequence can be determined from the gene sequence using computational methods. The structure of the SARS-CoV-2 virus could be predicted through AlphaFold [4] using the RNA sequence and knowledge of the SARS virus structure, or a 3D reconstruction from imaging techniques, like Cryo-EM, using the software RELION [5].

Bioinformatics technologies have also played a role in diagnostics during the COVID-19 pandemic. PCR has been firmly established as the diagnostic gold standard; however, this offers no additional insights on virus evolution, origin or mutations. ▶

Figure 1. James Shore of The Native Antigen Company.

Where next-generation and third-generation sequencing methods have been employed to detect SARS-CoV-2, sequences can be mapped against a reference genome, and bioinformatic analyses allow variant patterns within populations to be identified and infection routes traced.
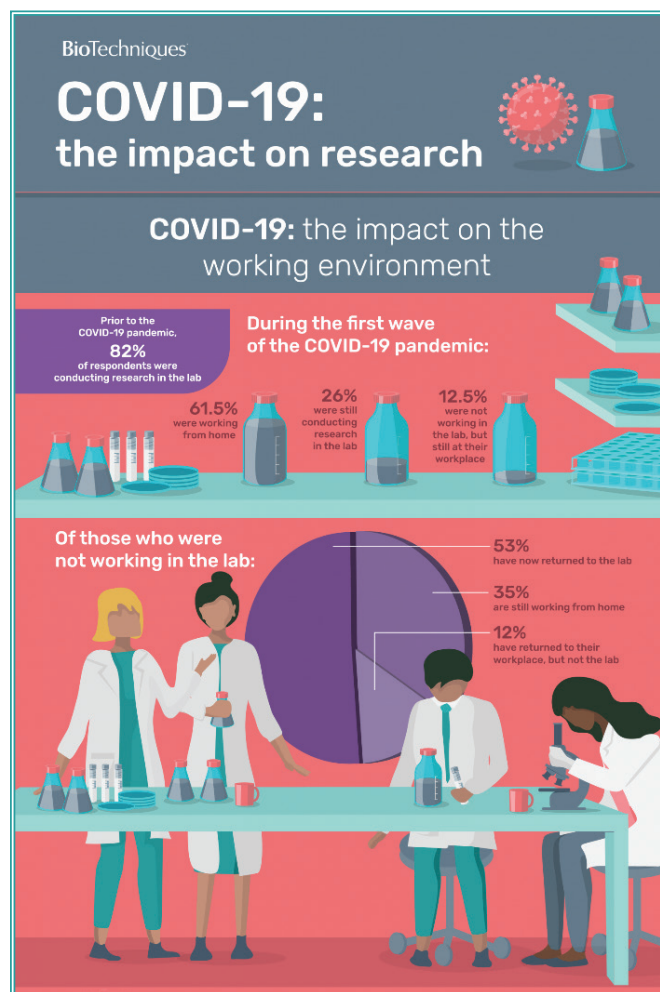

Figure 1. A section of the BioTechniques.com infographic demonstrating the results of a member survey from October 2020.

COVID-19 REPORTING

After the first cases of COVID-19 were confirmed and the SARS-CoV-2 virus genome was sequenced, the disease rapidly spread. The WHO declared the outbreak an "International Public Health Emergency" on 11 March 2020.

From this point, people immediately began working from home, including scientists. A member survey that we conducted in October 2020, with results published in December 2020 (see Figure 2), is a good representation of the state of lab research at that point in time. We discovered that during the first wave of the COVID-19 pandemic, only 26% of respondents were still conducting research in the lab, compared with 82% prior to the pandemic. A total of 35% had to alter their research focus, 60% of whom were working specifically on COVID-19, 83% said the pandemic delayed their research, and many began doing remote research – which involved a heavy focus on computational work [6].

There was a huge reliance on real-world data for scientists advising governments and both advising the public. COVID-19 case numbers and death numbers were reported daily (including on the WHO's own dashboard) [7], and extrapolations were taken to predict the near and distant future course of the pandemic, enabling decisions to be made by governments and communicated to the public.

While it may now feel like the COVID-19 pandemic is coming to a close, at the time of writing [15 March 2022], the WHO has yet to declare that the pandemic is over. In a 'World View' article published in Nature [8], the author requests that governments and organizations should "commit to transparent COVID data until the WHO declares the pandemic is over." They suggest that efforts have been focused on making data "look good" instead of just ensuring it is available and believe a published text file would suffice.

This raises the question: who is pandemic data most valuable to? Should it be presented in a format that is easily digested and interpreted by government officials and the public to avoid the potential of miscommunication and misinterpretation, enabling decisions to be made quickly and justly. Or, does the presentation of data not matter so much, and should data just be made publicly available in a format that enables scientists to leverage this for further research?

The answer is that both is necessary – for governments and the public, it is essential that bioinformaticians spend extra time formatting data in a way that is valuable for science communication and has no risk of misinterpretation – decisions for a nation are based on these data. For scientists, a global repository for immediate pandemic data that requires consistent standards in a format that can be easily analyzed further by bioinformaticians would be valuable.

TREATING COVID-19

The fast identification of the genome sequence and protein structures of SARS-CoV-2 enabled targeted drug screening for potential therapies that could prevent or limit viral propagation and infection to become a priority as the severity of the pandemic was realized [9]. The existence of close coronavirus relations also enabled scientists to draw on past research of potential therapeutics for MERS and SARS [10].

Drug repurposing, where predictions are made on which FDA-approved drugs might be effective and interactions between the drug and virus modelled computationally, has many benefits over novel drug discovery. This is the safest approach with the lowest risks for health, the environment and economically. There are likely to be few off-target results, resulting in minimal side effects as the drug has already been FDA approved. By modelling experiments computationally, this saves experimental time, resources and costs.

Since SARS-CoV-2 shares 96% RNA-dependent RNA polymerase in common with SARS, a primary route for COVID-19 drug discovery was focused on drugs that target viral RdRp proteins, of which remdesivir (Figure 3) – a broad-spectrum antiviral – is one [11]. Post-SARS outbreak in 2003, lopinavir was identified as a potential inhibitory drug, showing activity in *in vitro* screens [12]. Both drugs have since been used as approved treatments for COVID-19.

Many research groups used bioinformatic platforms to assess which repurposed drugs would have potential as COVID-19 treatments. ACE2 and TMPRSS2 were popular targets as they are proteins that are used by the virus to enter cells. Molecular dynamics (which simulates experimental conditions), molecular docking (which searches for ligands that bind with the receptor's active site) and artificial intelligence (which makes the search for drug repurposing more efficient by rapidly sifting through databases with the most likely successful matches) combined form high-throughput AI-based binding affinity prediction platforms, which are hugely beneficial in predicting therapeutics [3].

A study published in the Journal of Translational Medicine in June 2020 also investigated a novel bioinformatics Disease Canceling Technology platform, which focused on finding FDA-approved drugs that could halt coronavirus-associated gene expression changes [9].

VACCINATING THE WORLD
Creating an effective vaccine is one of the best ways to slow disease spread and prevent severe symptoms. Thus, this was a concerted focus of many major industry and academia players across the world up to and beyond the first COVID-19 vaccine's approval by the FDA under Emergency Use Authorization on 11 December 2020. This Pfizer-BioNTech vaccine was closely followed by Emergency Use Authorizations for the Moderna and Johnson & Johnson vaccines. All these vaccines have since been granted full FDA approval.

Bioinformatics has a crucial role to play in vaccines research and development through three key applications [3]:

- Reverse vaccinology, which can identify target antigens using knowledge of the virus' genome and proteome data;
- Immunoinformatics, which allows the screening and identification of epitopes with a high ability to elicit an immune response;
- Structural vaccinology, which combines reverse vaccinology with structural biology to design improved protein-based vaccine components [13].
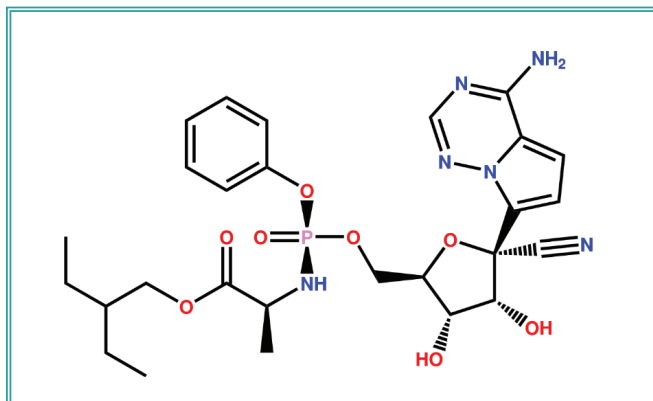

Figure 1. Structure of remdesivir.

Although these bioinformatic methods have been invaluable in the development of safe and effective vaccines for COVID-19, SARS-CoV-2 is an RNA virus and so is incredibly prone to mutations. This means that numerous iterations of the vaccine will be required, unlike the one vaccine approach used for stable DNA viruses, such as smallpox. So far, vaccines are still highly effective against all current variants, but this could change as the virus continues to mutate.

By furthering our understanding of how SARS-CoV-2 mutates and its different variants through the use of computational visualization technologies, we will be better prepared for both therapeutic and vaccine development. Vaccine development and administration programs are ongoing, and there are many innovative approaches currently in the works to mitigate the effects of variant emergence.

LIVING WITH COVID-19
The role of bioinformatics during the COVID-19 pandemic has been widespread. Bioinformaticians have been heavily relied on through every stage of pandemic research, from genome analysis, through to data reporting, drug discovery and vaccine development. As has been demonstrated during the pandemic, the ability to perform computational methods is incredibly valuable for researchers across all fields of research.

Moving forward, to live with COVID-19 and prepare for future pandemics, bioinformatics and *in silico* studies will continue to play a vital role in our developing understanding of virus structure, the impact of new mutations, predicting treatments and developing vaccines.

However, the life sciences will become even more digital, and rely more heavily on data. There is still a long way to go to enable connectivity across the entire healthcare system, from scientific discovery in academic laboratories, right through to drug administration in the clinic. The potential benefits of this connected science supply chain will be monumental, but will involve a world-wide, industry-wide mindset and approach shift. The future is going to be digital, and those who do not adapt will be left behind.

Written by Ebony Torrington

REFERENCES

1. Sawyer A, Free T, Martin J. Metagenomics: preventing future pandemics. *BioTechniques* 70(1), 1–4 (2021).

2. BioTechniques. Catching up with COVID-19: producing the first SARS-CoV-2 spike proteins. www.biotechniques.com/covid-19/catching-up-with-covid-19-producing-the-first-sars-cov-2-spike-proteins/

3. Ma L, Li H, Lan J *et al.* Comprehensive analyses of bioinformatics applications in the fight against COVID-19 pandemic. *Comput. Biol. Chem.* 95, 107599 (2021).

4. Al-Janabi A. Has DeepMind's AlphaFold solved the protein folding problem? *BioTechniques* 72(3), 73–76 (2022).

5. RELION 4.0. RELION Table of Contents. https://relion.readthedocs.io/en/release-4.0/

6. *BioTechniques*. How has the COVID-19 pandemic impacted lab researchers? www.biotechniques.com/molecular-biology/covid-19-pandemic-impacted-lab-researchers/

7. World Health Organization. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/

8. Mathieu E. Commit to transparent COVID data until the WHO declares the pandemic is over. *Nature* 602, 549 (2022).

9. Kim J, Zhang J, Cha Y *et al.* Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease-2019 (COVID-19). *J. Transl. Med.* 18, 257 (2020).

10. Abdelsattar AS, El-Awadly ZM, Abdelgawad M, Mahmoud F, Allam SA, Helal MA. The role of molecular modeling and bioinformatics in treating a pandemic disease: the case of COVID-19. *Open COVID J.* 1, 216–234 (2021).

11. Ko W-C, Rolain J-M, Lee N-Y *et al.* Arguments in favour of remdesivir for treating SARS-CoV-2 infections. *Int. J. Antimicrob.* 55(4), 105933 (2020).

12. Cao B, Wang Y, Wen D *et al.* A trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. *N. Engl. J. Med.* 382, 1787–1799 (2020).

13. MDPI. Vaccines special issue "Therapeutic and diagnostic applications of structural vaccinology". www.mdpi.com/journal/vaccines/special_issues/Structural_Vaccinology