# Approaches for Investigating Gene-Environment Interactions



Assistant Professor
Institute for Computational Biology
Department of Epidemiology and Biostatistics
Case Western Reserve University

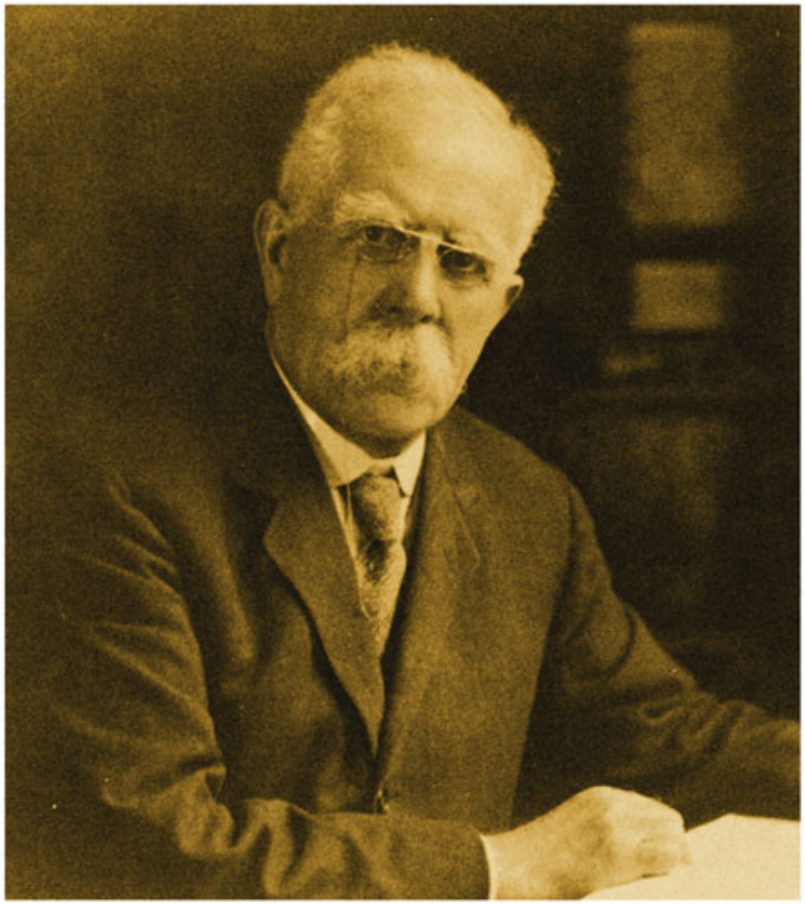
# Acknowledgements


- A BIG thanks to Dr. Dana Crawford for providing many examples and slides for this presentation


# What is an Interaction?

- Effect of one factor depends on the level of another factor (effect modification)
- A Biologist's View
- A Statistician's View

# Gene Environment Interaction



Sir Archibald Garrod

**Nature Reviews | Genetics**

Alkaptonuria (AKU):

- Inborn error of metabolism
- Autosomal recessive
- Urine is dark, depending on diet

variance with all that is known of the origin of species. Nor are direct evidences wanting of such minor chemical diversities as we have supposed to exist within the species. Such slight peculiarities of metabolism will necessarily be hard to trace by methods of direct analysis and will readily be masked by the influences of diet and of disease, but the results of observations on metabolism reveal differences which are apparently independent of such causes, as for example, in the excretion of uric acid by different human individuals. The phenomena of obesity and the

Garrod *Lancet* 2:1616-1620 (1902)

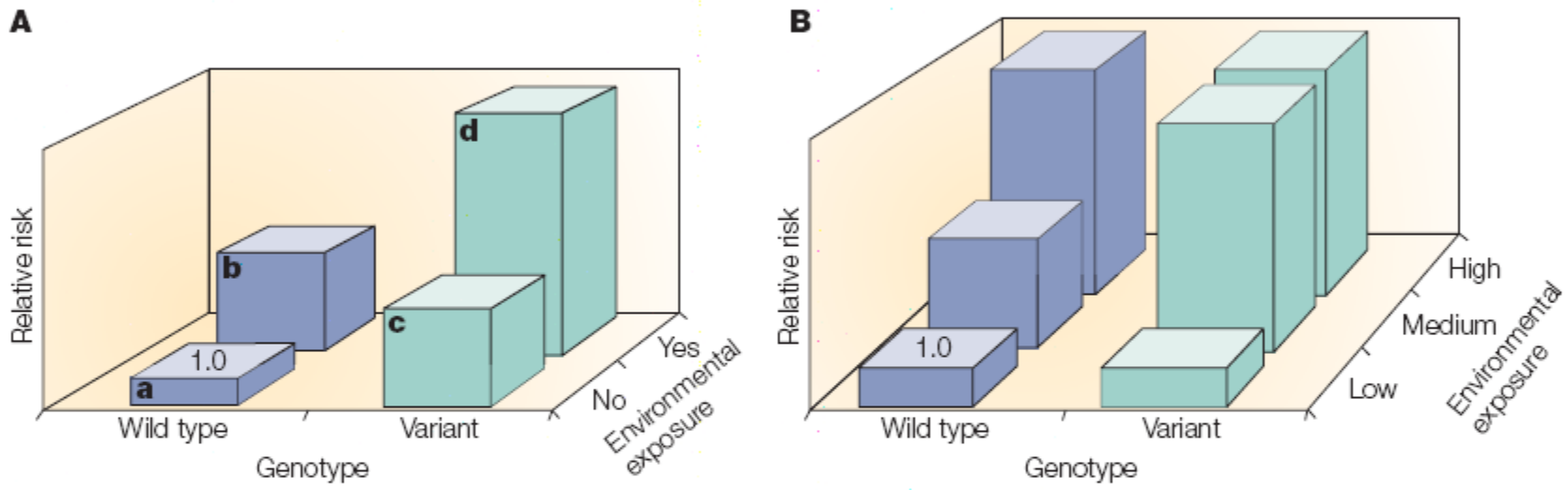
# Gene Environment Interaction

| Gene variant | Environmental exposure | Relative risk (XP) | Relative risk (PKU) | Relative risk (emphysema) |
|---|---|---|---|---|
| Absent | Absent | 1.0 | 1.0 | 1.0 |
| Present | Absent | ~1.0 | 1.0 | Modest |
| Absent | Present | Modest | 1.0 | Modest |
| Present | Present | Very high | Very high | High |

- Phenylketonuria (PKU): Need BOTH risk genotype and exposure
- Xeroderma pigmentosum (XP): Can develop skin cancer with exposure, but risk is much greater with risk genotype AND exposure
- Emphysema: Can develop disease with exposure (smoking) or risk genotype ($\alpha$-antitrypsin), but risk is much greater with BOTH

# Gene Environment Interactions

- How do we identify gene-environment interactions?
- Need Exposures AND Genotypes



# Gene Environment Interactions

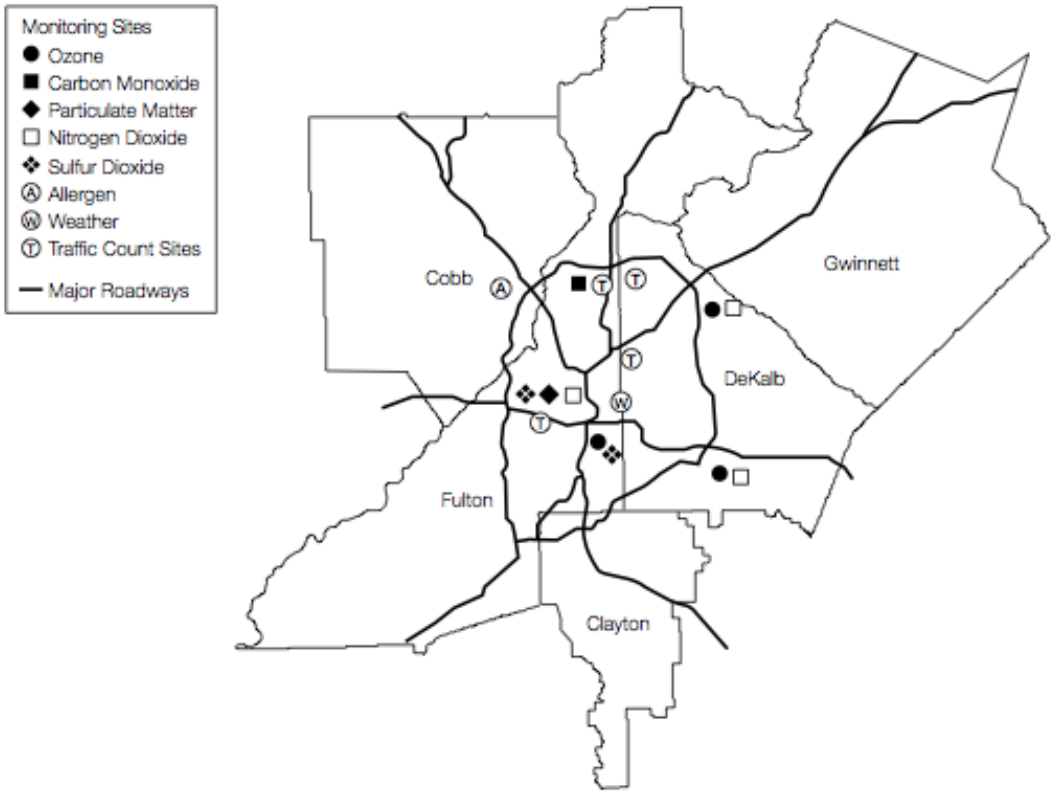
Common "environmental" exposures:

- Smoking/tobacco use (ETS)
- Diet
- Medication use
- Occupation
- Household environment
- Physical activity

# Gene Environment Interaction

- Regional exposures
- Geographic Information Systems (GIS)

**Figure 1.** Functioning Traffic Count and Air Quality Monitoring Sites in Atlanta, Ga, During the 1996 Summer Olympic Games



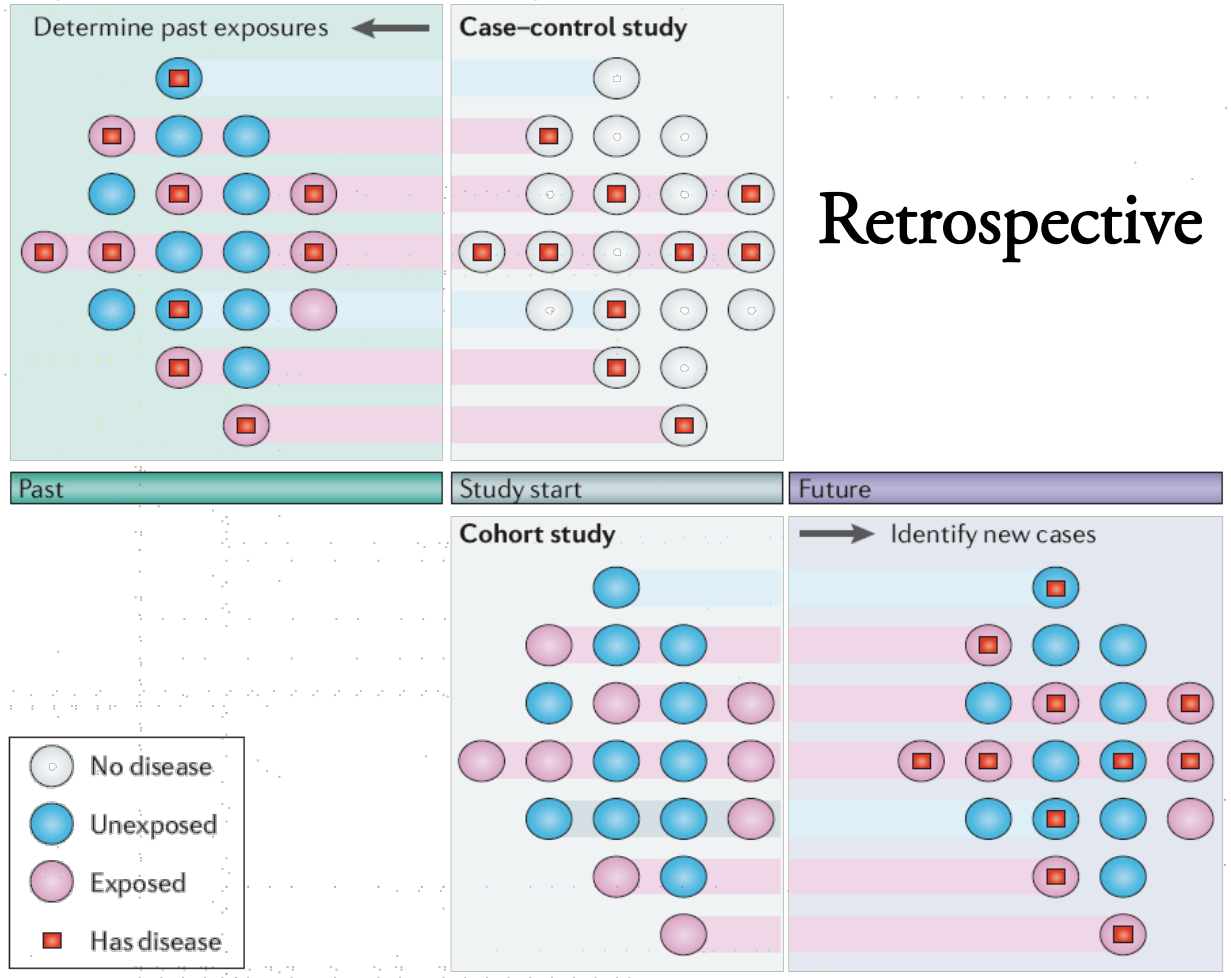

Borders indicate the 5 central counties of metropolitan Atlanta included in the study.

Friedman et al *JAMA* 285(7):897-905

# Collecting Exposures

- Ask participant about past exposures (surveys)
- Ask the participant's physician or pharmacist about past exposures (EMR, medication lists)
- Measure the exposure in blood, urine
- Ask participant to take medication, to exercise, etc

# Collecting Exposures




Retrospective

Prospective

Manolio, Bailey-Wlson, Collins *Nat Genet Review* 7:812-820 (2006)

# Retrospective Studies

- In a retrospective study, mothers who have infants with birth defects (cases) may "over recall" past exposures compared with mothers who have infants with no birth defects (controls)

Werler et al *Am J Epidemiol* 129:415-421 (1989)

# Bias

- All studies are subject to some kind of bias
- All studies may have inaccurate exposure assessment
- Examples:
  - Height and weight (BMI)
  - Energy intake
  - Family history
  - Smoking status

# Bias in Self-report Height

- Self-report vs measured (from NHANES III)
- Older individuals under report actual height
- Other studies suggest weight discrepancy is associated with gender, age, race/ethnicity, and BMI (Villanueva et al 2001)

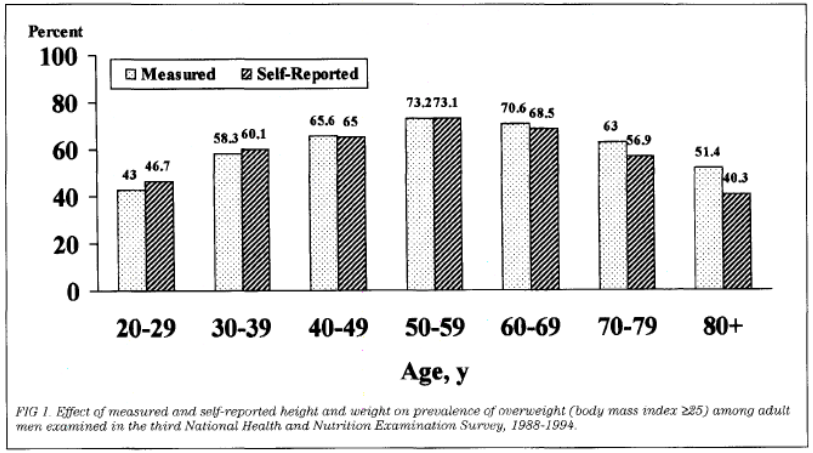



FIG 1. Effect of measured and self-reported height and weight on prevalence of overweight (body mass index ≥25) among adult men examined in the third National Health and Nutrition Examination Survey, 1988-1994.

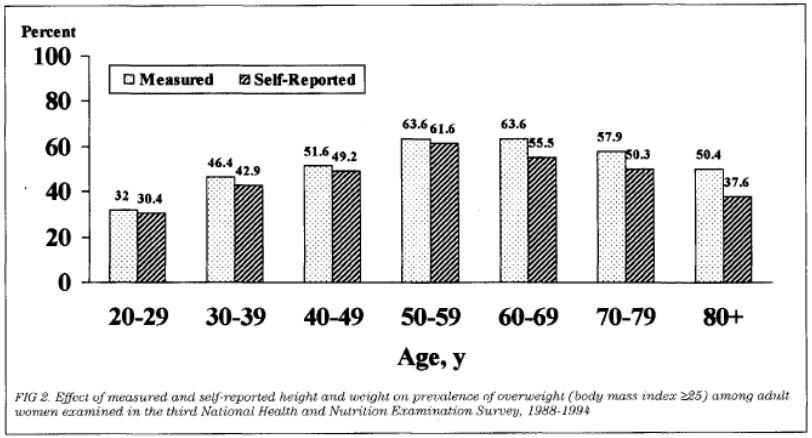



FIG 2. Effect of measured and self-reported height and weight on prevalence of overweight (body mass index ≥25) among adult women examined in the third National Health and Nutrition Examination Survey, 1988-1994

Kuczmarski, Kuczmarski, Najjar *J Am Diet Assoc* 101:28-34 (2001)

# Under-reporting Energy Intake

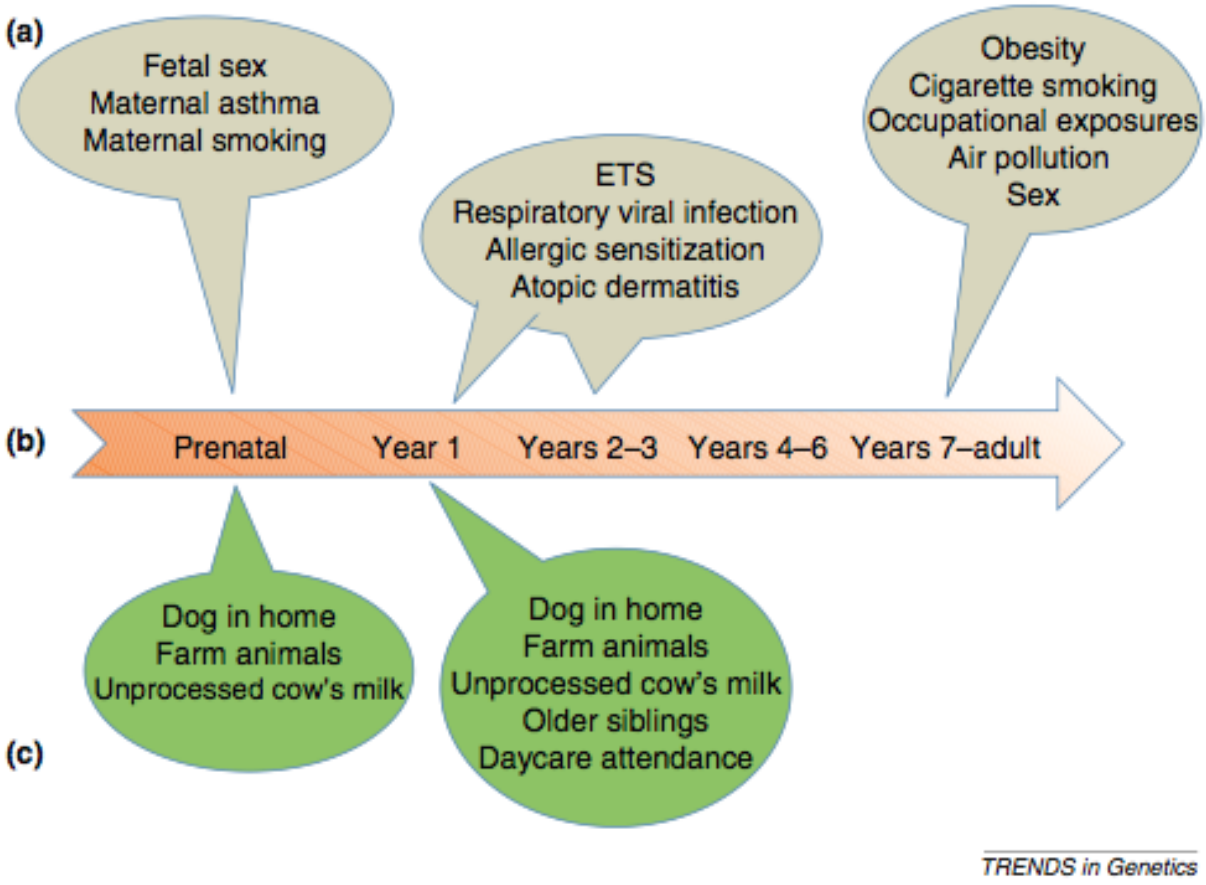
**TABLE 3**
Characteristics of male and female underreporters and adequate reporters of intake in the third National Health and Nutrition Examination Survey, 1988–1991

| | Men | | Women | |
|---|---|---|---|---|
| | Underreporters | Adequate reporters | Underreporters | Adequate reporters |
| *n* (%) | 874 (18.1) | 3082 (81.9) | 1189 (27.7) | 2624 (72.3) |
| EI:$BMR_{est}$[1] | 0.71 ± 0.01[2] | 1.64 ± 0.02[3] | 0.69 ± 0.01 | 1.48 ± 0.01[3] |
| Age (y) | 48.0 ± 1.0 | 42.7 ± 0.5[3] | 48.7 ± 1.0 | 45.0 ± 0.5[3] |
| Education (%) | | | | |
| 0–8 y | 20.4 ± 1.7 | 9.2 ± 0.7[3] | 15.1 ± 1.5 | 9.6 ± 1.0[3] |
| 9–11 y | 15.7 ± 1.6 | 14.4 ± 1.1 | 16.8 ± 1.8 | 13.3 ± 1.1 |
| 12 y | 28.7 ± 2.2 | 33.2 ± 1.6 | 34.6 ± 2.1 | 38.5 ± 1.1 |
| ≥ 13 y | 35.2 ± 2.4 | 43.3 ± 2.2[4] | 33.5 ± 2.0 | 38.6 ± 1.9 |
| Below poverty (%)[5] | 16.0 ± 1.6 | 9.4 ± 0.8[3] | 18.2 ± 1.9 | 11.9 ± 0.9[3] |
| Race-ethnic group (%) | | | | |
| Non-Hispanic white | 74.8 ± 2.7 | 79.9 ± 2.4 | 77.2 ± 2.8 | 80.1 ± 2.3 |
| Non-Hispanic black | 15.0 ± 1.9 | 9.0 ± 1.1[3] | 14.2 ± 1.8 | 10.0 ± 1.2 |
| Mexican American | 4.8 ± 0.5 | 4.7 ± 0.5 | 4.1 ± 0.5 | 3.8 ± 0.4 |

[1] Ratio of energy intake to estimated basal metabolic rate.
[2] $\bar{x}$ ± SEM.
[3,4] Significantly different from underreporters: [3] $P < 0.01$, [4] $P < 0.05$.
[5] Defined as a poverty income ratio < 1.00.

Briefel et al *Am J Clin Nutr* 65(suppl):1203S-1209S (2007)

# Timing of Exposure is Important

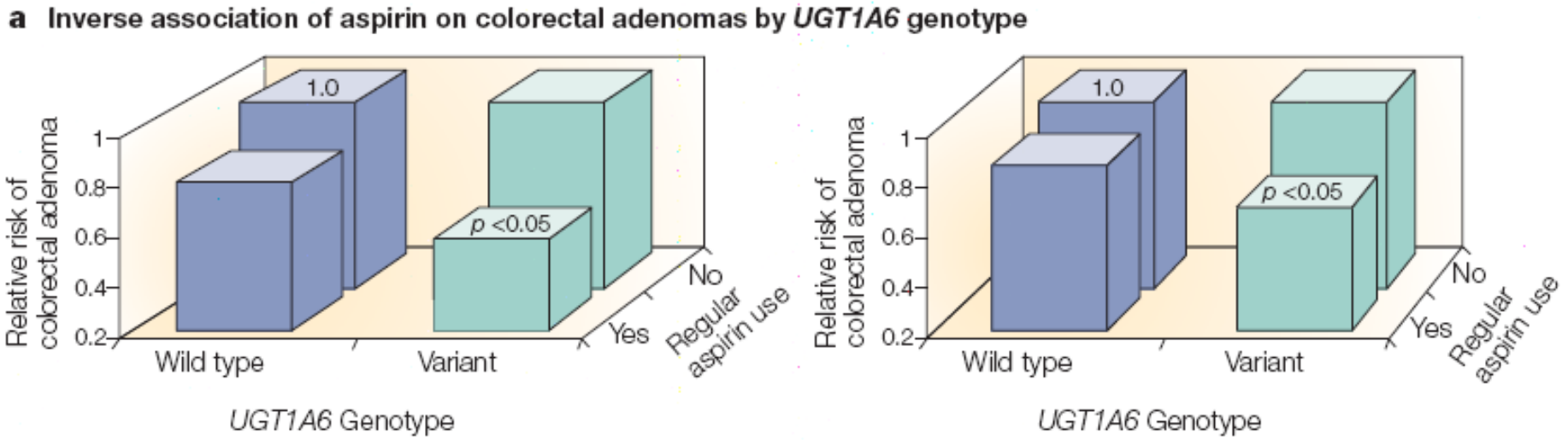



Ober and Vercelli *Trends in Genetics* (2011)

# Detection Strategies

- Stratify by Exposure and Genotype
- Include an Interaction Term in a Statistical Model
- Analysis of Covariance
- Mixed Model Approaches

# Stratified Analyses




Example of "pharmacogenetics"

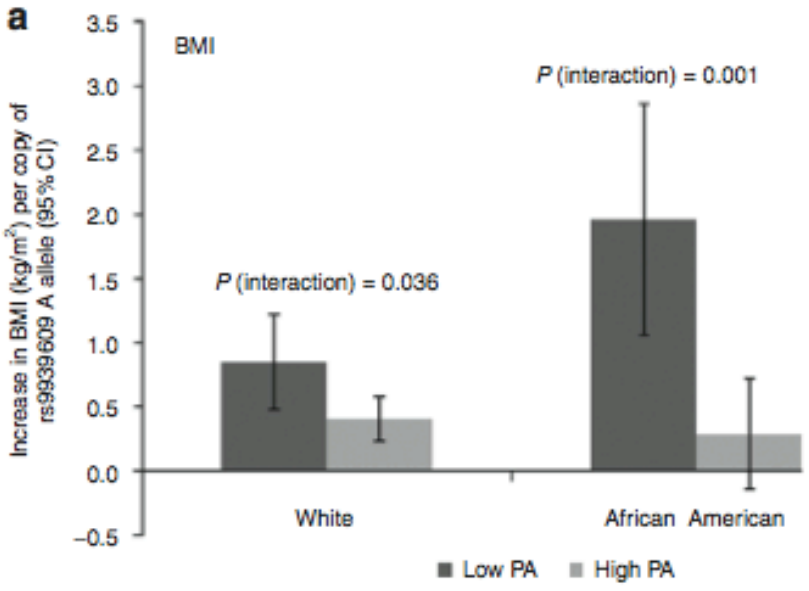
Hunter *Nat Genet Review* 6:287-298 (2005)

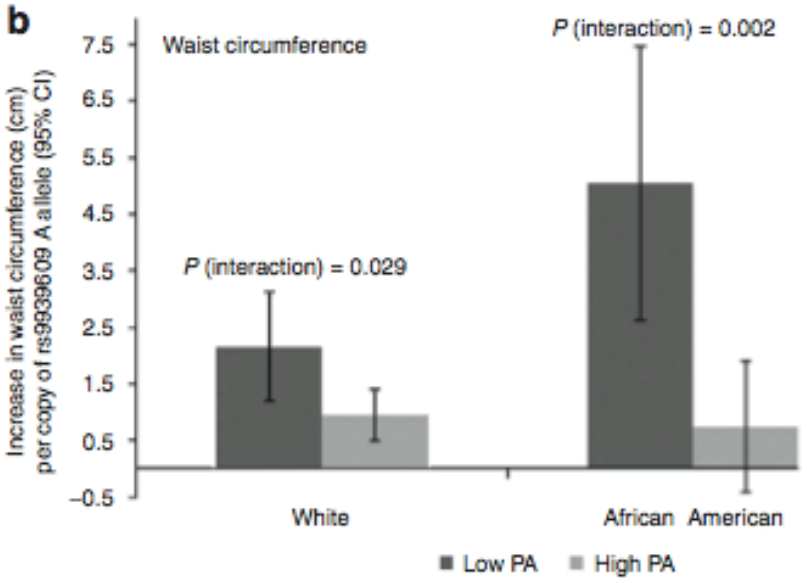
# Replication can be Difficult

| | | | | | | |
|---|---|---|---|---|---|---|
| *UGT1A6* | | | | | | |
| Thr181Ala+ Arg184Ser | Adenomatous polyps | 0.97 [0.74–1.26] | Het/hzv vs wt | | Risk reduction with ASA use stronger among those with any variant allele | Bigler 2001[61] |
| | Adenomatous polyps | 0.74 [0.39–1.41]<br>0.41 [0.24–0.71] | Wt user (>7 pills/ week) vs wt nonuser variant allele user (>7 pills/week) vs variant allele nonuser | $P=0.02$ | Risk reduction with ASA use stronger among those with variant allele | Chan 2005[62] |
| | Adenomatous polyp recurrence | 0.68 [0.52–0.89] | Het/hzv vs wt | $P=0.70$ | No interaction between ASA use and *UGT1A6* variant alleles for polyp recurrence | Hubner 2006[63] |
| | Cancer | 1.08 [0.94–1.24]<br>0.94 [0.76–1.15] | Het/hzv vs wt (colon cancer)<br>Het/hzv vs wt (rectal cancer) | $P=0.39$ (ibuprofen)<br>$P=0.40$ (ASA) | No interaction between *UGT1A6**ASA/ ibuprofen use and adenoma risk. | Samowitz 2006[49] |

Cross, Poole, Ulrich *Pharmacogenomics J* 8:237-47(2008)

# FTO and Physical Activity



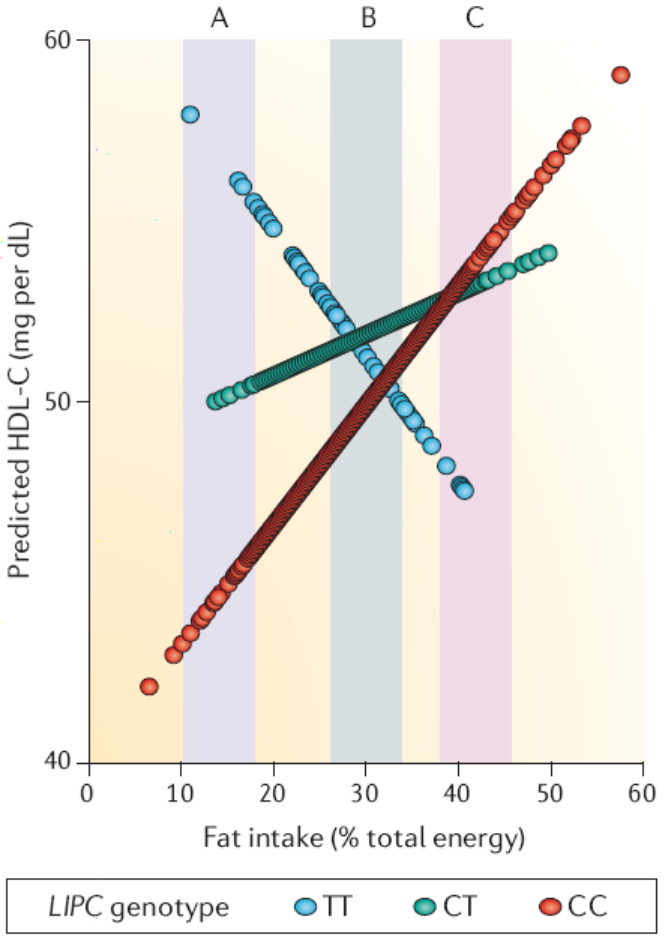

Demerath et al *Obesity (Silver Spring)* 19(9):1866-1872 (2011)

# Analysis of Covariance

- Similar to ANOVA
- Examines relationships between two continuous variables across continuous categories
- Null hypothesis = slopes are equal across all categories

# Analysis of Covariance

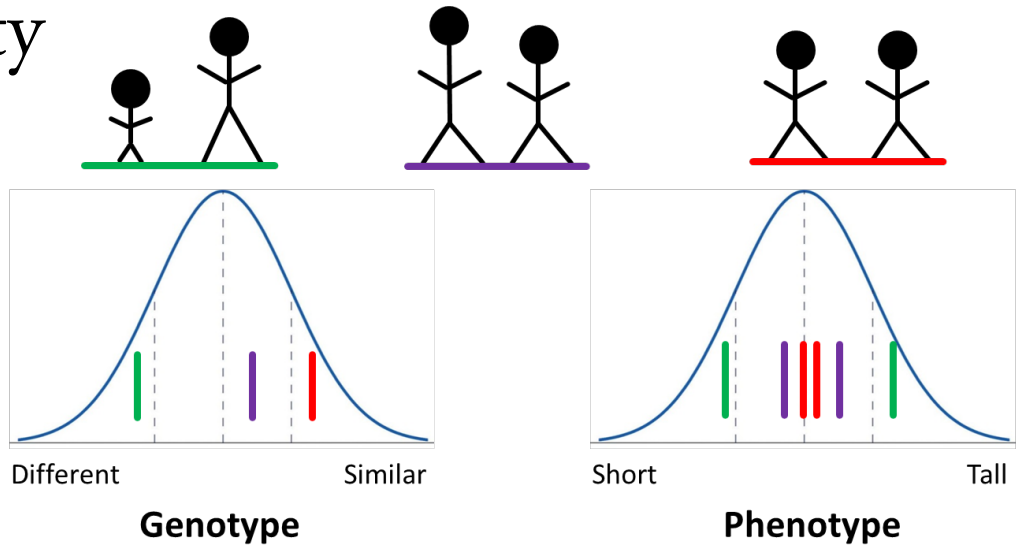
- Band A:
  Low fat intake and TT genotype
  Associated with high HDL
- Band B:
  Moderate fat intake and TT
  No association with HDL
- Band C:
  High fat intake and TT genotype
  Associated with low HDL




Manolio, Bailey-Wilson, Collins *Nat Genet Review* 7:812-820 (2006)

# Mixed Model Analysis

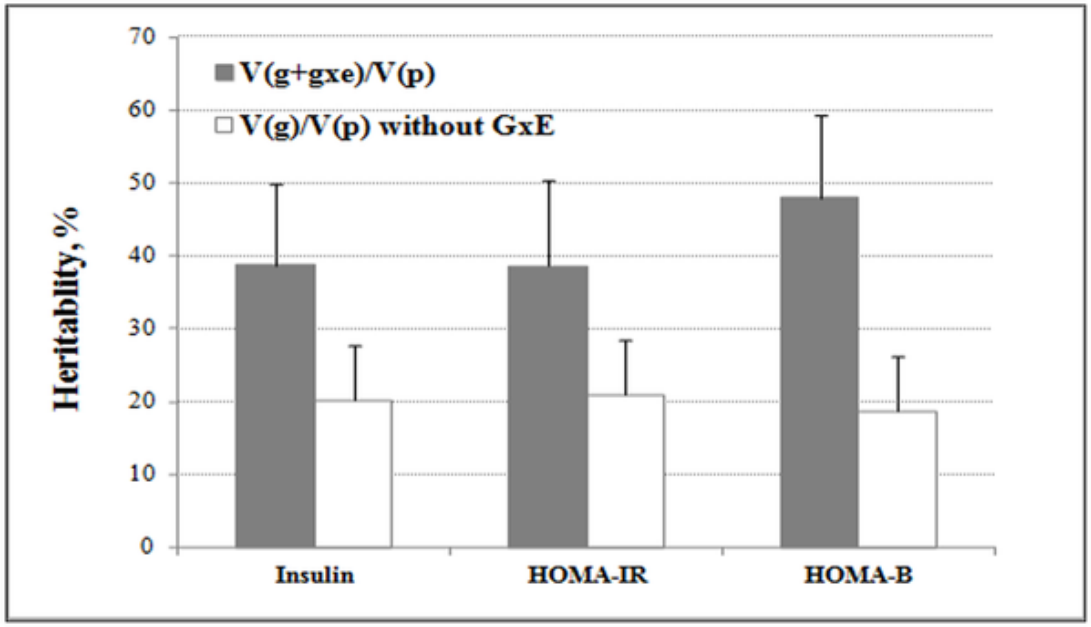
- Allows estimation of the Proportion of Trait Variance Explained (PVE) using GWAS data
- Relates genetic similarity to phenotypic similarity




Yang et al. Am J Hum Genet 88(1):76-82

# Mixed Model Analysis

- Mixed Models can include interaction terms to test for global genetic X environmental factor interaction




Zheng et al. *PLoS ONE* 8:10(2013)

# Difficulties with these Strategies

- Requires good exposure information
- Can be subject to bias (study design)
- Need large sample sizes
- Limited methods to identify all higher-order interactions

# Conclusions

- Gene Environment Interactions are likely an important part of the genetic architecture of human disease
- Systematically collecting environmental exposures in a manner to allows their identification is the primary challenge