Chapter 10

DCG: Disorders with Complex Genetics

10.1 Introduction: Complex genetic disorders and disorders with complex genetics

In medicine and medical genetics, one often encounters the phrase complex genetic disorder to refer to many common syndromes like diabetes, cancers, and hypertension. The definition of a complex genetic disorder is quite simple—it is a disorder that has something genetic going on but no one has a clue as to what the genetics are. The term complex genetic disorder is actually a misnomer because it implies that the disorder is genetic whereas empirical research usually shows that the environment can play a large etiological role. To insure that we do not neglect the environment, the term disorder(s) with complex genetics (DCG) will be used in this text.

There are two hallmarks for DCGs. We have already mentioned the first—all DCGs have an important environmental influence on them. The second hallmark is the prevalence of the disorder. The frequency of Mendelian disorders is one is several thousands of births. (Sickle cell anemia is a rare exception.) The frequency of DCG is one to two orders of magnitude higher. Indeed, if a disorder has prevalence of .1% or above, it is certain to be a DCG.

In terms of human behavior, all forms of psychopathology (schizophrenia, affective disorder, the anxiety disorders, etc.) are DCGs. Indeed, all forms of psychopathology have important—albeit largely unknown—environmental factors intimately tied up in their etiologies.

Do not let the term DCG engender pessimism about ever finding the causes and effective treatments for psychopathology (or any medical DCG for that matter). Indeed, research into DCG, although it is very difficult, is one of the exciting challenges to both the genetic and the epidemiological community. Finding and cloning the gene for a Mendelian disorder is much akin to following
the Oregon trail after a generation of pioneers have charted the land and founded communities along the way—just follow the road signs and you will get there. Research into DCG explores virgin territory.

10.2 Alzheimer’s Disease: A Model DCG

There are many model DCGs—breast cancer, diabetes, hypercholesterolemia, atherosclerosis and hypertension are but a few. Unfortunately, no forms of psychopathology can be used as a model DCG because our knowledge about their genetics is too embryonic at this stage. Perhaps the best model with behavioral sequelae is Alzheimer’s disease (AD), so we will use that to illustrate how a DCG is being researched.

10.3 AD: The Phenotype

AD is a progressive and irreversible disorder that involves damage and the eventual death of neurons in the central nervous system. The first sign of AD is usually memory loss for recent events—one may have trouble recalling the name of the current President but have no difficulty remembering who the President was when one was born. In normals, memory abilities begin a slow decline starting in the mid 20s and becoming noticeable in middle age—that familiar experience of recognizing a face but failing to recall the name just becomes more common with age. Indeed, it is quite difficult to distinguish the very early stages of AD from normal forgetfulness, and many family members fail to recognize that someone is having significant problems with memory.

As AD progresses, cognitive functioning and judgment deteriorate, and the dementia (a progressive and irreversible loss of intellectual functioning) becomes evident. It is not uncommon at this stage to see changes in personality—someone with a lifelong tendency to be patient and forbearing may become hostile and irritable or a reserved person may become outgoing and fatuous. Depression is also encountered. Eventually, all forms of reasoning are impaired, the person becomes unable to perform routine tasks like getting washed and dressed, and in the last stages may be incapable of communication. The course of the illness can range anywhere between two and twenty years, but on average people with AD survive around 8 to 10 years after the first diagnosis. Although AD is often listed as a cause of death, in many cases it is an indirect cause. For example, in the final stages, people become bedridden and physically weak, permitting opportunistic infections to occur. Many people with AD die from pneumonia.

AD is an adult disorder with an age of onset that can start as early as 30 to 40. The risk for developing AD increases with age and roughly doubles every 5 years after the age of 65. Indeed, age per se, is the major known risk factor for AD. Advances in medicine and public recognition of healthy diets and lifestyles have dramatically increased life expectancy over the past century,
greatly increasing the overall prevalence of AD, placing heavy economic burdens on the health care system, and creating heavy personal, emotional, and financial problems for family members and caregivers. For these reasons, AD is being heavily researched.

Although AD can be diagnosed using psychometric tests of memory and cognition, the only sure diagnosis is made on autopsy when the brain is examined under the microscope. AD brains show characteristic plaques and neurofibrillary tangles, particularly in the cortex and the hippocampus—an area of the brain deeply involved in memory.

In medical parlance, plaque simply means a deposit of foreign material. In AD, the plaques consist of misfolded pieces of a protein (an amyloid) that are chemically “sticky” and become attached to pieces of neurons and glial cells. The tangles derive from microtubules, long filaments in the neuron that act as train tracks in transporting molecules from the area around the nucleus down the axon to the terminal ends of the nerve. Tau proteins act as railway ties, preserving the integrity of the microtubule. In the formation of tangles in AD, the tau proteins are chemically changed and clump together, destroying the integrity of the microtubule. Figure 10.1 shows an amyloid plaque and a neurofibrillary tangle (NFT).

In addition to amyloid plaques and tangles, nerve cells die in AD leading to general atrophy of the brain.
10.4 Classification

AD is classified along two different dimensions—family history and age of onset. It has long been recognized that AD can run in some families, and early family studies suggested that a minority of families had transmission patterns consistent with a dominant gene. Today, family history is used to classify AD into two types, although the boundary between the two types is sometimes fuzzy. The first type is called familial Alzheimer’s disease or FAD and is diagnosed when the person with AD has at least one affected first-degree relative. The second and by far more common type is usually referred to as nonfamilial AD or sporadic Alzheimer’s disease (SAD) and is diagnosed when there is no clear pattern of familial aggregation.

Although age of onset is a continuous and quantitative variable, it is often dichotomized giving the two classes of early-onset AD (EOAD) and late-onset AD (LOAD). The age of onset distinguishing the two is usually taken as 65. The vast majority of AD patients are late-onset cases.

Familial Alzheimer’s disease is strongly—not perfectly—correlated with age of onset. Most FAD cases are early onset and also have a rapidly deteriorating course. However, not all early-onset cases are FAD and not all FAD are early-onset. Late-onset cases tend to be sporadic.

Do not get the impression that FAD represents the “genetic” cases of AD while sporadic, late-onset AD are due to the environment. As we will learn, there are a few rare mutations in single genes that can cause AD. These cases all fall into FAD. But twin data suggest that there is still heritability for LOAD (REFS).

10.5 Genetics of Alzheimer’s Disease

10.5.1 Mendelian forms

Some medical DCGs may be due to the effects of a single gene. These are called Mendelian forms of the DCG. Classic examples are the BRCA1 and BRCA2 genes and breast cancer. Some alleles in these genes may cause breast cancer and in some cases, uterine cancer. Typically, these are alleles are usually rare and account for a only a small proportion of all cases of breast and uterine cancers, but they have a high penetrance. up to 80% in women by age 70 (Foulkes and Shoen, 2013). There are three Mendelian forms of AD: (1) the amyloid precursor protein (APP); presenilin-1 (PSEN1); and presenilin-2 (PSEN2).

10.5.1.1 The APP gene

In AD, the fact that cases of Down’s syndrome (which involves the inheritance of extra material from chromosome 21) eventually show the AD-like pathology of plaques and tangles focused early linkage and association studies on chromosome 21. At the same time, it was reported that a few rare pedigrees with AD showed patterns of transmission consistent with a dominant gene. Linkage studies on
chromosome 21 markers in these pedigrees soon found evidence of a single gene that acts as a Mendelian form for AD. That gene turned out to be the amyloid precursor protein gene (APP).

Today, at least 24 different mutations in the amyloid precursor protein gene have been described that can cause AD (Kanzi, 20 X.X). With one exception, these rare variants act as autosomal dominants with very high penetrance and result in an early onset to the disorder. Study of these mutants has greatly increased understanding of the amyloid plaques.

In the brain, the amyloid precursor protein (APP) lives in the plasma membrane of synaptic neurons. A portion of the protein resides within the neuron while a much larger portion extends out and into the synapse. The problem for AD comes when this protein is broken down.

Figure 10.2 gives a schematic of the APP gene and its associated protein. The gene has 18 exons and most of the mutant alleles are located in exons 16 and 17. The end of exon 16 and beginning of exon 17 code for the section of the protein that becomes concentrated in the plaques. APP is broken down by a series of enzymes called secretases. One of these–β-secretase–slices APP into two parts (see Figure 10.2). Then, a second enzyme–γ-secretase–cleaves one of these fragments, but it can do so at one of two different locations (Figure 10.2). Slicing at the first site will result in a fragment that is called the Aβ40 peptide (for amyloid beta with 40 amino acids). The second cleavage site is close by and
results in a sequence of 42 amino acids called Aβ42. While both fragments can appear in plaques, it is Aβ42 that preferentially conglomerates in them. The mutant alleles that cause AD result in an increased ratio of Aβ42 to Aβ40.

10.5.1.2 The presenilin genes

Shortly after the discovery of the APP gene, other positive linkage results found a gene on chromosome 14. The gene has since been located and is called the presenilin 1 locus (PSEN1). A second locus on chromosome 1, the presenilin 2 gene (PSEN2), has been implicated in AD in a group of German pedigrees originating from the Volga Valley in Russia (Blacker & Tanzi, 1998; Pericak-Vance et al., 2000; St. George-Hyslop, 2000).

Over 100 mutations in the presenilin-1 gene (PSEN1) and over a dozen in presenilin-2 have been identified as Mendelian forms of AD (Kanzi, X.X). Like APP, the mutant alleles associated with AD are dominant and have high penetrance. What do these two genes do? Well, interestingly, they produce proteins that comprise the γ-secretase enzyme complex—the very enzyme that cleaves the amyloid precursor protein into the those “sticky” fragments that form the plaques! Also, just the APP mutations that cause AD, the mutations in the two presenilin genes result in an increased ratio of the less soluble Aβ42 fragment relative to the more soluble Aβ40 residue.

10.5.1.3 Perspective on Mendelian forms of AD

Together, the three Mendelian forms account for almost half of the families with FAD. But because FAD accounts for no more than 10% of all Alzheimer’s cases, the three genes are responsible for roughly 5% of all AD. This is an important lesson from AD and is an emerging pattern in the study of medical DCGs—Mendelian forms occur, but they are rare and account for only a few percent of all cases.

A second lesson—actually, a problem—from AD is the possibility of predicting who will develop a Mendelian form of AD. A person with AD in a pedigree for FAD could be tested to find if s/he has a deleterious locus at one of the three genes. Under current medical genetic practice, decisions about whether or not to test reside with the patient. Nevertheless, thorny ethical issues emerge. Suppose that the grandfather in an FAD pedigree tests positive for the presenilin 1 mutation, but his son Fred and his granddaughter Sally are unaffected. Fred is very adamant that he not be tested or given any other information about whether he carries the gene. Sally, on the other hand, is equally adamant about being tested so that she can plan her life according to the results. If Sally is tested and tests negative, there is no problem—the probability that her father Fred has the gene remains at .50. However, if Sally tests positive, then Fred must have the gene because that is the only way Sally could have gotten it. If Sally’s test results in this case become known within the family, Fred’s decision to remain uninformed is violated. In short, testing some people within a pedigree could violate issues of doctor-patient confidentiality of other people in the
pedigree. A third lesson of AD concerns the importance of finding a gene. You may quite rightly question the importance of isolating these three genes when they account for less than 1 in twenty Alzheimer’s cases. The major importance may actually be in the ability to learn more about the neurobiology and the disease process (AKA pathophysiology) of AD than in shouting to the world “Researcher Finds AD Gene.” This is such an important topic that a whole section is devoted to it later (Section X.X).

10.5.2 A major locus: the apolipoprotein E gene

It turns out that there is another gene, the APOE locus (for apolipoprotein E) that contributes to AD, especially the late onset form but its role is different fashion than the three Mendelizing forms. The three Mendelian forms are almost fully penetrant—if a person has the deleterious allele at any one of these three loci, then the probability of developing AD is close to 1.0. In contrast, the APOE locus is viewed as a susceptibility locus—i.e., it increases the likelihood of developing AD but does not guarantee that AD will occur. When a susceptibility gene has a relatively large effect on the phenotype, then it is often called a major locus.\footnote{There is not sharp distinction between a Mendelian form with moderate penetrance and a major locus. Indeed, some AD researchers consider the APOE gene a Mendelian form with reduced penetrance.}

The APOE locus codes for a protein that assists in the transportation of cholesterol throughout the body. It has long been of great interest to cardiology because of its role in heart disease, but the discovery of its association with AD was largely fortuitous. There are three alleles at the locus that are formally designated as ε2 (i.e., epsilon 2), ε3, and ε4 but are also called E2, E3, and E4, a notation that will be followed herein. E3 is the most frequent allele and E2 is the rarest, but the frequencies vary across populations. Among populations of European ancestry, the rounded-off frequencies are around .10 for E2, .75 for E3, and .15 for E4 (Gerdes et al., 1992).

The E2 allele acts as a protective factor against AD while the E4 allele increases the risk for AD in a dose dependent way—i.e., the E4 homozygote is at higher risk than the E4 heterozygote (Farrer et al., 1997; Rubinsztein & Easton, 1999). The E3/E4 heterozygote has about 3 time the risk of the E3/E3 homozygote and the E4 homozygote, around 8 times the risk of developing Alzheimer’s (Loy et al., ), but actual risk figures may vary according to ethnicity. The deleterious influence of E4 appears to be more pronounced in some Asian, Indian, and white populations than in African-Americans (Farrer et al., 1997; Ganguli et al., 2000; Tang et al., 1996). Among late onset cases, the E4 allele may also be associated with an earlier onset and more rapid course of illness. However, not everybody with the E4 allele develops AD, and many people without an E4 allele also develop AD. This is the reason why the APOE locus is referred to as a susceptibility or risk factor gene.
What are the roles of APOE in AD? Most suspect that they involve the protein’s role in degrading and clearing the Aβ fragments from the brain. Here, the E4 allele does not appear to work as efficiently as E2 and E3 in getting rid of Aβ (Holtzman et al., 2014; Jiang X.X. 2008). APOE may also influence other mechanisms such as cholesterol homeostasis and neural inflammation (Liu et al., 2013 X.X).

Testing for APOE is useful in research settings and in some clinical settings where a diagnosis of AD is already suspected. But should there be general population screening for APOE? This is still a controversial issue. Unlike many medical tests, knowledge of the APOE genotype in an otherwise healthy person is neither diagnostic nor prognostic. All statements about APOE are probabilistic, so the most a clinician could tell a person about his/her risk for AD is something akin to “maybe or maybe not.” Enthusiasm for general testing is also dampened by the fact that there are no known ways to prevent AD. If an effective but costly prevention were developed, then it would make sense to genotype healthy individuals and devote scarce resources to those at highest risk of developing AD. A further issue muddling the debate is confidentiality. If accepted medical practice were to screen, then insurance carriers could know the results with untoward consequences for those at risk for AD. Hence, although there are no laws preventing a person from getting genotyped, most public policy statements by medical groups and patient-advocacy associations do not recommend screening at the present time. However, the same public policy statements recognize that screening may eventually be useful as science progresses and issues of confidentiality and insurability are resolved.

10.5.3 Polygenes
Genome-wide association studies (GWAS) have begun to identify genes that contribute to AD, almost all to LOAD—the late onset variety.

10.5.4 Twin studies
There have been several twin studies of Alzheimer’s

10.5.5 Genetics of Familial Alzheimer’s Disease
The fact that cases of Down’s syndrome (which involves the inheritance of extra material from chromosome 21) eventually show the AD-like pathology of plaques and tangles focused early linkage and association studies on chromosome 21. Soon it was found that a small minority of FAD cases was due to mutations in the gene for the amyloid precursor protein (APP) on this chromosome. Shortly afterwards, other positive linkage results were found to a gene on chromosome 14. The gene has since been located and is called the presenilin 1 locus (PSEN1). A third locus on chromosome 1, the presenilin 2 gene (PSEN2), has been implicated in AD in a group of German pedigrees originating from the Volga Valley in Russia (Blacker & Tanzi, 1998; Pericak-Vance et al., 2000; St.
George-Hyslop, 2000). These three genes can be called Mendelian forms of AD. That is, they are single genes that will cause the disorder regardless of other genes and environmental factors.

The three loci—APP and presenilins 1 and 2—all show dominant transmission, so the risk to an offspring is 0.50. Together, they account for almost half of the families with FAD. But because FAD accounts for no more than 10% of all Alzheimer’s cases, the three genes are responsible for roughly 5% of all AD. This is an important lesson from AD and is an emerging pattern in the study of DCGs —Mendelizing forms occur, but they are rare and account for only a few percent of all cases.

A second lesson—actually, a problem—from AD is the possibility of predicting who will develop a Mendelizing form of AD. A person with AD in a pedigree for FAD could be tested to find if s/he has a deleterious locus at one of the three genes. Under current medical genetic practice, decisions about whether or not to test reside with the patient. Nevertheless, thorny ethical issues emerge. Suppose that the grandfather in an FAD pedigree tests positive for the presenilin 1 mutation, but his son Fred and his granddaughter Sally are unaffected. Fred is very adamant that he not be tested or given any other information about whether he carries the gene. Sally, on the other hand, is equally adamant about being tested so that she can plan her life according to the results. If Sally is tested and tests negative, there is no problem—the probability that her father Fred has the gene remains at 0.50. However, if Sally tests positive, then Fred must have the gene because that is the only way Sally could have gotten it. If Sally’s test results in this case become known within the family, Fred’s decision to remain uninformed is violated. In short, testing some people within a pedigree could violate issues of doctor-patient confidentiality of other people in the pedigree.

A third lesson of AD concerns the importance of finding a gene. You may quite rightly question the importance of isolating these three genes when they account for less than 1 in twenty Alzheimer’s cases. The major importance may actually be in the ability to learn more about the neurobiology and the disease process (AKA pathophysiology) of AD than in shouting to the world “Researcher Finds AD Gene.” It is helpful to digress here for a moment to see how the gene discoveries have assisted in the understanding of AD. Again, the details are unimportant—it is the logic of the scientific quest that requires careful attention.

The locus for amyloid precursor protein (APP) on chromosome 21 codes for a polypeptide that eventually gets cleaved into certain sections. One of the longer fragments is chemically “sticky,” so these fragments bind together (i.e., form multimers) and join with other chemicals. The long fragments were found to be heavily concentrated in the plaques and tangles of AD brains. Several of the mutations in the APP locus result in an increased proportion of the longer fragment. Also the effects of the two presenilin genes appeared to be associated with increased production of amyloid, especially the longer fragment. These findings (along with many others not reviewed here) have coalesced into the “amyloid cascade hypothesis” that holds that AD (or at least certain cases
of AD) is due to some combination of amyloid overproduction and failure to appropriately break down and get rid of the longer beta-amyloid fragments (Fine, 1999; Neve & Robakis, 1998; Small, 1998).

This working hypothesis has guided research in many different directions, three of which will be mentioned here. The first is in the area of therapeutics. Scientists are now exploring ways to interfere in the normal process of amyloid degradation so that the longer, “sticky” fragment can get broken down more easily and are also investigating ways to digest the amyloid deposits in the plaques. A second area of research is into the genetics and the enzymology of those enzymes (called proteases) that cleave APP. Finally, researchers have placed mutant human APP alleles and mutant presenilin alleles into mice and found that these transgenic mice strains show similar plaques and tangles as AD. This is an especially exciting development because an animal model permits research and drug testing that cannot be done with humans (Janss & Westway, 2001).

At the same time, work on basic neurobiology is guiding genetic research. The neurofibers that become tangled in AD may be viewed as flexible railroad tracks that are prevented from tangling because they are spiked into a series of railroad ties. In the neuron, the chemical equivalents of the railroad ties are tau proteins, so the genes for tau proteins are being screened to see if mutations here may play a role in AD.

10.5.5.1 The Apolipoprotein E Locus

The 3 dominant genes for FAD account for a small fraction of cases. What is responsible for all the other cases of AD—that half of FAD that is not due to the 3 loci and all of sporadic Alzheimer’s? It turns out that there is another gene, the APOE locus (for apolipoprotein E) that contributes to AD, especially the late onset form but it is expressed in a different fashion than the three Mendelizing forms. The three dominant loci are close to being fully penetrant—if a person has the deleterious allele at any one of these three loci, then the probability of developing AD is close to 1.0. In contrast, the APOE locus is viewed as a susceptibility locus—i.e., it increases the likelihood of developing AD but does not guarantee that AD will occur.

The APOE locus codes for a protein that assists in the transportation of cholesterol throughout the body. It has long been of great interest to cardiology because of its role in heart disease, but the discovery of its association with AD was largely fortuitous. There are three alleles at the locus that are formally designated as ε2 (i.e., epsilon 2), ε3, and ε4 but will be called E2, E3, and E4 herein. E3 is the most frequent allele and E2 is the rarest, but the frequencies vary across populations. Among populations of European ancestry, the rounded-off frequencies are around .10 for E2, .75 for E3, and .15 for E4 (Gerdes et al., 1992).

The E2 allele acts as a protective factor against AD while the E4 allele increases the risk for AD in a dose dependent way—i.e., the E4 homozygote is at higher risk than the E4 heterozygote (Farrer et al., 1997; Rubinsztein &
Easton, 1999). The E4 heterozygote has about 2 to 3 times the risk and the E4 homozygote, around 10 times the risk of developing Alzheimer’s, but actual risk figures may vary according to ethnicity. The deleterious influence of E4 appears to be more pronounced in some Asian, Indian, and white populations than in African-Americans (Farrer et al., 1997; Ganguli et al., 2000; Tang et al., 1996). In late onset AD, the E4 allele may also be associated with an earlier onset and more rapid course of illness. However, not everybody with the E4 allele develops AD, and many people without an E4 allele also develop AD. This is the reason why the APOE locus is referred to as a susceptibility or risk factor gene. Because it has a relatively large effect, it is sometimes called a major locus or major gene.

Testing for APOE is useful in research settings and in some clinical settings where a diagnosis of AD is already suspected. But should there be general population screening for APOE? This is still quite a controversial issue. Unlike many medical tests, knowledge of the APOE genotype in an otherwise healthy person is neither diagnostic nor prognostic. All statements about APOE are probabilistic, so the most a clinician could tell a person about his/her risk for AD is something akin to “maybe or maybe not.” Enthusiasm for general testing is also dampened by the fact that there are no known ways to prevent AD. If an effective but costly prevention were developed, then it would make sense to genotype healthy individuals and devote scarce resources to those at highest risk of developing AD. A further issue muddling the debate is confidentiality. If accepted medical practice were to screen, then insurance carriers could know the results with untoward consequences for those at risk for AD. Hence, although there are no laws preventing a person from getting genotyped, most public policy statements by medical groups and patient-advocacy associations do not recommend screening at the present time. However, the same public policy statements recognize that screening may eventually be useful as science progresses and issues of confidentiality and insurability are resolved.

What are the roles of APOE in AD? Most suspect that they involve the protein’s role in degrading and clearing the Aβ fragments from the brain. Here, the E4 allele does not appear to work as efficiently as E2 and E3 in getting rid of Aβ (Holtzman et al., 2014; Jiang X.X. 2008). APOE may also influence other mechanisms such as cholesterol homeostasis and neuroinflammation (Liu et al., 2013 X.X)

10.5.6 The Great Unknown

To summarize, there are four genes known to be involved in AD. Together, however, they account for only a small percentage of AD cases even within early onset AD. Twin studies suggest that genes play an important part in late onset AD (Bergem et al X.X; Gatz X.X) and GWAS are beginning to isolate important loci. Nevertheless, the etiology of the majority of cases is obscure, so the models for these are more on the theoretical than empirical level. There are four non-mutually exclusive models for the development of the majority of AD cases. They are: (1) environmental factors; (2) phenocopies; (3) the
multifactorial threshold model; and (4) heterogeneity. Most researchers suspect that all four play a role, but there is no consensus as to the relative importance of the them.

10.5.6.1 Environmental factors and phenocopies

We discuss environmental factors and phenocopies together because a pheno-copy may be considered an environmental factor of large effect. Twin studies suggest that MZ can be discordant for AD, so some environmental factors must be important. Today, there are several useful leads including severe head trauma, stroke, and other forms of brain damage (Johnson, Steward & Smith; Sivanandam) as well as cardiovascular and inflammatory syndromes (Sastre). Some of these risk factors may not be purely environmental. Genes, for example, contribute to vascular pathologies and hence may indirectly influence the risk for AD.

Currently, there is considerable research on the role of head injury. It has been clear for some time that repetitive brain trauma in some boxers (as well as some football players and soldiers injured in battle) can lead to a syndrome of cognitive deficits resembling the early stages of AD. Called in the past dementia pugilistica, it is now diagnosed as chronic traumatic encephalopathy (CTE). It is not clear, however, how specific the trauma is to AD as opposed to more general brain pathology. Here, animal studies may clarify the situation.

Usually a phenocopy is defined as an environmentally produced phenotype that resembles a genetic syndrome. For classic Mendelian disorders, several intrauterine effects can produce a neonate with physical features resembling a known Mendelian disorder. For example, an autosomal dominant mutation can lead to Holt-Oram syndrome, which results in incomplete and deformed limb development. The drug thalidomide, if taken during a sensitive period of pregnancy, will produce similar malformations.

The term phenocopy is chauvinistic. One could just as well invent the word genocopy to denote a genetic abnormality that results in the same phenotype as a well-described environmental syndrome. Holt-Oram syndrome might be regarded as a genocopy of thalidomide toxicity. The terminology becomes even more confusing when the concept of a phenocopy is applied to a DCG. Ingestion of large amounts of lead, arsenic and other heavy metals will lead to intellectual deficiency in an otherwise normal person. Calling this a “phenocopy” carries with it a tacit assumption that mental retardation is a genetic syndrome.

To avoid a useless, hair-splitting discussion about terminology, the following is suggested. The definition of a phenocopy given above is the one encountered in all textbooks on human and medical genetics. In practice, however, most researchers who work with DCGs that involve behavioral phenotypes (learning disabilities, psychopathology) use the term in a much looser sense. Here, a phenocopy refers to an environmental insult that will produce the syndrome or disorder in anyone regardless of their genotype. According to this definition, a psychosis resulting from prolonged amphetamine abuse may be regarded as a phenocopy of schizophrenia or brief reactive psychosis without any implication
that either of the disorders are 100% genetic syndromes. It is this loose definition that will be used herein.

In terms of Alzheimer’s disease, there are no firmly documented phenocopies. One cannot, however, rule out the possibility that, say, some forms of repeated and severe closed head injury may end up being a phenocopy.

10.5.6.2 Multifactorial Transmission: Genetics

Many traits are the result of the combined effects of several different gene products (peptides, proteins, and/or enzymes) where each product is the result of the DNA blueprint at a different locus. When the effects of several different genes add and/or interact together to produce a phenotype, then the mode of transmission is called oligogenic (when the number of genes is on the small side) or polygenic (when the number of genes is large). There is not an exact number that separates oligogenic from polygenic transmission except for a rough order of magnitude—six genes are clearly oligogenic while sixty genes are obviously polygenic, but sixteen genes could be either oligo- or polygenic. There should really be no debate about the extent to which DCG are oligogenic or polygenic. At present, we lack the technology to determine the number loci that contribute to, say schizophrenia or major depression. Speculation as to that number is more akin to medieval scholastic arguments about angels and pinheads than contemporary science. In the remainder of this text, the word “polygenic” will be used to avoid the awkward phrase “oligogenic or polygenic” with the tacit implication that the number of genes involved need not be large.

It is helpful to give a hypothetical illustration of polygenic inheritance. Consider a continuous phenotype such as height and assume that three different genes contribute to individual differences in height. Denote the three loci as the A, B, and C locus and let there be two alleles, denoted by an upper case and lower case letter, at each locus. At the A locus, assume that allele $A$ adds 1.9 cm to height while allele $a$ subtracts 1.9 cm from height. Hence, genotype $AA$ having two $A$ alleles will add 3.8 cm to height; genotype $Aa$ adds 0 cm to height; and genotype $aa$ adds -3.8 cm to height. Assume that allele $B$ adds .6 cm to height and $b$ subtracts .6 cm, and allele $C$ adds .3 cm while $c$ subtracts .3 cm.

The genotypes in the population will contain all combinations of the alleles at these loci—e.g. $AABbCc$, $AaBBcc$, etc. There will be 27 different genotypes. Assume that the influence of each locus adds together. Then the individual differences in height resulting from these three loci are given in Table 10.1. To find the predicted height of a person, take the average height in the population and then add the amount in Table 10.1 corresponding to the person’s genotype. For example, if mean height were 176.2 cm, then a person with genotype $AABbCC$ would be predicted to be $176.2 + 4.4 = 180.6$ cm tall.

Now it is time to learn some terminology from this example. Note that the three loci do not have equal effects. The ultimate influence of locus A on height is greatest and the influence of locus C is least. This would be called a weighted polygenic system. The alternative, an unweighted polygenic system, assumes
Table 10.1: Example of a polygenic model: Hypothetical effects of genotypes on height.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Amount</th>
<th>Genotype</th>
<th>Amount</th>
<th>Genotype</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>AABBCC</td>
<td>+5.6</td>
<td>AaBBCC</td>
<td>+1.8</td>
<td>aaBBCC</td>
<td>-2.0</td>
</tr>
<tr>
<td>AABBCc</td>
<td>+5.0</td>
<td>AaBBCc</td>
<td>+1.2</td>
<td>aaBBCc</td>
<td>-2.6</td>
</tr>
<tr>
<td>AABBcc</td>
<td>+4.4</td>
<td>AaBBcc</td>
<td>+0.6</td>
<td>aaBBcc</td>
<td>-3.2</td>
</tr>
<tr>
<td>AABcCC</td>
<td>+4.4</td>
<td>AaBcCc</td>
<td>+0.6</td>
<td>aaBcCC</td>
<td>-3.2</td>
</tr>
<tr>
<td>AAbCc</td>
<td>+3.8</td>
<td>AabCc</td>
<td>+0.0</td>
<td>aabCc</td>
<td>-3.8</td>
</tr>
<tr>
<td>AAbbCC</td>
<td>+3.2</td>
<td>AabbCC</td>
<td>-0.2</td>
<td>aabbCC</td>
<td>-4.4</td>
</tr>
<tr>
<td>AAbbCc</td>
<td>+2.6</td>
<td>AabbCc</td>
<td>-1.2</td>
<td>aabbCc</td>
<td>-5.0</td>
</tr>
<tr>
<td>AAbbcc</td>
<td>+2.0</td>
<td>Aabbc</td>
<td>-1.8</td>
<td>aabbc</td>
<td>-5.6</td>
</tr>
</tbody>
</table>

that each locus contributes the same amount to the phenotype; the unweighted model is important for theoretical genetics but is implausible in nature.

In the weighted system, locus \( A \) would be called a major gene or major locus in the sense that it has a relatively large influence on the phenotype. It would appear natural to refer to locus \( B \) and locus \( C \) as minor loci, but in the twisted logic of scientific language, the terms modifying genes, modifying loci, or even background genes are preferred. These are loci that have a relatively small quantitative effect on the phenotype.

Note how the effects of the alleles and the loci added together. For example, consider genotype \( AABBcc \). There are two \( A \) alleles, each contributing +1.9 cm, one \( B \) allele (adding .6 cm) and one \( b \) allele (subtracting .6 cm), and two \( c \) alleles, each of which subtract .3 cm. Hence, the deviation for this genotype is \( 2 \times 1.9 + (.6 - .6) + 2 \times (-.3) = +3.2 \) cm. Hence, this is called an additive model. Once again, the additive model is an assumption and not a necessary consequence of nature. Departures from the strictly additive model occur when there is any degree of dominance or epistasis (i.e., statistical interaction among several genes) and will be discussed in a later chapter.

10.5.6.3 Multifactorial Transmission: Environmental Factors

Recall that the phrase “predicted height” was used to describe the expected height of someone with a particular genotype. For example, someone with genotype \( AABbCC \) is predicted to be 108.6 cm. An actual person with this genotype may not be exactly 108.6 cm tall because various environmental factors could influence the phenotype. It is often assumed that several different environmental factors may contribute to the phenotype, so not only are there multiple genes but there are also multiple environmental factors that contribute to disorder. A classic example of multiple environmental factors influencing a DCG would be the role of smoking, exercise, and diet on cardiovascular disease.

When there are several genes and several different environmental factors contributing to a trait, then the model is usually referred to as multifactorial. The
phrase polygenic is also used instead of multifactorial with the tacit assumption that environment may play a role in addition to genes.

10.5.6.4 The Multifactorial Threshold Model

The multifactorial threshold model applies polygenic transmission and several etiological environmental factors to a discrete or dichotomous phenotype. A discrete or dichotomous phenotype is one that can be classified into two or more groups with no intermediate phenotypes between the groups. Pregnancy is a classic discrete phenotype—women are either pregnant or not pregnant and it is impossible to be in between. Note that a discrete phenotype may still have sub-phenotypes that are graded continuums. For example, length of pregnancy is an important continuous variable among the category of pregnant women.

Polydactyly (i.e., having an extra toe on a foot) in mice is a classic example of the multifactorial threshold model. Ordinarily mice have four digits on a paw, but under some circumstances, a fifth digit may appear. The extra digit may be poorly formed and nonfunctional, but in terms of phenotypic classification, it is an all or none trait. Although mutations in a few single genes can cause the trait, usually both genes and environment (especially prenatal stress) contribute to this phenomenon.

Originally developed by Falconer (1967), the multifactorial threshold model was first applied to human behavior by Gottesman and Shields (1967, 1972) in their classic twin study of schizophrenia. It has since become a very popular model of the transmission of psychiatric disorders and psychopathology and a version of it is depicted in Figure 10.3. The key concept of this model is liability (aka susceptibility, vulnerability, or predisposition). Liability is not the disorder per se. Instead, liability is defined as an unmeasured (i.e., latent) continuous variable that probabilistically relates to the disorder. The higher one is in liability, the greater the chances of developing the disorder.

Genes contribute to the genetic liability. For example, genotype $AABBbCcDd$ might contribute +1.3 units to genetic liability while genotype $aabbcCdd$ might contribute -2.3 units to genetic liability. Because genetic liability is unmeasured, any “units” are completely arbitrary. It is customary, however, to let 0 denote the mean liability, and let positive numbers denote risk (i.e., a greater propensity to develop the disorder) and negative numbers denote protection (a lower probability of developing the problem).

Environmental factors contribute
to environmental liability which has an analogous interpretation and measurement scale as the genetic liability but applies to environments instead of genes. Together, genetic and environmental liability contributes to total liability. It is the total liability that relates directly to the disorder. Once the total liability reaches a certain point (the threshold), then a disease process starts. Hence, individuals beyond the threshold develop the disorder while those below the threshold remain unaffected. Note how only the total liability correlates perfectly with the phenotype. Knowing a person’s genetic liability permits probabilistic statements of the form, “Joe is at high risk for bipolar disorder” or “Susie is at reduced risk for agoraphobia,” but does not allow one to predict with certainty that Joe will become bipolar and that Susie will never develop agoraphobia. Similarly, knowing a person’s environmental liability permits only probabilistic statements about the development of a disorder.

The threshold model is developed in more detail in a later chapter. Here we note that the important fact is that the model is once again a set of assumptions about how nature works. It is not the case that all the genetic and environmental factors for a number of disorders have been clearly elucidated and the multifactorial threshold model emerged as a parsimonious mathematical explanation for the empirical data.

\textbf{10.5.6.5 Varieties of multifactorial models}

Following Gibson (Rare and common variants: Twenty arguments, Nature Genetics 13:136-145, 2012), we identify three flavors of the multifactorial model for DCGs. The first is called the \textit{common disease, common variant} (CDCV) model. This model assumes that most alleles that confer risk for a DCG are common in the population, “common” usually being defined as having a prevalence of 1\% or greater\(^2\). It is usually further assumed that the effect of any single risk allele is very small and so that the genetic risk comes from the summation of a large number of these small effects.

The second model does not have a standard name so we call it the \textit{rare, large effect variant} (RLEV) model. This model assumes that risk alleles are rare (i.e., less than 1\%) but have a much larger effect on disease risk. In an extreme case, a risk allele (or risk genotype) may has such a large effect that it qualifies as a major locus. In most cases, however, genetic risk comes from the summation over a loci. In the RLEV model, however, that summation is over a much smaller number of loci.

The third model, also lacking a consensus name, may be termed the \textit{non additive, interactive, epigenetic} (NAIE) model. This model begins with the assumption that the additive effects of the CDCV model and the large effects of rare alleles are insufficient to explain genetic risk. The unpredicted risk,

\(^2\)Some geneticists use a cut off of 0.1\%.
however, can be accounted for by a slew of non-additive factors such as dominance, gene-gene-interaction (aka epistasis), and gene-environment interaction. (Gibson, 2012, p. 136).

Which model best fits the GWAS data? The answer is clear—none of the above. One suspects that the models are useful heuristically in terms of making predictions but that the real world turns out to be complicated. The first noteworthy point is that these models blend together rather than forming independent, non mutually exclusive paradigms. For example, schizophrenia might closely fit the CDCV model while early onset diabetes might best fit the RLEV model only because there are slight differences in the number of loci and the allelic effects underlying these traits. The underlying biological properties of the gene-phenotype relationship in schizophrenia and early onset diabetes could be very similar.

Second, time debating the merits of one model versus the other may be better spent identifying the genes contributing to a DCG. Once identified, the more productive approach would focus on the HOW (see Chapter X.X and Anastasi, 19xx)—how do these loci contribute to risk.

Finally, it is plausible that the genetic architecture for any DCG combines elements of each model, albeit in different proportions. At the end of the day, CDCV loci may contribute to, say, 54.3% of the genetic risk for schizophrenia, RLEV loci to 28.7% of the genetic risk, rare recessives (which create dominance variance) to 7.3%, interactions between prenatal viral infections and genotypes to 4.3%, and so on. Again, ask yourself, “What is more important? Seeing if the statistical impact of genes fits a model or identifying the genes and the mechanisms of their action?” In short, the HOW (Anastasi, ) wins once again.

10.5.6.6 Heterogeneity

In heterogeneity, there are different subsets of causal factors each of which is sufficient to cause AD. A more prosaic description would be that AD is really a collection of different disorders that end in a common syndrome. There are many examples of heterogeneity for simple Mendelian disorders—albinism, congenital adrenal hyperplasia (see Section X.X) and elliptocytosis are classic examples. Diabetes is probably the best example of a heterogeneous DCG.

To see how heterogeneity differs from the multifactorial model, suppose that LOAD can result from two different processes—cholesterol and the immune system. Under heterogeneity, if a person has sufficient risk in either the cholesterol or immune factors, then she will develop AD. It would be as if there were two multifactorial threshold models, one for cholesterol risk and the other for the immune factors, and if the threshold for either of the two is passed, then the AD process will begin. In a strict polygenic model, risk is a function of the addition of the two system—the person’s risk on cholesterol is added to the risk on the immune variables.

In the ordinary polygenic model, risk on one dimension can be compensated for by protective factors on the other dimension. For example, someone with very high risk on cholesterol but very low in the immune system will not de-
velop AD. The two cancel each other out. In the heterogeneity model on the other hand, that person would develop AD–cholesterol puts the person over the threshold and there is nothing that the immune system can do about that.

10.6 Lessons from GWAS

10.7 Animal models

Perhaps the most important implication of detecting Mendelian forms for a disorder has been to develop transgenic animal models of the disease. A transgenic animal is one in which a gene from another organism, often from another species entirely, is introduced. Many of the known allelic variants of the APP, PSEN1 and PSEN2 genes that cause AD in humans have been implanted in rodents (usually mice) yielding over two dozen transgenic strains (Hall & Roberson X.X, LaFerla X.X, McGowan X.X).

It is also possible to develop animal models where certain AD pathology can be induced and then corrected. For example, Hochgrafe, Sydow & Mandelkow X.X have developed transgenic mice that have various forms of human tau proteins. By manipulation the areas around the promotor regions for these genes, they can induce the correct taus to produce neurofibrillary tangles. Other manipulations can then stop this process.

Depending on the gene and allele, transgenic strains exhibit the Aβ plaques and/or neurofibrillary tangles characteristic of AD as they age. Most strains also show cognitive deficits. The transgens have been invaluable in elucidating mechanisms involved in plaque and tangle formations. They have also been useful in testing potential therapeutics, although to date no drug has been developed that successfully combats the disorder.

Despite their promise, there are limitations to animal models. “None of the existing models fully reproduces the complete spectrum of this insidious human disease” (LaFerla & Green X.X). In addition, most of the genes introduced into animals are for the familial form of the disease. Most human cases, in contrast, are LOAD–late onset with a lower familial loading.

10.8 Overall Perspective of DCG

If we join the empirical evidence about the genetics of AD along with the speculations about polygenic transmission and the multifactorial threshold model, we arrive at the overall model proposed by Gottesman (1991) about the mode of genetic transmission for schizophrenia. Figure 6.2 presents an adaptation of this model but will be discussed here in terms of AD.
Figure 10.4: Theoretical pathways for LOAD from GWAS.

Figure 1. Pathways implicated in Alzheimer's disease (AD). Familial Alzheimer's disease (FAD) is caused by one of three genes (PSEN1, PSEN2 and APP) involved in the amyloid processing pathway ("the amyloid cascade"). There is no single pathway that explains the more common nonfamilial, late-onset/sporadic Alzheimer’s disease (LOAD/sAD). Candidate genes are enriched for several pathways; Cholesterol Metabolism, Immune System and Endocytosis. These pathways may be downstream dependent or independent of amyloid processing.
The figure partitions the etiology of AD according to its major (but not necessarily exclusive) causes. There are three Mendelian forms of the disorder, corresponding to the APP, presenilin 1, and presenilin 2 mutations. Although these three forms have high penetrance, they account for only a small percent (4% in the cases of Figure 6.2) of all AD.

There are two major loci that contribute to AD. The first might be the APOE locus that is responsible for 8% of cases. This figure means that in 8% of AD, the APOE locus is the major cause for the disorder even though there may be many other genes and environmental factors that also contribute to the onset among this 8%. Note well that the 8% figure is completely hypothetical—we do not know how many AD cases have APOE as their major cause. In the future a second major locus may be uncovered that also contributes to susceptibility. This eventuality is portrayed in Figure X.2 as major locus 2 that is the major cause for 15% of cases.

The majority of cases—62% in Figure 6.2—may be due to the effects of many different loci and a variety of environmental factors that combine together and influence risk. This is the multifactorial threshold model that was discussed earlier and is denoted as simply “Multifactorial” in the figure. Once again, the estimate of 62% is fictitious. Most contemporary geneticists suspect that the multifactorial transmission will eventually account for the bulk of any DCG, but no one really knows that for certain.

Lastly, phenocopies account for the remaining 11%. Gross insults to the brain through trauma, viral infections, or a number of yet-unknown environmental factors might produce AD in people who are otherwise at low risk for
AD. Again, the percentage here is complete, uneducated guesswork. Taken together, this model implies that there is no single “cause” of AD. Rather, there are a number of different events, some large and some small, that lead to the same process or very similar processes. Just as many roads lead to Rome, there are different pathways to AD and the road traveled by one person may be very different from that taken by another. Some roads are superhighways (e.g., the three Mendelizing forms) that get one there in a hurry, but the majority are minor streets with twists, turns, detours, and dead ends. But everyone who enters Rome develops AD. As a result, the causative factors in one person could be quite distinct from the etiological events in someone else.

This overall model has long been applied to complex behavioral phenotypes such as mental retardation, specific learning disabilities, and every form of psychopathology. In every case, however, empirical knowledge about the relative contributions of Mendelizing forms, major loci, multifactorial transmission, and phenocopies is very rudimentary. For mental retardation, there are indeed many documented Mendelizing forms—PKU and Fragile X syndrome are but two of over one hundred—and several recognized phenocopies such as severe lead toxicity. And the ALDH-2 polymorphism illustrates the role of a major locus, at least in some Asian populations. However, slicing the rest of the pie is primarily guesswork. To complicate matters, the division of the pie for one DCG may be quite different from that of another disorder.

10.9 The Final Lesson for DCG: Genetics is a Tool, Not a Goal

The progress being made in AD research should reinforce a major lesson from our study of Mendelian traits. Finding a gene or genes for any disorder or behavior is not the goal or holy grail of science. Instead, it must be looked upon as reaching a milestone in a long journey, the majority of which is still in front of us. All of us have had the experience of being on a long road trip and spending hours passing cornfield after cornfield, desert and desert, or town after town. When we arrive at a salient point—whether it be the halfway mark, an important geographical landmark, or a major city—our lethargy gives way to a momentary sense of excitement and renewal. But do we stop and consider the journey complete? Of course not! We trudge on until we arrive at the intended destination.

So it is with finding a gene for a disorder like AD or any other human behavior. It is an important accomplishment. It should be celebrated, and kudos should be bestowed on the discoverers. But finding a gene for schizophrenia, intelligence, extraversion is not the raison d’être of genetics. It is an important step in the long journey towards unraveling the biology behind a phenotype.

The importance of the Alzheimer’s story is the interplay between the genetics and basic neurosciences. Genetics has lead to the development of transgenic animal strains that are exceptionally important for basic neuroscience research.
In turn, the results of neurobiology will aid in the search for other genes that may influence the disorder. In short, for DCG, genetics is a tool, not an end in itself.

10.10 References