Methods in Genetic Epidemiology

Lindon Eaves
VIPBG
March 2006

Link to basic genetics

http://psych-www.colorado.edu/
hgss/hgssfigures/hgssfigures.htm#Chapter10
“Epidemiology”

Identification of “causes from clusters”

Clusters and Causes

- Sources of water (cholera)
- Insect vectors (malaria)
- Working in mines (silicosis)
- Toxic waste (some cancers)
- Hypertension (ethnicity)

- Etc.
THE “FAMILY”

A primary cluster

Large pedigree from the isolate with a founder couple born in approximately 1650. Circles represent females; squares represent males. Blackened circles and squares represent individuals with at least two children affected with schizophrenia.

Published online September 3, 1999.
Copyright © 1999 by The American Society of Human Genetics. All rights reserved.
Insert Kuru pedigree here!

Draw your own pedigree

http://www.genesoc.com/nutrition/pedigree.htm
Familial disorders

- Familial hypercholesterolemia
- Bovine spongiform encephalopathy
- Schizophrenia
- Kuru
- Colon cancer
- Mental retardation
- Breast cancer
- Etc. etc. etc.

Family clustering may be genetic or environmental
“Genetic Epidemiology”

Systematic use of familial clustering to resolve and identify genetic and environmental causes of disease

Questions in GE

- Is “it” familial?
- Is it “genetic” (G), “environmental” (E) or both (G+E)?
- What are relative contributions of G and E?
- What kinds of genetic effects are there (one, few, many, additive/epistatic)?
- Effects of population structure, ethnicity, stratification, mate selection?
- What is role of E? Which? GxE interaction?
- Which specific genes are involved?
Measured Genotypes | Measured Environments

- $G_1$, $G_2$, $G_3$, $G_4$
- $E_1$, $E_2$, $E_3$, $E_4$

Endophenotypes

- $G_1'$, $G_2'$, $G_3'$, $E_1'$, $E_2'$, $E_3'$, $E_4'$

Outcome Phenotype

- $P_1$, $P_2$, $P_3$, $P_4$, $P_5$

Sequential ("complementary") genes

- A
- B
- C

Parallel ("duplicate") genes

- A
- B

= Pathway blocked by mutant gene
Answering the questions: some approaches

Is it familial?

- Risk to relatives (schizophrenia, colorectal cancer, mental retardation, hypertension)
- Correlation between relatives (IQ, diastolic blood pressure)
- Same trait may be category or dimension (hypertension = DBP>90mmHg)
Causes of Family Resemblance

Path diagram for the effects of genes and environment on phenotype

- Genotype (G)
- Environment (E)
- Measured variable (P)
- Latent variables
  - h
  - e
- Phenotype
- r

h  e
"r"  "r"
Distribution of stature corrected for age and sex
(Inches from mean)

Family Resemblance

Mother Father

Em Gm Gf Ef

Ec Gc

M F C

r e h ½ m h e ½ f f

-14 -10 -6 -2 2 6 10 14

0.16
0.12
0.08
0.04
0.00
0.04
0.08
0.12
0.16
-14 -10 -6 -2 2 6 10 14

IHT
Polygenic Inheritance (Fisher, 1918)
Estimating G and E (1)

Twin Studies (Galton 1865)
Stature in Twins

Scatterplot for corrected MZ stature

Scatterplot for age and sex corrected stature in DZ twins

Path diagram for twin resemblance
Four scenarios

Causes of Variation

Twin Correlation

No G  No C  G and C  G and I

Estimating G and E (2)

Adoption studies
Other approaches

- Separated Twins (e.g. Shields, 1966)
- Twins and parents
- Children of Twins
- Extended twin kinships
- Combinations of Methods
Identifying environments

Measure environment in family, twin or adoption study (e.g. twins discordant for exposure, characteristics of foster parents etc.)

Finding the genes

• Linkage studies – extended pedigrees, sib pairs
• Association studies (case-control, family-based – e.g. TDT, sib-pair)
“Linkage”

Genes that start together stay together – the closer they are, the more they stay together

Linkage

Are relatives who are more alike for a “marker” more alike in their phenotypes?
**Sib-pair similarity and linkage**

\[
r = R + \pi Q
\]
Linkage – plusses and minuses

- **PLUSES**
  - Depends only on marker location, not effect of marker on phenotype
  - Doesn’t require many markers to cover genome (100s)
  - Hard to “invent” linkage - robust

- **MINUSES**
  - Works best for simple (few-gene) traits
  - Gene effects need to be big
  - Specific localization poor (i.e. many genes under linkage “peak”)

Association

Do different forms of gene (“alleles”) have different phenotypes?
Association: Pluses and Minuses

**PLUSES**
- Statistical power
- “Tight” localization
- Can use “candidate genes”

**MINUSES**
- Association may not be causal (e.g. “linkage disequilibrium”, population stratification) – but can control/eliminate
- Needs large number of markers for genome-wide study (?500,000+)

![Diagram of Genome-Phenotype Marker-Phenotype Association](image)
Controlling for stratification

- Analyze within strata (ethnicity, SES etc)
- Use random genes to test, characterize and eliminate
- Family-based association – TDT, sib-pairs etc.

Complications – can’t always do genetics without environment

- GxE Interaction
- G-E correlation
- G x Age interaction
GxE Interaction

- Genes control sensitivity to the environment (some environments only affect particular genotypes)
- Environments modulate expression of genes (some genes only expressed in particular environments)
Analyzing GxE

- Family resemblance depends on environmental exposure
- Effect of gene contingent on environment (or vice-versa)

Genetic Variance and Shared Life Events in Adolescent Females (Silberg et al., 1999)
G-E Correlation:
Environmental exposure caused/influenced by genes

- “Active/Evocative” - environment depends directly on genes of individual (e.g. own smoking)
- “Passive” – environment depends on genes of relatives (e.g. parental smoking)
Analyzing rGE

- Include environmental measures in twin, adoption and family studies – build and test path models.

G x Age Interaction

- Genetic control of age of onset
- Different genes expressed at different ages
- Rates of growth/change depend on genes
  - A bit like GxE in some ways.