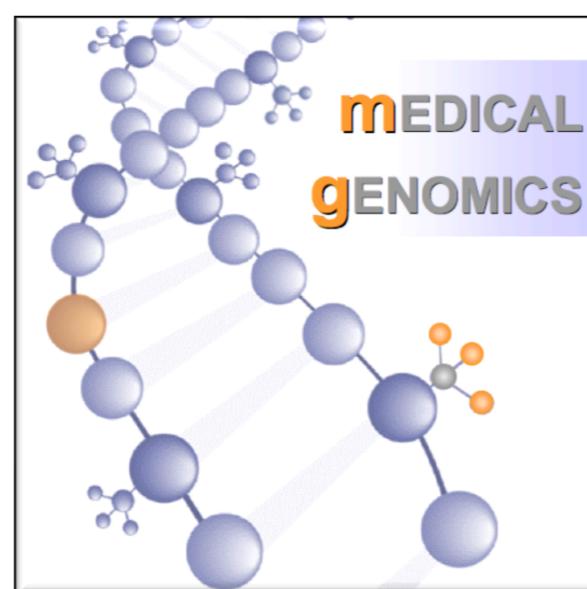




Personal Genome Project



Stephan Beck
Medical Genomics
UCL Cancer Institute
University College London
s.beck@ucl.ac.uk



opportunity

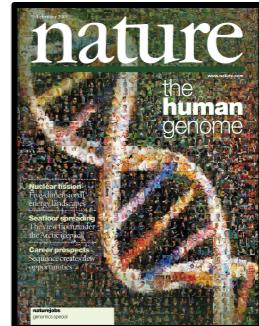


Consensus among researchers, clinicians, politicians and the public that ‘omics’ will transform biomedical research, healthcare and lifestyle choices

brief history



- **2001: First Human Genome Sequence**



open
access

- **2005: Personal Genome Project (PGP-US)**



open
access

- **2008: 1000 Genomes Project** (<http://www.1000genomes.org>)



managed
access

- **2010: UK10K Project** (www.uk10k.org)



managed
access

- **2013: UK100K Project** (www.genomicsengland.co.uk)



managed
access

- **2013: PGP-UK** (www.personalgenomes.org.uk)



open
access

- **Many other national and international projects**



Personal Genome Project UK

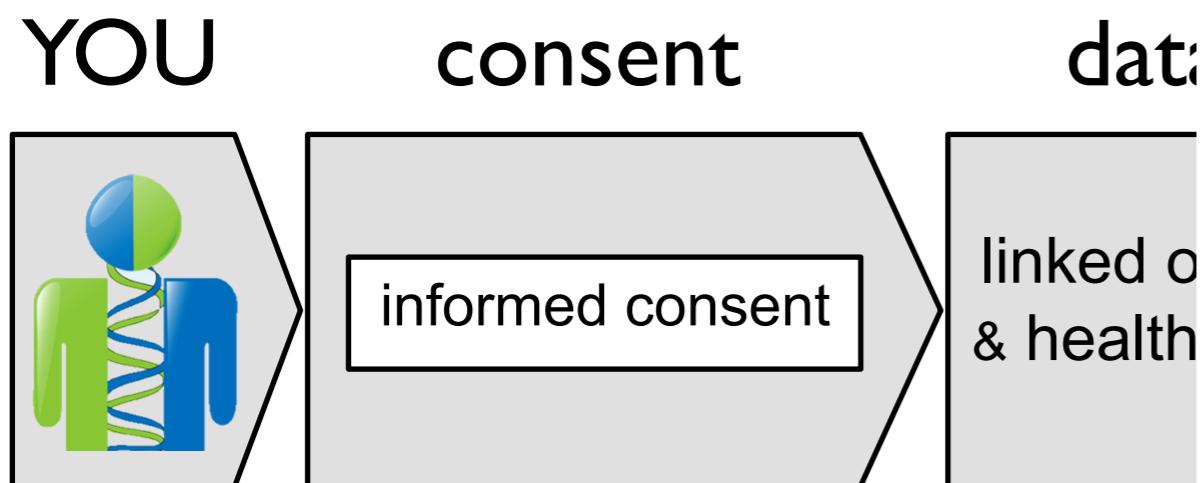
The Personal Genome Project UK is dedicated to creating public UK genome, health, and trait data. Sharing data is critical to scientific progress, but has been hampered by traditional research practices—our approach is to invite willing participants to publicly share their personal data for the greater good.



- Launched on Nov 6, 2013
- Open to all UK citizens and permanent residents over 21 years
- Uses Open Consent and Open Access policies
- Classified as research project
- Over 10,000 UK volunteers registered and first 1000 enrolled



Most Projects (including 1K, 10K and 100K Genome Projects)



- Promise of anonymity
- No or limited feedback to participants
- No or restricted public access

WORLD VIEW

A personal take on events



Be prepared for the big genome leak

It is only a matter of time until idealism sees the release of confidential genetic data on study participants, says Steven E. Brenner.

Most people in the United States could soon know someone whose genome is held in a research database. Concerns are growing about our ability to properly control access to that information. Also growing among some scientists is the feeling that restricting access to genomic data hinders research. How long will it be until an idealistic and technically literate researcher deliberately releases genome and trait information publicly in the name of open science?

Both the open-access literature and the open-source software movements began with idealists. It seems inevitable that there will be a major leak of genome information in the near future. Institutions and funders should consider react when this happens.

Some studies already gather the genetic data of individuals in a single analysis. Although this is posed to be highly protected, it is disseminated to various institutions that have inconsistent security and privacy standards. In practice, data protection often comes down to individual scientists. Once leaked, these data would be virtually impossible to contain.

What harm would come from a leak of personal and genomic data? The consent form for the Personal Genome Project (PGP) — which makes no attempt to keep genetic information secret — offers a guide. It lists a range of adverse consequences, from revealing non-paternity to being framed with synthesized DNA planted at a crime scene.

Most research genome data are de-identified, but given progress in re-identification and commercial genetic databases, will they stay that way? De-anonymized genomic data would be most likely to reveal health conditions relevant to the study for which they were collected. The effects might be uncomfortable but would probably reveal less than a typical Google search history. So far, no PGP participant who released genomes and traits has experienced adverse consequences that have been reported to the Institutional Review Board. In the longer term, the risk of harm may rise as our understanding of genetic variation increases.

Then there is the public outcry a genome breach might incite. The public often has an exaggerated perception of the links between genes and personal traits. Lacking contextual information, research participants could wonder whether their own genomes had been leaked and dread implausibly dire consequences.

Thus a genome leak might lead to a backlash. Volunteers might withdraw from research studies and refuse to join new ones. Research might even be subject to moratoriums and prohibitive restrictions. The harm to genetic research could be great, and study participants could be unsettled.

EJHG Open

European Journal of Human Genetics (2014), 1–5
© 2014 Macmillan Publishers Limited. All rights reserved 1018-4813/14
www.nature.com/ejhg

ARTICLE

Managing clinically significant findings in research: the UK10K example

Jane Kaye^{*1}, Matthew Hurles², Heather Griffin¹, Jasote Grewal³, Martin Bobrow⁴, Nic Timpson⁵, Carol Smee², Patrick Bolton⁶, Richard Durbin², Stephanie Dyke², David Fitzpatrick⁷, Karen Kennedy², Alastair Kent⁸, Dawn Muddyman², Francesco Muntoni⁹, Lucy F Raymond⁴, Robert Semple⁴, Tim Spector¹⁰ and UK10K¹¹

the possible harm of such leaks to better inform and protect research participants before and after leaks occur.

We should also take steps to minimize the frequency and extent of future genome leaks. Institutions could establish uniform protocols and reviews to ensure the safety of protected genomic data. All researchers using restricted genomic data should be trained regarding the ethics of and the technologies involved in protecting human data. Technical and legal strategies should be proactively deployed to help limit dissemination of leaked data to those who furtively hunt for them.

Augmented legal protections could reduce the harm from inappropriate use of such data. In the meantime, we need to address a quandary: research with leaked data would undoubtedly speed immediate scientific progress, but should scientists exploit them?

Most importantly, we must ensure that the necessary discussion about the risks of a genome leak is balanced with information about the tremendous benefits that collected genetic information has for all of us. Although the acceleration and promise of genomics makes a leak inevitable, it also guarantees medical progress. ■ **SEE EDITORIAL P.107**

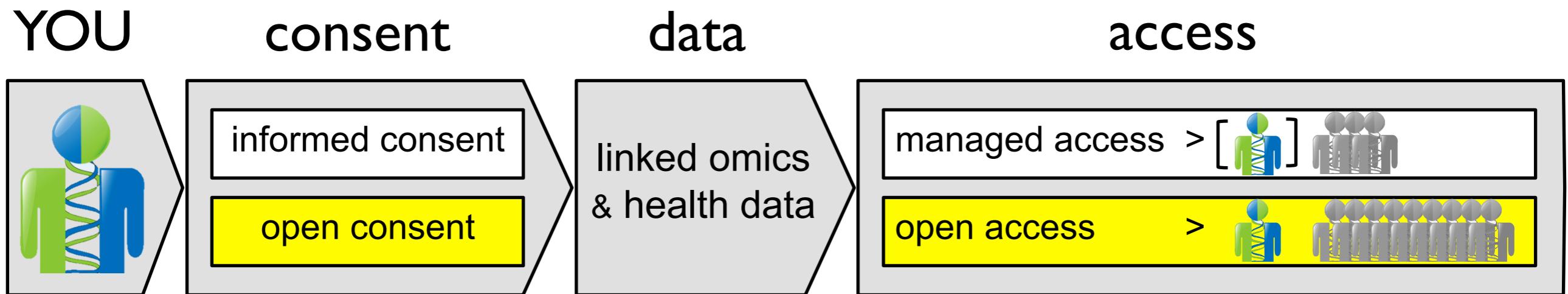
Steven E. Brenner is a Professor at the University of California, Berkeley.
e-mail: brenner@compbio.berkeley.edu

2014)

consent / access options



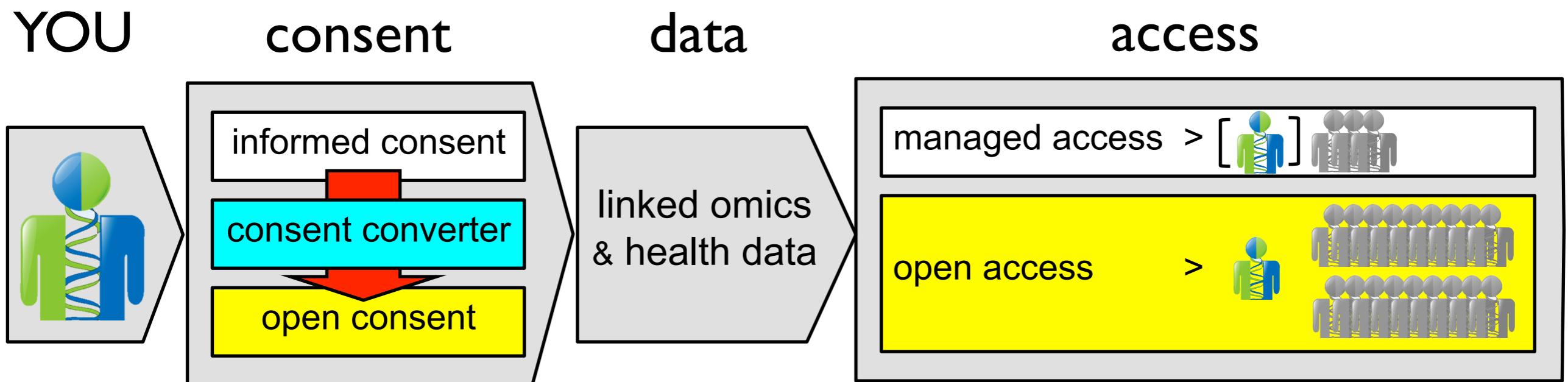
PGP-UK



- No promise of anonymity
- All data returned to YOU
- Free and unrestricted public access to data



PGP-UK (planned)



- Converts informed to open consent
- Converts managed to open data access
- Facilitates genome donations e.g. to PGP-UK

The European Genome-phenome Archive (EGA) allows you to explore **datasets** from genomic **studies**, provided by a range of **data providers**. Access to datasets must be approved by the specified **Data Access Committee (DAC)**.

managed access ~150 vs ~75,000,000 open access
[per month] [per month]

500,000 fold difference

majority of funding

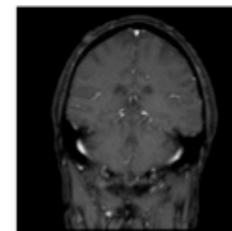
accelerated research



PGP enables *citizen science*



Genome In a Bottle Consortium
National Institute of Standards & Technology (NIST)



Longitudinal Body Microbiome
University of Colorado



Mapping Flu Incidence
Childrens Hospital Boston



Critical Assessment of
Genome Interpretation
Univ California Berkeley



Environmental Exposures
Silent Spring Institute



Home Microbiome
NC State University



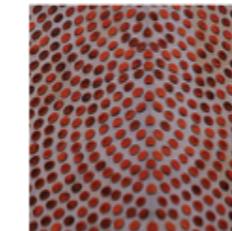
active NOT passive
collaborators



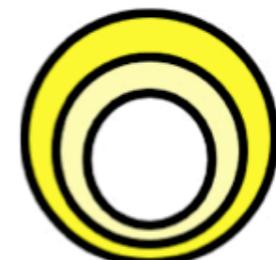
Olfactory Response
Rockefeller University



Absolute Pitch
Feinstein Institute



TestMyBrain.org
Harvard

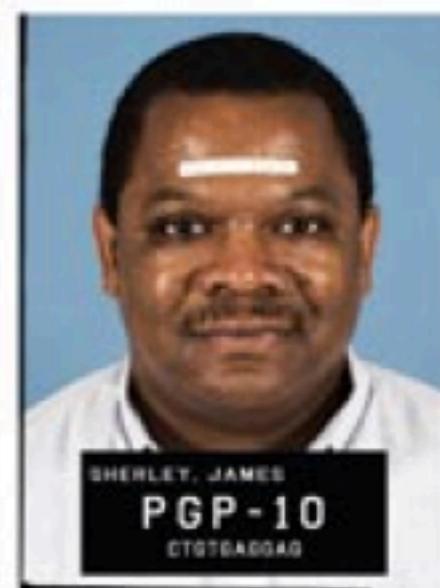
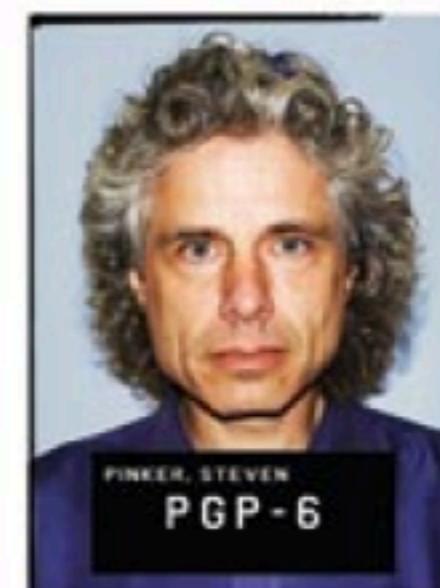
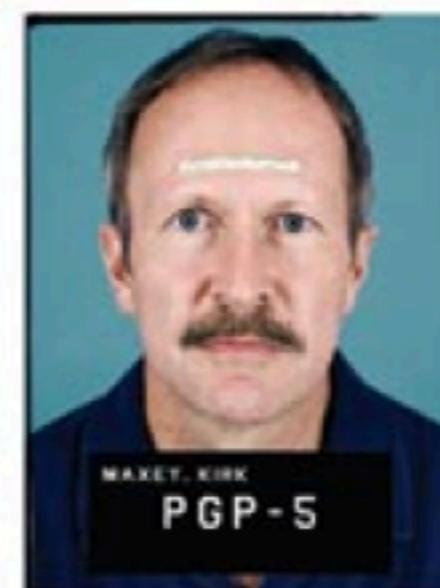
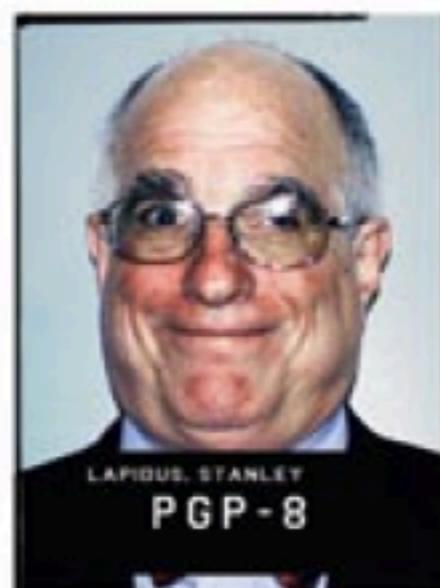
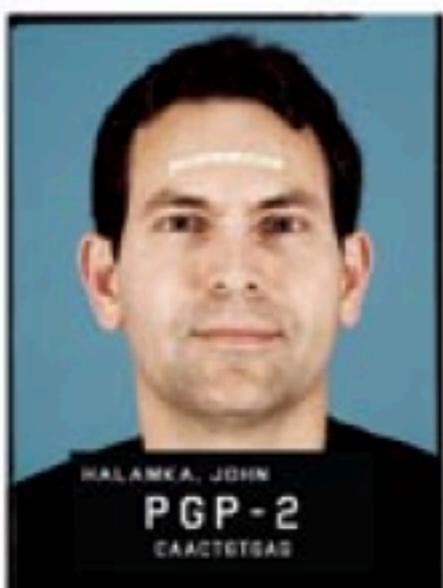
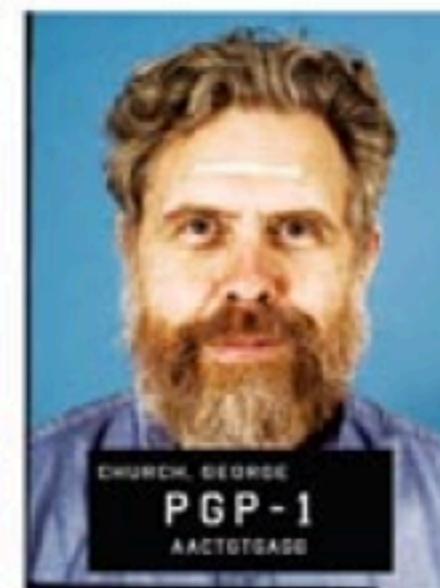
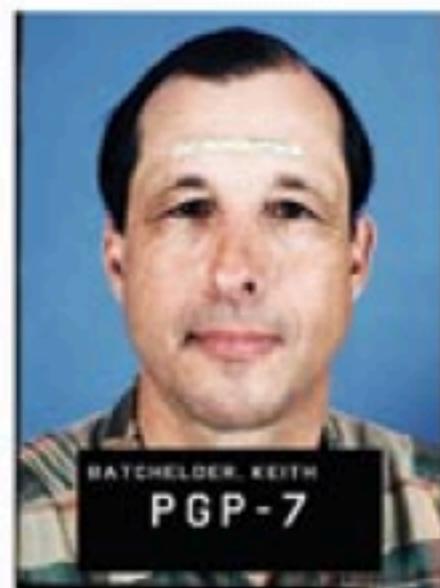
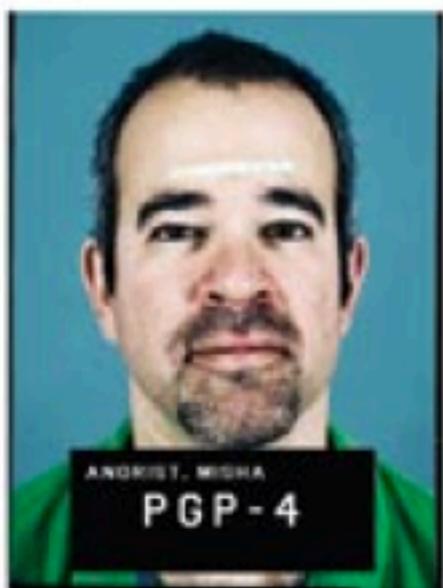


Circles in Evolution:Areola
Tabin Lab at Harvard



Activity Levels and Obesity
Center for Connected Health

PGP-10

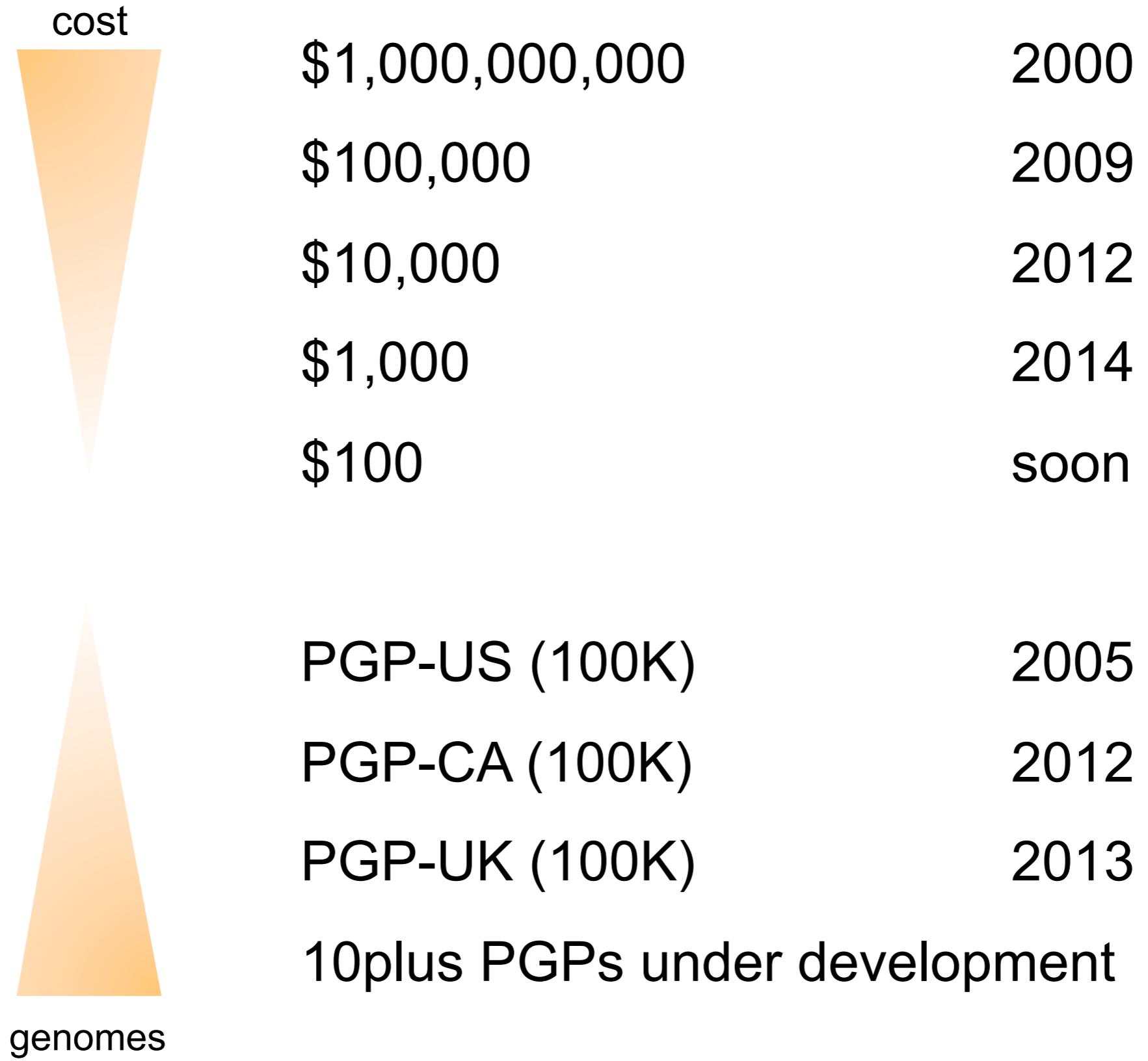


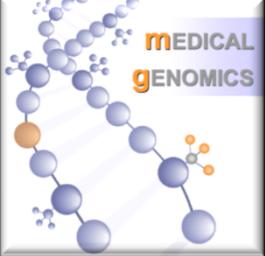
The first 10 to donate their genomes to the Personal Genome Project

Most analysed humans (<http://bit.ly/pgplinks>)



genomes for all





PGP stats



Ball et al. *Genome Medicine* 2014, **6**:10
<http://genomemedicine.com/content/6/2/10>



Genome **Medicine**

OPEN DEBATE

Open Access

Harvard Personal Genome Project: lessons from participatory public research

Madeleine P Ball^{1*}, Jason R Bobe², Michael F Chou¹, Tom Clegg³, Preston W Estep⁴, Jeantine E Lunshof^{1,5}, Ward Vandewege³, Alexander Wait Zaranek^{1,3} and George M Church¹

Abstract

Background: Since its initiation in 2005, the Harvard Personal Genome Project has enrolled thousands of volunteers interested in publicly sharing their genome, health and trait data. Because these data are highly identifiable, we use an 'open consent' framework that purposefully excludes promises about privacy and requires participants to demonstrate comprehension prior to enrollment.

Discussion: Our model of non-anonymous, public genomes has led us to a highly participatory model of researcher-participant communication and interaction. The participants, who are highly committed volunteers, self-pursue and donate research-relevant datasets, and are actively engaged in conversations with both our staff and other Personal Genome Project participants. We have quantitatively assessed these communications and donations, and report our experiences with returning research-grade whole genome data to participants. We also observe some of the community growth and discussion that has occurred related to our project.

Summary: We find that public non-anonymous data is valuable and leads to a participatory research model, which we encourage others to consider. The implementation of this model is greatly facilitated by web-based tools and methods and participant education. Project results are long-term proactive participant involvement and the growth of a community that benefits both researchers and participants.

PGP stats



Figure 2 Participant-initiated communications. Our website offers participants a 'Contact Us' button. From June 2012 to December 2013, we received 579 emails from participants. Few support requests derived from the return of genome data (3.6% of all emails, representing 11% of participants receiving genome data). Most of these were inquiries regarding data formats and additional files for the participants' own analyses of their data (2.9%), rather than requests for additional interpretation on our part (0.7%).



PGP stats

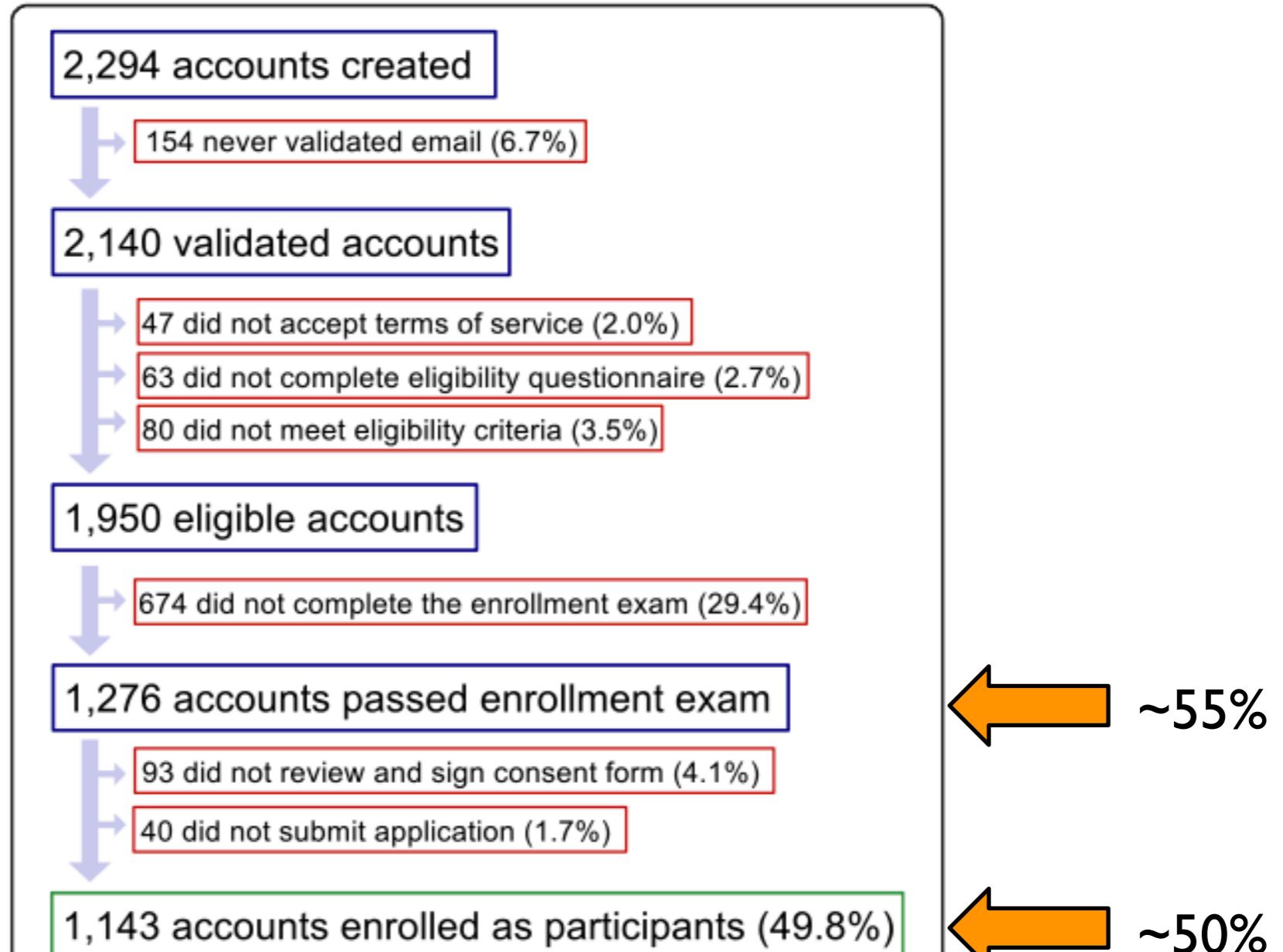


Figure 1 Status of 2,294 accounts created June 2012 to December 2013. About half of all accounts created on our site eventually complete the enrollment process to become participants (1,143 users, 50% of all accounts). Of the 1,151 accounts that did not complete the enrollment process, the majority (674 users, or 59% of incomplete enrollments) stopped at the enrollment examination stage.

PGP-UK summary



➤ Open Consent

- No promise of privacy or confidentiality
- First in UK and Europe

➤ Open Access

- CC0 public domain license
- Integrated genome, environmental, medical and personal data

➤ Participants

- Highly informed > study guide, entrance exam, regular feedback
- Receive all information, incl. CSFs
- Active (not passive) collaborator > *Citizen Science*
- Most analysed humans (<http://bit.ly/pgplinks>)

acknowledgments



PGP-UK Team

pgp-uk@ucl.ac.uk



Stephan Beck
Director



Jane Kaye
ELSI



Rifat Hamoudi
Software



Javier Herrero
Analysis



Emanuele Libertini
Analysis

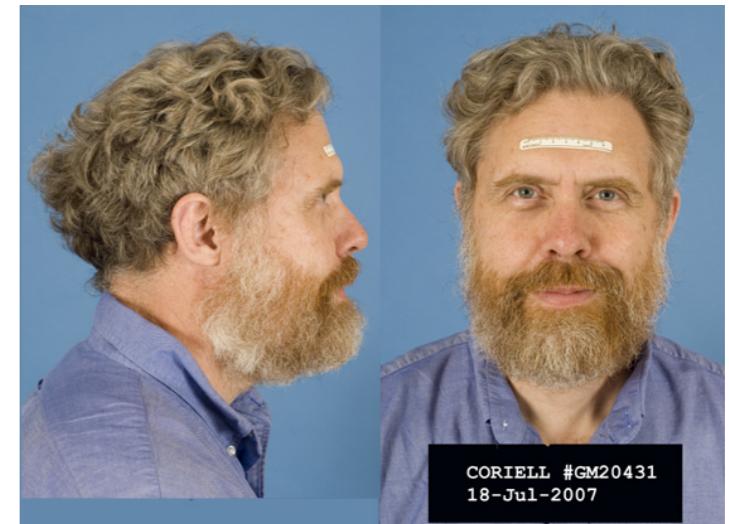


Paul Harrison
Outreach



Erica Jones
Admin

special thanks to



*"The time has come for Homo sapiens
to become our key model organism"*



Jason Bobe
Director, PersonalGenomes.org