Impact of Host Genomics on COVID-19: What We Know and Don’t Know

Priya Duggal, PhD, MPH
Johns Hopkins Bloomberg School of Public Health
July 23, 2020
CDC
Epidemiology and COVID-19
WHO informed of a cluster of cases of pneumonia of unknown cause in Wuhan City, Hubei Province, China

Pathogen Identified as novel coronavirus

Announced novel coronavirus. Virus is SARS-CoV-2, associated disease is COVID-19

Pulmonologist Nanshan Zhong announces human to human spread

Wuhan placed under quarantine

Early COVID-19 Timeline
Transmission

• This is a respiratory illness.
• Spread through droplets which includes close contact with people who are sick, including talking!
• It also resides on surfaces and may be airborne, both can impact exposure to another person even if they are not in the same room at the same time.
• Asymptomatic or pre-symptomatic transmission does occur. This means, unlike influenza, you cannot rely on coughing or sneezing to indicate someone is transmitting.
Typical Symptoms 2-14 days after exposure

- Fever
- Chills
- Cough
- Shortness of Breath
- Repeated Shaking with Chills
- Muscle Pain
- Headache
- Sore Throat
- New Loss of Taste or Smell
- Diarrhea
Typical severe presentation requiring hospitalization

Day 1: Patient may have mild symptoms

Day 5: Patient has difficulty breathing

Day 8: Patient may develop an acute respiratory distress syndrome (fluids builds up in lungs) 15% of cases will develop this.

Day 10: Disease progression may require ICU admittance for breathing. But also other manifestations (heart, kidney, stroke etc)
Asymptomatics & Mild Infections

- Some asymptomatic individuals have organ involvement but with no symptoms (see lungs)
- Mild infections report fatigue, shortness of breath 3 months post infection

"I suspect that, if you followed up with these asymptomatic people in several months, most of their CT scans would be completely normal unless they were known to later develop symptoms," Taylor-Cousar says.

In the midst of an infection, however, doctors can’t predict how an individual’s case will progress. "There’s no way to know who is going to stop with an asymptomatic infection and likely recover completely and who is likely to go on to more severe infection," Schluger says.
Multisystem Inflammatory Syndrome in Children

- Rare (n=1000)

- Temporally associated with COVID-19

- Median age 7.5-11.6 years in US, French and Italian reports

- Male: 62-70%

- ~70% previously healthy

- 20% received invasive mechanical ventilator support

- ~70% had 4 organ system involvement
But it's clear there is nothing typical about COVID-19

“I have never seen infection in which you have such a broad range literally no symptoms at all in a substantial proportion of the population to some who get ill with minor symptoms to some who get ill enough to be in bed for weeks,”

“Others get hospitalized, require oxygen, intensive care, ventilation and death. The involvement with the same pathogen is very unique.”

Anthony Fauci, July 21, 2020

...except disease heterogeneity
Host Genetics and COVID-19
Central Dogma of Genetics

DNA → RNA → Protein

Cells use DNA that codes genes to create proteins (Myoglobin, Collagen, etc)

These proteins work together and interactively to facilitate cell function
DNA, our building blocks

The human genome contains ~3 billion nucleotides
(= 200,000 page phone books)

Slides courtesy of NIH
Any two humans have 99.9% genetically identical sequence...so why are we studying our genes?

This results in 1 difference every 1000 nucleotides, resulting in ~3 million nucleotide differences between any 2 people.
How to Find Genes Responsible for Human Disease

Difficult challenge to find a disease gene—like finding a misspelled word in a set of encyclopedias!

“which chromosome?”

“chromosomal region”

“This is a sentence in a paragraph...”

“This it a sentence in a paragraph...”

“gene”

“mutation”
Genetic Epidemiology Approach

• Genotype individuals using large scale arrays or targeted candidate genes (Receptors, Cytokines etc)

• Sequence individuals (exome or whole genome)—ideal to get multiple affected families

• Compare polymorphisms (SNPs/SNVs) across the genome with case/control status

• We are looking for associations that may give us information on disease pathogenesis
How do we do genetics?

One example: Genome Wide Association Study

Two-stage design

**Stage 1:** 500K Illumina panel in 1,700 cases and 4,000 controls
  - Cochran-Armitage test for trend

**Stage 2:** replication in an independent cohort of 3,500 cases and 4,800 controls of the 384 most associated SNPs

If we can understand why people have different outcomes---we will better understand the immune response and eventually exploit this understanding for therapeutics and vaccines.

Our goals...are still to *better understand disease* so we can prevent and treat.
Things we look for to see if genetics might be playing a role…

• Disease Heterogeneity
• Familial aggregation
• Pathogen Dose and Environment
Reported Risk Factors for Severe Disease

- Older Age
- Male Sex
- Comorbidities
- Race
And yet…

Older people (even 102 years of age) recover.

Younger people require mechanical ventilation and even die with no comorbidities.

Race appears to be a risk factor—but as a social risk factor not biologic.
Things we look for to see if genetics might be playing a role....

- **Disease Heterogeneity**
  - Not explained by other risk factors (sex, age, comorbidities)

- **Familial aggregation**
  - Hard to assess with infectious diseases because of transmission

- **Pathogen Dose and Environment**
  - not major players (yet)
Driven by 2 main questions

Do you get infected?  
Do you get disease?
What makes infectious diseases different?

EXPOSURE
We need to know who was exposed.

Need careful and comprehensive characterization of infection and disease.

CONTROLS matter. If you call someone uninfected you need to know they were exposed & still uninfected.
Do you get infected (susceptibility)?

• This requires *detailed* information on the population under study
• We need to know the non-infected were exposed, and yet still did not get infected

**Study Design**

• Focus is typically on *highly exposed* individuals where you can document or infer that individuals were exposed
  
  • Health Care Workers/Essential Workers
  • Household members of known cases
  • Intensely followed cohorts (currently none for COVID-19)
Who gets infected? Who doesn’t?

>24 people in attendance, 14 infected at the party

13 sisters died, 18 others infected, 65 total in convent
Do you get disease (severity)?

- This is about the heterogeneity in disease.

**Study Design**

- We select cases and controls from among individuals who ALL have the infection
  - asymptomatic to severe/death
  - sample on distribution or extremes
- Otherwise, you are comparing a severe case to someone who may have never had the opportunity to become a severe case.
COVID-19 Clinical Spectrum

• 80% of cases are mild not requiring hospitalizations
  • Long term sequela?
  • Spectrum of symptoms

• Hospitalized cases heterogeneity
  • Oxygen needs
  • Penumonia
  • Organ involvement: Lungs, Hearts, Liver, Kidney, etc

• Children
  • Mild infections
  • Hospitalized Infections with oxygen needs
  • Multi-Inflammatory syndrome

• No free pass. Although some have no real outcomes. Infection across the severity spectrum associated with organ involvement and potentially long term sequelae.
Tissue specific expression and genetic regulation of SARS-CoV-2 receptors ACE2 and TMPRSS2

Yuan He, Marios Arvanitis, Princy Parsana, Ashton Omdahl, Jessica Bonnie, Zeyu Chen, Christopher D. Brown, Alexis Battle

- Rare variants identified across populations
- But, no direct evidence of binding-resistant mutants in different populations
- ACE2 has no significant cis-eQTLs in tissues with high ACE2 expression, including lung, heart, kidney, and small intestine
- ACE2 is predicted to be highly loss of function variant intolerant indicating little effect of genetic variation on ACE2 function or expression

Fig. 1 The coding-region variants and eQTL variants for ACE2 in East Asian and other populations. a Schematics of 52 coding variants in ACE2 identified in the Cohort and HGF databases. Yellow areas indicate the nonsense variants; blue indicate the missense variants. The number of samples with hotspots variants was marked. b The distribution of hotspot missense mutations of ACE2 in different populations. The colors indicate different populations, the distribution and the allele frequencies of representative eQTL variants for ACE2 in different populations. No charts depict the allele frequencies of an insert variant of ACE2 (c.428_429insC) in the world. Orange color denotes the frequency of alteration allele, and blue color denotes the reference allele. The allele frequencies of 19 eQTLs for ACE2 gene are shown in Tables. The color gradient from blue to red indicates the increasing of allele frequency. The allele frequencies of RLDs, variants 78930661-1 were annotated by the gnomAD database. SAS, East Asian; SAS European; APA, African; SAS, South Asian; NHS, Ad Mixed American.

Correspondence Open Access

Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations

Yuan He, Marios Arvanitis, Princy Parsana, Ashton Omdahl, Jessica Bonnie, Zeyu Chen, Christopher D. Brown, Alexis Battle
Cases: Hospitalized COVID+ (n=835 Italy/775 Spain)
Controls: Population based blood donors (1255 Italy/950 Spain)

*no information on comorbidities or treatment
The COVID19-HGI in numbers  https://www.covid19hg.org/

- 1128 members
- 69,255 unique users on the website
- 243,118 page view

- 201 studies from 51 countries
- 16 (+3) "contributing studies"

Scientists
Andrea Ganna and Mark Daly

Studies

Countries that have shared results:

UK (4 studies), Iceland, Italy (2), Spain, Netherlands (2), Finland, USA (2), Qatar, Brazil, Belgium, Scotland, Korea
What are we doing at Johns Hopkins?

Main Aim to identify genes that are associated with severe or mild disease.
We need to be driven by the question.

<table>
<thead>
<tr>
<th>What else should we look at?</th>
</tr>
</thead>
<tbody>
<tr>
<td>We need to be driven by the question.</td>
</tr>
</tbody>
</table>

| Rare cases of particular outcomes (MIS-C, clotting, kidney damage) |
| What makes these individuals different? |

| Associations with cytokines, inflammatory markers. |
| Can we identify early associations with biomarkers? |

| Associations with humoral and T cell immunity? |
| Does genetics influence the antibody or T cell response to infection or reinfection? |

| Pharmacogenetics for clinical trials. |
| Does host genetics alter clinical trial meds? |

| Linking genetics with genomics. |
| Evaluating genetics, epigenetic, transcriptomics and others to tell a more complete story |

| Associations with viral pathogen. |
| Is there host-pathogen interaction? |
Acknowledgements

**Genes and ID research team at JHU Epidemiology**

*PhD Students:* Cristian Valencia, Dylan Duchen, Rebecca Munday, Steven Clipman  
*Faculty:* Poonum Korpe, Ana Valencia (visiting), Candelaria Vergara, Genevieve Wojcik

**COVIDGene Team**

Cristian Valencia, Leon Hsieh, Dylan Duchen, Rebecca Munday, Steven Clipman, Candelaria Vergara, Genevieve Wojcik, Andrea Cox, Chloe Thio, Poonum Korpe, Benjamin Larman, Shruti Mehta, Hannah Manley, David Thomas

**Funding**

NIAID, JHU Provost Award