

Gatekeeping the Pandemic

“The reluctance to share data before publishing has become a tricky bottleneck.”



P06 FEATURED ARTICLE

--

THE MASSIVE EFFORT TO
COLLECT COVID-19 DATA --
AND THE INCENTIVE TO
KEEP IT BEHIND CLOSED
DOORS

Society News



BINFSOC

BINF3010.

This coming Thursday, BINF3010's mid-term exam will be held. We wish everyone doing the course the best! Make sure to have plenty of breaks, go over the lecture notes, and maybe create some practise questions for yourself to answer for the essay questions.

R Workshop.

BINFSOC will be hosting a collaboration with MathSoc in week 6 of this term to go over the basics of the R software package. R is a free and powerful statistical language, and is widely used in both uni courses and the workforce. The workshop will cover what you can do with R, how easy it is to use it, and show key functions in data handling and plotting. This will be a two hour revision session ideal for students in BINF3010 this term, or anyone doing comp sci / engineering and want to brush up on their statistical analysis skills.

> Date: Tuesday 6th July

> Time: 2 - 4 PM

> Location: Zoom

Merch.

BINFSOC are considering releasing merchandise on our website in the form of posters. We're still in the design stage, but stay tuned for some bioinformatics designs of a similar style to our BINFsights newsletters. If there's a particular design you'd like to see us to make into a poster, tell us at binfsights@unswbinfsoc.com

Student Guide.

We are now in the planning stages of a Student Guide for first year bioinformatics students. We're aiming to include info about degree structure, where to find help and advice regarding your courses, and a general overview on what to expect from the bioinformatics world. If you're interested in giving us advice on what to include, as someone thinking of changing to bioinformatics or as a past student who wishes they had been told something about their journey beforehand, contact us at binfsights@unswbinfsoc.com

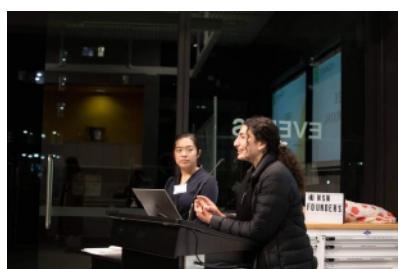
B Networking Night.

BINFSOC's Networking Night saw a great turnout of biotech students and seasoned industry representatives alike to network with each other, discuss career advancements, give advice on how to navigate the biotech industry landscape, and of course to enjoy the provided catering.

We'd like to say a huge thank-you to our representatives for taking the time outside of their busy schedules to help out the next generation of computational biologists. In attendance were Warren Kaplan and John Reeves from the Garvan, Evelyn Wu and Jake Richards from Novartis, and UNSW's very own Bruno Gaeta.

We hope you all had a wonderful time!

– The BINFSOC team



“

Computers are to biology
what mathematics is to
physics.

— Harold Morowitz

gatekeeping the pandemic

Writer Anthony Nguyen *Editor* Cam McMenamie

Open data sharing, open doors?

From universities to pharmaceutical companies, small and large laboratories from around the world are generating huge amounts of data each day. Any information being collected and used to validate original research findings plays an important role in moving science forward.

What does data sharing enable?

Research data is valuable. Headed by big data, the calls for collaborative and open data initiatives have grown louder, arguing that the open sharing of data is key to accelerating scientific progress. For example, previous data might be used to underpin a new hypothesis, to confirm experimental results, or to reduce the repetition of existing experiments based on the same data. However, in the public domain, it could also enable large scale analyses across different databases or enable easy access to the data by resharing the complete datasets in the form of supplementary materials or public repository links in related studies. All of this points to results being delivered seamlessly and efficiently, with remarks that the benefits have not only a national reach, but a global one too.

What kind of data is being collected during the COVID-19 pandemic?

In the context of molecular biology, the data in question might be genomic sequencing data of

SARS-CoV-2, the virus responsible for COVID-19, and its emerging strains. Over a year after the beginning of the pandemic, an international consortium named the Global Initiative on Sharing Avian/All Influenza Data (GISAID) has accumulated over 1.2 million SARS-CoV-2 genome sequences, with contributions from researchers from over 170 different countries. While it is not the only repository for viral genomes, with unrestricted databases ENA and GenBank also in the picture, it has become the most popular database among researchers due to its policies which incentivise data sharing.

“The reluctance to share data before publishing has become a tricky bottleneck.”

One concern within the scientific community is having their data scooped by other teams before having had the chance to analyse it themselves. That is to say, anonymous access to data generated by a laboratory before it is published could open doors for its unrestricted use and analysis by others without proper acknowledgement or credit. Such frustrations are often observed by smaller laboratories whose lab-work efforts may be easily accessed and used by others who were not burdened with

collecting and preparing those results. The reluctance to share data before publishing has become a tricky bottleneck. This comes as insights drawn from big data experiments often lead to public recognition, reputation, patents and grants.

However, those who access data from GISAID must first confirm their identities and agree to credit the laboratories who generated the data in the first place. Not only this, but it is also encouraged that users contact and collaborate with the original data providers in multi-author efforts.

From past to present

In previous health crises such as the outbreak caused by the Ebola virus, real-time and widespread sharing of both genomic and epidemiological data in the public domain before formal publication enabled the rapid publication of further studies.

Similarly, with recent months of the COVID-19 pandemic seeing a climbing of dangerous viral strains and an interest in using genomic surveillance to guide national policies (for

example, on vaccine development and distribution) the amount of sequence data submitted to GISAID has also climbed.

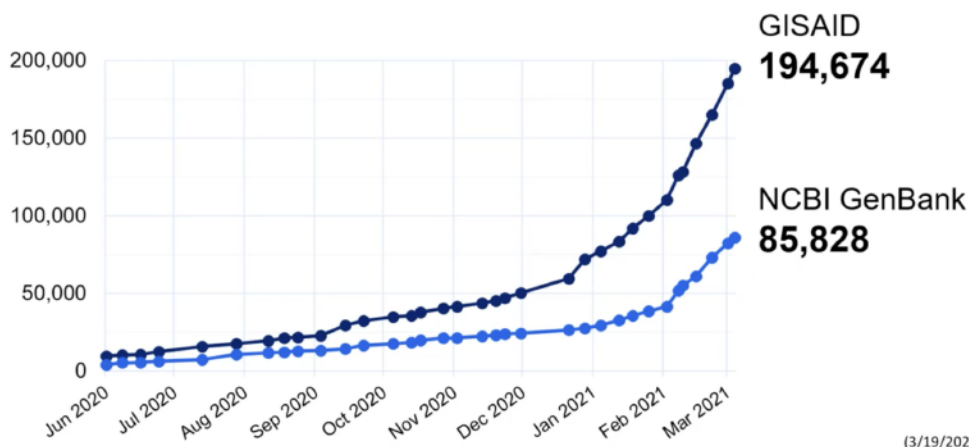
If this climb can be partially attributed to the incentives GISAID provides for smaller laboratories, such as those in lower to middle income countries where SARS-CoV-2 strains of interest have become prevalent due to lower vaccination rates, to share their data in real-time, then the importance and robustness of incentivising data sharing must be realised.

In light of this, there has been dialogue surrounding the need for any form of “gatekeeping” during a global health pandemic. Concerns revolve around the impact these restrictions may have on the flow of scientific progress should crucial data be subject to the nuances of intellectual property or confidentiality.

In both cases, the ultimate goal is to ameliorate our response to health crises such as the COVID-19 pandemic. Whether or not open or restricted sharing of genome data is the correct approach to this is still up for debate.

SARS-CoV-2 submissions (GISAID + NCBI)

From US laboratories



SOURCE: UNITED STATES CENTERS FOR DISEASE CONTROL AND PREVENTION

Open Position @ UNSW

PhD student in single cell computational biology



-
SUPERVISOR

--

Dr. Fabio Zanini
Data Driven Biomedicine lab @ UNSW Sydney

> fabilab.org

ABOUT THE PROJECT

--

While Physics has a standard model that explains most phenomena in the universe in terms of interactions between elementary particles, such a model is still missing in biology. To fill this gap, we are looking for a talented PhD student to create a standard model of cell biology.

-

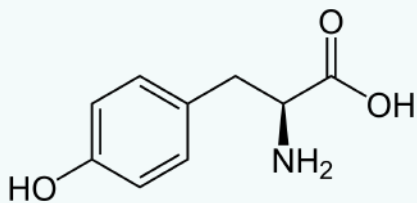
PLANNED START: third/last quarter of 2021

MORE INFO > unswbinfsoc.com/phd-position

AMINO ACID OF THE WEEK

[TYROSINE]

CHEMICAL STRUCTURE



TYROSINE

Y

Tyr

181.19

RNA CODONS

U A Y

POLAR.

DISCOVERED IN 1846 BY J. VON LIEBIG, IN CASEIN
(ISOLATED FROM CHEESE)

PHENOL FUNCTIONALITY: USEFUL IN PROTEINS USED FOR
SIGNAL TRANSDUCTION

NON-ESSENTIAL; CAN BE MADE FROM PHENYLALANINE.

USED IN SYNTHESIS OF NEUROTRANSMITTERS: EPINEPHRINE,
NOREPINEPHRINE, DOPAMINE.

Contact us



IF YOU HAVE ANY COMMENTS or feedback regarding BINFsights, please write to us at binfsights@unswbinfsoc.com

We also encourage anyone to share with us anything you'd like us to take a look at, be it a bioinformatics tool that you have made or find useful; or news in the bioinformatics world that you'd like to see written about in future issues.



TO VIEW PAST AND PRESENT issues of BINFsights, check out our website at unswbinfsoc.com/binfsights

Stay tuned on our Facebook page for updates regarding events and society news.

-- The BINF SOC Team

BINF
sights.