Polygenic and multifactorial inheritance

Dr S M H Ghaderian, MD, PhD
Associate Professor of Medical Genetics
8 floor, Medical Genetics Department
School of Medicine
Shahid Beheshti University of Medical Sciences
Introduction

• Many disorders demonstrate familial clustering that does not conform to any recognized pattern of Mendelian inheritance.

• As it is likely that many factors, both genetic and environmental, are involved in causing these disorders, they are generally referred to as showing *multifactorial* inheritance.

• According to this theory, individuals are affected if they lie at the wrong end of the distribution curve.

• This concept of a normal distribution generated by many genes, known as *polygenes*, each acting in an additive fashion, is plausible for physiological characteristics such as height and possibly blood pressure.
• Sequencing of the human genome has shown that the 3 billion base pairs are 99.9% identical in every person.

• This also means that individuals are, on average, 0.1% different genetically from every other person on the planet.

• And within this 0.1% lies the mystery of why some people are more susceptible to a particular illness, or more likely to be healthy, than other members of the population.
Disorders which show multifactorial inheritance

**Congenital malformations**
- Cleft lip/palate
- Congenital dislocation of the hip
- Congenital heart defects
- Neural tube defects
- Pyloric stenosis
- Talipes

**Acquired diseases of childhood and adult life**
- Asthma
- Autism
- Diabetes mellitus
- Epilepsy
- Glaucoma
- Hypertension
- Inflammatory bowel disease (Crohn disease and ulcerative colitis)
- Ischaemic heart disease
- Ischaemic stroke
- Manic depression
- Multiple sclerosis
- Parkinson disease
- Psoriasis
- Rheumatoid arthritis
- Schizophrenia
POLYGENIC INHERITANCE AND THE NORMAL DISTRIBUTION

• Before considering the impact of recent research in detail, it is necessary to outline briefly the scientific basis of what is known as polygenic or quantitative inheritance.

• 'Additive' implies that the effects of the genes are cumulative, i.e. no one gene is dominant or recessive to another.

• Several human characteristics show a continuous distribution in the general population that closely resembles a normal distribution.

• This takes the form of a symmetrical bell-shaped curve distributed evenly about a mean.
• The spread of the distribution about the mean is determined by the standard deviation.

• Approximately 68%, 95% and 99.7% of observations fall within the mean plus or minus one, two or three standard deviations, respectively.

• For example, hypertension (high blood pressure) and diabetes affect roughly 23% and 5% of the American population, respectively.

• Epilepsy, schizophrenia, and bipolar disorder, all of which are brain disorders, each affect approximately 1% of the population.

• Multiple sclerosis and autism have frequencies of about 0.1%.
The normal (gaussian) distribution

- Mean
- ± 1SD (≈68%)
- ± 2SD (≈95%)
- ± 3SD (≈99.7%)

© Elsevier: Turpenny & Ellard: Emery’s Elements of Medical Genetics 13e - www.studentconsult.com
A, The distribution of height in a population, assuming that height is controlled by a single locus with genotypes AA, Aa, and aa.

B, The distribution of height, assuming that height is controlled by two loci. There are now five distinct phenotypes instead of three, and the distribution begins to look more like the normal distribution.

C, Distribution of height, assuming that multiple factors, each with a small effect, contribute to the trait (the multifactorial model).
Further support for this concept comes from the study of familial correlations for characteristics such as height and, to a lesser extent, intelligence.

**Correlation** is a statistical measure of the degree of association of variable phenomena, or, in more simple terms, a measure of the degree of resemblance or relationship between two parameters.

As first-degree relatives share, on average, 50% of their genes, it would be reasonable to predict that, if a parameter such as height were polygenic, the correlation between first-degree relatives such as siblings would be 0.5.

Several studies have shown that the sib-sib correlation for height is indeed close to 0.5.

In reality, human characteristics such as height and intelligence are also influenced by environment, and possibly also by genes that are not additive in that they exert a dominant effect.
These factors probably account for the observed tendency of offspring to show what is known as 'regression to the mean'.

This is demonstrated by tall or intelligent parents (the two are not mutually exclusive!) having children whose average height or intelligence is slightly lower than the average or mid-parental value.

Similarly, parents who are very short or of low intelligence tend to have children whose average height or intelligence is lower than the general population average, but higher than the average value of the parents.

If a trait were to show true polygenic inheritance with no external influences, the measurements in offspring would be distributed evenly around the mean of their parents' values.
## Degrees of relationship

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Proportion of genes shared</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First degree</strong></td>
<td>1/2</td>
</tr>
<tr>
<td>Parents</td>
<td></td>
</tr>
<tr>
<td>Siblings</td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td></td>
</tr>
<tr>
<td><strong>Second degree</strong></td>
<td>1/4</td>
</tr>
<tr>
<td>Uncles and aunts</td>
<td></td>
</tr>
<tr>
<td>Nephews and nieces</td>
<td></td>
</tr>
<tr>
<td>Grandparents</td>
<td></td>
</tr>
<tr>
<td>Grandchildren</td>
<td></td>
</tr>
<tr>
<td>Half-siblings</td>
<td></td>
</tr>
<tr>
<td><strong>Third degree</strong></td>
<td>1/8</td>
</tr>
<tr>
<td>First cousins</td>
<td></td>
</tr>
<tr>
<td>Great-grandparents</td>
<td></td>
</tr>
<tr>
<td>Great-grandchildren</td>
<td></td>
</tr>
</tbody>
</table>
Efforts have been made to extend the polygenic theory for the inheritance of quantitative or continuous traits to try to account for discontinuous multifactorial disorders.

According to the liability/threshold model, all of the factors that influence the development of a multifactorial disorder, whether genetic or environmental, can be considered as a single entity known as liability.
The liabilities of all individuals in a population form a continuous variable, which has a normal distribution in both the general population and relatives of affected individuals.

However, the curves for these relatives will be shifted to the right, and the extent to which they are shifted is directly related to the closeness of their relationship to the affected index case.
Hypothetical liability curves in the general population and in relatives for a hereditary disorder in which the genetic predisposition is multifactorial.
1. CONSEQUENCES OF THE LIABILITY/THRESHOLD MODEL

Part of the attraction of this model - and it should be emphasized once again that this is a *hypothesis* rather than proven fact - is that it provides a simple explanation for the observed patterns of familial risks in conditions such as cleft lip/palate, pyloric stenosis and spina bifida.

**For example:**

1. The incidence of the condition is greatest among relatives of the most severely affected patients, presumably because they are the most extreme deviants along the liability curve.

   - For example, in cleft lip/palate the proportion of affected first-degree relatives (parents, siblings and offspring) is 6% if the index patient has bilateral cleft lip and palate, but only 2% if the index patient has a unilateral cleft lip.
Recurrence Risks (%) for Pyloric Stenosis, Subdivided by Gender of Affected Probands and Relatives*

<table>
<thead>
<tr>
<th>Relatives</th>
<th>Male Probands</th>
<th>Female Probands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>London</td>
<td>Belfast</td>
</tr>
<tr>
<td>Brothers</td>
<td>3.8</td>
<td>9.6</td>
</tr>
<tr>
<td>Sisters</td>
<td>2.7</td>
<td>3.0</td>
</tr>
</tbody>
</table>

* Note that the risks differ somewhat between the two populations.
2. The risk is greatest among close relatives of the index case and decreases rapidly in more distant relatives.

- For example, in spina bifida the risks to first-, second- and third-degree relatives of the index case are approximately 4%, 1% and less than 0.5%, respectively.

3. If there is more than one affected close relative, the risks for other relatives are increased.

- In spina bifida, if one sibling is affected the risk to the next sibling (if folic acid is not taken by the mother periconceptionally) is approximately 4%; if two siblings are affected, the risk to a subsequent sibling is approximately 10%.
4. If the **condition is more common** in individuals of a particular sex, then relatives of an affected individual of the less frequently affected sex will be at higher risk than relatives of an affected individual of the more frequently affected sex.

- This is illustrated by the condition pyloric stenosis.
- Pyloric stenosis shows a male to female ratio of 5 : 1.
- The proportions of affected offspring of male index patients are 5.5% for sons and 2.4% for daughters, whereas the risks to the offspring of female index patients are 19.4% for sons and 7.3% for daughters.
The probable explanation for these different risks is that in order for a female to be affected she has to lie at the extreme of the liability curve, so that her close relatives will also have a very high liability for developing the condition.

As males are more susceptible to developing the disorder, risks in male offspring are higher than in female offspring regardless of the sex of the affected parent.

5. The risk of recurrence for first-degree relatives (i.e. siblings and offspring) approximates to the square root of the general population incidence.

Thus, if the incidence is 1 in 1000, the sibling and offspring risk will equal approximately 1 in 32, or 3%. 
If the prevalence of the disease in a population is $f$ (which varies between zero and one), the risk for offspring and siblings of probands is approximately $\sqrt{f}$.

This does not hold true for single-gene traits, because their recurrence risks are largely independent of population prevalence.

It is not an absolute rule for multifactorial traits either, but many such diseases do tend to conform to this prediction.
Severe (A) and mild (B) forms of cleft lip/palate
Recurrence Risks for First-, Second-, and Third-Degree Relatives of Probands

<table>
<thead>
<tr>
<th>Disease</th>
<th>Prevalence in General Population</th>
<th>Degree of Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First Degree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Second Degree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Third degree</td>
</tr>
<tr>
<td>Cleft lip/palate</td>
<td>0.001</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>Club foot</td>
<td>0.001</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.002</td>
</tr>
<tr>
<td>Congenital hip dislocation</td>
<td>0.002</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.004</td>
</tr>
</tbody>
</table>
The distribution of height, assuming the presence of a major gene (genotypes AA, Aa, and aa) combined with a multifactorial background. The multifactorial background causes variation in height among individuals of each genotype. If the distributions of each of the three genotypes were superimposed, then the overall distribution of height would be approximately normal, as shown by the dotted line.
Nature and Nurture: Disentangling the Effects of Genes and Environment

- Family members share genes and a common environment.
- Family resemblance in traits such as blood pressure therefore reflects both genetic and environmental commonality ("nature" and "nurture," respectively).
- For centuries, people have debated the relative importance of these two types of factors.
- It is a mistake, of course, to view them as mutually exclusive.
- Few traits are influenced only by genes or only by environment.
- Most are influenced by both.
- Determining the relative influence of genetic and environmental factors can lead to a better understanding of disease etiology.
- It can also help in the planning of public health strategies.
- A disease in which hereditary influence is relatively small, such as lung cancer, may be prevented most effectively through emphasis on lifestyle changes (avoidance of tobacco).
- When a disease has a relatively larger hereditary component, as in breast cancer, examination of family history should be emphasized in addition to lifestyle modification.
2. HERITABILITY

- Although it is not possible to assess an individual's liability for a particular disorder, it is possible to estimate what proportion of the etiology can be ascribed to genetic factors as opposed to environmental factors.

- This is referred to as *heritability*, which can be defined as the proportion of the total phenotypic variance of a condition that is caused by additive genetic variance.

- In statistical terms, variance equals the square of the standard deviation.

- Heritability is often depicted using the symbol $h^2$ and is expressed either as a proportion of 1 or as a percentage value.

- Estimates of the heritability of a condition or trait provide an indication of the relative importance of genetic factors in its causation, so that the greater the value for the heritability the greater the role of genetic factors.
This ratio of sib risk to population incidence is known as $\lambda_s$.

For example, in type 1 diabetes, where the UK population incidence is 0.4% and the risk to siblings is 6%, $\lambda_s$ is 15.

For type 2 diabetes in Europe, $\lambda_s$ is estimated at a more modest 3.5 (35% sibling risk; 10% population risk).

Correlations and concordance rates in MZ and DZ twins can be used to measure the heritability of multifactorial traits.

Essentially, heritability is the percentage of population variation in a trait that is due to genes (statistically, it is the proportion of the total variance of a trait that is caused by genes).
Heritability is estimated from the degree of resemblance between relatives expressed in the form of a correlation coefficient, which is calculated using statistics of the normal distribution.

In practice it is desirable to try to derive heritability estimates using different types of relatives, and to measure the disease incidence in relatives reared together and living apart so as to try to disentangle the possible effects of common environmental factors.

The degree of familial clustering shown by a multifactorial disorder can be estimated by measuring the ratio of the risk to siblings of affected individuals compared with the general population incidence.
A simple formula for estimating heritability (\( h^2 \)) from twin correlations or concordance rates is as follows:

\[
h^2 = \frac{c_{MZ} - c_{DZ}}{1 - c_{DZ}}
\]

where \( c_{MZ} \) is the concordance rate (or intraclass correlation) for MZ twins and \( c_{DZ} \) is the concordance rate (or intraclass correlation) for DZ twins.

As this formula illustrates, traits that are largely determined by genes result in a heritability estimate that approaches 1.0 (i.e., \( c_{MZ} \) approaches 1.0, and \( c_{DZ} \) approaches 0.5).

As the difference between MZ and DZ concordance rates becomes smaller, heritability approaches zero.

Correlations and concordance rates in other types of relatives (e.g., between parents and offspring) can also be used to measure heritability.
Concordance Rates in Twins for Selected Traits and Diseases*

<table>
<thead>
<tr>
<th>Trait or Disease</th>
<th>Concordance Rate</th>
<th>DZ Twins</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective disorder (bipolar)</td>
<td>0.79</td>
<td>0.24</td>
<td>&gt;1.0†</td>
</tr>
<tr>
<td>Affective disorder (unipolar)</td>
<td>0.54</td>
<td>0.19</td>
<td>0.70</td>
</tr>
<tr>
<td>Alcoholism</td>
<td>&gt;0.60</td>
<td>&lt;0.30</td>
<td>0.60</td>
</tr>
<tr>
<td>Autism</td>
<td>0.58</td>
<td>0.23</td>
<td>0.70</td>
</tr>
<tr>
<td>Blood pressure (diastolic)†</td>
<td>0.58</td>
<td>0.27</td>
<td>0.62</td>
</tr>
<tr>
<td>Blood pressure (systolic)†</td>
<td>0.55</td>
<td>0.25</td>
<td>0.60</td>
</tr>
<tr>
<td>Body fat percentage†</td>
<td>0.73</td>
<td>0.22</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>Body mass index†</td>
<td>0.95</td>
<td>0.53</td>
<td>0.84</td>
</tr>
<tr>
<td>Cleft lip/palate</td>
<td>0.38</td>
<td>0.08</td>
<td>0.60</td>
</tr>
<tr>
<td>Club foot</td>
<td>0.32</td>
<td>0.03</td>
<td>0.58</td>
</tr>
<tr>
<td>Dermatoglyphics (finger ridge count)†</td>
<td>0.95</td>
<td>0.49</td>
<td>0.92</td>
</tr>
<tr>
<td>Diabetes mellitus (type 1)</td>
<td>0.30-0.70</td>
<td>0.05-0.10</td>
<td>~0.80</td>
</tr>
<tr>
<td>Diabetes mellitus (type 2)</td>
<td>0.35-0.60</td>
<td>0.15-0.20</td>
<td>~0.60</td>
</tr>
<tr>
<td>Epilepsy (idiopathic)</td>
<td>0.69</td>
<td>0.14</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>Height†</td>
<td>0.94</td>
<td>0.44</td>
<td>1.0</td>
</tr>
<tr>
<td>IQ‡</td>
<td>0.76</td>
<td>0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>Measles</td>
<td>0.95</td>
<td>0.87</td>
<td>0.16</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>0.28</td>
<td>0.03</td>
<td>0.50</td>
</tr>
<tr>
<td>Myocardial infarction (males)</td>
<td>0.39</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>Myocardial infarction (females)</td>
<td>0.44</td>
<td>0.14</td>
<td>0.60</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>0.47</td>
<td>0.12</td>
<td>0.70</td>
</tr>
<tr>
<td>Spina bifida</td>
<td>0.72</td>
<td>0.33</td>
<td>0.78</td>
</tr>
</tbody>
</table>

DZ, Dizygotic; IQ, intelligence quotient; MZ, monozygotic.

* These figures were compiled from a large variety of sources and represent primarily European and U.S. populations. Heritability is estimated here using the formula $2(c_{MZ} - c_{DZ})$.

† Because these are quantitative traits, correlation coefficients are given rather than concordance rates.

‡ Several heritability estimates exceed 1.0. Because it is impossible for >100% of the variance of a trait to be genetically determined, these values indicate that other factors, such as shared environmental factors, must be operating.
## Estimates of heritability of various disorders

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Frequency (%)</th>
<th>Heritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>Asthma</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>Cleft lip ± cleft palate</td>
<td>0.1</td>
<td>76</td>
</tr>
<tr>
<td>Pyloric stenosis</td>
<td>0.3</td>
<td>75</td>
</tr>
<tr>
<td>Ankylosing spondylitis</td>
<td>0.2</td>
<td>70</td>
</tr>
<tr>
<td>Club foot</td>
<td>0.1</td>
<td>68</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>Hypertension (essential)</td>
<td>5</td>
<td>62</td>
</tr>
<tr>
<td>Congenital dislocation of the hip</td>
<td>0.1</td>
<td>60</td>
</tr>
<tr>
<td>Anencephaly and spina bifida</td>
<td>0.3</td>
<td>60</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td>Congenital heart disease</td>
<td>0.5</td>
<td>35</td>
</tr>
</tbody>
</table>
IDENTIFYING GENES THAT CAUSE MULTIFACTORIAL DISORDERS

- Multifactorial disorders are common and make a major contribution to human morbidity and mortality.

- It is therefore not surprising that vigorous efforts are being made to try to identify genes that contribute to their etiology.

- A number of strategies have been used to search for disease susceptibility genes.
Strategy to find disease susceptibility genes for type 2 diabetes mellitus (T2DM). Candidate genes may be selected from human models (e.g. monogenic forms of diabetes), knowledge of biology (insulin secretion or action), positional cloning or animal models. The candidate gene is screened to find variants, which are then tested for association with T2DM by genotyping cohorts of subjects with T2DM and controls. (Modified from Gloyn A L 2003 The search for type 2 diabetes genes. Ageing Res Rev 2: 111-127, with permission.)
1. LINKAGE ANALYSIS

- Linkage analysis has proved extremely valuable in mapping single-gene disorders by studying the co-segregation of genetic markers with the disease.

- However, this type of approach is much more difficult in multifactorial disorders, for the following reasons:

1. If a multifactorial disorder has a true polygenic underlying genetic susceptibility, in theory it is unlikely that alleles at a single locus will make a major contribution.

   - It is extremely difficult mathematically to develop strategies for detecting linkage of additive 'polygenes', each of which makes only a small contribution to the phenotype.

2. Many multifactorial diseases show a variable age of onset so that the genetic status of unaffected family members cannot be known with certainty.
3. Most families in which a multifactorial disease is, or has been, present have only one or two living affected members so that the number of 'informative meioses' available for study is usually very small.

4. Some apparent multifactorial disorders, such as coronary artery disease and schizophrenia, are probably etiologically heterogeneous, with different genetic and environmental mechanisms involved in different subtypes that cannot be easily distinguished at the phenotypic level.

- This makes analysis of linkage results very difficult.
Despite these limitations, progress is being made towards identifying susceptibility loci using modifications of the approaches utilized for mapping single gene loci. It has been recognized for some time that one of the best approaches would be to undertake disease association studies and linkage analysis utilizing a so-called 'ideal' population. Such a population would be relatively large yet historically isolated and therefore genetically homogeneous, with extensive medical records dating back for many generations, a large tissue bank, good medical services and a cooperative willing citizenship. The 270 000 citizens of Iceland have been deemed to represent such an ideal population, and a genomics company, known as DeCODE Genetics, has been granted a licence to set up a national medical database and undertake genetic research.
Similar initiatives are likely in other populations; recently, for example, the Center for Arab Genomic Studies (CAGS) has been established in Dubai.

Although on the one hand these initiatives have raised serious concern about the issue of informed consent, on the other hand they could lead to the relatively rapid isolation of genes that make a significant contribution to human morbidity and mortality.
a. Affected sibling-pair analysis

- Standard linkage analysis requires information regarding the mode of inheritance, gene frequencies and penetrance.
- For multifactorial disorders this information is not usually available.
- One solution to this problem is to use a model-free method of linkage analysis that seeks to identify alleles or chromosome regions shared by affected individuals.
- A common approach is to look for regions of the genome that are 'identical by descent' (IBD) in affected sibling pairs.
If affected siblings inherit a particular allele more or less often than would be expected by chance, this indicates that the allele or its locus is involved in some way in causing the disease.

Consider a set of parents with alleles AB (father) and CD (mother) at a particular locus.

The probability that any two of their children will have both alleles in common is 1 in 4.

The probability that they will have one allele in common is 1 in 2, and the probability that they will have no alleles in common is 1 in 4.

If siblings who are affected with a particular disease show deviation from this 1 : 2 : 1 ratio for a particular variant, this implies that there is a causal relationship between the locus and the disease.
The probability that siblings will have 2, 1 or 0 parental alleles in common. Significant deviation from the 1: 2 : 1 ratio indicates that the locus is causally related to the disease.
Many genome-wide scans have been performed for various disorders and, although a number of loci have been mapped, the number of disease susceptibility genes identified by this approach is disappointingly small.

One reason, probably, is the complex nature of multifactorial disease, with numerous genes of modest effect interacting with one another and the environment.

Some studies are simply underpowered and recent efforts have concentrated on large collections of carefully phenotyped affected sibling pairs.
b. Linkage disequilibrium mapping

✓ Once a chromosome region that appears to confer susceptibility has been identified, the next step is to reduce the genetic interval by 'fine mapping'.

✓ The most powerful method uses *linkage disequilibrium* (LD) mapping to construct haplotypes by genotyping SNPs within the region.

✓ Historical cross-over points reduce the genetic interval by defining LD 'blocks'.

✓ Candidate genes within the region are then sequenced to find DNA variants that can be tested for association with the disease.
The LD structure of glucokinase. $r^2$ values between the 84 SNPs across a 116-kb region are presented. An $r^2$ value of 1 indicates that two SNPs are linked. There are two blocks of LD within the glucokinase gene (outlined in red).
Calculation of odds ratio for a disease association

<table>
<thead>
<tr>
<th></th>
<th>Allele 1</th>
<th>Allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Controls</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>( \frac{a}{c} \div \frac{b}{d} )</td>
<td>( \frac{ad}{bc} )</td>
</tr>
</tbody>
</table>
Transmission disequilibrium test

- One way to overcome population stratification problems is to use family-based controls.

- The *transmission disequilibrium test* (TDT) requires a collection of trios that consist of an affected proband and both parents (regardless of affection status).

- Parents who are heterozygous for the marker allele in question are selected and the number of times this allele is transmitted to their affected offspring is compared to the number of times the other allele is transmitted.

- Overtransmission of the marker allele strengthens the evidence for association, but definitive evidence that a variant is a predisposing allele usually requires functional studies.
2. ASSOCIATION STUDIES

- The study of disease associations is undertaken by comparing the incidence of a particular variant in affected patients with the incidence in a carefully matched control group.

- This approach is often described as a 'case-control' study.

- If the incidences in the two groups differ significantly, this provides evidence for a positive or negative association.

- The polymorphic system that has frequently been studied is the HLA (human leukocyte antigen) histocompatibility complex on chromosome 6.

- One of the strongest known HLA associations is that between ankylosing spondylitis and the B27 allele.
This is present in approximately 90% of all patients and in only 5% of controls.

The strength of an HLA association is indicated by the ratio of the risk of developing the disease in those with the antigen to the risk of developing the disease in those without the antigen.

This is known as the **odds ratio** and it gives an indication of how much more frequently the disease occurs in individuals with a specific marker than in those without that marker.
One of the major difficulties with disease associations is to establish how they should best be interpreted.

In particular it is important to try to rule out a chance or spurious observation by ensuring as far as possible that the proposed association is biologically plausible and that the patient and control groups are closely matched.

If evidence for a strong association is forthcoming, this suggests that the allele encoded by the marker locus is directly involved in causing the disease (i.e. a susceptibility locus) or that the marker locus is in linkage disequilibrium with a closely linked susceptibility locus.

When considering disease associations it is important to remember that the identification of a susceptibility locus does not mean that the definitive disease gene has been identified.
This is illustrated by the association of HLA-B27 with ankylosing spondylitis.

Although this is one of the strongest disease associations known, only 1% of all individuals with HLA-B27 develop ankylosing spondylitis, so that many other factors, genetic and/or environmental, must be involved in causing this condition.

Positive results from association studies require replication in other cohorts.

A common reason for false-positive association is population stratification, where the population contains a number of subsets and both the disease and the allele happen to be common within that subset.

A famous example, reported in a study by Lander and Schork, showed that in a San Francisco population HLA-A1 is associated with the ability to eat with chopsticks.

This association is simply explained by the fact that HLA-A1 is more common amongst Chinese people than caucasians!
3. WHOLE-GENOME ASSOCIATION STUDIES

- In *whole-genome association* studies, researchers compare the entire genome in a case-control study, rather than looking at just one variant at a time.

- This powerful new method can therefore be used to identify new disease susceptibility genes.

- Technological advances mean that it is now possible simultaneously to test up to 500,000 SNPs on a single microarray (a 500K 'SNP Chip'). In the UK, the Wellcome Trust has funded a large project to perform whole-genome association studies in approximately 3000 controls and 2000 patients affected with tuberculosis, coronary heart disease, type 1 diabetes, type 2 diabetes, rheumatoid arthritis, Crohn disease and ulcerative colitis, bipolar disorder or hypertension (http://www.wtccc.org.uk).
A. International HapMap Project
(http://www.hapmap.org)

- Whilst it is estimated that there may be up to 10 million SNPs in the human genome, many SNPs are in linkage disequilibrium and are therefore co-inherited.

- Regions of linked SNPs are known as haplotypes.

- A single SNP can be chosen that 'tags' a haplotype; these are described as tag SNPs.

- The International HapMap Project is identifying haplotypes in different populations and it is estimated that the total number of haplotype-tagging SNPs is between 300 000 and 600 000 depending on the population studied.
This means that whole-genome association studies of approximately 500,000 tag SNPs can test for the majority of genetic variation in the human genome.

The HapMap Project will provide a valuable resource to learn more about the genetic predispositions that underlie common diseases such as cardiovascular disease, diabetes, cancer, autoimmune and psychiatric disorders.
Populations studied in the International HapMap Project

<table>
<thead>
<tr>
<th>Town/country</th>
<th>Ancestry</th>
<th>Samples analysed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibadan, Nigeria</td>
<td>Yoruba</td>
<td>30 trios (adult and both parents)</td>
</tr>
<tr>
<td>Tokyo, Japan</td>
<td>Japanese</td>
<td>45 unrelated individuals</td>
</tr>
<tr>
<td>Beijing, China</td>
<td>Chinese</td>
<td>45 unrelated individuals</td>
</tr>
<tr>
<td>USA</td>
<td>Northern and western European</td>
<td>30 trios (adult and both parents)</td>
</tr>
</tbody>
</table>
CONCLUSION

- The term multifactorial has been coined to describe the pattern of inheritance displayed by a large number of common disorders that show familial clustering and are probably caused by the interaction of genetic and environmental factors.
- The genetic mechanisms underlying these disorders are not well understood.
- The liability/threshold model should be viewed as an attractive hypothesis rather than as proven scientific fact.
- Research in molecular biology is beginning to unravel some of the mysteries of multifactorial inheritance.
- The past 10 years has seen the recruitment of large numbers of patients and controls to create valuable DNA resources for study, and further collections are in progress.
For example, the UK Biobank Project (http://www.ukbiobank.ac.uk) aims to collect DNA samples and information on the health and lifestyle of 500,000 volunteers aged between 40 and 69 years.

Over the next 20-30 years, approved researchers will be able to use these resources to study the progression of illnesses such as cancer, heart disease, diabetes, and Alzheimer disease.

From this they hope to develop new and better ways of preventing, diagnosing and treating such problems.
Technological developments in SNP typing, together with an increased understanding of genetic variation, mean that the next few years are likely to prove very exciting as these new approaches are applied to polygenic disease.

This emphasis on the underlying genetic contribution to multifactorial disorders should not in any way detract from the importance of trying to identify major environmental causal factors.

This is amply demonstrated by the beneficial effect of folic acid supplementation in preventing neural tube defects.
1. The concept of multifactorial inheritance has been proposed to account for the common congenital malformations and acquired disorders that show non-Mendelian familial aggregation.

- These disorders are thought to result from the interaction of genetic and environmental factors.

2. Human characteristics such as height and intelligence, which show a normally distributed continuous distribution in the general population, are probably caused by the additive effects of many genes, i.e. polygenic inheritance.
3. According to the liability/threshold model for multifactorial inheritance, the population's genetic and environmental susceptibility, which is known as liability, is normally distributed.

   - Individuals are affected if their liability exceeds a threshold superimposed on the liability curve.

4. Recurrence risks to relatives for multifactorial disorders are influenced by disease severity, degree of relationship to the index case, number of affected close relatives and, if there is a higher incidence in one particular sex, the sex of the index case.

5. Heritability is a measure of the proportion of the total variance of a character or disease that is due to the genetic variance.
6. Loci that contribute to susceptibility for multifactorial disorders can be identified by
(a) a search for disease associations with variants in candidate genes,
(b) linkage analysis looking, for example, for chromosomal regions that are identical by
descent in affected sibling pairs
(c) whole-genome association studies to compare genetic variation across the entire
genome in large case-control studies.
   by [Peter D Turnpenny](https://www.peterdturnpenny.com)

   by [Lynn B. Jorde](https://www.lynnjorde.com), John C. Carey, [Michael J. Bamshad](https://www.bamshad.org)