


By Louis J. Přeček, Christopher R. Jones, and Ying-Hui Fu

Abstract

Tremendous progress in the field of human genetics has made a major impact over the last two decades into our understanding of many Mendelian disorders affecting humans. It is much more difficult to approach the genetics of complex disorders and is particularly challenging in the case of dissecting genetics of the wide range of behavioral variation in the general population. Knowledge of the biology and genetics of Mendelian traits is beginning to inform studies in complex genetics. Furthermore, the convergence of this growing knowledge with increasingly powerful new tools now put complex genetics of behavior within reach. The bottleneck in such studies and the greatest challenge to investigators is the rigorous phenotyping of human research subjects.

Introduction

The realization that recombination mapping could allow localization and ordering of genes on a chromosome first came in the field of phage genetics. The mathematical algorithms for statistical analysis of mapping data were established over 50 years ago (Haldane, 1934; Morton, 1955). However, it was not until the latter part of the last century that application of these principles to the mapping and cloning of genes causing human diseases was proposed (Botstein et al., 1980).

This effort was led by Ray White, whose group collected a majority of the 60 multigenerational CEPH families that are used worldwide for the mapping of human genetic markers. Ascertainment of four living grandparents, two parents, and a sibship of at least 10 children was required for inclusion. These families are necessary for tracing the segregation of polymorphic alleles in families and recognition of whether alleles from two markers are traveling together or separated by recombination during meiosis. It is critical to "set phase" by determining which contributions of an individual's DNA came from mother and father. While simple in concept, the collection of these families was the major effort that enabled all future mapping and cloning efforts in human genetics. Arguably, these initial gene-mapping efforts marked the beginning of the Human Genome Project.

Following this paradigm, scientists began to make progress late in the 1980s and early 1990s with the identification of genes for diseases such as Duchenne muscular dystrophy (Koenig et al., 1987), neurofibromatosis type 1 (Cawthon et al., 1990; Viskochil et al., 1990), fragile X mental retardation (Fu et al., 1991), and myotonic dystrophy (Fu et al., 1992), to name a few. During the past 15 years, these reagents have led to an explosion of the molecular characterization of many Mendelian disorders.

The paradigm is one of phenotyping large families segregating Mendelian disease alleles and then using genetic markers to map their location in the genome. Subsequently, positional cloning in the critical region defined by recombination events in these families can lead to identification of the disease gene.

Thus, a genetic marker that is close to a disease allele in the genome is physically linked and can only be separated in the event of rare recombination events between the marker and the disease gene itself. Even if recombination has occurred between a marker allele and the disease mutation, linkage can still be seen in large families as one can calculate the probability of a marker allele to segregate with the disease phenotype as a function of being linked vs traveling with the disease allele by sheer coincidence.

Human behavior is genetically programmed to an extent. In addition, genes interact with the environment to manifest as the rich spectrum of human behaviors that we recognize in the population. Because of the complexity (presumably many genes contributing to a particular behavior and the interaction of those genes with other genes and with the environment), the genetics of behavior was a topic that could not be addressed until recently.

The first efforts at characterizing genetic aspects of behavior began in fruit flies in the late 1960s by Seymour Benzer and colleagues (Konopka and Benzer, 1971). Benzer and colleagues hypothesized that by devising simple assays of fruit fly behavior (circadian activity rhythms, learning and memory paradigms, mating behavior), genetic screens could be done to...
identify phenotypic variants caused by single genes. This approach was remarkably successful and continues to develop in modern behavioral genetics. The earliest mutant recognized as associated with a Mendelian behavioral phenotype was the period mutation (Konopka and Benzer, 1971). Additional screening identified an allelic series of mutants causing both long and short periods. These investigations culminated a number of years later in identification of the period gene and subsequent characterization of the genetic variants in the mutant flies (Bargiello et al., 1984; Zehring et al., 1984). Since that time, additional Drosophila circadian rhythm genes have been identified, and such studies have also expanded to include rodent circadian rhythm genes (Dunlap, 1999; King and Takahashi, 2000; Reppert, 1998; Wager-Smith and Kay, 2000).

The challenges of translating this work into the study of genetic factors contributing to human behavioral variation are onerous. It is not ethical to lock humans in a dark room and observe their activity rhythm for weeks at a time without their permission. The modern-day free running experiments in humans are extremely laborious and costly. Furthermore, because of the commitment of time and energy of such study subjects, it is difficult to recruit particular individuals of interest at any high frequency. Finally, it is not possible or desirable to do mutagenesis in humans and screening for mutant phenotypes.

However, extending these kinds of studies into humans is absolutely critical to understanding human circadian function. There is no doubt by experts in the field that many elements of circadian clocks are conserved among mammals and even vertebrates and into invertebrates. However, humans are unique and distinctly different from other organisms, including rodents. While the genomes of humans and rodents are quite homologous, the differences that allow us to function in uniquely human endeavors promise to be an exciting and productive area for research over the coming decades. Finally, because of the ability of human subjects to communicate far more extensively than nonhuman research organisms, it is possible to ask research questions in human populations that go far beyond simple behavioral paradigms used to study other organisms. Thus, it is critically important to translate current knowledge to a study of human behaviors, including circadian behavior.

Historically, it has only been possible to gain insights into genetically programmed phenotypes through the study of genetic variants in an organism that have a causal relationship with the phenotype. This has, until recently, been impossible in human circadian function as there were no recognized Mendelian variants in circadian function. This changed, however, with the description of the first Mendelian circadian rhythm disorder called familial advanced sleep phase syndrome (FASPS) (Jones et al., 1999).

Because genetic screens are not available in humans, one must begin by identifying individuals with phenotypes of interest. For long- and short-period circadian rhythm variants, one would predict that such individuals should have an advanced or delayed sleep phase, respectively. Because the variation in sleep- and wake-time preference is so large in the general population, selecting individuals with the most dramatic phenotypes has the highest probability of identifying a single-gene circadian trait. Then, through examination and phenotyping of first- and second-degree relatives, it is possible to establish whether the trait is segregating in a Mendelian fashion. However, one must be mindful that the wide spectrum of variation in both general population and FASPS families can lead to errors in phenotyping (Fig. 1).

**Identification of Probands**

The challenge of translating this work into the study of genetic factors contributing to human behavioral variation are onerous. It is not ethical to lock humans in a dark room and observe their activity rhythm for weeks at a time without their permission. The modern-day free running experiments in humans are extremely laborious and costly. Furthermore, because of the commitment of time and energy of such study subjects, it is difficult to recruit particular individuals of interest at any high frequency. Finally, it is not possible or desirable to do mutagenesis in humans and screening for mutant phenotypes.

However, extending these kinds of studies into humans is absolutely critical to understanding human circadian function. There is no doubt by experts in the field that many elements of circadian clocks are conserved among mammals and even vertebrates and into invertebrates. However, humans are unique and distinctly different from other organisms, including rodents. While the genomes of humans and rodents are quite homologous, the differences that allow us to function in uniquely human endeavors promise to be an exciting and productive area for research over the coming decades. Finally, because of the ability of human subjects to communicate far more extensively than nonhuman research organisms, it is possible to ask research questions in human populations that go far beyond simple behavioral paradigms used to study other organisms. Thus, it is critically important to translate current knowledge to a study of human behaviors, including circadian behavior.

Historically, it has only been possible to gain insights into genetically programmed phenotypes through the study of genetic variants in an organism that have a causal relationship with the phenotype. This has, until recently, been impossible in human circadian function as there were no recognized Mendelian variants in circadian function. This changed, however, with the description of the first Mendelian circadian rhythm disorder called familial advanced sleep phase syndrome (FASPS) (Jones et al., 1999).

Because genetic screens are not available in humans, one must begin by identifying individuals with phenotypes of interest. For long- and short-period circadian rhythm variants, one would predict that such individuals should have an advanced or delayed sleep phase, respectively. Because the variation in sleep- and wake-time preference is so large in the general population, selecting individuals with the most dramatic phenotypes has the highest probability of identifying a single-gene circadian trait. Then, through examination and phenotyping of first- and second-degree relatives, it is possible to establish whether the trait is segregating in a Mendelian fashion. However, one must be mindful that the wide spectrum of variation in both general population and FASPS families can lead to errors in phenotyping (Fig. 1).

**Enrollment and Sampling of Research Subjects**

Research subjects can be enrolled after appropriate informed consent. A huge bureaucracy has evolved around research in humans, particularly in genetics. Rules and regulations vary from institution to institution, and further discussion of this area would be fodder for an entire Methods in Enzymology series. Once subjects are enrolled, DNA must be obtained from each, usually by phlebotomy. DNA is purified from leukocytes through very standard methods (Piatcck et al., 1991).

**FIG. 1.** Normal distributions of sleep time preferences in FASPS families and the general population. The spectrum of sleep and wake time in the general population is quite large and spans from morning larks to night owls. Within FASPS kindreds that have been identified, DNA must be obtained from each, usually by phlebotomy. DNA is purified from leukocytes through very standard methods (Piatcck et al., 1991).
Phenotyping Subjects

To date, there is only one Mendelian circadian rhythm trait recognized in humans. FASPS is characterized by onset in infancy, childhood, or young adult life as a tendency to sleep before 8:30 PM and to wake spontaneously before 5:30 AM (in the absence of alarm clocks, light therapy, medications or other chemicals, etc.). These individuals have a normal quality and quantity of sleep, but their sleep bout is "advanced" in the solar day relative to that of more conventional sleepers (Jones et al., 1999). Inherent challenges exist because of the large variation of such traits in the general population and the masking of innate biological sleep/wake tendencies by psychosocial and familial/cultural factors to which humans are subject. Thus phenotyping must focus on disentangling the biological tendency from the milieu of factors affecting human behavior.

One approach is to assess the circadian period (\(\tau\)) in research subjects. \(\tau\) can be estimated under constant, so-called "free-running" (Kleitman and Kleitman, 1953) or forced desynchrony conditions (Dijk and Czeisler, 1991). One FASPS subject was shown to have a remarkably short \(\tau\) under constant conditions (Fig. 2, Jones et al., 1999). Performing such studies on all research subjects would be ideal but is not possible because not all subjects would submit to such time-consuming studies and because they are prohibitively difficult and expensive. The constant routine (Czeisler and Brown, 1985) is less time-consuming than estimating \(\tau\) but still requires 1–2 full days in the laboratory. The dim light melatonin onset (DLMO) is a reliable phase marker of the suprachiasmatic nucleus and can be accomplished in one evening using saliva samples (Lewy and Sack, 1989). Like all measures of circadian phase, the DLMO can theoretically be shifted somewhat by self-imposed light-dark cycles; the magnitude of this effect in research subjects is difficult to predict. The phase of the daily core body temperature nadir can now be measured conveniently and reliably in the laboratory. The dim light melatonin onset (DLMO) is a reliable phase marker of the suprachiasmatic nucleus and can be accomplished in one evening using saliva samples (Lewy and Sack, 1989). Like all measures of circadian phase, the DLMO can theoretically be shifted somewhat by self-imposed light-dark cycles; the magnitude of this effect in research subjects is difficult to predict. The phase of the daily core body temperature nadir can now be measured outside the laboratory by radiotelemetry using swallowed temperature transducers (Hamilos et al., 1998) but seems to be a noisier signal than the DLMO.

The rest-activity rhythm can be measured conveniently and reliably in ambulatory subjects using small, wrist-worn movement detectors known as actigraphs (Ancoli-Israel et al., 2003). Actigraphy correlates well with sleep recorded in the laboratory and is thought to be complemented by sleep logs. Again, many individuals are capable of behaviorally shifting their activity rhythm so as to mask or attenuate the magnitude of their "biological" circadian preference. Self-assessment instruments such as the Horne-Ostberg and Munich Chronotype questionnaires have been validated as tools for assessing sleep/wake preferences in humans (Horne and Ostberg, 1976; Roenneberg et al., 2001). The Horne-Ostberg morningness–eveningness score has been shown to correlate rather well with measured \(\tau\) (Duffy et al., 2001).

It is also critical to rule out other reasons for early sleep and wake times. Any cause of excessive daytime sleepiness (e.g., obstructive sleep apnea and narcolepsy) could manifest with early sleep times. Major depression causes early-morning awakening in many individuals. Thus, a general history and a depression index are critical to rule out other sleep disorders. Numerous tools are available for these purposes, and there is no clear consensus which would be of highest sensitivity and specificity for clinical circadian studies. At least for now, this must be left to the judgment of individual clinicians evaluating the subjects.

Our tendency to sleep and wake is further complicated by exogenous compounds that are commonly consumed by humans, including alcoholic beverages and caffeinated products. Furthermore, many of us make value judgments about our own or others' tendencies to sleep or wake at various times, and these may affect reporting by research subjects. Phenotyping of such subjects is extremely challenging for the reasons just mentioned and efforts are underway to improve on these assessments as discussed elsewhere in this volume (Chapter 10).
Mapping and Cloning Human Circadian Genes

Genetic linkage mapping requires polymorphic markers that can be genotyped in DNA samples from research families being studied. This is now usually done using polymerase chain reaction amplification of highly polymorphic repeat markers. The notion that such markers could be used to map genes causing human traits is predicated on the fact that genetic loci will segregate independently if they are on different chromosomes or far apart on the same chromosome. Marker alleles close to the genetic variant causing a trait will segregate with the trait (Fig. 3). The statistical measure of whether a marker allele is segregating with a trait because it is linked (vs by chance) has been studied beginning with the work of Haldane (1934) and development of the LOD score method based on sequential testing procedures (Morton, 1955). This method was incorporated in a computer algorithm (LINKAGE) that continues to be used today (Lathrop et al., 1985).

Once a locus has been localized, subsequent mapping with a dense array of markers allows more precise localization of recombinational events that delimit the critical region containing the gene of interest. The likelihood of narrowing the critical region increases with the number and size of the families being studied. Other methods can sometimes be used to further narrow the region and include homozygosity mapping and linkage disequilibrium mapping but are beyond the scope of this article (Hastbacka et al., 1992; Lander and Botstein, 1987).

After narrowing the critical interval to the extent possible, identification of the causative gene and mutation requires identification of genetic variants. This can be done using mRNA if the gene is expressed in available tissue, but generally, mutation analysis is performed with genomic DNA; this has the added advantage that splice site mutations in introns can sometimes be seen.

Many techniques for mutation detection have been employed, including single-strand conformation polymorphism (SSCP) (Orita et al., 1989) and denaturing HPLC (O'Donovan et al., 1998). However, decreased cost and increased throughput for sequencing make this the best approach in many cases. Typically, mutation screening is first performed on exons and flanking intronic sequence of coding exons and then on noncoding exons. This technique is not sensitive to gene duplications or deletions and this caveat must be considered if mutations cannot be found. Also, mutations in introns and intragenic regions may contribute to differences in expression levels and would not be detected by sequencing of exons.

Proof that a gene is causative of a Mendelian trait requires that the variant is not found in control samples and that it segregates with the phenotype in the pedigree(s). Causation is further supported (although not proven) if the residue is highly conserved and if the variant results in a dramatic change (dramatic charge or size change, truncation of protein, etc.). The strongest genetic evidence of causation is (1) identification of multiple independent mutations in the same gene in different families with the same phenotype and/or (2) occurrence of a de novo mutation in a sporadic case of a rare phenotype. These cases represent convergences of multiple rare events, thus decreasing the likelihood that they occurred by chance.

Ultimately, demonstration of functional consequences of the genetic variant in vitro or recapitulation of the phenotype in vivo (generation of animal models) provides very strong evidence for a causal relationship. However, differences between human and rodent physiology lead to some human disease gene mutations that do not cause phenotypes in mice.

Approaches to Complex Genetics

In contrast to genetic approaches for identifying causative genes in Mendelian disorders, characterization of complex genetic contributions to human behavior or complex diseases is not searching for the single
causative mutation. Rather, genetic variations are sought that contribute to risk of a disorder, such as hypertension or Alzheimer's, for example. With regard to traits that are not diseases themselves, genetic variation contributes to normal variations in the population. This is the case with traits such as sleep/wake time preferences, for example. Variations at multiple loci summate, and the resulting phenotype is the result of contributions from multiple loci.

In the field of complex genetics, investigators typically pursue either case control or family-based tests for linkage. In family-based studies, data are generated within a family with unaffected individuals representing controls for their affected relatives. In this case, some of the potentially confounding factors are better controlled, as family members generally share more similar environmental factors, as well as sharing a more significant percentage of their DNA (depending on the relationship between such individuals). Transmission disequilibrium tests have been used to this end and examine the frequencies of alleles transmitted to affected individuals vs alleles that are not transmitted (Spielman et al., 1993).

Case control studies investigate the association between a trait and genotypes using data generated from a cohort of unrelated patients and a population of unrelated control individuals. For example, if one allele in a candidate gene occurs at a higher rate in affected subjects versus controls, this would raise the possibility that this variant itself (or another variant in linkage disequilibrium) might contribute to that phenotypic trait. The likelihood of this association being significant can be calculated statistically. The assumption made here is that the matching between patients and controls is appropriate and that these groups do not differ in marker frequencies for reasons other than the locus being correlated with the trait. This is a relatively new area and the specific tools for statistically analyzing such data are still evolving. There are multiple approaches to searching for associations between genetic variations and quantitative traits. No one of the tools currently available is clearly the best approach for dealing with data in such experiments.

Challenges of Behavioral Genetics

The single largest limitation to identification of molecular determinants of behavior is the robustness of the observed phenotype. This is simpler in model organisms where it is possible, for example, to measure activity rhythms under conditions of continuous darkness. Given the broad spectrum of variation in behaviors that are considered "normal," it has been very difficult to find strong behavioral phenotypes segregating as highly penetrant Mendelian traits.

Even within a large family segregating an autosomal-dominant trait for FASPS, phenocopies (other causes of the same phenotype) can be present. The ASPS of aging, for example, is separate from FASPS but could lead one to make the diagnosis of FASPS in an aged individual who is not a carrier for an FASPS gene variant. For this reason, it is critical that the early morning awakening of FASPS must be present before the age of 40 (Jones et al., 1999). There are individuals in the population who do not carry a single gene causing early bird phenotype who, because of the constellation of genes in their genome, are in the tail of distribution of sleep time preferences (Fig. 1). This turned out to be the case in a small branch of kindred 2174, the family in which FASPS was first described.

FIG. 4. FASPS kindred 2174. Circles represent women and squares represent men. Filled symbols show individuals with the FASPS phenotype. Empty symbols are those considered unaffected by FASPS, and a cross-hatched symbol marks an individual who cannot be classified as definitely affected or definitely unaffected. Horne-Ostberg scores, where available, are shown underneath each individual. The mean Horne-Ostberg scores were 75.1 (n = 18, SD = 6.45) for individuals meeting FASPS criteria, 61.9 (n = 15, SD = 9.9) for those meeting non-FASPS criteria, and 64.4 (n = 22, SD = 9.1) for individuals whose phenotype was indeterminate. One branch includes three subjects considered affected in the initial genetic studies (Horne-Ostberg scores of 67, 68, and 61). However, these individuals turned out not to carry the allele linked to FASPS nor the genetic variant in the hPer2 gene. Note that a woman marrying into this branch of the family has a high Horne-Ostberg score (72) and presumably carries multiple genetic factors, leading her to manifest an early bird phenotype.
Jones et al., 1999) and in whom the molecular basis of FASPS was first established (Toh et al., 2001) (Fig. 4).

Thus, there are many challenges to characterizing human behavioral phenotypes. These pose significant challenges for the molecular characterization of human behavior. Still, as shown in FASPS, it is now possible to identify the genetic basis of human behavior and sharpen our focus on the critical issues of precise phenotyping required for such accomplishments.

Acknowledgments

This work was supported by Grant HL59596 (from NIH). Dr. Ptacek is an Investigator of the Howard Hughes Medical Institute.

References


