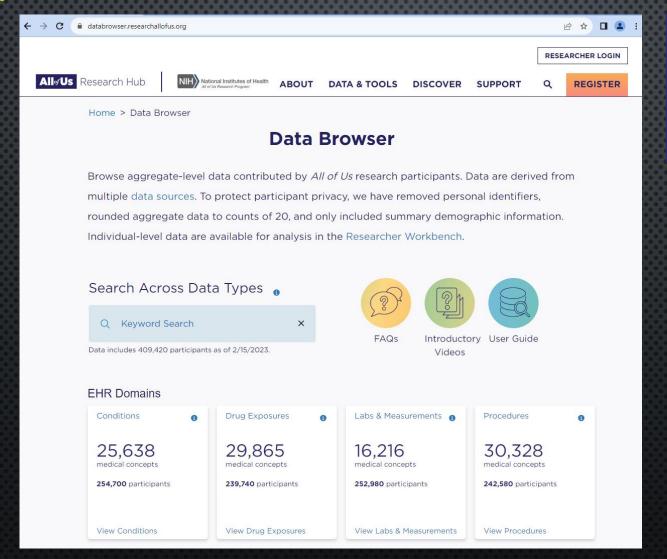# GENOTYPE TO PHENOTYPE WITH BIG DATA / NIH ALL OF US

# http://databrowser.researchallofus.org

# GENOME →TRAIT

**DOES PERSON HAVE TRAIT?**

**GENOME REPRESENTED BY 3 BILLION  BASES . . .**

| | |
|---|---|
| TRUE | GAAGTCAGGAGTTCAAGACCAGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCA |
| FALSE | AGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCAC |
| TRUE | ACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG |
| FALSE | ACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAGCTTGGGCAAC |
| TRUE | GAAGTCAGGAGTTCAAGACCAGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCA |
| FALSE | AGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCAC |
| TRUE | ACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG |
| TRUE | ACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAGCTTGGGCAAC |

# GENOME → TRAIT

**DOES PERSON HAVE TRAIT?**

**GENOME REPRESENTED BY 3 MILLION SNPS . . .**

| | |
|---|---|
| TRUE | GAAGTCAGGAGTTCAAGACCAGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCA |
| FALSE | AGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCAC |
| TRUE | ACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG |
| FALSE | ACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAGCTTGGGCAAC |
| TRUE | GAAGTCAGGAGTTCAAGACCAGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCA |
| FALSE | AGCCTGGACAACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCAC |
| TRUE | ACATAGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG |
| TRUE | ACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCACAATGGTACATGCCTGTAGGTTGCAGTGAGCCAAGATCGCACCACTGCACTCCAGCTTGGGCAAC |

# IS THERE A WAY TO EFFECTIVELY REPRESENT THE GENOME WITH A SMALL SET OF NUMBERS?

# GENOME→TRAIT

**DOES PERSON HAVE TRAIT?**

**GENOME REPRESENTED BY 88 NUMBERS . . .**

| DOES PERSON HAVE TRAIT? | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRUE | 0.6692 | 0.1110 | 0.8492 | 0.3999 | 0.7105 | 0.7826 | 0.3924 | 0.1968 | 0.3401 | 0.5229 | 0.3563 |
| FALSE | 0.7797 | 0.5124 | 0.5825 | 0.5165 | 0.2953 | 0.8429 | 0.9047 | 0.5270 | 0.6030 | 0.3484 | 0.7326 |
| TRUE | 0.9704 | 0.2992 | 0.9449 | 0.2745 | 0.7832 | 0.6782 | 0.6906 | 0.5825 | 0.2656 | 0.7008 | 0.5836 |
| TRUE | 0.7293 | 0.4957 | 0.9162 | 0.7633 | 0.2475 | 0.7382 | 0.4629 | 0.9832 | 0.1649 | 0.2918 | 0.9134 |
| FALSE | 0.2912 | 0.3831 | 0.9482 | 0.9368 | 0.8467 | 0.8955 | 0.2205 | 0.9957 | 0.8946 | 0.9266 | 0.1125 |
| TRUE | 0.2266 | 0.1279 | 0.3359 | 0.2456 | 0.4121 | 0.0897 | 0.7801 | 0.0087 | 0.7738 | 0.7936 | 0.4557 |
| FALSE | 0.5799 | 0.8747 | 0.2202 | 0.5659 | 0.5211 | 0.4174 | 0.4021 | 0.3522 | 0.8971 | 0.3544 | 0.9425 |
| TRUE | 0.3455 | 0.0724 | 0.7952 | 0.2513 | 0.7598 | 0.5140 | 0.5037 | 0.1641 | 0.0331 | 0.6201 | 0.8587 |
| FALSE | 0.6937 | 0.0253 | 0.2690 | 0.7587 | 0.2655 | 0.7809 | 0.5909 | 0.6215 | 0.0112 | 0.7535 | 0.6059 |
| TRUE | 0.7507 | 0.4902 | 0.1657 | 0.8345 | 0.7747 | 0.7518 | 0.0202 | 0.8121 | 0.1391 | 0.8865 | 0.0797 |

# RESEARCH QUESTIONS

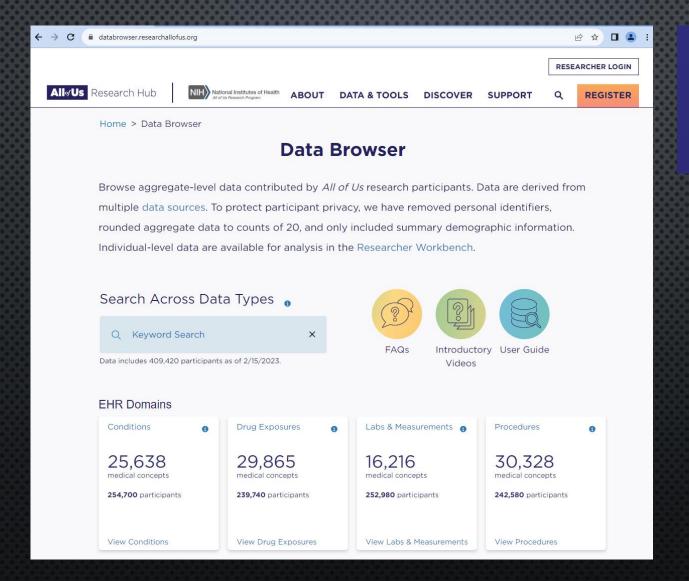1. IS THERE ANY CORRELATION BETWEEN _____ AND A PERSON'S GERM LINE GENETICS?

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**

# DATABROWSER.RESEARCHALLOFUS.ORG

## EHR Domains

### Conditions ⓘ

**25,638**
medical concepts

**254,700** participants

View Conditions

### Drug Exposures ⓘ

**29,865**
medical concepts

**239,740** participants

View Drug Exposures

### Labs & Measurements ⓘ

**16,216**
medical concepts

**252,980** participants

View Labs & Measurements

### Procedures ⓘ

**30,328**
medical concepts

**242,580** participants

View Procedures

## Genomics

## Measurements and Wearables

### SNP/Indel Variants ℹ️

**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 ℹ️

**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

### Physical Measurements ℹ️

**8**
Physical Measurements

**337,540** participants

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

View Physical Measurements

### Fitbit ℹ️

**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

# 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?