



Research for the Consumer, Powered by the Consumer

Mike Polcari, Joyce Tung



Who is 23andMe?

- **Direct to consumer genetic testing company**
- Founded in 2006 by Linda Avey and Anne Wojcicki
- Customers receive uninterpreted raw genotype data and genetic results on their ancestry
- Granted authorization by FDA to market a direct-to-consumer genetic test (Bloom syndrome)
- 1M+ genotyped customers





Our Mission:
To help people
access, understand
and benefit from the
human genome.

How can people benefit?

What can we gain from research into human genetics?



- Help find the **right treatments for the right people**
- Use genetics to personalize treatment
- Understand risk to maximize **prevention**



- Help develop **better drugs** using genomic data from **everyday people**

The background is a solid light green color. On the left side, there is a cluster of overlapping, semi-transparent, rounded geometric shapes in various colors including blue, green, orange, yellow, purple, and pink. The shapes are layered, creating a sense of depth and movement. The text 'The Research Platform' is positioned on the right side of the image, centered vertically relative to the cluster of shapes.

The Research Platform

How do people get started?



1 Order a kit online

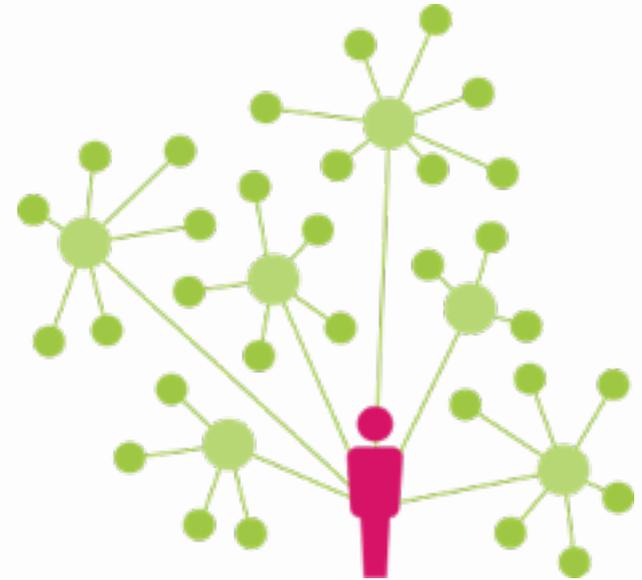
2 Provide your DNA Sample

3 Register your sample

4 Mail your kit back

Consenting to participate

- Research protocol and consent document are approved by an **external ethics review board** (IRB)
- People make an **informed decision** to contribute their data to research
- Their genotype and phenotype data will be used in a **broad range of studies**
- **80%** of customers consent to participate in research



Advantages of online research



- **Participation is easy**

- People can get involved at any time of day, from their couch



- **Geography is not a barrier**

- Proximity to a clinical study site is not a requirement

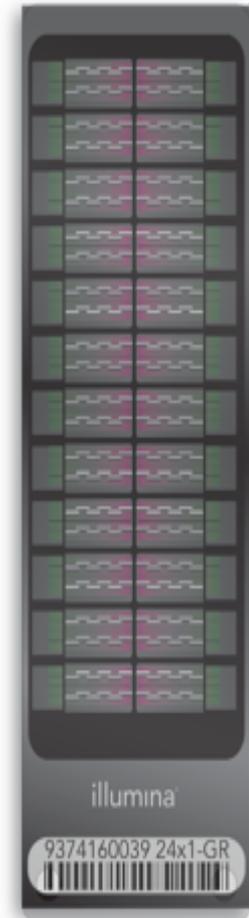


- **Everyone can be in multiple studies at once**

- The average 23andMe participant is genotyped only once, but is part of 200+ studies

Genotype Data

- Currently based on a **fully customized Illumina HumanOmniExpress+ Beadchip**, with SNPs selected to provide good genome-wide coverage for imputation and coverage of variants of known or suspected medical relevance
- Currently imputing with 1000G phase 1, looking into 1000G phase 3 and UK10K
 - Phase with Beagle, impute with Minimac
- IBD calculated by HAPLOSCORE
- Ancestry estimated using our Ancestry Composition algorithm



Phenotype Data

- Specially trained survey methodologists develop **online surveys** to collect high quality phenotypic data for research
- Topics range from **eye color** to **caffeine consumption** to **Parkinson's disease**
- Customers can **choose** to take as many surveys as they want whenever they want
- Now **thousands** of questions asked, organized into over a thousand “phenotypes” for analysis
- Over **250 million data points** collected



Online Research Surveys

23andMe HOME MY RESULTS FAMILY & FRIENDS RESEARCH & COMMUNITY Jaysa Tong

23andMe HOME MY RESULTS FAMILY & FRIENDS RESEARCH & COMMUNITY Jaysa Tong

SMOKING AND TOBACCO 1 Smoking and Tobacco Use < PREVIOUS PAGE NEXT PAGE >

Have you ever smoked tobacco products other than cigarettes, including cigars, pipes, water pipes or hookahs, bidis, or cigarillos?

Yes No I'm not sure

Have you smoked at least 100 cigarettes in your entire life?

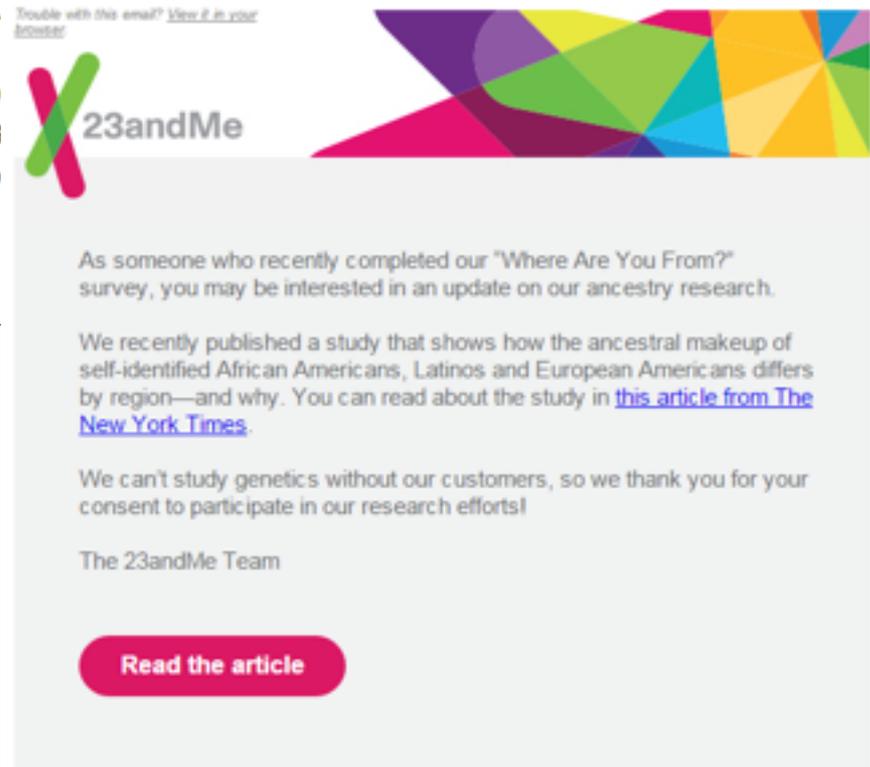
Yes No I'm not sure

Quality of Self-Reported Data

- Adhere to **best practices** in survey methodology
- **Replicated** genetic associations at hundreds of loci for 50+ conditions
- **Large sample sizes** help mitigate challenges of misclassification, though data quality will always be an important consideration
- Data are structured for research and collected all in the same way



A relationship with our participants

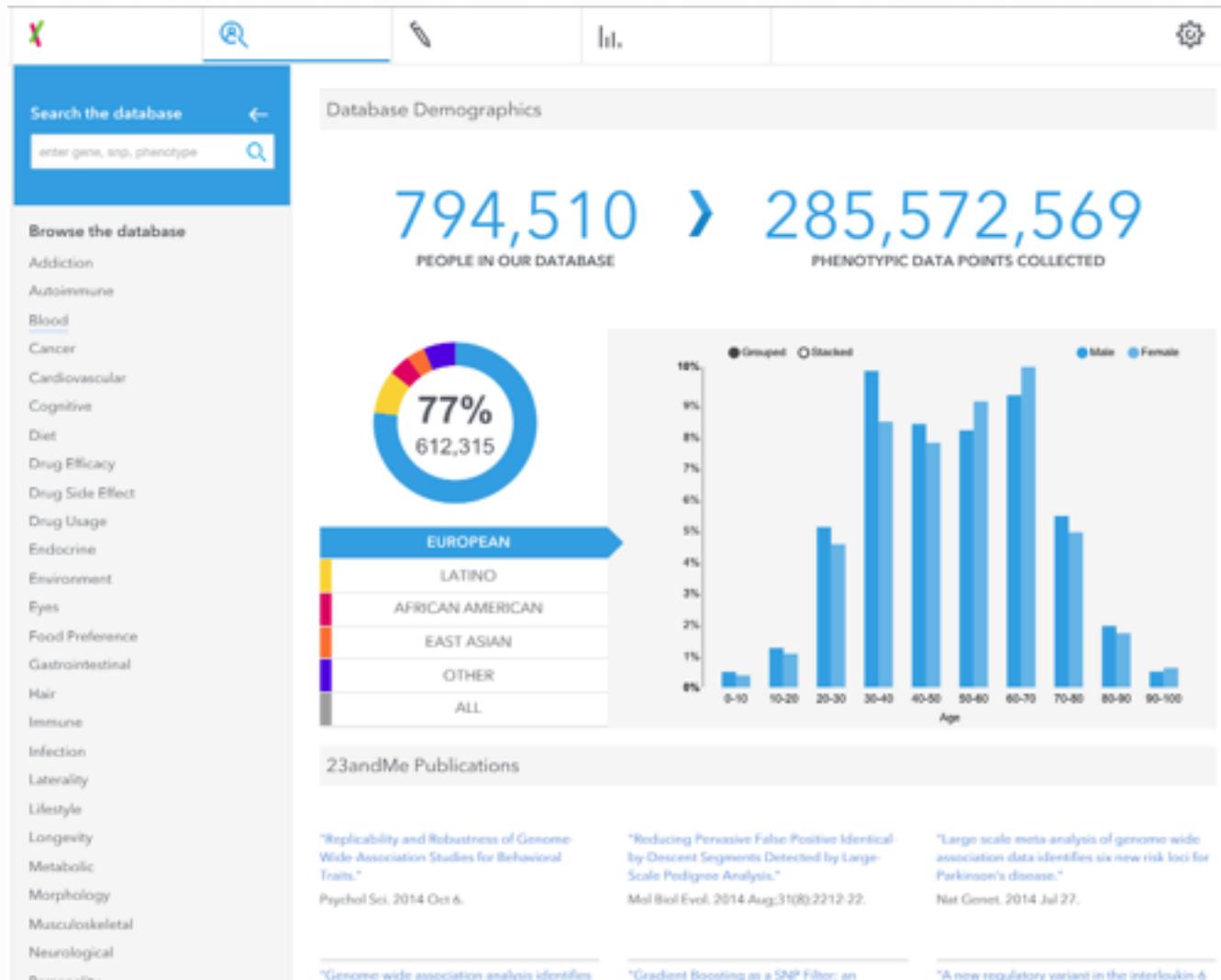


API

- Customers **may** use OAuth2 to authorize 3rd parties
- 3rd parties include: EMR's, PHR's, Research Studies, Apps...
- Flexible, **secure** REST API includes access most account data: raw genetics, higher-level interpretations, computed ancestry, relationships between users, PII, phenotypes, etc.
- 3rd parties request only relevant subset of data
- api.23andme.com



Research Portal - Overview



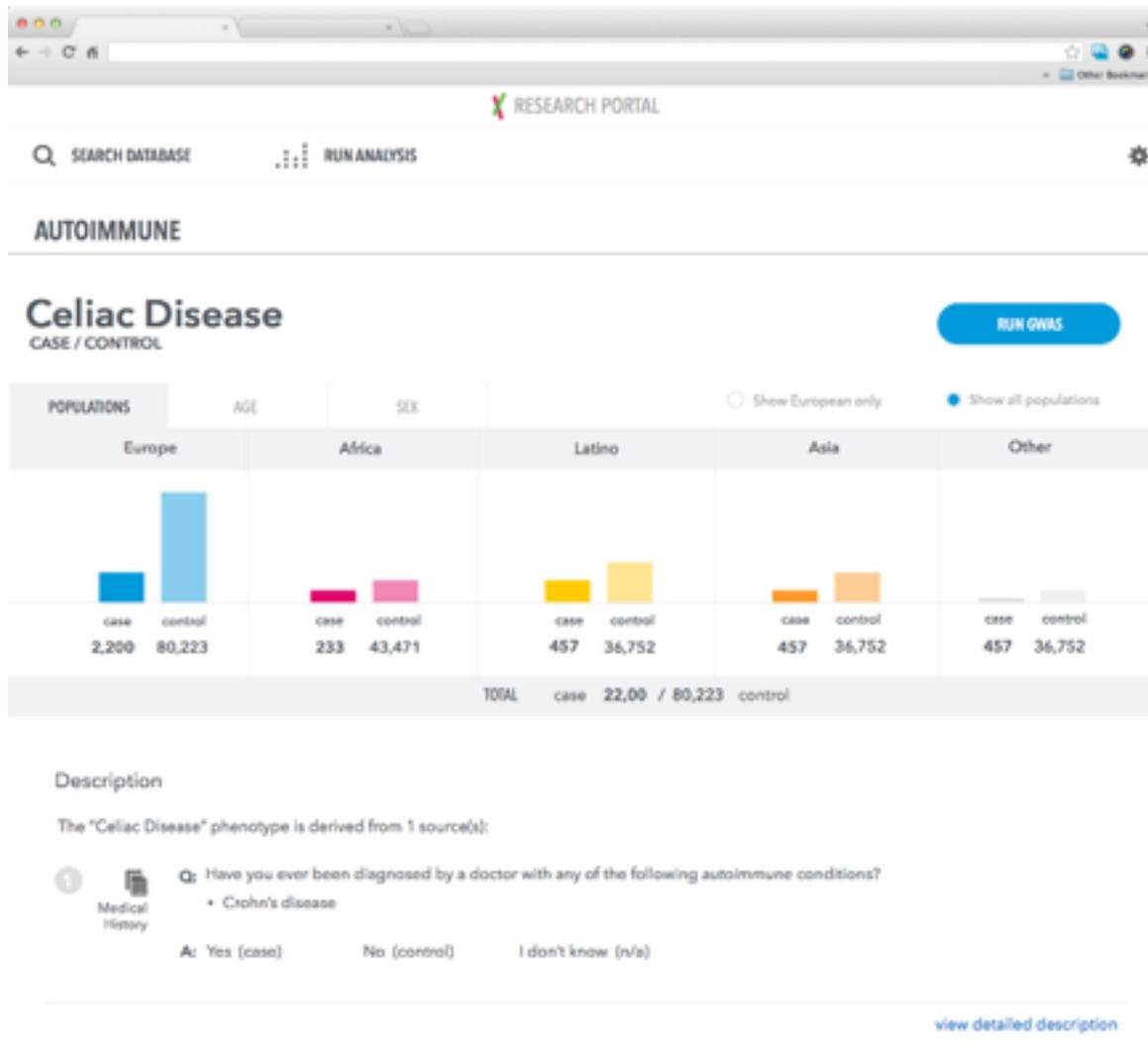
Research Portal - SNP

GENOTYPED SNP STATISTICS

ALL V4 V3 V1/V2

POP	DIPLOID GENOTYPES			HAPLOID GENOTYPES		ALLELE FREQUENCY	HW PVALUE	CALL RATE
<input checked="" type="radio"/> Show diploid frequency <input type="radio"/> Show diploid counts								
all	AA	.4123		A	0.6490		0.00e+00	0.9996
	AG	.4781		G	0.3510			
	GG	.1791						
europe	AA	.4586		A	0.6490		1.52e-01	0.3222
	AG	.3998		G	0.3510			
	GG	.1997						
africa	AA	.0587		A	0.8918		9.25e-03	0.9996
	AG	.6901		G	0.1082			
	GG	.3214						
asia	AA	.8713		A	0.4057		1.59e-01	0.9997
	AG	.0971		G	0.5943			
	GG	.0087						
latino	AA	.4556		A	0.4057		2.43e-03	0.9996
	AG	.3241		G	0.5943			
	GG	.0843						

Research Portal - Phenotype



Publications and Collaborations

30+

Total publications

40+

Academic collaborations

www.23andme.com/for/scientists

NIH Grants

- SBIR Catalyst (NHGRI): Web-based Phenotyping for Genome Wide Association Studies of Drug Response (2010) 1R43HG005807-01
- SBIR Phase I (NHGRI): Development of DNA Sequence Data-Quality Metrics for Personal Genomics (2012) 1R43HG006976-01
- SBIR Phase I (NHLBI): Genetics of Allergic Disease in a Participatory Research Cohort (2012) 1R43HL115873-01
- SBIR Phase I (NHGRI): Development of a web-based database and research engine for genetic discovery (2012) 1R44HG006981-01
 - Phase II: 2013 @R44HG006982-02

