University of Jordan
Faculty of Medicine

GENETICS & Molecular Biology

Number: 36

Done By: Abdullah Qaswal.

Subject: Multifactorial inheritance and common diseases.

Doctor: Mazin Al-Salihi.

Price:

Date: 27/4/2016

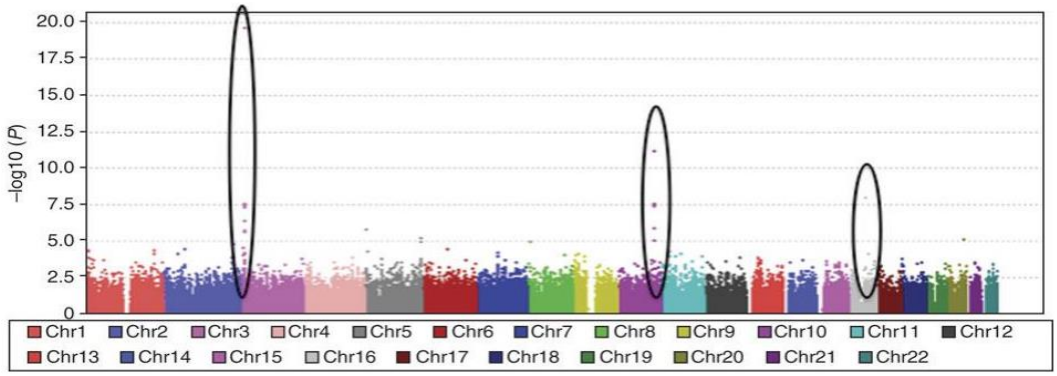
Multifactorial inheritance and common diseases

Multifactorial means that there are multiple factors or reasons including genetic and environmental factors which contribute in the occurrence of a certain disease.

The difference between a **single gene disorder**, with modifying genes or environmental modification, is that once you have that gene it will be the cause of the disease, and the rest factors or modifiers just increase of decrease the severity of the disease, while in **multifactorial diseases** there might be multiple genetic influences, multiple environmental factors and each one of those factors singly is not going to produce the disease.

Typically, multifactorial diseases corporate multiple genes, so if we were to produce a Manhattan plot for one of those diseases, what would it look like?

It would look like the picture below, with multiple associations over multiple chromosomes in multiple genes.



Multifactorial is typically polygenic and multifactorial diseases follow a bell shaped normal population distribution.

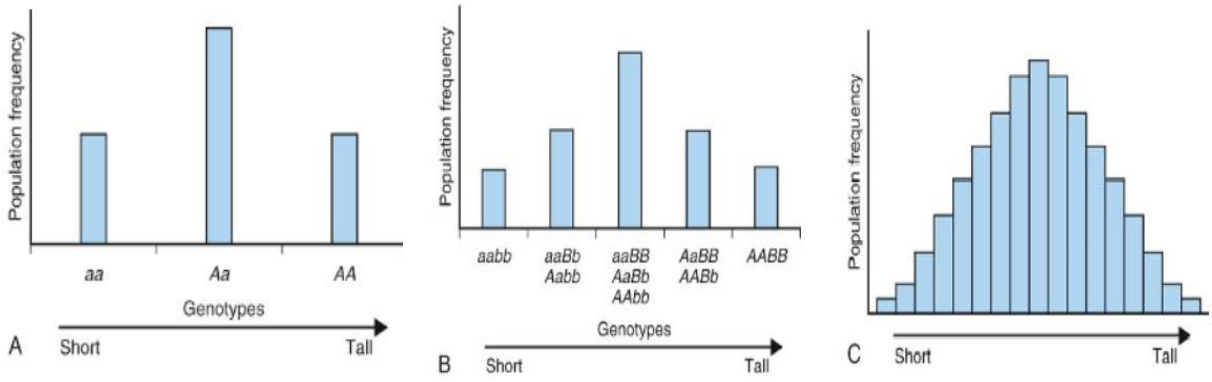


Figure 1

Figure 1 represents 3 types of distributions of a height in a population; A, B and C.

Figure A: represents a distribution of the height in a certain population assuming that there is only one single gene for height controlled by 2 alleles (A and a) with genotypes and phenotypes as the following:

AA is tall, *Aa* is intermediate in height and *aa* is short because the allele (A) is the tall-related gene and the allele (a) is the short-related gene.

So, according to figure (1, A) I will expect that the percentages of distribution of these phenotypes are: tall (25%), intermediate (50%) and short (25%), but this actually will not occur when you examine a population and class the people according to their height quantitatively (130cm, 131cm, 132cm…etc.). Instead, what will occur is a bell-shaped distribution which is represented as continuous columns of population frequency for each quantitative height, and this manner of distribution is illustrated in **figure C.**

Figure B: represents the distribution of height in a population assuming that the height is controlled by two genes; that is 4 alleles (A, a, B and b). Notice that the number of population-frequency columns increased in comparison with those in figure A.

Figure C: represents the distribution of height assuming that multiple factors (genetic and environmental factors) affect height, and this model is the *multifactorial model.*

According to **figure C**, you can notice how much the columns are packed together due **to the continuity of quantitative heights for populations** and this is what will actually occur when you examine a population and class them according to their heights as we mentioned before indicating that multiple factors are involved in the contribution of such distributions.

Note: most of quantitative or measured traits such as height and blood pressure follow a continuous manner forming a **bell-shaped distribution.**

From that we conclude that height is a multifactorial trait; that is, it does not only depend on genetics, as one gene won't make you really tall or extremely short, and it also depends on many environmental factors and hormonal effects, like what you eat, your growth pattern, when did you hit puberty…etc. So, it will follow a normal population distribution. And the combination of all the factors (genetic and environmental) creates the continuity (diversity) of the quantitative heights.

Another difference between multifactorial diseases and single gene disorders could be concluded from a pedigree, like the recurrence risk.

In multifactorial diseases, **recurrence risk** varies depending on how many family members are affected because environment plays a role here, too. The

more family members affected the higher the risk of other family members getting it because you're adding environmental influences that may increase your risk of having a certain disease, but this doesn't always hold true for single gene disorders, as there the percentages are fairly clear and there's no in between (If it's autosomal recessive you will have a 25% risk, and it doesn't matter whether you have many other siblings that have that AR disease, and if it's autosomal dominant then your chance is 50%). Also, the more severe the disease is in the **proband**, which is the first person who is identified with the disease, the more likely subsequent children will have the disease, while in an AR or AD, where there is only one mutation, the severity of the disease doesn't really tell you whether the recurrence risk of the next person is going to be higher or lower.

Now, in AD and AR the chance of getting the disease drops more rapidly with **distant relatives**, so as you move one away it will be reduced by a certain percentage. So, the likelihood of an, for example, AR disease happening in the next relative becomes less by losing one allele and so on. It's a stepwise process (the risk will be 25% at first, then 12.5%, 6.25...etc.). As for multifactorial inheritance, the more distant the relatives are the less likely they're sharing genetic material and the more different their environments are, and those two factors make the drop in the recurrence risk more severe.

Finally, and most importantly, the relative risk varies with **population prevalence**. In single gene disorders, population prevalence has no relation, unless they are intermarrying, whereas here, in multifactorial diseases, the higher the prevalence in the population the more likely one would get the disease, and that is because of environmental effect.

While most multifactorial diseases and traits follow the normal population distribution some of them do not, and those diseases follow the **threshold model**.

Here, multiple genetic influences and environmental effects come into play, and at a certain point they would tip the balance towards having the disease. That point or threshold limit may not be the same for males and females, one such example is **pyloric stenosis**, which it is more likely to occur in males than it is to occur in females. This also brings us back to the recurrence risk, if your proband case was a female, would it be more likely or less likely for the siblings coming after to be affected?

It would be more likely if the proband case is a female (here we're talking about pyloric stenosis) because the threshold for them is higher, and that means that more genetic influence and more environmental factors went into producing that diseased patient, which means that this family is in a particularly bad environment for that particular disease, and also that family has a lot of the genetic traits required to produce that disease because it took a lot more for the female to cross that threshold into becoming diseased. So, that means it will be more likely to occur in other siblings, especially males because you've passed well beyond the point of threshold for males, and if they received the same genetic influences that the female who was diseased had, and the environment is that similar then they're more likely to get the disease. If the case was the other way around (higher threshold for males), and if the proband (primary case) was a male then all the following siblings will have a higher risk of getting that disease.

Figure 2 represents a *liability distribution*, which looks the same as a normal distribution. Let's say there are 10 things that cause the disease and they're either negative or positive, and if we thought of them as alleles, what would be the chance of an individual having all ten positive or all ten negative (high or low)?

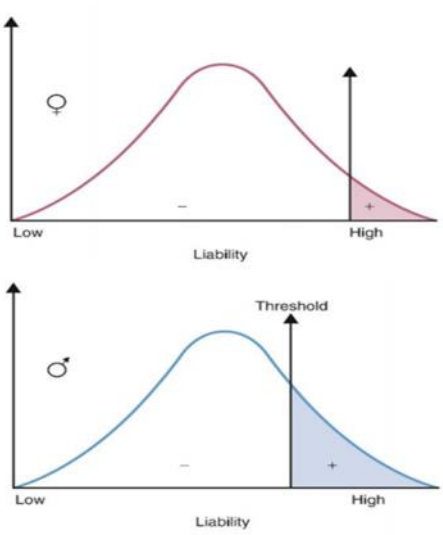


Figure 2

It would be low either way, and the higher chance is to get a mix of both, positive and negative, and that is what makes a normal distribution. So, people on the severe end of the positive (+) area are those that are going to be liable to get the disease, and those on the severe end of negative (-) are unlikely to get the disease, but their percentage is just as small as those that get the disease, which is why it ends up looking like a normal distribution.

Note: If you have a mutation in CFTR gene you would have *cystic fibrosis*, and all of the other genes, bacterial infections, environmental effects, and other factors, such as smoking, make your disease worse or better, but that **single gene** identified your disease. On the other hand, when we're talking about *heart diseases*, in which there's a genetic influence, and also you would be more likely to get the disease if you have a relative who has cardiac problems, but that is not just because you share genetic material, but also because you potentially share the same environmental effects, as your diets, drug use, habits and general atmosphere may be similar, and that makes it a **multifactorial disease**.

Nature (genetic) versus Nurture (environmental):

It is a mistake to view these two factors as mutually exclusive as only few traits are influenced by genes alone, or only by the environment. Most are influenced by both.

To estimate the relative influence of genes and environment, two strategies will be reviewed; twin studies and adoption studies.

Twin studies:

Twins can be Monozygotic (MZ) or dizygotic (DZ). In MZ twins, they are genetically identical, however, differences may occur at the level of **epigenetics** in MZ twin, while in DZ twin, they share **50%** of their DNA.

How can we benefit from DZ and MZ twins to estimate the relative influence of genes and environment?

DZ twins provide a convenient comparison: their environmental differences should be similar to those of MZ twins, but their genetic differences are as great as those between normal siblings because they share 50% of their DNA.

So, if both members of a twin pair share a trait or a disease, they are said to be **concordant.** If they do not share the trait, then they are **discordant.**

Now, we are going to use the **concordance rate**, which reflects how much the twins share a certain trait or disease for both MZ and DZ twins and make a comparison between the two rates.

Since MZ twins are genetically identical, we can set them as a standard group and compare their concordance rate with that of DZ twins, so if the concordance rate for DZ twins is similar or close to concordance rate for MZ twins, that means that the genetic factor is less likely to cause a major effect on the occurrence of a certain trait or disease and the environmental factors are more likely to cause that major effect.

On the other hand, if the rate of DZ is not similar (with a relative big difference) to the rate of MZ that means the genetic factor is more likely to cause a major effect and environmental factors are less likely to do so.

Concordance rate applies for yes/no questions (does the patient have the disease?), but it does not apply to continuous **quantitative traits**, such as height and blood pressure, so we use something called *intraclass correlation coefficient*. Correlation coefficient has a range between -1.0 and +1.0 and it measures the degree of homogeneity of a trait. For twin studies, we use it to assess the degree of similarity between twins for a trait such as height, but you have to notice (with regard to intraclass correlation coefficient) that if a

trait is totally dependent on genes, the correlation coefficient for **MZ twins equals 1**, but **equals 0.5 for DZ** twins since they share only 50% of their DNA.

The interclass correlation coefficient will be one if the twins are very similar, zero if they are not similar, and -1 if they are total opposites, as for concordance rate it will be 1 if they are concordant and zero if they are completely separated, so it's from 1 to -1 versus 1 to 0.

Let's take examples to make the concept clear:

Example 1:

Traits	**Concordance rate**
Measles:	MZ (0.95), DZ (0.87).
Epilepsy:	MZ (0.69), DZ (0.14).

With regard to **measles**, the rate for DZ is fairly close to the rate of MZ indicating that measles is unlikely to be influenced markedly by genes, but rather by environmental effects because measles is a contagious disease caused by a virus from the environment, but by applying a comparison between the two rates for **epilepsy**, you will find that the rate for DZ is far away from the rate for MZ indicating that there's a massive genetic influence for epilepsy compared to measles.

Example2:

Trait	**Intraclass correlation coefficient (ICC)**
Height	MZ (0.94), DZ (0.44)

You can notice that ICC for MZ is close to 1 (*not* equal to 1 because there are other factors that determine the height) indicating that height is markedly dependent on genes, and the ICC for DZ is close to 0.5 indicating also that height is markedly dependent on genes.

Note also that if there is a minus sign in the ICC that would indicate a perfect disagreement in the trait (if one member of the twins got a certain trait, the other will get the opposite).

By looking to the pictures in figure (4), you can notice the difference between the two members of the MZ twin by aging and growing up even if the genetic material for both is mostly the same, actually the one to left is a smoker and the other one is not, so environmental factors, like smoking, can affect and somewhat change the person even from their MZ identical twin.

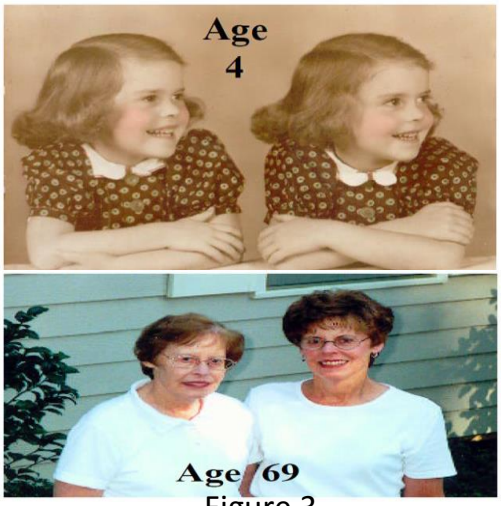


Figure 3

Our homework for this lecture is **heritability** (refer to chapter 12, pages; 245 - 246).

Adoption studies:

Studies of adopted children are also used to estimate the genetic contribution to a multifactorial trait.

If you put two identical twins in two very different environment you would have taken out the likelihood of the environments being identical, and you're assuming that they're genetically identical, so any differences in those twins are going to be more likely environmental. Keep in mind that those twins shared the same womb (uterus), which is typically an identical environment, during prenatal development, also children are sometimes adopted after they are several years old, indicating that some non-genetic influences have been imparted by the natural parents.

The other more useful form of adoption studies is when a child who is born to parents that have the disease, and those parents can't take care of the child, so that child would be adopted to a family who does not have that disease, so now whether that child gets and develops that kind of disease or not is based on two things; genetics or environment. We've potentially taken out the environmental effect by removing the child from the diseased family to a non-diseased one, but if that child develops the disease here, it won't be precise to say that it all depends on the environment. If the disease is environmental the rate of the disease in this family compared to the **general population** (if we expand it to the general population) is going to be, supposedly, almost the same. Whereas, if the child gets the disease (don't forget that we study multiple children in multiple adoption cases), and he/they had an increased rate of disease compared to the general population that's when we think that it's genetic rather than environmental.

So you now have a multifactorial disease in your hands with a genetic component, **how do you find the gene?**

By using GWAS (genome-wide association studies), multiple polymorphisms are shown up and by this technique we can detect multiple genes which contribute in the occurrence of a *multifactorial disease.*

Now, let's take some diseases applying the multifactorial model:

1. **Heart disease (CAD):**
   There are environmental factors and genetic factors affecting both, **lipid profile** and **inflammation**; several genes are involved in lipid metabolism

and transport, also several genes whose protein products contribute to inflammation have been associated with CAD.

2. **Familial hypercholesterolemia (FH):**
   It is autosomal dominant (AD), where a mutation in LDLR (low density lipoproteins receptor) affects the level of cholesterol (if you lose 50% of that receptor then you are not equipped to uptake LDL). Other mutations in certain cases may occur affecting cholesterol levels, too.
   Note: we mainly look for xanthomas and corneal arcus to indicate that there's peripheral deposition of cholesterol.

   You absorb fatty acids from your intestine and inside the enterocytes these fatty acids along with glycerol form triacylglycerol (TAG), which are then packaged in chylomicrons (lipoproteins), removed through the lymphatics, and then go to the liver which starts packaging remnant chylomicrons with cholesterol (most of cholesterol is synthesized in your body) to get low-density lipoprotein (LDL), also known as "bad cholesterol".
   LDL particle is bound to the LDL receptors to get inside the cell, and then cholesterol is going to be released. Excess cholesterol can be stored in the cell as cholesterol ester or removed from cell by association with high-density lipoprotein (HDL); that is, good cholesterol.
   Note: the reason it's called low and high density lipoproteins refers to the content of lipids to proteins. So, HDL is high in density because it contains less lipids and more proteins and that is why it's called good cholesterol because it tells us that the body is producing enough cholesterol to be incorporated peripherally onto cells, but not in excess amounts that will get back to the liver. So, the higher your HDL and the lower you LDL content is the better of you are.

   As cholesterol levels in the cell rise, cellular cholesterol synthesis is reduced by inhibiting the rate limiting enzyme, HMG-CoA reductase, in addition, the number of LDL receptors is decreased by lowering the transcription rate of the LDL receptor gene itself and this decreases cholesterol intake.

According to what is mentioned above, a mutation in LDL-receptor gene reduces the number of effective LDL receptors, resulting in decreased LDL uptake and hence elevated levels of circulating cholesterol. The number of effective receptors is reduced by about half in FH heterozygotes, and homozygotes have virtually no functional LDL receptors.

Statin drugs inhibit HMG-CoA reductase and so reduce cholesterol synthesis. Decreased synthesis leads to further production of LDL receptors. Using these drugs along with other therapies, serum cholesterol levels in FH heterozygotes can often be reduced to approximately normal level.
On the other hand, these therapies are largely ineffective in homozygotes because these persons have fewer or no LDL receptors.

What is the problem associated with the accumulation of cholesterol in blood vessels?
Accumulation of cholesterol will lead to attraction of macrophages to ingest these crystals, so these activated macrophages are going to secrete a lot of chemical mediators activating more and more inflammatory cells, finally you end up with chronic inflammation and fibrosis (deposition of collagen). As result of fibrosis, calcium ions are going to be deposited on the deposited collagen (dystrophic calcification) making it fragile, producing atherosclerosis plaques and eventually cardiovascular disease (CVD). The previous events would lead to a restriction in the blood supply, and at some point, with high blood pressure the calcium deposits would be ruptured, and this calcium will deactivate and reduce the coagulation factors and platelets creating a thrombus, and eventually a heart attack occurs.

Also, a mutation(s) in the gene encoding apolipoprotein B is another genetic cause of elevated LDL cholesterol. These mutations occur in the portion of the gene that is responsible for binding of apolipoprotein B to LDL receptor, and they increase circulating LDL cholesterol levels.

In addition, a small number of FH cases are caused by a mutations in the gene that encodes PCSK9, an enzyme that plays a key role in degrading LDL receptors. People who have a mutation in PCSK9 actually have a very good lipid profile, very low LDL and very high HDL because the LDL receptors won't be destroyed and would remain on the surface, so now a new drug has been developed that inhibits PCSK9, which can help the general population (those with high cholesterol levels), as it would increase LDL receptors and thus decrease the circulating LDL levels.

3. **Diabetes Mellitus (DM):**
   A heterogeneous group of disorders of glucose metabolism.

A. **DM1:**
   Is characterized by T-cells infiltrating the pancreas and destroying β-cells, or that the T-regulator cells will not inactivate and inhibit cytotoxic T-

lymphocytes and so they would start killing the β-cells. Also, it may be caused by mutations that affect insulin transcription. Patients with type 1 diabetes must receive exogenous insulin to survive.

In addition to T-cells infiltration of the pancreas, autoantibodies are formed against pancreatic cells, insulin and enzymes.

There are dozens of genes that are associated with DM1 susceptibility. One of these is CTLA4 (cytotoxic lymphocyte associated-4) gene, which encodes an inhibitory T-cell receptor. Another gene associated with DM1 susceptibility is PTPN22, which is involved in T-cell regulation.

Also, it is characterized by abnormality in antigen presentation through MHC1, and variations in insulin region that encodes insulin accounts in familial clustering of DM1.

B. DM 2:

Persons with type 2 diabetes usually have some degree of endogenous insulin production, at least in the earlier stages of the disease, and they can sometimes be treated successfully with dietary modifications, oral drugs, or both.
In contrast to those with type 1 diabetes, patients with type 2 diabetes have insulin resistance (their cells have difficulty using insulin) and more likely to be obese. Because of that resistance they are not treated with insulin, and instead by drugs like Glucophage, Metformin…etc.

Multiple genes have been identified contributing to DM 2 susceptibility:

1. *TCF7L2 gene*, which encodes a transcription factor involved in **insulin secretion.**
2. *KCNJ11 gene*, which encodes a potassium channel necessary for glucose-stimulated **insulin secretion.**
3. *PPARγ gene*, which encodes a nuclear receptor that is involved in adipocyte differentiation and glucose metabolism, so it is well-related to obesity which is common in DM2.

C. Maturity-onset diabetes of the young (MODY):

Typically occurs before 25 years of age and follows an autosomal dominant mode of inheritance. In contrast to type2 diabetes, it is not associated with obesity.

Mutations in a gene which encodes glucokinase, a rate-limiting enzyme in the conversion of glucose to glucose-6-phosphate in the pancreas, can cause MODY.

There are some other five mutations in genes which encode transcription factors involved in pancreatic development or insulin regulation, which can also cause this type of diabetes.

Mutations in these genes, all of which are expressed in pancreatic β-cells, lead to β-cell abnormality and thus to diabetes.

4. **Obesity:**
   People who have a BMI that is over 40 were found to have leptin mutations. Leptin is a hormone that is produced by adipocytes, then it goes to the hypothalamus in the brain and inhibits the feeding behavior by binding to its receptors there, as presented in figure (4).

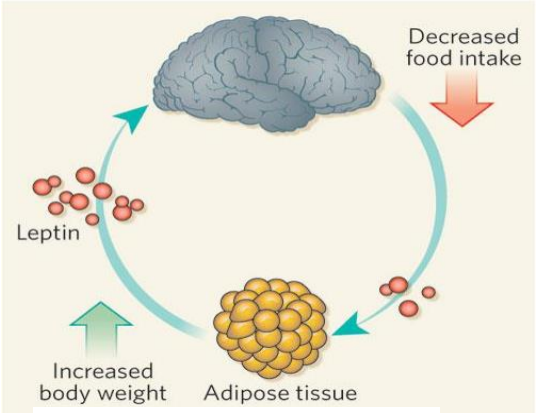


Figure 4

Does this mean that obese people have low level of leptin hormone or a mutation in its receptor?

No, actually, most obese people have **high levels of leptin**, indicating that the leptin gene is functioning normally. Leptin receptor defects were then suspected, but these are also very rare in humans. And that leads us to think that there are other mutations and genes that affect the body's fat mass and how much we store, such as FTO-IRX.

So, by studying **extreme cases** of obesity, the genetic background will be more obvious.

5. Alzheimer disease (AD):

Alzheimer disease is a genetically heterogeneous disorder. Approximately half of early-onset cases can be attributed to mutations in any of the three genes, all of which affect amyloid-β deposition. Two of the genes, *presenilin1 (PS1)* and *presenilin2 (PS2)*, are very similar to each other, and their protein products are involved in cleavage of the amyloid-β precursor protein **(APP)** by *γ-secretase.*

Gain-of-function mutations in PS 1or PS2 affect the cleavage of APP such that amyloid-producing forms of it accumulate excessively producing amyloid plaques which are deposited in the brain.
A small percentage of cases of **early-onset** Alzheimer disease are caused by mutations of the gene (APP) that encodes APP itself, which is located on **chromosome 21.** These mutations disrupt normal secretase cleavage sites in APP, again leading to the accumulation of the protein product.

It is interesting that this gene **(APP gene)** is present in three copies in people with **trisomy 21**, where the extra gene copy leads to amyloid deposition and the frequent occurrence of early onset Alzheimer disease in **down syndrome** patients and this is also an example on **extreme cases** through which we can make the genetic background for AD more obvious.

An important risk factor for the more common **late-onset form of AD** is allelic variation in the *apolipoproteinE (APOE)* locus, which encodes protein product associated with *clearance of amyloid* from the brain. So, if you get a mutation in this locus that means less clearance of amyloid and more deposition.

Test yourself!

A **male (1)** affected by type A hemophilia married a **female (2)** whose father has type A hemophilia and they had a **baby (3).** After 18 years, the baby became an adult and decided to study human medicine. Unfortunately, he left the faculty after 4 years of studying and decided to marry. He married a **female (4)**, but because he had medical background about the diseases and the detection, etc. they did early detection before marriage. The results showed that 50% of males of their offspring will have type A hemophilia and there is no chance to have any female with type A hemophilia. Assuming that X_1 : is the chromosome with normal allele and X_2 : is the chromosome with disease allele, **Answer the questions 1 and 2:**

1. The genotype and phenotype for 2 and 4 respectively are:

    a. $X_1 X_1$ (normal) , $X_2 X_2$ (diseased)
    b. $X_1 X_2$ (normal) , $X_1 X_2$ (normal)
    c. $X_2 X_2$ (diseased) , $X_1 X_1$ (normal)
    d. $X_1 X_2$ (diseased) , $X_1 X_2$ (diseased)
    e. There is something missing in the question.

2. The percentage to have a healthy female (without type A hemophilia) among females of offspring of **1and 2** is:
   a. 50%
   b. 100%
   c. 0%
   d. 25%
   e. 60%

3. For a certain disease, the difference between concordance rate for MZ and DZ is 0.55 , and the heritability equals 0.64, then the concordance rate for MZ is:
   a.0.69
   b. 0.54
   c.0.95
   d.0.14
   e. 0.41

Hints: with regard to Q1 and 2, refer to sheet 3, while Q3 is a question related to the homework for this lecture (chapter 12, pages; 245, 246).

Answers:
1. B 2. A 3. A

The previous questions were written by Abdullah Qaswal.
And as Albert Einstein says: “Imagination is more important than Knowledge”.

The sheet is over

This sheet has been corrected and edited.