

MIND THE GAP

RESOURCES REQUIRED TO RECEIVE,
PROCESS, & INTERPRET RESEARCH-
RETURNED WHOLE GENOME DATA



FARREN BRIGGS, PHD SCM
Assistant Professor, Case Western Reserve University



OVERVIEW OF RETURNING **WHOLE GENOME DATA**

- WHERE IT ALL BEGAN
- WHO ARE POTENTIAL STUDY PARTICIPANTS?
- WHAT GENETIC RESULTS/DATA MIGHT BE RETURNED?
- HOW MIGHT GENETIC RESULTS/DATA BE RETURNED?
- HOW TO INTERPRET RETURNED GENETIC DATA?
- EXPLORE MY GENETIC DATA



U.S. FEDERAL INVESTMENT

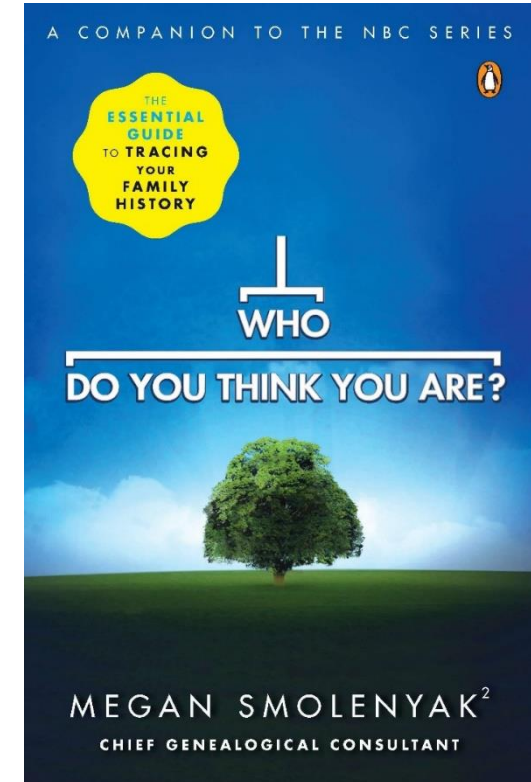
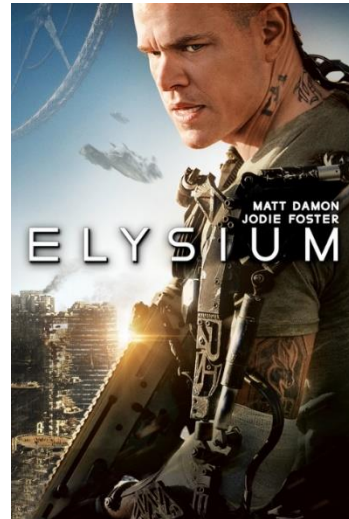
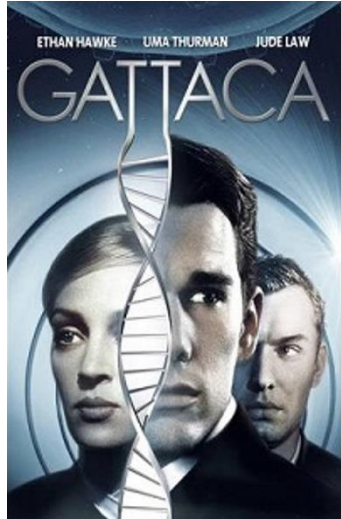


“Tonight I’m launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes.

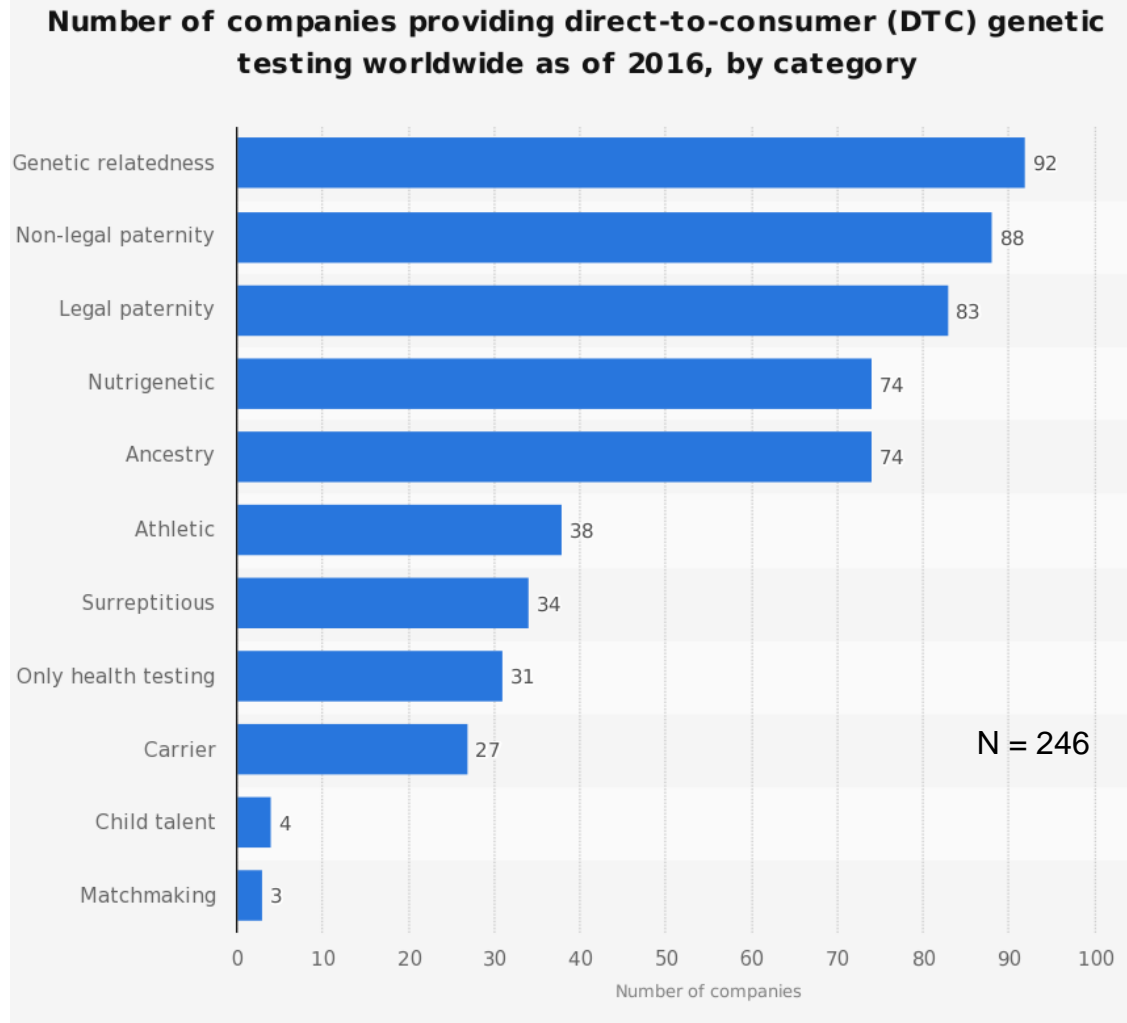
And to give us all access to the personalized information we need to keep ourselves and our families healthier.”

President Barack Obama
2015 State of the Union Address | January 20, 2015

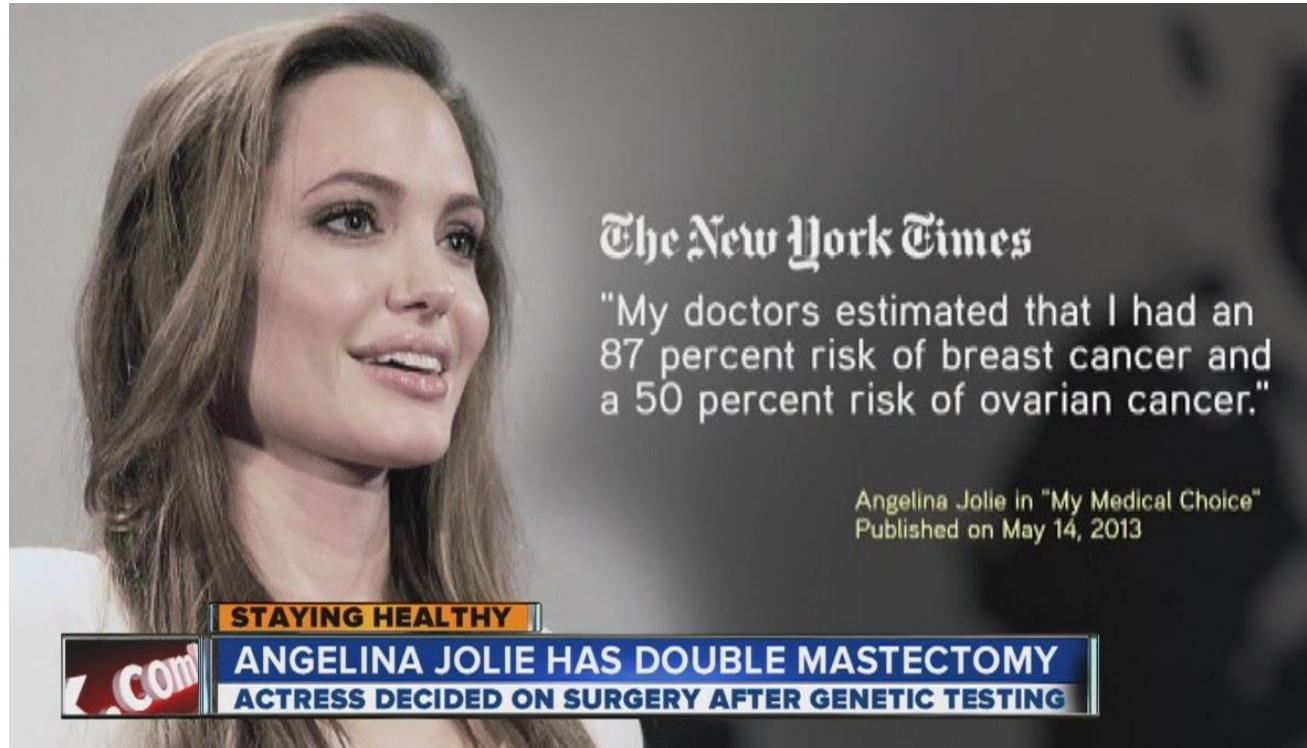
GENETICS GOES HOLLYWOOD TO DOCU-SERIES



DIRECT TO CONSUMER GENETIC TESTING (DTC-GT)



CELEBRITY TESTIMONIALS



The New York Times

"My doctors estimated that I had an 87 percent risk of breast cancer and a 50 percent risk of ovarian cancer."

Angelina Jolie in "My Medical Choice"
Published on May 14, 2013

STAYING HEALTHY

ANGELINA JOLIE HAS DOUBLE MASTECTOMY
ACTRESS DECIDED ON SURGERY AFTER GENETIC TESTING

com

This image is a screenshot from a video testimonial featuring Angelina Jolie. On the left, she is shown from the chest up, looking slightly to her right with a gentle smile. The background is a dark, out-of-focus grey. To her right, white text is overlaid on the background. At the bottom of the frame, there is a news-style banner with a red and white logo on the left and a blue bar with white text containing the headline. Below the headline, there is a smaller yellow bar with the words 'STAYING HEALTHY' in black.



<https://www.youtube.com/watch?v=W5nylAJlky4>

<https://www.aarp.org/health/healthy-living/info-2014/sheryl-crow-melissa-etheridge-beat-cancer.html>

https://www.etonline.com/news/209150_exclusive_christina_applegate_youth_oregon_bad_moms_married_with_children

POTENTIAL RESEARCH PARTICIPANTS WANT THEIR RAW GENETIC DATA

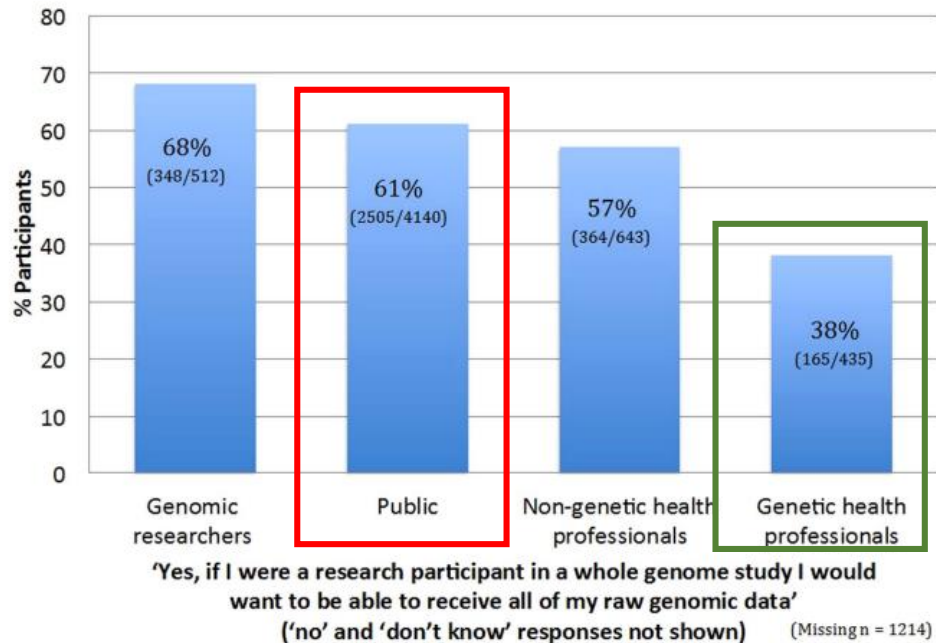


Figure 1 Interest in receiving raw genomic data.

Q: If you were given all of your raw genomic data from a research study, what would you do with this? (n=6944)

A: "I would seek out an interpretation of it" 62% (n=4320)

- ▶ 60% said "I'd analyse it myself" (n=2581)
(Within the different professional groups, these are the percentages who would analyse the data themselves: 81% genomic researchers, 68% genetic health professionals, 56% other health professionals, 56% public)
- ▶ 57% said "I would ask for a referral to my local clinical genetics service" (n=2459)
- ▶ 43% said "I would ask my GP or Primary Care Physician" (n=1844)
- ▶ 41% said "would find a genomics researcher and ask them" (n=1775)
- ▶ 15% said "I would pay a commercial genetics company to analyse the data" (n=658)
- ▶ 5% had other suggestions: eg, "use google", "I would ask my bioinformatician colleagues", "I would share it on GitHub", "ask a genetic counsellor", "I would open source it to anyone online", "I would refer the raw data in a zip file to a company like 23andMe", "I would want information about interpreting the data before deciding how to proceed" (n=237)

RESEARCH PROGRAMS ARE NOW RETURNING GENETIC DATA

Project	Dates	# genomes sequenced to-date	Platform	Results Returned	Report to Health Record	Uninterpreted Data to Participants	Accredited Lab	Last updated
Harvard PGP	2005-	352	WGS	Filtered Variants w/ Lit Annot	No	Yes (Variants)	No	Nov 2017
BWH/Harvard MedSeq	2011-	110	WGS	Monogenic, Common, PGx	Yes	FASTQ	Yes	Nov 28, 2017
Mount Sinai HealthSeq	2012–2015	40	WGS	Monogenic, Common, PGx	No	BAM, VCF	No	paper
Mayo "10 scientists"	2012–2014	10	WES	Monogenic	No	Yes	No	paper
Institute for Systems Biology (ISB) Pioneer 100	2014	108	WGS	Monogenic, Common, PGx	No	BAM, VCF	No	paper
BWH/BCH/Harvard BabySeq Project	2015–	160	WGS	Monogenic, PGx	Yes	FASTQ	Yes	Nov 2017
Nevada Institute of Personalized Medicine	2015–	0	WES	Monogenic, PGx	No	BAM, VCF	No	paper
NYGC Seeq.io	2016-	~500	WGS	ancestry, microbiome	No	BAM	No	Feb 2017
NIH All of Us	2017-	?	WGS	ACMG 59, PGx	?	?	Yes	Paper 8/15/2019
100,000 Genomes Project (UK)	2015-	44,633	WGS	Monogenic, PGx	Yes	Yes	Yes	January 2018

12/10/2018 NIH All of Us Webinar:

Genomic Data returned includes: PGx, AMCG pathogenic findings, Ancestry data, and **Raw data files**

WHO ARE POTENTIAL STUDY PARTICIPANTS?

EXAMPLE



WHO WILL BE A PART OF ALL OF US?



- Longitudinal cohort of $\geq 1,000,000$ or more **people** living in the United States
- **Goal:** “to accelerate research and improve health.. Uncover(ing) paths toward delivering **precision medicine** – or individualized prevention, treatment, and care – for all of us”
- **How:** By taking into account individual differences in **lifestyle, socioeconomic, environment, and biology.**

Persons in underrepresented populations will be **prioritized.**

The target percentage of persons in racial and **ethnic minorities** is **>45%** and that of persons in **under-represented** populations is **>75%.**



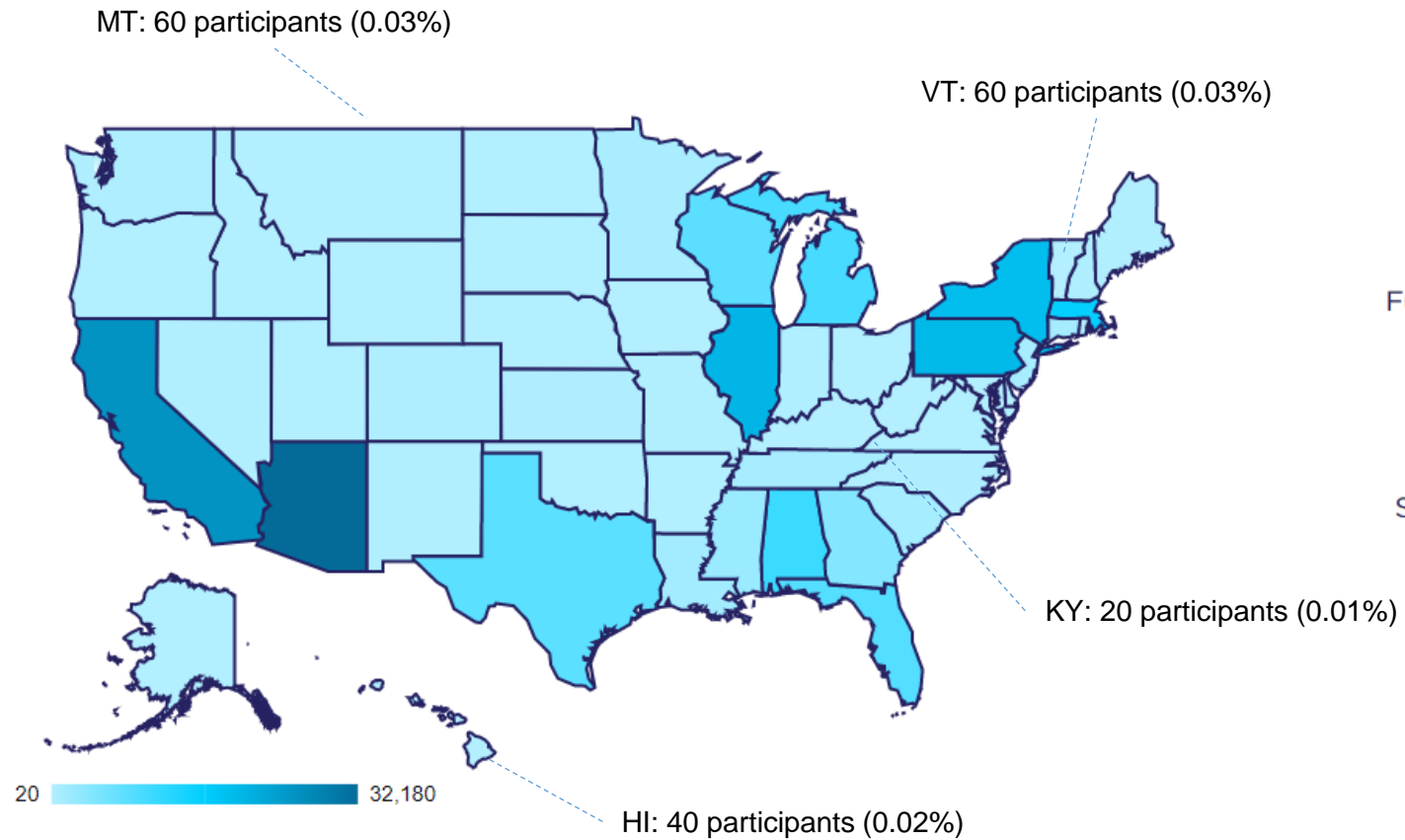
GEOGRAPHY

Geography

Outreach



256,000+
Participants

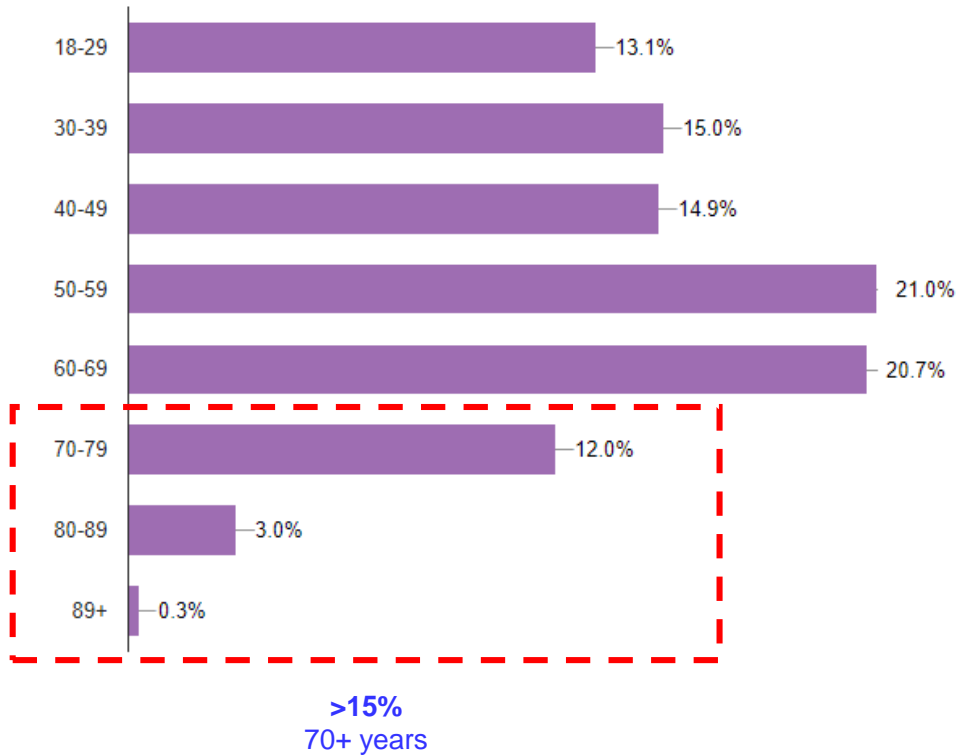


100+
Funded Partner Organizations

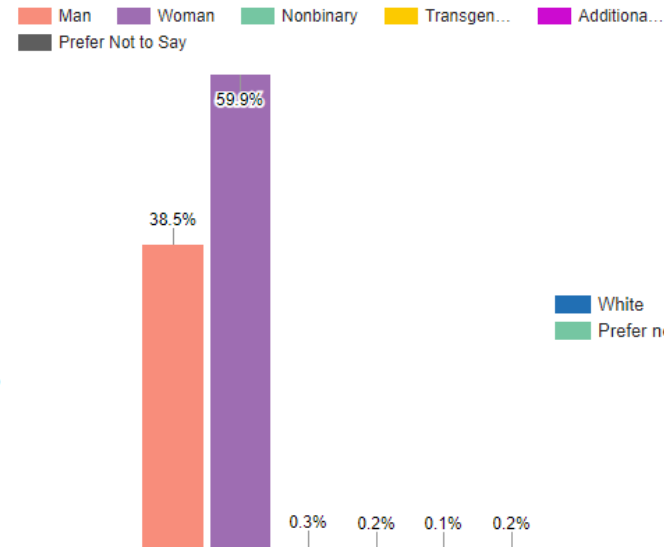
370+
Sites Collecting Samples and Measurements

AGE, SEX, & RACE/ETHNICITY

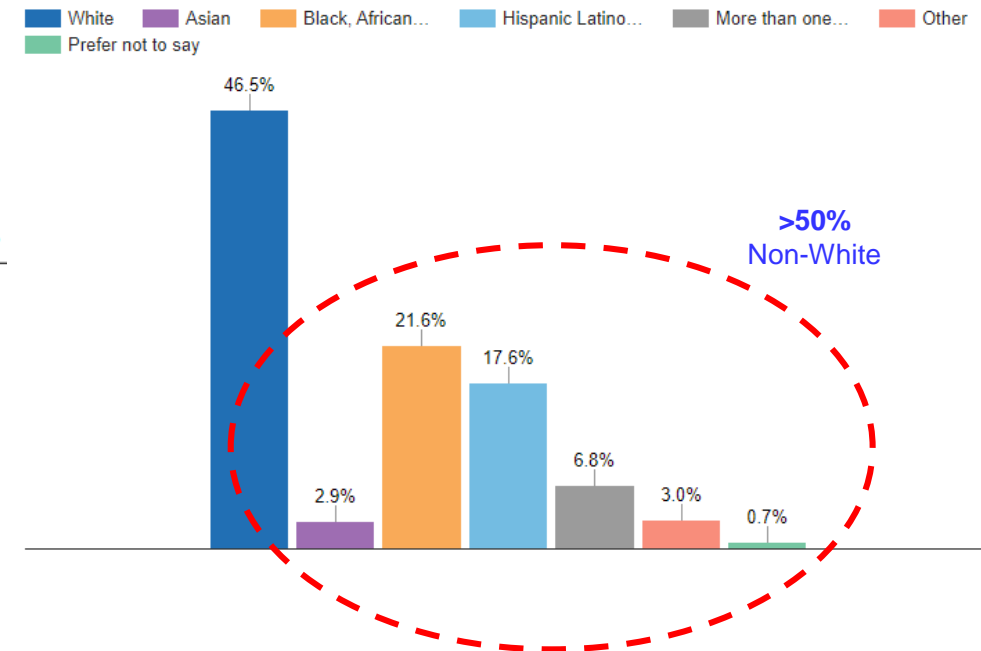
Age



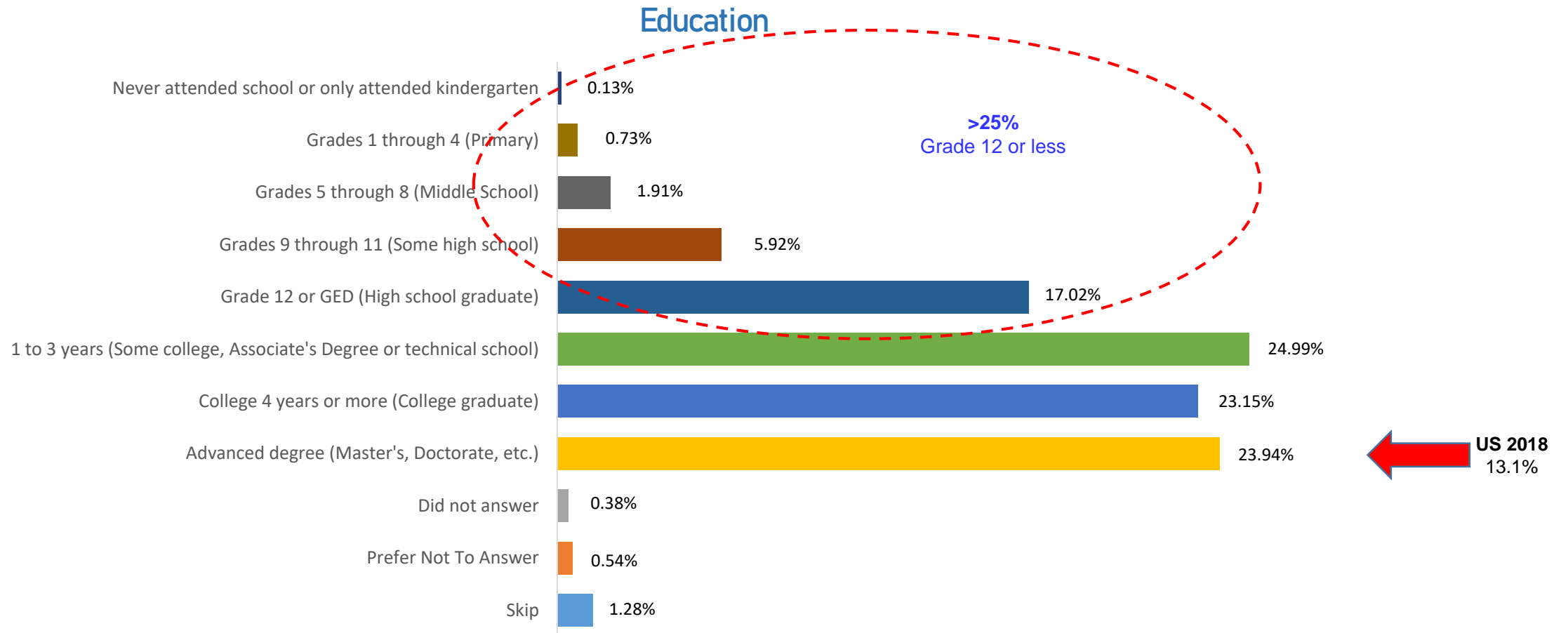
Gender Identity



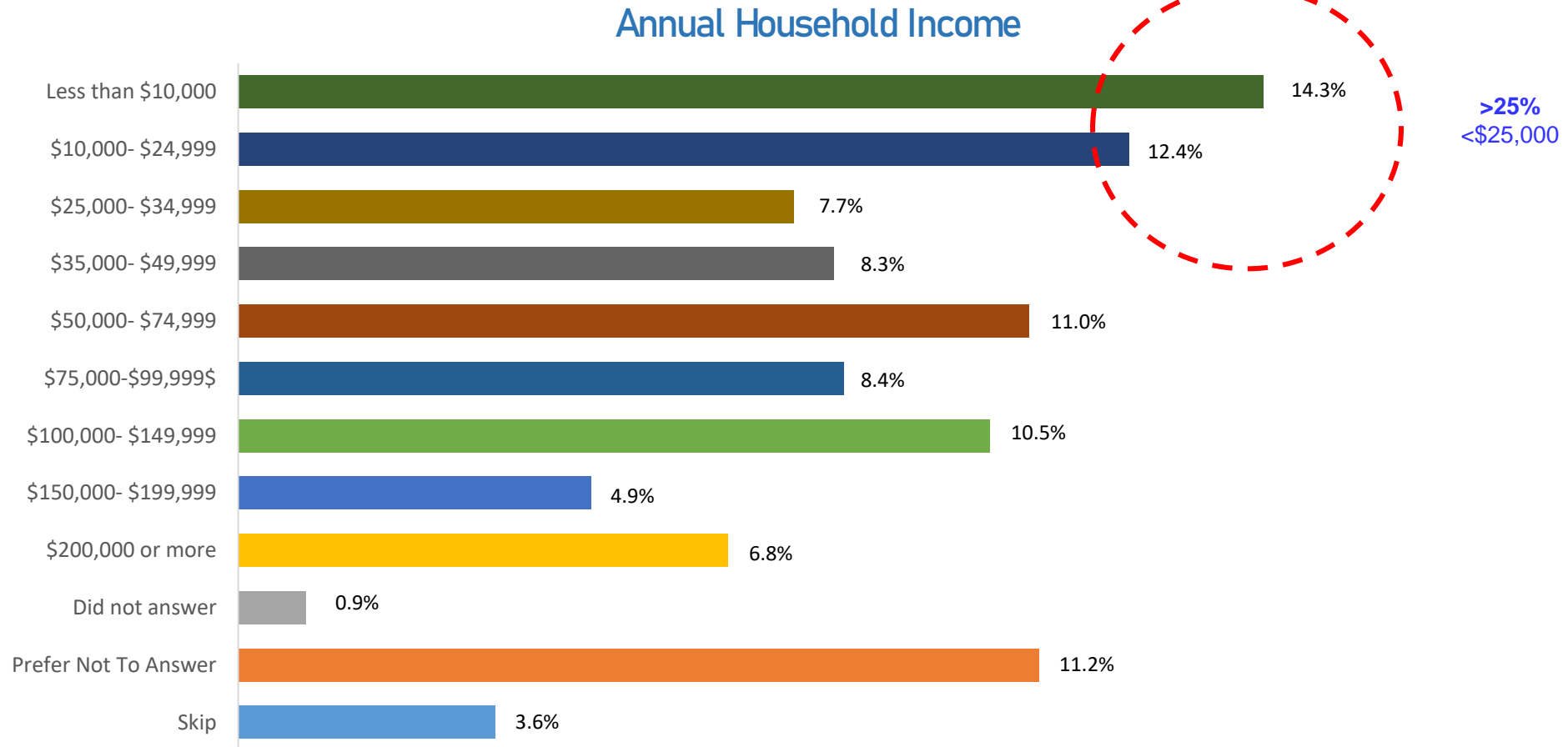
Race & Ethnicity



EDUCATION

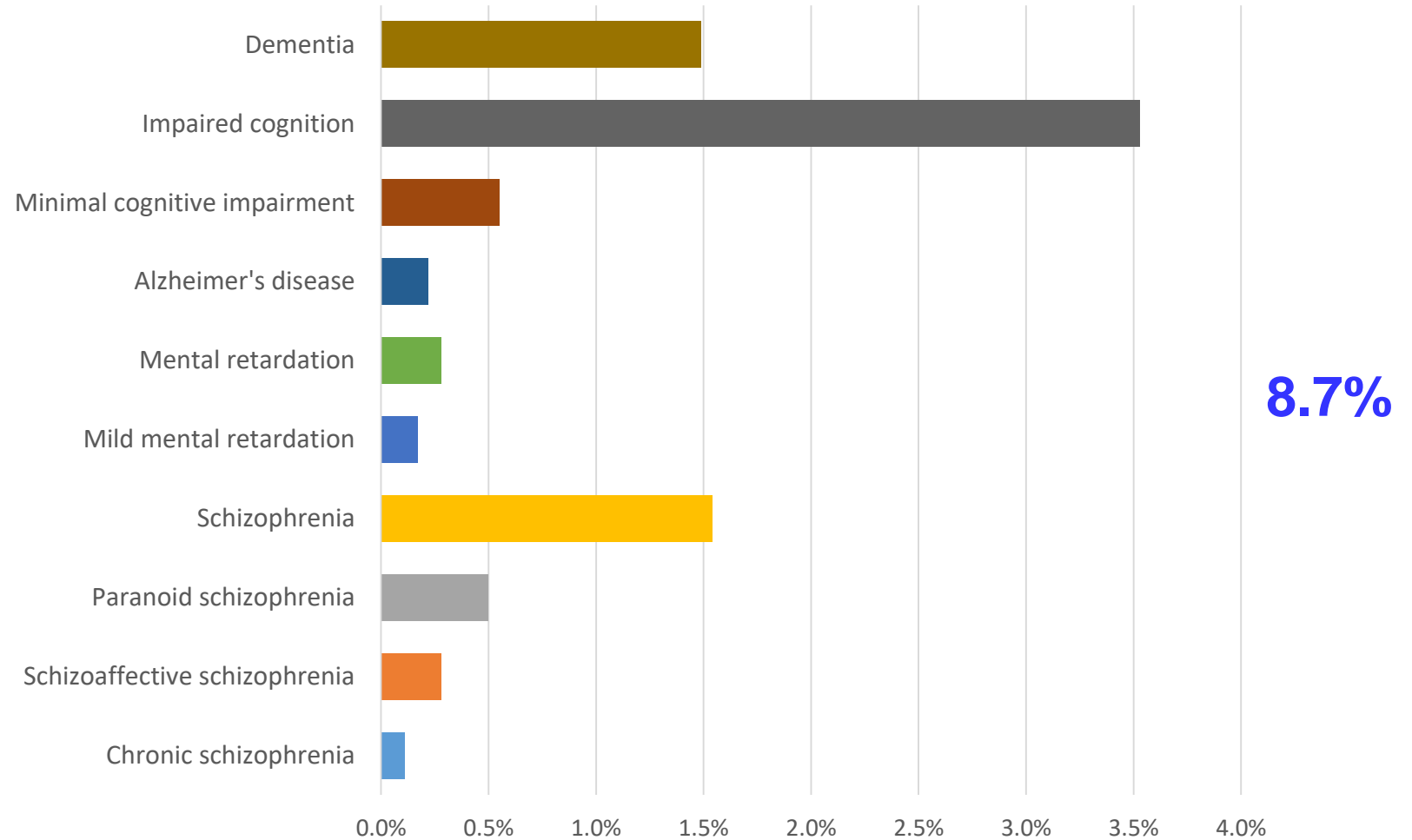


ANNUAL HOUSEHOLD INCOME



HEALTH STATES

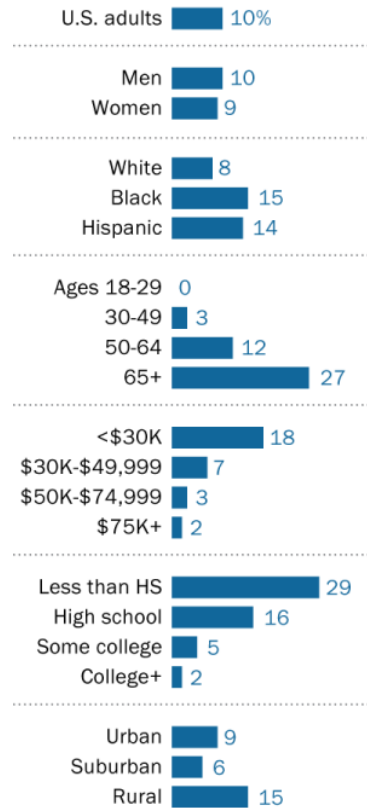
36,260 Participants



OTHER MAJOR USA POPULATION TRAITS

2019 INTERNET ACCESS

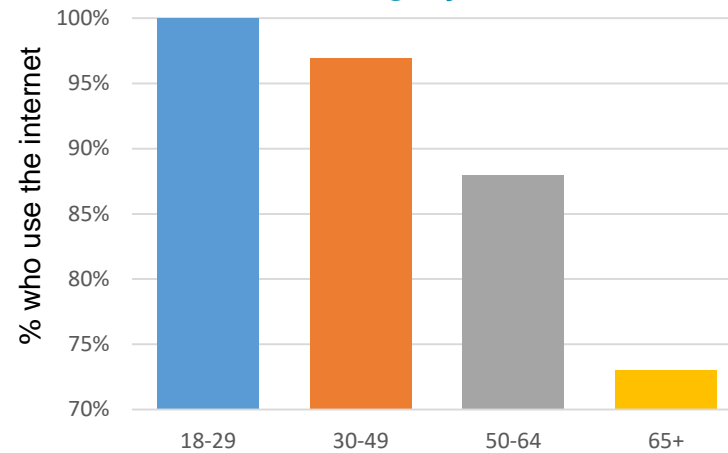
% of U.S. adults who say they do not use the internet



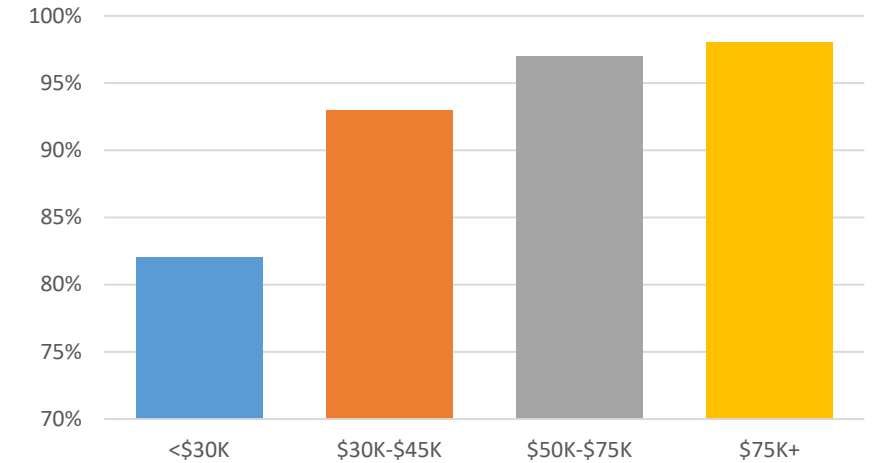
Note: Whites and blacks include only non-Hispanics. Hispanics are of any race.
Source: Survey conducted Jan. 8-Feb. 7, 2019.

PEW RESEARCH CENTER

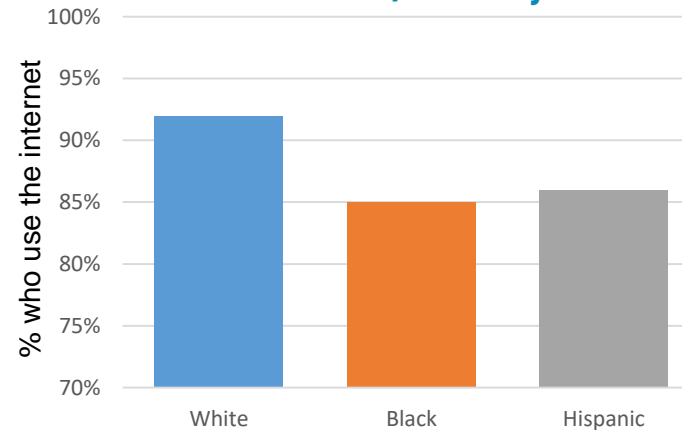
Age (years)



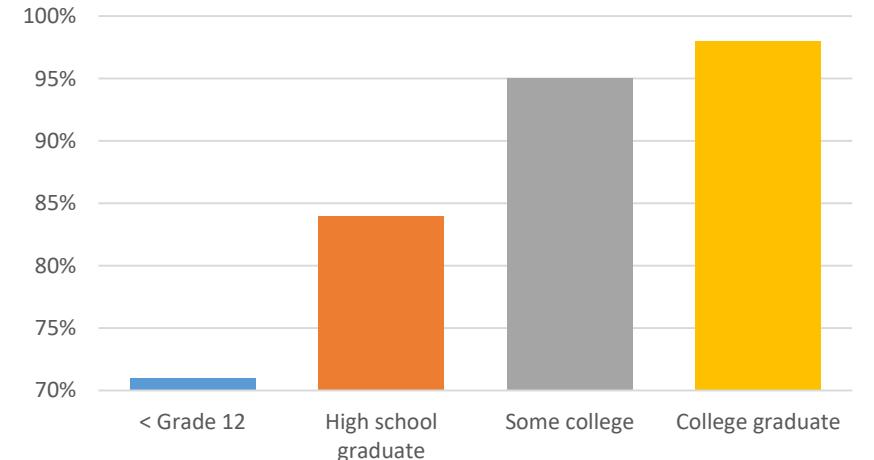
Income



Race/Ethnicity



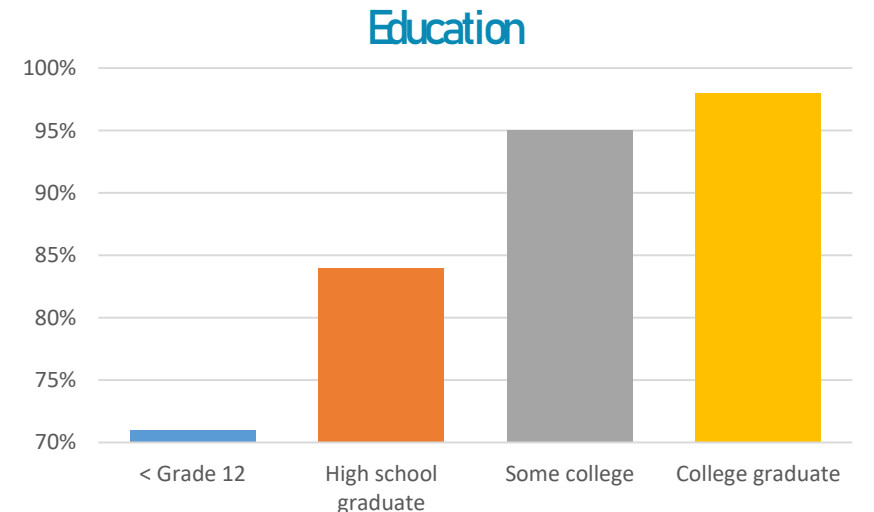
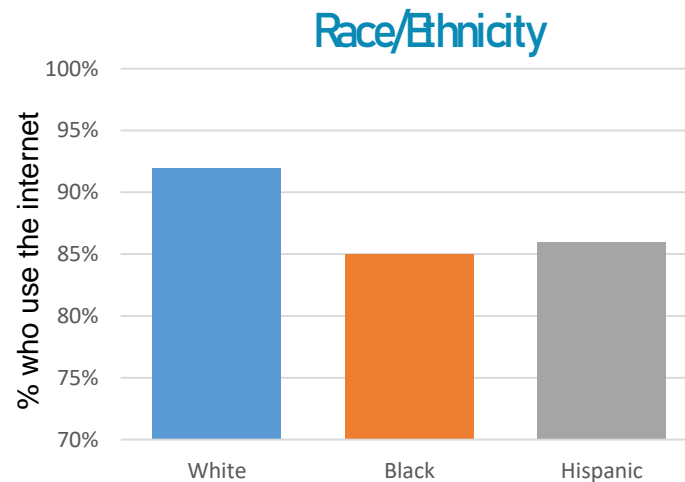
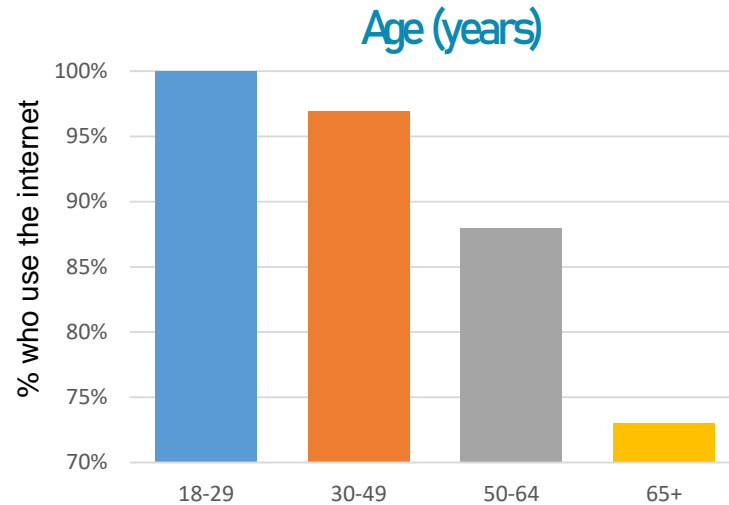
Education



OTHER MAJOR USA POPULATION TRAITS

2019 INTERNET ACCESS

Access to
broadband and non-
handheld computers
much lower



OTHER MAJOR USA POPULATION TRAITS

DIGITAL ILLITERACY



16%

16-65 yrs

35%

Hispanics

41%

< Grade 12

28%

55-65 yrs

22%

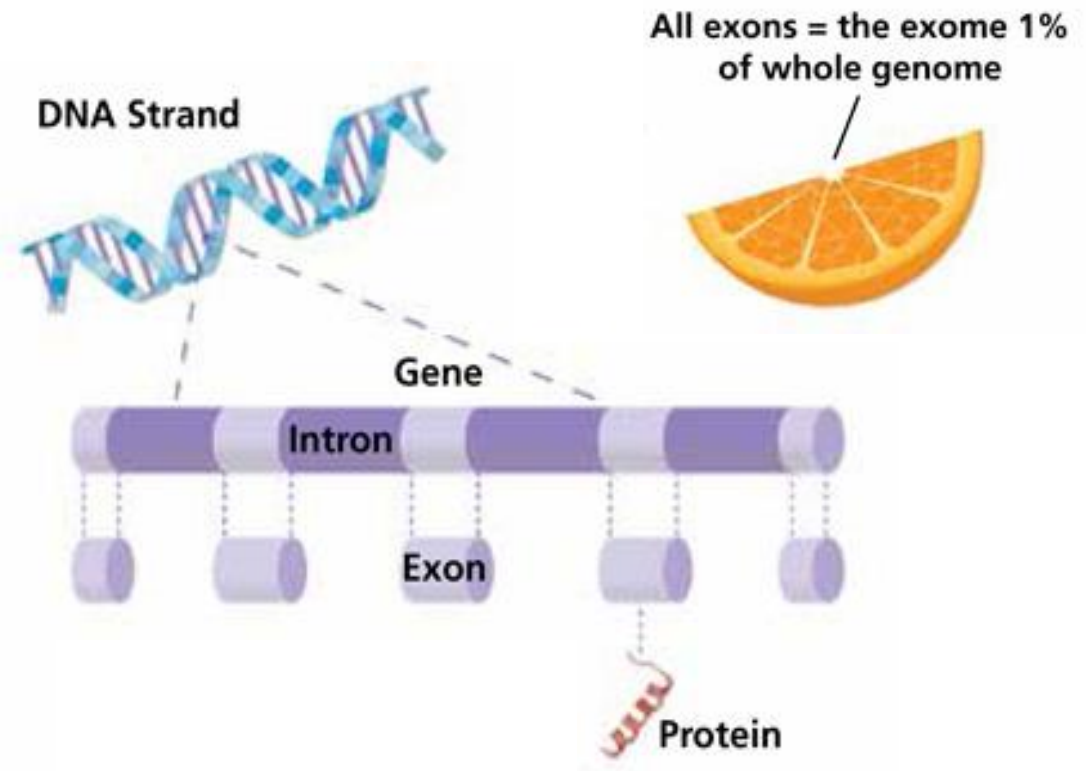
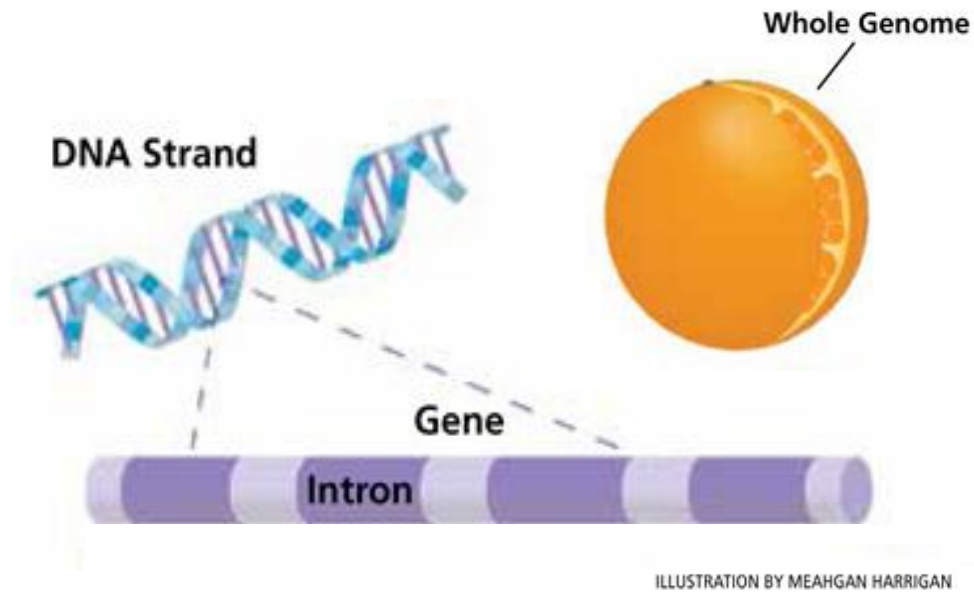
Blacks

17%

High school
Grad

WHAT CURATED GENETIC RESULTS/DATA MIGHT BE RETURNED?

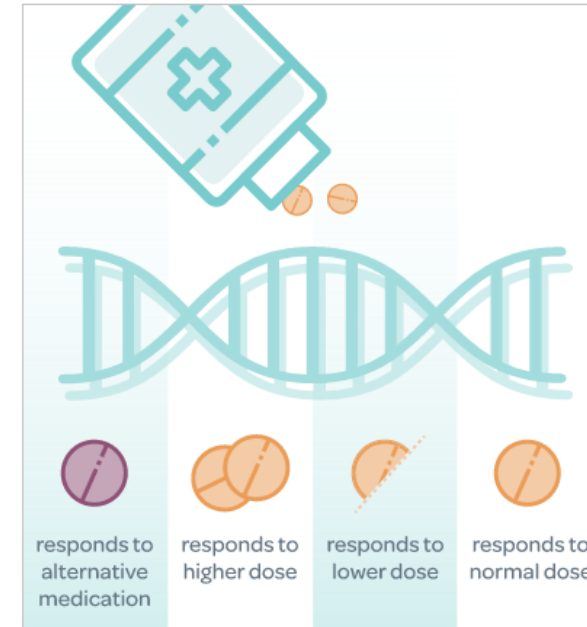
WHOLE GENOME SEQUENCING (WGS) AND WHOLE EXOME SEQUENCING (WES)



RETURNING CURATED GENETIC FINDINGS



ACMG 59



Pharmacogenes

CURRENT RECOMMENDATIONS FOR 2^o FINDINGS IN CLINICAL WES/WGS



Secondary finding – a result not related to test indication

Example: WES was ordered for a female patient with family history of breast and ovarian cancer

- Order was likely for **BRCA1/BRCA2** variants
- Data also reveal an **LDLR** mutation associated with familial hypercholesterolemia
- ACMG recommends reporting both **BRCA1/BRCA2 AND LDLR** findings to patient

CURRENT RECOMMENDATIONS FOR 2^o FINDINGS IN CLINICAL WES/WGS

Characteristics of “ACMG 59”:

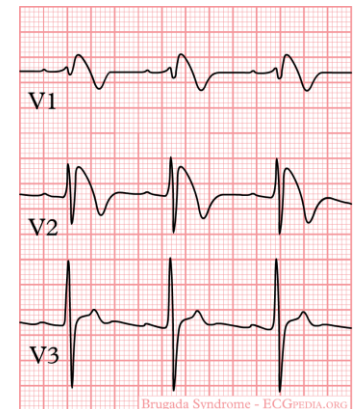
Condition has established **prevention**, **intervention**, or **treatment** strategies

Gene has **known** highly penetrant mutations

Example: Brugada syndrome and **SCN5A** Ile176Val (LOF)

Cardiac arrhythmia associated with sudden death

Implantable cardioverter-defibrillator is established prevention strategy



CHALLENGES FOR SECONDARY FINDINGS IN CLINICAL WES OR WGS

ACMG recommends reporting **known** and **expected** pathogenic variants

ClinVar Variant Classification



© American College of Medical Genetics and Genomics **ACMG STANDARDS AND GUIDELINES** | **Genetics in Medicine**

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD¹, Nazneen Aziz, PhD^{2,16}, Sherri Bale, PhD³, David Bick, MD⁴, Soma Das, PhD⁵, Julie Gastier-Foster, PhD^{6,7,8}, Wayne W. Grody, MD, PhD^{9,10,11}, Madhuri Hegde, PhD¹², Elaine Lyon, PhD¹³, Elaine Spector, PhD¹⁴, Karl Voelkerding, MD¹³ and Heidi L. Rehm, PhD¹⁵; on behalf of the ACMG Laboratory Quality Assurance Committee

KNOWN/EXPECTED PATHOGENIC VARIANTS: CLINVAR RESOURCE

Variants are classified by consequence



The **star rating** represents variant's review status, an indication of classification confidence



Practice Guideline

KNOWN/EXPECTED PATHOGENIC VARIANTS: CLINVAR RESOURCE

Variants are classified by consequence



The **star rating** represents variant's review status, an indication of classification confidence



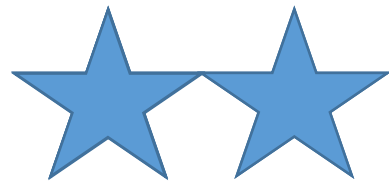
Reviewed by Expert Panel
(e.g., CPIC)

KNOWN/EXPECTED PATHOGENIC VARIANTS: CLINVAR RESOURCE

Variants are classified by consequence



The **star rating** represents variant's review status, an indication of classification confidence



Criteria provided; multiple submitters; no conflicts

KNOWN/EXPECTED PATHOGENIC VARIANTS: CLINVAR RESOURCE

Variants are classified by consequence



The **star rating** represents variant's review status, an indication of classification confidence



Criteria provided; multiple submitters; conflicts



Criteria provided; single submitter

KNOWN/EXPECTED PATHOGENIC VARIANTS: CLINVAR RESOURCE

Variants are classified by consequence



The **star rating** represents variant's review status, an indication of classification confidence

No assertion criteria provided

AN APPLICATION OF REPORTING 2^o FINDINGS IN CLINICAL WES/WGS

Example: Geisinger required “pathogenic” variants to have

1) ★★ or ★★★ ClinVar rating

OR

2) Predicted loss of function

OR

3) Both

JAMA
Network | **Open**[™]

Original Investigation | Genetics and Genomics

Exome Sequencing-Based Screening for *BRCA1/2* Expected Pathogenic Variants Among Adult Biobank Participants

CHALLENGES WITH RETURN OF ACMG RESULTS

ACMG recommendations evolve over time



2013

56 gene-condition pairs

including

MYLK and
Familial thoracic aortic
aneurysm and
dissection

2016

59 gene-condition pairs

now including

ATP7B and Wilson disease
BMPR1A and Juvenile polyposis
SMAD4 and Juvenile polyposis
OTC and ornithine transcarbamylase deficiency

CHALLENGES WITH RETURN OF ACMG RESULTS

The use of ACMG secondary findings recommendations for general population screening: a policy statement of the American College of Medical Genetics and Genomics (ACMG)

ACMG Board of Directors¹

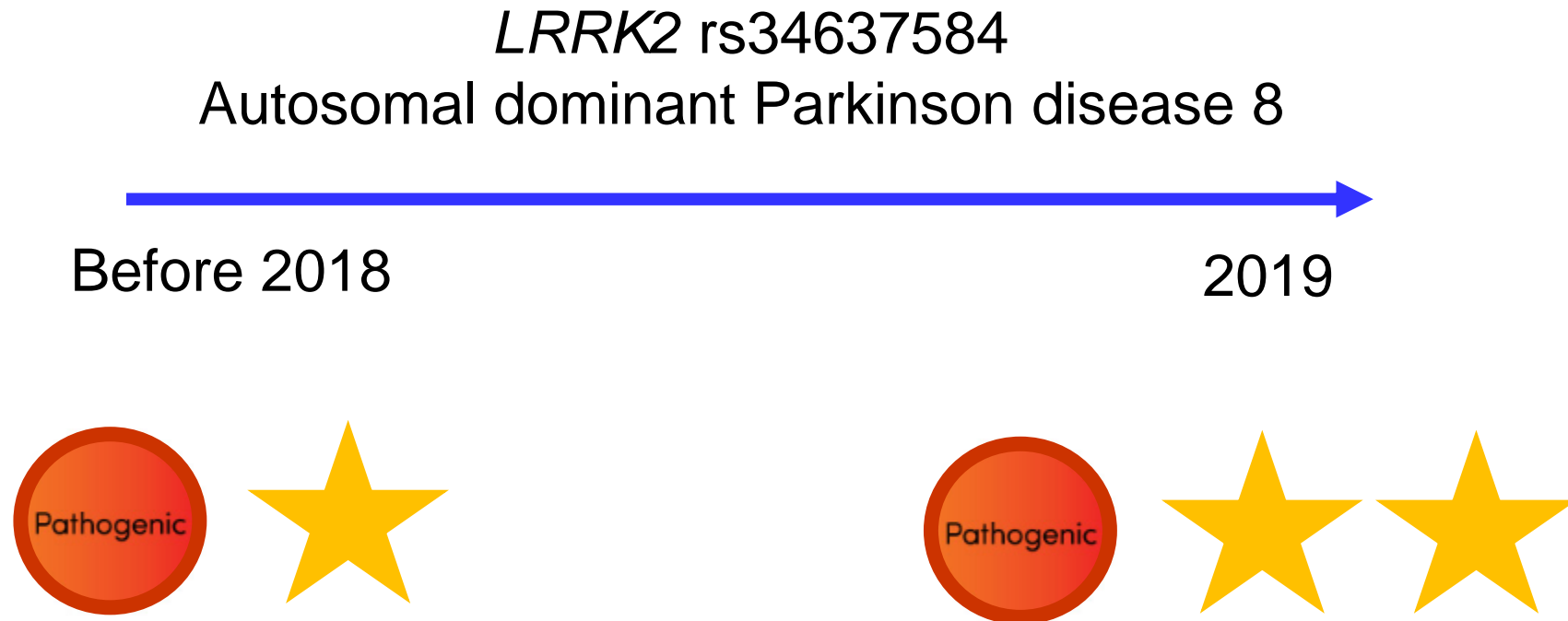


Population screening (?)

The American College of Medical Genetics and Genomics (ACMG) has previously published policy statements on the reporting of secondary findings in clinical exome and genome sequencing (ACMG SF v1.0 and ACMG SF v2.0), also known as the “ACMG 56” and “ACMG 59,” respectively.^{1,2} These recommendations specifically stated that “reporting some incidental [a.k.a. secondary] findings would likely have medical benefit for the patients and families of patients undergoing *clinical sequencing*” (ACMG board’s emphasis). The ACMG SF v2.0 list of genes was not validated for general population screening. The use of ACMG SF v2.0 for purposes other than reporting incidental findings after clinical sequencing is not endorsed by ACMG.

CHALLENGES WITH RETURN OF ACMG RESULTS

ClinVar annotations change over time

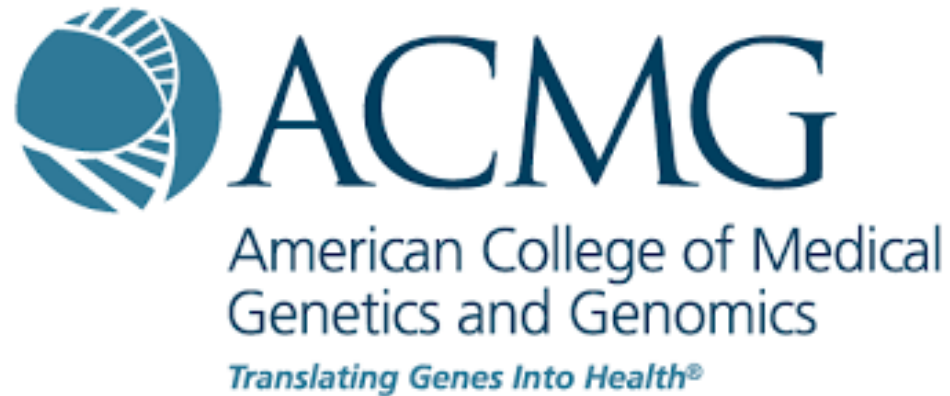


CHALLENGES WITH RETURN OF ACMG RESULTS

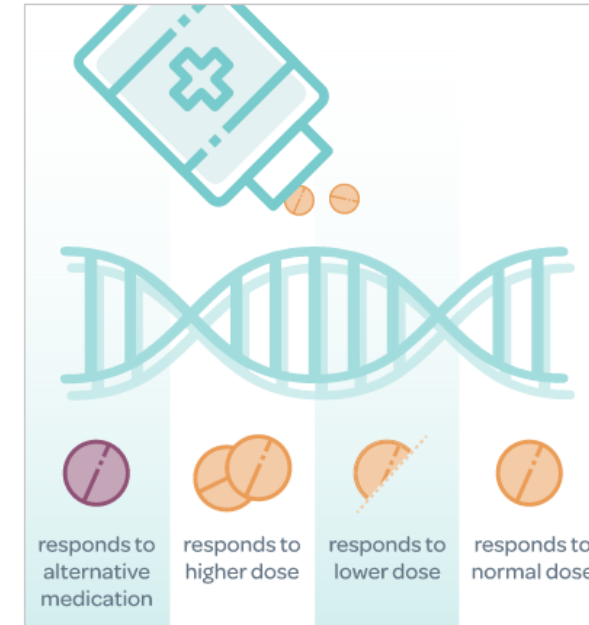
ClinVar annotations change over time



RETURNING CURATED GENETIC FINDINGS



ACMG 59



Pharmacogenes

THE GENETICS OF DRUG RESPONSE

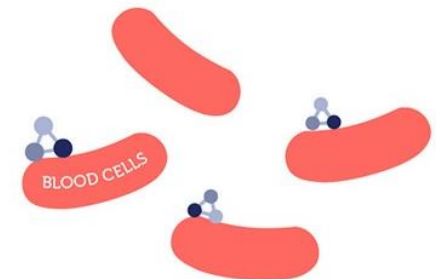
PHARMACOGENOMICS

FDA (2017) Drug Label for Warfarin dosing
 (Prevention and treatment of various thrombotic disorder)

VKORC1	CYP2C9				
	*1/*1	*1/*2	*1/*3	*2/*2	*2/*3
GG	5-7 mg	5-7 mg	3-4 mg	3-4 mg	3-4 mg
AG	5-7 mg	3-4 mg	3-4 mg	3-4 mg	0.5-2 mg
AA	3-4 mg	3-4 mg	0.5-2 mg	0.5-2 mg	0.5-2 mg



Warfarin reduces the body's ability to make Vitamin K which interferes with protein creation



Lower levels of clotting protein makes blood cells less likely to clot

THE GENETICS OF DRUG RESPONSE

PHARMACOGENOMICS

CPIC is an expert panel that

- Creates
- Curates
- Posts

Gene/drug Clinical Practice Guidelines that are

- Freely available
- Peer-reviewed
- Evidence-based
- Updatable

ClinVar



61 dosing guidelines

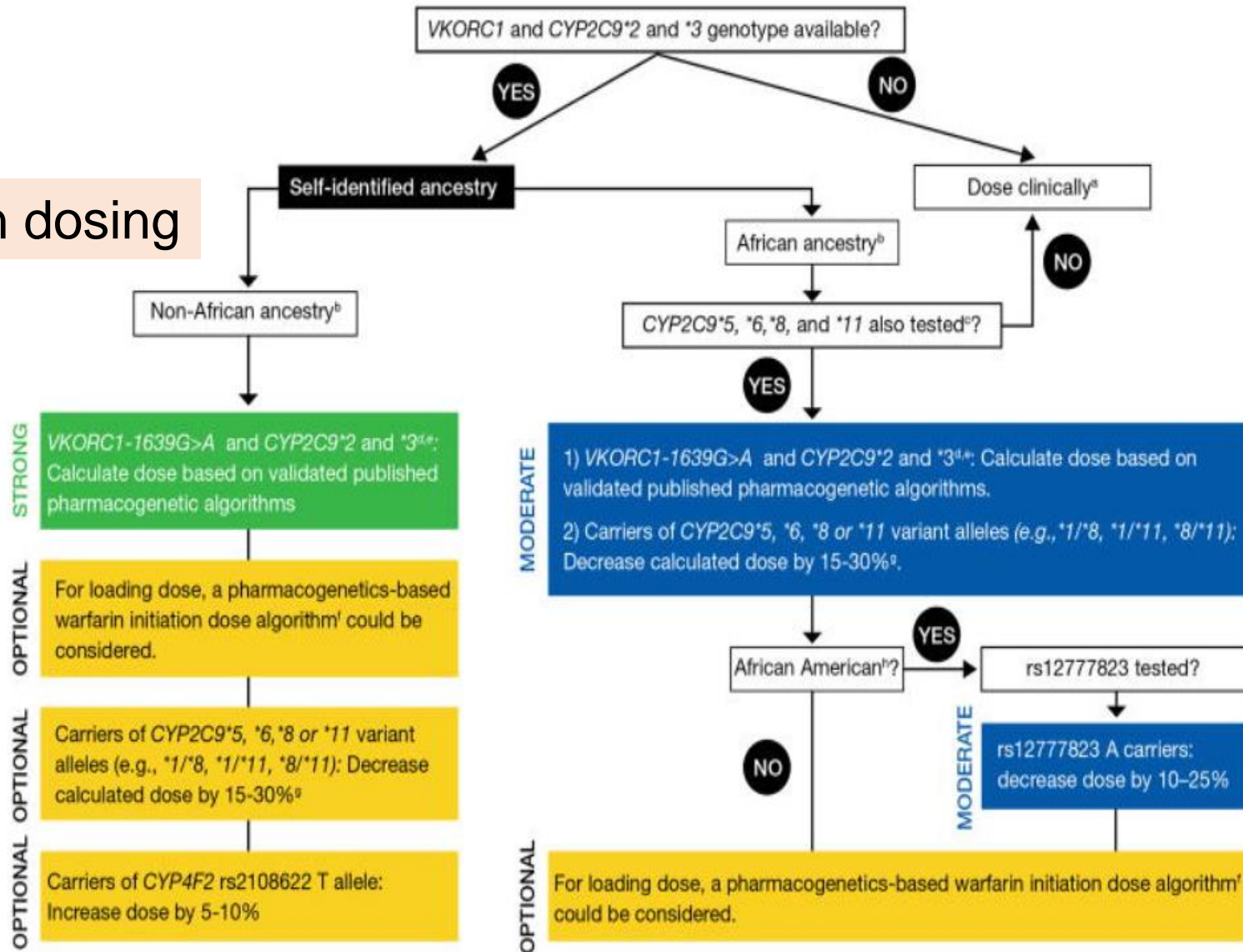
19 genes

46 drugs

As of 9/8/2019

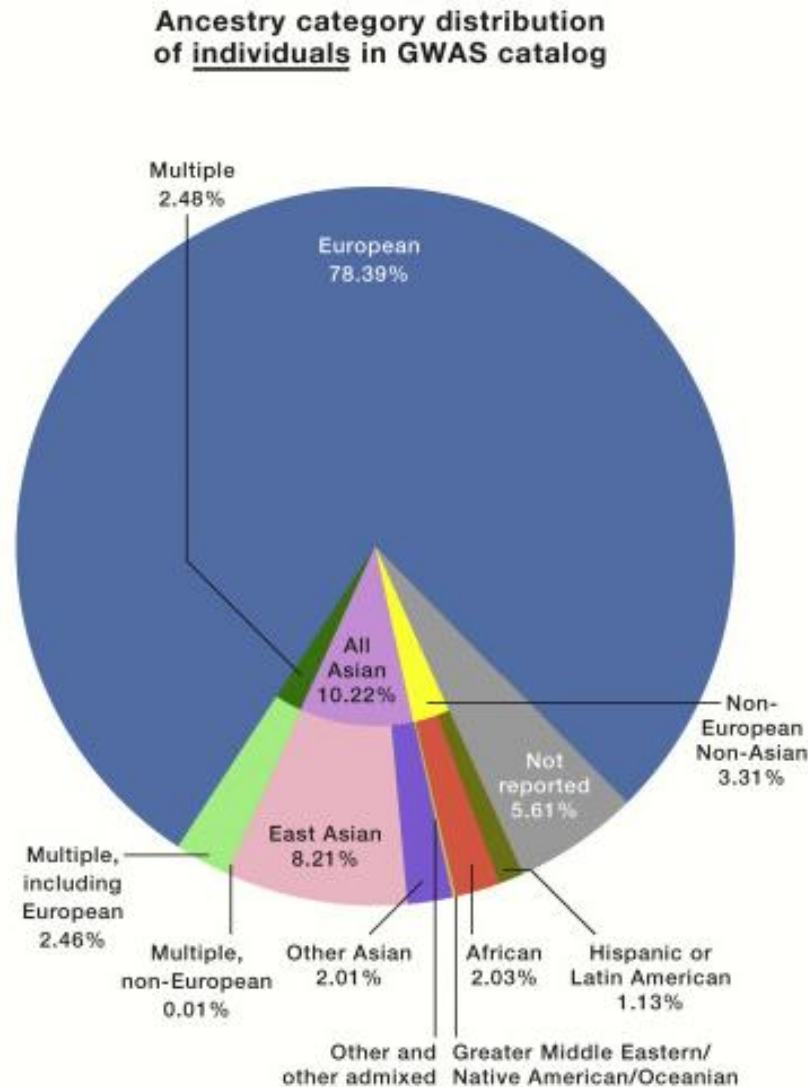
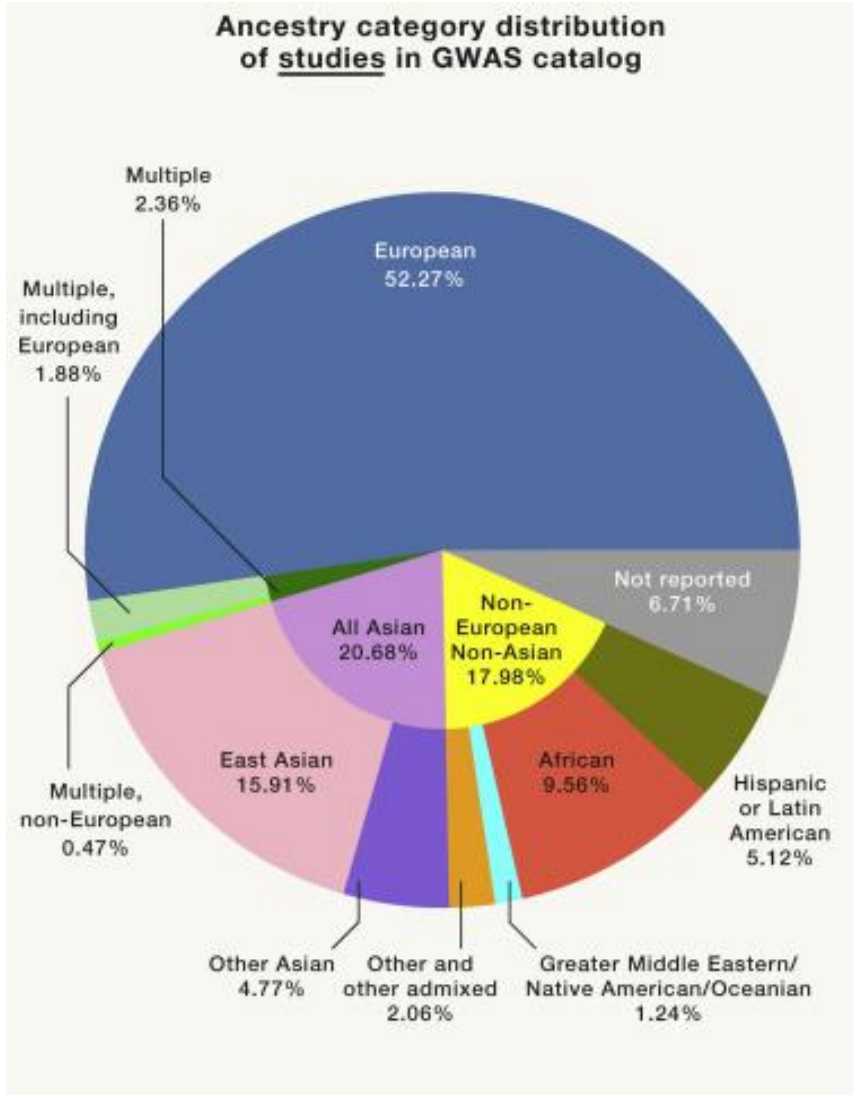
CHALLENGES WITH RETURN OF PGx RESULTS

Warfarin dosing



Pharmacogenomic recommendations **VARY** by race/ethnicity or genetic ancestry

CHALLENGES WITH RETURN OF PGx RESULTS



Not enough data available in non-Europeans; therefore, the **discovery** of pharmacogenes for diverse populations are lagging

**HOW MIGHT CURATED GENETIC
RESULTS/DATA BE RETURNED?**

RETURN OF CURATED RESEARCH GENETIC RESULTS

CONFIDENTIAL

Integrated BRCAAnalysis® with Myriad myRisk™ Hereditary Cancer

myRisk Genetic Result

MYRIAD myRisk Hereditary Cancer Powered by myVision

RECEIVING HEALTHCARE PROVIDER	SPECIMEN	PATIENT
Physician Name, MD Myriad Oncology Partners 320 Wakara Way Salt Lake City, UT 84108	Specimen Type: Buccal Draw Date: Apr 8, 2012 Accession Date: Apr 9, 2012 Report Date: Apr 30, 2012	Name: Patient Name Date of Birth: Jan 12, 1950 Patient ID: 1144 Gender: Female Accession #: 00001144-BLD Requisition #: 000000

ORDERING PHYSICIAN: Physician Name, MD

RESULT: POSITIVE—CLINICALLY SIGNIFICANT MUTATION IDENTIFIED

Note: "CLINICALLY SIGNIFICANT," as defined in this report, is a genetic change that is associated with the potential to alter medical intervention.

GENE	MUTATION	INTERPRETATION
BRCA1	c.68_69del (p.Glu23Valfs*17) Heterozygous	HIGH CANCER RISK This patient has Hereditary Breast and Ovarian Cancer (HBOC) syndrome.

DETAILS ABOUT: BRCA1 c.68_69del (p.Glu23Valfs*17); NM_007294.3; AKA: 187delAG

Functional Significance: Deletions – Abnormal Protein Production and/or Function
 The heterozygous germline BRCA1 mutation c.68_69del is predicted to result in the premature truncation of the BRCA1 protein at amino acid position 39 (p.Glu23Valfs*17).

Clinical Significance: High Cancer Risk
 This mutation is associated with increased cancer risk and should be regarded as clinically significant.

ADDITIONAL FINDINGS: VARIANT(S) OF UNCERTAIN SIGNIFICANCE (VUS) IDENTIFIED

GENE	VARIANT(S) OF UNCERTAIN SIGNIFICANCE	INTERPRETATION
CDKN2A (p16INK4a)	c.2T>A	UNCERTAIN CLINICAL SIGNIFICANCE There are currently insufficient data to determine if these variants cause increased cancer risk.

Additional Details About CDKN2A (p16INK4a): The heterozygous germline CDKN2A (p16INK4a) variant c.2T>A is located within the CDKN2A (p16INK4a) gene translation start codon and is predicted to result in abnormal protein translation. Start codon mutations are known to disrupt normal initiation of protein synthesis and are interpreted as pathogenic according to the recommendations of the American College of Medical Genetics (Richards CS et al. Genet Med, 10:294-300, 2008). However, an alternative in-frame methionine is located 9 amino acids downstream of the normal start codon. If this methionine were to be utilized as an alternative initiation codon, it would result in the deletion of the first 8 amino acids of the CDKN2A (p16INK4a) protein. At this time, there is insufficient information to determine whether or not this alternative methionine is utilized, and if the resulting shortened protein would be fully functional.

Details About Non-Clinically Significant Variants: All individuals carry DNA changes (i.e., variants) and most variants do not increase an individual's risk of cancer or other diseases. When identified, variants of uncertain significance (VUS) are reported. Likely benign variants (Favor Polymorphisms) and benign variants (Polymorphisms) are not reported and available data indicate that these variants most likely do not cause increased cancer risk. Present evidence does not suggest that non-clinically significant variant findings be used to modify patient medical management beyond what is indicated by the personal and family history and any other significant clinical findings.

Variant Classification: Myriad's myVision™ Variant Reclassification Program continuously performs ongoing evaluations of variant classifications. In certain cases, healthcare providers may be contacted for more clinical information or to arrange family testing to aid in variant classification. When new evidence about a variant is identified and determined to result in clinical significance and management change, that information will automatically be made available to the healthcare provider through an amended report.

MYRIAD, © 2014, Myriad Genetic Laboratories, Inc. | 320 Wakara Way, Salt Lake City, Utah 84108 | PH: 800-488-7422 FX: 801-584-3815 myRisk Genetic Result Page 1 of 2

The genetic test result

The genetic test result indicates whether a clinically actionable mutation is identified from the 25 genes analyzed.

If positive, the genetic mutation is detailed with appropriate nomenclature, and its clinical and functional significance.

Presence of genetic variants of uncertain significance (VUS) that are not currently considered clinically actionable, are reported.

Personal report via snail mail



but

Not easily scalable
 Expensive (?)
 Mobile population

RETURN OF CURATED RESEARCH GENETIC RESULTS

Font Size - +

Welcome | Try My46 | Secure Login | Create Account | Help | FAQ

My46 | Manage My Genome | Learning Center | How My46 Works | Research

Print

Make your **genome** work for you.

Start

About Us

My46 is an innovative web-based tool that enables individuals to manage their own genetic testing results.

What is a Genome?

Your genome is your entire genetic code or all of the DNA in a cell.

What My46 Means

Most human cells have 46 chromosomes that provide the genetic instructions for a body to live, grow, and develop.

Through a Web-based tool

but



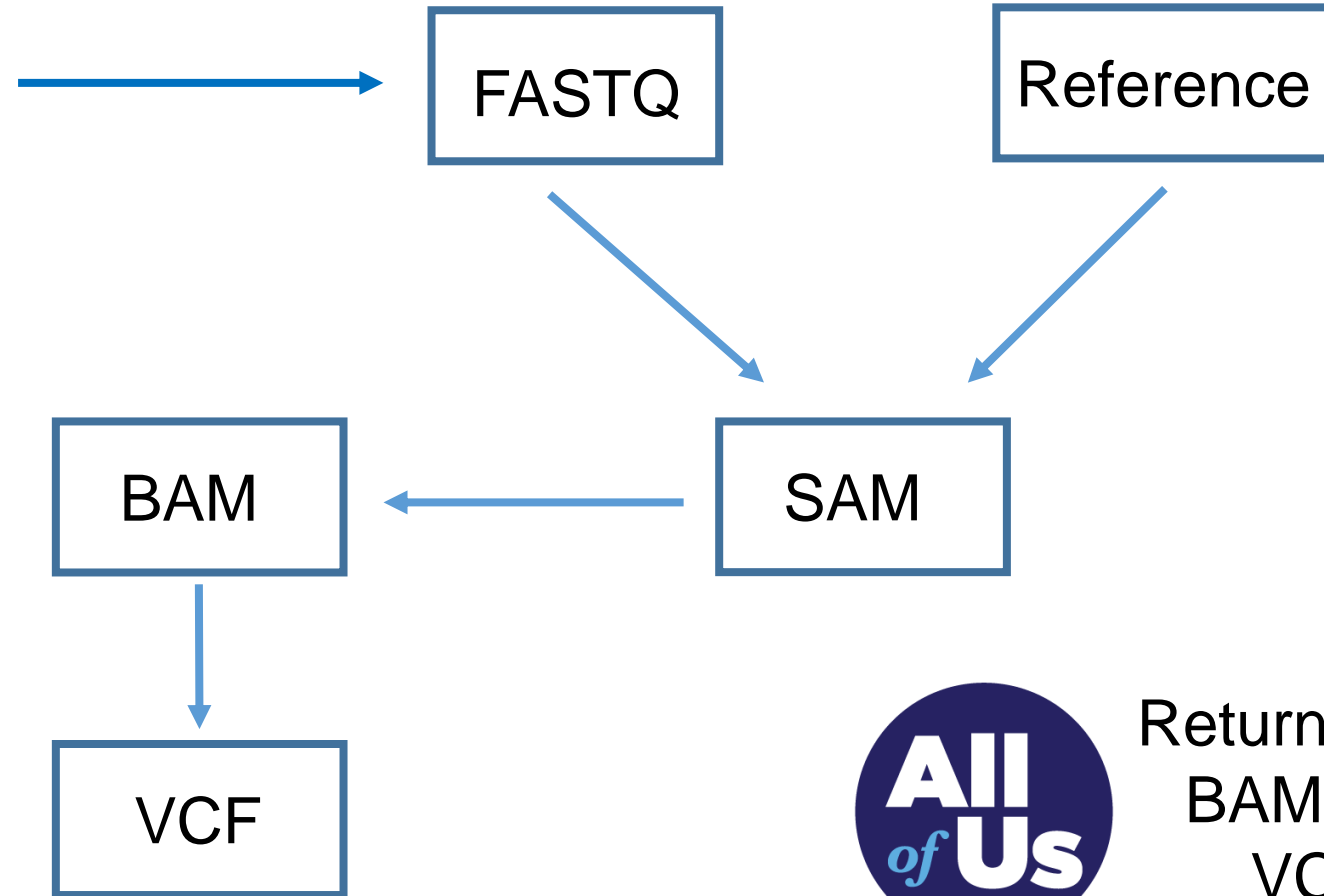
My46 enables individuals to manage their results from genetic testing, whether it is a single result being offered for return or the hundreds of results that can be offered for return from tests based on WES and WGS

**WHAT RAW WGS/WES DATA
MIGHT BE RETURNED?**

RETURNING **RAW** WGS/WES DATA



Sequencer



Returning SAM, BAM, and/or VCF (?)

RETURNING **RAW** WGS/WES DATA

Name	Description	WES file size (GB)	WGS files size (GB)
BAM	<i>Binary Alignment/Map format</i> : nucleotide sequence with corresponding quality scores mapped to the reference genome, derived from raw data files using an alignment algorithm	~5–15	~150–250
VCF	<i>Variant Call Format</i> : files for storing <u>variant bases relative to the reference genome</u> , which are derived from sequence files using a variant calling algorithm; usually annotated with allele frequency and predicted consequence	~0.02	~0.2

+ 1GB reference

RETURNING **RAW** WGS/WES DATA

VCF FORMAT

```
##fileformat=VCFv4.2
##FORMAT=<ID=GT,Number=1,Type=Integer,Description="Genotype">
##FORMAT=<ID=GP,Number=G,Type=Float,Description="Genotype Probabilities">
##FORMAT=<ID=PL,Number=G,Type=Float,Description="Phred-scaled Genotype Likelihoods">
#CHROM  POS      ID       REF      ALT      QUAL     FILTER  INFO    FORMAT  SAMP001  SAMP002
20      1291018  rs11449  G        A        .        PASS    .       GT      0/0      0/1
20      2300608  rs84825  C        T        .        PASS    .       GT:GP   0/1:.   0/1:0.03,0.97,0
20      2301308  rs84823  T        G        .        PASS    .       GT:PL   ./.:.   1/1:10,5,0
```

VCF considered “interpreted” data files and **may** require an investigational device exemption from the FDA

**HOW MIGHT RAW WGS/WES
BE RETURNED?**

RETURN OF **RAW** WGS/WES DATA

Font Size - +


Welcome | Try My46 | Secure Login | Create Account | Help | FAQ

My46 Manage My Genome Learning Center How My46 Works Research


Print

Make your **genome** work for you.

Start

About Us  My46 is an innovative web-based tool that enables individuals to manage their own genetic testing results.

What is a Genome?  Your genome is your entire genetic code or all of the DNA in a cell.

What My46 Means  Most human cells have 46 chromosomes that provide the genetic instructions for a body to live, grow, and develop.

Through a Web-based tool

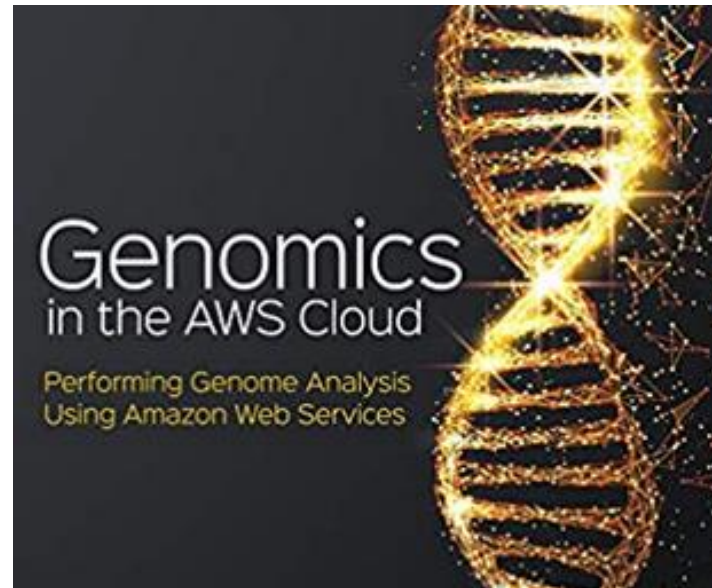
but



My46 enables individuals to manage their results from genetic testing, whether it is a single result being offered for return or the hundreds of results that can be offered for return from tests based on WES and WGS

RETURN OF **RAW** WGS/WES DATA

Other private storage tools



but

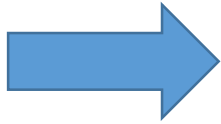


WHAT TO DO WITH RAW WGS/WES DATA?

WHAT TO DO WITH RAW WGS/WES DATA

Q: If you were given all of your raw genomic data from a research study, what would you do with this? (n=6944)

A: "I would seek out an interpretation of it" 62% (n=4320)



▶ 60% said "I'd analyse it myself" (n=2581)
(Within the different professional groups, these are the percentages who would analyse the data themselves: 81% genomic researchers, 68% genetic health professionals, 56% other health professionals 56% public)

▶ 57% said "I would ask for a referral to my local clinical genetics service" (n=2459)

▶ 43% said "I would ask my GP or Primary Care Physician" (n=1844)

▶ 41% said "would find a genomics researcher and ask them" (n=1775)

▶ 15% said "I would pay a commercial genetics company to analyse the data" (n=658)

▶ 5% had other suggestions: eg, "use google", "I would ask my bioinformatician colleagues", "I would share it on GitHub", "ask a genetic counsellor", "I would open source it to anyone online", "I would refer the raw data in a zip file to a company like 23andMe", "I would want information about interpreting the data before deciding how to proceed" (n=237)

WHAT TO DO WITH RAW WGS/WES DATA

gyanwali



Converting .vcf to .csv??

Personally I use vcftools

<http://vcftools.sourceforge.net/index.html>

The quickest way to convert a vcf to 23andMe assuming you have Linux, and your vcf already has dbSNP IDs assigned is with Plink, using the following plink command:

```
./plink --vcf example.vcf --recode 23 --chr 1-23 --out example.txt
```

How do you use that? I tried opening it but it doesnt exactly run.

gyanwali

WHAT TO DO WITH RAW WGS/WES DATA

gyanwali ◉



How do you use that? I tried opening it but it doesnt exactly run.

1. Download Plink 1.9 from here: <https://www.cog-genomics.org/plink/1.9/>.
2. Extract the zip file to the same folder where your VCF file is located.
3. Download these two text files to the same folder: [v3snps.txt](#) & [Xv3.txt](#)
4. Start the command prompt (type cmd in your windows search bar) and change the directory to where your file is located.
Command example (just copy-paste your file directory after cd): `cd your_file_directory`
5. Then type: `plink --vcf your_filename.vcf.gz --extract v3snps.txt --snps-only --out 23andme.txt`
6. Open the 23andme.txt, go to the end of the document and copy-paste the content of Xv3.txt, and save.
Done. Shouldn't take more than 5 minutes.

WHAT TO DO WITH **RAW** WGS/WES DATA

Using the UNIX/LINUX Environment...

A screenshot of the Coursera website showing the Genomic Data Science Specialization course. The page includes a search bar, a navigation menu, and a list of course features. The features listed are: 100% online courses, Flexible Schedule, Intermediate Level, Approx. 6 months to complete, and English. The course is described as being for the next generation sequencing data scientist and has 16,383 already enrolled students.

Genomic Data Science Specialization
Be a next generation sequencing data scientist.. Master the tools and techniques at the forefront of the sequencing data revolution.

Enroll for Free
Starts Sep 09

16,383 already enrolled

- 100% online courses**
Start instantly and learn at your own schedule.
- Flexible Schedule**
Set and maintain flexible deadlines.
- Intermediate Level**
- Approx. 6 months to complete**
Suggested 7 hours/week
- English**
Subtitles: English, Arabic

Q: Who of the public will be able to do this?

WHAT TO DO WITH RAW WGS/WES DATA

DNA Genics - DNA Kit Studio - <https://www.dnagenics.com> - v2.7

DNA Kit Studio v2.7
<https://www.dnagenics.com/> [Please, make a donation to support the development of this tool](#)

FamilyTreeDNA
Everyone has a story... what's yours?

RAW Tools | VCF Tools | Genome Tools | Support

VCF Converter | VCF MTDNA

VCF CONVERTER

This tool converts a VCF/gVCF to RAW file formats such as 23andme, Ancestry, FTDNA, MyHeritage, LivingDNA.

VCF/gVCF to RAW/VCF/gVCF Converter

VCF File Input *

Raw Data Output *

Output format: 23andme

Use Raw Data Template

Annotate Using Reference

Options

When using a RAW template match by by RSID by Position

Write SNPs when the RSID is not identified

Fix genotype orientation

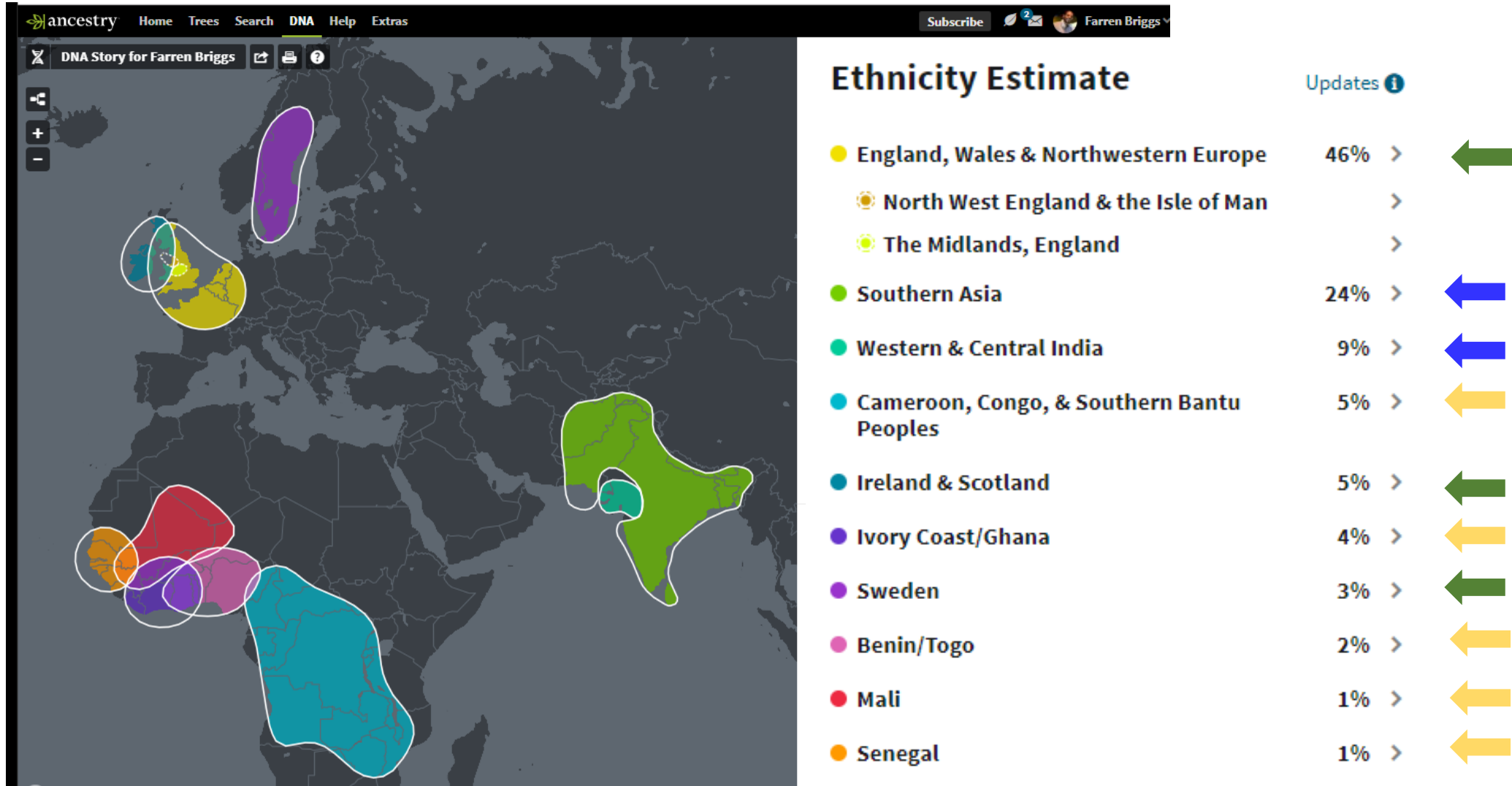
Extract all chromosomes

```
> Started conversion
> Matching by Position
> The source RAW file contains 668942 SNPs
> Conversion in progress...
> The output RAW file contains 668011 SNPs
> Finished conversion
```

HOW TO INTERPRET GENETIC DATA?

MY GENETIC PROFILE

* Immediate family approved *public* sharing



MY GENETIC TRAITS

Facial Hair Thickness

Your DNA suggests you (or your close male relatives) **have less thick facial hair**.

Does your beard—or the beards of the men in your family—match what your genes suggest? We looked at one genetic marker, or location in your DNA, to make our estimate. But other things can influence facial hair.

Genetics and Other Factors

Here's a breakdown of how genetic and non-genetic factors may influence this trait.

Known Genetics

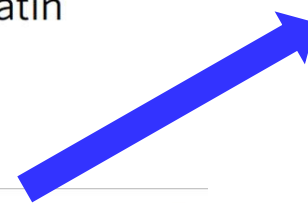
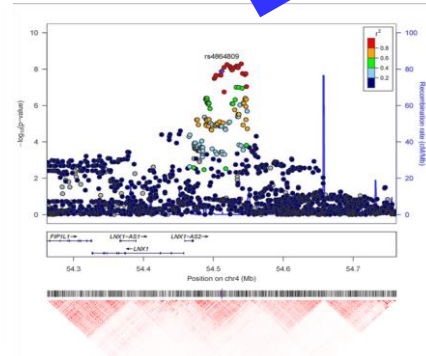
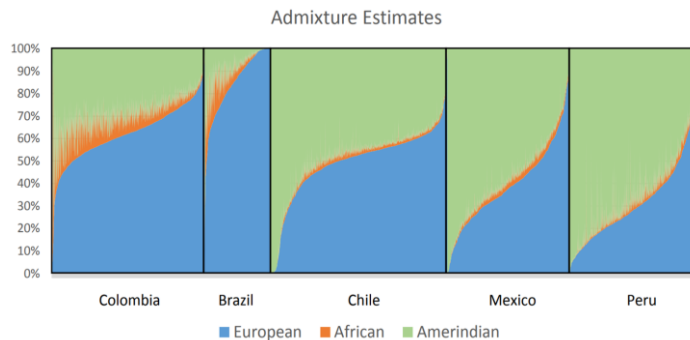
For this test we looked at one gene, *LMX1*, which seems to influence facial hair thickness.

ARTICLE

Received 13 Jul 2015 | Accepted 25 Jan 2016 | Published 1 Mar 2016

DOI: 10.1038/ncomms10815 OPEN

A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features



Hello, Farren

This test is shown to matches as Farren Briggs • Linked to Farren Briggs

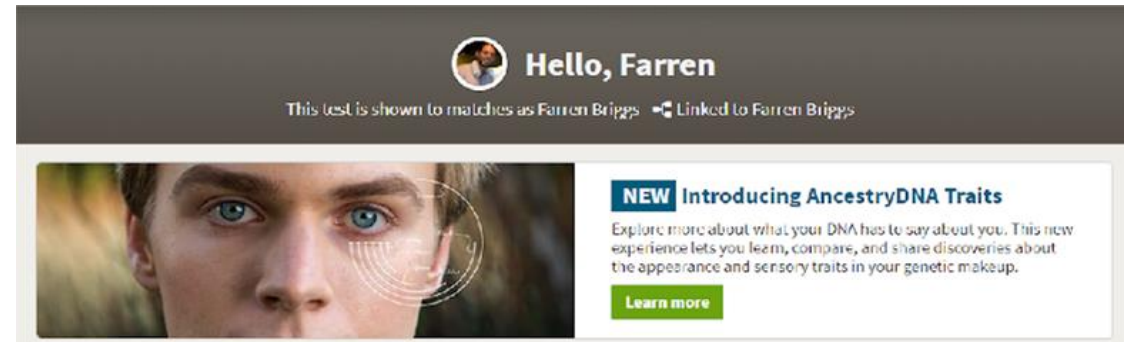
NEW Introducing AncestryDNA Traits

Explore more about what your DNA has to say about you. This new experience lets you learn, compare, and share discoveries about the appearance and sensory traits in your genetic makeup.

Learn more

Study	Population	Sample Size	G (ref allele)
The PAGE Study	Native Hawaiian	4,478	0.37
The PAGE Study	Asian	8,232	0.38
1000Genomes	South Asian	978	0.41
1000Genomes	East Asian	1,008	0.41
1000Genomes	African	1,322	0.41
gnomAD - Genomes	East Asian	1,546	0.42
The PAGE Study	South Asian	846	0.42
gnomAD - Genomes	African	8,676	0.45
The PAGE Study	African American	32,212	0.46
The PAGE Study	Dominican	3,782	0.52
The PAGE Study	Cuban	4,198	0.57
The PAGE Study	Central American	2,412	0.58
gnomAD - Genomes	Ashkenazi Jewish	288	0.58
The PAGE Study	Puerto Rican	7,844	0.58
The PAGE Study	Native American	1,246	0.61
gnomAD - Genomes	American	846	0.61
The PAGE Study	Mexican	10,684	0.61
1000Genomes	Europe	1,006	0.64
The PAGE Study	South American	1,958	0.64
gnomAD - Genomes	Other	1,086	0.65
1000Genomes	American	694	0.65
gnomAD - Genomes	European	18,842	0.67

MY GENETIC TRAITS



Cleft Chin

The DNA we tested tells us you probably **have a cleft chin**.

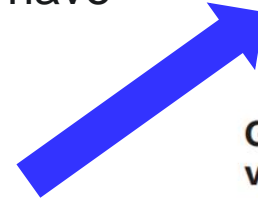
Omega-3

People with DNA like yours can sometimes have **average omega-3 levels**

Vitamin E

People with DNA like yours can sometimes have **slightly lower vitamin E levels**

I also probably have a **unibrow, skin pigmentation, lighter hair...**



Genome-Wide Association Study Identifies Three Common Variants Associated with Serologic Response to Vitamin E Supplementation in Men¹⁻⁴

Jacqueline M. Major,⁵ Kai Yu,⁵ Charles C. Chung,⁶ Stephanie J. Weinstein,⁵ Meredith Yeager,⁶ William Wheeler,⁷ Kirk Snyder,⁷ Margaret E. Wright,⁸ Jarmo Virtamo,⁹ Stephen Chanock,^{5, 6} and Demetrius Albanes^{5*}

Human Molecular Genetics, 2011, Vol. 20, No. 19 3876-3883
doi:10.1093/hmg/ddr296
Advance Access published on July 5, 2011

Genome-wide association study identifies common variants associated with circulating vitamin E levels



THIRD PARTY INTERPRETATION (TPI) TOOLS

"Personalize your list of supplements unique to your genetics"



"Learn how to fix your brain according to your DNA make-up."



HOW MANY TPI TOOLS WOULD PEOPLE USE?

Survey of 870 DTC-GT consumers who downloaded their genetic data

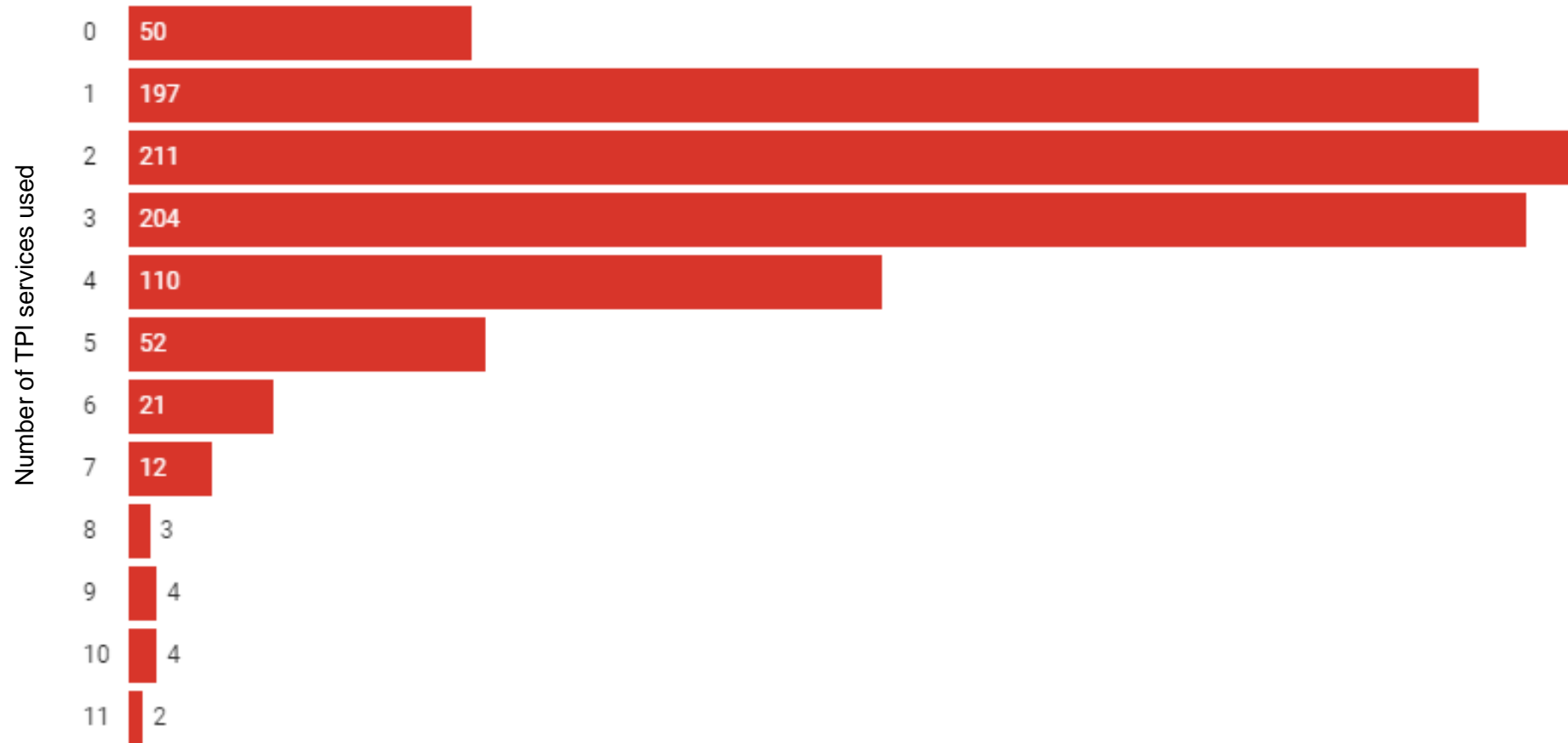
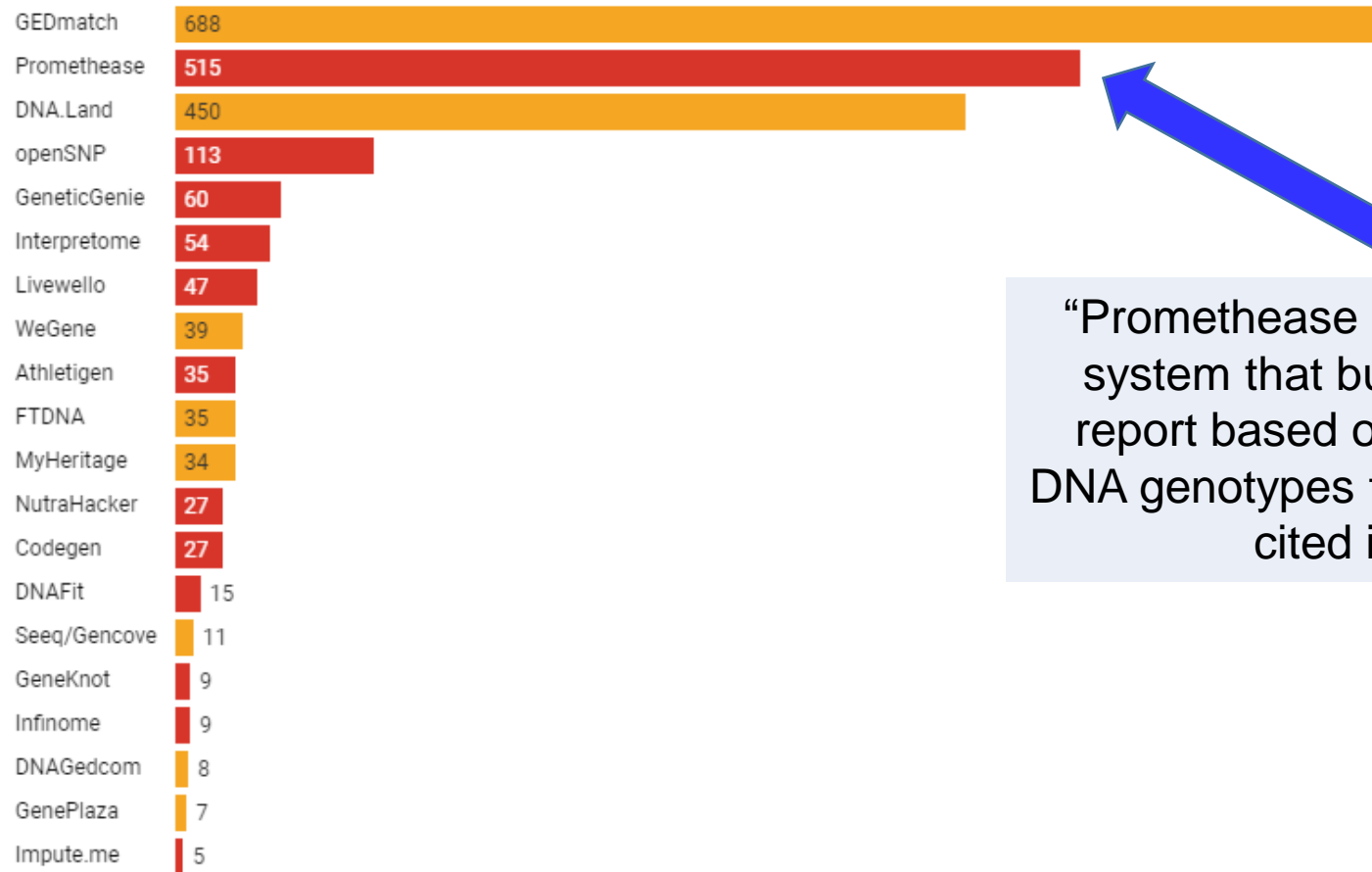


Chart: The Conversation, CC-BY-ND • Source: Nelson SC, et al. AJHG 105, 122-133, July 3, 2019

WHICH TPI TOOLS DID USERS TURN TO?

Survey of 870 DTC-GT consumers who downloaded their genetic data

Non-health-related Includes health-related



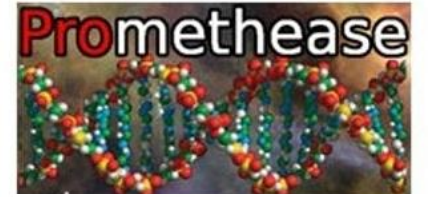
“Promethease is a literature retrieval system that builds a personal DNA report based on connecting a file of DNA genotypes to the scientific findings cited in SNPedia.”

Chart includes tools mentioned by five or more users.

Chart: The Conversation, CC-BY-ND • Source: [American Journal of Human Genetics](#), Nelson SC, et al. AJHG 105, 122-133, July 3, 2019

MY HEALTH-RELATED GENETIC TRAITS

\$12 USD + 47K SNPs from AncestryDNA



SNPedia

SNP	Summary	Repute	Publications	Magnitude
rs1426654	probably light-skinned, European ancestry		18	2.7
rs189798	normal high myopia risk		1	2
rs10825992	decreased high myopia risk	Good	1	2
rs2802292	Less likely to live to 100.	Bad	10	2.5
rs1935949	1.27X likelihood of 'exceptional longevity'	Good	1	2



WHERE TO SEEK HELP?



↑ 2 ↓
r/promethease · Posted by u/Smlanza5 8 days ago

Can someone please explain this to me? Does this mean I carry a mutation on the STXBP2 gene? I carry a separate mutation on the MUNC13-4 gene, which also causes HLH (my son passes away from this disorder). I need to know if I carry a separate HLH associated mutation or not.

AT&T 87% 7:58 AM

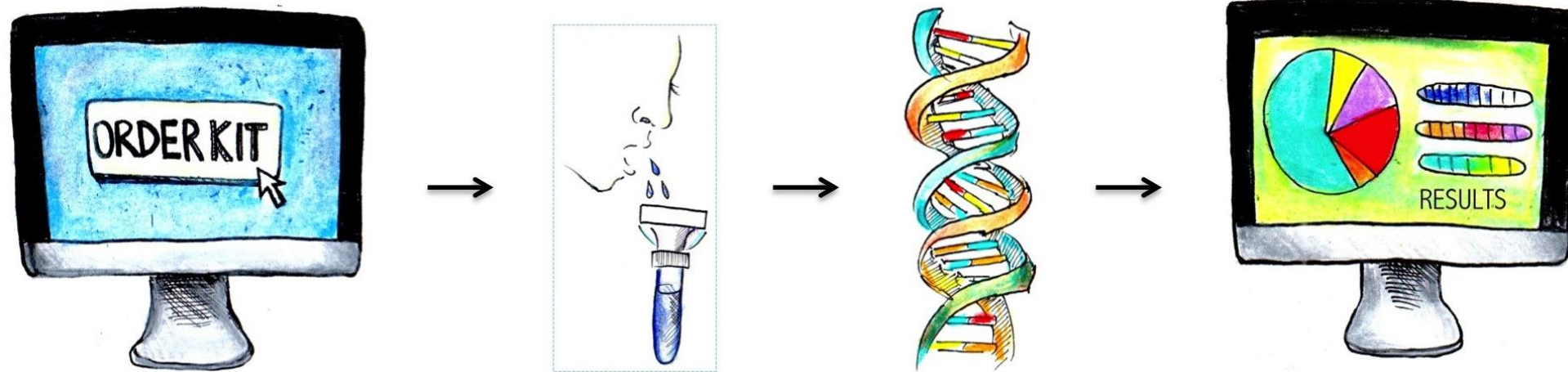
rs61736587(G;G)

common in clinvar

aka c.1621G>A (p.Gly541Ser) considered pathogenic for familial hemophagocytic lymphohistiocytosis (HLH) in ClinVar

[more info](#)

WHERE TO SEEK HELP?



Next steps:

PCP, Genetic Health Professional, Test family members, Share data.... Re-test ???

MACRO SUMMARY – *MIND THE GAPS*

WHO ARE OUR STUDY PARTICIPANTS?

EDUCATION, HEALTH STATUS, RESOURCES, LITERACY

HOW DO WE ENSURE EQUITY IN ACCESS, INTERPRETATION, & OPPORTUNITIES FOR ACTION?

DOES IT NEED TO BE EQUAL? WHAT ABOUT PRIVACY? RELATIVES?

HOW DO WE SUPPORT PARTICIPANTS OVER TIME?

ACMG/PGx/CLINVAR FINDINGS CHANGE...

WHO BEARS THE ONUS?

PARTICIPANT, RESEARCHER, FUNDING AGENCY, GOVERNMENT?



ACKNOWLEDGEMENTS



DANA CRAWFORD, PHD

Associate Professor
Department of Population and Quantitative Health Sciences
Cleveland Institute for Computational Biology
Case Western Reserve University



JESSICA COOKE BAILEY, PHD

Assistant Professor
Department of Population and Quantitative Health Sciences
Cleveland Institute for Computational Biology
Case Western Reserve University

Human Genetics (2019) 138:691–701
<https://doi.org/10.1007/s00439-019-02033-5>

REVIEW

Mind the gap: resources required to receive, process and interpret research-returned whole genome data

Dana C. Crawford^{1,2,3} · Jessica N. Cooke Bailey^{1,3} · Farren B. S. Briggs^{1,3}

Received: 4 April 2019 / Accepted: 27 May 2019 / Published online: 3 June 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Thank you
(and mind the gap)