Design and Selection

Patricia Hartge

Workshop on Sequencing in Cohort Studies
and Large Sample Collections

June 28-29, 2012

Background

- GWAS
  - demanded large samples with DNA
- Lethal and uncommon diseases
  - Pancreatic cancer, glioma
- Uncertainties on vitamin D enrichment
  - Demanded large samples with serum, std. method
- Controversial effects of obesity
  - Demanded large samples of non-smokers
- Minority populations

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Performance of Common Genetic Variants in Breast-Cancer Risk Models

Sholom Wacholder, Ph.D., Patricia Hartge, Sc.D., Ross Prentice, Ph.D., Montserrat Garcia-Closas, M.D., Ph.D., Heather Spencer Feigelson, Ph.D., W. Ryan Diver, M.S.P.H., Michael J. Thun, M.D., David G. Cox, Ph.D., Susan E. Hankinson, Ph.D., Peter Kraft, Ph.D., Bernard Rosner, Ph.D., Christine D. Berg, M.D., Louise A. Brinton, Ph.D., Jolanta Lissowska, Ph.D., Mark E. Sherman, M.D., Rowan Chlebowski, M.D., Charles Kooperberg, Ph.D., Rebecca D. Jackson, M.D., Dennis W. Buckman, Ph.D., Peter Hui, B.S., Ruth Pfeiffer, Ph.D., Kevin B. Jacobs, B.S., Gilles D. Thomas, M.D., Robert N. Hoover, M.D., Sc.D., Mitchell H. Gail, M.D., Ph.D., Stephen J. Chanock, M.D., and David J. Hunter, M.B., B.S., Sc.D.

ABSTRACT

BACKGROUND

Genomewide association studies have identified multiple genetic variants associated with breast cancer. The extent to which these variants add to existing risk-assessment models is unknown.

A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33

Gloria M Petersen[*1,54], Laufey Amundadottir[2,3,54], Charles S Fuchs[4,5,54], Peter Kraft[6,7,54], Rachael Z Stolzenberg-Solomon[3,54], Kevin B Jacobs[3,8,9], Alan A Arslan[10–12], H Bas Bueno-de-Mesquita[13], Steven Gallinger[14], Myron Gross[15], Kathy Helzlsouer[16], Elizabeth A Holly[17], Eric J Jacobs[18], Alison P Klein[19,20], Andrea LaCroix[21], Donghui Li[22], Margaret T Mandelson[21,23], Sara H Olson[24], Harvey A Risch[25], Wei Zheng[26], Demetrius Albanes[3], William R Bamlet[1], Christine D Berg[27], Marie-Christine Boutron-Ruault[28], Julie E Buring[29,30], Paige M Bracci[17], Federico Canzian[31], Sandra Clipp[32], Michelle Cotterchio[33], Mariza de Andrade[1], Eric J Duell[34], J Michael Gaziano[35,36], Edward L Giovannucci[5,6,37], Michael Goggins[38], Göran Hallmans[39], Susan E Hankinson[5,6], Manal Hassan[22], Barbara Howard[40], David J Hunter[5,6], Amy Hutchinson[3,8], Mazda Jenab[34], Rudolf Kaaks[31], Charles Kooperberg[21], Vittorio Krogh[41], Robert C Kurtz[42], Shannon M Lynch[43], Robert R McWilliams[1], Julie B Mendelsohn[3], Dominique S Michaud[6,44], Hemang Parikh[2,3], Alpa V Patel[18], Petra H M Peeters[44,45], Aleksandar Rajkovic[46], Elio Riboli[44], Laudina Rodriguez[47], Daniela Seminara[43], Xiao-Ou Shu[26], Gilles Thomas[3,48], Anne Tjønneland[49], Geoffrey S Tobias[3], Dimitrios Trichopoulos[6,50], Stephen K Van Den Eeden[51], Jarmo Virtamo[52], Jean Wactawski-Wende[53], Zhaoming Wang[3,8], Brian M Wolpin[4,5], Herbert Yu[25], Kai Yu[3], Anne Zeleniuch-Jacquotte[11,12], Joseph F Fraumeni Jr[3], Robert N Hoover[3,54], Patricia Hartge[3,54] & Stephen J Chanock[2,3,54]

We conducted a genome-wide association study of pancreatic cancer in 3,851 affected individuals (cases) and 3,934 unaffected controls drawn from 12 prospective cohort studies and 8 case-control studies. Based on a logistic regression model for genotype trend effect that was adjusted for study, age, sex, self-described ancestry and five principal components, we identified eight SNPs that map to three loci on chromosomes

only a small fraction of familial aggregation of pancreatic cancer can be explained by previously identified, highly penetrant mutations in *BRCA2*, *CDKN2A* (also known as *p16*), *STK11* (also known as *LKB*), *APC*, *BRCA1*, *PRSS1* and *SPINK*[2,3]. Truncating mutations and deletions in *PALB2* have also recently been shown to be involved in familial pancreatic cancer[4,5].

We recently reported common risk variants for pancreatic cancer

American Journal of Epidemiology

Vol. 172, No. 1
DOI: 10.1093/aje/kwq116
Advance Access publication:
June 18, 2010

Original Contribution

Circulating 25-Hydroxyvitamin D and the Risk of Rarer Cancers: Design and Methods of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers

Lisa Gallicchio*, Kathy J. Helzlsouer, Wong-Ho Chow, D. Michal Freedman, Susan E. Hankinson, Patricia Hartge, Virginia Hartmuller, Chinonye Harvey, Richard B. Hayes, Ronald L. Horst, Karen L. Koenig, Laurence N. Kolonel, Francine Laden, Marjorie L. McCullough, Dominick Parisi, Mark P. Purdue, Xiao-Ou Shu, Kirk Snyder, Rachael Z. Stolzenberg-Solomon, Shelley S. Tworoger, Arti Varanasi, Jarmo Virtamo, Lynne R. Wilkens, Yong-Bing Xiang, Kai Yu, Anne Zeleniuch-Jacquotte, Wei Zheng, Christian C. Abnet, Demetrius Albanes, Kimberly Bertrand, and Stephanie J. Weinstein

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Body-Mass Index and Mortality among 1.46 Million White Adults

Amy Berrington de Gonzalez, D.Phil., Patricia Hartge, Sc.D., James R. Cerhan, Ph.D., Alan J. Flint, Dr.P.H., Lindsay Hannan, M.S.P.H., Robert J. MacInnis, Ph.D., Steven C. Moore, Ph.D., Geoffrey S. Tobias, B.S., Hoda Anton-Culver, Ph.D., Laura Beane Freeman, Ph.D., W. Lawrence Beeson, Dr.P.H., Sandra L. Clipp, M.P.H., Dallas R. English, Ph.D., Aaron R. Folsom, M.D., D. Michal Freedman, Ph.D., Graham Giles, Ph.D., Niclas Hakansson, Ph.D., Katherine D. Henderson, Ph.D., Judith Hoffman-Bolton, Jane A. Hoppin, Sc.D., Karen L. Koenig, Ph.D., I-Min Lee, Sc.D., Martha S. Linet, M.D., Yikyung Park, Sc.D., Gaia Pocobelli, M.S., Arthur Schatzkin, M.D., Howard D. Sesso, Sc.D., Elisabete Weiderpass, Ph.D., Bradley J. Willcox, M.D., Alicja Wolk, Dr.Med.Sci., Anne Zeleniuch-Jacquotte, M.D., Walter C. Willett, M.D., Dr.P.H., and Michael J. Thun, M.D.

ABSTRACT

BACKGROUND

A high body-mass index (BMI, the weight in kilograms divided by the square of the height in meters) is associated with increased mortality from cardiovascular disease and certain cancers, but the precise relationship between BMI and all-cause mortality remains uncertain.

Cancer Control and Population Sciences Home

Epidemiology and Genomics Research Home

Key Research Resources and Initiatives

Consortia (EGRP Facilitated and Funded)

Cohort Consortium:

Cohort Consortium Home
Mission
Leadership
Members
Publications
⊞ Projects
Strategic Plan for 2009-2010
⊞ Annual Meetings

Cohort Consortium

Search: GO

Save the Date

The 2012 Cohort Consortium Annual Meeting will be October 24-26, 2012 at Natcher Conference Center on the NIH Campus in Bethesda, Maryland. The main meeting will be October 25th and the ancillary project meetings will be October 24-26. More details coming soon!

Overview

The Cohort Consortium is an extramural-intramural partnership formed by the National Cancer Institute (NCI) to address the need for large-scale collaborations to pool the large quantity of data and biospecimens necessary to conduct a wide range of cancer studies. The Consortium, through its collaborative network of investigators, provides a coordinated, interdisciplinary approach to tackling important scientific questions, economies of scale, and opportunities to quicken the pace of research. View the Cohort Consortium's mission statement.

The Cohort Consortium includes investigators responsible for 43 high-quality cohorts involving more than 4 million people. The cohorts are international in scope and cover large and diverse populations. Extensive risk factor data are available on each cohort, and biospecimens including germline DNA collected at baseline, are available on more than 2 million individuals. Investigators team up to use common protocols and methods, and to conduct coordinated parallel and pooled analyses.

View a list of Cohort Consortium publications.

[Return to Top]

Membership

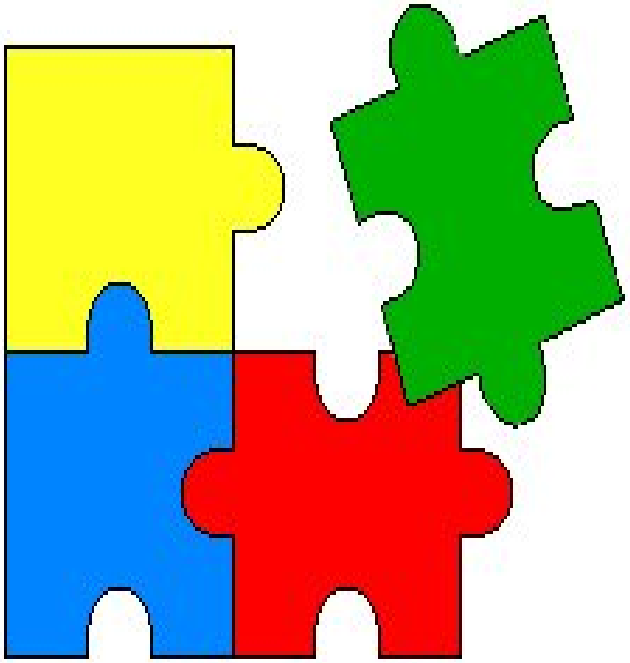
What questions will sequencing answer for this disease or trait?

- Will it reveal novel genetic influences?
- Will it reveal how a known or suspected exposure fits with a gene to influence risk?
- Will it implicate a biologic intermediate?
- Will it bend the risk prediction curve?
- Will it connect one disease to another?

Voyage of Discovery: Expect the Unexpected

- Q: Will it reveal novel genetic influences?
  - A: Yes.
  - But will there be many or important hits?
- Heritability was a good clue to GWAS hits.
  - E.g., Prostate cancer, height
- Suppose a powerful environmental agent (tobacco, nutrition, radiation) affects risk
  - Big E has not predicted how many genetic hits are found
  - Tobacco: Lung, bladder, pancreatic cancer
  - Diet: Menarche, menopause, BMI
- Are tissue comparisons relevant, available?

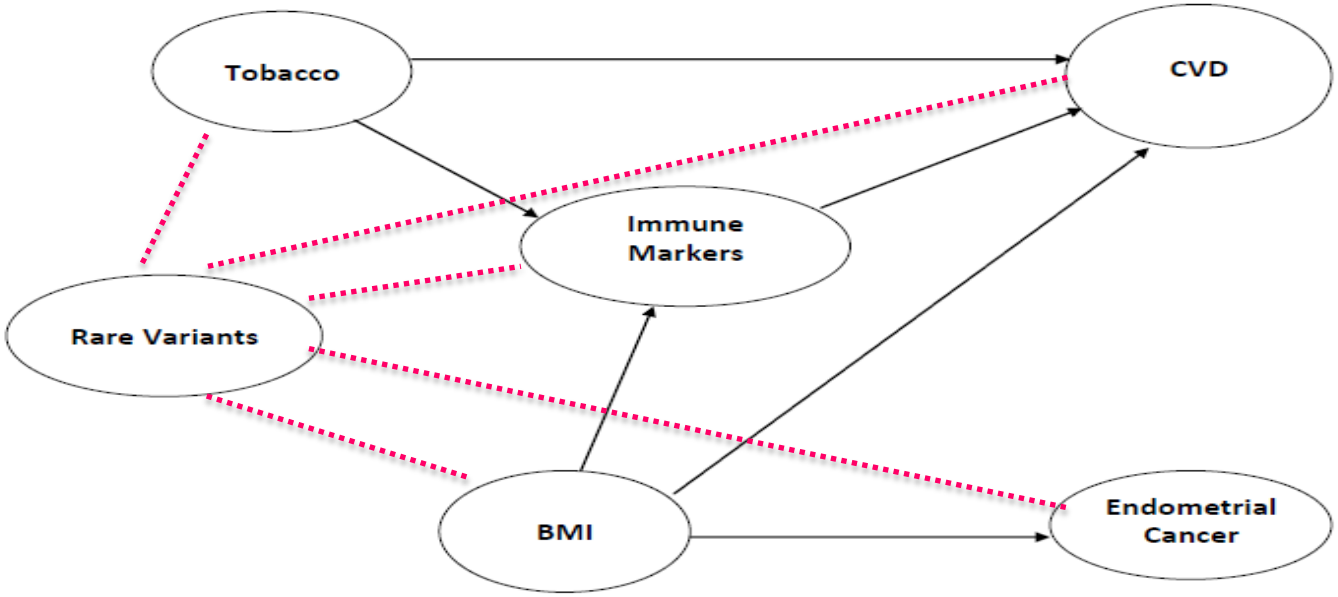
Interlocking Parts: Gene and Environment



- Need big enough for G|E
- Must have measured E
- Good to have large range of E
  - General population for tobacco, obesity, residence, activity
  - May need specialty cohorts for ionizing radiation
  - E.g., 6q21 variants implicate PRDM1 in second cancer after Hodgkin lymphoma

Couriers and other middle men

- Suppose novel rare variants reveal gene linked to endometrial cancer and to CVD.
- Known: Tobacco strongly linked to CVD, not to endometrial cancer.
- Serum inflammation markers follow tobacco, precede BMI, CVD.
- Serum hormone markers follow weight gain, precede endometrial cancer, not CVD.



Bending the Curve

- We want to have a very bent curve.
- Start with the things we know from questions.
- Predict or explain the risk in the population.
- Add the effects of known genetic risks.

➢ Will sequencing help to predict? How much?

In Search of a Cohort…

- Modern (sturdy consent, rich biobank, smooth access, easy informatics)
- Mature (we have questions to ask now)
- Large (a million is good)
- With data on phenotypes, behaviors, environments at baseline

…With Enriched Subsets

- With later questionnaire data
- With serial samples
- With all or subset already genotyped for GWAS

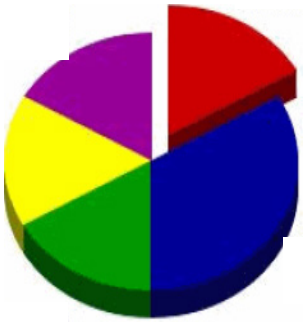
…Or a “Synthetic” Cohort: Cohorts Combined for Power

- Fraction of the cost
- Fraction of the time
- Experience in sharing
- Harmonized data
- We know it works

Combined for Larger Base, Sampled for Focused Questions

Subset selected for Sequencing

Larger subset with GWAS done,
or with serum ready, etc.



Full set, e.g. 1 million in pooled cohort base

Example: Stratified Case-Cohort

- Sample from all cohorts (for cross-sectional comparisons now and incident illnesses later)
- Sample of cases with disease X
- Sample of cases with disease Y
- Sample of obese participants, or heavy smokers
- Participants with family history

➢ Piece of cake!

On Balance

Mature versus new

- Many events have occurred, time for full spectrum of effects of an exposure
- Cross-sectional analyses of common phenotypes: great versus good

Combined versus single cohort

- Mosaic versus uniform, harmonized data (and investigators)
- Size requirements vary greatly

Case-cohort versus other designs

Taking our best guess on what sequencing will tell us

Thank You