Update on COVNET (Genetic Determinants of COVID-19 Outcomes & Susceptibility)

COVNET Team
Division of Cancer Epidemiology and Genetics

May 27, 2021
Three NCI/NIH COVID-19 genetic efforts in progress

1. **COVNET**: a large-scale study of the germline genetics of COVID-19 susceptibility and manifestations
   - 40,000 GWAS & 5,000 Whole Genome Sequencing (WGS) of COVID-19 cases

2. **COVIDcode Study**: NIH Clinical Center IRB approved study of 2,500 cases
   - Collaboration between investigators in NHGRI, NCI, and NIAID
   - Genetic analyses and extensive Immunologic assessment

3. **NCI COVID-19 in Cancer Patients Study (NCCAPS)**: Prospective Study of COVID-19 in Cancer Patients
   - Target accrual of 2,000 patients with biospecimens
   - Follow-up and survivorship evaluations by early 2022
Clinical course of COVID-19 is highly variable

### Clinical Manifestations

<table>
<thead>
<tr>
<th>A</th>
<th>Asymptomatic</th>
<th>B</th>
<th>Mild</th>
<th>C</th>
<th>Moderate</th>
<th>D</th>
<th>Severe</th>
<th>E</th>
<th>Critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence of clinical manifestations</td>
<td>Fever, dry cough, sore throat, runny nose, sneezing, fatigue, myalgia, tiredness, muscle pain, headache, and smell loss.</td>
<td>Symptoms of mild to moderate pneumonia: fever (usually persistent, &gt;37.8°C), dry cough, dyspnea, fast breathing, ( \text{SpO}_2 \geq 90% ) on room air.</td>
<td>Dyspnea, hypoxia, &gt;50% lung involvement, diarrhea, vomiting, nausea, and one of the following manifestations: respiratory rate &gt;30 breaths/min; severe respiratory distress; or ( \text{SpO}_2 &lt; 90% ) on room air. Patients show clinical worsening.</td>
<td>Severe shortness of breath, chest pain, movement impairments, loss of speech. Complications: Acute respiratory distress syndrome or respiratory failure, myocardial injury, arrhythmia, heart failure, acute kidney injury, acute liver injury, encephalopathy, disseminated intravascular coagulation, rhabdomyolysis, septic shock, multiple organ dysfunctions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Frequency

- Asymptomatic: 40%
- Mild: 40%
- Moderate: 15%
- Severe: 5%

Factors associated with the clinical course of COVID-19

- Hospitalization and death consistently associated with gender (male), age, obesity, lower socioeconomic status, and select pre-existing health conditions*

- Non-European ancestry

*P-value < 0.001 for the association between each of these factors and COVID-19 death. Similar observations were made in translational studies by Williamson et al., Nature 2021; Shelton et al., Nature Genetics 2021.
COVID-19 cases in the U.S. by age

Percentage of Cases

Age Group (Years) | Percentage | Cases
--- | --- | ---
0-4 | 2% (415,947) | 
5-17 | 9.5% (2,011,703) |
18-29 | 22.4% (4,758,873) |
30-39 | 16.4% (3,470,285) |
40-49 | 14.9% (3,170,090) |
50-64 | 20.5% (4,358,712) |
65-74 | 7.8% (1,648,484) |
75-84 | 4.1% (868,242) |
85+ | 2.4% (516,803) |

Percentage of Deaths

Age Group (Years) | Percentage | Cases
--- | --- | ---
0-4 | <0.1% (91) |
5-17 | 0.1% (252) |
18-29 | 0.5% (1,877) |
30-39 | 1.2% (4,322) |
40-49 | 2.8% (10,465) |
50-64 | 14.5% (53,741) |
65-74 | 21.1% (78,436) |
75-84 | 27.6% (102,357) |
85+ | 32.2% (119,457) |

CDC COVID Data Tracker: [https://covid.cdc.gov/covid-data-tracker/#demographics](https://covid.cdc.gov/covid-data-tracker/#demographics)
COVID-19 cases in the U.S. by age

Shelton et al., Nature Genetics 2021.
**Primary objectives**

- To identify germline susceptibility variants that determine host responses to COVID-19 disease in case-case design.
- Identify common and rare germline variants associated with host susceptibility to severe or fatal COVID-19 disease.

**Secondary objective**

- Exploratory analyses of epigenetic signatures, serologic immune markers and antibody profiles using whole blood collected.
- Collect *(when possible & distributed carefully)*:
  - Whole blood specimens for sera, DNA, RNA
  - Explore B and T cell repertoire
  - Serum or plasma to explore humoral response and soluble mediators
  - RNA sample tubes for transcriptomic analysis
**Principal Investigator:** Les Biesecker, M.D., NHGRI

**Associate Investigators:**
- NCI, DCEG
  - Stephen Chanock, Sharon Savage, Lisa Mirabello, Renee Bremer, Mandy Black, Lisa McReynolds, Margarita Aryavand
- NIAID
  - Stephen Holland, Helen Su, Luigi Notorangelo, and others

---

**Data collection**

- Target COVID-19 patient recruitment: 2500
  - NIH intramural program: CC and OMS participants
  - Extramural collaborators
  - Participants can self enroll
- Uses CGB’s CHARMS (communications hub and research management system)
- Collection of risk factors, rapid oral/on-line questionnaire (30 questions)
  - Extended Questionnaire including more clinical details (150 questions)
COVIDcode progress
Target: 2500

Enrollment by Group

- Group 2 = 621 (78.71%)
- Group 3 = 167 (21.17%)
- Group 1 = 1 (0.13%)

Specimen Counts
- Blood Draw: 468
- Saliva: 153

COVID-19 Diagnoses Over Time
NCI COVID-19 in Cancer Patients Study (NCCAPS)

1. Cohort of cancer patients infected with COVID-19 comprising all age groups for collection of a comprehensive dataset on the cancers, treatments, medications, symptoms, course, and recovery, and co-morbidities with longitudinal follow-up until return to pre-morbid status;

2. Follow subset of pts for >1 yr to assess impact of COVID-19 on survivorship and cancer outcomes;

3. Collect blood samples at study entry and then every 2-3 months for 1 yr to estimate antibody and cellular immune response, genetic susceptibility, coagulation abnormalities, and for biomarker development;

4. Collect imaging and QOL data longitudinally;

5. Public database/biospecimens.
Critical Study Milestones

• **Study opened in late May 2020**
  • 6 weeks from idea to patient entry

• Enroll the first 500 patients within 3 months of trial activation

• Target accrual of 2,000 patients with biospecimens @ Nationwide

• Follow-up and survivorship evaluations by early 2022

• Begin biomarker studies on blood samples soon after initial 500 patients accrued—supported by new Congressional appropriation

• GWAS and NGS on all 2,000 to be done through CGR/DCEG
<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Not Hispanic or Latino</th>
<th>Hispanic or Latino</th>
<th>Not Reported / Unknown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian or Alaskan Native</td>
<td>8</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Asian</td>
<td>16</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Black or African American</td>
<td>123</td>
<td>2</td>
<td>2</td>
<td>127</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>White</td>
<td>820</td>
<td>72</td>
<td>10</td>
<td>902</td>
</tr>
<tr>
<td>Not Reported/Unknown</td>
<td>18</td>
<td>45</td>
<td>6</td>
<td>69</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>990</strong></td>
<td><strong>119</strong></td>
<td><strong>18</strong></td>
<td><strong>1,127</strong></td>
</tr>
</tbody>
</table>
NCI COVID-19 in Cancer Patients Study (NCCAPS)

- NCCAPS pediatric amendment to specifically accrue COVID-19 pediatric patients without the requirement of longitudinal follow-up.
  - Maximize accrual to determine the consequences of COVID-19 on kids with cancer
Germline genetics of COVID-19 susceptibility and manifestations

COVNET weblink: https://dceg.cancer.gov/research/how-we-study/genomic-studies/covnet
Status of COVNET, May 2021

• Over 150 programs/Geneticists in US approached
  • >100 Teleconferences

• MTAs (marker of commitment in US)
  • 20 Signed
  • 5 in process

• Sampled received (as of 5/19/2021)
  • ~9,000 (15 studies)
  • ~4500 expected in coming weeks
  • ~4000 longer term

• Genotyped
  • 3,764 (phenotypes received for 2,654)
COVNET - Current Collaborators

**International Collaborators**
- Innsbruck Medical University, Austria
- Seoul National University Hospital, Korea
- Academy of Athens, Greece
- Koc University, Istanbul, Turkey
- NRCRM, Kiev, Ukraine
- Latin American countries (next slide)

**Internal Collaborators**
- COVNET
- UCD
- UCSD
- CDC-Utah
- KP-Colorado
- UAB
- Emory
- MGH/Harvard
- Weill Cornell
- CDC-Columbia
- UPenn
- MSKCC
- COVIDcode
- NIAID
- NCI studies

NIH NATIONAL CANCER INSTITUTE
COVNET Latin America
Large-scale genome-wide association study and whole genome sequencing of COVID-19 severity

May 2021

LATIN AMERICA - COLLABORATORS

<table>
<thead>
<tr>
<th>Country</th>
<th>Investigator</th>
<th>Planned sample size</th>
<th>Current samples</th>
<th>MTA</th>
<th>Updates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chile/Argentina/Guatemala/Colombia</td>
<td>Luis A. Quiñones, Matias Olguín</td>
<td>1,000</td>
<td>700</td>
<td>Executed</td>
<td>700 done + 300 from Guatemala/Argentina/South Chile</td>
</tr>
<tr>
<td>Chile</td>
<td>Catterina Ferreccio, Vanessa Van de Wyngard</td>
<td>1,500</td>
<td>600</td>
<td>Executed</td>
<td>448 done + 900 to DNA extraction</td>
</tr>
<tr>
<td>Colombia/Venezuela</td>
<td>Bladimiro R Orozco</td>
<td>158</td>
<td>158</td>
<td>In progress</td>
<td>158 done</td>
</tr>
<tr>
<td>Chile</td>
<td>Alvaro Cerda, Monica Aguilar</td>
<td>300</td>
<td>100</td>
<td>Executed</td>
<td>100 done + 200 to be collected</td>
</tr>
<tr>
<td>Peru</td>
<td>Meddly Santolalla</td>
<td>1,100</td>
<td>700</td>
<td>In progress</td>
<td>DNA extraction kits Children? Blood?</td>
</tr>
<tr>
<td>Brazil</td>
<td>Eduardo Tarazona, Maria Cássia, Leandro Colli</td>
<td>400</td>
<td>400</td>
<td>In progress</td>
<td>207 done + 193 DNA extraction</td>
</tr>
</tbody>
</table>

Total expected 4,458
Distinctive Opportunities in COVNET

• First look at susceptibility to SARS-CoV2
  • CDC Studies x 2
  • Austrian Isghl Study
• Special Populations
  • Liquidator Cohort in Ukraine (post Chernobyl)
  • Long term consequences of radiation exposure
• Population Genetics of Latin America
COVNET: Key principles

• Emphasis on harmonization of:
  • Phenotype definitions
  • Genotype analyses
  • Meta-analyses
• Aligned with COVID-19 Host Genetics Initiative: https://www.covid19hg.org
  • Importance of phenotype data (lagging behind)
• Emphasis on disparities
• Data sharing is central to identify and validate risk alleles
  • Hard to share until we have adequate phenotype data
Sample requirements and processing

Amy Hutchinson
Director of Operations, CGR
Sample Requirements Requested by COVNET (pre WGS)

**DNA**
- Mass: 1.0 – 1.5ug (no WGS)
- Mass: 2.5 – 3.5 ug (with WGS)*
- Volume: ≥ 30ul

**Blood**
- Whole Blood, Buffy Coat, or PBMCs
- Volume: ≥ 150ul

**Buccal**
- Oragene, Mouthwash, or Saliva
- Volume: ≥ 1000ul

* Sample Kit use is encouraged.

* Currently exploring low input options for WGS
Sample Processing

Samples are processed in a highly-automated, high-throughput laboratory environment following established SOPs.

**Laboratory**
Tracking via a highly-integrated, customized Laboratory Information Management System (LIMS)

**Extraction**
KingFisher Flex

**DNA QC**
Volume verification
PicoGreen quantification

**Genotyping**
Standard input = 200ng
Minimum input = 50ng
1-3% QC replicates

**Whole Genome Sequencing**
Standard input = 1200ng
Exploring option of low input WGS @ 300ng

Approved/consented studies only with phenodata

Clifton Dalgarg
The American Genome Center
## Samples Genotyped

<table>
<thead>
<tr>
<th>Institute</th>
<th>Unique subjects</th>
<th>Subjects with phenotype classification</th>
<th>Samples genotyped (includes replicates)</th>
<th>Expected replicates (samples with lower call rate excluded)</th>
<th>Sex-discordant</th>
<th>Contaminated</th>
<th>Low call rate</th>
<th>Unexpected Replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Athens</td>
<td>306</td>
<td>305</td>
<td>315</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Austria</td>
<td>342</td>
<td>NA</td>
<td>346</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Carrington/Harvard</td>
<td>495</td>
<td>433</td>
<td>500</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>COVIDcode</td>
<td>45</td>
<td>40</td>
<td>47</td>
<td>2</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KP - Colorado</td>
<td>549</td>
<td>443</td>
<td>554</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MSKCC</td>
<td>388</td>
<td>NA</td>
<td>391</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>NCI/CCR</td>
<td>59</td>
<td>55</td>
<td>58</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NIAID*</td>
<td>392</td>
<td>378</td>
<td>396</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Northwestern</td>
<td>86</td>
<td>50</td>
<td>88</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South Korea</td>
<td>105</td>
<td>82</td>
<td>109</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UAB</td>
<td>576</td>
<td>561</td>
<td>592</td>
<td>16</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>UCSD</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UPenn</td>
<td>358</td>
<td>304</td>
<td>364</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>90*</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,704</strong></td>
<td><strong>2,654</strong></td>
<td><strong>3,764</strong></td>
<td><strong>67</strong></td>
<td><strong>30</strong></td>
<td><strong>11</strong></td>
<td><strong>8</strong></td>
<td><strong>152</strong></td>
</tr>
</tbody>
</table>

* Includes provided WGS data
Post-Genotyping Data QC for GWAS on GSA v2

• Sample-level QC
  • Array processing: remove samples that fail to generate valid idat/gtc files
  • Completion rate: cutoff 0.8 for samples and 0.8 for loci, followed by 0.95 for samples and 0.95 for loci
  • Sample contamination: exclude samples with >10% contamination as predicted by VerifyIDintensity
  • Expected replicate removal: include replicate with higher call rate

• Subject-level QC
  • Sex verification: reported vs. observed sex based on chrX method-of-moments F coefficient. Cutoff at 0.5 (expected to be 0.0 for females and 1.0 for males)
  • Unexpected replicates: phenotypes are assessed before filtering unexpected replicates
Data Sharing Plan

• Following primary data QC, a delivery package will be provided to each site via NIH secured transfer on the Box platform.
• Data access is restricted to you (and those you designate) plus designated NCI staff only.
• Delivery package includes:
  • QC Reports (.xls and .doc) – These reports are generated from the genotyping batch of which your samples were a part. You will therefore notice additional samples included in both reports, this is expected and samples that are not your own can be ignored (no proprietary information is provided here).
  • Called Genotypes – Called genotypes for your subjects from CGR’s pipeline in standard PLINK format.
  • Sample ID Linkage File – This is the master sample ID file and provides a listing of the various IDs contained in the reports and genotyping files for your samples only. This provides the link between IDs we generated internally as part of our laboratory pipelines and the IDs you provided with your samples. Any samples excluded from genotyping are listed at the top of the spreadsheet with a comment regarding the reason for exclusion.
  • Illumina Infinium Genotyping_V3 – This is CGR’s standard material and methods (M&M) document for Illumina genotyping and describes genotyping and data QC workflows currently in place. Specific details related to data QC of this project are included in the QC Report.
Analyses and preliminary data: GWAS

Lisa Mirabello
Senior Investigator, DCEG

Meredith Yeager
Scientific Director, CGR
Published germline genetics of severe COVID-19 disease

• Several studies suggest an underlying genetic component to severe disease

• One genetic locus—on chromosome 3p21.31—has been repeatedly associated with hospitalization\(^1-^3\)

• \(ABO\) locus has been reported as a risk factor for both COVID-19 susceptibility\(^3\) and severity\(^1\) in some studies but not in others\(^2\)

• Recent GWAS identified 3 new loci on: chromosome 12q24.13 (\(OAS1-3\)); chromosome 19p13.3 (\(DPP9\)); and, chromosome 21q22.1 (\(IFNAR2\))\(^2\)

• Compared patients with severe COVID-19 disease to the general population

• Rare variants in IFN genes associated with severe disease\(^4\)

Published germline genetics of severe COVID-19 disease

- 23andMe research cohort: 15,434 COVID-19 patients
- Chromosome 3p21.31 contains multiple genes that could be functionally implicated in COVID-19 pathology: *SLC6A20, LZFTL1, CCR9, CXCR6, XCR1, FYCO1*
Published germline genetics of severe COVID-19 disease

- COVID-19 Host Genetics Initiative: 3 case-control GWAS including 49,562 COVID-19 patients from 46 studies
  1. Critically ill cases requiring respiratory support in hospital or who were deceased, \( N=6,179 \)
  2. Cases with moderate or severe COVID-19 defined as those hospitalized, \( N=13,641 \)
  3. All cases with reported SARS-CoV-2 infection with or without any symptoms, \( N=49,562 \)
- Each compared to ~2M population controls: genetically ancestry-matched without SARS-CoV-2
Published germline genetics of severe COVID-19 disease

Hospitalized COVID-19+ (N.cases=13,641, N.controls=2,070,709)
GWAS Analysis plan

**Post-Genotyping QC**
- Assess completion rate, sample contamination, duplicates, sex verification

**Compute ancestry**
- GRAF, plink

**Split data by ancestry**
- plink

**QC within each ancestry group**
- plink
  - **SNP-level QC**:
    - Genotyping rate, MAF, HWE, differential missingness
  - **Sample-level QC**:
    - Heterozygosity, call rate, prune relatives and duplicates
  - **PCA on filtered and LD pruned dataset**
    - gcta

**GWAS on imputed data**
- Include age, sex, PCs and study-specific covariates
  - SAIGE LMM, plink

**Post-imputation QC**
- QCtools, bcftools, plink

**Submit data for imputation using TopMed reference panel**
- bcftools, plink

Adapted from Pairo-Castineira et al., *Nature* 2020.
### Samples received: Phenotype data required for analyses

<table>
<thead>
<tr>
<th>Samples</th>
<th>N</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total received</td>
<td>8,898</td>
<td></td>
</tr>
<tr>
<td>Total genotyped</td>
<td>3,704</td>
<td></td>
</tr>
<tr>
<td>Total with phenotype</td>
<td>2,654</td>
<td></td>
</tr>
<tr>
<td>Unable to classify disease</td>
<td>213</td>
<td>8%</td>
</tr>
<tr>
<td>Missing covariate data (e.g., smoking info)</td>
<td>1,351</td>
<td>51%</td>
</tr>
<tr>
<td>Missing co-morbidity data</td>
<td>2,032</td>
<td>77%</td>
</tr>
</tbody>
</table>

- Phenotype data needed with requested covariates using our provided template in order to harmonize disease severity classification across studies
Analysis Plan: Primary association analyses for disease severity

Laboratory confirmed SARS-CoV-2 infection

**Mild disease**: no hospitalization after a positive test

**Moderate disease**: hospitalization for COVID19 and non-mechanical respiratory support

**Severe disease**: death or hospitalization for COVID19 and mechanical respiratory support

**Analysis 1**: Very severe respiratory confirmed COVID19 disease
- Mild vs. Severe

**Analysis 2**: Moderate/severe respiratory confirmed COVID19 disease
- Mild vs. Moderate + Severe (Not hospitalized vs. hosp.)

**Analysis 3**: Severity ordinal

**Analysis 4**: Mortality

For EUR ancestry individuals: generalized linear mixed models to evaluate imputed SNP associations for the specified outcomes; models adjusted for PCs, sex, age, and appropriate covariates.
**COVNET COVID-19 cases with phenotype data**

<table>
<thead>
<tr>
<th>COVID patients</th>
<th>N</th>
<th>Column %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mild</td>
<td>1486</td>
<td>47.8%</td>
</tr>
<tr>
<td>moderate</td>
<td>792</td>
<td>25.5%</td>
</tr>
<tr>
<td>severe</td>
<td>830</td>
<td>26.7%</td>
</tr>
<tr>
<td><strong>missing</strong></td>
<td>213</td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>1382</td>
<td>44.5%</td>
</tr>
<tr>
<td>female</td>
<td>1726</td>
<td>55.5%</td>
</tr>
<tr>
<td><strong>Vital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alive</td>
<td>1132</td>
<td>82.7%</td>
</tr>
<tr>
<td>deceased</td>
<td>236</td>
<td>17.3%</td>
</tr>
<tr>
<td><strong>missing</strong></td>
<td>1740</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>current</td>
<td>400</td>
<td>22.1%</td>
</tr>
<tr>
<td>former</td>
<td>298</td>
<td>16.4%</td>
</tr>
<tr>
<td>never</td>
<td>1116</td>
<td>61.5%</td>
</tr>
<tr>
<td><strong>missing</strong></td>
<td>1351</td>
<td></td>
</tr>
</tbody>
</table>

*Severity could not be assigned from phenotype data*
COVNET COVID-19 cases with phenotype data

Bar graphs showing the distribution of mild, moderate, and severe COVID-19 cases by sex and smoking status.

- **Sex**: Male (M) and Female (F)
- **Smoking**: Current, Former, Never

The graphs indicate higher percentages of severe cases among males compared to females. Smoking status also shows differences across disease severities, with higher percentages of severe cases among those who are current smokers or have a history of smoking.
Ancestry computation with GRAF

- Assumes that each subject is an admixture of three ancestries:
  - European
  - Asian
  - African
- Estimates proportions of the 3 ancestries and assigns each subject to a population
- PCA performed on filtered (MAF>0.1, HWE p-value>10^{-6}, missingness<1%) and LD-pruned SNPs (1000 markers, a step size of 80 markers and an r2 threshold of 0.1)

- African American=586
- European=2,093
- Hispanic 2=220
Analysis of EUR samples with COVID disease severity classification available (N=1,611 unique subjects)

<table>
<thead>
<tr>
<th>Institute</th>
<th>Samples genotyped (includes replicates)</th>
<th>EUR ancestry</th>
<th>EUR: Sample-level exclusions</th>
<th>Final EUR sample counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing phenotype data/classification</td>
<td>QC: kinship, duplicates, sex-discordant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athens</td>
<td>315</td>
<td>257</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Austria</td>
<td>346</td>
<td>129</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>Carrington Lab</td>
<td>500</td>
<td>331</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>COVIDcode</td>
<td>47</td>
<td>18</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>KP - Colorado</td>
<td>554</td>
<td>268</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>MSKCC</td>
<td>391</td>
<td>234</td>
<td>234</td>
<td></td>
</tr>
<tr>
<td>NCI/CCR</td>
<td>58</td>
<td>40</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>NIAID</td>
<td>396</td>
<td>351</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Northwestern</td>
<td>88</td>
<td>58</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>South Korea</td>
<td>109</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UAB</td>
<td>592</td>
<td>317</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>UCSD</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>UPenn</td>
<td>364</td>
<td>89</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,764</strong></td>
<td><strong>2,093</strong></td>
<td><strong>482</strong></td>
<td><strong>66</strong></td>
</tr>
</tbody>
</table>
Analysis of EUR samples with COVID disease severity classification available (N=1,611 unique subjects)

SNP-level QC

• Select variants with call rate > 99%, minor allele frequency (MAF) > 1%, Hardy-Weinberg equilibrium (HWE) p-value > 10^{-6}

• Exclude SNPs with significant differential missingness (P<0.001) across disease states

• Compute PCs with LD-pruned SNPs using gcta1.9

• Submit QC’d dataset for imputation using TopMed reference panel

• Download and QC imputed data (imputation quality score > 0.9, ~6.8M SNPs remain) and convert to BGEN format
Case-Case GWAS: mild vs. moderate + severe European ancestry cases

- 614 mild cases, 997 moderate/severe cases
- **Top SNP:** chr12, 12q13.12; p-value = 5.09e-07
  - Covariates: age, sex, PC1-10

![QQ plot and Manhattan plot](image-url)
WGS selection and data sharing

Stephen Chanock
Principles for Whole Genome Sequencing (30X)

- Target: 15-20% of each COVNET study
- All of NCCAPS (extra funds)
- Larger fraction of CovidCode (extra funds)
- Selection of samples
  - NCI COVNET team will work with each study
    - Must meet requirements for adequate DNA and phenotypes
  - Highest priorities
    - Extreme phenotypes (mortality as well as survival with many comorbidities)
    - Population genetics
- Agreement (@ reagent cost)
  - American Genome Center at Uniformed Services University
  - HudsonAlpha Institute for Biotechnology with UAB samples only
Whole Genome Sequencing

- Collaboration with The American Genome Center (TAGC) at Uniformed Services University (USU)
- Currently Pilot 384 subject in progress
- Selection for WGS requires proper study approvals/consent AND completed phenotype data delivery

<table>
<thead>
<tr>
<th>Institute</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrington Lab</td>
<td>1</td>
</tr>
<tr>
<td>NCI/CCR</td>
<td>45</td>
</tr>
<tr>
<td>COVIDcode</td>
<td>32</td>
</tr>
<tr>
<td>Northwestern</td>
<td>51</td>
</tr>
<tr>
<td>South Korea</td>
<td>3</td>
</tr>
<tr>
<td>UAB</td>
<td>10</td>
</tr>
<tr>
<td>UPenn</td>
<td>242</td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td><strong>384</strong></td>
</tr>
</tbody>
</table>
Whole Genome Sequencing: batch 1

<table>
<thead>
<tr>
<th>WGS selection criteria</th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenotype extremes</td>
<td>40</td>
</tr>
<tr>
<td>aged &lt;50 years and severe disease</td>
<td>8</td>
</tr>
<tr>
<td>aged &gt;65 years and mild disease</td>
<td>32</td>
</tr>
<tr>
<td>Diverse population (AfrAm, His, Asn)</td>
<td>214</td>
</tr>
<tr>
<td>EUR, disease groups for comparison</td>
<td>125</td>
</tr>
<tr>
<td>Duplicates</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>384</strong></td>
</tr>
</tbody>
</table>
The COVID-19 Host Genetics Initiative

Aims

The COVID-19 host genetics initiative is a bottom-up collaborative effort that has three main goals:

1. Provide an environment to foster the sharing of resources to facilitate COVID-19 host genetics research (e.g. protocols, questionnaires).
2. Organize analytical activities across studies to identify genetic determinants of COVID-19 susceptibility and severity.
3. Provide a platform to share the results from such activities, as well as the individual-level data where possible, to benefit the broader scientific community.
Secondary analyses with COVNET genotype data

• Use the genotyped SNPs within chromosome 6 for HLA imputation to determine if specific HLA alleles are associated with COVID-19 disease severity

• Evaluate large structural copy number alterations (i.e., clonal mosaicism)
  • Acquired somatic mutations (mosaic gain, loss, or copy neutral LOH) in a clonal subset of cells that differs from the inherited germline genome
Next Major Steps

• Increase Accrual - target 40,000 for GWAS
• Selecting NGS/WGS for American Genome Center (USUHS)
  • 4-5,000 (COVNET)
  • Selection - extreme phenotypes, URM and population genetics
  • 2,000 (NCCAPS)
• Data sharing with dbGap/AnVil
• COVID Human Genetics Initiative
Important Take Home Points

• Critical need for providing *minimal phenotype data*- focus on acute COVID19 outcomes
  • Major impediment to analyses
• Potential for studying ‘Long-COVID19’ when follow-up data is available
  • Challenge of defining phenotype
  • NIH Strategic Initiative
• Large capacity for GWAS and NGS
  • Targets
    • 40,000 GWAS
    • 5,000 NGS/WGS
QUESTIONS?

https://dceg.cancer.gov/research/how-we-study/genomic-studies/covnet