

# hail

## Sequencing and Introduction to Hail: Practical Session Primer

2021 Virtual Workshop on Statistical Genetics Methods for Human Complex Traits

June 16<sup>th</sup>, 2021 (practical)

*Hosted by the Institute for Behavioral Genetics, University of Colorado, Boulder*

Kumar Veerapen, PhD  
*Hail Support and Community Outreach Manager*  
Tim Poterba, Carolin Diaz, Dan Howrigan, John  
Compitello, Cotton Seed, Dan Goldstein



<https://hail.is>  
@mkveerapen / @hailgenetics  
veerapen@broadinstitute.org  
#scalableGenomics  
#hailGenetics #ATGUstrong

# Learning Objectives

- Learn and see basic Hail capabilities
- Run a simple GWAS on sequencing data in Hail
- Learn about resources to start using Hail on your own data

# Learning Outline

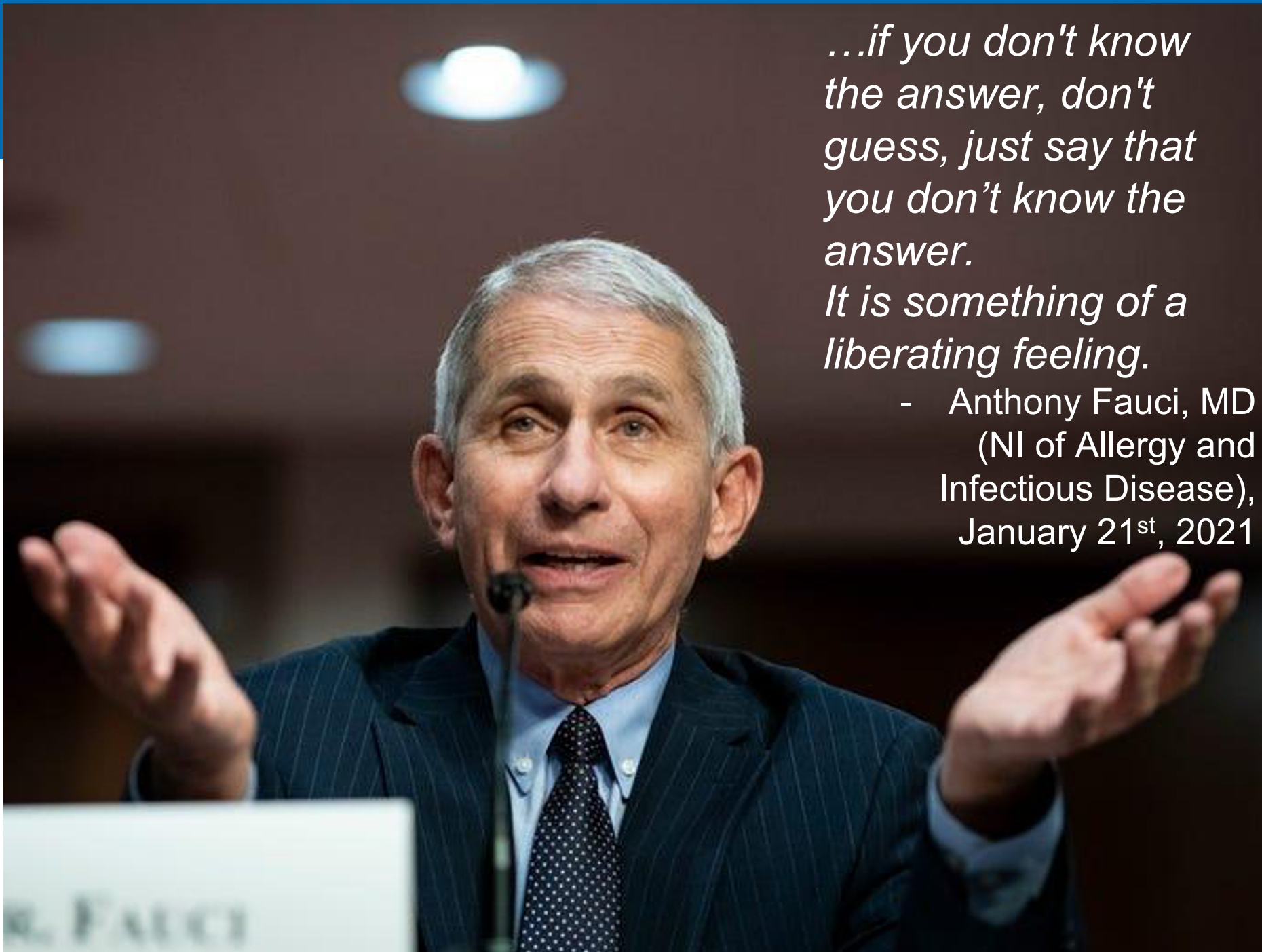
Recap and intro (10 minutes)

Logging into the notebook workspace (5 minutes)

Breakout rooms: GWAS practical using Hail (1.5 hours)

Main room discussion (15 minutes)

Closing remarks (5 minutes)



*...if you don't know  
the answer, don't  
guess, just say that  
you don't know the  
answer.*

*It is something of a  
liberating feeling.*

- Anthony Fauci, MD  
(NI of Allergy and  
Infectious Disease),  
January 21<sup>st</sup>, 2021

# Who are your Tutors?

Instructor: [Kumar Veerapen, PhD](#)

Instructor: [Tim Poterba](#)

Tutors:

Session A	Session B
Dan Goldstein	Dan Goldstein
Tim Poterba	Tim Poterba
Carolin Diaz	Carolin Diaz
Cotton Seed	Cotton Seed
Dan Howrigan	Kumar Veerapen
John Compitello	John Compitello

When poll is active, respond at [Pollev.com/kumar2021](https://Pollev.com/kumar2021)

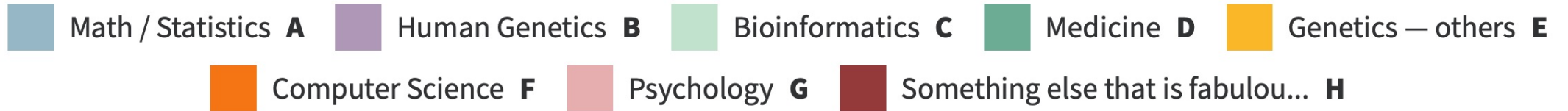
## Where are you from?



Respond at [PollEv.com/kumar2021](https://poll-ev.com/kumar2021)

Text **KUMAR2021** to **37607** once to join, then **A, B, C, D, E...**

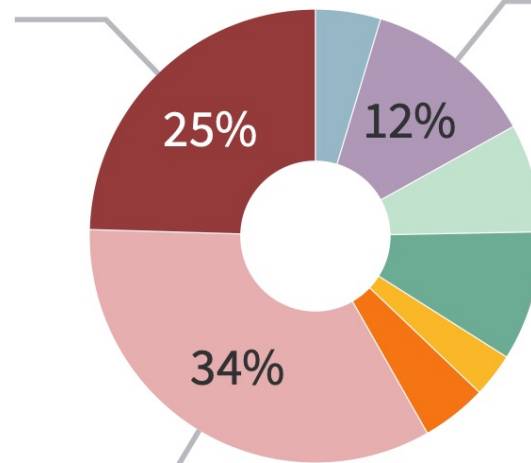
# What is your academic background?



Something else that is fabulous

Human Genetics

Psychology



# Hail vs PLINK (and other cmd line tools)

- You could run most of the tasks covered in this workshop in Hail!
  - Some exceptions, e.g. SEM, SAIGE
  - Steeper learning curve!
- If PLINK works for you, use it!
- If you find yourself:
  - needing `dp1yr` / `dataframes` for genomics...
  - with sequencing data...
  - with **BIG DATA...**

... reach for Hail!



# What is Hail?

*“On a scale from zero to dplyr, the Hail 0.2 interface scores an 8/10 for general-purpose data analysis.” - Konrad K., lead analyst, gnomAD*

Open-Source Data  
Science Library

Slice, dice, query,  
and model any  
kind of data

Scalability

Easy to use with  
both small and  
**biobank-scale**  
genomic data

Unified Genomic  
Data Representation

The MatrixTable is a  
single interface for  
working with all kinds  
of genomic data

Community

Forum and chatroom  
for people interested in  
thinking + talking  
about genomic data  
analysis

The logo for Hail, featuring the word "hail" in a blue, lowercase, sans-serif font. The letter "i" is stylized with a DNA double helix structure integrated into its vertical stroke.

Learn more at [Hail.is](https://hail.is)

**\*We can't read your  
minds, so talk to us**  
[discuss.hail.is](https://discuss.hail.is)

# Hands on using workshop.hail.is workshop name: `ibg2021` password: `IBG`


1. Introduce yourselves  
(even if you have met, give each other a smile and wish them well 😊)
2. Elect a note taker and time-keeper
3. Elect a screen sharer
4. Elect a troubleshooter
5. ASK FOR HELP!  
If you have faculty in your room, they are learning with you!



Activities <

< 2021\_IBG

 Delete

 Visual settings

 Edit

 Present



My activities



Shared activities

 When poll is active, respond at [Pollev.com/kumar2021](https://Pollev.com/kumar2021)

## How did you feel about the Hail practical?



# Workshop service


- The Hail workshop service will stay on for another 2 hours
- If you want to keep going after that, you can install Hail on your own system!



Activities <

< 2021\_IBG

 Delete

 Visual settings

 Edit

 Present



My activities



Shared activities

 When poll is active, respond at **PollEv.com/kumar2021**

 Text **KUMAR2021** to **37607** once to join

## Which section did you feel best about?

1. Using Jupyter
2. Import and initialize Hail
  - Importing and using HGDP
  - Annotation of HGDP
  - Interrogating SNP distribution
  - Sample and Variant QC
  - Association testing
  - PCA

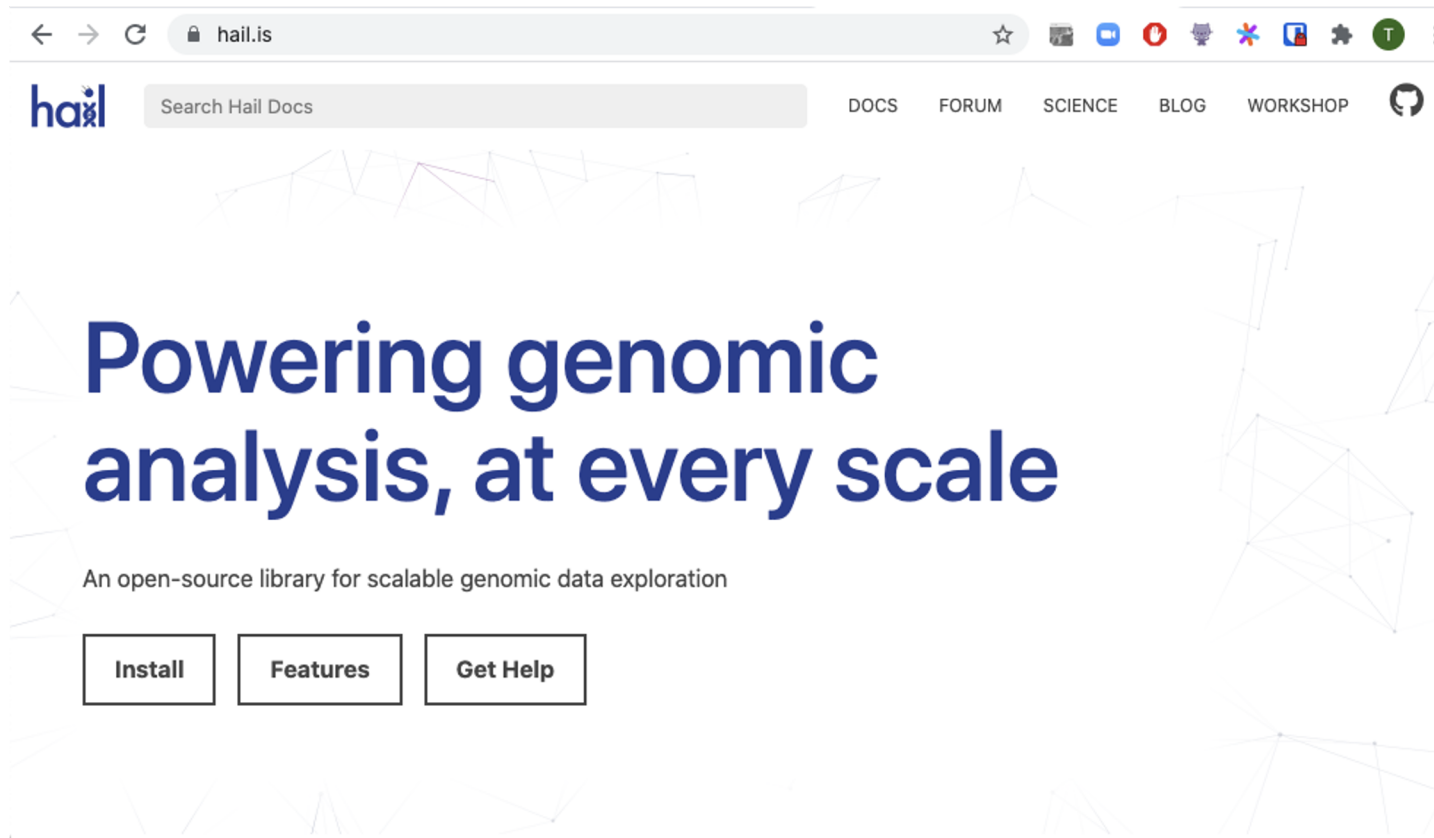
# Hail on various hardware

- Laptop/Desktop  
development, small data (10s of genomes, 100s of exomes)
- Server  
medium data (1Ks of genomes, 10Ks of exomes)
- High Performance Computing (HPC) cluster  
large (1M genomes, 10M exomes)
- Cloud  
large (1M genomes, 10M exomes)

# Your next steps

```
pip install hail
```

Learn more at the website: <https://hail.is>



The screenshot shows the Hail website homepage. The browser address bar displays 'hail.is'. The website header includes the Hail logo, a search bar labeled 'Search Hail Docs', and navigation links for 'DOCS', 'FORUM', 'SCIENCE', 'BLOG', and 'WORKSHOP'. The main content area features a large blue heading: 'Powering genomic analysis, at every scale'. Below this heading is the text 'An open-source library for scalable genomic data exploration'. At the bottom of the main content area, there are three buttons: 'Install', 'Features', and 'Get Help'.

Visit the forum!



The screenshot shows the Hail discussion forum page. The browser address bar displays 'discuss.hail.is'. The page header includes the Hail logo, a search bar, and buttons for 'Sign Up' and 'Log In'. Below the header, there are links for 'About', 'FAQ', 'Terms of Service', and 'Privacy'. The main content area features a heading 'About Hail Discussion' and a paragraph of text: 'Discussion forum for Hail, an open-source, scalable framework for exploring and analyzing genomic data (https://hail.is)'.

Chat with the user and developer community!

<https://hail.zulipchat.com>



- All messages 1
- Private messages
- @ Mentions
- ★ Starred messages
- Recent topics

