PRECISION MEDICINE / PRECISION HEALTH

Karen L. Edwards, PhD
Professor, Dept. of Epidemiology
School of Medicine
University of California, Irvine
Irvine, CA

Outline

- Two National Initiatives
  - Precision Medicine Initiative
  - Cancer Moonshot Initiative

The Precision Medicine Initiative

2011: National Academies of Sciences

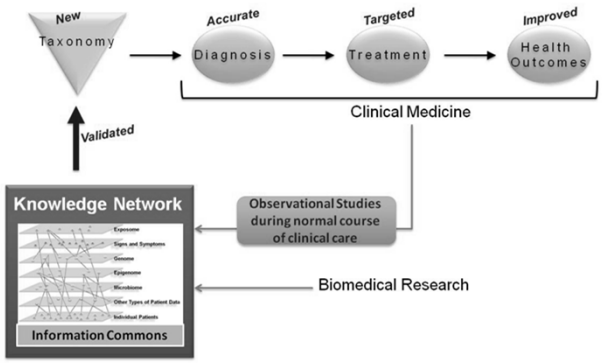


Figure S-1: Creation of a New Taxonomy first requires an "Information Commons" in which data on large populations of patients become broadly available for research use and a "Knowledge Network" that adds value to these data by highlighting their interconnectedness and integrating them with evolving knowledge of fundamental biological processes.

a "new taxonomy" that defines disease based on underlying molecular and environmental causes, rather than on physical signs and symptoms

http://dels.nas.edu/Report/Toward-Precision-Medicine-Building-Knowledge/13284

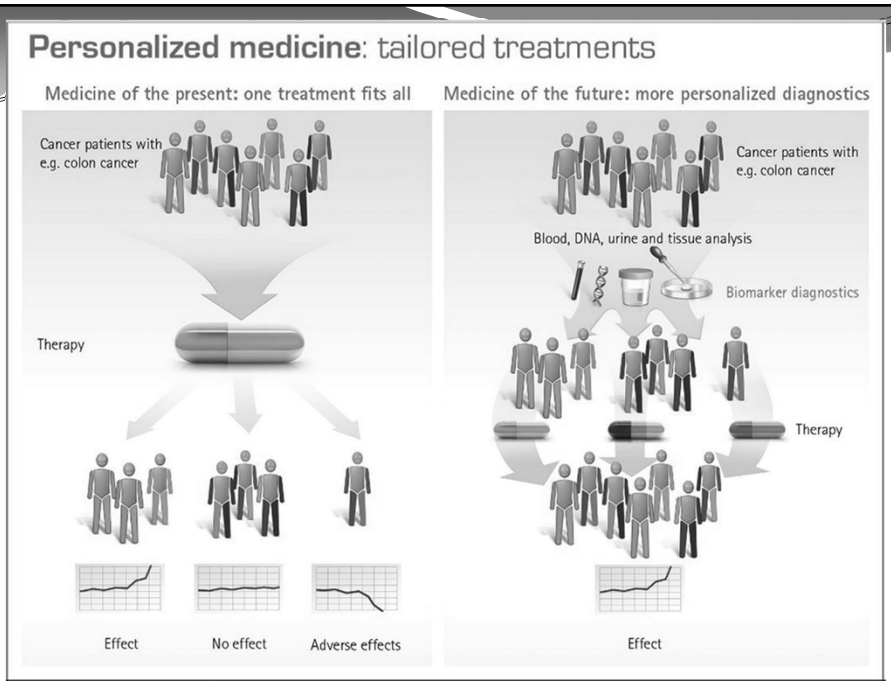
What is the Precision Medicine Initiative?

Mission statement:

To enable a new era of medicine through research, technology, and policies that empower patients, researchers, and providers to work together toward development of individualized care.

The future of precision medicine will enable health care providers to tailor treatment and prevention strategies to people's unique characteristics, including their genome sequence, microbiome composition, health history, lifestyle, and diet. To get there, we need to incorporate many different types of data, from metabolomics (the chemicals in the body at a certain point in time), the microbiome (the collection of microorganisms in or on the body), and data about the patient collected by health care providers and the patients themselves. Success will require that health data is portable, that it can be easily shared between providers, researchers, and most importantly, patients and research participants.

Agencies across the Federal government are doing important work to support the President's vision. This is an "all of government" effort, leveraging the unique expertise and history of each agency to carry forward the President's vision of individualized treatments for every American. Here's how each participating agency is moving ahead to implement PMI:



Credit: www.bayerpharma.com/en/research-and-development/research-focus/oncology/personalized-medicine/index.php

Pharmacogenomics

- Pharmacogenomics (sometimes called pharmacogenetics) is a field of research focused on understanding how genes affect individual responses to medications. The long-term goal of pharmacogenomics is to help doctors select the drugs and dosages best suited for each person.
  - NIGMS Pharmacogenomics Fact Sheet (https://www.nigms.nih.gov/education/pages/factsheet-pharmacogenomics.aspx)

https://precisionmedicine.duke.edu/researchers/precision-medicine-programs/pharmacogenomics

Genomic Medicine

- Genomic Medicine: *An emerging medical discipline that involves using genomic information about an individual as part of their clinical care (e.g., for diagnostic or therapeutic decision-making) and the other implications of that clinical use.*
  - *NHGRI Definition (https://www.genome.gov/27552451/what-is-genomic-medicine/)*

Precision Medicine

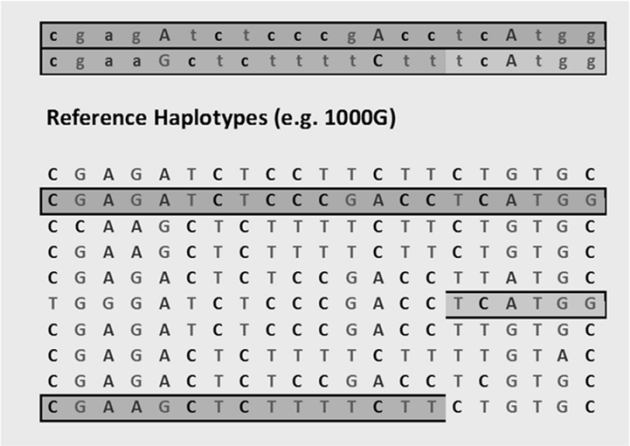
- Goal of the national Precision Medicine Initiative (PMI) is to provide precise health advice, diagnoses, and treatments for each individual in the population
- Approach: identify subsets of (homogeneous) patients most likely to benefit from a particular treatment or intervention
- Molecular profiling to create diagnostic, prognostic and therapeutic strategies tailored to each patient
- Individualized or *Precision Prevention* is in the future
  - Will require understanding Gene -Environment interactions
  - Role of Epigenetics
  - Approaches to change behavior

National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington (DC): National Academies Press (US); 2011. Available from: http://www.ncbi.nlm.nih.gov/books/NBK91503/

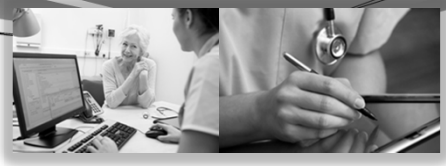
Precision Medicine Cohort Program

- Goal is to enroll 1 million people
- Health Systems
- Volunteers
- Diverse sample that represents the US population
  - Age, gender, race, ethnicity
  - Without regard to disease or health status
- Participation will require
  - Provision of a biological sample for genetic analysis
  - Clinical examination
  - Access to medical records and health information
  - Regular updates of information and longitudinal participation
  - Willingness to enroll in other studies
  - Collection of data via other mechanisms and new technology
- Return of Results and Access to data by Participants

Sequence Data



Genomics



EHRs

Technologies

Data Science

Patient Partnerships

Requires participation on a scale we have not seen before

Precision Medicine

- Genomics is an integral component
- Genetic Epidemiology will play a central role

Annotation

- Annotation is the process of marking genes or other biological features in a DNA sequence
- In genetic epi it is used to identify genes, predict function of genes and their variants
  - Study design – focus assessment on genes / variants that are likely functional
  - Interpretation – what is the function of the gene / variant that we have identified
- Continually updated and changing

Bioinformatics

- Primary goal is to increase understanding of biological processes
- With growing amount of (genomic) data, it is not practical to analyze manually
  - Develops and applies techniques to manage and analyze large volumes of data
  - Create and promote databases, algorithms, computational and statistical techniques and theory related to management and analysis of biologic (genomic) data
  - End users of the data
- Some overlaps with genetic epidemiology

Precision Medicine: What is needed (a genetic epidemiologist's perspective)

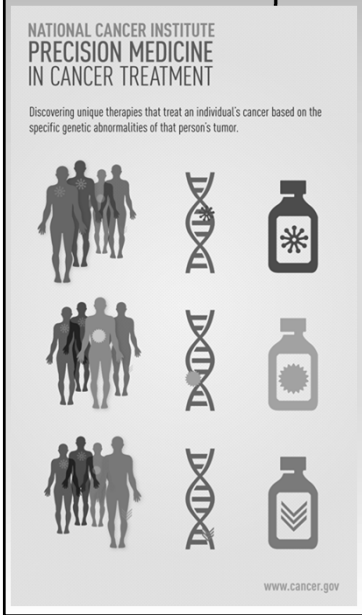
- Ability to stratify patients, tumors, population into more homogeneous subgroups
  - Genomic and epigenomic profiling
  - Biomarkers
  - Evidence for Clinical Validity and Utility
- Large samples – 1 million person national PM cohort
  - Big Data – linking with EHR and other data sources
  - Will rely on existing studies and data
  - May be a need for new data collection to fill in gaps
  - Engaging research participants
- Stakeholder input
  - Incorporate input from broad stakeholder groups, including research participants
- Incorporate environmental and lifestyle factors for Prevention
  - Focus on Gene x Environment interactions
    - Modifiable environmental and lifestyle factors
  - Approaches for behavior change that promote adoption of risk reducing factors

Precision Oncology

Carolyn M. Hutter, PhD
Program Director
Division of Genomic Medicine, NHGRI

SISG Lecture July 20, 2016

2015 Precision Medicine Initiative: Two Components: Cancer & National Cohort

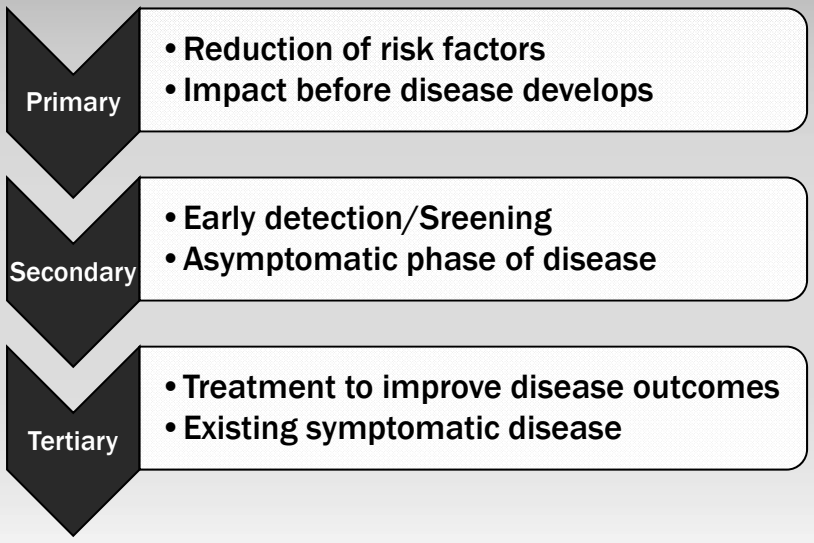


"An emerging approach for disease prevention and treatment that takes into account people's individual variations in genes, environment, and lifestyle"

Building a Knowledge Base

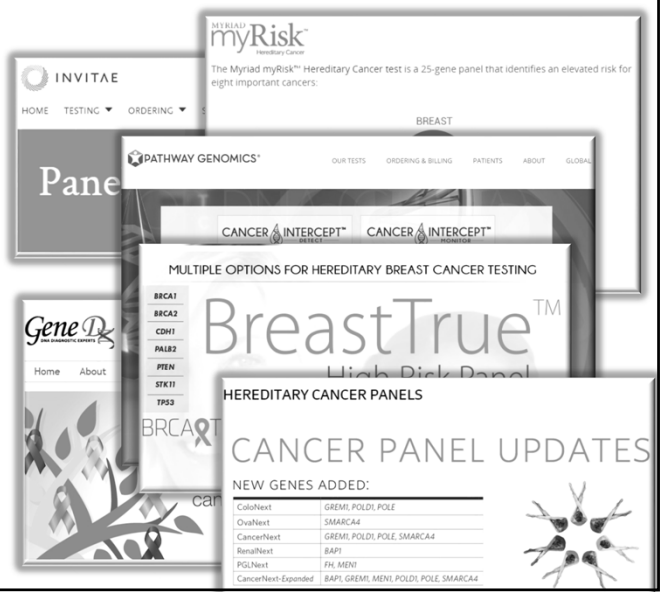
- Key Concept underlying most Precision Medicine Efforts
- Collect multi-dimensional information on large numbers of participants
- Share this knowledge
- Analyze the information
- Apply it in ways that impact individual and population health

Levels of Prevention



Multi-Gene Panels

- Increasingly commonly used in cancer risk setting
- But how strong is the evidence that variation in the genes on the panel cause the disease in question?



ClinGen Clinical Validity Summary Matrix

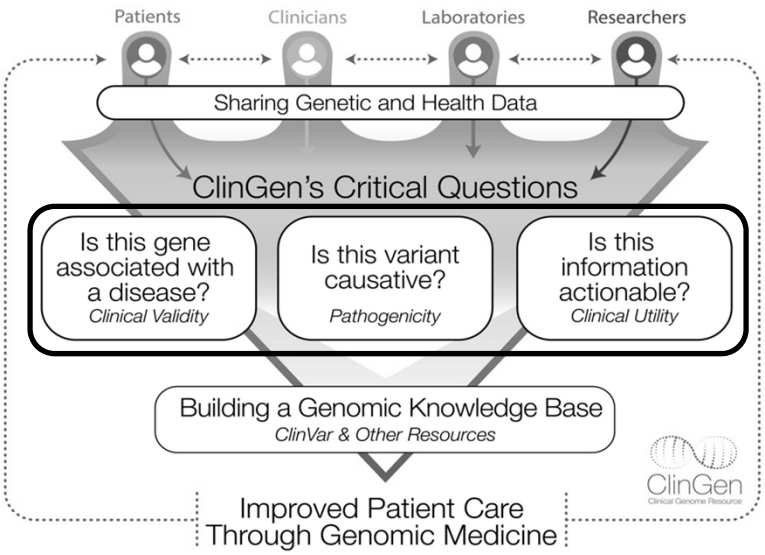
Assertion criteria	Description	Number of Points 0	1	2	3	4	5	6	7
# Probands	Total # of *unrelated* probands with variants that provide convincing evidence for disease causality across all curated literature	N/A	1-3	4-6	7-9	10-12	13-15	16-18	19+
Functional evidence	Points given based on the **gene-level** functional evidence supporting a role for this gene in disease	0	1	2	3	4	5	6+	
# Publications	# of curated Independent publications reporting **human** variants in the gene under consideration	N/A	1	2	3	4	5+		
Time (yrs)	# of years since first publication reporting a disease association **(if ≤2 publications --> then 1 is max score for time)**	this yr	1-3 yr	≥3 yr					
Is there valid contradictory evidence?		Y/N?							
Description of Contradictory Evidence:									

Classification	Total Score
Limited:	2-8
Moderate:	9-12
Strong:	13-16
Definitive:	17-20

Assertion:

Explanatory video available at: http://calculator.clinicalgenome.org/ashg-2015

"Building a genomic knowledge base to improve patient care."



Hereditary Cancer WG: Preliminary classifications suggest 50% of curated genes on clinical multigene panels for pancreatic cancer demonstrate only limited evidence

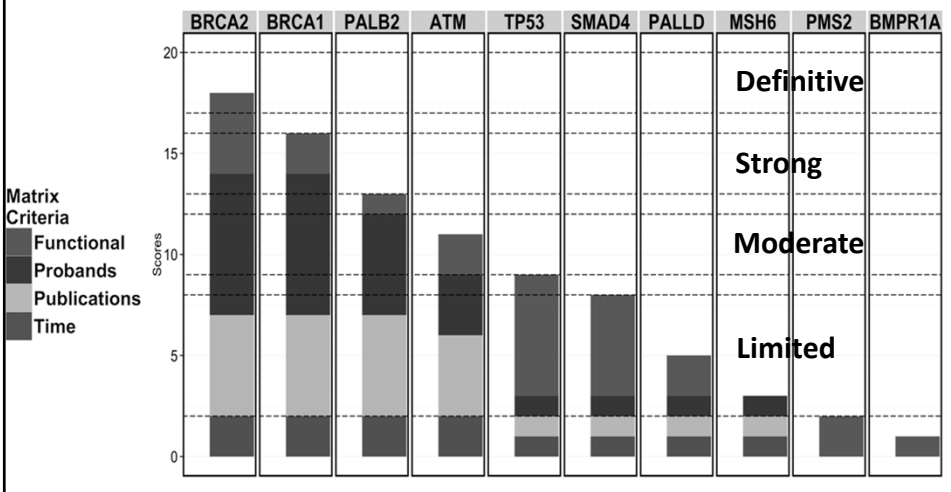


Figure by Raj Ghosh

* Pending expert review

BRCA1 Counseling/Testing Recommendations of the U.S. Preventive Services Task Force (USPSTF, 2013)

- The USPSTF recommends that primary care providers screen women who have family members with breast, ovarian, tubal, or peritoneal cancer with one of several screening tools designed to identify a family history that may be associated with an increased risk for potentially harmful mutations in breast cancer susceptibility genes (*BRCA1* or *BRCA2*). Women with positive screening results should receive genetic counseling and, if indicated after counseling, BRCA testing.
  Grade: B Recommendation.

- The USPSTF recommends against routine genetic counseling or BRCA testing for women whose family history is not associated with an increased risk for potentially harmful mutations in the *BRCA1* or *BRCA2* genes.
  Grade: D Recommendation.

http://www.ahrq.gov/clinic/uspstf/uspsbrgen.htm

Underutilization of *BRCA1/2* testing to guide breast cancer treatment: Black and Hispanic women particularly at risk

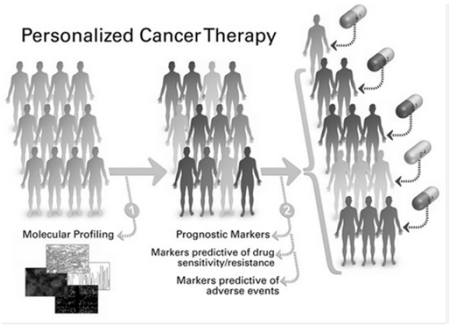
Douglas E. Levy, PhD[1,2,3], Stacey D. Byfield, PhD, MPH[4], Catherine B. Comstock, MPH,[5] Judy E. Garber, MD, MPH[3,6], Sapna Syngal, MD, MPH[3,6,7], William H. Crown, PhD[4], and Alexandra E. Shields, PhD[1,2,3]

Purpose: Women with early-onset (age ≤40 years) breast cancer are at high risk of carrying deleterious mutations in the *BRCA1/2* genes; genetic assessment is thus recommended. Knowledge of *BRCA1/2* mutation status is useful in guiding treatment decisions. To date, there has been no national study of *BRCA1/2* testing among newly diagnosed women. **Methods:** We used administrative data (2004–2007) from a national sample of 14.4 million commercially insured patients to identify newly diagnosed, early-onset breast cancer cases among women aged 20–40 years (n = 1474). Cox models assessed *BRCA1/2* testing, adjusting for covariates and differential lengths of follow-up. **Results:** Overall, 30% of women aged 40 years or younger received *BRCA1/2* testing. In adjusted analyses, women of Jewish ethnicity were significantly more likely to be tested (hazard ratio = 2.83, 95% confidence

to assess risk of hereditary breast and ovarian cancer (HBOC) is among the most established genetic tests in clinical use.[1–3] Guidelines and commercial testing for *BRCA1/2* mutations have been available for more than a decade,[4] and most health insurers now reimburse at least partially for these tests in individuals at high risk for mutations.[5] National guidelines recommend that women diagnosed with early-onset breast cancer receive *BRCA1/2* testing to guide treatment decisions.[6] Among patients newly diagnosed with cancer, a positive test result will often prompt more aggressive surgical treatment (e.g., bilateral salpingo oophorectomy or prophylactic contralateral mastectomy) with the goal of minimizing the potential for second primary cancers.[3,7,8] A positive test result may also prompt consideration

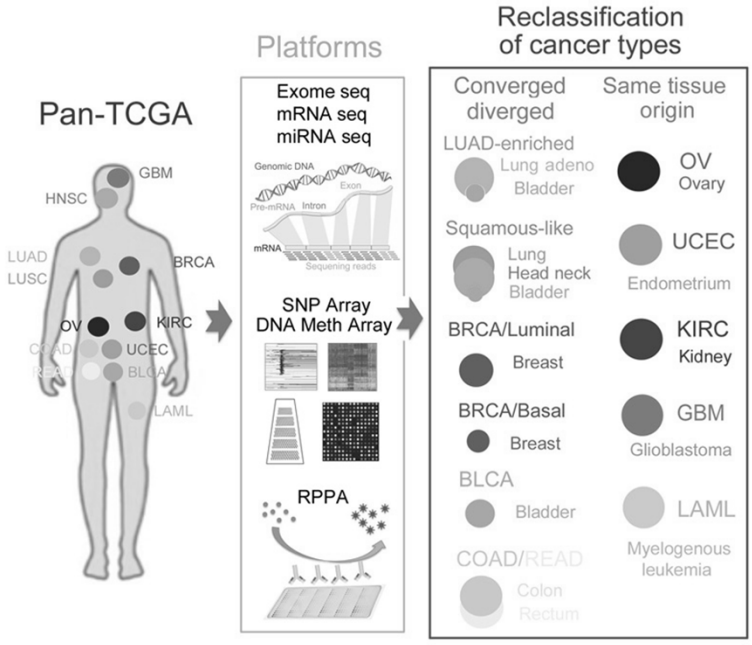
Genetics in Medicine 2011

Precision Cancer Genomics



Applications of Genomics in Clinical Cancer Care

- Diagnostic sub-classifications
- Novel prognostic biomarkers
- Genomic predictors of response to therapies
- Genomic mechanisms of resistance to therapies

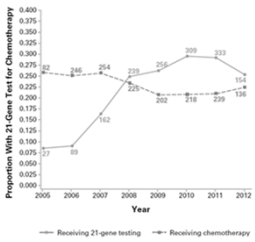


CLINICAL

Breast Cancer Multigene Testing Trends and Impact on Chemotherapy Use

G. Thomas Ray, MBA; Jeanne Mandelblatt, MD; Laurel A. Habel, PhD; Scott Ramsey, MD, PhD; Lawrence H. Kushi, ScD; Yan Li, MD; and Tracy A. Lieu, MD, MPH

Figure 1. Proportion of 7004 Eligible Patients With Breast Cancer Receiving 21-Gene Testing and Chemotherapy by Year (Kaiser Permanente Northern California)



Solid line indicates receiving 21-gene testing; dotted line indicates receiving chemotherapy. Values at plot points are the number of women with 21-gene test or chemotherapy (2005 and 2012 are partial years).

Table 3. Chemotherapy Treatment by Patient Group and 21-Gene Test Recurrence Score (Kaiser Permanente Northern California, September 2005-June 2012)[a]

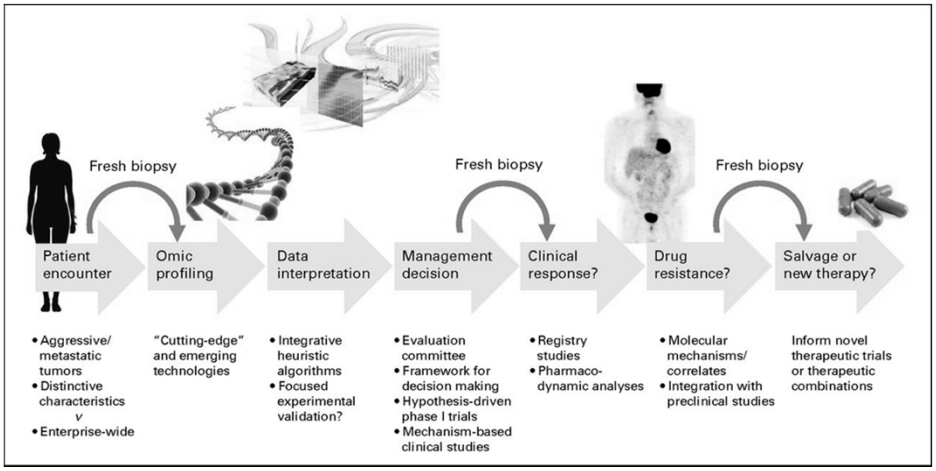
Characteristic	Total Patients	Patients Receiving Chemotherapy, n (%)
All patients	7004	1600 (23)
Patients who did not have Oncotype DX testing	5437	1190 (22)
Patients who had Oncotype DX testing	1567	410 (26)
By RS[b]		
Low risk (RS ≤18)	820	68 (8)
Intermediate risk (RS = 18 to 30)	606	241 (40)
High risk (RS ≥31)	141	101 (72)

RS indicates Recurrence Score.
[a]Among breast cancer cases meeting National Comprehensive Cancer Network criteria for consideration of the 21-gene test (Oncotype DX). See text for inclusion criteria.
[b]Differences in the percent of patients receiving chemotherapy among the 3 Recurrence Score groups was statistically different at $P \leq .01$.

- 21 gene test that predicts recurrence for HR+ women
- Only 22% of those who meet guidelines were actually tested
- Recurrence risk scores correlated with use of chemotherapy
- Introduction of test associated with modest decrease in overall chemotherapy use

Am J Manag Care. 2016;22(5):e153-e160

Paradigm for Cancer Precision Medicine



Levi Garraway, *Journal of Clinical Oncology, 2013*

Assigning Meaning: What is an Actionable Tumor Genomic Alteration?

An "actionable" tumor genomic alteration is a Clinically Relevant genomic finding in a patient's tumor sample that has implications for clinical care.

- Therapeutic implications
  - Sensitivity to Therapies
  - Resistance to Therapies
- Prognostic implications
- Diagnostic implications

Slide by Nick Wagle

Examples:

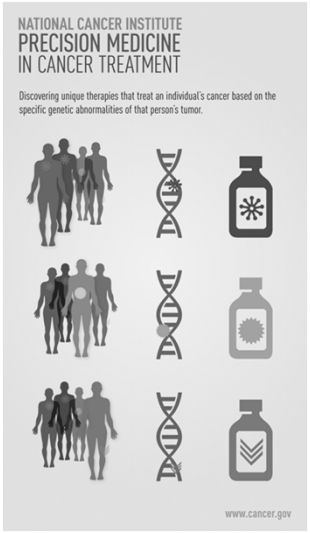
BRAF V600E mutations may predict sensitivity to RAF inhibitors in melanoma

KRAS codon 12 and 13 mutations may predict resistance to anti-EGFR antibody treatment in colorectal cancer

Certain IDH1 mutations may predict favorable prognosis in gliomas

BCR-ABL translocations are diagnostic, prognostic, and predictive of response to treatment with imatinib

NCI and the Precision Medicine Initiative®



- NCI is focusing its PMI activities on four broad areas:
  - **Expanding Precision Medicine Clinical Trials**
  - **Overcoming Drug Resistance**
  - **Developing New Laboratory Models for Research**
  - **Developing a National Cancer Knowledge System**

Genomically Based Clinical Trials

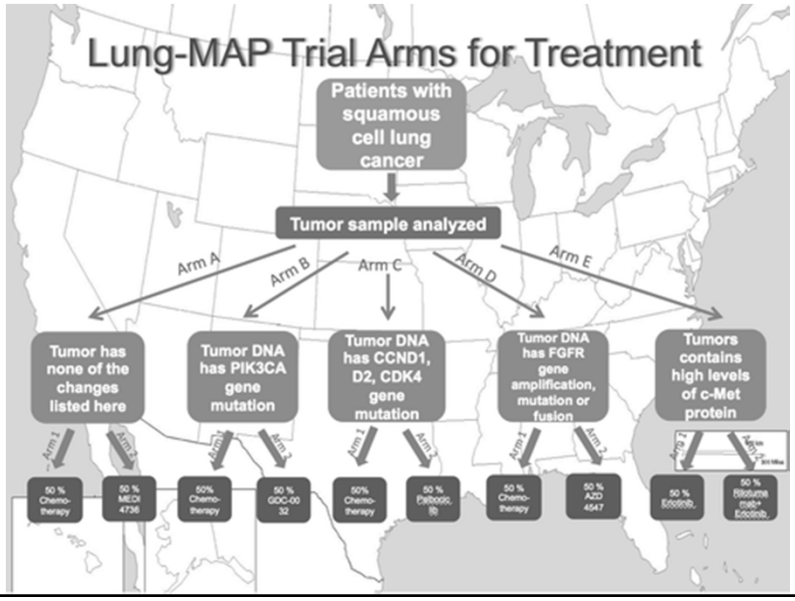
Umbrella Trials

- Same Tumor
- Different mutations
- Test impact of different drugs on different mutations in a single cancer type

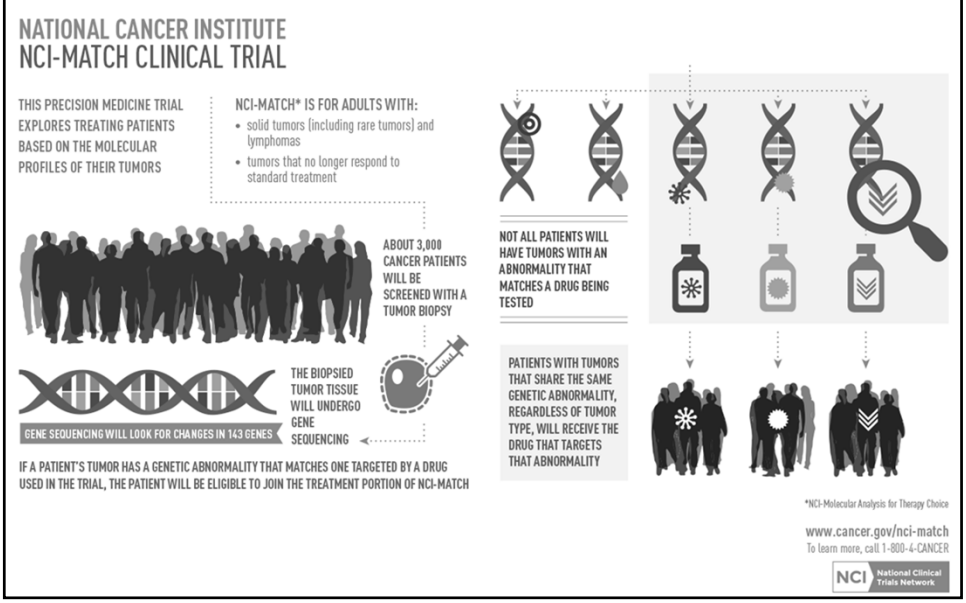
Basket Trials

- Different tumors
- Similar mutations
- Test effect of a drug on single mutation in a variety of cancer types

NCI Precision Medicine Trials: Lung-MAP



NCI Precision Medicine Trials: NCI Match



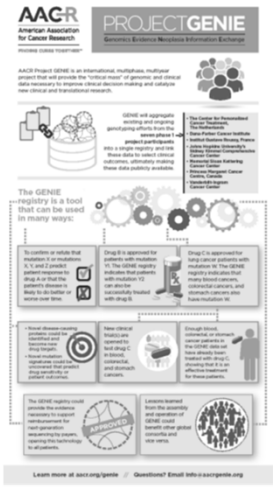
Cancer Moonshot Initiative

- **Cancer Vaccine Development**
- **Early Cancer Detection**
- **Immunotherapy and Combination Therapies**
- **Genomic Profiling of Tumor and Surrounding Cells**
- **Enhanced Data Sharing**
- **Pediatric Cancer**
- **Oncology Center of Excellence**
- **Vice President's Exceptional Opportunities in Cancer Research Fund**

http://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/research-areas

Sharing Clinical Interpretation in Cancer Genomics

Global Alliance for Genomics & Health



MY CANCER GENOME
GENETICALLY INFORMED CANCER MEDICINE

Find a Cancer Mutation

Disease (required): Select Disease

Gene (optional): Select Disease First

Variant (optional): Select Disease First

Project

CIViC BETA
CLINICAL INTERPRETATIONS OF VARIANTS IN CANCER

CIViC: Clinical Interpretations of Variants in Cancer

Discover supported clinical intepretations of mutations related to cancer and collaborate with colleagues to add variants and support for cancer-related mutations.

CIViC: Clinical Interpretations of Variants in Cancer Details

Precision medicine refers to the use of prevention and treatment strategies that are tailored to the unique features of each individual and their disease. In the context of cancer this might involve the identification of specific mutations shown to predict response to a targeted therapy. The biomedical literature describing these associations is large and growing rapidly. Currently these interpretations exist largely in private or

2. Approve: Community Editors and Moderators may approve submitted evidence items, after which the community may view and edit the item.
3. View: Make use of the community-created content in your own research by browsing, searching, and examining detailed evidence items. All CIViC data and source code are provided freely for almost any use.

Tension Between Precision Medicine and Public Health

Public Health in the Precision-Medicine Era

Ronald Bayer, Ph.D., and Sandro Galea, M.D., Dr.P.H.

That clinical medicine has contributed enormously to our ability to treat and cure sick people is beyond contention. But whether and to what extent medical care has transformed morbidity and mortality patterns at a population level and what contribution, if any, it has made to the well-being and life expectancy of the least-advantaged people have been matters of contention for more than a century. This debate has taken on renewed importance as the scientific leadership at the National Institutes of Health (NIH), National Academy of Medicine, and U.S.

JAMA, June 2015

VIEWPOINT

Seven Questions for Personalized Medicine

Michael J. Joyner, MD
Department of Anesthesiology, Mayo Clinic, Rochester, Minnesota.

Nigel Paneth, MD, MPH
Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing; and Department of Pediatrics and Human Development, College of Human Medicine, Michigan State University, East Lansing.

Personalized or precision medicine maintains that medical care and public health will be radically transformed by prevention and treatment programs more closely targeted to the individual patient. These interventions will be developed by sequencing more genomes, creating bigger biobanks, and linking biological information to health data in electronic medical records (EMRs) or obtained by monitoring technologies. Yet the assumptions underpinning personalized medicine have largely escaped questioning. In this Viewpoint, we seek to stimulate a more balanced debate by posing 7 questions for the advocates of personalized medicine.

Does the Human Genome Contribute to Disease Risk Prediction?

Personalized medicine builds on the Human Genome Project, which was forecasted to revolutionize disease risk prediction, with projected relative risks as high as 6 for gene variants linked to specific diseases. However, the relative risks for the vast majority of gene variants are offered enhanced screening and preemptive surgery. In the 25 years since *BRCA1/2* was discovered, breast cancer mortality in the United States has declined by nearly one-third; however, little of this decline stems from the discovery of *BRCA1/2*. Moreover, *BRCA1/2* is a unique story because the gene variants account for such a substantial amount of the variance in outcome for a limited number of patients.

In CF, 2 drugs (ivacaftor and lumacaftor) have recently been developed based on the CF transmembrane conductance regulator gene (*CFTR*), but they are useful only in patients with specific *CFTR* mutations, in whom they increase, singly or in combination, maximum forced expiratory volume (FEV) by 5% to 10% and improve weight gain.[4] However, since the discovery of this gene in the 1980s, CF survival has improved substantially as a result of strict adherence to clinical management guidelines originating in pulmonary physiology and infectious diseases, but not genomics.

Although well-deserved recognition has accompa-

Too much emphasis on healthcare!

"We worry that an unstinting focus on precision medicine… is a mistake — and a distraction from the goal of producing a healthier population."

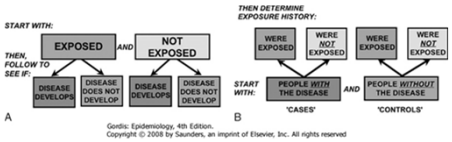
Bayer and Galea, NEJM, 2015

Summary

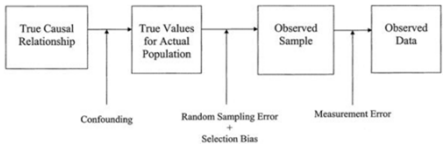
- Precision Medicine takes into account individual differences in people's genes, environments, and lifestyles.
- Precision Medicine may impact cancer at primary, secondary and tertiary prevention
- Cancer genomics is on the leading edge of genomic medicine (as are pharmacogenomics and rare diseases applications)
- Current precision medicine efforts include a focus on building knowledge basis and knowledge systems, including genomics

Epidemiology 101

- Epidemiology is the study of the distribution and determinants of health-related states in populations
- Study design is a key component of epidemiology
- Relative risks, risk differences and odds ratios are used to measure association
- It is important to consider and address bias in epi studies
- Understanding confounding and effect modification are important in studies of association

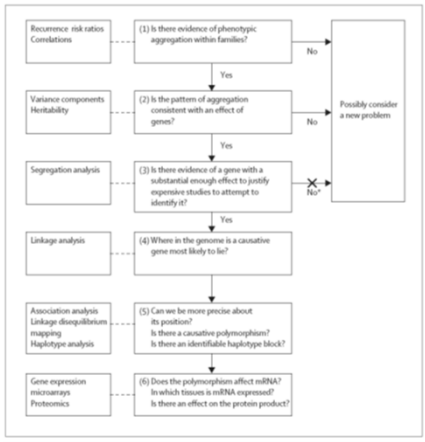


	Disease		
Exposure	+	-	Total
+	a	b	a+b
-	c	d	c+d

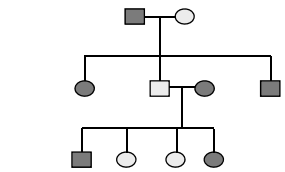


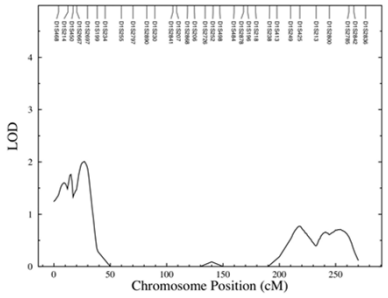
Overview of Genetic Epidemiology

- Different study designs in genetic epidemiology answer different questions about genetic basis of disease.
- Collecting family data and family history data is important but challenging.
- Pedigrees provide visual representations of family data.



Segregation and Linkage Analysis

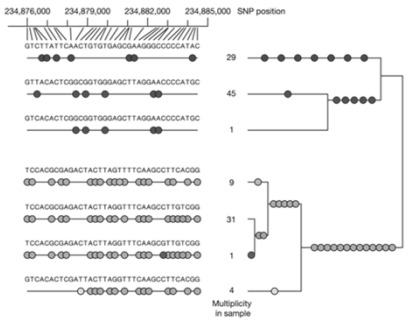




- Complex segregation analysis is a modeling approach for evaluating the transmission of a trait within a pedigree.
- Linkage analysis focuses on the co-segregation between a marker locus and a putative disease locus using family data.

Linkage Disequilibrium

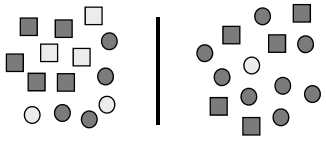
- LD refers to nonindependence of alleles
- Measures of LD include D, D' and r^2
- LD is exploited for association studies, imputation and GWAS
- LD in turn causes challenges in interpreting association findings and identifying the causal allele(s)



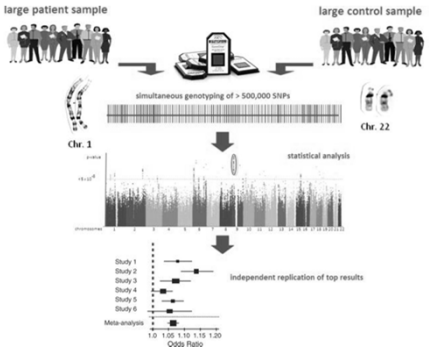
Association Studies

- Focus on association between a particular allele and disease in a population.
- Need to consider potential impact of population stratification.
- Family based association studies.
- Need to consider power and sample size in design and analysis of association studies.

Population based Association studies



Genome Wide Association Studies

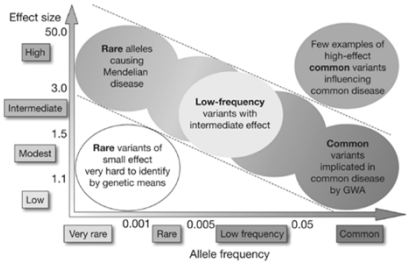


- GWAS perform association analysis on markers across the genome.
- Standards for QC, analysis and presentation.
- GWAS have been highly successful at identifying common variants with modest effect sizes associated with disease.
- Post-GWAS studies focus on functional characterization and epidemiologic interpretation.
- Increasing examples of applications of GWAS findings.

Gene-Environment Interactions

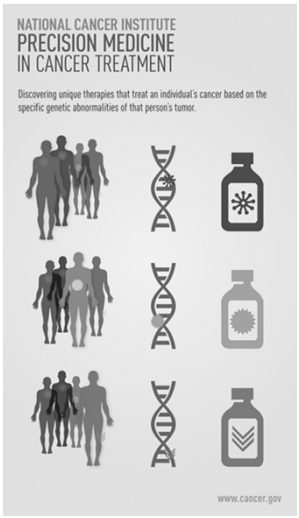
- Focus on interplay between genetic and environmental factors
- Need to consider scale (additive and multiplicative)
- Novel methods to improve power
- Attention needs to be given to measuring and analyzing both the G and the E

Sequencing Studies



- Next-generation sequencing allows us to examine rare and low-frequency variants.
- Gene-level tests are sometimes needed to evaluate this data.
- Family based study designs can be useful for identifying rare variants associated with disease phenotypes.

Precision Medicine



SISG Module Genetic Epidemiology
Optional Exercise: Measures of Association

Table 1: Association between binge drinking and atherosclerosis

	Disease Status	
Exposure Status	Case	Control
Exposed (+)	106	585
Unexposed (-)	186	2165

1. What is the incidence of outcome in the exposed group? In the unexposed group?
2. What is the RR for the association between exposure and outcome? The AR? The OR?

Table 2: Association between, stratified by gender

Men	**Disease Status**	
Exposure Status	Case	Control
Exposed (+)	89	374
Unexposed (-)	118	801

Women	**Disease Status**	
Exposure Status	Case	Control
Exposed (+)	17	211
Unexposed (-)	68	1364

3. What is the RR in men? What is the RR in women?
4. Is there evidence for confounding or effect modification?

SISG Module Genetic Epidemiology
Optional Exercise: Measures of Association
ANSWER KEY

Table 1: Association between binge drinking and atherosclerosis

	Disease Status	
Exposure Status	Case	Control
Exposed (+)	106	585
Unexposed (-)	186	2165

1. What is the incidence of outcome in the exposed group? In the unexposed group?

ANSWER: Incidence in exposed = 106/(106+585)=15.3%
Incidence in unexposed=186/(186+2165)=7.9%

2. What is the RR for the association between exposure and outcome? The AR? The OR?

ANSWER: RR = 1.9 (1.6 – 2.4); AR=7.4%; OR = 2.1 (1.6 – 2.7)
RR = 1.9 (1.6 – 2.4)

Table 2: Association between, stratified by gender

Men	**Disease Status**	
Exposure Status	Case	Control
Exposed (+)	89	374
Unexposed (-)	118	801

Women	**Disease Status**	
Exposure Status	Case	Control
Exposed (+)	17	211
Unexposed (-)	68	1364

3. What is the RR in men? What is the RR in women?

ANSER: RR in men= 1.50 (1.16-1.93); RR in women= 1.57 (0.94-2.62)

4. Is there evidence for confounding or effect modification?

ANSER: The RR is similar between men and women, so there is not strong evidence for effect modification. The RR in the stratified analysis is lower (closer to the null of 0) than in the combined analysis, indicating evidence of confounding.

1000 Genomes Browser: The 1000 Genomes Browser provides a visual access to the data and information gathered in the 1000 Genomes Project.

1. Go to the main page for the 1000 Genomes (http://www.1000genomes.org/). Click on the "About" tab.
    a. **QUESTION:** What is the stated Goal of this Project?
    b. **QUESTION:** How many samples were sequenced in total? How many of African Ancestry?
2. Go to the 1000 Genomes Browser Webpage (http://browser.1000genomes.org/index.html)
3. Click on "Browse Human"
4. Search for "UCHL1"
5. Click on "Region in Detail" under 3. UCHL1 - ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)
6. Click on "Linkage Data"
7. Click on "Select Populations"
8. Choose the "1000Genomes:phase_1_CEU" and "1000Genomes:phase_1_YRI"
    a. QUESTION: Which population seems to have more LD among variants in this region?
9. For more details, go to the 1000 Genomes Tutorials:
    a. http://www.1000genomes.org/using-1000-genomes-data
    b. ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/browser/1000genomes_browser_main_project_20110521/The_1000_Genomes_Browser_Tutorial.ensembl_65.doc

Genome variation Server: The Genome Variation Server (GVS) enables rapid access to human genotype data found in dbSNP, and provides tools for analysis of genotype data. It is particularly useful for examining LD between variants in particular genes or regions, and for selecting tagSNPs.

1. **Go to the website http://gvs.gs.washington.edu/GVS138/**
2. **Search for "UCHL1" under gene name**
3. **Select "HapMap-CEU"**
4. **Click the tab that says "Display Linkage Disequilibrium", and open the graphical display**
5. **Click the tab that says "Display tagSNPs", and open the table view.**
    a. **QUESTION: How many variants are there in total for this gene in the HapMap CEU population?**
    b. **QUESTION: How many tagSNPs would you need to genotype to capture the variation in this gene?**
6. **Click the "Display SNP Summary" tab and see what information is provided.**
7. **Repeat 3-6, but selecting the HapMap-YRI rather than the HapMap-CEU population.**
8. **Repeat the above for a gene of interest.**
9. **Go back to the main website. Click on the side tab that says "BatchGVS138" Batch GVS allows you to perform large automated searches.**
10. **Click on How to use Batch GVS to see details and examples for using this feature in Batch Mode "http://gvs.gs.washington.edu/GVSBatch138/HelpHowToUse.jsp"**

SISG Module Genetic Epidemiology
Optional Exercise: Linkage Disequilibrium

Table 1: Haplotype and allele frequencies

		Locus 2		
		B	b	
Locus 1	A	0.04	0.50	0.54
	a	0.27	0.19	0.46
		0.31	0.69	1.0

1. The table above provides hypothetical data, showing allele and haplotype frequencies from a population sample. Note, the designation of capital vs. lower case letters are arbitrary. At locus 2 the "b" allele is more frequent than the "B" allele in this population.
    a. Using the notation presented in the slides, state the values for p_A, p_a, p_B, p_b, p_{AB}, p_{Ab}, p_{aB}, and p_{ab} for this population sample.
    b. Describe in words what p_A and p_b represent.
    c. What is the most frequent haplotype in this sample? The least frequent?
    d. Based on this data , calculate D, r^2 and D' between Locus 1 and Locus 2.

2. Assume Locus 1 is the (untyped) disease causing SNP, and Locus 2 is the SNP you are genotyping in your case-control study. You have performed power calculations assuming that you had genotyped the disease causing allele. These power calculations indicate a sample size of 500 cases and 500 controls. Estimate the sample size you will actually need for your study.

ANSWER KEY

Table 1: Haplotype and allele frequencies

		Locus 2		
		B	b	
Locus 1	A	0.04	0.50	0.54
	a	0.27	0.19	0.46
		0.31	0.69	1.0

1. The table above provides hypothetical data, showing allele and haplotype frequencies from a population sample. Note, the designation of capital vs. lower case letters are arbitrary. At locus 2 the "b" allele is more frequent than the "B" allele in this population.
    a. Using the notation presented in the slides, state the values for p_A, p_a, p_B, p_b, p_{AB}, p_{Ab}, p_{aB}, and p_{ab} for this population sample.
    ANSWER: p_A=0.54, p_a=0.46, p_B=0.31, p_b=0.69, p_{AB}=0.04, p_{Ab}=0.50, p_{aB}=0.27, and p_{ab} =0.19
    b. Describe in words what p_A and p_b represent.
    ANSWER: p_A is the frequency of the A allele at locus 1. p_b is the frequency of the b allele at locus 2
    c. What is the most frequent haplotype in this sample? The least frequent?
    ANSWER: The most frequent haplotype is the Ab haplotype (50% of the population. The least frequent is the AB haplotype (4% of the population).
    d. Based on this data, calculate D_{AB}, r^2 and D' between Locus 1 and Locus 2.
    ANSWER:
    D_{AB} =$\mathbf{p_{AB} - p_A p_B}$=0.04-(0.54*0.46)= -0.12
    r^2 =$\mathbf{D^2 / (p_A\ p_B\ p_a\ p_b)=(}$-0.12^2)/(0.54*0.46*0.31*0.69)=0.27
    D'=If D_{AB}<0: $D'_{AB} = D_{AB}/(\min(P_AP_B,P_aP_b))$=-0.12/min(0.54*0.31, 0.46*0.69)
    =-0.12/min(0.16, 0.31)=-0.12/0.16=-0.75

2. Assume Locus 1 is the (untyped) disease causing SNP, and Locus 2 is the SNP you are genotyping in your case-control study. You have performed power calculations assuming that you had genotyped the disease causing allele. These power calculations indicate a sample size of 500 cases and 500 controls. Estimate the sample size you will actually need for your study.

ANSWER: $N_2=N_1/r^2$

N_2=500/0.27=1851 cases and 1851 controls.

In-Class Exercise: Calculating genetic, environmental, and gene-environment interaction effects

The Fargo-Akron-Kalamazoo Epidemiology (FAKE) study is a prospective cohort study of 80,000 adults from the U.S. Midwest, recruited at primary care visits to several large HMOs. All participants contributed a blood sample and completed a diet questionnaire at enrollment. FAKE investigators hypothesize that subjects carrying a nonsense mutation in *rhubarbulin 1* (*RBLN1-X*) will be more sensitive to the effects of heavy rhubarb consumption (equivalent to more than three slices of rhubarb pie per day), which can lead to chronic diarrhea (CD).[1] The following table shows the distribution of *RBLN1-X*, heavy rhubarb consumption, and incidence of CD over the first 5 years of follow up (assume no censoring).

RBLN1-X	Rhubarb pie	Cases	Controls	TOTAL
Non-Carrier	Unexposed	6,990	50,660	57,650
Carrier	Unexposed	2,468	11,939	14,407
Non-Carrier	Exposed	1,293	5,098	6,391
Carrier	Exposed	461	1,091	1,552
	TOTAL	11,212	68,788	80,000

1. If the investigators had not collected data on *RBLN1-X*, would they have detected an association between rhubarb consumption and CD? I.e. what is the marginal measure of association for CD comparing exposed to unexposed?

2. If the investigators had not collected data on rhubarb consumption, would they have detected an association between *RBLN1-X* and CD? I.e. what is the marginal measure of association for CD comparing carriers to non-carriers?

3. What are the genotype-stratum-specific effects of rhubarb consumption?

4. Would you say there is gene-environment interaction?

*Modified from exercise developed by Peter Kraft.

[1] I doubt that excessive rhubarb consumption can cause CD. But according to wikipedia, rhubarb can be used as a laxative.

In Class Exercise: GxE Harmonization

You continue to work with collaborators on the FAKE study. They decide to follow-up on their candidate gene study with a genome-wide association study (GWAS). They were only able to afford genome-wide genotyping on a subset of the subjects, so they decide to reach out to their collaborators in the Meta-Analysis of Diet and Environment for Understanding Phenotypes (MADE-UP) consortia. The next page has "table 1" for the 8 studies in this consortia. Brainstorm with your group about the following:

1. What are potential issues/challenges that you might encounter in analyzing this data?
2. What are solutions might you use for some of these challenges?
3. What additional information would be most helpful for you to have?

Table 1: Overview of Subject Characteristics

	Study 1: Cohort		**Study 2: Case-control**		**Study 3: Case-control**		**Study 4: Cohort**		**Study 5: Case-control**		**Study 6: Case-control**		**Study 7: Cohort**		**Study 8: Cohort**	
	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control
N	931	1,435	1,410	1,666	2,031	2,044	69	238	465	465	5,450	5,475	1,064	1,202	1,381	1,303
% Female	32.0%	43.2%	32.1%	44.1%	19%	16.5%	34.8%	26.2%	26.7%	26.8%	56%	56%	60%	55%	0%	0%
Mean Age (yrs)	65.5	65.8	65.1	67.5	59.8	61.3	58.1	57.8	62.4	62.8	64.0	64.2	61.3	62.8	65.4	65.4
% Strawberry eaters	47.7%	45.8%	45%	40%	65.2%	56.2%	60.9%	65.2%	55.4%	55.6%	59.3%	52.1%	58.2%	59.0%	65.3%	66.4%
% Rhubarb eaters	21.6%	15.1%	25.6%	24.5%	36.7%	34.5%	12.1%	7.1%	14.1%	10.2%	NA	NA	28.4%	33.4%	14.9%	10.7%
Instrument for dietary assessment	FFQ	FFQ	FFQ	FFQ	5 Q survey	5 Q survey	FFQ	FFQ	24 hour recall	24 hour recall	5 Q survey	5 Q survey	FFQ	FFQ	24 hour recall	24 hour recall
Country	USA	USA	USA	USA	China	China	Japan	Japan	Germany	Germany	USA	USA	Canada	Canada	USA	USA
Genotyping Platform	Illumina 550K		Affymetrix 6.0		Illumina 550K		Illumina 1M		Illumina Omni Express		Illumina Omni Express		Affymetrix Axiom-CEU		Affymetrix Axiom-CEU	

SISG Module
Genetic Epidemiology
Suggested Readings:

General

- Series on Genetic Epidemiology in The Lancet http://www.thelancet.com/series/genetic-epidemiology
- Austin MA. Genetic Epidemiology: Methods and Applications . CABI; 1 edition (August 2013)ISBN-10: 1780641818

Pedigree Drawing

- Bennett RL, et al Recommendations for standardized pedigree nomenclature.AJHG 1995; 56; 745-52
- Bennett RL et al,. Standardized human pedigree nomenclature: Update and assessment of the recommendations of the National Society of genetic counselors. J Genetic Counsel (2008); 17:424-433

Family Based Studies

- Jarvik GP. Complex segregation analysis: uses and limitations. Am J. Hum. Genet. 1998; 63: 942-046.
- Bellis C, Cox HC, Dyer TD, Charlseworth JC, Begley KN, Quinlan S, Lea RA, Heath SC, Blangero J, Griffiths LR. Linkage mapping of CVD risk traits in the isolated Norfolk Island population. Hum. Genet. 2008; 124(5): 543-552.

Linkage Disequilibrium

- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009;10:387-406
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65.
- Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9, 477-485 (June 2008)

GWAS

- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014 Jan;42.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL,

Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53.

- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008 May;9(5):356-69.

GxE Interactions

- Khoury MJ and Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies--challenges and opportunities. Am J Epidemiol. 2009; 169: 227-30.
- Thomas D. Gene--environment-wide association studies: emerging approaches. Nat Rev Genet. 2010 Apr;11(4):259-72.
- Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM; NCI Gene-Environment Think Tank. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. Genet Epidemiol. 2013 Nov;37(7):643-57.

Rare Variant/Next Generation Sequencing Association Studies

- Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, Daly MJ, Neale BM, Sunyaev SR, Lander ES. Searching for missing heritability: designing rare variant association studies. Proc Natl Acad Sci U S A. 2014 Jan 28;111(4):E455-64.
- Cirulli ET1, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010 Jun;11(6):415-25.
- Wu MC, Seunggeun L, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. AJHG 2011 July 15; 89:82-93
- Auer PL, Lettre G. Rare variant association studies: considerations, challenges and opportunities. Genome Med 2015 Feb 23;7(1):16.

Precision Medicine

- Collins FS, Varmus H. A New Initiative on Precision Medicine. N Engl J Med. 2015 372:793-795
- The Precision Medicine Initiative Working Group Report to the Advisor Committee to the Director, NIH https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf