Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

A Genetics Primer for Social Health Research

Baldwin M. Way[1]* and Brian M. Gurbaxani[2,3]
[1] *University of California*
[2] *Centers for Disease Control and Prevention*
[3] *Georgia Institute of Technology*

Abstract

The recently completed sequencing of the human genome heralds a new era in the study of social influences upon health. Because the interactions between genes and the environment are bidirectional, expertise in the areas of psychosocial processes and health behaviors will be necessary for elucidating how genes, behavior, and the environment interact to affect health outcomes. For investigators whose primary background is in social health research, the terminology used by geneticists may seem like a foreign language. To help navigate this foreign territory, the nature of genetic variation and gene action is presented in non-technical terms using the serotonin transporter gene as an example because it is thought to influence sensitivity to the social environment. In addition, we describe several methodological pitfalls to be avoided when associating genetic variation with psychosocial and behavioral risk factors for poor health outcomes.

Introduction

Now that the Human Genome Project has catalogued the full extent of human genetic material (International Human Genome Sequencing Consortium, 2004), the next scientific step will be to determine the effects of this genetic information upon health outcomes. Because the action of genes is highly dependent upon the environment (Kandel, 2001), consideration of social, behavioral, and psychological variables will be necessary for detecting the genetic contributions to health. Integrating across these multiple levels of analysis will require unprecedented interdisciplinary collaboration. For social health researchers, one hindrance to such collaborations is understanding the terminology, concepts, and methodology used by geneticists.

Therefore, this paper is written for psychologists and sociologists interested in incorporating genetic measures into their studies, or at least gaining a working knowledge of the literature, but who do not want to become geneticists in the process. Thus, we provide an introductory description of genetic concepts in non-technical language. For the more interested

reader, many of these concepts are described in greater detail in the appendix. The paper is structured around understanding the information provided by an individual's genotype, which refers to his/her genetic identity (defined in greater detail in the next section). The technology for determining an individual's genotype (e.g., genotyping) has become increasingly automated in recent years and made it financially feasible (~$50/subject) to supplement psychological studies with genotype data.

Interpreting the results from such studies is contingent upon understanding the nature of the relationship between genes and psychology. Therefore, this paper discusses the nature of gene action and its influence upon cellular processes, which is the critical biological level of analysis in between the genetic and psychological levels. After an introduction to the nature of gene action, genetic variation and its influence upon cellular function is discussed. Finally, because associating genetic variation with psychosocial and behavioral health risks is a new enterprise, several methodological recommendations are made for conducting such studies. To provide a common thread across the multiple levels of analysis between gene and behavior, we use the serotonin system as an illustrative example. This neurochemical system seems to play a critical role in regulating responses to social stimuli and is thus likely to have broad influence across the diverse subdomains of social health research.

Variation in the Serotonin Transporter Gene and Sensitivity to the Social Environment

The serotonin transporter is probably best known as the site of action for the drug Prozac and related anti-depressants (Wong, Bymaster, & Engleman, 1995). Within a portion of the serotonin transporter gene, there is a segment of DNA that is longer in some individuals than others (Lesch et al., 1994). In straightforward fashion, the short version is called the short allele, whereas the long version is called the long allele. Thus, the term allele simply refers to a difference in the genetic (e.g., DNA) sequence. Because each of us receives one allele from our mother and one allele from our father, we have two copies of the serotonin transporter gene. Hence, an individual has either two copies of the short allele (short/short), one copy of each allele (short/long), or two copies of the long allele (long/long). The combination of alleles one has at this location in the DNA is referred to as one's genotype (e.g., long/long), and the process of determining which alleles are present in a particular individual is called genotyping. With respect to the serotonin transporter gene, the location where this variation occurs has been given the name 5-HTTLPR (this acronym will be defined later) and seems to moderate the effects of both social and behavioral influences upon health.

For example, one of the most robust social effects upon both physical and mental health is that of socioeconomic status (SES; Adler et al., 1994;

Singh-Manoux, Marmot, & Adler, 2005). Naturally, the quality of health care, nutrition, and housing contribute to this health gradient; yet, psychosocial stress is also a major mediator that links SES to health (Cohen, Kaplan, & Salonen, 1999). There are individual differences in response to the stress of low SES, a portion of which can be explained by the 5-HTTLPR. Individuals with a copy of the short allele (short/short or short/long) are more sensitive to this social stressor than those with two long alleles, as measured by a biochemical assay (Manuck, Flory, Ferrell, & Muldoon, 2004). Similarly, individuals with a short allele are at higher risk for major depressive disorder when exposed to life stressors such as unemployment and divorce (Caspi et al., 2003). Although the mechanisms for this effect are unclear, it seems that having a short allele (particularly two copies; e.g., short/short) renders one more sensitive to the social environment because having good social supports has greater influence upon mood state in these individuals than in individuals who have the long allele (Kaufman et al., 2004).

Beyond influences upon mood state, the 5-HTTLPR may also influence the adoption of health behaviors as well. For example, it may influence the number of sexual partners, which is related to risk for sexually transmitted diseases. Adolescents with a short allele (short/short or short/long) who regularly attend religious services have 30% fewer sexual partners than individuals who attend religious services as often, but have two copies of the long allele (Halpern, Kaestle, Guo, & Hallfors, 2006). As frequency of participation in religious activities influences the number of sexual partners (Miller & Gur, 2002) as well as age of sexual debut (Rostosky, 2004), one could interpret these findings within the context of social control theories (Durkheim, 1951; Rohrbaugh & Jessor, 1975) where individuals with the short allele are more responsive to their social milieu and thus are more prone to adopt the conservative norms concerning sexual behavior associated with many religious organizations.

This example of the 5-HTTLPR illustrates three conceptual points about the relationship between genetics and psychology.

First, genes do not code for particular behavioral or psychological outcomes, but rather code for how an individual responds to the environment. Because the short allele increases sensitivity to the environment, if the early family environment is loving and caring, short/short individuals are at *decreased* risk for depression. Conversely, as described above, short/short individuals raised in a harsh, abusive home are at *increased* risk for depression, relative to long/long individuals. Thus, neither the gene nor the environment is the sole predictor of the outcome, but the two interact to do so (Eley et al., 2004; Taylor et al., 2006). Put in colloquial terms, it is not nature *or* nurture but rather nature *and* nurture that shape health outcomes.

Second, genetic effects are probabilistic, not deterministic. Hence, there are individuals with the short/short genotype that experience significant life stressors and do not become depressed. Thus, the short/short genotype

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

and life stressors are not diagnostic predictors of depression, but rather indicators of the probability of becoming depressed (also see Appendix: When are genetic effects deterministic?).

Further challenging the notion of genetic determinism, the 5-HTTLPR is not even a marker that is specific for depression risk. In meta-analyses, the short allele has also been associated with alcohol dependence (Feinn, Nellissery, & Kranzler, 2005), suicide (Li & He, 2007), bipolar disorder (Cho et al., 2005), and obsessive-compulsive disorder (Lin, 2007). Thus, the 5-HTTLPR, like many genetic variants, is neither necessary nor sufficient for a particular disease outcome.

Third, the effect sizes in these meta-analyses are small (for a further discussion of the effect sizes in genetic association studies, see Appendix: How large are the typical effect sizes of genetic variants on social health outcomes?). In fact, the effects are so small that they can be difficult to detect without the large sample sizes provided by meta-analyses. The effect sizes for typical environmental risk factors for illness (e.g., life stressors and depression) are several times larger (Kendler, 2005). Hence, for health outcomes where environmental influences have been shown (e.g., cardiovascular disease, type II diabetes, and chronic fatigue syndrome), the total genetic effect seems to be divided amongst many genes, each of small effect (Goertzel et al., 2006; Gottesman & Shields, 1967; Risch, 1990). Individuals with a high probability of suffering such health outcomes will thus have many of these genes, whereas individuals with low risk are presumed to have few of them. Geneticists refer to such health outcomes as having complex genetic bases.

A major reason for the weak and non-specific effects of the 5-HTTLPR is that there are many intervening biological levels between a gene and psychological outcome. At each of these levels – ranging from the microscopic level of gene transcription to the macroscopic level of emergent complexity in the brain – there is extensive regulation and influence from the environment that dilutes the influence of a gene upon a particular health outcome. Hence, in order to understand the information provided by a genotype, it is necessary to understand how genes affect processes at other levels in the biological hierarchy and how environmental influences are overlaid upon these. The next section introduces the basics of how genes affect biological function using the serotonin transporter as an example.

Genes code for Proteins: The Central Dogma

The genetic level can be viewed as a repository of instructions, much like a library, that contains the directions for conducting the biological functions of an organism. In this genetic library, the information is stored in the form of DNA, which consists of a four-letter alphabet, A, T, C, and G. (Technically, these four letters are called either bases or nucleotides.) The complete sequence of these four letters is called the human genome

and is about 3 billion letters long. Much like the letters in words, the sequence of these letters provides biological meaning. Hence, when the sequence is different, the biological meaning is different (e.g., CAT is different from TAT).

A specific segment of DNA sequence is referred to as a gene and contains the instructions for making a particular functional product (Snyder & Gerstein, 2003). (Also see Appendix: How are genes named and where are genes located?) It is these functional products, particularly proteins, that are involved in regulating processes in the next level of the biological hierarchy, the cellular level.

Many processes occur at the cellular level, but most relevant for this discussion is the process of neurotransmission, whereby serotonin serves as a signal to communicate between cells. Each of the principle steps in the process of neurotransmission is performed by a protein. The enzymes that make and break down serotonin, the receptors that bind serotonin, and the transporter that reuptakes serotonin into neural cells are all proteins. Hence, the performance and quantity of these proteins determine the duration and magnitude of serotonin's actions. Similarly, it is at this level that drugs such as Prozac or Ecstasy change the function of the serotonin transporter protein and thereby alter serotonin signalling.

In differentiating between the genetic level and the cellular level, it is helpful to think of the genes as containing the blueprints for a protein and the proteins as being the actual product that rolls off the assembly line. How the genetic blueprints get translated into a functional product is often referred to as the central dogma of biology and is summarized in the statement by Francis Crick (co-discoverer of the structure of DNA), 'DNA makes RNA, RNA makes protein, and proteins make us' (Crick, 1958). In addition to the brief description of the central dogma below, this process is depicted in Figure 1A. More detailed information can be found in a genetic or biochemistry textbook (e.g., Robinson, 2005).

Central dogma: Mechanisms

To affect biological function, the library of information stored in the DNA sequence must be disseminated to the cell. This dissemination process is referred to as gene expression and is analogous to taking a book off the shelf of the DNA library and reading it (Figure 1A). Hence, a protein is formed when a gene is expressed.

More formally, the initiation of gene expression is called transcription and involves making a copy of the DNA sequence using RNA. Because this process is how DNA disseminates its instructions to the rest of the cell, the RNA is referred to as *messenger* RNA (mRNA; Brenner, Jacob, & Meselson, 1961).

The initial transcript of RNA is merely a first draft of the message and undergoes an editing process whereby certain portions of the RNA are

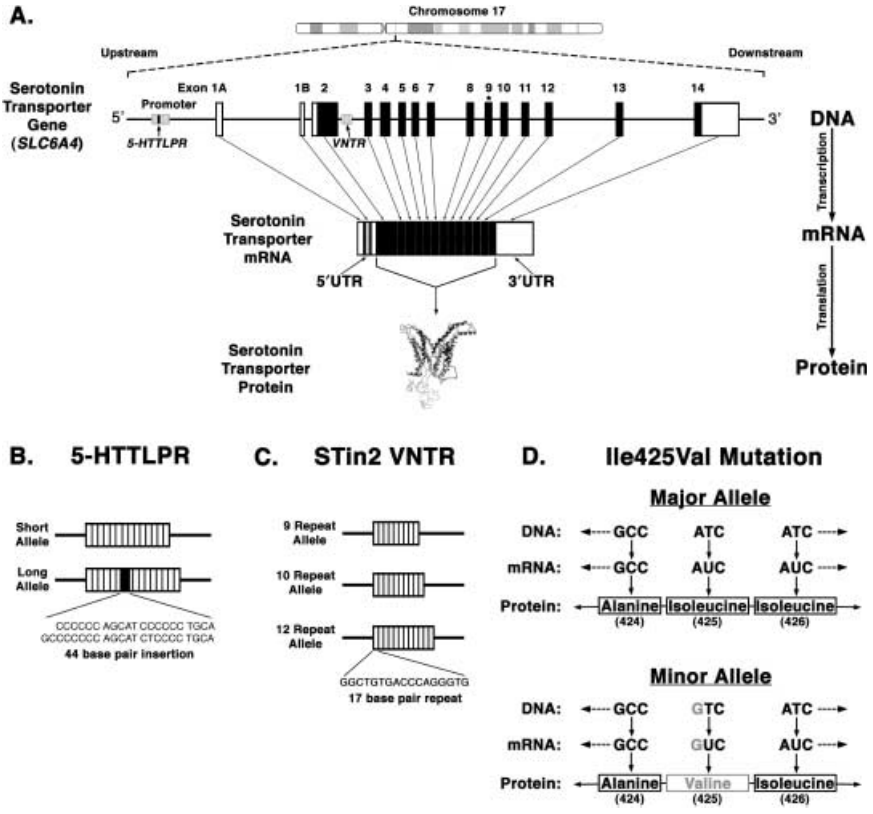

Figure 1. Schematic representation of serotonin transporter gene organization, polymorphisms, and expression. (A) Top*:* Map of chromosome 17. The location of the serotonin transporter gene is shown in expanded form below the chromosome map. The horizontal line represents the strand of DNA with the serotonin transporter gene. The upstream end of the gene (5′) is located on the left. Progressing to the right, the first portion of the gene is the promoter. Within the promoter is the *5-HTT* gene linked polymorphic region (5-HTTLPR) depicted with a horizontal rectangle and shown in greater detail in B. After the promoter, the position of the *ex*ons (expressed sequence) is depicted along the DNA strand with vertical rectangles that separate the intervening intronic sections. Each exon is numbered. STin2 VNTR is denoted with a horizontal rectangle and is actually located in intron 3, as exon 1B (Bradley & Blakely, 1997) had not yet been identified when STin2 was named. Star denotes location of the Ile425Val Mutation in exon 9. Expressed portions of the gene are indicated by arrows extending downward from the DNA to the mRNA. The open rectangles depict sequence that is transcribed into mRNA, but not translated into protein. UTR denotes untranslated region. As in the DNA depiction, the darkened boxes within the mRNA depiction represent sequence that is translated into the final protein. Bottom*:* Three-dimensional model of the serotonin transporter protein (adapted from Ravna, Sylte, Kristiansen, & Dahl, 2006; reprinted with permission from Elsevier) that resides in the membrane of the cell and performs the transport of serotonin. (B) Schematic of the principal 5-HTTLPR alleles (adapted from Lesch et al., 1996; reprinted with permission from American Association for the Advancement of Science). Each rectangle represents the repeating sequence with the 44 base pair insertion/deletion denoted by a black box. Listed below is the nucleotide sequence, as originally reported by Heils et al. (1996). Note: consensus has not been achieved concerning the location of the specific deleted sequence (Wendland, Martin, Kruse, Lesch, & Murphy, 2006). (C) Schematic of STin2 VNTR poly-morphism, showing the most frequent alleles. Rectangles represent the repeating sequence.

deleted. The stretches of RNA that are removed are called introns, whereas the stretches that remain in the final message are called exons. (The corresponding portions of the DNA sequence from which the mRNA is copied are also called introns or exons).

The process of converting the information contained in the mRNA to protein is called translation because it translates the language of RNA, which consists of the letters A, C, G, and U into the language of proteins, which are built from 20 different amino acids. Thus, proteins consist of a sequence drawn from the 20-letter alphabet of amino acids, rather than the four-letter alphabet of DNA or RNA. Unlike transcription, where there is a 1 : 1 relationship between DNA and RNA, there is a 3 : 1 relationship between RNA and an amino acid. Thus, the three RNA bases spell a word, or codon, that specifies a particular amino acid (see Figure 1D; Nirenberg & Matthaei, 1961). Similar to DNA and RNA (and words), the particular sequential order of these amino acids gives a protein its identity and determines its function (but, see Appendix: Does the same DNA sequence always code for the same protein?). For example, changing just one of the 630 amino acids in the serotonin transporter can change its function. Changing the amino acid (from serine to leucine) at position number 372 can eliminate the activity of the transporter, meaning it will no longer transport serotonin (Forrest, Tavoulari, Zhang, Rudnick, & Honig, 2007). Alternatively, changing the amino acid (from isoleucine to methionine) at position number 172 can eliminate the ability of Prozac to exert its pharmacological effect upon the serotonin transporter (Henry et al., 2006). Or, as discussed in greater detail in Figure 1A and D as well as the section on how genetic variation relates to biological function, changing the amino acid at position 425 from an isoleucine to a valine increases the efficiency of serotonin transport (Kilic, Murphy, & Rudnick, 2003). Hence, it is the action of proteins that is critical for regulating the processes at the cellular level that are thought to constrain psychological processes.

Conceptual Background on Gene Expression

With this background on the respective roles of genes and proteins, the next step that is necessary for understanding the information that a genotype

Figure 1. (*Continued*) Bottom: 17 nucleotides in the first repeat, according to Kaiser et al. (2001). (D) Depiction of isoleucine to valine non-synonymous mutation at amino acid position 425 of the serotonin transporter protein. For both the minor and major alleles, a portion of the DNA sequence is depicted on top (letters representing each base are separated for presentation purposes, but actually are continuous). Middle: Segment of the mRNA sequence transcribed (depicted with downward arrows) from the DNA sequence. Bottom: Amino acid sequence coded for by the mRNA (translation from mRNA to amino acid is depicted by the downward arrows). For the minor allele, the substitution of a G for an A results in valine (depicted in gray) rather than isoleucine.

can and cannot provide is to understand the process of how a gene gets turned on (expressed). Hence, the next section differentiates between genetic sequence and genetic expression. A helpful example to keep in mind when thinking about this difference is that of the caterpillar and butterfly. The DNA sequence (e.g. genotype) does not change when a caterpillar becomes a butterfly, but the radically different morphology (e.g., phenotype) is a result of differences in gene expression. When the insect is a caterpillar, there are certain sets of genes that are expressed in order to make it a caterpillar. Then, when it is time to become a butterfly, these genes get shut off and the genes that make it a butterfly are turned on.

More directly applicable to this discussion is the activation of the serotonin transporter gene as a function of stress. One of the hormones released in response to a stressor is cortisol, which binds to the glucocorticoid receptor. When cortisol is bound to this receptor, the bound receptor moves into the nucleus where it can bind the DNA and can then actually increase expression of the serotonin transporter gene (Glatz, Mossner, Heils, & Lesch, 2003). This regulation process occurs for all genes in a cell whereby their expression is increased, decreased, turned on, or shut off in response to specifically targeted signals (e.g., cortisol). Typically, changes in gene expression will lead to corresponding changes in the amount of protein produced, which then affects processes at the cellular level. Thus, if more serotonin transporter protein is made, the cell has the capability to transport more serotonin.

Some of the signals influencing gene expression come from other genes, but many of these signals are triggered by the different environments an individual encounters. Thus, signals from the environment can initiate a cascade of events that regulate the information disseminated from a gene by influencing its expression. (For a discussion of long-term environmental influences upon gene expression, see Appendix: How do environmental signals trigger lasting changes in gene expression?). This is why complex psychological processes are a function of both genes and environment: nature and nurture.

The particular sets of genes that are expressed in a cell are what give a cell its unique character. Although the DNA sequence is the same in each cell taken from the same person (immune system cells are an exception), cells in the eye, skin, liver, and brain express different portions of the DNA sequence. It is these differentially expressed genes that then enable each respective cell type to conduct its respective job. In a similar fashion, a particular cells' ability to perform its job is determined by the current state of gene expression in that cell.

Therefore, assessing the degree of activation or suppression of specific genes provides a window into the functional capabilities of a cell. However, to measure gene expression, one needs to measure the amount of RNA or protein in the specific cell type of interest. With present technology, gene expression in areas such as the brain can only be measured

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

post-mortem, which creates obvious research challenges. Alternatively, measuring an individual's genotype is much less invasive because the DNA sequence is present in each and every cell, including cells in the saliva or blood. However, the information contained in the DNA sequence is limited. It is only an indicator of function, not an actual measure of functional capability (this difference will be elaborated upon in the section on genetic variation).

Gene expression: Mechanisms

Although the primary focus of this article is upon the nature of the information provided by a genotype, it is important to understand the different portions of a gene involved in regulating gene expression because variation (e.g., 5-HTTLPR) can occur anywhere in a gene. (For further details on the mechanisms underlying gene expression, please see Appendix: What molecular mechanisms control gene expression?).

The switch for turning on a gene is called the promoter, which resides at the front end of all genes. This front end, or upstream portion, of the gene is referred to as the 5′ (pronounced 'five prime') end, a notation derived from the chemical nature of DNA, and the back end of the gene is referred to as the 3′ end (Figure 1A). The actual act of transcription is performed by the assembly of dozens of proteins at the gene's transcription start site. (Previously, we had described the relationship between gene and protein as a one-way street, but in actuality, there is bidirectional communication between the protein and genetic levels). The degree of activation of this switch can be modulated by DNA sequences that are typically located in the vicinity of the promoter that can act to enhance or repress the degree of expression. For example, binding to such enhancers and silencers by the glucocorticoid receptor is one of the ways in which cortisol can modulate gene expression (Dostert & Heinzel, 2004).

Terminology for Describing Genetic Variation

Having completed a description of basic genetic processes, we now turn to variations upon this theme by discussing variations in the DNA sequence and how they can affect cellular function. As mentioned earlier, the entirety of genetic information resides in each cell and is referred to as the human genome, which is 99.9% identical among all humans (Kruglyak & Nickerson, 2001). The differences in this genetic sequence, the remaining 0.1%, are thought to underlie much of the physiological component of individual uniqueness. These changes can be referred to as variants, mutations, or polymorphisms. Although there is some disagreement concerning this terminology, the term 'polymorphism' is generally used to refer to a mutation that is present in greater than 1% of the population. In other words, if 20% of the population were to have a particular sequence variation,

it would be called a polymorphism, whereas if this variation were present in less than 1% of the population, it would be called a mutation. The term variant, or allelic variant, is generally used irrespective of frequency (den Dunnen & Antonarakis, 2001). For a more detailed description of genetic variation, see Gibson and Muse (2004).

Polymorphisms are primarily found through the process of sequencing DNA, which refers to identifying each and every base (e.g., letter). Once the complete sequence and thus the precise location of a polymorphism is known, a different methodology can be used to identify which base occupies a certain position. This is called genotyping, which gives information on the particular form of a polymorphism an individual has. Thus, genotyping is only used to determine which alleles are present at a particular locus, not discover new allelic variants.

Single nucleotide polymorphism

The most prevalent and perhaps the easiest type of polymorphism to understand is the single nucleotide polymorphism (SNP; pronounced 'snip'). As the name implies, one single nucleotide, or base, is substituted for another. Using the analogy of each base in the DNA sequence being represented by a letter, this corresponds to the difference between 'goal' and 'goat.' Thus, 'goal' and 'goat' would be the two alleles at this locus, or location within the genome. When an individual is homozygous at this locus, her genotype would be 'goal/goal' or 'goat/goat.' Heterozygous individuals would have the 'goal/goat' genotype, meaning they have one version of each allele. Each SNP is given a reference number (rs#) and is catalogued in the National Center for Biotechnology information (NCBI) SNP database (dbSNP). SNPs occur about every 500 to 2000 bases, so each of us has about three million SNPs. [What we have been referring to as letters are typically referred to as base pairs (abbreviated bp); the term 'pair' is used because DNA consists of two strands of letters and each letter has a corresponding partner that it pairs with].

Length polymorphism

The 5-HTTLPR is a different type of polymorphism where the variation is not in a single nucleotide, but rather in the length of the DNA sequence. To continue with the language analogy, this would be akin to the difference between 'goal' and 'goooooooooal.' In the case of the 5-HTTLPR, the long allele has a sequence of 44 nucleotides that is not present in the short allele (Heils et al., 1996; Lesch et al., 1996). This is referred to as an insertion/deletion polymorphism because depending on whether one is referring to the long or the short allele, the 44 base pairs is either inserted or deleted (Figure 1B). As an aside, the acronym 5-HTTLPR stands for Serotonin Transporter Linked Polymorphic Region.

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

The notation '5-HT' is an abbreviation for the chemical name of serotonin, 5-hydroxytryptamine. The molecule was named 'serotonin' before the discovery of its chemical structure (Rapport, Green, & Page, 1948).

Variable number tandem repeats polymorphism

Another variation on this theme is a variable number tandem repeats polymorphism (VNTR). Again, using the language analogy, these would be analogous to the difference between 'gogogogogogoal' and 'gogoal' where a specific sequence is repeated a different number of times. Such a polymorphism also occurs within the serotonin transporter gene (although at a different location, or locus, than the 5-HTTLPR; Figure 1C) where a 17-base-pair sequence is repeated either 9, 10, or 12 times (Lesch et al., 1994), yielding three different alleles. Thus, unlike SNPs, there can be more than two VNTR alleles in the population (although any one person will have no more than two, one on each chromosome. For recently discovered caveats, see Appendix: Is there always two copies of every gene?). This polymorphism is called STin2.10, for Serotonin Transporter intron 2, 10 repeat allele and STin2.12 for the 12 repeat allele, etc. (Confusingly, the 5-HTTLPR can also be referred to as a VNTR because the short allele has 14 repeats of a 20–23 base sequence, whereas the long allele has 16 repeats of this same sequence).

How Do Genetic Polymorphisms Affect Biological Function?

With this background, it is now possible to put all of the pieces together in order to understand how the position of a polymorphism within a gene can potentially alter its functional effects.

Polymorphisms in the coding sequence

Polymorphisms that occur in the portions of the genetic sequence that code for proteins are thought to have the greatest probability of affecting cellular function because they can change the particular amino acid coded for, which will typically change the structure of the protein (Figure 1D). These are called by several interchangeable terms: replacement, non-synonymous, missense, or structural polymorphisms (For a discussion of synonymous polymorphisms, see Appendix: Can synonymous SNPs affect protein function?). In referring to such polymorphisms, they are typically denoted by the amino acid that is changed. For example, there is a rare variant in the portion of the serotonin transporter gene that codes for the 425th amino acid in the serotonin transporter protein (see Figure 1D; Glatt et al., 2001; Sutcliffe et al., 2005). The variant leads to a change from isoleucine (Ile), which is present in 99.8% of the population, to Valine (Val). Thus, it is referred to as Ile425Val, with the major (e.g., more

common) allele listed first and the minor allele listed second (den Dunnen & Antonarakis, 2000). In this case, the change increases the ability of the protein to reuptake serotonin (Kilic et al., 2003).

Typically, the functional effects of coding polymorphisms are assessed in a biochemical assay that measures the functional activity of the protein produced by each allele or genotype. In the case of the serotonin transporter, the amount of serotonin transported by cells that carry the different alleles is measured in a test tube. In general, the behavioral and psychological effects of replacement polymorphisms are the most widely accepted because of their well-documented effects upon protein structure and thus function.

Polymorphisms in genetic regulatory regions

Polymorphisms that occur in the regulatory areas of the gene are likely to affect cellular function in a different way than polymorphisms in the coding region. Rather than changing how the protein functions (e.g., how the job gets done), they are likely to affect whether or not a gene gets turned on in response to a particular stimulus and to what degree. Thus, regulatory polymorphisms influence how many proteins are produced to do the job and when they report for duty. Regulatory polymorphisms are most likely to be located in the promoter or first few introns, but occasionally, regulatory elements in the downstream 3' region can engage in some backseat driving.

As regulatory regions are responsible for determining when and how much protein is made, variation in this region of the gene will affect the responsiveness of a gene to cellular signals, including those that were triggered by environmental stimuli. Such variation is likely to be particularly relevant for social health researchers interested in individual differences in response to the same environment. For example, the previously described effects of the glucocorticoid receptor upon serotonin transporter gene expression depend on 5-HTTLPR allelic status. The short allele seems to be less sensitive to the effects of glucocorticoid receptor binding. Thus, because of variation in the promoter, the gene responds differently to the same signal and is one hypothesis of how the 5-HTTLPR moderates response to stressors (Glatz et al., 2003).

The functional effect of a regulatory polymorphism is generally assessed by determining how much mRNA or protein is produced. This can be assessed in post mortem samples of brain tissue, in cells (i.e., lymphocytes) isolated from individuals or in *in vitro* model systems, such as engineered cells.

Demonstrating functional effects of regulatory polymorphisms has been more challenging to document than coding polymorphisms because the effects are of a smaller magnitude (1–2 fold) and are highly dependent on the cellular context and environmental signals that are present (Knight,

2005). Recent methodological developments have created the opportunity to measure the expression of each allele (Yan, Yuan, Velculescu, Vogelstein, & Kinzler, 2002; Yan & Zhou, 2004) and provide the most compelling evidence that a regulatory variant indeed has functional effects. With respect to the 5-HTTLPR, homozygotes for the short allele produce less serotonin transporter mRNA and protein in lymphocytes (Bradley, Dodelzon, Sandhu, & Philibert, 2005; Lesch et al., 1996; Martin, Cleak, Willis-Owen, Flint, & Shifman, 2007). However, the results in post mortem neural tissue have been more equivocal (Lim, Papp, Pinsonneault, Sadee, & Saffen, 2006; Mann et al., 2000; Parsey et al., 2006). Perhaps this is because only a main effect of the 5-HTTLPR upon gene expression has been assessed. Yet, based on the psychological data, one would expect to see allelic variation in gene expression only in subjects who had experienced significant life stressors, which has not been assessed. The lack of attention to the role of the environment reflects the biases of geneticists and is indicative of the field's need for health psychologists and their greater attention to environmental influences.

Such functional assays are also used to determine whether the expression of a particular allele is dominant over the other allele or whether both alleles are expressed, which is referred to as codominance. Whether the short allele is dominant or both alleles are codominant is an issue that has not been clearly resolved for the 5-HTTLPR.

In summary, polymorphisms in the regulatory regions (e.g., 5-HTTLPR) tend to have effects upon gene expression. If a polymorphism increases gene expression, then more protein will be produced, and there will be a corresponding change in cellular function because there are more proteins to perform the task. In the case of the serotonin transporter, greater expression would lead to more protein, which could then transport more serotonin. In contrast, polymorphisms within coding regions (e.g., Ile425Val) affect cellular processes by changing how the protein functions. In other words, more serotonin is transported by the protein coded for by the Val^{425} allele not because the cell is producing more protein, as is the case with the 5-HTTLPR long allele, but because the protein is more efficient at transporting serotonin (Kilic et al., 2003).

Non-functional Polymorphisms

Having just described the way that polymorphisms can affect cellular function, we turn to the opposite case: when polymorphisms do not affect function. In fact, the vast majority of polymorphisms fall into this category, presumably because there is no evolutionary reason to get rid of them (Kimura, 1983). These polymorphisms tend to fall in intronic and intergenic areas of the genome.

Researchers should be aware of such polymorphisms because they can be spuriously associated with the behavioral or psychological outcome

(e.g., phenotype) of interest. This is because nonfunctional polymorphisms can 'hitchhike' along with the polymorphism that is actually causing the phenotype in a process called linkage disequilibrium (discussed in the appendix). Therefore, additional experimental approaches, such as animal models or functional assays akin to those described in the previous section, are needed to ensure that the polymorphism associated with the phenotype of interest actually has a functional effect at the cellular level. When effects at the cellular level have been demonstrated, such variants are denoted as functional polymorphisms. Because most psychologists are likely to only genotype subjects for a few polymorphisms, it is prudent to make sure they are functional.

How Do I Genotype Participants in My Studies?

Technological developments have led to falling genotyping costs, creating the possibility to incorporate genetic measures into studies not previously focusing on genetics. Such studies require three steps: (a) obtaining and extracting the DNA, (b) genotyping, and (c) analysis. It is now possible to extract large quantities of DNA from saliva samples that can be obtained using a kit that can be stored at room temperature for up to a year. Extraction and genotyping can be done commercially, at core facilities of major universities, or in a standard genetics lab. However, consistent with the interdisciplinary perspective of this paper, such research is probably best done in collaboration with a research team that includes a molecular geneticist and a statistical geneticist or bioinformatics expert. A molecular geneticist will be well versed in genotyping procedures, which is a critical step in light of the fact that standard genotyping protocols may yield erroneous results with markers that are difficult to genotype. For example, the concentration of magnesium, which serves as a co-factor for the enzyme normally used to prepare DNA for genotyping, can lead to an artifactually inflated presence of the 5-HTTLPR short allele (Kaiser, Tremblay, Roots, & Brockmoller, 2002; Yonan, Palmer, & Gilliam, 2006). The statistical geneticist can aid in study design and data analysis in order to prevent repeating many of the mistakes of early association studies.

Methodological Considerations for Genetic Association Studies

The methodological approach most likely to be of utility in social health research is the association study (Cardon & Bell, 2001; Cordell & Clayton, 2005; Hattersley & McCarthy, 2005; Lander & Schork, 1994). In this hypothesis-driven approach, participants are genotyped for a particular polymorphism within a gene, called a candidate gene, that is hypothesized to affect the phenotype of interest. Selection of the 5-HTTLPR as a potential candidate gene is an example of this approach, whereby it was

reasoned that because the serotonin transporter protein is the primary target of many antidepressants, variation in the gene coding for this protein may influence risk for depression. If a case–control design is used, the differential frequency of alleles between cases and controls is assessed to determine if the allele is more prevalent in the disease group. If a continuous measure such as neuroticism is used, alleles are correlated with the trait measure.

The association approach is different from the classical methodology used in medical genetics called linkage analysis, which is to follow the occurrence of a disease causing trait through a family pedigree. Although the methodology underlying linkage has been well worked out (Collins, 1995), it is less suitable for complex traits where there are many genes with small effect as well as robust environmental influences (Altmuller, Palmer, Fischer, Scherb, & Wjst, 2001; Risch & Merikangas, 1996).

The methodology for association studies, however, is still being optimized, as evidenced by the disappointingly low replication rate of such investigations (Ioannidis, 2006). According to a review of 55 meta-analyses of genetic associations with a diverse array of health outcomes that included psychiatric, cardiovascular, and cancer-related conditions, only 49% of the associations were significant and when heterogeneity and publication bias were considered this number declined to 16% (Ioannidis, Trikalinos, Ntzani, & Contopoulos-Ioannidis, 2003). In a different study across a similar array of disease outcomes, only 6 of 166 already replicated associations were replicated in 75% of the identified studies (Hirschhorn, Lohmueller, Byrne, & Hirschhorn, 2002). As these authors (p. 60) eloquently concluded, 'until complete meta-analyses can be performed using data from multiple large studies, we will be left with a scenario in which the majority of reported associations are in genetic purgatory, neither convincingly confirmed or refuted, awaiting future judgment.' Based on such data, the most competitive genetics journals require replicating an association in an additional sample before accepting a manuscript (Nature Genetics, 2005).

Such data are sobering, and there are many potential explanations for the poor replicability of the results (e.g., genetic background, epistasis, epigenetics, imprinting, allelic heterogeneity, and genetic heterogeneity) that are beyond the scope of this review. Here, we discuss two methodological pitfalls (multiple comparisons and population stratification) that have contributed to the replication problem and then discuss three potential approaches that have been proposed for increasing the reliability of association studies.

Statistical considerations

A major contributor to the poor replication rate is likely to be insufficient correction for multiple comparisons (Colhoun, McKeigue, & Davey Smith, 2003). According to simulation data of a standard candidate gene study, a *P*-value with a significance threshold of 0.05 produced a 'publishable'

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

false positive result greater than 95% of the time (Sullivan, 2007). As most investigators have invested significant blood, sweat, and tears into acquiring their data, they naturally exhaustively analyze their data set, which typically includes multiple dependent and moderating variables (e.g., neuroticism, depression, self-esteem inventories, etc.). Such analyses should be tracked and reported so that both reviewers and investigators can correct for such comparisons accordingly.

The most conservative genetics journals require special justification for declaring a *P*-value greater than 1×10^{-7} significant (Freimer & Sabatti, 2005). The stringency of this threshold derives from correcting the association for all possible candidate genes in the genome (~25,000) that could have been selected, which yields a prior probability of 1 of 25,000 of a phenotype–genotype association (Freimer & Sabatti, 2004).

At present, a consensus concerning appropriate methods for determining significance in the face of the multiple comparisons issue has not been achieved, but promising approaches include Bayesian methods (Wacholder, Chanock, Garcia-Closas, El Ghormli, & Rothman, 2004), permutation testing (Hirschhorn & Daly, 2005), distance matrices (Zapala & Schork, 2006), and use of the false discovery rate (discussed in greater detail later; Benjamini & Hochberg, 1995).

Population stratification

Population stratification is a form of confounding that occurs when a sample contains two or more population subgroups that have different allele frequencies as well as different distributions of a particular phenotype (Cardon & Palmer, 2003; Thomas & Witte, 2002; Wacholder, Rothman, & Caporaso, 2002). As an illustration of this confound, we draw from the classic example of Lander and Schork (1994). If one were looking for genes associated with the quantitative trait of chopstick dexterity in an ethnically diverse city such as Los Angeles, one is likely to find that 5-HTTLPR short/short individuals are more facile with chopsticks. This is not because being short/short improves dexterity, but rather because 70% of the East Asian population is short/short, whereas only 20% of the European-American population has this genotype (Gelernter, Kranzler, Coccaro, Siever, & New, 1998). In this case, cultural background is confounded with genotype. This population stratification problem is particularly acute with polymorphisms in genes related to the immune system, as these vary substantially according to ancestral geographic origins (Hoffman, Hansson, Mezey, & Palkovits, 1998).

At the very least, association analyses should be performed separately according to ethnicity. However, even associations within ethnically homogeneous samples (Campbell et al., 2005; Helgason, Yngvadottir, Hrafnkelsson, Gulcher, & Stefansson, 2005; Marchini, Cardon, Phillips, & Donnelly, 2004) have been confounded by population structure. At present, it is unclear

whether these are the exception or the rule (Hutchison, Stallings, McGeary, & Bryan, 2004).

Nonetheless, with the reducing cost of genotyping it is prudent to employ one of several methods that have been developed to control for confounding by population stratification (Bacanu, Devlin, & Roeder, 2002; Devlin & Roeder, 1999). The most common approach is to genotype subjects on a set of ancestry informative markers, which are alleles that are only present in groups with particular geographic origins (Pritchard & Rosenberg, 1999). When analyzing the data, this information can then be used as a covariate or for separating outliers with respect to ancestral origin.

Intermediate phenotypes

In addition to avoiding such methodological problems as population stratification and multiple comparisons, reliability of association studies can potentially be improved by using designs that are more sensitive to genetic effects.

One approach for increasing the reliability of association studies is to focus on intermediate phenotypes, which are called endophenotypes when they are heritable (Gottesman & Gould, 2003). This approach is based upon the idea of reducing the complexity of the phenotype in order to increase the potential effect of a given gene. A phenotype such as major depression can vary in etiology, severity, duration, specificity, age of onset, familiality, and comorbidity. Hence, the broadness of this category may dilute the effects of a particular gene on a specific facet of this heterogeneous classification. In light of the role of stressful life events in precipitating depression, an example of an intermediate phenotype would be response to social stress, for which there is evidence of a main effect of the 5-HTTLPR (Jabbi et al., 2007).

In addition, the use of biological measurements is a way of bridging the gap between DNA sequence and health outcome. As there are fewer steps in the pathway between gene and a biomarker than between gene and health outcome, it is reasoned that genetic effects upon such markers will be more robust (Meyer-Lindenberg & Weinberger, 2006). The use of such a biological intermediate phenotype has been employed most successfully using functional imaging to show effects of the 5-HTTLPR upon amygdala reactivity while viewing threatening and fearful faces (Hariri et al., 2005).

Gene-environment interactions

As described in the opening example, assessing gene–environment interactions is another potential way to increase the reliability of association studies. The theoretical perspective underlying this approach is that a

specific polymorphism may not influence health outcomes in all situations, but only in particular environments (Moffitt, 2005). The 5-HTTLPR is the prototypical example of this perspective, as there was no main effect of the 5-HTTLPR upon depression risk in the same studies that reported a significant interaction with stressful life events. Hence, a likely contributor to the low replication rate of genetic association studies is the lack of control for environmental variables. As health psychologists have specialized in measuring such environmental variables, their expertise in determining which variables to measure could help improve the low replication rate of genetic association studies. Additional recommendations for designing studies of gene-environment interactions are nicely laid forth by Moffitt, Caspi, and Rutter (2005).

Multi-method convergence

Ultimately, the most convincing evidence will come from a convergence of findings acquired using multiple methodologies. The incorporation of genotyping into traditional social health investigations provides an unprecedented opportunity for integration across multiple biopsychosocial levels. This is because the identification of a genetic association implicates a particular biochemical pathway. For example, the 5-HTTLPR is presumed to affect mood via its effects upon serotonin signalling. Hence, serotoninergic neurotransmission can be experimentally manipulated (using administration of a drug like Prozac) in order to provide corroborative evidence to data that was obtained through a correlation with the 5-HTTLPR.

In addition, animal models can be used. There is a polymorphism that is analogous to the 5-HTTLPR in rhesus monkeys (Lesch et al., 1997; Suomi, 2006), which provides the opportunity to study the interaction of genes and social status or stress in an environment that can be better controlled and manipulated than is possible in human studies. Thus, genetics can serve as the common thread for integrating across behavioral, biochemical, psychological, and social levels to provide a clearer understanding of health.

Accordingly, our perspective is that as long as there is careful selection and measurement of phenotypes and environmental variables, supplementing health psychology studies with genotype data is likely to be a fruitful approach for identifying novel factors affecting health in both experimental psychological studies as well as general population based samples. (For a discussion of future directions of genetics studies, please see Appendix).

Acknowledgements

While conducting this work, B.M.W. was supported by an NIMH institutional training grant fellowship (MH15750) as part of the UCLA Health

Psychology Program. B.M.G. wishes to thank his CDC and Emory University colleagues in the CFS group for helpful discussions on genetics and gene-environment interactions.

Short Biographies

Baldwin Way is currently a post-doctoral fellow in Health Psychology at the University of California at Los Angeles. After graduating from Dartmouth College, he received his doctorate in neuroscience at the University of California at Los Angeles. His research focuses upon integrating genetic, pharmacological, psychological, and neuroimaging methodologies to better understand the role of the serotonin system in emotion regulation and sociality. Reflecting this interdisciplinary approach, he has published work in the fields of genetics, pharmacology, anatomy, law, and political psychology.

Brian Gurbaxani is a Senior Scientist at the Centers for Disease Control and Prevention in Atlanta, GA, in the Chronic Viral Diseases Branch. He is also an adjunct assistant professor of electrical and computer engineering at Georgia Tech, where he supervises graduate students doing research for the CDC as part of their theses. He is active in writing computational algorithms for analyzing large datasets related to chronic fatigue syndrome (CFS), human papillomavirus (HPV), and cervical cancer. He received a Ph.D. from UCLA in molecular biology with emphases in immunology, bioinformatics, and computational biology under two fellowships from the NSF. In a previous life, after receiving a B.S. in applied and engineering physics from Cornell University, he worked as a systems engineer for Hughes Space and Communications company in southern California. There, he wrote satellite mission control algorithms and large-scale simulations of space systems and automotive electronics. He is interested in data mining, visualization, and mathematical modeling techniques for high-dimensional biological and epidemiological datasets created using various 'omics' technologies and/or during large population studies.

Endnote

* Correspondence address: Department of Psychology, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095-1563, USA. Email: bway@ucla.edu

References

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., et al. (1994). Socioeconomic status and health. The challenge of the gradient. *American Psychologist*, **49**, 15–24.

Altmuller, J., Palmer, L. J., Fischer, G., Scherb, H., & Wjst, M. (2001). Genomewide scans of complex human diseases: True linkage is hard to find. *American Journal of Human Genetics*, **69**, 936–950.

Bacanu, S. A., Devlin, B., & Roeder, K. (2002). Association studies for quantitative traits in structured populations. *Genetic Epidemiology*, **22**, 78–93.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bird, A. (2007). Perceptions of epigenetics. *Nature*, **447**, 396–398.

Bradley, C. C., & Blakely, R. D. (1997). Alternative splicing of the human serotonin transporter gene. *Journal of Neurochemistry*, **69**, 1356–1367.

Bradley, S. L., Dodelzon, K., Sandhu, H. K., & Philibert, R. A. (2005). Relationship of serotonin transporter gene polymorphisms and haplotypes to mRNA transcription. *American Journal of Medical Genetics. Part B Neuropsychiatric Genetics*, **136**, 58–61.

Brenner, S., Jacob, F., & Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, **190**, 576–581.

Campbell, C. D., Ogburn, E. L., Lunetta, K. L., Lyon, H. N., Freedman, M. L., Groop, L. C., et al. (2005). Demonstrating stratification in a European American population. *Nature Genetics*, **37**, 868–872.

Cardon, L. R., & Bell, J. I. (2001). Association study designs for complex diseases. *Nature Reviews Genetics*, **2**, 91–99.

Cardon, L. R., & Palmer, L. J. (2003). Population stratification and spurious allelic association. *Lancet*, **361**, 598–604.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science*, **301**, 386–389.

Cho, H. J., Meira-Lima, I., Cordeiro, Q., Michelon, L., Sham, P., Vallada, H., et al. (2005). Population-based and family-based studies on the serotonin transporter gene polymorphisms and bipolar disorder: A systematic review and meta-analysis. *Molecular Psychiatry*, **10**, 771–781.

Cohen, S., Kaplan, G. A., & Salonen, G. T. (1999). The role of psychological characteristics in the relation between socioeconomic status and perceived health. *Journal of Applied Social Psychology*, **29**, 445–468.

Colhoun, H. M., McKeigue, P. M., & Davey Smith, G. (2003). Problems of reporting genetic associations with complex outcomes. *Lancet*, **361**, 865–872.

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. *Nature Genetics*, **9**, 347–350.

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *Lancet*, **366**, 1121–1131.

Crick, F. H. C. (1958). On protein synthesis. *Symposia of the Society for Experimental Biology, XII*, **139**, 138–163.

den Dunnen, J. T., & Antonarakis, S. E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Human Mutation*, **15**, 7–12.

den Dunnen, J. T., & Antonarakis, S. E. (2001). Nomenclature for the description of human sequence variations. *Human Genetics*, **109**, 121–124.

Devlin, B., & Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, **29**, 311–322.

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, **55**, 997–1004.

Dewan, A., Liu, M., Hartman, S., Zhang, S. S., Liu, D. T., Zhao, C., et al. (2006). HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science*, **314**, 989–992.

Dostert, A., & Heinzel, T. (2004). Negative glucocorticoid receptor response elements and their role in glucocorticoid action. *Current Pharmaceutical Design*, **10**, 2807–2816.

Draghici, S., Khatri, P., Eklund, A. C., & Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends in Genetics*, **22**, 101–109.

Duerr, R. H., Taylor, K. D., Brant, S. R., Rioux, J. D., Silverberg, M. S., Daly, M. J., et al. (2006). A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*, **314**, 1461–1463.

Durkheim, E. (1951). *Suicide*. New York: The Free Press.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., et al. (2004). Gene-environment interaction analysis of serotonin system markers with adolescent depression. *Molecular Psychiatry*, **9**, 908–915.

Feinn, R., Nellissery, M., & Kranzler, H. R. (2005). Meta-analysis of the association of a functional serotonin transporter promoter polymorphism with alcohol dependence. *American Journal of Medical Genetics. Part B Neuropsychiatric Genetics*, **133**, 79–84.

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, **7**, 85–97.

Fischer, A., Sananbenesi, F., Wang, X., Dobbin, M., & Tsai, L. H. (2007). Recovery of learning and memory is associated with chromatin remodelling. *Nature*, **447**, 178–182.

Forrest, L. R., Tavoulari, S., Zhang, Y. W., Rudnick, G., & Honig, B. (2007). Identification of a chloride ion binding site in Na^+/Cl -dependent transporters. *Proceedings of the National Academy of Sciences, USA*, **104**, 12761–12766.

Freimer, N. B., & Sabatti, C. (2004). The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. *Nature Genetics*, **36**, 1045–1051.

Freimer, N. B., & Sabatti, C. (2005). Guidelines for association studies in Human Molecular Genetics. *Human Molecular Genetics*, **14**, 2481–2483.

Gelernter, J., Kranzler, H., Coccaro, E. F., Siever, L. J., & New, A. S. (1998). Serotonin transporter protein gene polymorphism and personality measures in African American and European American subjects. *American Journal of Psychiatry*, **155**, 1332–1338.

Gibson, G., & Muse, S. V. (2004). *A Primer of Genome Science* (Vol. 2nd edn). Sunderland, MA: Sinauer Associates.

Glatt, C. E., DeYoung, J. A., Delgado, S., Service, S. K., Giacomini, K. M., Edwards, R. H., et al. (2001). Screening a large reference sample to identify very low frequency sequence variants: Comparisons between two genes. *Nature Genetics*, **27**, 435–438.

Glatz, K., Mossner, R., Heils, A., & Lesch, K. P. (2003). Glucocorticoid-regulated human serotonin transporter (5-HTT) expression is modulated by the 5-HTT gene-promotor-linked polymorphic region. *Journal of Neurochemistry*, **86**, 1072–1078.

Goertzel, B. N., Pennachin, C., de Souza Coelho, L., Gurbaxani, B., Maloney, E. M., & Jones, J. F. (2006). Combinations of single nucleotide polymorphisms in neuroendocrine effector and receptor genes predict chronic fatigue syndrome. *Pharmacogenomics*, **7**, 475–483.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., et al. (2005). The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science*, **307**, 1434–1440.

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, **160**, 636–645.

Gottesman, I. I., & Shields, J. (1967). A polygenic theory of schizophrenia. *Proceedings of the National Academy of Sciences, USA*, **58**, 199–205.

Halpern, C. T., Kaestle, C. E., Guo, G., & Hallfors, D. D. (2006). Gene-environment contributions to young adult sexual partnering. *Archives of Sexual Behavior*, **36**, 543–554.

Hardy, J., & Gwinn-Hardy, K. (1998). Genetic classification of primary neurodegenerative disease. *Science*, **282**, 1075–1079.

Hariri, A. R., Drabant, E. M., Munoz, K. E., Kolachana, B. S., Mattay, V. S., Egan, M. F., et al. (2005). A susceptibility gene for affective disorders and the response of the human amygdala. *Archives of General Psychiatry*, **62**, 146–152.

Hattersley, A. T., & McCarthy, M. I. (2005). What makes a good genetic association study? *Lancet*, **366**, 1315–1323.

Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: Physiological, pathological and therapeutic implications of human membrane transport proteinsIntroduction. *Pflugers Archiv*, **447**, 465–468.

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D., et al. (1996). Allelic variation of human serotonin transporter gene expression. *Journal of Neurochemistry*, **66**, 2621–2624.

Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*, **316**, 1491–1493.

Helgason, A., Yngvadottir, B., Hrafnkelsson, B., Gulcher, J., & Stefansson, K. (2005). An Icelandic example of the impact of population structure on association studies. *Nature Genetics*, **37**, 90–95.

Henry, L. K., Field, J. R., Adkins, E. M., Parnas, M. L., Vaughan, R. A., Zou, M. F., et al. (2006). Tyr-95 and Ile-172 in transmembrane segments 1 and 3 of human serotonin transporters interact to establish high affinity recognition of antidepressants. *Journal of Biological Chemistry*, **281**, 2012–2023.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**, 95–108.

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, **4**, 45–61.

Hoffman, B. J., Hansson, S. R., Mezey, E., & Palkovits, M. (1998). Localization and dynamic regulation of biogenic amine transporters in the mammalian central nervous system. *Frontiers in Neuroendocrinology*, **19**, 187–231.

Hutchison, K. E., Stallings, M., McGeary, J., & Bryan, A. (2004). Population stratification in the candidate gene study: Fatal threat or red herring? *Psychological Bulletin*, **130**, 66–79.

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., et al. (2004). Detection of large-scale variation in the human genome. *Nature Genetics*, **36**, 949–951.

Insel, T. R., & Lehner, T. (2007). A new era in psychiatric genetics? *Biological Psychiatry*, **61**, 1017–1018.

International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature* **437**, 1299–1320.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

Ioannidis, J. P. (2006). Commentary: Grading the credibility of molecular evidence for complex diseases. *International Journal of Epidemiology*, **35**, 572–578; discussion 593–576.

Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *Lancet*, **361**, 567–571.

Jabbi, M., Korf, J., Kema, I. P., Hartman, C., van der Pompe, G., Minderaa, R. B., et al. (2007). Convergent genetic modulation of the endocrine stress response involves polymorphic variations of 5-HTT, COMT and MAOA. *Molecular Psychiatry*, **12**, 483–490.

Kaiser, R., Tremblay, P. B., Roots, I., & Brockmoller, J. (2002). Validity of PCR with emphasis on variable number of tandem repeat analysis. *Clinical Biochemistry*, **35**, 49–56.

Kaiser, R., Tremblay, P. B., Schmider, J., Henneken, M., Dettling, M., Muller-Oerlinghausen, B., et al. (2001). Serotonin transporter polymorphisms: No association with response to antipsychotic treatment, but associations with the schizoparanoid and residual subtypes of schizophrenia. *Molecular Psychiatry*, **6**, 179–185.

Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between genes and synapses. *Science*, **294**, 1030–1038.

Kaufman, J., Yang, B. Z., Douglas-Palumberi, H., Houshyar, S., Lipschitz, D., Krystal, J. H., et al. (2004). Social supports and serotonin transporter gene moderate depression in maltreated children. *Proceedings of the National Academy of Sciences, USA*, **101**, 17316–17321.

Kendler, K. S. (2005). 'A gene for ...': The nature of gene action in psychiatric disorders. *American Journal of Psychiatry*, **162**, 1243–1252.

Kilic, F., Murphy, D. L., & Rudnick, G. (2003). A human serotonin transporter mutation causes constitutive activation of transport activity. *Molecular Pharmacology*, **64**, 440–446.

Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., et al. (2007). A 'silent' polymorphism in the MDR1 gene changes substrate specificity. *Science*, **315**, 525–528.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge, MA.: Cambridge University Press.

Knight, J. C. (2005). Regulatory polymorphisms underlying complex disease traits. *Journal of Molecular Medicine*, **83**, 97–109.

Kruglyak, L., & Nickerson, D. A. (2001). Variation is the spice of life. *Nature Genetics*, **27**, 234–236.

Lander, E. S., & Schork, N. J. (1994). Genetic dissection of complex traits. *Science*, **265**, 2037–2048.

Lesch, K. P., Balling, U., Gross, J., Strauss, K., Wolozin, B. L., Murphy, D. L., et al. (1994). Organization of the human serotonin transporter gene. *Journal of Neural Transmission General Section*, **95**, 157–162.

Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., et al. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science*, **274**, 1527–1531.

Lesch, K. P., Meyer, J., Glatz, K., Flugge, G., Hinney, A., Hebebrand, J., et al. (1997). The 5-HT transporter gene-linked polymorphic region (5-HTTLPR) in evolutionary perspective: Alternative biallelic variation in rhesus monkeys. Rapid communication. *Journal of Neural Transmission*, **104**, 1259–1266.

Li, D., & He, L. (2007). Meta-analysis supports association between serotonin transporter (5-HTT) and suicidal behavior. *Molecular Psychiatry*, **12**, 47–54.

Lim, J. E., Papp, A., Pinsonneault, J., Sadee, W., & Saffen, D. (2006). Allelic expression of serotonin transporter (SERT) mRNA in human pons: Lack of correlation with the polymorphism SERTLPR. *Molecular Psychiatry*, **11**, 649–662.

Lin, P. Y. (2007). Meta-analysis of the association of serotonin transporter gene polymorphism with obsessive-compulsive disorder. *Progress in Neuro-psychopharmacology & Biological Psychiatry*, **31**, 683–689.

Lupski, J. R., de Oca-Luna, R. M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B. J., et al. (1991). DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell*, **66**, 219–232.

Mann, J. J., Huang, Y. Y., Underwood, M. D., Kassir, S. A., Oppenheim, S., Kelly, T. M., et al. (2000). A serotonin transporter gene promoter polymorphism (5-HTTLPR) and prefrontal cortical binding in major depression and suicide. *Archives of General Psychiatry*, **57**, 729–738.

Manuck, S. B., Flory, J. D., Ferrell, R. E., & Muldoon, M. F. (2004). Socio-economic status covaries with central nervous system serotonergic responsivity as a function of allelic variation in the serotonin transporter gene-linked polymorphic region. *Psychoneuroendocrinology*, **29**, 651–668.

Marchini, J., Cardon, L. R., Phillips, M. S., & Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nature Genetics*, **36**, 512–517.

Martin, J., Cleak, J., Willis-Owen, S. A., Flint, J., & Shifman, S. (2007). Mapping regulatory variants for the serotonin transporter gene based on allelic expression imbalance. *Molecular Psychiatry*, **12**, 421–422.

McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science*, **316**, 1488–1491.

Meaney, M. J., & Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: Life at the interface between a dynamic environment and a fixed genome. *Dialogues Clinical Neurosciences*, **7**, 103–123.

Meyer-Lindenberg, A., & Weinberger, D. R. (2006). Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nature Reviews Neuroscience*, **7**, 818–827.

Miller, L., & Gur, M. (2002). Religiousness and sexual responsibility in adolescent girls. *Journal of Adolescent Health*, **31**, 401–406.

Moffitt, T. E. (2005). The new look of behavioral genetics in developmental psychopathology: Gene-environment interplay in antisocial behaviors. *Psychological Bulletin*, **131**, 533–554.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry*, **62**, 473–481.

Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., et al. (2006). Human catechol-*O*-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science*, **314**, 1930–1933.

Nature Genetics. (2005). Framework for a fully powered risk engine. *Nature Genetics*, **37**, 1153.

Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences, USA*, **47**, 1588–1602.

Otsu, M., & Sitia, R. (2007). Diseases originating from altered protein quality control in the endoplasmic reticulum. *Current Medicinal Chemistry*, **14**, 1639–1652.

Palmer, L. J., & Cardon, L. R. (2005). Shaking the tree: Mapping complex disease genes with linkage disequilibrium. *Lancet*, **366**, 1223–1234.

Parsey, R. V., Hastings, R. S., Oquendo, M. A., Hu, X., Goldman, D., Huang, Y. Y., et al. (2006). Effect of a triallelic functional polymorphism of the serotonin-transporter-linked promoter region on expression of serotonin transporter in the human brain. *American Journal of Psychiatry*, **163**, 48–51.

Pearson, H. (2006). Genetic information: Codes and enigmas. *Nature*, **444**, 259–261.

Philibert, R., Madan, A., Andersen, A., Cadoret, R., Packer, H., & Sandhu, H. (2007). Serotonin transporter mRNA levels are associated with the methylation of an upstream CpG island. *American Journal of Medical Genetics. Part B Neuropsychiatric Genetics*, **144**, 101–105.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, **264**, 1733–1739.

Pritchard, J. K., & Rosenberg, N. A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics*, **65**, 220–228.

Rapport, M. M., Green, A. A., & Page, I. H. (1948). Crystalline Serotonin. *Science*, **108**, 329–330.

Ravna, A. W., Sylte, I., Kristiansen, K., & Dahl, S. G. (2006). Putative drug binding conformations of monoamine transporters. *Bioorganic & Medicinal Chemistry*, **14**, 666–675.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature*, **444**, 444–454.

Risch, N. (1990). Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genetic Epidemiology*, **7**, 3–16; discussion 17–45.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.

Robinson, T. R. (2005). *Genetics for Dummies*. Hoboken, NJ: Wiley.

Rohrbaugh, J., & Jessor, R. (1975). Religiosity in youth: A personal control against deviant behavior. *Journal of Personality*, **43**, 136–155.

Rostosky, S. (2004). The impact of religiosity on adolescent sexual behavior: A review of the evidence. *Journal of adolescent research*, **19**, 677–697.

Schwartz, D., & Collins, F. (2007). Medicine. Environmental biology and human disease. *Science*, **316**, 695–696.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.

Shahbazian, M. D., & Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annual Review of Biochemistry*, **76**, 75–100.

Singh-Manoux, A., Marmot, M. G., & Adler, N. E. (2005). Does subjective social status predict health and change in health status better than objective status? *Psychosomatic Medicine*, **67**, 855–861.

Snyder, M., & Gerstein, M. (2003). Genomics. Defining genes in the genomics era. *Science*, **300**, 258–260.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, **315**, 848–853.

Sullivan, P. F. (2007). Spurious genetic associations. *Biological Psychiatry*, **61**, 1121–1126.

Suomi, S. J. (2006). Risk, resilience, and gene x environment interactions in rhesus monkeys. *Annals of the New York Academy of Sciences*, **1094**, 52–62.

Sutcliffe, J. S., Delahanty, R. J., Prasad, H. C., McCauley, J. L., Han, Q., Jiang, L., et al. (2005). Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. *American Journal of Human Genetics*, **77**, 265–279.

Taylor, S. E., Way, B. M., Welch, W. T., Hilmert, C. J., Lehman, B. J., & Eisenberger, N. I. (2006). Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive symptomatology. *Biological Psychiatry*, **60**, 671–676.

Thomas, D. C., & Witte, J. S. (2002). Point: Population stratification: A problem for case-control studies of candidate-gene associations? *Cancer Epidemiology, Biomarkers & Prevention*, **11**, 505–512.

Trinh, R., Gurbaxani, B., Morrison, S. L., & Seyfzadeh, M. (2004). Optimization of codon pair use within the (GGGGS)3 linker sequence results in enhanced protein expression. *Molecular Immunology*, **40**, 717–722.

Wacholder, S., Rothman, N., & Caporaso, N. (2002). Counterpoint: Bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. *Cancer Epidemiology, Biomarkers & Prevention*, **11**, 513–520.

Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., & Rothman, N. (2004). Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, **96**, 434–442.

Wendland, J. R., Martin, B. J., Kruse, M. R., Lesch, K. P., & Murphy, D. L. (2006). Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR and rs25531. *Molecular Psychiatry*, **11**, 224–226.

Wolfsberg, T. G., Wetterstrand, K. A., Guyer, M. S., Collins, F. S., & Baxevanis, A. D. (2003). A user's guide to the human genome. *Nature Genetics*, **35**(Suppl 1), 4.

Wong, D. T., Bymaster, F. P., & Engleman, E. A. (1995). Prozac (fluoxetine, Lilly 110140), the first selective serotonin uptake inhibitor and an antidepressant drug: Twenty years since its first publication. *Life Sciences*, **57**, 411–441.

Yan, H., & Zhou, W. (2004). Allelic variations in gene expression. *Current Opinion in Oncology*, **16**, 39–43.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., & Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science*, **297**, 1143.

Yang, Z., Camp, N. J., Sun, H., Tong, Z., Gibbs, D., Cameron, D. J., et al. (2006). A variant of the HTRA1 gene increases susceptibility to age-related macular degeneration. *Science*, **314**, 992–993.

Yonan, A. L., Palmer, A. A., & Gilliam, T. C. (2006). Hardy-Weinberg disequilibrium identified genotyping error of the serotonin transporter (SLC6A4) promoter polymorphism. *Psychiatric Genetics*, **16**, 31–34.

Zapala, M. A., & Schork, N. J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences, USA*, **103**, 19430–19435.

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**, 1336–1341.

Appendix

When are genetic effects deterministic?

The notion of genes having deterministic effects is largely a result of historical residue from early medical genetics studies of rare, debilitating disorders such as Huntington's disease, cystic fibrosis, or sickle-cell anemia. These highly heritable disorders are often sardonically referred to as one gene one disorder (O GOD) diseases because they are caused by alterations in a single gene that inevitably lead to the onset of disease (Hardy & Gwinn-Hardy, 1998; Plomin, Owen, & McGuffin, 1994). Although genetic effects in O GODs are deterministic, O GODs are extremely rare. Unfortunately, the environment is largely irrelevant in such disorders. However, genetic effects in other health outcomes are more subtle. For most health outcomes, however, genetic effects are complex; that is, health outcomes are influenced by many genes each of small effect.

How large are the typical effect sizes associating genetic variants with social health outcomes?

The effects sizes for associations between a particular genetic variant and a health outcome are generally small. For example, across meta-analyses associating the 5-HTTLPR with alcohol dependence (Feinn, Nellissery, & Kranzler, 2005), suicide (Li & He, 2007), bipolar disorder (Cho et al., 2005), and obsessive-compulsive disorder (Lin, 2007), the odds ratios range between 1.12 to 1.2. The odds ratio is often used as a measure of effect size in such case–control association studies due to the non-continuous nature of the variables. Technically speaking, the odds ratio reflects the probability that a group possessing a certain gene variant (e.g., short allele) has increased or decreased risk for a given health outcome, such as depression, than a control group. Thus, an odds ratio of 2 means that group A has twice the risk of group B. For the odds ratio to be significant, researchers look for the 95% lower (or upper) bound of the odds ratio to be greater than (or less than) 1.0 as an indication of increased (or decreased) risk. By way of comparison, the odds ratio between exposure to significant life stressors and risk for a diagnosis of depression is around 12 (Kendler, 2005). Hence, genetic effects can be difficult to detect unless the environment is carefully evaluated or the sample sizes are extremely large.

How are genes named?

As one can imagine, with around 24,000 currently identified genes a systematic nomenclature is necessary to avoid confusion. Hence, a nomenclature committee formed by the Human Genome Organization, an international scientific organization, approves a specific name and abbreviation for every gene (http://www.gene.ucl.ac.uk/nomenclature/). Typically, these names consist of a string of Latin letters or Arabic numbers that is no more than six symbols in length. For example, the serotonin transporter gene is referred to as *SLC6A4* to indicate that within solute carrier family 6, it is the fourth transporter (the A simply acts as a spacer between numbers), which is the family of neurotransmitter transporters (Hediger et al., 2004). Other members of this family include the dopamine transporter (*SCL6A3*) and the norepinephrine transporter (*SLC6A2*). Because the gene and the protein often have the same name, confusion is avoided by referring to the gene in italics and the protein for which it codes in standard type.

Where are genes located?

The entire genome is partitioned across 46 chromosomes, which are simply long stretches of DNA stored in the nucleus of each cell. Each

gene resides at a particular address on both pairs of a particular chromosome, either the 22 autosomes (non-sex chromosomes) or the sex chromosomes (X and Y). This address is referred to with differing levels of specificity, much like the difference between a zip code and a street address. The more broad and commonly used address is derived from classical cytogenetic methods where the chromosomes were stained with a particular dye that produces a characteristic pattern of light and dark bands. Numbering of these bands gives the general location, or zip code, of the gene. For example, the location of the serotonin transporter gene (*SLC6A4*) is 17q11.2. The 17 indicates that the gene is located on chromosome 17, whereas the q indicates that it is located on the long arm. Each chromosome has a short (denoted with a p) and a long (q) arm that are separated by the centromere, which refers to a position near the center of the chromosome where the DNA is compacted. The final number indicates distance from the centromere and the higher the number, the further from the centromere (and closer to the end of the chromosome) the gene is.

With the information provided by the recently completed sequencing of the human genome, it is possible to specify a more precise location for each gene, the genomic location. In this case, the serotonin transporter gene is located on chromosome 17 between bases 25,549,032 and 25,586,831 (Build, or version number, 35) where the numbering of the base pairs begins at the tip of the short end of the chromosome. This numbering changes slightly with each new revision of the reference sequence that is stored in the NCBI database. For a guide to genetic information stored in the NCBI database, see Wolfsberg, Wetterstrand, Guyer, Collins, and Baxevanis (2003).

Does the same DNA sequence always code for the same protein?

Although the dogma that amino acid sequence determines the structure and thus the function of a protein usually holds, violations of the dogma are known to exist. For example, some proteins, especially those secreted from the cell, bound to the cell surface (e.g., the serotonin transporter), or ejected from the cell (e.g., antibodies in the immune system), need to be translated in the presence of accessory proteins, such as 'chaperones' (proteins that surround or otherwise protect thermally less stable proteins from other proteins while they are folding) and glycotransferases (that attach sugars to proteins at key places in their sequence) in order to fold properly (Otsu & Sitia, 2007). Proteins can assume different conformations, and possibly even functions, for the same sequence of amino acids depending on the presence of these neighboring proteins. These variant structures are most often either non-functioning or pathological.

Can synonymous SNPs affect protein function?

Although the bulk of research attention has focused on polymorphisms that change the amino acid sequence or are located in regulatory regions, there are also SNPs in the coding region of a gene that do not affect the amino acid sequence, which are called synonymous SNPs. With such SNPs the mRNA sequence is altered, but they still code for the same amino acid (i.e., 'UUA' and 'CUA' both code for the amino acid Leucine). Long thought to have no functional effect, such 'silent' SNPs have recently been shown to affect the degree of protein expression by altering the physical structure of the mRNA message or by affecting the speed of translation (Nackley et al., 2006; Trinh, Gurbaxani, Morrison, & Seyfzadeh, 2004). Some silent mutations have recently been shown to change protein structure and thus function (Kimchi-Sarfaty et al., 2007), which is requiring a revision of conventional theories of protein synthesis.

How do environmental signals trigger lasting changes in gene expression?

The effects of glucocorticoids upon serotonin transporter gene expression described in the text have effects lasting on the order of hours or days. A different process is thought to underlie changes in gene expression that can last for years and possibly even be passed on to future generations (Bird, 2007). These changes are often called epigenetic ('epi' for 'upon' or 'after') because they have enduring effects upon genes without changing the genetic sequence. An example of such a change is the permanent addition of a chemical (methyl) group to a particular DNA base (C when adjacent to a G, called a CpG site), a process referred to as methylation. The addition of the methyl group acts like a block or plug to prevent gene expression.

Perhaps the most relevant example for social health research is the effect of early maternal care upon adulthood stress responses in rats (Meaney & Szyf, 2005). The quality of care influences the methylation status of the glucocorticoid receptor, which leads to differences in expression of this receptor that persist into adulthood and are thought to underlie the lasting alterations in stress responsivity. Similar epigenetic effects are likely to be involved in the long-term effects of childhood stressors such as low SES. Furthermore, the moderating influence of the 5-HTTLPR upon the effects of SES may also be influenced by epigenetic changes as there is evidence that the different 5-HTTLPR alleles can be differentially methylated (Philibert et al., 2007). However, this latter observation is based upon studies in lymphoblasts and whether or not it occurs in neural tissue is unknown. This highlights one of the challenges in studying epigenetic mechanisms, which, like measuring gene expression, requires sampling the actual cells one is interested in studying. Hence, this is a methodology

that is unlikely to become as widely used by health psychologists as is genotyping.

What molecular mechanisms control gene expression?

The actual process of how gene expression is initiated is only beginning to be understood. For biophysical reasons the process is something of an enigma because DNA is highly negatively charged and hence tends to repel itself. Left to its own devices, this charge repulsion would lead DNA to stretch out–each copy of the DNA inside each human cell, if fully extended, would stretch to about 2 m in length. Compacting all of this material into the cell nucleus, which is one million times smaller, requires that DNA be wrapped tightly around positively charged proteins called histones. A chemical process called histone acetylation causes these proteins to release the DNA enough for it to be read during transcription. Because of this, the timing of the production of proteins that do histone acetylation and deacetylation is critical to proper gene expression (Shahbazian & Grunstein, 2007). These are not simple on and off processes, but rather are modulated in a subtle way by the exact pattern of letters in the histone genes themselves (Pearson, 2006). At present, drugs that disrupt the process of histone deacetylation have only been documented to affect cognitive processes in animal models (Fischer, Sananbenesi, Wang, Dobbin, & Tsai, 2007), but it is likely that such an intervention could modulate stress responses as well.

Are there always two copies of every gene?

In addition to the well-studied forms of genetic variation described in the main text, there is another form of variation that was originally thought to be limited to a few rare diseases (Lupski et al., 1991) but has recently been discovered (Iafrate et al., 2004; Sebat et al., 2004) to be surprisingly prevalent in normal individuals. This form of variation is called copy number variation (CNV; copy number polymorphism when present in >1% of the population) and refers to the deletion or duplication of a DNA segment 1000 base pairs or larger. In other words, a copy of a particular gene can be either missing or duplicated. The prevalence of CNVs is difficult to determine because of current technological limitations, but recent evidence indicates that CNV occurs in at least 12% of the genome (Redon et al., 2006), and there are likely to be several hundred CNVs in each individual (Feuk, Carson, & Scherer, 2006). CNV has been associated with risk for HIV infection as well as disease progression (Gonzalez et al., 2005), but it has not yet been associated with psychosocial variables or health risk behaviors. In light of the robust effects of CNV documented at the cellular level (Stranger et al., 2007), this is likely to be a promising area of future research.

What is linkage disequilibrium?

To understand linkage disequilibrium, imagine that a base (e.g., an A, G, T, or C) of your genetic code becomes mutated (sorry to be the bearer of bad news, but you probably have several). This mutation will be perfectly correlated with each and every other SNP on this chromosome. As you pass this mutation on to succeeding generations, the relationship between this mutation and the SNPs on this chromosome will change because you do not pass on an exact replica of this chromosome to your child. In the process of forming a sperm (or egg), there is a shuffling of DNA between the chromosome pairs (one of which came from your mother and one of which came from your father) that is referred to as recombination or crossing-over. Much like shuffling a new deck of cards, the insertion of chromosomal material from the other paired chromosome can separate the mutation from SNPs that were originally on the same chromosome. With each successive generation, there is another shuffle that is likely to reduce the association between the mutation and the SNPs that were on the original chromosome. The decrease in the relationship between the mutation and the other SNPs is primarily a function of the number of base pairs separating them on the chromosome (e.g., genetic distance). Thus, over a number of generations, the relationship between this mutation and a SNP located distally on the original chromosome will become random or independent; that is, they are in linkage *equilibrium*. To use the analogy of an ordered deck of cards, if the mutation were the King of Hearts, the relationship with a distantly located card (e.g., the three of hearts) is likely to become independent, whereas the Queen of Hearts would be the least likely to be separated by a shuffle and would remain in linkage disequilibrium with the King of Hearts. Statistically, the measure of linkage disequilibrium is D' or r^2 (Devlin & Risch, 1995). For a more thorough discussion of LD, see Palmer and Cardon (2005).

Where is the field of genetics headed?

Recent technological developments have created the opportunity for a different approach to conducting genetic association studies. Rather than genotyping a few markers in a few genes, it is now possible to genotype an individual for 500,000 markers simultaneously using SNP Chips (Hirschhorn & Daly, 2005). In this approach, probes for each SNP are etched onto a glass slide using technology very similar to that used in etching computer chips. Fifteen years ago, it would have taken a researcher decades to genotype 500,000 markers and now it can be done in an afternoon. Although it is not yet technologically possible to sample every SNP in the genome, 500,000 SNPs is sufficient to sample nearly every haplotype. (A haplotype refers to a cluster of adjacent SNPs on a single chromosome). Because of the high correlation (linkage disequilibrium) among all SNPs

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

in a particular haplotype block, only one SNP is needed to delineate a particular genetic neighborhood from other areas in the genome. Thus, each haplotype can be represented by a single SNP, which is called a tagging SNP. These tagging SNPs were identified as part of the international haplotype mapping project (http://www.HapMap.org) which has identified the haplotype blocks throughout the genome for four different ancestral groups (International HapMap Consortium, 2005).

The use of SNP chips is part of the currently unfolding transformation of biomedical research. At each level of the central dogma (DNA, RNA, and protein), it is possible to use such high-throughput technologies to generate data on hundreds of thousands of markers relatively quickly and economically. Referred to as genomics, proteomics, epigenomics, or a host of other '-omics,' this approach is not only changing the technology used to do research, but also the theoretical approach as well.

With the ability to screen so many genes and proteins at once, medical research no longer has to be hypothesis driven, but can be data driven instead. Rather than testing a particular hypothesis, one can instead collect volumes of '-omics' data and 'listen' to it in order to ascertain what the data itself is saying about the disease–an approach known as data mining in other contexts. One no longer has to have a hypothesis as to where to look to find the needle in the haystack, rather you can assay the entire haystack in an afternoon and try to pick apart the output with computers. These haystack-sifting computer algorithms form another emerging discipline called 'bioinformatics'. As one can imagine, identifying and analyzing the data associated with 500,000 SNPs requires new analytic approaches in order to trust the numbers.

One such approach is to use the False Discovery Rate (Benjamini & Hochberg, 1995). Simply put, this is the proportion of alternate hypotheses the researcher is accepting which should in fact be rejected. It is especially useful in genomics studies when hundreds of thousands or even millions of independent variables are involved, and criteria such as the Bonferroni correction that minimizes the number of Type 1 errors wind up being too strict. FDR effectively loosens up the Type 1 error rate in multiple hypothesis testing situations where one is generating many alternate hypotheses, and the cost of pursuing a false hypothesis is not large. Thus, FDR is typically used in large genomics studies, dealing with thousands of genes, that are in reality 'fishing expeditions' and the important thing is to identify a lot of candidate genes that can later be 'cleaned up' by more accurate methods on much smaller subsets of the data. As such, it is not uncommon for genomics researchers to analyze their data with an FDR set to 30% to 50%, yielding hundreds of candidate genes that are later reduced by 50% or more. Were a Bonferroni correction used, it is likely that no candidate genes would be identified.

Preliminary indications from several recent studies have generated optimism that the whole genome association approach will reliably identify

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x

polymorphisms influencing disease. In the last year, new causal variants or markers in linkage disequilibrium with them have been identified for myocardial infarction (Helgadottir et al., 2007; McPherson et al., 2007), Type II diabetes (Zeggini et al., 2007), inflammatory bowel disease (Duerr et al., 2006), and adult onset macular degeneration (Dewan et al., 2006; Yang et al., 2006). In addition to the technological and theoretical shifts underlying such studies, they also represent a shift towards large consortiums (e.g., Genetic Association Information Network, GAIN, (Insel & Lehner, 2007), or the Genes and Environment Initiative (Schwartz & Collins, 2007)) in order to enroll sufficient subjects necessary for identifying the weak effects of any one particular variant. Whether or not this Wal-Martification of science will generate more than just optimism remains an open question.

Based on recent experience using high-throughput technologies for the measurement of gene expression (Microarrays), there is reason to temper such optimism. Many of the initial, high-profile findings produced by such microarrays could not be replicated (Draghici, Khatri, Eklund, & Szallasi, 2006). The noisiness of the data resulting from such technology as well as the sheer number of comparisons may contribute to this irreproducibility. The seductive appeal of high classification accuracies and visually striking data clustering can actually be artifacts of overfitting: when one has a large number of independent variables to work with (100,000's or more, as discussed above) on small sample sizes (typically less than a few hundred individuals), overfitting can be accomplished quite easily, often without the investigator being aware of it. In many ways, the technology is outpacing the ability to reliably analyze it and new analytical approaches are needed.

This is not to say that 'omics' and bioinformatics results should never be trusted, but only that they should be evaluated carefully and with a wary eye, and this can be done even without being statistically or computationally savvy. One should look for results not that simply have a good *P*-value when evaluated one dimension at a time or that seem to cluster the data with high accuracy, but results that are reproducible and stand up to cross-validation, bootstrapping, or Monte Carlo analysis.

 Social and Personality Psychology Compass 2/2 (2008): 785–816, 10.1111/j.1751-9004.2008.00084.x