The power of billions: data-driven innovation

Shirley Wu, Ph.D. | Director of Health Product @ 23andMe

Bay area Biotech-pharma Statistical Workshop
November 7, 2019

23andMe overview

- Founded in 2006
- Product helps individuals learn more about their health, traits, and ancestry through their DNA
- Customers can choose to participate in online research, fill out surveys
- 10M+ kits sold, ~80% of customers consent to participate in research
- **The largest research platform with both genetic and phenotypic information**

Unique interconnected business

Consumer

Providing direct access to learn about own genetics: health, traits, wellness, and ancestry

Empowering easy research participation

Research

Recruiting consenting participants from diverse ancestries and particular diseases of interest

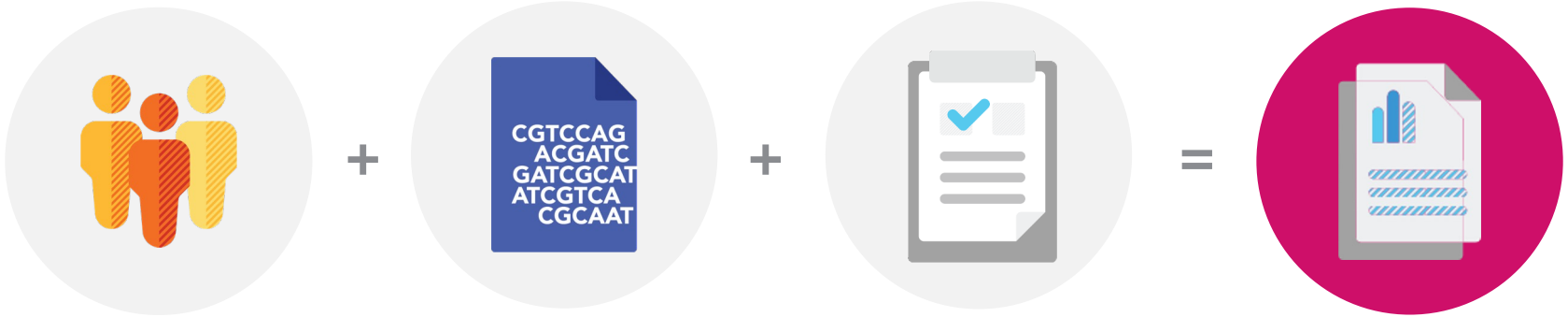
Making discoveries that help Therapeutics and power new insights for customers

Therapeutics

Leveraging data to accelerate drug discovery

Addressing unmet medical needs and rare conditions

The 23andMe research platform is a new paradigm in biomedical research



Each research participant provides 500,000+ genotypic data points

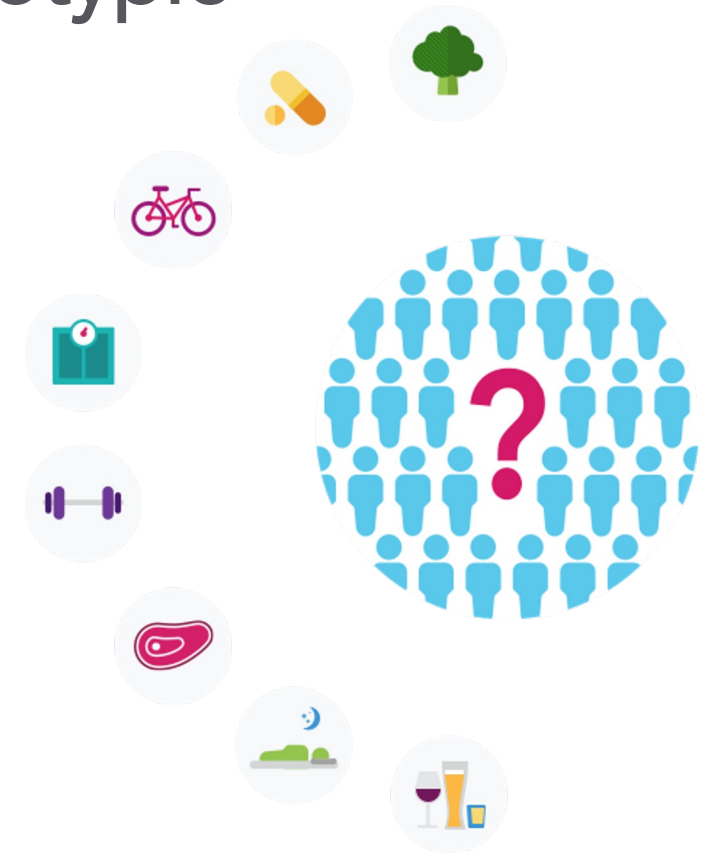
TCGTGCAT**A**CGCATCAGTAT

TCGTGCAT**T**CGCATCAGTAT

TCGTGCAT**A**CGCATCAGTAT

Participants contribute phenotypic data via online surveys

- Disease diagnoses
- Medication usage, side effects
- Behavior and lifestyle
- Environmental exposures
- Family history



Engaged consumers create the foundation for crowdsourced science

7M+

customers consented to research

75%

take ≥ 1 survey

~50%

log in quarterly

2B+

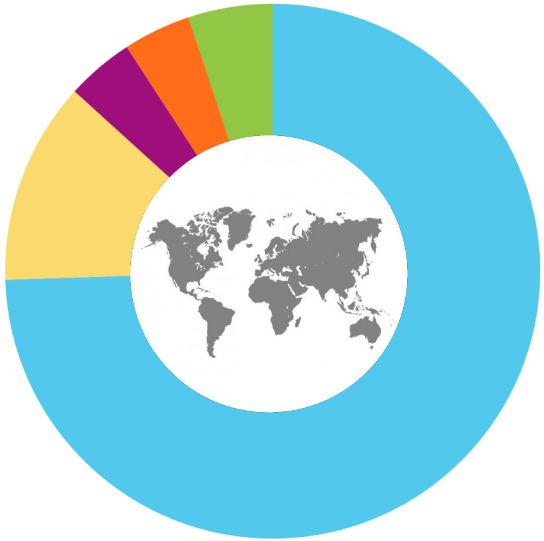
survey questions answered

What's unique about 23andMe?

Sheer size of disease cohorts

19,111 with Parkinson's	**1,019,600 APOE e4 carriers**	**107,005 with cancer (non skin)**	**744,463 with depression**
2,860,246 confirmed controls	2,954,800 confirmed controls	2,603,630 confirmed controls	2,177,226 confirmed controls
293,853 with asthma	**115,622 with psoriasis**	**819,371 with CVD**	**11,425 with colorectal cancer**
1,386,943 confirmed controls	2,806,720 confirmed controls	2,258,687 confirmed controls	2,945,316 confirmed controls

Database diversity and efforts to improve



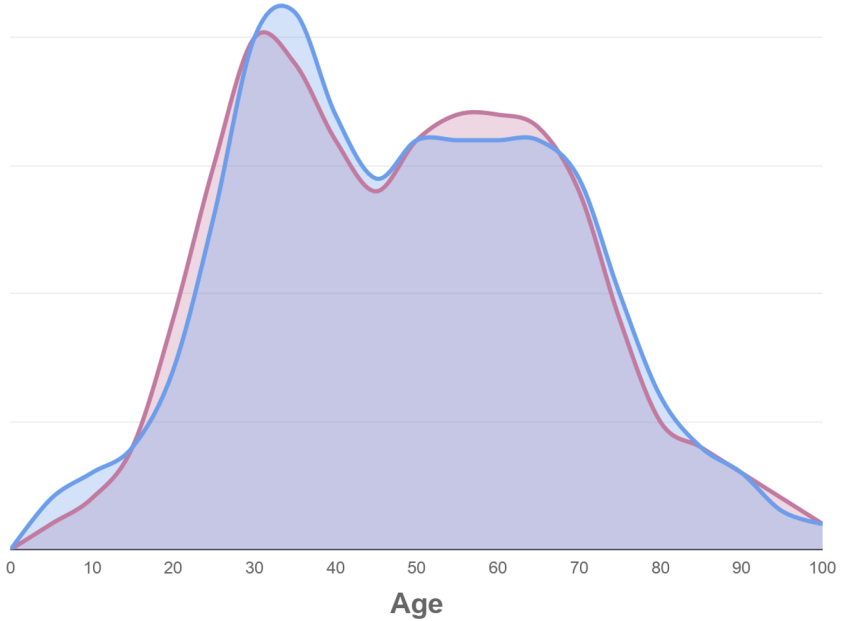
73% European

12% Latino

4% African American

4% East & South Asian

7% Other



55% Female

45% Male

Phenotypic breadth

Data sources

Longitudinal, high priority surveys

- Health intake
- Wellness

Stand-alone surveys and one-off questions (targetable)

Emerging data sources

- Passive data
- Lab values

Topics

Addiction
Autoimmune
Blood
Cancer
Cardiovascular
Cognitive
Diet
Drug usage
Drug efficacy
Drug side effects
Endocrine
Environment
Eyes
Food preferences
Gastrointestinal
Hair
Immune
Infection
Laterality
Lifestyle, sleep, sports
Longevity
Metabolic
Morphology
Musculoskeletal
Neurological
Personality
Pigmentation
Pregnancy
Psychiatric
Renal
Reproductive
Respiratory
Sensation
Sexual orientation
Skin
Teeth

What can you do with billions of data points?

Faster, better informed drug discovery

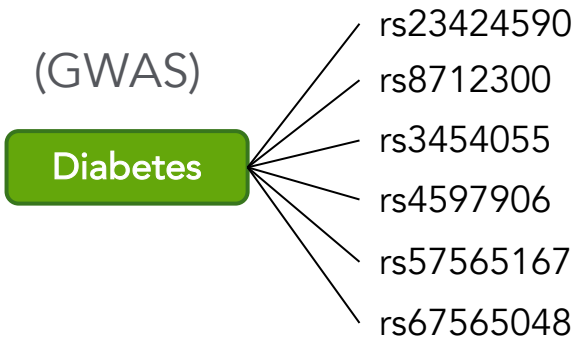
- Detect genetic associations with disease and identify potential drug targets
- Understand pathways, drug mechanism of action, and potential side effects to improve success rate
- (Future) Precisely stratify patient populations to make clinical trials more efficient and successful

GWAS Fine-mapping PheWAS Environmental data

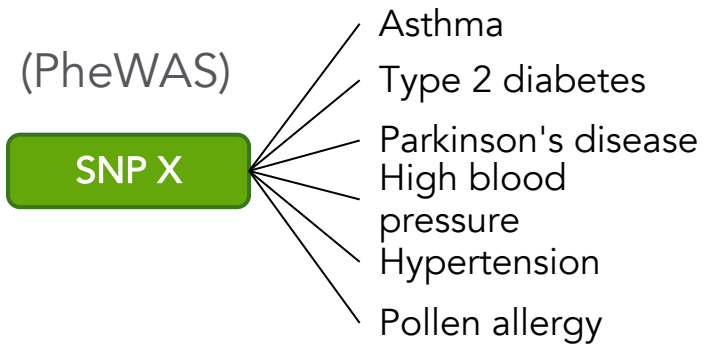
Application of public data Polygenic risk scores

GWAS and PheWAS at scale

Identifies SNPs associating with phenotype of interest

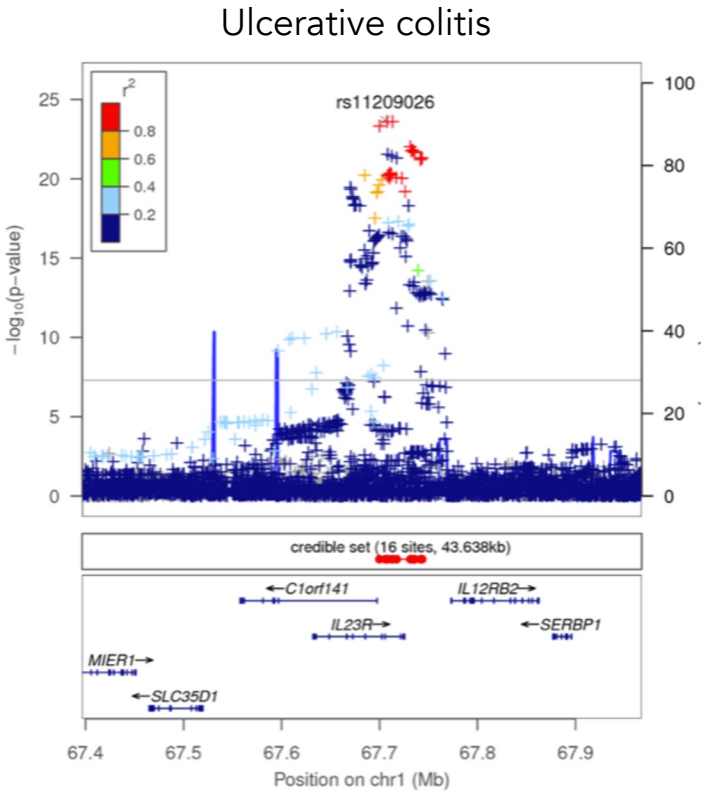


Identifies phenotypes that associate with SNP or gene of interest



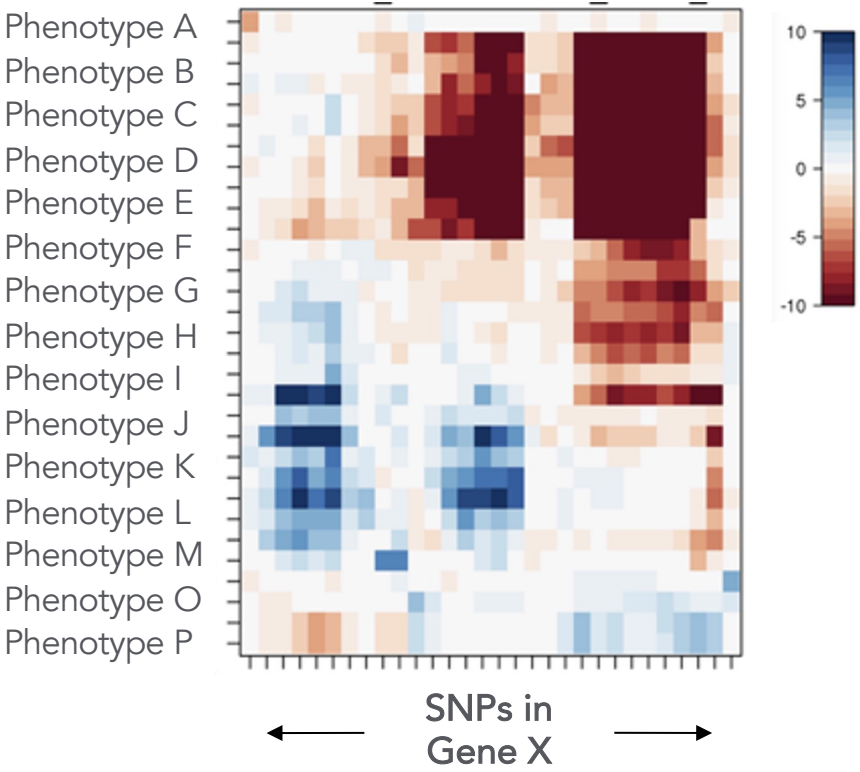
We regularly perform GWAS with millions of SNPs for each of hundreds of phenotypes relevant to human disease, and perform PheWAS on all 20,000+ protein-coding genes in the human genome.

Use all the data!



- **Interrogate all available sources of data:** external and internal, genetics, expression, biochemical, *in vitro*, *in vivo* etc.
- **Huge sample sizes** result in narrow LD regions and sharp peaks, making fine mapping to associated eQTLs and identifying likely target much easier

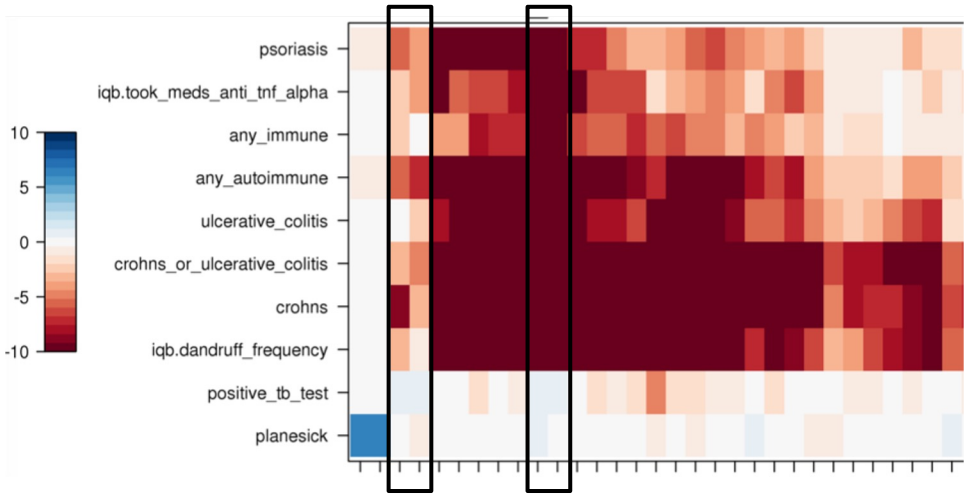
PheWAS for drug target analysis



- Clustering of related phenotypes can reinforce relevance of gene to disease biology and provide insight on potential therapeutic indications
- Opposing directionality of SNPs across different phenotypes can reinforce drug concept MOA and can reveal potential side effects of modulating gene function

IL23R PheWAS Reveals Potential Side Effects of IL-23 Blockade

SNPs identified with opposite directionality for autoimmune and infectious disease (tuberculosis) – reveals potential for increased risk of infectious disease with IL-23 antagonism.



ustekinumab, an anti-p40 antibody (IL-12/IL-23 antagonist) contains warnings for infectious disease risk, with specific warning for tuberculosis:

Serious Infections

STELARA® may lower your ability to fight infections and may increase your risk of infections. While taking STELARA®, some people have serious infections, which may require hospitalization, including tuberculosis (TB), and infections caused by bacteria, fungi, or viruses.

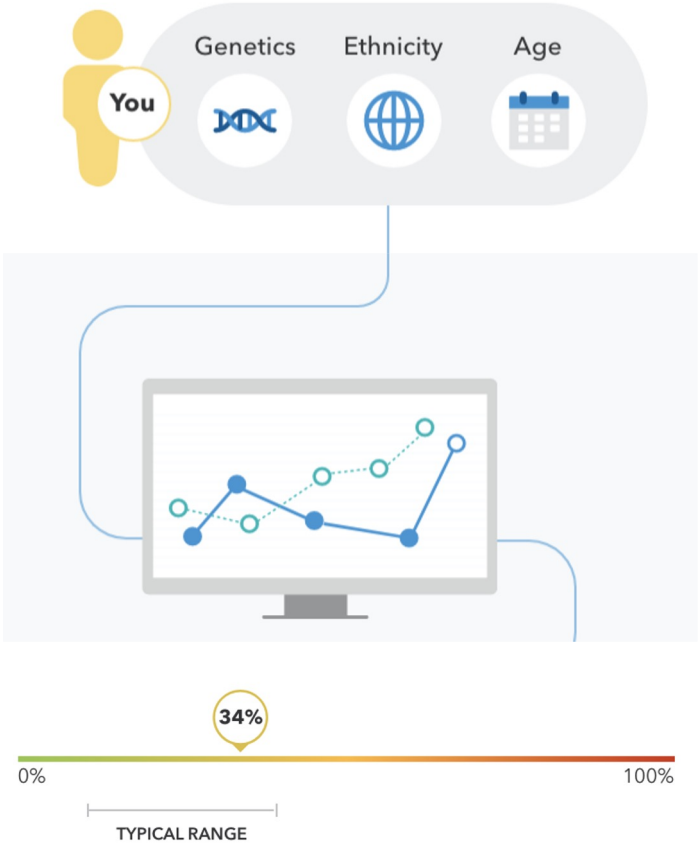
- Your doctor should check you for TB before starting STELARA® and watch you closely for signs and symptoms of TB during treatment with STELARA®.
- If your doctor feels that you are at risk for TB, you may be treated for TB before and during treatment with STELARA®.

You should not start taking STELARA® if you have any kind of infection unless your doctor says it is okay.

ustekinumab drug safety information: https://www.stelarainfo.com/

Polygenic scores

A Polygenic Score (PGS) captures the cumulative effects of many genetic factors, optionally with other factors like age, sex, PCs, and ethnicity, to predict how likely a person is to have a trait or condition.



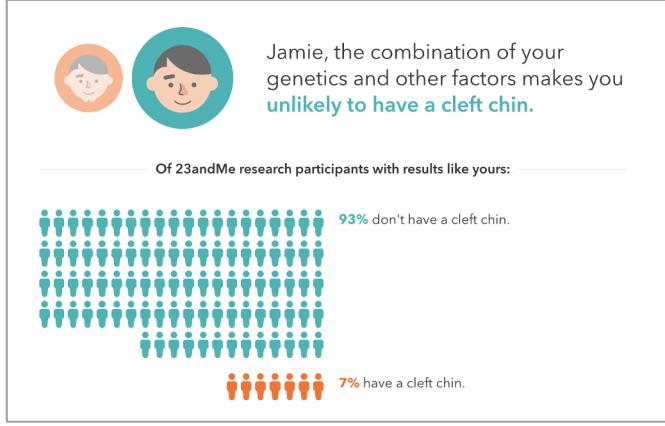
Delivering polygenic scores to consumers since 2015

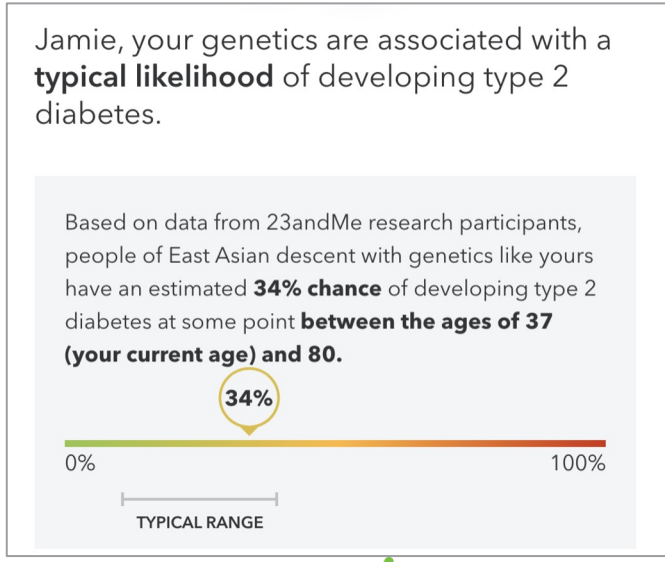
2015 — **16 polygenic scores for common traits** including cleft chin, sweet/salty taste preference, and male baldness

2017 — **Genetic Weight score, including gene-by-environment** analysis against lifestyle factors

2018 — **12 additional polygenic scores for traits** including wake-up time, mosquito attraction, and dandruff

2019 — **Type 2 Diabetes polygenic score** powered by data from more than **2.5 million research participants**





Delivering polygenic scores to consumers since 2015

2015 — **16 polygenic scores for common traits** including cleft chin, sweet/salty taste preference, and male baldness

2017 — **Genetic Weight score, including gene-by-environment** analysis against lifestyle factors

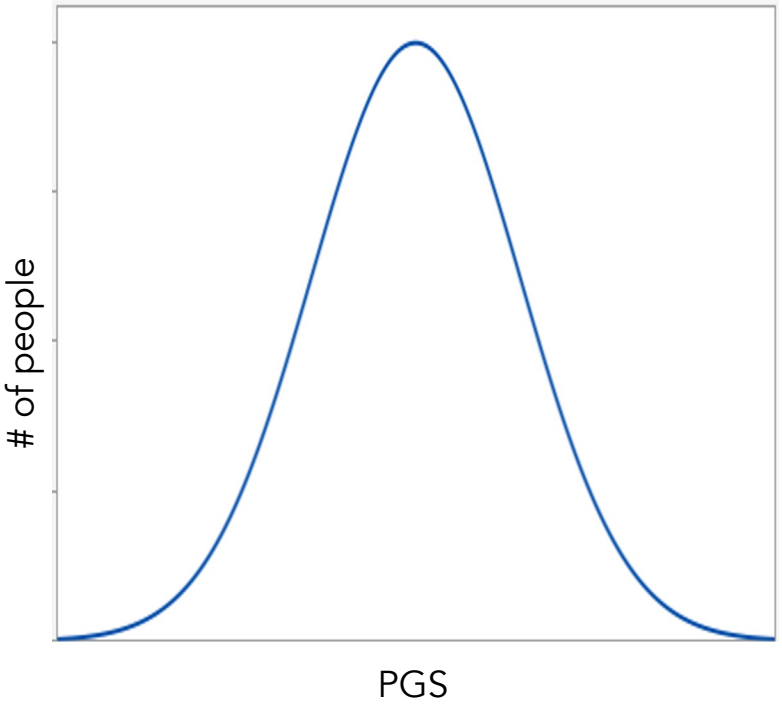
2018 — **12 additional polygenic scores for traits** including wake-up time, mosquito attraction, and dandruff

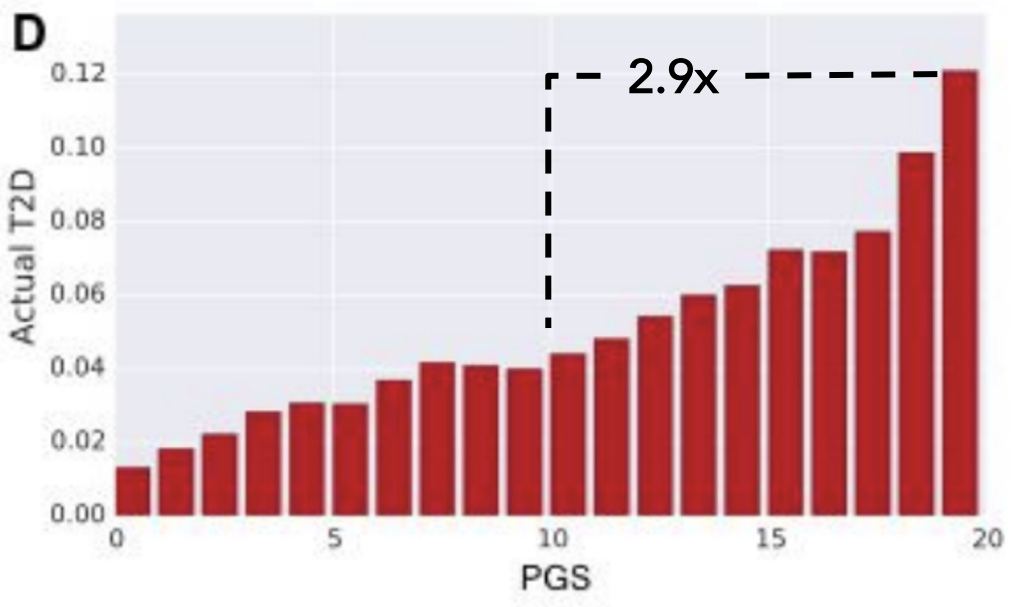
2019 — **Type 2 Diabetes polygenic score** powered by data from more than **2.5 million research participants**

All derived from 23andMe's research platform

Our polygenic scores now routinely use data from over 1M+ consented individuals

Polygenic scores for risk prediction





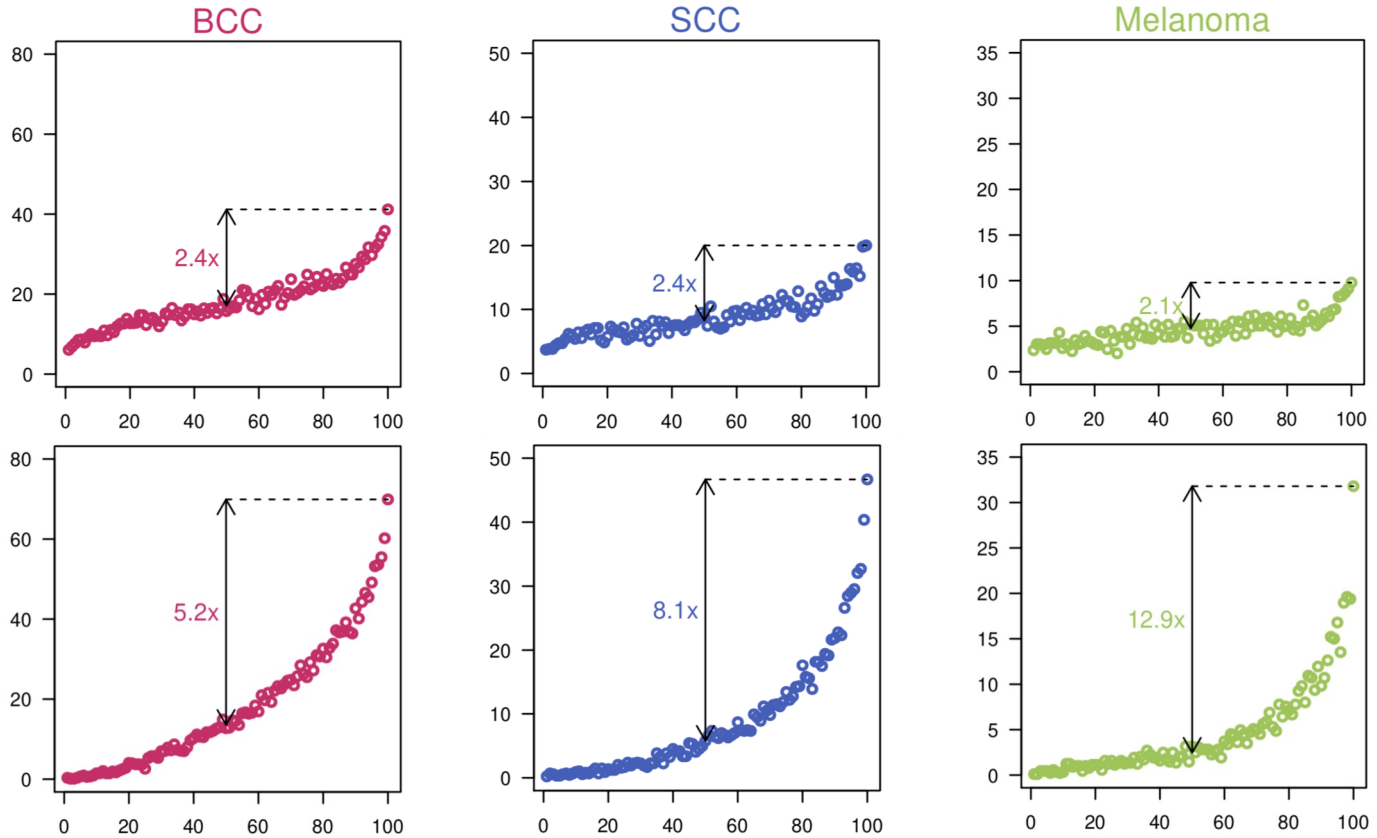
N=176,630

https://research.23andme.com/wp-content/uploads/2019/03/23_19-Type2Diabetes_March2019.pdf

Including exposures improves stratification

Genetics only

Genetics + exposure



N=88,985

Manuscript in progress, preliminary data presented at American Society of Human Genetics 2018 meeting

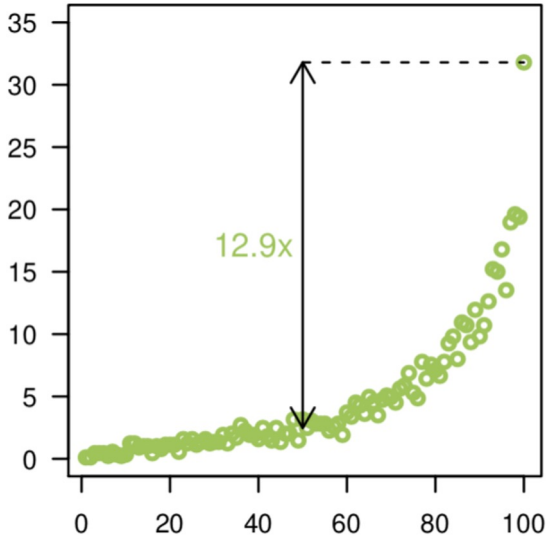
Risk stratification can identify interesting sub-cohorts

Why do some individuals > 60 yo in the top risk band with high exposure not have skin cancer?

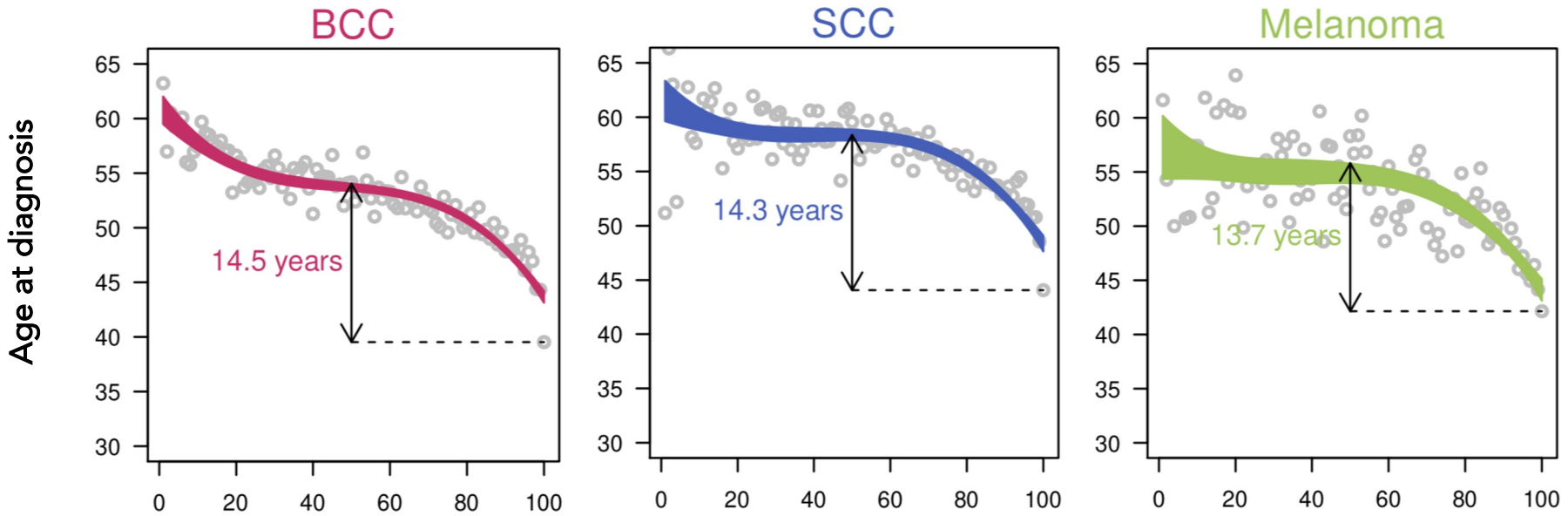
Why did some individuals in the lowest risk band with minimal exposure develop skin cancer in their 30s?

- Investigating either extreme can reveal novel biology
- Finding protective effects especially attractive for drug development

Melanoma prevalence by risk score percentile

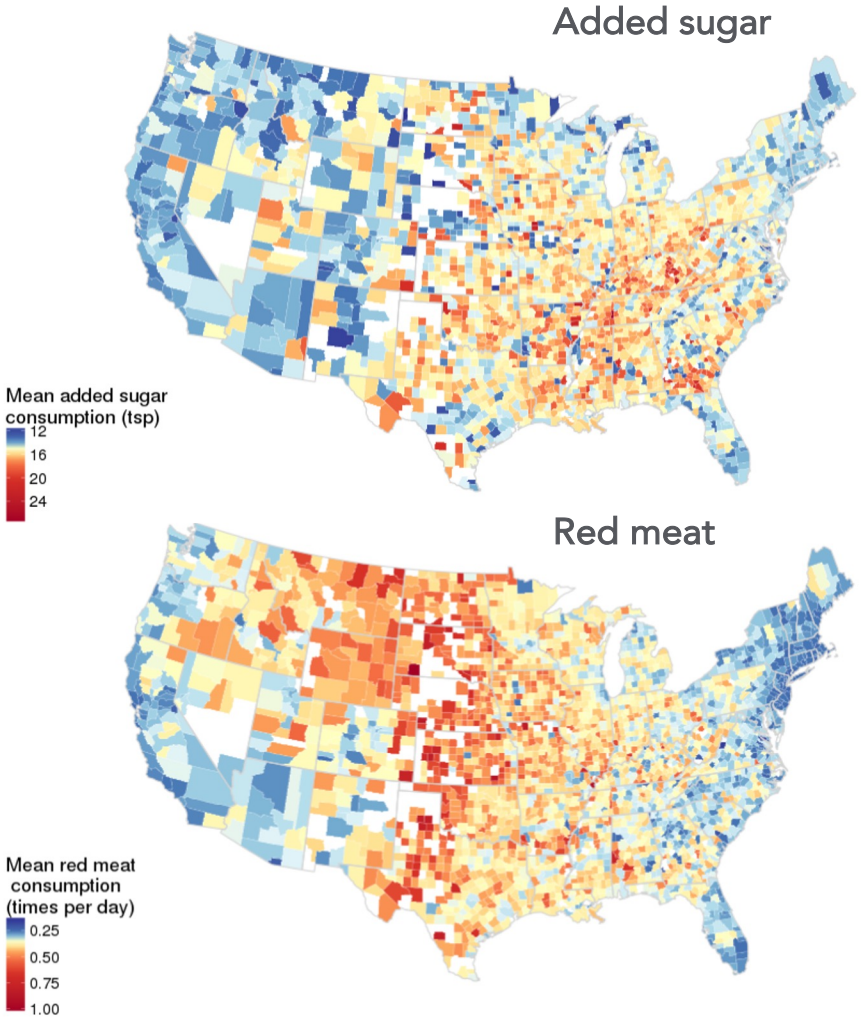


Risk stratification identifies interesting sub-cohorts



Identifying fast or early progressors, or individuals with extreme risk profiles, can enable targeted recruitment and more efficient clinical trials.

Manuscript in progress, preliminary data presented at American Society of Human Genetics 2018 meeting



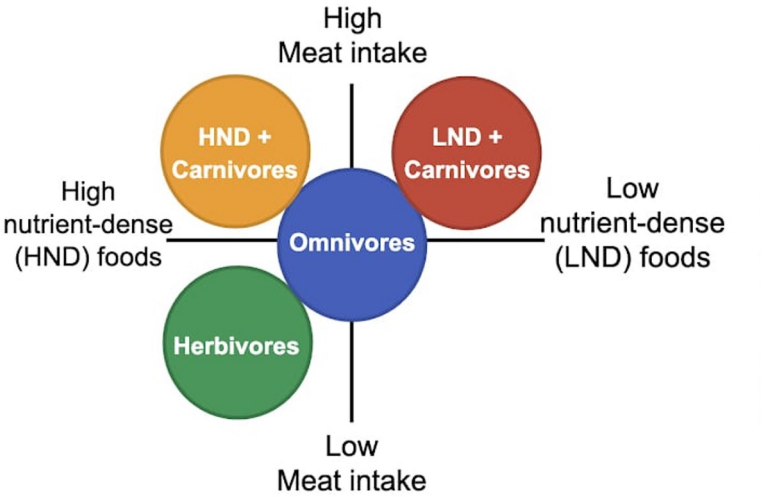
Detailed epidemiological data on behavioral phenotypes and health outcomes

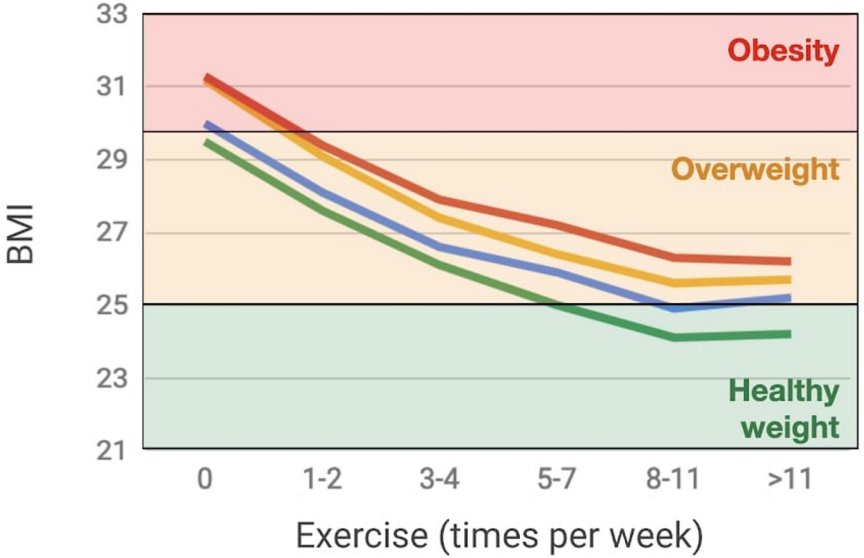
Daily dietary habits in 23andMe customers by US county

Derived from a 23andMe administered NHANES-style survey,

N ~ 700,000, only counties with >5 respondents shown

23andMe's data enable deeper analyses and insights into patterns of human behavior and health





Four diet types projected in the two main behavioral axes defining those types. Behaviors assessed via food frequency questionnaire in ~ 850,000 23andMe research participants.

https://blog.23andme.com/23andme-research/23andme-explores-dietary-habit-and-health-outcomes/

Future directions

Drug discovery:

- Leverage environment and other non-genetic data to further understand biology and drug mechanism of action and potentially develop drugs that are more effective for people with certain exposures
- Apply risk prediction to define precise cohorts that make clinical trials faster and more successful

Consumer products:

- Improve prediction through more sophisticated integration of non-genetic factors
- Provide actionable information about exposures to avoid, as long as we can reasonably establish causality
- Deliver deeper insights into relatable behaviors and their relationship to outcomes to drive individual change

Acknowledgments

23andMe research participants

23andMe Therapeutics team, including Computational Biology

23andMe Research team, including Data Collection, Statistical Genetics, and Research Ethics teams

23andMe Product, Product Science, and R&D teams


23andMe®