How to Use an Article About Genetic Association
A: Background Concepts

John Attia, MD, PhD

John P. A. Ioannidis, MD, PhD

Ammarin Thakkinstian, PhD

Mark McEvoy, MMedSc

Rodney J. Scott, PhD

Cosetta Minelli, PhD

John Thompson, PhD

Claire Infante-Rivard, MD, PhD

Gordon Guyatt, MD, MSc

This is the first in a series of 3 articles serving as an introduction to clinicians wishing to read and critically appraise genetic association studies. We summarize the key concepts in genetics that clinicians must understand to review these studies, including the structure of DNA, transcription and translation, patterns of inheritance, Hardy-Weinberg equilibrium, and linkage disequilibrium. We review the types of DNA variation, including single-nucleotide polymorphisms (SNPs), insertions, and deletions, and how these can affect protein function. We introduce the idea of genetic association for both single-candidate gene and genome-wide association studies, in which thousands of genetic variants are tested for association with disease. We use the *APOE* polymorphism and its association with dementia as a case study to demonstrate the concepts and introduce the terminology used in this field. The second and third articles will focus on issues of validity and applicability.

JAMA. 2009;301(1):74-81   www.jama.com

CLINICAL SCENARIO

A 55-year-old man consults you, worried about his risk of developing Alzheimer disease. His grandfather had dementia in his 70s, and his own father was diagnosed with dementia at 65 years. He has been a smoker since he was 20 years of age, works as an electrician, and has been taking antihypertensive medication (thiazide and β-blocker) for the last 5 years. He has never had his cholesterol level checked. He has recently read a news story about genetic tests and asks you whether he should have any genetic test for dementia risk, in particular for a gene called *APOE*.

THE SEARCH

Consulting your electronic medical texts, you find a section addressing genetic risk factors for dementia. You come across words such as *allele*, *genotype*, and *genetic association* and encounter many abbreviations, including *APOE*, e3/e4, and e4/e4 (Box). A colleague with an interest in genetics steers you toward a helpful Web site, the HuGE Navigator,[1] and you find that there are apparently more than

1000 publications on genetic associations of Alzheimer disease that discuss several hundred genes; the whole endeavor seems daunting. A quick Internet search reveals that many companies offer genetic tests for Alzheimer disease, often as part of testing hundreds of thousands of genetic variations. You realize that you need to review your basic genetics knowledge and discover more about how to read genetic association studies.

The Human Genome Project has greatly stimulated interest in genetic determinants of disease. The determinants of common mendelian diseases that involve a single gene (eg, cystic fibrosis, Huntington disease) are well established. Current research addresses the role of genetics in the chronic diseases composing the major causes of human morbidity and mortality, diseases that result from the concomitant effect of environmental, behavioral, and

Author Affiliations: Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, Hunter Medical Research Institute, and Department of General Medicine, John Hunter Hospital, Newcastle, Australia (Dr Attia and Mr McEvoy); Department of Hygiene and Epidemiology, University of Ioannina, School of Medicine, Ioannina, Greece, and Center for Genetic Epidemiology and Modeling, Tufts Medical Center, Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts (Dr Ioannidis); Clinical Epidemiology Unit, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand (Dr Thakkinstian); Division of Genetics, Hunter Area Pathology Service, John Hunter Hospital, New Lambton, Australia, and Centre for Information Based Medicine, Faculty of Health, University of Newcastle, Hunter Medical

Research Institute, Newcastle, Australia (Dr Scott); Respiratory Epidemiology and Public Health, National Heart and Lung Institute, Imperial College, London, England (Dr Minelli); Department of Health Sciences, University of Leicester, Leicester, England (Dr Thompson); Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montreal, Canada (Dr Infante-Rivard); and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada (Dr Guyatt).
Corresponding Author: John Attia, MD, PhD, Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, Level 3, David Maddison Bldg, Newcastle 2300, Australia (john.attia@newcastle.edu.au).
Users' Guides to the Medical Literature Section Editor: Drummond Rennie, MD, Deputy Editor, *JAMA*.

Box. Glossary

Additive

Describes any trait that increases proportionately in expression when comparing those with no copy, 1 copy, or 2 copies of that allele, ie, those with 1 copy of the allele show more of the trait than those without, and in turn, those with 2 copies show more of the trait than those with 1 copy

Allele

One of several variants of a gene, usually referring to a specific site within the gene

Candidate gene study

A study that evaluates association of specific genetic variants with outcomes or traits of interest, selecting the variants to be tested according to explicit considerations (known or postulated biology or function, previous studies, etc)

Chromosome

Self-replicating structures in the nucleus of a cell that carry the genetic information

Dominant

Describes any trait that is expressed in a heterozygote, ie, one copy of that allele is sufficient to manifest its effect

Genome

The entire collection of genetic information (or genes) that an organism possesses

Genome-wide association study

A study that evaluates association of genetic variation with outcomes or traits of interest by using 100 000 to 1 000 000 or more markers across the genome

Genotype

The genetic constitution of an individual, either overall or at a specific gene

Haplotype

Alleles that tend to occur together on the same chromosome because of SNPs being in proximity and therefore inherited together

Heterozygous

An individual is heterozygous at a gene location if he or she has 2 different alleles (one on the maternal chromosome and one on the paternal) at that location

Homozygous

An individual is homozygous at a gene location if he or she has 2 identical alleles at that location

Isoform

Variant in the amino acid sequence of a protein

Linkage

The tendency of genes or other DNA sequences at specific loci to be inherited together as a consequence of their physical proximity on a single chromosome

Linkage disequilibrium

A measure of association between alleles at different loci

Locus/loci

The site(s) on a chromosome at which the gene for a particular trait is located or on a gene at which a particular SNP is located

Messenger RNA

A ribonucleic acid–containing single-strand copy of a gene that migrates out of the cell nucleus to the ribosome, where it is translated into a protein

Mutation

A rare variant in a gene, occurring in <1% of a population; cf *polymorphism*

Pedigree

A diagram depicting heritable traits across 2 or more generations of a family

Phenotype

The observable characteristics of a cell or organism, usually being the result of the product coded by a gene (genotype)

Polymorphism

The existence of 2 or more variants of a gene, occurring in a population, with at least 1% frequency of the less common variant (cf *mutation*)

Recessive

Describes any trait that is expressed in a homozygote but not a heterozygote, ie, 2 copies of that allele are necessary to manifest its effect

Ribosome

The protein synthesis machinery of a cell where messenger RNA translation occurs

SNP

Abbreviation for single-nucleotide polymorphism, a single base pair change in the DNA sequence at a particular point compared with the "common" or "wild-type" sequence

Synonymous SNP

A SNP that does not lead to a change in the amino acid sequence compared with the common or wild-type sequence; cf *nonsynonymous*, in which there is a change in the amino acid sequence as a result of the SNP

Variant allele

The allele at a particular SNP that is the least frequent in a population

Wild-type allele

The allele at a particular SNP that is most frequent in a population, also called "common" allele

genetic factors. In the last year alone, major studies have tested hundreds of thousands of genetic variations ("polymorphisms"), trying to establish the genetic determinants of coronary artery disease,[2] type 2 diabetes,[3] stroke,[4] multiple sclerosis,[5] breast cancer,[6] bipolar disorder,[7] rheumatoid arthritis,[8] Crohn disease,[7] and Alzheimer disease.[9-11]
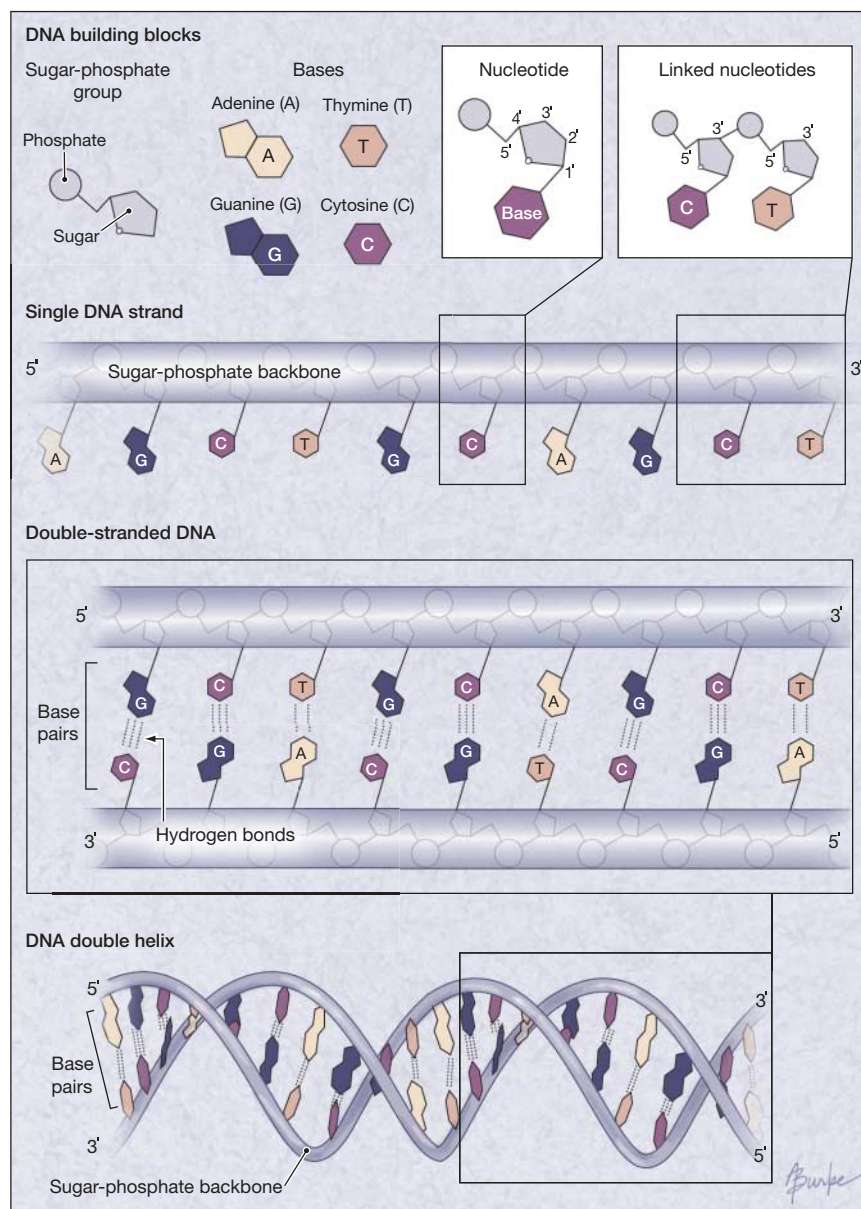
There are many reasons why a clinician may wish to understand the results of genetic association studies. This genetic information may shed light on pathways involved in disease and identify new targets for therapeutic interventions, advances that are of interest in basic and applied research. Genetics might be able to improve diagnosis or improve the ability to use therapeutic agents more efficiently and with less risk, an area termed *pharmacogenomics*. Although few genomic applications have reached this level of utility, despite the claims of companies marketing genetic tests, this might change in the future. The most likely short-term clinical effect of this genetic information is to enhance risk stratification and provide patients with information about their prognosis.

This series of 3 articles aims to provide clinicians with an introduction to reading and understanding genetic studies in medicine. This first article summarizes the key genetic concepts clinicians must understand to read articles offering genetic associations for possible clinical use; those familiar with basic genetics may wish to move straight to the second article. That article, along with the third, will focus on issues of validity and applicability fundamental to critical appraisal of such studies. Throughout the series, our presentation will be simplified but sufficiently detailed to allow readers to judge the relevant issues.

THE GENETIC BLUEPRINT

In 1953, James Watson and Francis Crick proposed a winding staircase (double helix) structure of DNA (**FIGURE 1**). The sides of the staircase or ladder, called *strands*, are formed by alternating sugar (deoxyribose) and phosphate molecules; the rungs of the ladder are formed by 4 nitrogen-containing ring compounds called *bases*: adenine (A), thymine (T), guanine (G), and cytosine (C). A pair of these bases forms each rung of the ladder; adenine always binds to thymine and cytosine always binds to guanine to form the full rung. Thus, each rung of the helix ladder is called a *base pair*. A single base plus its associated sugar and phosphate groups is called a *nucleotide*.

One long stretch of double-stranded DNA, forming the spiral staircase, constitutes 1 chromosome, on which there are many genes, a gene being a stretch of

Figure 1. Components and Structure of DNA

The building block of DNA is the nucleotide—a sugar (deoxyribose) with a phosphate group at the 5′ carbon and a base (adenine, thymine, guanine, or cytosine) at the 1′ carbon. Nucleotides link together by a bond between the phosphate group of one nucleotide and the 3′ carbon of the previous nucleotide, to form a single DNA strand with a resulting directionality of 5′ to 3′. Two strands with opposite directionality combine to form a double helix that is held together by hydrogen bonds across the bases. Adenine always binds to thymine and guanine always binds to cytosine. The sequence of base pairs encodes the genetic information.

DNA that typically codes for 1 protein. Twenty-three chromosomes in the sperm and the corresponding 23 chromosomes in the egg come together at fertilization to form the entire "DNA set" of a human, called the *genome*. Each person therefore has 23 pairs of chromosomes, of which 1 pair is the sex chromosomes that determine sex. The other 22 pairs are numbered 1 to 22, providing each person with 2 versions of each gene, one on the maternally inherited and one on the paternally inherited chromosome (**FIGURE 2**).

DNA is the blueprint for making proteins that build cells and tissues and enzymes that catalyze biochemical reactions within a cell. The information on the DNA guides the production of proteins through a 2-step process of transcription and translation.

The first step involves transcribing DNA into messenger RNA (mRNA) (**FIGURE 3**). The mRNA molecule migrates out of the nucleus to the cytoplasm to reach the protein-building machinery of the cell, known as the *ribosome*. Here, during the second step, the mRNA molecule is translated into protein, using the code to link amino acids one at a time. These processes of transcription and translation convert the genetic information at one gene, called the *genotype*, into the protein that, in concert with other genes, their proteins, and environmental exposures, determines the final attribute, the *phenotype* (eg, hair color, height, thrombophilia).

HUMAN VARIATION

Sequencing the human genome—identifying the entire sequence of base pairs in the 25 000 genes that constitute human DNA—has revealed that the sequence is more than 99% identical across different people.[12] However, the human genome includes 3.3 billion base pairs; thus, even with this high level of similarity, there are still more than 12 million potential variations between 2 people's genomes.[13-15] Differences in gene structure that occur infrequently in populations, ie, less than 1%, are called *mutations*, whereas differences that occur more frequently, ie, greater than or equal to 1%, are called *poly-*

morphisms (**FIGURE 4**). These polymorphisms may take a number of different forms:

1. the presence or absence of an entire stretch of DNA (insertion/deletion polymorphisms), a variation of which involves DNA duplication, called *copy number variation*, or CNV;

2. repeating patterns of DNA that vary in the number of repeats; each repeating "unit" may vary from 2 to 3 to hundreds of base pairs long and may repeat a few times to hundreds of times; and

3. a single–base pair change, called a *single-nucleotide polymorphism*, or SNP (pronounced "snip"). This is by far the most common polymorphism; scientists have cataloged more than 12 million SNPs to date.[13,15] Some SNPs are in parts of the gene that are translated, ie, code for protein: among them, non-synonymous SNPs lead to a change in amino acid sequence of the resultant protein, whereas synonymous SNPs do not result in amino acid change. Other SNPs are in areas of the chromosome that do not directly code for protein but may still influence cell function through other means, such as controlling the amount of the protein that the cell builds. Given the very large number of SNPs, their nomenclature can be confusing, but the most common system uses a number with the prefix "rs" (for *r*eference *S*NP), eg, rs1228756.

The different forms or variants that a particular polymorphism may take are called alleles. For example, the *APOE* gene in the clinical scenario has 3 alleles, named e2, e3, and e4 (*e* for epsilon; for historical reasons, there is no e1). The location along the DNA strand at which a particular allele is present is called a *locus*. In a feral moment, geneticists decided to call the form of the gene that is most common in the population the *wild type*; we will use the less colorful term *common* allele; and the less common allele(s), *variant* allele(s). The different alleles—in the case of SNPs, only the nonsynonymous SNPs—result in production of different forms of the protein for which the gene is responsible. These different proteins are called *isoforms*.

Of the *APOE* alleles, allele e3 is most common in white populations (78%) and thus represents the common allele; e2 and e4 are variant alleles (with a frequency of 6% and 16%, respectively, among white individuals). Physiologically, the apoE protein carries a form of cholesterol and binds to the apoE receptor on the surface of cells for the cholesterol to be metabolized. Of the 3 protein isoforms that result from the 3 corresponding *APOE* alleles, the e2 isoform has decreased strength of binding, or affinity, to the apoE receptor. The proteins resulting from the e3 and e4 alleles, the e3 and e4 isoforms, have higher affinity compared with e2. (For the sake of simplicity, we have presented the 3 alleles of the *APOE* gene, leading to the 3 isoforms, as if they are changes at one point in the DNA. In fact, these isoforms are defined by changes at 2 points in the DNA, leading to changes in 2 amino acids (positions 112 and 158) that make up the apoE protein: alleles e2, e3, and e4 at the DNA level lead to amino acids cysteine/cysteine, cysteine/arginine, and arginine/arginine at these 2 sites, respectively).

Because each person has 2 versions of each chromosome, one from the mother and one from the father, individuals have 2 *APOE* genes. The 2 relevant chromosomes (in the case of apoE, the 19th chromosome) may carry the e3 allele (denoted e3/e3) or 1 each of e3 and e2 alleles (denoted e2/e3), or any other combination, eg, e2/e4, e3/e3. This shorthand denotes the "genotype" of the individual. Individuals with 2 copies of the same allele are said to be *homozygous*, or a *homozygote*; ie, an individual with e3/e3 is homozygous for e3, or an e3 homozygote. Conversely, an individual with 2 different alleles, eg, e2/e3, is a *heterozygote*, or *heterozygous*.

Individuals with e2/e3 produce some e2 protein and some e3 protein. The question then follows: which protein function wins? If the variant allele is *dominant*, it need be present in only 1 of the 2 genes to carry all the biological activity. In such cases, the allele on the other chromosome will remain biologically "silent." Conversely, an al-

lele that is *recessive* will need to be present on both genes, ie, homozygous, to affect function; it will otherwise remain silent. If 2 differing protein isoforms resulting from 2 different alleles (eg, e2, e3) share function, the model of protein function is known as an *additive*, or *per-allele*, model. These dominant, recessive, and additive/per-

Figure 2. Human Male Karyotype, Chromosome Structure and Mapping, and Location of *APOE*
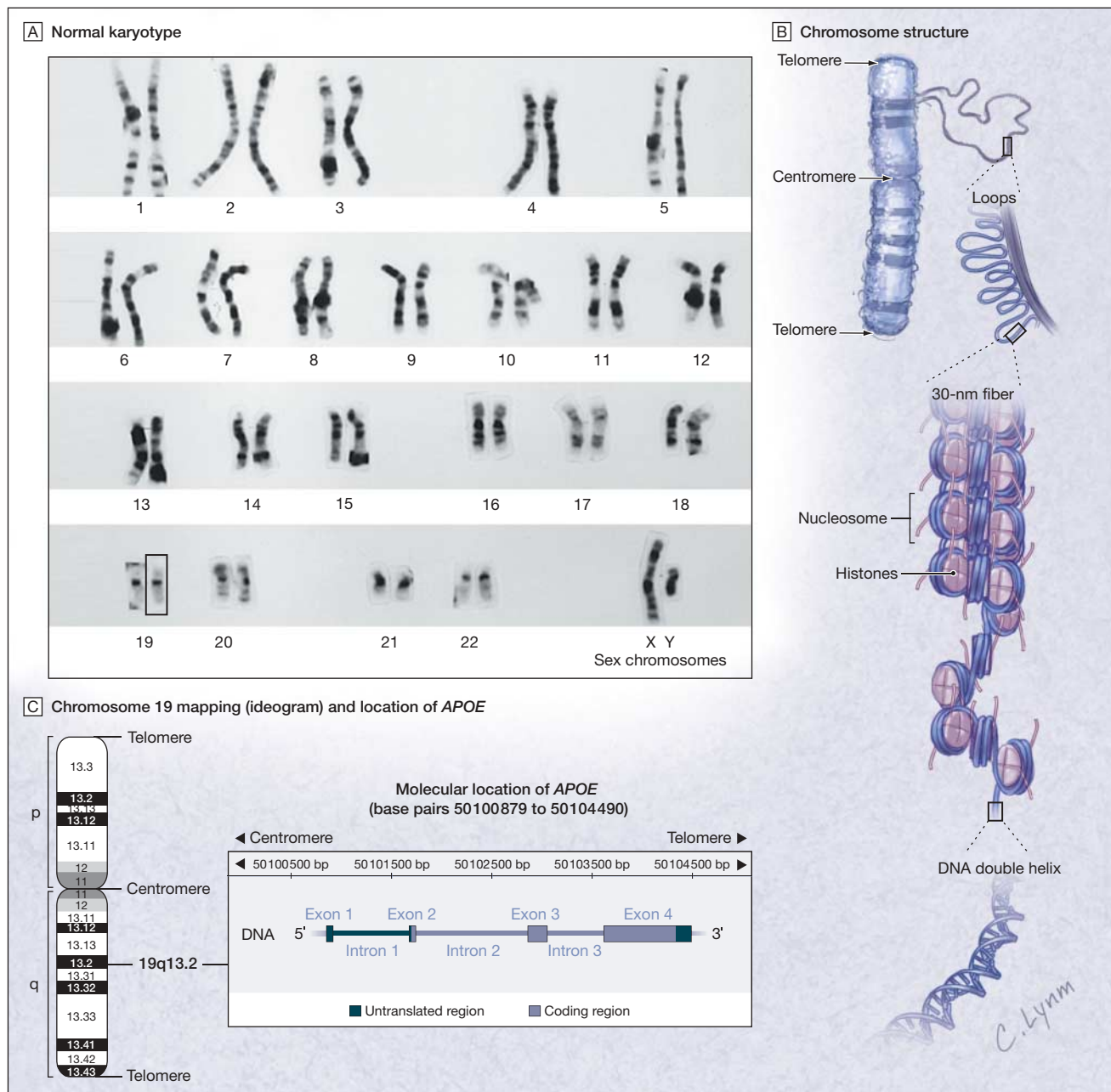
A, Typically, an individual has 23 pairs of chromosomes. One member of each pair is inherited from the mother and one from the father. Chromosomes shown in the karyotype were obtained when the cell was not dividing, stained using Giemsa, and ordered by size. B, The DNA double helix is wound around proteins called histones to form small packages called nucleosomes. The nucleosomes in turn are wound around themselves to form loops that make up the chromosome. The region of the chromosome near the center is called the centromere, and each end is called a telomere. C, The centromere is not exactly at the center of the chromosome, resulting in a shorter arm, named p for *petit* (French for small) and a longer arm, named q. Chromosome 19 is the site of the *APOE* gene, which is composed of sequences with regulatory functions (untranslated regions) and sequences with coding functions. Regions of the gene that are spliced out during transcription to messenger RNA are called introns. The remaining regions, exons, contain the sequences that code for the final protein product.

allele models are called *models of inheritance*, or *genetic models*.

As it turns out, the *APOE* alleles act pretty much in an additive manner; thus, e2/e3 individuals have overall apoE function that is partway between the reduced affinity of an e2 homozygote and the higher affinity of an e3 homozygote. With most genetic association studies for complex diseases, the inheritance model remains unknown.

EXAMINING GENES AT THE POPULATION LEVEL

In genetics, it is usual to describe the distribution of the alleles of interest in the population. In the same way that most continuous variables in medicine observe a normal distribution, most allele distributions observe what is called Hardy-Weinberg equilibrium (HWE). The Hardy-Weinberg law states that if there are 2 alleles at a particular locus, named *A* and *a*, with frequency

p and *q*, respectively, then after 1 generation of random mating the genotype frequencies of the AA, Aa, and aa groups in the population will be p^2, $2pq$, and q^2, respectively. Given that there are only 2 alleles possible, A or a, then $p + q = 1$, and $p^2 + 2pq + q^2 = 1$. It is general practice in a genetic association study to check whether the allele frequencies observe HWE proportions. Deviations from HWE in the population may be due to the following:

Figure 3. Transcription and Translation
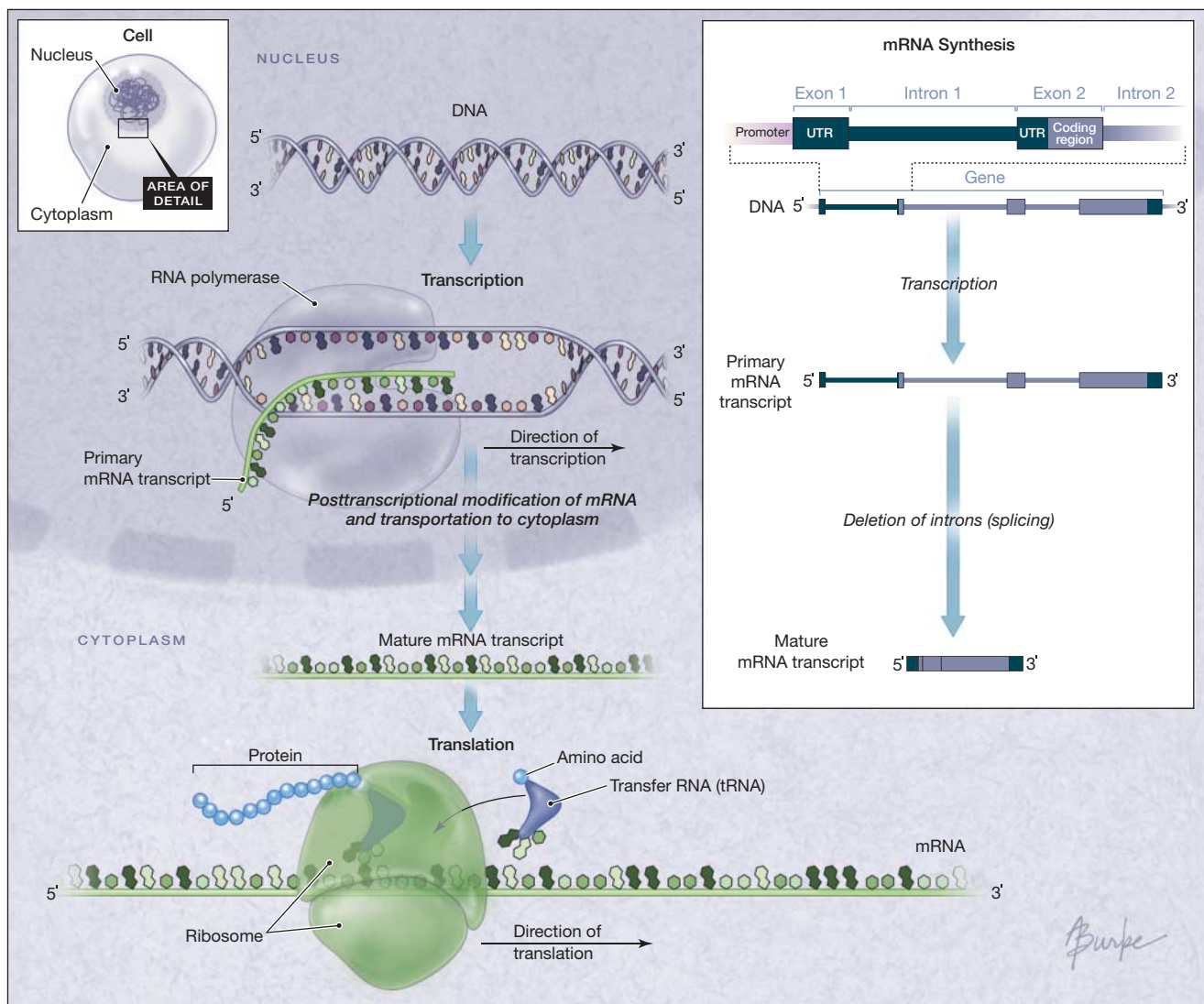
During transcription, the DNA double helix is split apart, and RNA polymerase synthesizes messenger RNA (mRNA) using one DNA strand as a template. Sections of the primary mRNA transcript, called introns, are spliced out to form the mature mRNA, which moves into the cytoplasm. The ribosome uses the mRNA sequence to build the protein. A specific sequence of 3 bases codes for each amino acid, which is delivered to the ribosome by transfer RNA. UTR indicates untranslated region.

- Inbreeding, ie, marrying close relatives, because HWE depends on random (with respect to the relevant gene) mating

- Genetic drift, a process in which a population is isolated, with a limited number of possible matings
  - Migration

- New mutations; only very new mutations upset HWE because equilibrium is usually reached within 1 generation in a sufficiently large population (this is where the "equilibrium" in HWE comes from)

- Selection, eg, a selective disadvantage of a particular allele that leads to fetal death.

Deviations from HWE may also signal methodological problems with the genetic study (eg, error in genotyping or population stratification), a possibility discussed in the next article in this series.

Figure 4. Common (Wild-Type) Allele and 4 Types of Genetic Polymorphisms

DNA polymorphisms include deletions, in which a DNA sequence is missing compared with the common allele, and insertions, in which a DNA sequence is added compared with the common allele. Repeats may also occur, in which the same sequence repeats multiple times. Depending on the size of the repeating unit and the number of repeats, these variants may have different names, such as satellites, microsatellites, minisatellites, or copy number variants. Single-nucleotide polymorphisms (SNPs), variations at a single base-pair location, are the most common type of polymorphism in the human genome.

CANDIDATE GENE VS GENOME-WIDE APPROACHES

Investigators conducting genetic association studies may target genes for investigation according to the known or postulated biology and previous results, an approach known as candidate gene association. Alternatively, they may screen the entire genome for associations, an approach that has, in recent years, transformed the field of genetic association studies. Researchers now investigate hundreds of thousands of SNPs across the entire genome, without any previous hypotheses about potential mechanisms or candidates. This type of "agnostic" study, "genome-wide association," or GWA, has greatly accelerated the pace of discovery of genetic associations.[16,17]

As discussed in the next article in this series, testing so many potential genes simultaneously carries the risk of finding many spurious associations. For this reason, SNPs that seem to have strong or suggestive statistical signals in an initial GWA study are then tested for replication in other large data sets or studies. To ensure that the discoveries are not just statistical flukes, replication studies are often published along with the initial data.[7] The boundaries between candidate gene studies and agnostic GWA studies can become blurred, and the 2 types of studies are not mutually exclusive: GWA studies propose new candidates for replication but may also interrogate traditional candidates.

Whether hypothesis-driven or GWA studies, genetic association studies usually represent population-based investi-

gations in which diseased and nondiseased individuals are unrelated. Genetic studies (either candidate or genome-wide) may also be carried out among family members in large pedigrees affected by a rare disease caused by a rare mutation. In such studies, the presence of certain stretches of DNA or genes in family members with disease and their absence in those without disease is called *linkage analysis*. The methods and interpretation of family studies differ radically from those of population-based studies,[18,19] and this series is restricted to discussion of the latter.

Because detecting SNPs is relatively easy and they are responsible for most of the genetic variation in humans, they have been the focus of most of the research exploring gene-disease association. The basic idea of a gene-disease association study is relatively simple. In the same way that variation in an exposure is linked to an outcome (eg, cholesterol and myocardial infarction) in a traditional epidemiologic study, variation in a gene is linked to an outcome (eg, a SNP in the cholesteryl ester transferase gene and myocardial infarction) in a genetic association study.

LINKAGE DISEQUILIBRIUM

One goal of traditional epidemiology, elucidating causation, can be frustrated by noncausal associations. Traditional epidemiology tries to deal with this problem by adjusted or multivariate analysis. For example, when examining the association of hypertension with stroke, investigators will simultaneously adjust for other variables, including age, sex, smoking, and obesity.

In genetic association studies, one goal may be to establish whether a SNP is causally associated with the outcome (eg, presence or absence of disease). This requires isolating the function of a particular SNP from the other SNPs that may be nearby in the gene. In practice, because stretches of the genome tend to be inherited together as a unit (a situation known as *linkage disequilibrium*), this may be difficult. Thus, the association of a SNP with an outcome, no matter how strong, may be noncausal. It is possible that some other SNP in linkage disequilibrium with the

one under study is the true causal variable. Although this is an important distinction if the aim is to understand the underlying biology to develop a new therapeutic agent, it is not critical if the aim is to use the SNP simply as a marker of risk.

The linkage disequilibrium phenomenon results in "haplotype blocks," stretches of DNA defined by the presence of high linkage disequilibrium among the SNPs present. Two or more SNPs that are in linkage disequilibrium in the same haplotype block can define haplotypes, specific combinations of variants across these SNPs. The combinations of variants that can arise are dictated by the extent of linkage disequilibrium.

As an example, consider SNP A with a common allele frequency (A) of 80% and SNP B with a common allele frequency (B) of 60%. If there is no linkage, allele A at SNP A and allele B at SNP B will be found together in the same person $0.80 \times 0.60 = 36\%$ of the time, ie, consistent with chance. With perfect linkage disequilibrium, eg, where the SNPs are very close together, it may happen that allele A is always found with allele B. The extent of this linkage disequilibrium may be expressed a number of different ways, with metrics such as r^2 (a plain correlation coefficient) or D′. Describing the properties of each metric goes beyond the scope of this article[16]; however, in general, the greater these measures, the greater the degree of linkage between the variants, with a D′ of 1 indicating that 2 alleles are always found together.

Armed with this background knowledge of genetic concepts and terminology, you are now ready to embark on an evaluation of validity, results, and applicability in genetic association studies. The remaining 2 articles in this series will address these issues.

REFERENCES

1. Yu W, Gwinn M, Clyne M, et al. A navigator for human genome epidemiology. *Nat Genet*. 2008; 40(2):124-125.
2. Samani NJ, Erdmann J, Hall AS, et al; WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357(5):443-453.
3. Zeggini E, Weedon MN, Lindgren CM, et al; Wellcome Trust Case Control Consortium (WTCCC). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829):1336-1341.
4. Matarín M, Brown WM, Scholz S, et al. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol*. 2007;6(5):414-420.
5. Hafler DA, Compston A, Sawcer S, et al; International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genome-wide study. *N Engl J Med*. 2007;357(9):851-862.
6. Easton DF, Pooley KA, Dunning AM, et al; SEARCH collaborators; kConFab; AOCS Management Group. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447(7148): 1087-1093.
7. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447(7145):661-678.
8. Plenge RM, Seielstad M, Padyukov L, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genome-wide study. *N Engl J Med*. 2007;357(12):1199-1209.
9. Coon KD, Myers AJ, Craig DW, et al. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry*. 2007; 68(4):613-618.
10. Li H, Wetten S, Li L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol*. 2008; 65(1):45-53.
11. Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron*. 2007;54(5):713-720.
12. Lander ES, Linton LM, Birren B, et al; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921.
13. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437 (7063):1299-1320.
14. Frazer KA, Ballinger DG, Cox DR, et al; International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007;449(7164):851-861.
15. Sachidanandam R, Weissman D, Schmidt SC, et al; International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001; 409(6822):928-933.
16. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008;9(5):356-369.
17. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA*. 2008;299(11):1335-1344.
18. Dawn Teare M, Barrett JH. Genetic linkage studies. *Lancet*. 2005;366(9490):1036-1044.
19. Risch N. Evolving methods in genetic epidemiology, II: genetic linkage from an epidemiologic perspective. *Epidemiol Rev*. 1997;19(1):24-32.

intellectual input from Forest Laboratories. We sincerely regret this lack of transparency in our initial disclosures that resulted from these errors of memory.

Robert G. Robinson, MD
robert-robinson@uiowa.edu
Stephan Arndt, PhD
Department of Psychiatry
University of Iowa
Iowa City

1. Robinson RG, Jorge RE, Moser DJ, et al. Escitalopram and problem-solving therapy for prevention of poststroke depression: a randomized controlled trial. *JAMA*. 2008; 299(20):2391-2400.

CORRECTIONS

Incomplete Financial Disclosures: In the Original Contribution entitled "Escitalopram and Problem-Solving Therapy for Prevention of Poststroke Depression: A Randomized Controlled Trial," published in the May 28, 2008, issue of *JAMA* (2008; 299[20]:2391-2400), the financial disclosures on page 2399 were incomplete. They should have appeared as follows: "All of the authors received salary contributions from the grant supporting this study. Over the past 5 years, Dr Robinson reports serving as a consultant to the former Hamilton Pharmaceutical Company and Avanir Pharmaceutical Company; receiving support or honoraria from Lubeck, Forest Laboratories, and Pfizer; being on the speakers' bureau for Forest Laboratories and Pfizer; and receiving grant support from the National Institute of Mental Health. Dr Small reports that he conducted a research study funded by Northstar Neuroscience that was unrelated to this prevention study. Dr Arndt reports inheriting Pfizer stock, which he owned from January 6, 2005, to December 23, 2006, and receiving grant support from the National Institute of Mental Health. The former Hamilton Pharmaceutical Company and Avanir Pharmaceutical Company had no financial interest in this prevention study. No other authors reported any financial disclosures."

Author Name Omitted: In the review of the book *HIV*, published in the August 6, 2008, issue of *JAMA* (2008;300[5]:584-585), the name and affiliation information of an author were omitted. In the byline at the end of the book review, Claudia Cosmineanu, MD, should have been listed following the information provided for Richard A. Stein, MD, PhD. Dr Cosmineanu is affiliated with the Department of Pediatrics, Elmhurst Hospital Center-Mount Sinai Service, New York, New York (claudia.cosmineanu@mssn.edu).

Incorrect Figure Label: In the Clinical Review titled "Inhaled Corticosteroids in Patients With Stable Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-analysis," published in the November 26, 2008, issue of *JAMA* (2008; 300[20]:2407-2416), a column heading was incorrectly labeled in a figure. In the bottom panel of Figure 2 on page 2413, the "Odds Ratio (95% CI)" column heading should have read "Relative Risk (95% CI)."

Data Error in Text: In the Users' Guides to the Medical Literature article entitled "How to Use an Article About Genetic Association: A: Background Concepts," published in the January 7, 2009, issue of *JAMA* (2009;301[1]:74-81), an incorrect number appeared on page 81. In the penultimate paragraph, the second sentence should have read, "If there is no linkage, allele A at SNP A and allele B at SNP B will be found together in the same person $0.80 \times 0.60 = 48\%$ of the time, ie, consistent with chance."

Incorrect Units: In the Special Communication titled "Treating Drug Abuse and Addiction in the Criminal Justice System: Improving Public Health and Safety," published in the January 14, 2009, issue of *JAMA* (2009;301[2]:183-190), data were reported with incorrect units on page 185. The cost of incarceration should have been reported as $22 000 per year rather than per month; likewise, the average cost of methadone should have been reported as $4000 per year rather than per month.