

# Sharing Genomic (and Other) Data: The Perspective of the Kaiser Permanente Research Program on Genes, Environment and Health

Catherine Schaefer, PhD

Executive Director, RPGEH  
Research Scientist / Epidemiologist  
KPNC Division of Research  
Oakland, CA



Secretary's  
Advisory  
Committee on  
Genetics, Health  
and Society

February 4, 2010

Washington, DC

Division of Research

## Overview

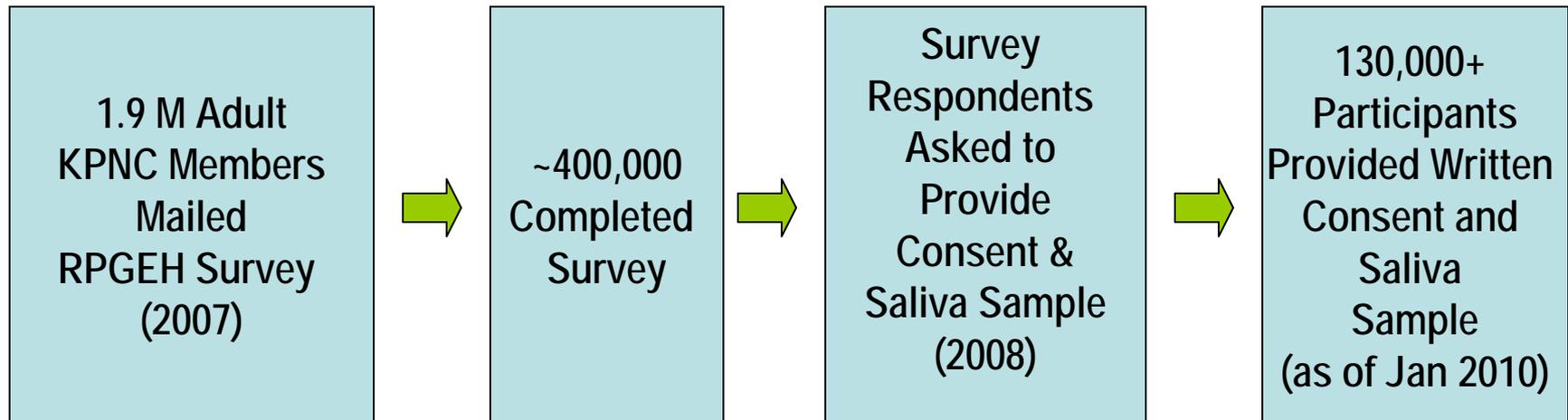
- Kaiser Permanente Northern California (KPNC) is developing a comprehensive resource for research on genetic and environmental influences on health
- The resource will link data on 500,000 members of KPNC, including
  - Comprehensive, continuously updated, clinical data from electronic medical records (EMR)
  - Participant surveys
  - Environmental exposures based in GIS database
  - Genetic, biomarker, and environmental data from biospecimens

- The aims are to enable scientists to conduct research on genetic and environmental influences on
  - **Disease susceptibility**
  - **Disease course, prognosis, and outcomes**
  - **Response to treatment (pharmacogenetics)**
- Conduct research to translate findings into improvements in medical care and public health
- Conduct research on the ethical, legal and social implications of genetic research and the use of genomic information in medical care

## A Brief History of the RPGEH

- Initial funding in 2005-2006
  - Engaged KPNC membership through focus groups, internal communications, and media
  - Organized Community, Scientific, and Bioethics Advisory Panels
  - Organized EMR data by disease groups to facilitate research
    - 10 registries covering over 100 diseases and conditions
  
- Development of the RPGEH has been funded by Kaiser Permanente and grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, and The Ellison Medical Foundation.

## Development of the RPGEH Cohort (survey & samples)



*RPGEH survey includes questions regarding demographics, health history, family history, smoking, alcohol, diet, physical activity, and reproductive history and health (available in English, Spanish, and Chinese).*

## Key Current Activities

- Continuing enrollment and sample collection
  - 200 K by year-end 2010; 500 K by year-end 2013
- Collection of blood samples using KPNC clinical infrastructure
- Continuing work on funded GWA studies: multi-ethnic study of bipolar disorder; study of prostate cancer among African Americans
- Development of a collaborations portal and access review committee, receiving applications in 2010

## Key Current Activities

- Genome-wide genotyping of 100,000 participants by year-end 2011, funded by GO grant from NIH
  - Designed as a resource for study of age-related diseases, healthy aging, and longevity; average age is 65;
  - 650 K SNPs will be genotyped
  - Linked to data from EMR, survey, and environmental databases
  - Accessible through dbGaP and collaborations with RPGEH
  - Will require re-consent for deposit of data in dbGaP

## Considerations for Data Sharing

- **RPGEH is a rich resource**
  - Large, diverse and generally representative of population
  - Comprehensive, continuously updated EMR enables excellent phenotypic characterization and follow-up
- **KPNC sees that RPGEH can make an important contribution and wants to ensure that the best and broadest use is made, consistent with commitment to members**
- **KPNC / RPGEH perspective on data sharing is shaped by commitment to members; we are invested in them and they determine the future of KPNC**
  - Over 50% of first 100K participants in RPGEH have been members for more than 20 years
  - Trust in KPNC by our members enables us to do research

## Factors Affecting Data Sharing

- **Written, informed consent is broad**
  - No restrictions on kinds of health problems that can be studied
  - Health information can be updated from EMR
  - All studies must be IRB-approved
  - Data can be shared with scientists outside KPNC who agree to protect confidentiality and follow rules for use
  
- **Use and share genomic data for research only**
  - Research results will not be placed in EMR; participation is confidential
  - Genomic data will not be returned to individuals or their providers
  - Participants may be contacted if information develops that has significance for health
  - Participant may withdraw and ask that sample be destroyed
    - How do we ensure latter commitments are met when data are used through public database

### Data Sharing and Social Harms

- **Concern has been expressed about data sharing through federal database such as dbGaP**
- **Community Advisory Panel, focus groups, and some survey respondents are concerned that the government may “take” and misuse data**
  - **Building of other federal DNA databases increases perceived vulnerability of NIH database to re-identification or misuse**
  - **Use of DNA to deny treaty rights or label immigrants is community concern**
- **Concern that data may be misused to stigmatize vulnerable group**
- **Perception that storage and control of data by KPNC / RPGEH gives participants better recourse and control**

## Sharing Phenotypic Data

- **In most research contexts, sharing genomic data means sharing phenotypic data, since interest is in linking genomic data with health**
  - Must also consider factors affecting sharing other types of data (e.g., re-identification and discrimination or stigma)
  - Health plans with EMRs have huge investments in these data
- **Quality of phenotypic data is critical to best use of genomic data and resource**
  - Challenging to abstract high density data to be useful to all
  - Best use of data depend on knowledge of system that generated the data -- important to involve scientists who know the data

## Stakeholder Interviews

- **Goal to inform the development of RPGEH access and collaboration policies and procedures**
- **Qualitative interviews and focus groups with multiple RPGEH stakeholders**
  - **Researchers**
  - **IRB/HRPPP members**
  - **CAP members**
  - **Participants**
  - **KPNC leadership**
  - **RPGEH funders**
- **Research questions**
  - What are the specific data sharing, benefit sharing and governance issues inherent in a biobank that is situated in an integrated delivery system?
  - How are these issues perceived in different ways by different biobank stakeholders?
  - What are the different stakeholders' values and preferences regarding protection of participants' privacy, data sharing, governance, and benefit sharing with respect to the RPGEH?

## Key Personnel

Catherine Schaefer, PhD, Director

Neil Risch, PhD, Co-Director

Carol Somkin, PhD, Lead of ELSI and Community Cores

Stephen Van Den Eeden, PhD, Environmental Exposures Core

Charles Quesenberry, PhD, Biostatistics Core

Lisa Croen, PhD, Biorepository Design Core

Larry Kushi, ScD, Access and Collaborations Core

Mary Henderson, MBA, MPH, Managing Director

Sarah Rowell, MPH, Research Operations

Larry Walter, MA, Informatics

Diane Olberg, Communications Manager