Genetics of Complex Disease: Heritability

Sorin Istrail

Department of Computer Science
Brown University, Providence
sorin@cs.brown.edu

November 2, 2010

Outline

Genetic characters: mendelian and polygenic

- Mendelian conditions are due to single-gene effects subject to simple loss of function or gain of function mutations.
- Polygenic conditions depend on the combined action of a very large number of genetic factors, each of which makes only a very small contribution to the final phenotype.
- Both categories, mendelian and polygenic, are conceptual models, useful tools for thinking about inheritance – in reality, all conditions fall somewhere between the two.
- For any particular character, you need to decide what mix of the two conceptual models gives the best description, they you need to add environmental effects.
- The overall etiology of any disease might be represented by a point somewhere inside the triangle with three corners: single-gene, polygenic, environmental.

Multifactorial or complex

- "Multifactorial" or "complex" are useful but non committal terms for conditions that fall in the interior of the triangle.
- Genes are always mendelian, but characters are never wholly dependent on the genotype at a single locus. In depth analysis of Mendelian disease show that these "simple" conditions are not so simple.
- The analysis of the genotype-phenotype correlation seldom yields desired clear cut results.
- The long and winding road between a DNA sequence variant and an observable phenotype allows many additional players, environmental, genetic and stochastic chance, to join the action.

Penetrance

- The concept of penetrance allows you to lump all these other factors together and not try to analyze them.
- Modifier genes are genes that influence the expression of a basically mendelian character.
- One can divide human characters into two types:
  - discontinuous or dichotomous characters, like diseases and malformations, that you either have or not have, and
  - continuous or quantitative characters, that everybody has, but to different degrees; e.g., height, weight, blood pressure. Continuous characters cannot be mendelian.

QTL

- Gene loci involved in determining continuous characters are called quantative trait loci (QTL)
- QTL has been mainly developed by animal and plant geneticists concerned with improving continuous characters like milk yields or corn production.
- QTL variants are likely to have more effects on gene function.
-
  - Variants in promoters or other control elements could result in small changes in the level of expression
  - Sequence changes might slightly alter the stability of the mRNA or the balance of splice isoforms
  - The same gene could cause a mendelian condition if it has a mutation of major effect, and act as a QTL with more subtle changes.
  - A genetic change may have a major effect on one character and be a QTL for another.

Polygenic Theory

- Models of polygenic inheritance consider characters that depend on the combined action of a large number of genetic factors, each of which makes only a small contribution to the final phenotype
- The main use of these models in human genetics is not to make predictions, but to provide a framework for thinking about more complex genetic conditions. For any actual condition the framework needs to be populated by empirical empiric epidemiological data.
- The two concepts of the framework that are most relevant to clinical genetics are:
  - Heritability, and
  - Threshold

Heritability

- Heritability is estimated by comparing the incidence of a character in relatives of affected people with the incidence in the general population.
- It is denoted by h^2 reflecting its origin in polygenic theory as a correlation coefficient.
- Heritability is a useful measure for researchers investigating a complex condition because it gives a guide to the relevance of genetic factors.
- The heritability of a condition is not a measure of how the condition is caused physiologically or biochemically in a person. In fact, it is not a statement about individuals at all, but about a population.
- For continuous characters it is calculated from the correlation between relatives.
- It is a number between 0 (= no genetic involvement), ad 1 (= no non-genetic factors present)

Heritability

- Heritability is the proportion of the phenotypic variance that is attributable to genetic variance.
- It answers the following question: How much of the difference between people, in this particular society and at this particular time, are the result of the genetic difference between them, and how much are they the result of difference in people's environments?
- Heritability is not a fixed property of a condition. If occurence of a condition is affected by both genetic and environmental factors, its heritability will be high if most members of the population share a relatively uniform exposure to the relevant environmental factors, and low if exposure differs greatly between people.
- A part-genetic condition that is also associated with poverty and social deprivation will have higher heritability in an egalitarian society than in one with great inequalities.

Heritability - an example

- 100 years ago phenylketonuric mental retardation was wholly genetically determined.
- Nowadays, in societies where neonatal screening and dietary treatment are available, its causes are almost entirely environmental - lost samples, a family that couldn't be traced or that are unable to make sure a homozygous child sticks rigidly to the diet.
- This example underlies the fact that heritability is a property of a character at a particular place and time, and not an intrinsic property of a gene.

Threshold

- The second category of concepts of basic importance to clinical genetics are the Thresholds. They are used to explain how discrete or dichotomous characters can be polygenic.
- In its simplest form, polygenic theory applies only to continuously quantitative characters that show a Gaussian distribution in a population.
- In clinical genetics, thinking about thresholds helps make sense of the way risks for complex conditions, unlike risks for mendelian conditions, depend on family history.

D.S. Falconer's Thresholds and Polygenic Susceptibility to Disease

- Although diseases are dichotomous characters - some people have the disease and some do not - threshold theory postulates that susceptibility to the disease is a continuous character. The susceptibility is a quantitative character that depends on the combined small effects of many genes.
- It shows a bell-shaped distribution in the population.
- Most people have a middling susceptibility;
- A few people have very high or very low susceptibility.
- Only people whose susceptibility exceeds a certain threshold develop the disease.
- A person having an affected relative must have a high genetic susceptibility. They got a set of the high-susceptibility genes. Because you share genes with relatives, you may be lucky to share only a few of your relative high susceptibility genes, or you may be unlucky and share many.

D.S. Falconer's Thresholds and Polygenic Susceptibility to Disease

- Overall, there will be a wide distribution of susceptibility among relatives of affected people, but the curve will be displaced towards the high-susceptibility end compared to the distribution in the general population.
- Relatives of affected people are more likely than unrelated people to end up above the threshold. The closer the relationship, the more likely this is.
- Therefore complex diseases tend to run in families, but the tendency is much weaker than with mendelian diseases.

Further extension of the Thresholds model

- The idea can be extended in several ways
  - To allow for environmental factors, we could suppose that, rather than a single fixed threshold, there will be an increasing risk of developing the disease across some range of high susceptibilities
  - Thresholds can be sex-specific. The classic example is pyloric stenosis. This is much more common in boys than in girls (5M:1F). On average, affected girls have higher susceptibility genes than affected boys. Therefore, relatives of affected girls are at higher risk of pyloric stenosis than relatives of affected boys.

The General Rule

- Although these models can be developed mathematically, they are mainly useful as qualitative tools for thinking about recurrent risks.
- **The General Rule**:*In complex disease, the worse luck (many affected people; the less common sex affected) a family have had in the past, the greater the risk of a recurrence.*

The Heritability h^2

- Galton invented the concept of *regression* which is key for what we need.
- If we plot a quantitative trait such as the heights of pairs of fathers (the x values) and sons (the y values) as points in the 2-d plane, we have a *scattergram*.
- The magnitude of the scatter for either variable, is measured by variance, $\sigma_x^2 = E[x - E[x]]^2$ where E stands for expectation or mean or average.
- the best-fit of a line for the scatter is the *regression line*.

The Heritability h^2

- The slope of the regression line is expressed in terms of the *covariance* of x and y defined as $\sigma_{xy} = E(xy) - E(x)E(y)$.
- An estimate of the covariance is $Cov(x, y) = \bar{xy} - \bar{x}\bar{y}$
- The variance of x is given by $Var(x) = \bar{x^2} - \bar{x}^2$
- The slope of the of the regression line is called the *regression coefficient* $\hat{b} = \frac{Cov(x,y)}{Var(x)}$
- The correlation coefficient is given by $r = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$
- Assuming that the environmental covariance is zero, the regression coefficient b of offspring on one parent can be calculated for a random mating population, and it indicates the degree to which the phenotypic variation of the trait is determined by genetic variation.
- We then have $b = \frac{1}{2}h^2$