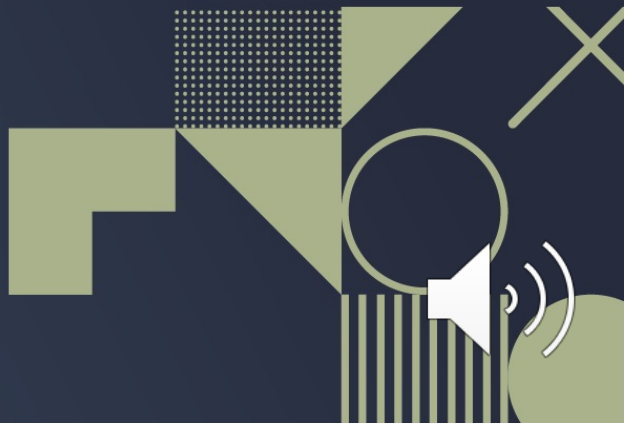
HPC Response to the COVID-19 Pandemic

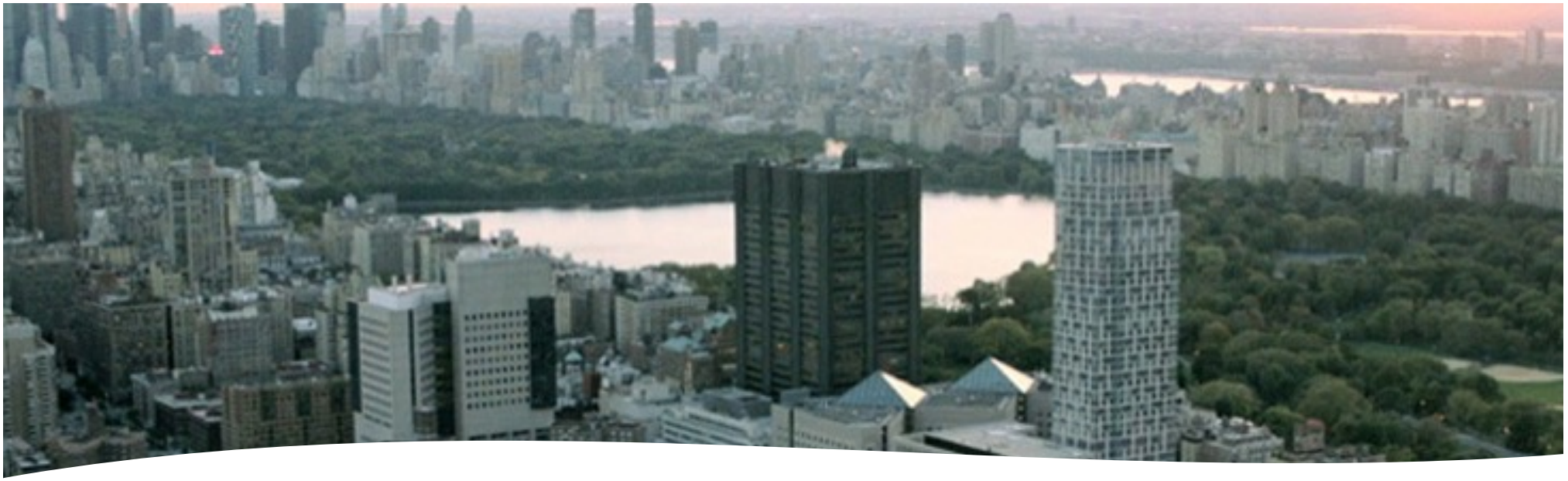
Patricia Kovatch
Dean for Scientific Computing and Data
Professor, Genetics and Genomic Sciences and Systems Pharmacology
Icahn School of Medicine at Mount Sinai



About me

- Founded Scientific Computing and Data at Mount Sinai in October 2011
- Oversee HPC, clinical data warehouse, research data services
  - Director for the NIEHS-funded Human Health Exposure Analysis Resource Data Repository and Management Core
  - Co-Director for the NIA-funded Alzheimer's Disease Research Center Data Management and Statistical Core
  - Continuously funded by NIH since 2014
- Enjoying over 30 years in High Performance Computing (HPC)
  - Deployed world's 1st academic petaflop and world's 3rd fastest machine in 2009
  - Directed over 15 top 10 supercomputers for the National Science Foundation
  - Directed National Institute for Computational Sciences for the University of Tennessee at Oak Ridge National Laboratory

Icahn School of Medicine at Mount Sinai



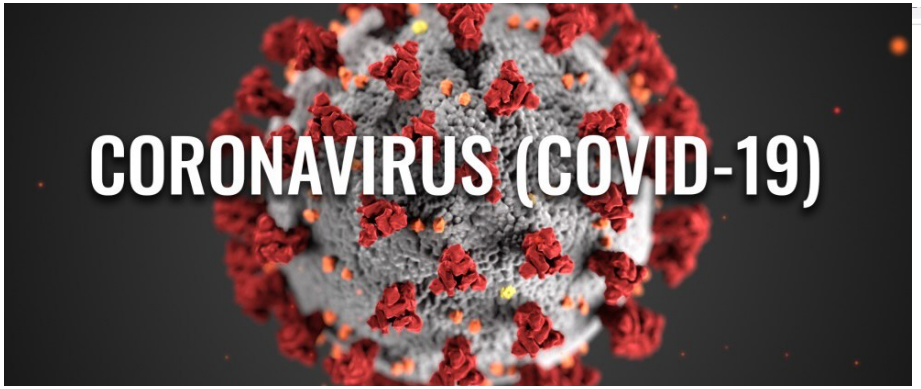
The Mount Sinai Health System

- School of Medicine (1968) and eight hospitals (1820)
  - Both ranked in the top 20 by US News & World Report*
- 6th largest employer in NYC (42,000 employees)
- >$8B/year revenue
  - Over 8 million patients, 7,480 physicians, 2,000 residents/fellows
  - 3,815 hospital beds and 3.5 million outpatient visits/year
  - Over 18,000 babies born each year
  - 147 operating rooms plus 47 additional specialty ORs

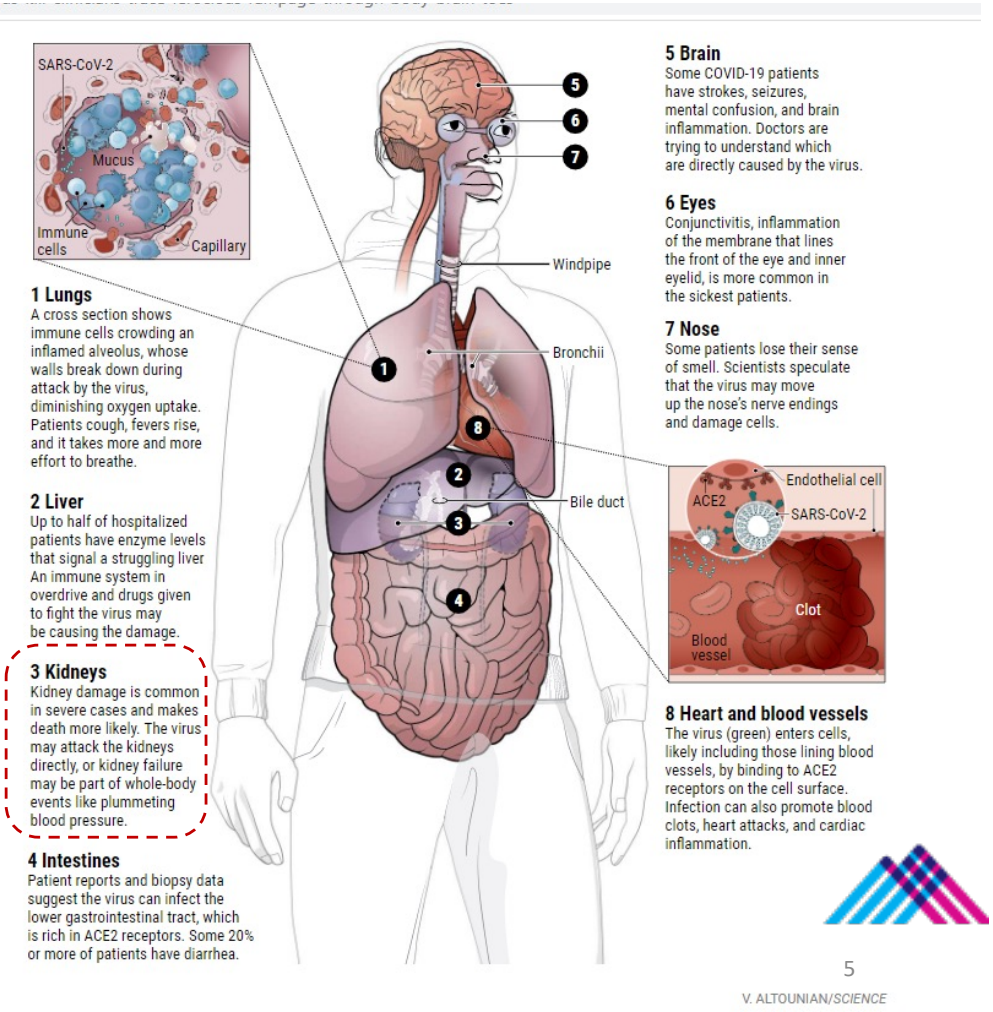
Icahn School of Medicine at Mount Sinai

Scientific Computing and Data Mission

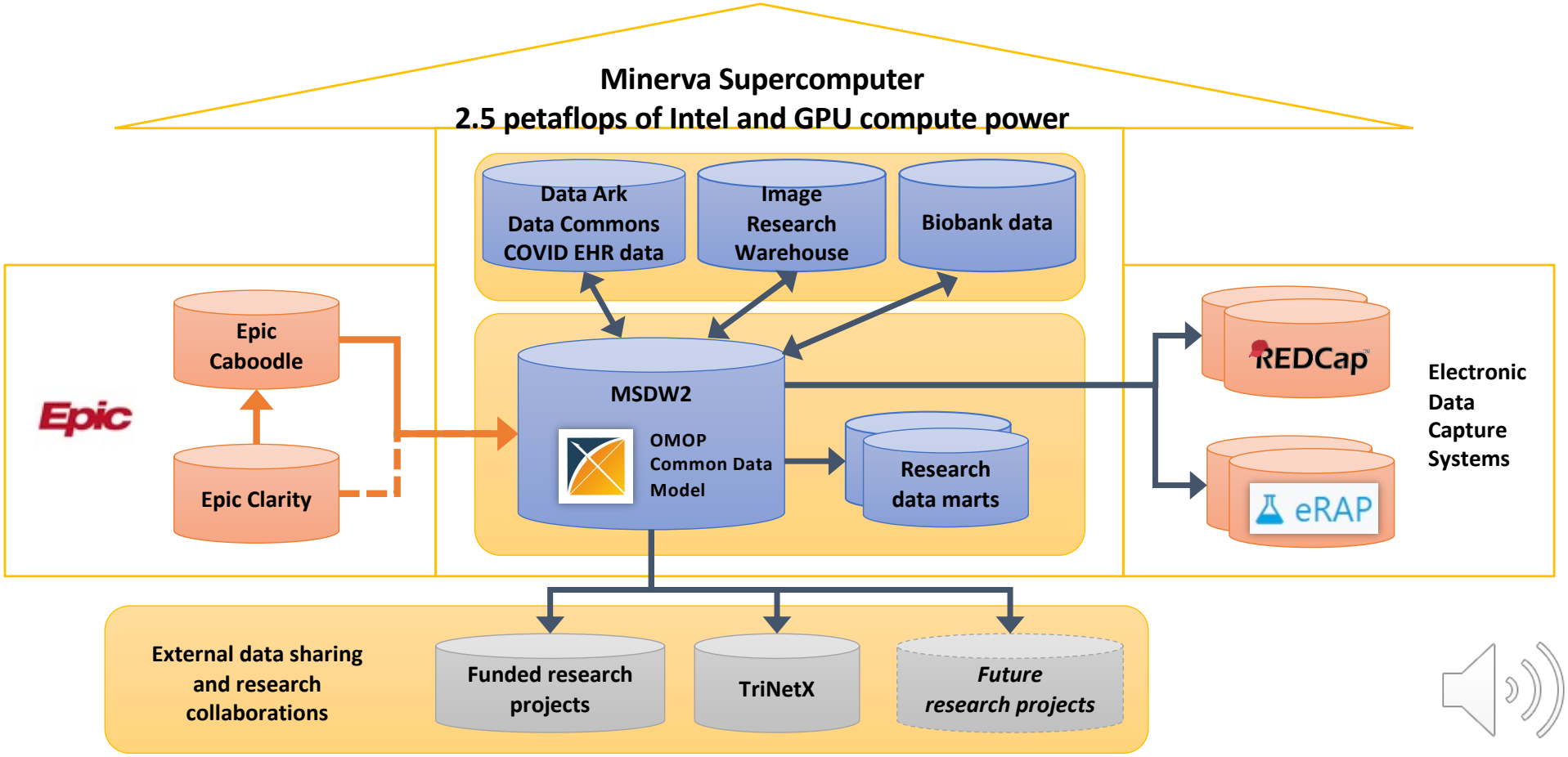
To enable and accelerate biomedical discovery through collaborative research using high performance computing, biomedical informatics and data science

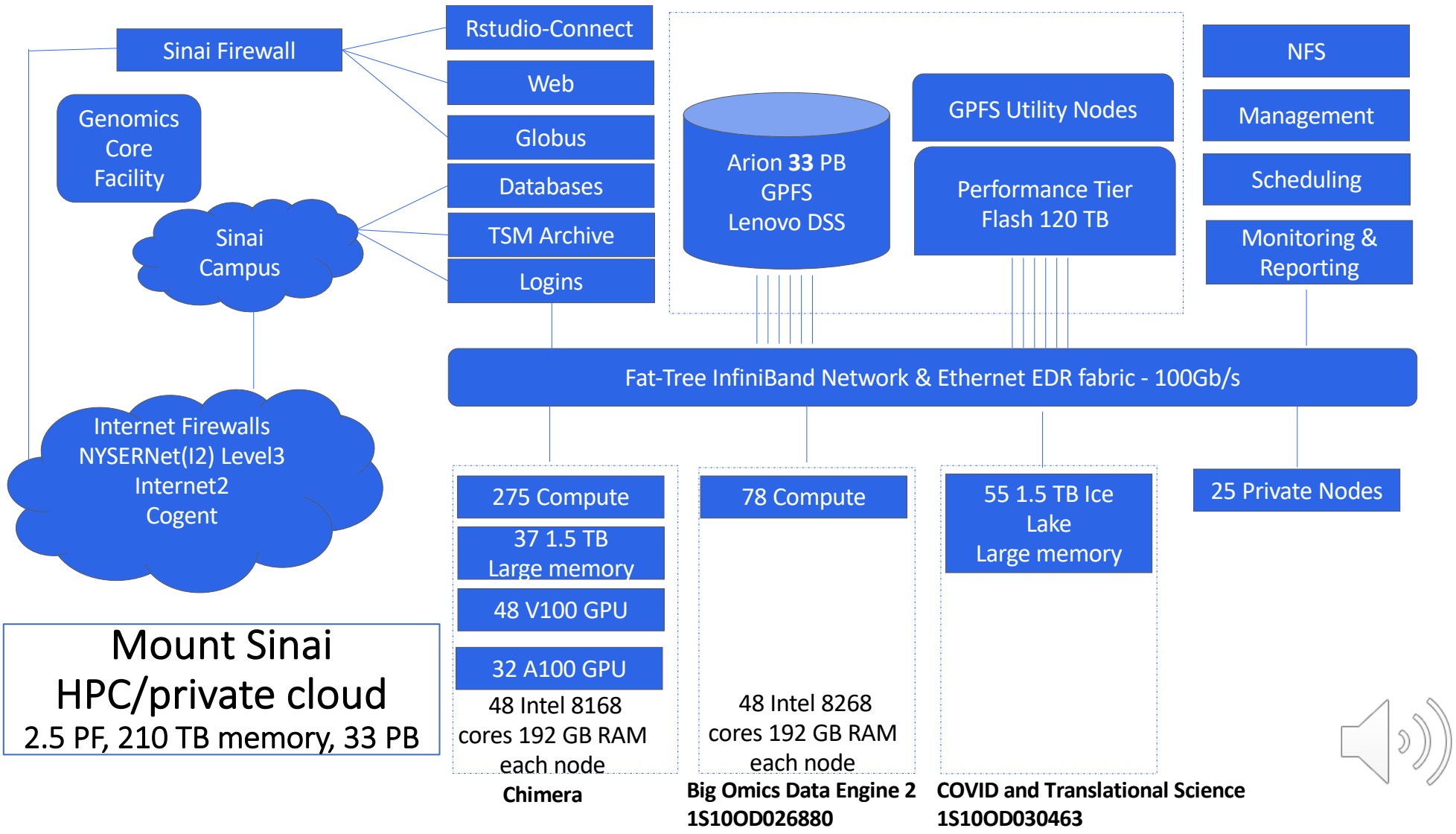


SARS-CoV-2 is Devastating to Numerous Organ Systems

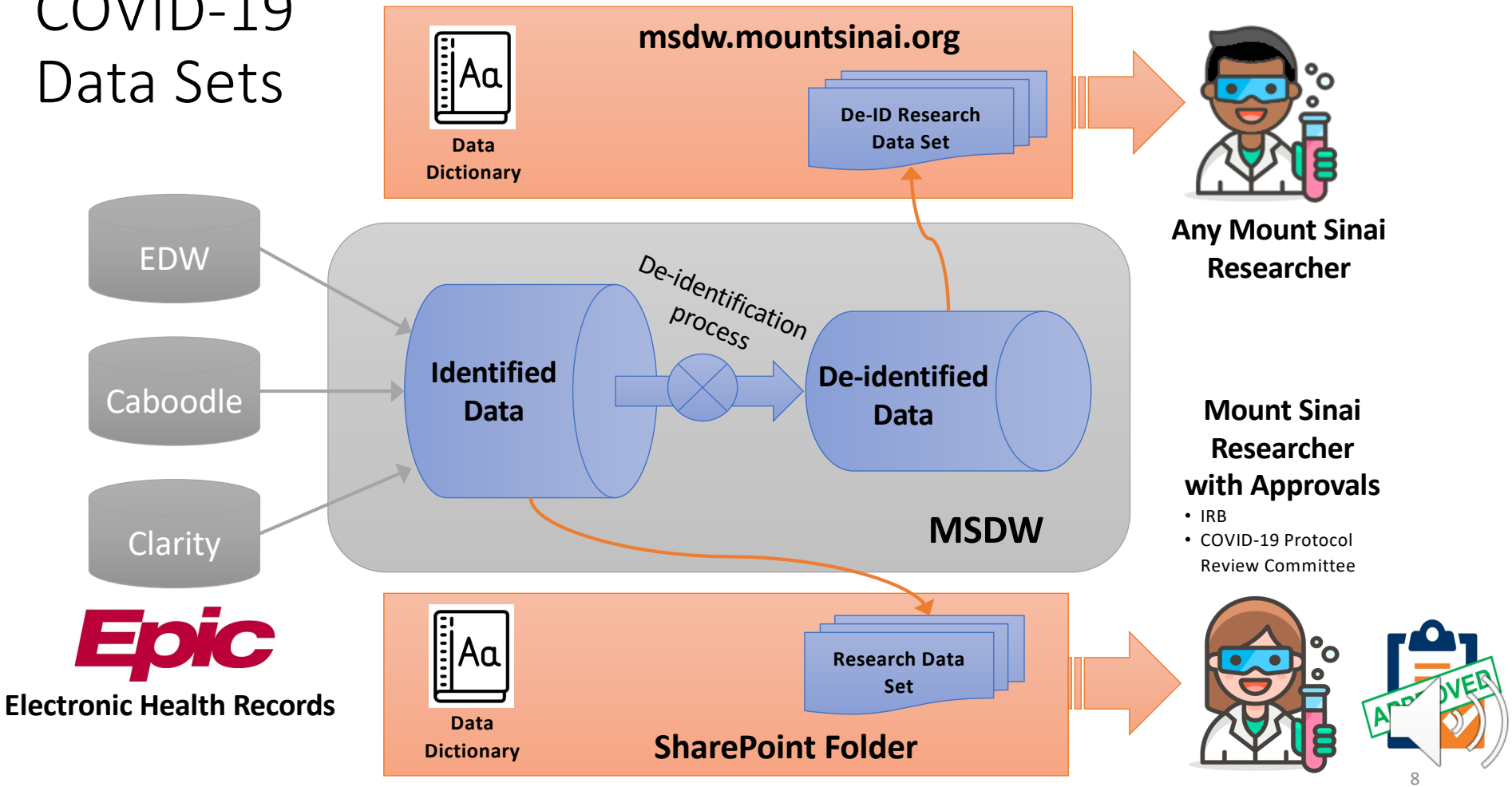


Integrated Computational and Data Ecosystem




Sinai Firewall
Genomics Core Facility
Sinai Campus
Internet Firewalls
NYSERNet(I2) Level3
Internet2
Cogent
Rstudio-Connect
Web
Globus
Databases
TSM Archive
Logins
Arion 33 PB GPFS Lenovo DSS
GPFS Utility Nodes
Performance Tier Flash 120 TB
NFS
Management
Scheduling
Monitoring & Reporting
Fat-Tree InfiniBand Network & Ethernet EDR fabric - 100Gb/s
275 Compute
37 1.5 TB Large memory
48 V100 GPU
32 A100 GPU
48 Intel 8168 cores 192 GB RAM each node
78 Compute
48 Intel 8268 cores 192 GB RAM each node
55 1.5 TB Ice Lake Large memory
25 Private Nodes
Mount Sinai HPC/private cloud
2.5 PF, 210 TB memory, 33 PB
Chimera
Big Omics Data Engine 2
1S10OD026880
COVID and Translational Science
1S10OD030463

COVID-19 Data Sets



Data de-identification for US Health Insurance Portability And Accountability (HIPAA) Privacy Rule Compliance

- **Process for COVID-19 Data Sets**
  - Mask or remove all 18 Personal Health Identifiers (PHI)
  - Birth date converted to age at encounter, with all ages over 89 set to 90
  - Convert all dates to elapsed days relative to encounter date
  - Zip codes truncated to 3 digits (or masked if population < 20k)
- **Process for Other medical record data sets**
  - Same as above, but all dates for each patient are shifted by a random integer to preserve chronology

1. Patient names
2. Geographic elements like address street, city, county, zip code (first 3 digits allowed)
3. All elements of any dates (except year), including dates of birth, admission, discharge, death, and exact ages over 89
4. Telephone numbers
5. Fax numbers
6. E-mail addresses
7. Social Security Number (SSN)
8. Medical Record Number (MRN)
9. Health insurance beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers, including license plate numbers
13. Device identification and serial numbers
14. Digital identifiers, including Uniform Resource Locators (URLs)
15. Internet Protocol (IP) addresses
16. Biometric identifiers, including finger, retinal, and voiceprints
17. Full-face photographic images and other comparable images
18. Any other uniquely identifying numbers or codes

Scientific Computing
De-identified COVID-19 Research Data Set

Over 7,000 downloads at Mount Sinai since April 2020

Data Sources: Epic, NYSDOH Wadsworth Lab
Inclusion Criteria: Visit Type *OR* Encounter Dx *OR* Lab Test Order/Result related to COVID-19
Purpose: Research and clinical questions

Patient Encounters

- **Demographics**
  Age, Sex, Race, Ethnicity, Zip Code (3 digits), Preferred Language
- **Outcomes**
  Infection Status, Mortality
- **Comorbidities/Outcomes**
  Asthma, COPD, CKD, HTN, Obesity, DM, HIV, Smoking, Cancer, CAD, CHF, OSA, A Fib, Liver Disease, AKI, VTE, Cerebral Infarction
- **Encounter**
  Encounter Type, Visit Type, Patient Class, Department, Location of Care, Admission, Discharge

Vital Signs

- *BMI*
- *Max Temp*
- *Heart Rate*
- *Respiratory Rate*
- *Max Blood Pressure*
- *Min Blood Pressure*
- *Min O2 Saturation*

Lab Tests

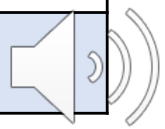
- *SARS-CoV-2 PCR, Antibody Assay*
- *IL-6, IL-8, IL-1 Beta*
- *TNF Alpha*
- *D-Dimer*
- *Ferritin*
- *LDH*
- *Fibrinogen*
- *C-Reactive Protein*
- *Procalcitonin*
- *WBC, RBC*
- *Blood Culture*
- *AST*
- *ALT*
- *Alk Phos*
- *Sodium*
- *Potassium*
- *eGFR*
- *Hemoglobin*
- *BUN*
- *Creatinine*
- *Lymphocyte*
- *Eosinophil*
- *Basophil*
- *Monocyte*
- *Neutrophil*
- *Platelet*
- *Hematocrit*
- *Troponin*
- *aCL Antibody*
- *B2GPI Antibody*
- *Phospholipid Antibody*
- *ESR*
- *BNP*
- *MCV, MCH, MCHC*
- *MPV*
- *PTT, PT, INR*
- *Albumin*
- *Calcium*
- *Uric Acid*
- *Bilirubin*
- *pH*
- *pCO2, pO2*
- *O2 Sat*
- *HCO3*
- *CK, CK-MB*
- *Troponin I*

Medication Administrations

- *Tocilizumab*
- *Remdesivir*
- *Sarilumab*
- *Hydroxychloroquine*
- *Anakinra*
- *Azithromycin*
- *Rivaroxiban*
- *Apixaban*
- *Enoxaparin*
- *tPA*
- *Heparin*
- *Etanercept*
- *Nitric oxide*
- *Dopamine*
- *Vasopressin*
- *Norepinephrine*
- *Epinephrine*
- *Milrinone*
- *Dobutamine*
- *Phenylephrine*
- *Methylprednisolone*
- *Prednisone*
- *Dexamethasone*

Select COVID-19 Data Set Research Projects

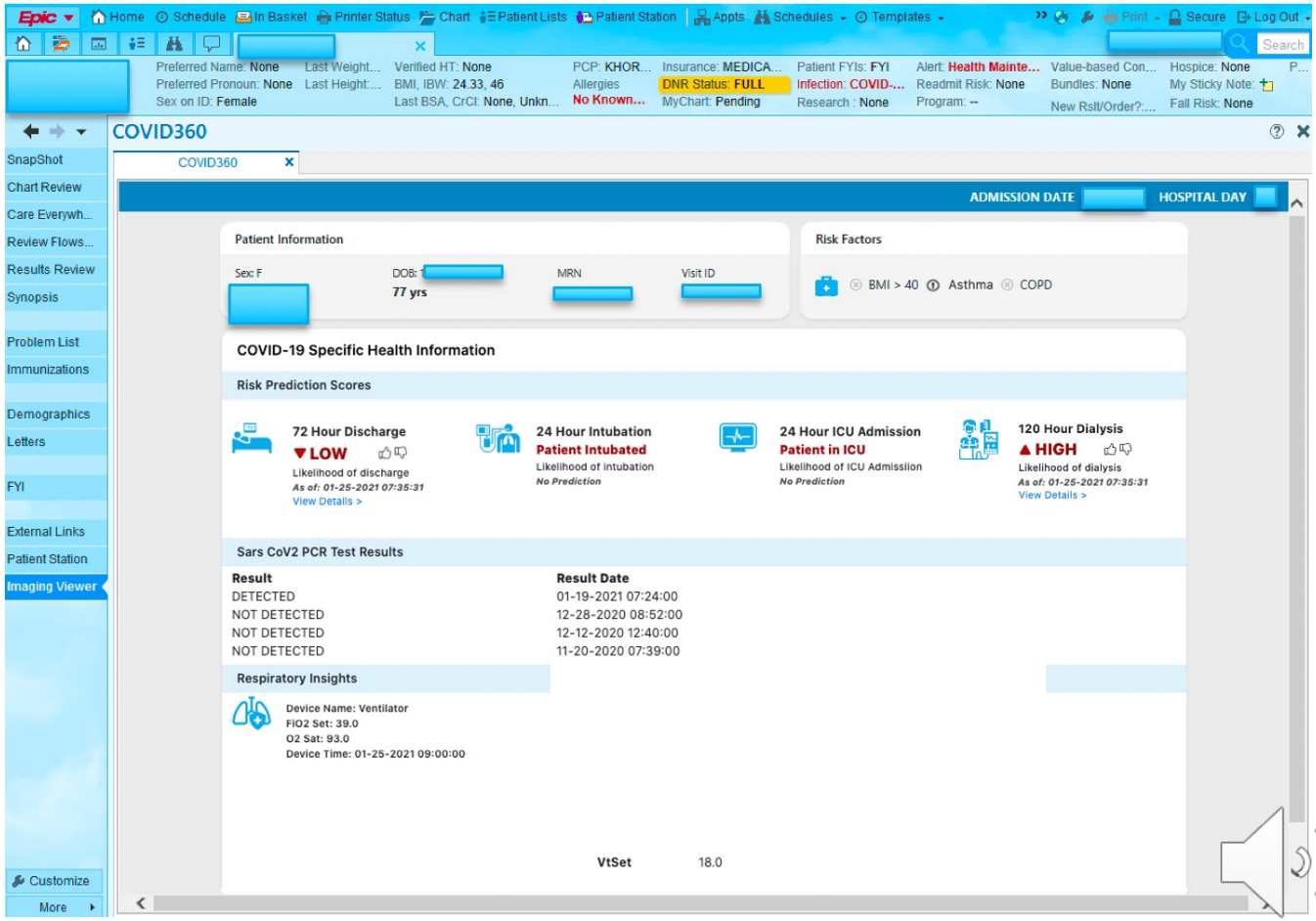
Genetics	Social and genetic determinants for COVID-19 disease susceptibility	Noura Abul-Husn, MD, PhD Eimear Kenny, PhD
	COVID-19 Predictive Modeling & Patient Tool	Girish Nadkarni, MD Benjamin Glicksberg, PhD
	Modeling of COVID Course of Care	Matthew Levin, MD
Emergency Medicine	National NHLBI ED COVID registry with PETAL/ CORAL	Lynne Richardson, MD
Cardiology	Cardiac arrhythmias in patients with COVID-19	Vivek Reddy, MD
Population Health	CDRN COVID-19 database for INSIGHT	Carol Horowitz, MD, MPH
Internal Medicine	Determining COVID-19 Natural History and Predicting Adverse Clinical Outcomes	Juan Wisnivesky, MD DrPH Keith Sigel, MD
Ob/Gyn	COVID perinatal quality database	Elizabeth Howell, MD



COVID360 Real-time Dashboard

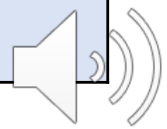
Goal

To combine electronic medical record information, risk factors, and the output of a predictive model into an easy-to-use tool for rapid insights



Select COVID-19 Research Projects using HPC

Tracking & Survelliance	Nextstrain COVID-19 tracking	Harm Van Bakel
	Computational support for HPI COVID-19 initiative	Girish Nadkarni
Drugs & Therapeutics	Systems to identify and validate novel therapeutics for COVID-19	Bin Zhang
	Drug repurposing and crowdsourcing for COVID-19	Avi Ma'ayan
Patient Outcomes	Provide a web application to analyze & predict COVID-19 patient outcomes	Towfique Raj
	Fast-tracking COVID-19 detection test by fluidigm microfluidics technology	Stuart Sealfon
	Gene Burden Analysis of BioMe™ Donor Whole Exome Sequencing Data	Yuval Itan
Molecular Dynamics	COVID-19 viral replication and host immune response	Ming-Ming Zhou, Martin Walsh, Marta Filizola, Adolfo Garcia-Sastre
	Small Molecule Inhibitors Blocking Virus-ACE2 Interactions	Marta Filizola, Jian Jin, Florian Krammer, AdolfoGarcia-Sastre



Coronavirus in N.Y.: Manhattan Woman Is First Confirmed Case in State

A woman in her 30s who had traveled in Iran contracted the virus and is now isolated at home.

The New York Times

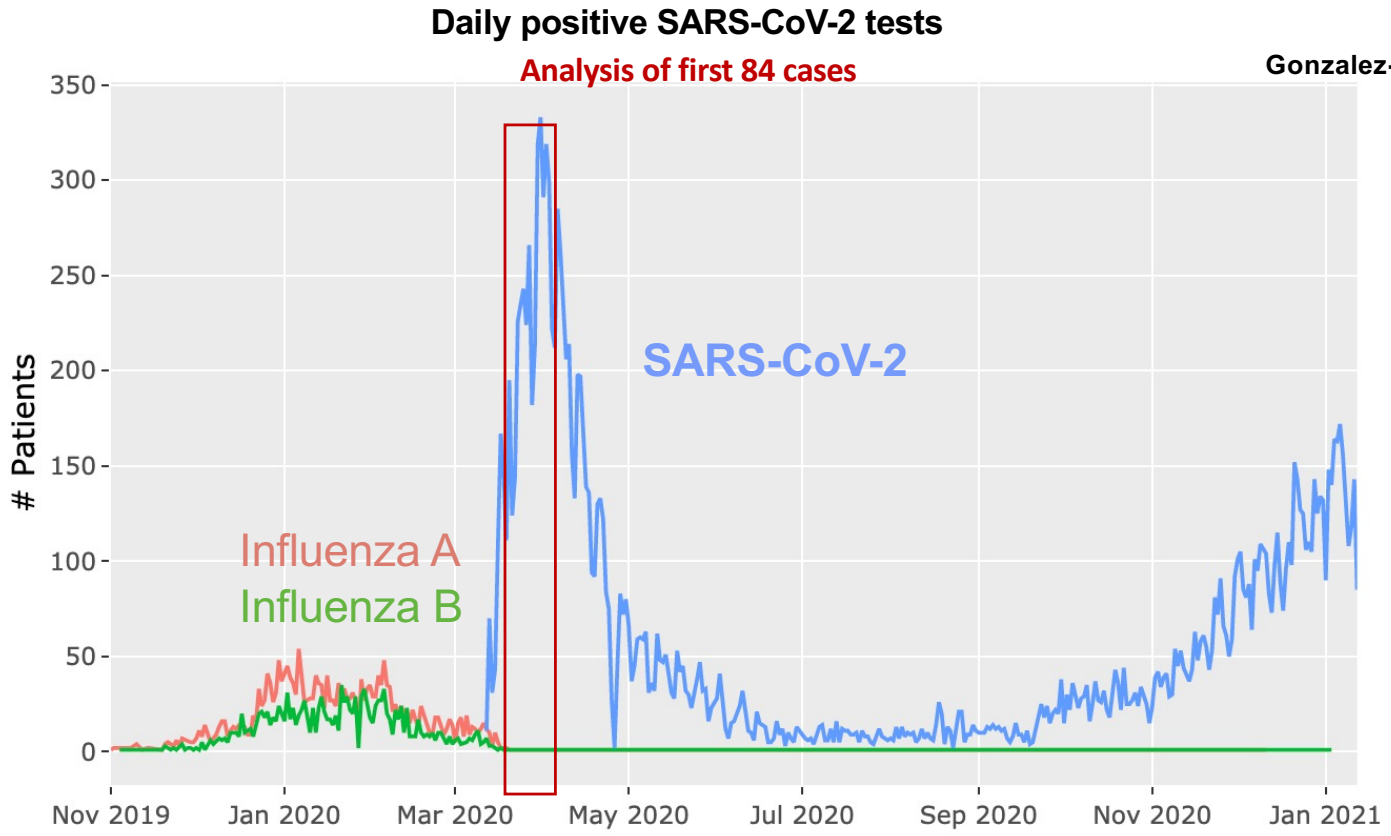
Early SARS-CoV-2 transmission in New York City linked to Europe

We analyzed the viral genomes of these cases and compared them to the background sequences deposited by other countries across the world in a massive global sequencing effort.

The picture that emerged from there is that the initial introduction of the virus into NY mainly came through Europe where there had been a large peak during a time when travel restrictions were not in place yet.

Gonzalez-Reiche & Hernandez e.. *Science*, 2020

Genomic surveillance of SARS-CoV-2 across the Mount Sinai Health System



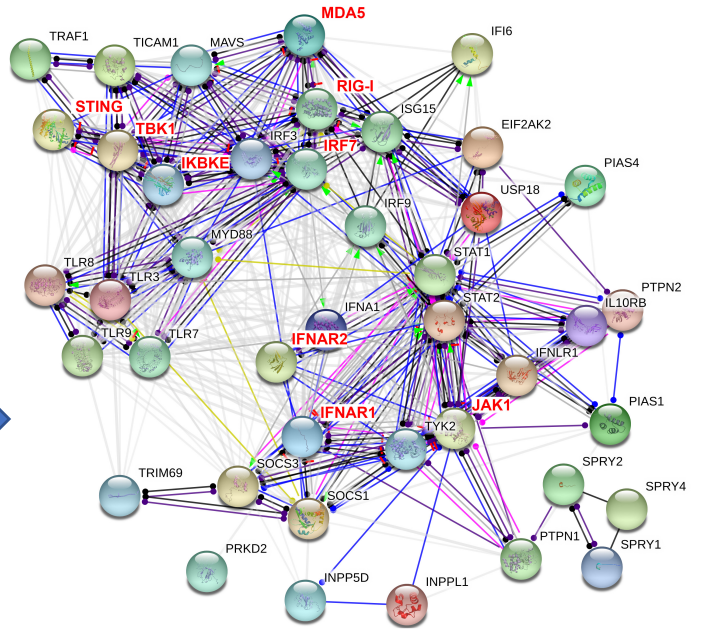
This shows a plot of the positive COVID cases along with a typical flu season to give a perception of the scale of the pandemic; within weeks we had as many cases as normally seen in an entire flu season.

COVID-19 gene burden analyses in BioME biobank

- Severe (hospitalized+ventilated): 252
- Mild (non-hospitalized): 464
- Negative: 1,268
- Now >2,500 COVID-19-positive in BioMe

p-SKAT(COVID_all vs BioMe)HM_impact	p-SKAT(COVID_hosp vs 1000G)High_impact
p-Fisher's(COVID_all vs BioMe)HM_impact	p-Fisher's(COVID_hosp vs 1000G)High_impact
p-SKAT(COVID_all vs BioMe)High_impact	p-SKAT(COVID<50 vs 1000G)HM_impact
p-Fisher's(COVID_all vs BioMe)High_impact	p-Fisher's(COVID<50 vs 1000G)HM_impact
p-SKAT(COVID_hosp vs BioMe)HM_impact	p-SKAT(COVID<50 vs 1000G)High_impact
p-Fisher's(COVID_hosp vs BioMe)HM_impact	p-Fisher's(COVID<50 vs 1000G)High_impact
p-SKAT(COVID_hosp vs BioMe)High_impact	p-SKAT(UKBB_hosp vs UKBB_nonhosp)HM_impact
p-Fisher's(COVID_hosp vs BioMe)High_impact	p-Fisher's(UKBB_hosp vs UKBB_nonhosp)HM_impact
p-SKAT(BioMe_hosp vs BioMe_nonhosp)HM_impact	p-SKAT(UKBB_hosp vs UKBB_nonhosp)High_impact
p-Fisher's(BioMe_hosp vs BioMe_nonhosp)HM_impact	p-Fisher's(UKBB_hosp vs UKBB_nonhosp)High_impact
p-SKAT(BioMe_hosp vs BioMe_nonhosp)High_impact	p-SKAT(UKBB_allpos vs 1000G)HM_impact
p-Fisher's(BioMe_hosp vs BioMe_nonhosp)High_impact	p-Fisher's(UKBB_allpos vs 1000G)HM_impact
p-SKAT(COVID<50 vs BioMe)HM_impact	p-SKAT(UKBB_allpos vs 1000G)High_impact
p-Fisher's(COVID<50 vs BioMe)HM_impact	p-Fisher's(UKBB_allpos vs 1000G)High_impact
p-SKAT(COVID<50 vs BioMe)High_impact	p-SKAT(UKBB_hosp vs 1000G)HM_impact
p-Fisher's(COVID<50 vs BioMe)High_impact	p-Fisher's(UKBB_hosp vs 1000G)HM_impact
p-SKAT(COVID<50 vs nonhosp_BioMe)HM_impact	p-SKAT(UKBB_hosp vs 1000G)High_impact
p-Fisher's(COVID<50 vs nonhosp_BioMe)HM_impact	p-Fisher's(UKBB_hosp vs 1000G)High_impact
p-SKAT(COVID<50 vs nonhosp_BioMe)High_impact	p-SKAT(UKBB_hosp vs UKBB_neg)HM_impact
p-Fisher's(COVID<50 vs nonhosp_BioMe)High_impact	p-Fisher's(UKBB_hosp vs UKBB_neg)HM_impact
p-SKAT(COVID_hosp vs 1000G)HM_impact	p-SKAT(UKBB_hosp vs UKBB_neg)High_impact
p-Fisher's(COVID_hosp vs 1000G)HM_impact	p-Fisher's(UKBB_hosp vs UKBB_neg)High_impact

- Analyses performed at the Yuval Itan lab:
  - SKAT-O analyses done by Yiming Wu
  - HGC simulations done by Cigdem Sevim Bayrak



SKAT-O P=1.32E-6

HGC P<1.0E-7

- 43 Interferon genes (IFN, with Dusan Bogunovic): $P=1.32\times10^{-6}$
  - $P=3.04\times10^{-4}$ in UKBB
- 21 ACE pathway genes: 3.26×10^{-6}
  - $P= 2.91\times10^{-5}$ in UKBB

There is a group of genes that contain relatively more mutations in severe COVID patients when compared to controls, which were identified by a method called gene burden analysis.

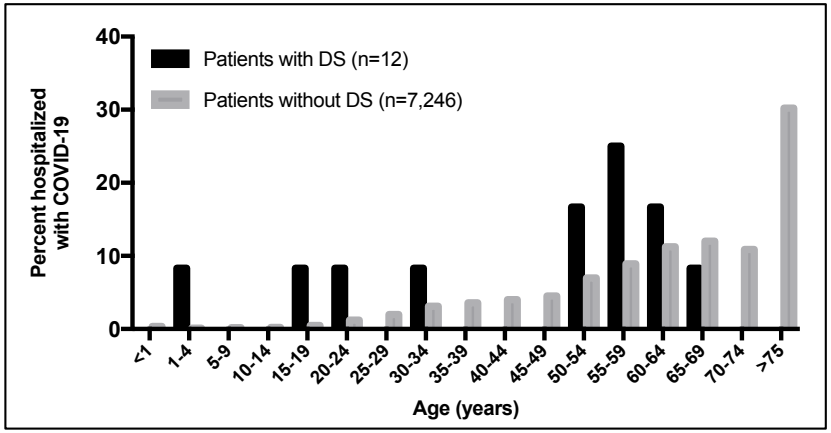
Impact of COVID-19 on individuals with Down syndrome (DS)

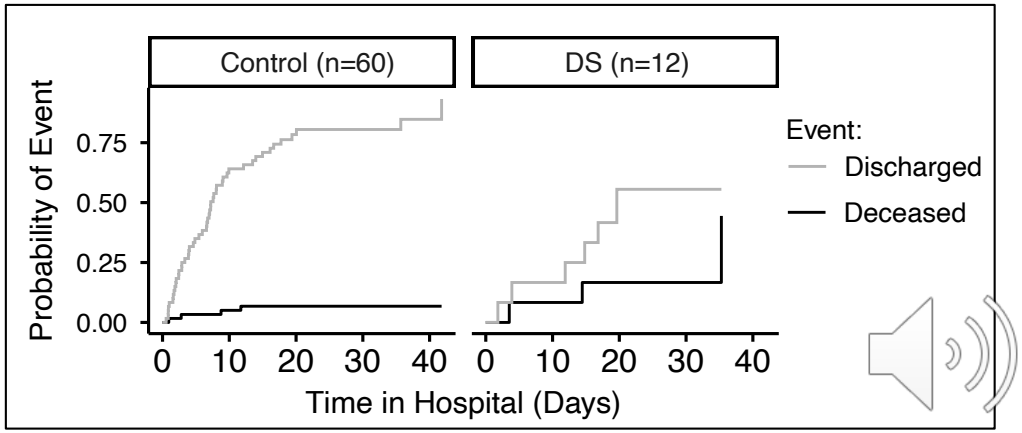
- In patients who were hospitalized with COVID-19, those with DS were on average 10 years younger than those without DS
- Patients with DS had more severe disease than age-matched controls, particularly an increased incidence of sepsis and mechanical ventilation
- An appreciably higher proportion of patients died in the DS group than in the control group

BRIEF COMMUNICATION Genetics inMedicine

Individuals with Down syndrome hospitalized with COVID-19 have more severe disease

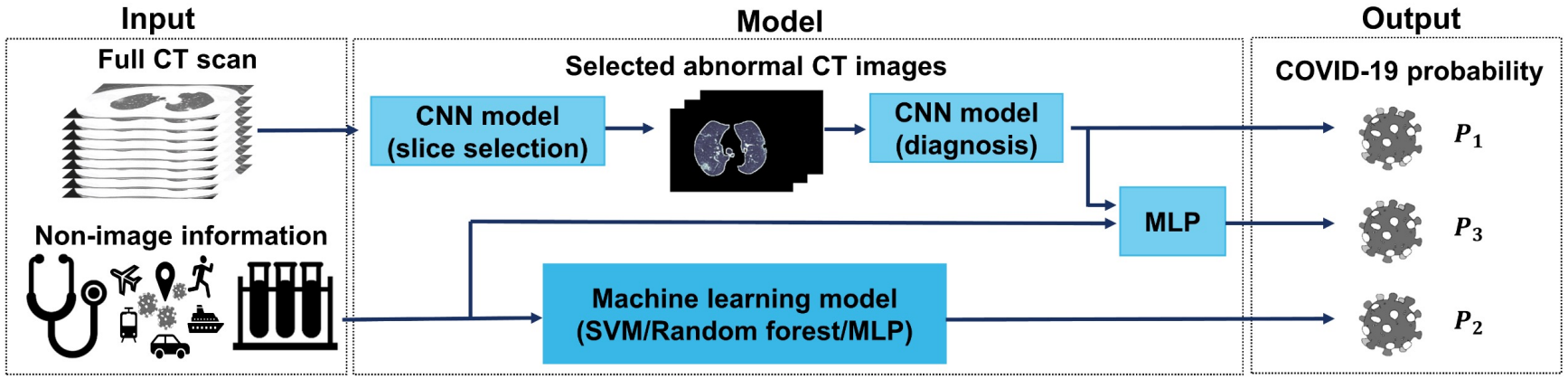
Louise Malle, BA[1,2,3,4], Cynthia Gao, BS[5], Chin Hur, MD, MPH[6], Han Q. Truong, BS[6], Nicole M. Bouvier, MD[1,5], Bethany Percha, PhD[5], Xiao-Fei Kong, MD, PhD[6] and Dusan Bogunovic, PhD[1,2,3,4,7]

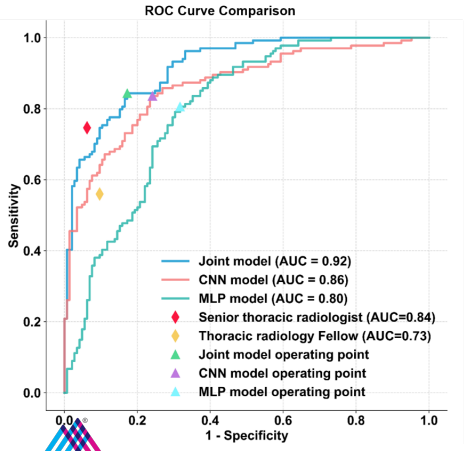




Artificial intelligence-enabled rapid diagnosis of COVID-19 patients

N=905

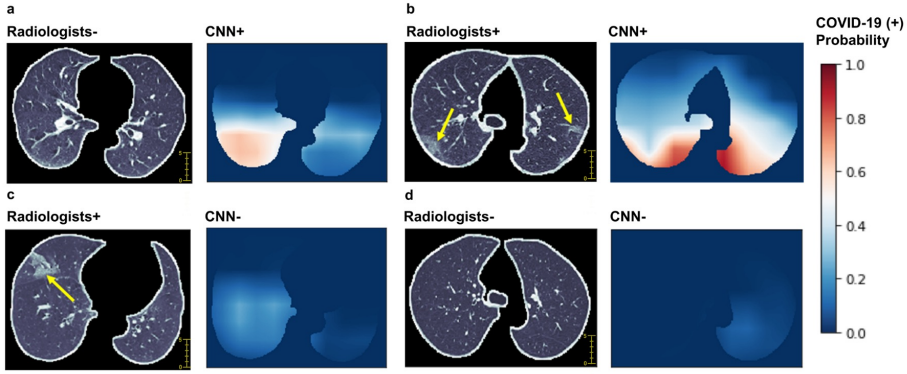




Icahn School of Medicine at Mount Sinai

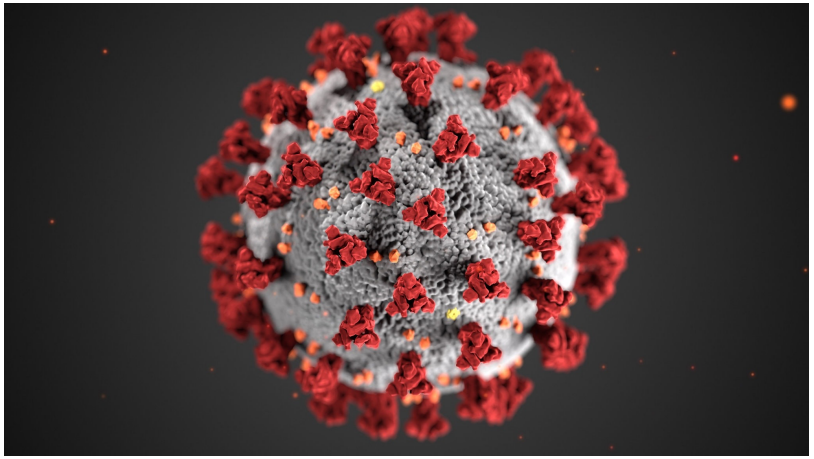
BioMedical Engineering and Imaging Institute

- The AI system achieved an area under the curve of 0.92 and had equal sensitivity as compared to a senior thoracic radiologist.
- The AI system also improved the detection of patients who were positive for COVID-19 via RT–PCR who presented with normal CT scans, correctly identifying 17 of 25 (68%) patients, whereas radiologists classified all of these patients as COVID-19 negative.
- When CT scans and associated clinical history are available, the proposed AI system can help to rapidly diagnose COVID-19 patients (faster than RT-PCR).



Mei, X; Yang Y; Fayad ZA et al. Nat Med 0

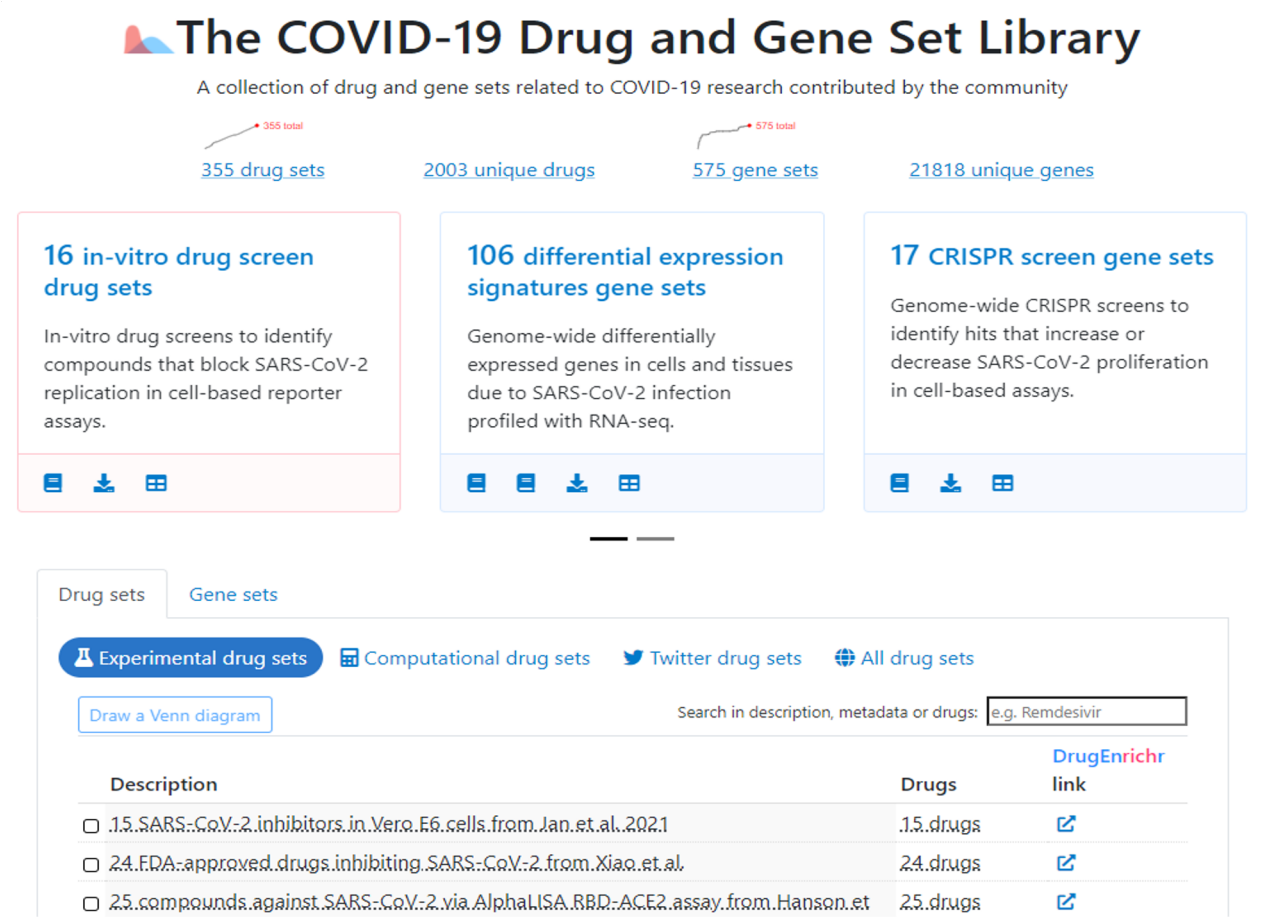
National Institutes of Health (NIH) funded R01 DA04547-02S1: "A structure-based drug discovery effort to combat COVID-19 and future outbreaks"



Aim 1. To identify potential virus-ACE2 disruptors among FDA-approved drugs.

Aim 2. To identify novel chemotypes as backup or additional virus-ACE2 interface disruptors.

Collaboration with Marta Filizola, Adolfo Garcia-Sastre, Jian Jin, and Florian Kramer



Kuleshov MV, Stein DJ, Clarke DJB, Kropiwnicki E, Jagodnik KM, Bartal A, Evangelista JE, Hom J, Cheng M, Bailey A, Zhou A, Ferguson LB, Lachmann A, Ma'ayan A. The COVID-19 gene and drug set library. *Patterns* 1(6):100090 (2020) PMID: 32838343

https://maayanlab.cloud/covid19/

Thank you!

- **We have open positions!!**
  - **HPC Architect**
  - **Computational and Data Ecosystem Manager**
  - **Clinical Data Specialist**
  - **Bioinformatician**
  - **Application Support Specialist**
  - **Healthcare SQL Analyst**
  - **ETL Developer**

careers.mountsinai.org