CUSynBioSoc COVID-19 Hackathon

Hope everyone is staying safe and well during this time, and we hope to see everyone - in some form - back in the new academic year. Now that we are out of exam season, we are running a virtual hackathon to keep everyone busy. This document will lay out the guidelines for this hackathon.

WHAT IS A HACKATHON - for first-timers
A hackathon is a competition where teams produce coding projects in a short period of time - usually a weekend. The name is a misnomer - there is no hacking involved! Most large hackathons do not have a focused theme, and teams are free to work on any project they wish. Other hackathons will have a particular theme and will provide datasets for teams to use. These projects are then presented at the end of the period and judged. There are also side-prizes to be won based on using particular sponsoring companies' products or programs. Hackathons are a great way to practice coding and meet up with like-minded programmers!

CONTENT
There are five files provided for analysis from 3 separate projects:
1. Single-cell expression data of patient lung biopsies (prepared by Jarrod Shilts)
2. Bulk-RNA expression data from human cell lines and ferret lungs (Alex Baker)
3. Protein spike mutagenesis data (prepared by Georgeos Hardo)

We recommend that each person in a group works on one or two datasets over the course of the weekend, but teams as a whole are welcome to work on as many datasets as they have time/inclination to. As a Hackathon centred around exploratory data analysis, we encourage teams to get creative in looking at the data through different angles and visualizations. However, if your exploratory work finds something particularly interesting, we would encourage making that result thorough as opposed to trying to pursue as many superficial analyses as possible.

To that end, the overall goal is simple: to see how much you can uncover about the biology of the novel coronavirus in the time available. The "background info" documents offer some suggestions for questions to look at to get started, but these are only meant to help give a starting point and you can answer other scientific questions you can think of. The literature on COVID-19 is rapidly growing but mostly based on young papers without any replication, so even if you later find that what you discovered has already been proposed, that's still a valuable finding!

TEAMS
We have assembled four teams of around 5 people to give each of you a group to work together with and help support each other. We have tried to match up experienced and non-experienced people into teams. Be aware that some of your teams will be operating across time zones! Each team will have their own private Slack channel for communicating.

TIMING
The GitHub link to the files will be shared with the teams shortly before the competition start.

The competition will start at **12 noon BST, Saturday 20 June**.

The competition will finish at **12 midnight BST, Sunday 21 June**.

The judges will announce the results and superlatives by the latest on **26 June.**

SUBMISSIONS
We are expecting two pieces of work for your submission:
- A Google Slides file which lays out your analysis with results and some discussion of what findings you made, within the further discussion in the notes section if needed.
- The code used to generate the graphs and analysis within your Slides file - these can be emailed to us or you can submit a pull request to the Github.

RESTRICTIONS
Feel free to use papers, blogs, StackOverflow, Biostars etc. for code and ideas as long as entire scripts are not copied verbatim.

CONTACT
If you have questions about the datasets, then feel free to post a Slack message to the Hackathon_hq channel where a member of our team can answer it. If your question relates to some discovery you've made that you'd like to keep private, you can instead send a Slack private message or email.

Jarrod Shilts: js2320 (at) cam.ac.uk
Alex Baker: ab2529 (at) cam.ac.uk
Georgeos Hardo: gh464 (at) cam.ac.uk
Anamay Shetty: ahs53 (at) medschl.cam.ac.uk
Rui Li: rl574 (at) cam.ac.uk
Kavi Shah: khs37 (at) cam.ac.uk

The blue-ribbon panel will be using the following framework to rank projects (note that we appreciate "4" is a high bar, which is why we set it to our maximum score):

Scientific value
1. Little work at a very basic level (e.g. submitting a couple simple bar graphs on the most obvious analyses)
2. Relatively limited but showing some understanding of relevant areas (e.g. pursuing basic questions to limited detail)
3. Good work which implements analysis beyond the most obvious questions, showing creativity
4. Excellent work which makes new discoveries of genuine scientific merit, which could even warrant follow-up by COVID-19 researchers

Technical rigor
1. Missing key elements for evaluating the results, like ignoring experimental replicates and showing neither the replicates nor error bars
2. Plots show basic attempts at conveying the data's certainty and presenting whether different plots/analyses agree with each other or disagree
3. Makes good use of independent lines of analysis or multiple datasets to support findings
4. All key results and claims are backed by high-quality validation, including statistical testing with correct procedures as would be typically found as part of a publication

Presentation
1. Poor presentation and no code provided
2. Clear presentation and code provided
3. Excellent presentation and code provided
4. Excellent presentation and fully reproducible code provided

In the spirit of normal hackathons, there will also be the following side-prizes:
● The *Hadley Wickham Prize* for 'coolest' `ggplot2`/`plotnine` graph (as defined by Anamay)
● The *Heebie Jeebies Prize* for the finding that makes our judges feel most afraid about the virus
● The *Weasel Award* for best contribution towards improving the health of weasels (based on analysis of the ferret dataset)
● The *Element of Surprise Award* for the most random and out-of-the-box analysis that actually yielded a meaningful result