GENOTYPE TO PHENOTYPE
WITH BIG DATA / NIH ALL OF US
Data Browser

Browse aggregate-level data contributed by All of Us research participants. Data are derived from multiple data sources. To protect participant privacy, we have removed personal identifiers, rounded aggregate data to counts of 20, and only included summary demographic information. Individual-level data are available for analysis in the Researcher Workbench.

Search Across Data Types

EHR Domains

- Conditions: 25,638 medical concepts, 264,700 participants
- Drug Exposures: 29,865 medical concepts, 239,740 participants
- Labs & Measurements: 16,216 medical concepts, 252,980 participants
- Procedures: 30,328 medical concepts, 242,500 participants
**GENOME → TRAIT**

**DOES PERSON HAVE TRAIT?**

| TRUE | 3AAGTCAGGAGTTCAAGACAGCTGGACAATAGTGAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| FALSE | AGGCTGGACAACATAGTGAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| TRUE | ACGTAGTAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| FALSE | ACCCGCTCTCTACTAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| TRUE | 3AAGTCAGGAGTTCAAGACAGCTGGACAATAGTGAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| FALSE | AGGCTGGACAACATAGTGAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| TRUE | ACGTAGTAACCCCGTCTCTACAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |
| TRUE | ACCCGCTCTCTACTAAAAATACAAAAATTCAGCCAAGGCAATGATGACATGCCTGTAGTGGCA |

**GENOME REPRESENTED BY 3 BILLION BASES . . .**
DOES PERSON HAVE TRAIT?

<table>
<thead>
<tr>
<th>TRUE</th>
<th>GENOME REPRESENTED BY 3 MILLION SNPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3AAGTCAGGAGTTCAAGACACAGCCTGGAACACATAGTGAATAATCCTTGACTATAATTGAATAATGCAAGGCCATATTGACATGCACTGCGCTGTCAGTTGCA</td>
<td></td>
</tr>
<tr>
<td>FALSE</td>
<td>AGCCTGGAACACATTAGTGAACCCCTCTCTACTTAAATATTCAAATGCAAATGAGAAATGCAAGCTGCTGAGCCCAGAGATCGACAGTGCCAGTCCTCAGG</td>
</tr>
<tr>
<td>TRUE</td>
<td>ACGATGGAACCCCGTCTCTCTACTTTATAGAAAAATAGCAAACAGAGGTAGCTGCAATGTCAGGCTGCGCTGAGAAGCTGTCAGACCTTATGAGAA</td>
</tr>
<tr>
<td>FALSE</td>
<td>ACCCGTCTCTCTACTTTATAGAAAAATAGCAAACAGAGGTAGCTGCAATGTCAGGCTGCGCTGAGAAGCTGTCAGACCTTATGAGAA</td>
</tr>
<tr>
<td>TRUE</td>
<td>3AAGTCAGGAGTTCAAGACACAGCCTGGAACACATAGTGAATAATCCTTGACTATAATTGAATAATGCAAGGCCATATTGACATGCACTGCGCTGTCAGTTGCA</td>
</tr>
<tr>
<td>FALSE</td>
<td>AGCCTGGAACACATTAGTGAACCCCTCTCTACTTAAATATTCAAATGCAAATGAGAAATGCAAGCTGCTGAGCCCAGAGATCGACAGTGCCAGTCCTCAGG</td>
</tr>
<tr>
<td>TRUE</td>
<td>ACGATGGAACCCCGTCTCTCTACTTTATAGAAAAATAGCAAACAGAGGTAGCTGCAATGTCAGGCTGCGCTGAGAAGCTGTCAGACCTTATGAGAA</td>
</tr>
<tr>
<td>TRUE</td>
<td>ACCCGTCTCTCTACTTTATAGAAAAATAGCAAACAGAGGTAGCTGCAATGTCAGGCTGCGCTGAGAAGCTGTCAGACCTTATGAGAA</td>
</tr>
</tbody>
</table>
IS THERE A WAY TO EFFECTIVELY REPRESENT THE GENOME WITH A SMALL SET OF NUMBERS?
### GENOME → TRAIT

**Does person have trait?**

<table>
<thead>
<tr>
<th></th>
<th>0.6692</th>
<th>0.1110</th>
<th>0.8492</th>
<th>0.3999</th>
<th>0.7105</th>
<th>0.7826</th>
<th>0.3924</th>
<th>0.1968</th>
<th>0.3401</th>
<th>0.5229</th>
<th>0.3563</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>0.7797</td>
<td>0.5124</td>
<td>0.5825</td>
<td>0.5155</td>
<td>0.2953</td>
<td>0.8429</td>
<td>0.9047</td>
<td>0.5270</td>
<td>0.6030</td>
<td>0.3484</td>
<td>0.7326</td>
</tr>
<tr>
<td>FALSE</td>
<td>0.9704</td>
<td>0.2992</td>
<td>0.9449</td>
<td>0.2745</td>
<td>0.7832</td>
<td>0.6782</td>
<td>0.6906</td>
<td>0.5825</td>
<td>0.2656</td>
<td>0.7008</td>
<td>0.5836</td>
</tr>
<tr>
<td>TRUE</td>
<td>0.7293</td>
<td>0.4957</td>
<td>0.9162</td>
<td>0.7633</td>
<td>0.2475</td>
<td>0.7382</td>
<td>0.4629</td>
<td>0.9832</td>
<td>0.1649</td>
<td>0.2918</td>
<td>0.9134</td>
</tr>
<tr>
<td>FALSE</td>
<td>0.2912</td>
<td>0.3831</td>
<td>0.9482</td>
<td>0.9368</td>
<td>0.8467</td>
<td>0.8955</td>
<td>0.2205</td>
<td>0.9557</td>
<td>0.8946</td>
<td>0.9266</td>
<td>0.1125</td>
</tr>
<tr>
<td>TRUE</td>
<td>0.2266</td>
<td>0.1279</td>
<td>0.3359</td>
<td>0.2456</td>
<td>0.4121</td>
<td>0.0897</td>
<td>0.7801</td>
<td>0.0087</td>
<td>0.7738</td>
<td>0.7936</td>
<td>0.4557</td>
</tr>
<tr>
<td>FALSE</td>
<td>0.5799</td>
<td>0.8747</td>
<td>0.2202</td>
<td>0.5659</td>
<td>0.5211</td>
<td>0.4174</td>
<td>0.4021</td>
<td>0.3522</td>
<td>0.8971</td>
<td>0.3544</td>
<td>0.9425</td>
</tr>
<tr>
<td>TRUE</td>
<td>0.3455</td>
<td>0.0724</td>
<td>0.7952</td>
<td>0.2513</td>
<td>0.7598</td>
<td>0.5140</td>
<td>0.5037</td>
<td>0.1641</td>
<td>0.0331</td>
<td>0.6201</td>
<td>0.8587</td>
</tr>
<tr>
<td>FALSE</td>
<td>0.6937</td>
<td>0.0253</td>
<td>0.2690</td>
<td>0.7687</td>
<td>0.2655</td>
<td>0.7809</td>
<td>0.5909</td>
<td>0.6215</td>
<td>0.0112</td>
<td>0.7535</td>
<td>0.6059</td>
</tr>
<tr>
<td>TRUE</td>
<td>0.7507</td>
<td>0.4902</td>
<td>0.1657</td>
<td>0.8345</td>
<td>0.7747</td>
<td>0.7518</td>
<td>0.0202</td>
<td>0.8121</td>
<td>0.1391</td>
<td>0.8665</td>
<td>0.0797</td>
</tr>
</tbody>
</table>
RESEARCH QUESTIONS

1. IS THERE ANY CORRELATION BETWEEN __________________ AND A PERSON’S GERM LINE GENETICS?

2. HOW WELL CAN YOU PREDICT THAT A PERSON WILL HAVE __________________ BASED ON THEIR GERM LINE GENETICS?
WE HAVE FILLED IN THE BLANK WITH THESE CONDITIONS.

COVID-19 SEVERITY
OVARIAN CANCER
GLIOBLASTOMA MULTIFORME
BREAST CANCER
ALCOHOLISM
SCHIZOPHRENIA
Data Browser

Browse aggregate-level data contributed by All of Us research participants. Data are derived from multiple data sources. To protect participant privacy, we have removed personal identifiers, rounded aggregate data to counts of 20, and only included summary demographic information. Individual-level data are available for analysis in the Researcher Workbench.

Search Across Data Types

- Keyword Search
- FAQs
- Introductory Videos
- User Guide

Data includes 400,420 participants as of 2/15/2023.

EHR Domains

- **Conditions**: 25,638 medical concepts, 264,700 participants
- **Drug Exposures**: 29,865 medical concepts, 239,740 participants
- **Labs & Measurements**: 16,216 medical concepts, 262,980 participants
- **Procedures**: 30,328 medical concepts, 242,580 participants

View Conditions
View Drug Exposures
View Labs & Measurements
View Procedures
Genomics

SNP/Indel Variants

245,400
Participants in Short-Read Whole Genome Sequencing (WGS) dataset

1,074,881,214
SNP/Indel Variants

View SNP/Indel Variants

Genomic data only in Researcher Workbench

1,040 participants in the Long-Read WGS dataset

11,400 participants in the Short-Read WGS Structural Variants dataset

312,940 participants in the Genotyping Arrays dataset

Register for access

Measurements and Wearables

Physical Measurements

8
Physical Measurements

337,540 participants

Participants have the option to provide a standard set of physical measurements.

View Physical Measurements

Fitbit

6
Fitbit Measurements

15,620 participants

Fitbit data includes heart rate and activity summaries.

View Fitbit
Survey Questions

The Basics
28 questions available
409,420 participants
This survey includes participant demographic information.
View Complete Survey

Overall Health
21 questions available
400,760 participants
Survey includes information about how participants report levels of individual health.
View Complete Survey

Lifestyle
26 questions available
398,200 participants
Survey includes information on participant smoking, alcohol and recreational drug use.
View Complete Survey

Health Care Access & Utilization
57 questions available
190,220 participants
Survey includes information about a participant's access to and use of health care.
View Complete Survey

Personal and Family Health History
614 questions available
185,240 participants
Survey includes information about the medical history of a participant and their immediate family members.
View Complete Survey

COVID-19 Participant Experience (COPE)
191 questions available
100,220 participants
Survey includes information about the impact of COVID-19 on participant mental and physical health.
View Complete Survey

COPE Minute Survey
141 questions available
127,240 participants
Survey includes information regarding a participant's COVID-19 vaccination experience.
View Complete Survey

Social Determinants of Health
80 questions available
117,800 participants
Survey includes information about the social determinants of health, including a participant's neighborhood, social life, stress, and feelings about everyday life.
View Complete Survey
1. IS THERE ANY CORRELATION BETWEEN __________________ AND A PERSON’S GERM LINE GENETICS?

2. HOW WELL CAN YOU PREDICT THAT A PERSON WILL HAVE __________________ BASED ON THEIR GERM LINE GENETICS?