On the genetics of sleep disorders: genome-wide association studies and beyond

Michael J Parsons
Mammalian Genetics Unit, MRC Harwell, Harwell, Oxfordshire, UK

Abstract: Sleep is an essential behavior, yet much of its underlying functions are still unknown. The disruption of sleep can lead to a variety of health consequences. Family and twin studies have together shown that genetic factors underlie variation in sleep characteristics and sleep disorders. Given the importance of sleep to our well-being, understanding its underlying genetic factors is essential to both the prevention and treatment of these disorders. Recently, genome-wide association studies (GWAS) have helped to provide evidence of associations of both known and novel genetic variants with sleep disorders. This review outlines the findings from GWAS for a number of sleep disorders, including insomnia, restless leg syndrome, obstructive sleep apnea, and narcolepsy, and discusses these findings in the context of supporting evidence from independent methodologies. Finally, the limitations of GWAS approaches are outlined, along with the future directions of the genetics of sleep in the post-GWAS era.

Keywords: genome-wide association studies, sleep, sleep disorder, insomnia, restless leg syndrome, obstructive sleep apnea, narcolepsy

Introduction
Sleep is a fundamental state in which we spend approximately one-third of our lives. Our dedication to this activity likely reflects its profound significance in many areas of our functioning and well-being, and the consequences of its disruption can be profound. Sleep disorders are common, and range from: disturbances of sleep duration and quality, such as insomnia; sleep-related movement disorders, such as restless leg syndrome (RLS); breathing difficulties during sleep, such as obstructive sleep apnea (OSA); and the dissociation between sleeping and waking states, such as narcolepsy. Given the importance of sleep to our well-being, understanding the genetic and environmental factors that underlie individual differences with regard to sleep and sleep disorders is essential to put us in a stronger position to prevent and treat these disorders.

Traditional “family studies” investigate whether a particular trait runs in families, thus providing a starting point to determine the extent of genetic influences underlying specific sleep phenotypes and sleep disorders. Family studies have highlighted a genetic component in a number of sleep disorders, including insomnia,1–3 RLS,4 OSA,5 and narcolepsy.6 Twin studies have also detected a genetic component for a range of sleep phenotypes, including sleep duration, insomnia, RLS, OSA, and narcolepsy.7

A variety of approaches have been used to identify genes underlying the heritability of sleep and sleep disorders, including the use of animal models and candidate gene association studies in human cohorts. Candidate gene association studies for sleep and sleep disorders using human genetic cohorts have identified a number of...
potential loci; however, they are limited because they focus on known or suspected candidate genes for a disorder.8 Animal models have implicated various molecular pathways involved in sleep, including neurotransmitter systems (adenosine, histamine, dopamine, acetylcholine, norepinephrine, gamma-aminobutyric acid, and serotonin), ion channels, genes associated with immune function, intracellular messengers, and genes involved in regulating circadian rhythms.5,9 Traditionally, mammalian models (eg, mice, rats, and dogs) were used, although recently simpler model systems have been adopted (eg, drosophila, Caenorhabditis elegans, and zebra fish). Choice of the model used is often balanced between a number of factors, including the ease of fine genetic manipulation, similarity in sleep architecture to humans, and the resources required for maintaining these models.

More recently, researchers are turning to unbiased approaches to find genetic variants in novel genes implicated in these disorders, such as genome-wide association studies (GWAS).10 By comparing the genome-wide genetic variation across individuals, usually using case and control groups, GWAS provide powerful evidence to support the role of a genetic variant in a disorder. GWAS range in both sample size and number of single nucleotide polymorphisms (SNPs) investigated, with the trend being for larger sample sizes, (thousands if not tens of thousands) and a greater coverage of SNPs (typically 300–500,000’ SNPs). GWAS have provided evidence for the involvement of numerous, often novel, genes underlying a range of sleep disorders. The purpose of this review is to summarize these findings, discuss their implications for sleep research, and outline the future directions of the genetics of sleep in the post-GWAS era.

GWAS for sleep duration
Sleep duration varies widely among individuals and has been associated with numerous health outcomes. Twin studies have shown that self-reported sleep duration has a heritability ranging from 31% to 37%.11,12 Candidate gene association studies for sleep duration have nominated a number of potential loci, although many have not been replicated. Strong/substantial evidence has been found for an association of sleep duration with polymorphisms in the CLOCK gene13-15 and in a functional variant of the circadian gene DEC2.16,17

In the first GWAS that focused on sleep duration and usual bedtime using a self-reported questionnaire, the authors genotyped over 70,000 SNPs in approximately 750 individuals and found five suggestive peaks (logarithm of odds (>2); four associated with usual bedtime and one with sleep duration.18 These linkage peaks contained a number of genes, including the circadian genes Casein kinase II subunit alpha 2 (CSNK2A2), a known component of the core circadian clock, and Prokineticin2 (PROK2), a neuropeptide that is important in circadian signaling.19 The most significant association with sleep duration was a SNP within the intron of cyclic adenosine monophosphate-specific 3′,5′-cyclic phosphodiesterase 4D (PDE4D) (rs1823068, \( P = 2.5e^{-8} \)), a gene that is expressed throughout the brain (see Table 1 for a summary of all genome-wide significant loci). A nonsynonymous SNP in neuropeptide S receptor 1 (NPSR1; rs324981, \( P = 1.8e^{-5} \)) was associated with a delayed bedtime of 15 minutes per Ile107 allele.

A second GWAS investigating sleep duration using the Munich Chronotype Questionnaire in a large European sample (n=4,251).20 This study found an association between an intronic variant (rs11046205, \( P = 4.0e^{-8} \)) in the adenosine triphosphate-binding cassette, sub-family C member 9 (ABCC9) gene (an adenosine triphosphate-sensitive potassium channel) and sleep duration. This SNP explained 5% of the variation in sleep duration, such that having at least one G allele was associated with a decrease in sleep duration of approximately 30 minutes. Although not replicated in the overall sample of an independent cohort, there was a significant association in a subsample limited to morning chronotypes collected during the winter (both chronotype and seasonality independently affect sleep duration). There was significant replication in an independent sample of rs11046205 and sleep duration (\( P < 0.05 \)), but sleep duration was only lower in those individuals with two copies of the G allele.21 Neuronal-specific knockdown of the drosophila homolog (Sur2) resulted in an almost complete loss of sleep during the first 3 hours of the night.20

Another GWAS investigating sleep duration in a Finnish population (n=1,941) failed to find any genome-wide signals, although it did find 31 SNPs with a lower statistical cut-off (\( P < 5e^{-5} \)). Three of these SNPs were nominally associated with sleep duration in a replication sample (n=6,834): rs10914351 near receptor-type tyrosine-protein phosphatase PCP-2 (\( P = 0.049 \)), rs1037079 in protocadherin-7-CENTD1 (\( P = 0.011 \)), and rs2031573 in Krueppel-like factor 6 (KLF6), a transcription factor (\( P = 0.044 \)).22 Interestingly, both the rs2031573 and rs1037079 SNPs were associated with decreased KLF6 mRNA expression in blood (\( P < 0.05 \)), suggesting that these SNPs may have functional consequences. As single-gene analysis misses the effects of a set of genes on a shared biological pathway, many researchers additionally investigate whether there is enrichment of specific gene ontology (GO) terms for a given gene list. Such a pathway
enrichment analysis using nominally significant associations found enrichment of the mitogen-activated protein kinase signaling pathway (including CACNA1C), ABC transporters (including ABCC9), and long-term depression (P<0.05). However, this study failed to replicate the previous associations of ABCC9 (rs11046205) or PDE4D (rs1823068) with sleep duration.

Finally, a recent GWAS investigating sleep duration in a large European sample (n=47,180) found two novel genome-wide significant (GWS) loci, one on chromosome 2, in an intergenic region (rs1191685, P=1.1e-9) upstream of Paired box gene 8 (PAX8), a thyroid specific transcription factor, and another on chromosome 6, in an intergenic region (rs4587207, P=2.0e-8) upstream of radiation-inducible immediate-early gene IEX-1 (IER3), flotillin-1 (FLOT1), and a non-coding RNA (LINC00243). The association of the chromosome 2 loci SNPs with sleep duration were successfully replicated, whereas the chromosome 6 loci SNPs were not. This study failed to find any associations with core circadian genes and failed to replicate the previous association of the ABCC9 gene with sleep duration.

No single genome-wide association for sleep duration has been consistently replicated across these studies. The inconsistencies are likely due to, among other things, the use of a wide range of self-reported questionnaires to access sleep duration. Measures of sleep assessed using questionnaires do not always marry well with measures assessed in other ways, such as actigraphy or electroencephalographic recordings. Furthermore, while one third of self-reported sleep duration is heritable, sleep power measured using electroencephalography has a heritability estimated to be as high as 95%, and may be better suited to detect genetic variation associated with sleep characteristics. Future GWAS for sleep duration would be aided by the use of a more consistent and biologically relevant assessment of sleep duration.

**GWAS for insomnia and sleep quality**

Insomnia is a sleep disorder characterized by the inability to fall asleep or maintain sleep (for more comprehensive criteria, see the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition). It is one of the most common sleep disorders, affecting as many as 10%–15% of the adult population, although the prevalence depends on the definition of insomnia used. Although a large twin study found that genetic heritability for insomnia was estimated at 57%, little is known about the underlying genetic variation, which is likely due to inconsistencies in defining insomnia. Candidate gene studies for insomnia have nominated genetic variants within the circadian gene CLOCK, the GABAergic system, the adenosinergic system, and the serotonergic system. The first GWAS for insomnia investigated just over 81,000 SNPs in a Korean population (n=8,842). The strongest association was found for a SNP within the receptor tyrosine kinase-like orphan receptor 1 (ROR1) gene (rs11208305, P<5.6e-6), which is involved in synaptic formation. Interestingly, GO term analysis of genes nominally associated with insomnia found an enrichment for neuronal function (P<0.05). The next most significant association was for a SNP within 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase beta-1 (PLCBI; rs7187712, P<8.5e-6), one of 16 SNPs in this gene nominally associated with insomnia. PLCBI was one of many genes that were nominally associated with insomnia that are also known to function in learning and memory-related calcium signaling.

A GWAS that investigated self-reported sleep habits in an Australian sample (n=2,323) assessing more than 2,000,000 SNPs failed to find any GWS hits; however, it did find a group of seven SNPs in perfect linkage disequilibrium within the third intron of CACNA1C, a gene that encodes the 1C subunit of a 1-type voltage-gated calcium channel, that were strongly associated with sleep latency (the most significant SNP being rs7304986, P=1.3e-6) and another CACNA1C variant (rs7304986) associated with sleep quality (P=4.4e-6). Another SNP, downstream of neuropeptide S (NPS; rs10734107), was significantly associated with sleep latency (P=1.1e-5). In a separate analysis, the authors investigated the candidate SNPs from a previous study, and only one SNP in EPB41 (rs2985334) survived multiple testing correction (P=0.0062, β=5.3 minutes). Although the CACNA1C association with sleep latency was not replicated in a Chronogen Consortium sample, another replication study found a CACNA1C variant (rs16929277, P<0.05) that was significantly associated with both sleep latency and sleep quality.

Byrne et al also performed a GWAS specifically investigating caffeine-related insomnia alone and after controlling for a general insomnia score in the same Australian sample (n=2,402). This study failed to find GWS SNPs, and many of the most significant associations were in intergenic regions for both analyses. The SNP most strongly associated with caffeine-related insomnia without controlling for general insomnia was an intronic SNP in Proline-rich membrane anchor 1 (PRIMA1) (rs6573232, P=1.4e-6), a gene important for anchoring acetylcholinesterase to neuronal membranes. Although a number of the other most significant associations...
were SNPs in intergenic regions (eg, rs7628219, \( P = 1.7e-6 \); rs2065779, \( P = 2.5e-6 \); and rs11706236, \( P = 3.6e-6 \)), the authors also found associations with a set of intronic SNPs in microtubule-associated tumor suppressor Candidate 2 (\( MTUS2 \)); including rs2388082, \( P = 1.4e-6 \) and a SNP downstream of a melatonin receptor 1B (\( MTNR1B \); rs10830964, \( P = 5.2e-6 \)). When general insomnia was controlled for, two of the most significant associations were in or near the guanylate binding protein 4 (\( GBP4 \)) gene (rs521704, \( P = 1.9e-6 \); rs561042, \( P = 6.2e-6 \)), which hydrolyzes guanosine triphosphate, a nucleic acid that is structurally similar to adenosine. This study successfully replicated a previous association with an adenosine receptor SNP.\(^{32}\) A GO enrichment analysis also found a significant overrepresentation of genes associated with nucleic acid metabolism (\( P = 1e-4 \)), suggesting that genes directly or indirectly affecting adenosine signaling and metabolism may be involved in caffeine-induced insomnia.

While these GWAS investigating insomnia/sleep quality failed to identify any variants that were GWS, the results highlighted pathways that may underlie the genetic etiology of insomnia, including genes involved in calcium signaling, neuronal function and adenosine metabolism. The use of a consistent and clinically validated definition of insomnia may aid the success of future GWAS for insomnia.

**GWAS for RLS**

RLS, one of the most common neurological disorders, is characterized by an irresistible urge to move one’s body, usually the legs, and is accompanied by an uncomfortable and unpleasant sensation.\(^{40}\) Iron deficiency and disruption of the dopamine system are two of the most common medical conditions associated with RLS,\(^{41}\) with an age-independent prevalence of approximately 10%.\(^{42}\) Twin studies have found that the symptoms of RLS are highly heritable, ranging from 44% to 63% (95% confidence interval).\(^{43-45}\) Linkage analysis studies in families have revealed seven unique genetic loci for RLS: \( RLS1 \), chromosome 12q12–q21; \( RLS2 \), chromosome 14q13–21; \( RLS3 \), chromosome 9p24–p22; \( RLS4 \), chromosome 2q33; \( RLS5 \), chromosome 20p13; \( RLS6 \), chromosome 19p13; and \( RLS7 \), chromosome 16p12.1.\(^{46-52}\)

To better characterize the genetic variation underlying RLS, Winkelmann et al conducted a GWAS for RLS.\(^{53}\) They found three loci significantly associated with RLS, including intronic variants in Myeloid ecotropic viral integration site 1 homeobox 1 (\( MEIS1 \); rs2300478, 8.1e-23) (chromosome 2p14), a homeobox gene; variants in BTB domain containing 9 (\( BTBD9 \); rs9262409, 9.4e-13; rs9357271, 1.5e-12), a gene recently implicated in synaptic plasticity;\(^{54}\) and seven variants (the most significant being rs1026732, 1.4e-11) over a region including both Mitogen-activated protein kinase kinase 5 (\( MAP2K5 \)), a protein kinase, and SKI family transcriptional corepressor 1 (\( SKOR1 \)), a transcription factor. Together these variants accounted for over 50% of the risk for RLS. Another GWAS for RLS revealed two additional GWS variants (rs4626664, \( P = 5.4e-10 \); rs1975197, \( P = 5.8e-9 \)) that lie within the 5’-untranslated region of the protein tyrosine phosphatase receptor type D (\( PTPRD \)) gene,\(^{55}\) shown to be important for correct targeting of the motor neuron axon during mammalian axonogenesis.\(^{56}\) Interestingly, a different variant within \( PTPRD \) was significantly associated with periodic leg movements of sleep, a symptom of RLS, in an independent GWAS.\(^{57}\) A larger GWAS for RLS (approximately 5,000 cases) replicated these four peaks and found an additional two novel GWS loci, one (rs6747972, \( P = 9.0e-11 \)) in an intergenic region 1.3 Mb downstream of \( MEIS1 \) and the other (rs3104767, \( P = 9.4e-19 \)) in an area of linkage disequilibrium containing TOX high mobility group box family member 3 (\( TOX3 \)) and the non-coding RNA \( BC034767.38 \) This study found that these six loci together explain 6.8% of the known heritability of RLS, suggesting that additional loci were missed using this approach.

A number of replication studies have been conducted to confirm the associations of these loci with RLS, particularly focusing on those discovered initially (\( BTBD9 \), \( MEIS1 \), \( MAP2K5 \)/\( SKOR1 \)). Two case-control studies, one using a US sample and another using three European populations, successfully replicated the association of all three of these loci with RLS.\(^{59,60}\) When limited to a subset of sporadic cases, thus limiting power, the second study only replicated the \( BTBD9 \) association.\(^{61}\) Another case-control study in an American sample successfully replicated the association with both \( BTBD9 \) and \( MEIS1 \), but only found a trend toward significance for the variants in \( MAP2K5 \)/\( SKOR1 \).\(^{61}\) A recent study investigating \( BTBD9 \), \( MEIS1 \), \( MAP2K5 \)/\( SKOR1 \), and \( PTPRD \) in a Korean case-control sample only successfully replicated the association with \( BTBD9 \).\(^{62}\) Another study only investigating the two nominated \( PTPRD \) SNPs in both RLS families and in a case-control sample, replicated the association between RLS and rs1975197 (\( P = 0.02 \)), but not rs4626664 (\( P = 0.6 \)).\(^{63}\) A study using a small Czech sample did not replicate any of the previous associations with RLS, but this may be due to power issues.\(^{64}\) There is thus the strongest evidence for an association between \( BTBD9 \) and RLS, followed by \( MEIS1 \) and \( MAP2K5 \)/\( SKOR1 \).

It has recently been shown that SNPs found to have GWAS associations with complex diseases are more likely
Table 1: Summary of genome-wide significant associations

<table>
<thead>
<tr>
<th>Genetic variant</th>
<th>Location</th>
<th>Gene</th>
<th>P-value</th>
<th>Pathology</th>
<th>Reference</th>
<th>Replication</th>
<th>Non-replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1823068</td>
<td>Intron</td>
<td>cAMP-specific 3',5'-cyclic phosphodiesterase 4D</td>
<td>2.5e-08</td>
<td>Sleep duration</td>
<td>19</td>
<td>NA</td>
<td>20, 22, 23</td>
</tr>
<tr>
<td>rs11046205</td>
<td>Intron</td>
<td>ATP-binding cassette, sub-family C member 9</td>
<td>4.0e-08</td>
<td>Sleep duration</td>
<td>20</td>
<td>21</td>
<td>22, 23</td>
</tr>
<tr>
<td>rs1191685</td>
<td>Intergenic</td>
<td>Paired box gene 8</td>
<td>1.1e-09</td>
<td>Sleep duration</td>
<td>23</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs4587207</td>
<td>Intergenic</td>
<td>Radiation-inducible immediate-early gene IEX-1, flotillin-1, non-coding RNA (LINCO0243)</td>
<td>2.0e-08</td>
<td>Sleep duration</td>
<td>23</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs2300478</td>
<td>Intron</td>
<td>Myeloid ecotropic viral integration site I homeobox 1</td>
<td>8.1e-23</td>
<td>RLS</td>
<td>53</td>
<td>55, 58–61</td>
<td>62, 64</td>
</tr>
<tr>
<td>rs9296249</td>
<td>Intron</td>
<td>BTB domain containing 9</td>
<td>9.4e-13</td>
<td>RLS</td>
<td>53</td>
<td>55, 58–62</td>
<td>64</td>
</tr>
<tr>
<td>rs1026732</td>
<td>Intron</td>
<td>Mitogen-activated protein kinase 5, SKI family transcriptional corepressor 1</td>
<td>1.4e-11</td>
<td>RLS</td>
<td>53</td>
<td>55, 58–60</td>
<td>61, 62, 64</td>
</tr>
<tr>
<td>rs4626664</td>
<td>5'-UTR</td>
<td>Protein tyrosine phosphatase receptor type D</td>
<td>5.9e-10</td>
<td>RLS</td>
<td>55</td>
<td>57, 58, 63</td>
<td>NA</td>
</tr>
<tr>
<td>rs6747972</td>
<td>Intergenic</td>
<td>Myeloid ecotropic viral integration site I homeobox 1</td>
<td>9.0e-11</td>
<td>RLS</td>
<td>58</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs3104767</td>
<td>Intergenic</td>
<td>TOX high mobility group box family member 3, non-coding RNA (BC034767)</td>
<td>9.0e-11</td>
<td>RLS</td>
<td>58</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>DQB1*06:02</td>
<td>Coding</td>
<td>Major histocompatibility complex, class II, DQ beta 1</td>
<td>4.7e-47</td>
<td>Narcolepsy</td>
<td>98</td>
<td>87, 107, 108</td>
<td>NA</td>
</tr>
<tr>
<td>rs5770917</td>
<td>Intergenic</td>
<td>Carnitine palmitoyltransferase 1B, choline kinase β</td>
<td>4.4e-07</td>
<td>Narcolepsy</td>
<td>98</td>
<td>102</td>
<td>106</td>
</tr>
<tr>
<td>rs16826005</td>
<td>Intron</td>
<td>NCK-associated protein 5</td>
<td>1.0e-07</td>
<td>Narcolepsy</td>
<td>102</td>
<td>NA</td>
<td>108</td>
</tr>
<tr>
<td>rs1154551</td>
<td>Intron</td>
<td>T-cell receptor alpha</td>
<td>2.8e-22</td>
<td>Narcolepsy</td>
<td>103</td>
<td>87, 106–108</td>
<td>105</td>
</tr>
<tr>
<td>rs2305795</td>
<td>3'-UTR</td>
<td>Purinergic receptor subtype 2Y_5</td>
<td>6.1e-10</td>
<td>Narcolepsy</td>
<td>105</td>
<td>106, 108</td>
<td>87</td>
</tr>
<tr>
<td>rs34593439</td>
<td>Intron</td>
<td>Cathepsin H</td>
<td>1.8e-08</td>
<td>Narcolepsy</td>
<td>107</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs7537111</td>
<td>Intergenic</td>
<td>Tumor necrosis factor superfamilly member 4</td>
<td>4.1e-08</td>
<td>Narcolepsy</td>
<td>107</td>
<td>NA</td>
<td>87</td>
</tr>
<tr>
<td>rs9648789</td>
<td>Intergenic</td>
<td>T-cell receptor γ locus</td>
<td>3.7e-09</td>
<td>Narcolepsy</td>
<td>108</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs10995245</td>
<td>Intron</td>
<td>Zinc finger gene 365</td>
<td>1.2e-11</td>
<td>Narcolepsy</td>
<td>108</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>rs2252931</td>
<td>Intergenic</td>
<td>Interleukin-10 receptor β, interferon (α, β, γ) receptor 1</td>
<td>2.2e-09</td>
<td>Narcolepsy</td>
<td>108</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Notes: All genetic variants are designated by the dbSNP ID, minus DQ01*06:02, which is designated by the HLA ID. If more than one SNP was identified that reached RWS for a study, only the most significant SNP in the study is represented; *DQ01*06:02 is thought to mediate the majority of the effects of the HLA locus, alone or as a haplotype, but other variants also play a role (see references 6, 84–86); †the SNPs rs5770917 and rs16826005 are included in the graph as they reached GWS within the individual studies (P<6.4e-5 and 5e-7, respectively), but it is important to note that they do not survive the P<5e-8 now more commonly accepted as a statistical cut-off for GWS. *Abbreviations: RLS, restless leg syndrome; SNP, single nucleotide; NA, not available; UTR, untranslated region; GWS, genome-wide significance.

to affect gene expression, and in turn represent expression quantitative trait loci (eQTLs). The presence of eQTLs has been used successfully to prioritize subthreshold GWAS loci for a range of complex diseases, such as Crohn’s disease. Using a similar approach in two independent RLS cohorts, researchers used mRNA from peripheral blood to prioritize the 332 most significantly associated SNPs from the largest RLS GWAS. Of the 23 cis-eQTLs that overlapped between the two cohorts, eleven were located in known RLS-associated loci. None of the 12 novel eQTL SNPs were significantly associated with RLS, but the transcription factor, de-etiolated homolog 1 (DETI), a gene involved in ubiquitination and subsequent degradation of target proteins, was nominally associated with RLS (P=0.04).

It is often difficult to determine if a GWS SNP is in fact the causal variant underlying the association or instead is in linkage disequilibrium with the true causal variant/s. Recently, it has been proposed that rare, functional variants may underlie these associations as well as explaining some of the missing heritability in GWAS. Sequencing of the coding regions of BTBD9 and MEIS1 revealed a rare novel non-synonymous variant in MEIS1, p.R272H (ss107795884), which segregated, although not significantly, in a small RLS pedigree. Resequencing of the coding regions and exon-intron boundaries...
from the previously described GWAS for RLS revealed 49 novel, rare alleles (minor allele frequency <5%), which together were overrepresented in the RLS cases compared with a control sample. Non-synonymous SNPs within MEIS1 were significantly enriched in the RLS cases, with nominally significant enrichment for the other six investigated genes. The 13 non-synonymous variants found within the canonical MEIS1 isoform (ENST00000272369) were functionally characterized using an assay for neuronal development in zebrafish; four were found to be hypomorphic and six to be null variants. There was a significant excess of these loss-of-function alleles in RLS patients. Though further research is required to determine the functional variants and their associated mechanisms of action, together these studies have implicated genes associated with neuronal development and synaptic plasticity, among other pathways, as the potential basis of the genetic etiology of RLS.

**GWAS for OSA**

OSA is a common disorder characterized by repetitive episodes of upper airway obstruction or airway collapse that occurs during sleep, which leads to oxygen desaturation, disrupted/fragmented sleep, and daytime sleepiness. Many factors may contribute to this airway obstruction, including obesity, large necks and nasal obstruction caused by a deviated septum. Genetic factors are known to contribute to OSA, with heritability estimates for the apnea hypopnea index (AHI), a measure of OSA, of 32% in one study. Several genetic variants have been identified in candidate gene studies of OSA, yet few have been successfully replicated. Three genes with the strongest evidence of involvement in OPA include apolipoprotein E (APOE), angiotensin-converting enzyme (ACE), and tumor necrosis factor alpha (TNF\(\alpha\)), but recent meta-analyses have only validated the association with TNF\(\alpha\).

Given the close association of OPA with obesity, commonly measured using body mass index, it is necessary to control for obesity in any association models for OPA to determine if the association is truly unique to OPA. Genome-wide univariate family linkage studies for OPA have revealed only weak candidate loci for AHI, independent of body mass index, while bivariate analysis revealed a linkage peak for respiratory disturbance index and AHI at 19p13.4 (logarithm of odds, 3.04), which includes a killer cell immunoglobulin-like receptor gene cluster.

A large candidate gene association study for OSA using the Cleveland Family Study, consisting of European-American and African-Americans subsamples, investigated over 45,000 SNPs from approximately 2,100 candidate genes. Within the African American subset, two SNPs survived multiple testing corrections: rs11126184 \((P=1.4e-6)\), a SNP in the pleckstrin (PLEK) gene was associated with OSA, and rs7030789 \((P=4.6e-6)\), a SNP in the lysophosphatidic acid receptor 1 (LPAR1) gene was associated with AHI. PLEK has a role in protein kinase C signaling and LPAR1 is a G protein-coupled receptor. Within the European subsample, rs1409986 \((P=2.2e-4)\), a SNP in the prostaglandin E2 receptor (PTGER3) gene, was significantly associated with OSA. All three of these SNPs remained significant in analysis models including body mass index. LPAR1 and PTGER3 associations, but not PLEK, were replicated in an independent sample. A significant enrichment in phosphorylation of the STAT protein family and insulin signaling GO functional categories was found, suggesting that these pathways might warrant further investigation for genetic etiology of OPA.

As GWAS often miss rare, causal variants that might underlie the heritability for a complex disorder, researchers are increasingly turning to other methods, such as exome sequencing and selected resequencing of GWAS loci to detect these variants. A recent study used exome sequencing in four probands that presented with intellectual disability, developmental delay, and mildly dysmorphic facial features, of which three probands also had OSA. This study revealed four truncating mutations in the AT-hook, DNA-binding motif containing 1 (AHDC1) gene, an uncharacterized human transcription factor, to be associated with these traits.

As a range of factors influence OSA, genes underlying OSA can affect one or more of these factors, making it critical to carefully consider which factors to include in any GWAS approach for OSA. This consideration may help to reveal additional loci for the disorder that may be specific or shared across these associated factors.

**GWAS for narcolepsy**

Narcolepsy is a sleep disorder that affects approximately 1 in 2,000 individuals. It is characterized by irresistible excessive daytime sleepiness as well as cataplexy, the sudden loss of muscle tone in response to strong emotions. Family studies have shown up to a 40 times higher relative risk of narcolepsy in first-degree family members of patients and twin studies estimate approximately 35% heritability; together this suggests that the etiology of narcolepsy has a genetic component.

Of all the sleep disorders, narcolepsy is the one for which there has been particular progress in terms of specifying
candidate genes. It has been reported that a very high proportion (>85%) of those suffering narcolepsy with cataplexy carry a specific human leukocyte antigen (HLA) allele (HLA-DQβ1*) compared with just 12%–38% in the general population. Additional studies show that a number of HLA class II region alleles play a complex role in the genetic predisposition to narcolepsy. A recent study found that the DQBI*06:02 allele confers an extremely high risk for narcolepsy with cataplexy (odds ratio 251). Homozygotes for this allele had higher levels of HLA-DQ mRNA and protein expression in white blood cells, suggesting a potential functional consequence of this allele. Due to its association with HLA alleles, it is thought that narcolepsy could be an autoimmune disease. This hypothesis has been strengthened by the findings of auto-antibodies close to the time of disease onset and the increase in the number of narcolepsy cases following the 2009 H1N1 vaccination.

Two independent animal models have implicated the hypocretin (orexin) system in narcolepsy: a knock-out mouse missing the hypocretin gene displayed narcolepsy-like symptoms and disruption of a hypocretin receptor caused narcolepsy in a canine model. It is now believed that the majority of human cases of narcolepsy are due to the loss of hypocretin-producing neurons in the hypothalamus. The loss of these neurons is likely caused by a combination of environmental factors and a genetic susceptibility based, at least in part, on the HLA locus, which might cause these cells to be the target of autoimmune attack.

GWAS in narcolepsy have generally served to strengthen the associations with HLA as well as nominate potential novel pathways. One of the first GWAS for narcolepsy replicated the association with the HLA locus and also found GWS association with a SNP between the carnitine palmitoyltransferase 1B (CPT1B) and choline kinase β genes (CHKB; rs5770917, P = 4.4e-7) that decreases the expression of these two genes. CPT1B is an enzyme involved in β-oxidation of long-chain fatty acids, a pathway involved in regulating theta frequency during rapid eye movement sleep. CHKB is an enzyme involved in the metabolism of the acetylcholine precursor choline. Both rs5770917 (P = 3.6e-3) and variants in the HLA locus (P = 9.2e-11) were also significantly associated with essential hypersomnia, a group of disorders related to narcolepsy. Consistent with these findings, rs5770917 was associated with decreased serum levels of CPT1B mRNA expression and oral administration of L-carnitine decreased daytime sleep times in narcoleptic patients in a double-blind study. In addition to CPT1B, a GWAS for HLA-DQβ1*06:02 in negative essential hypersomnia found another novel significant association with NCK-associated protein 5, a protein coding gene with no known function (rs16826005, P = 1e-7).

GWAS have also strengthened the autoimmune hypothesis for narcolepsy, with a number of significant associations found in genes important for regulating immune function. One GWAS for narcolepsy with individuals that were DQBI*0602-positive found significant associations with variants in the T-cell receptor alpha (TRCα) locus (rs1154155, P = 1e-21), a receptor responsible for responding to HLA-peptide presentation. This variant was also significantly associated with essential hypersomnia specifically in individuals with the DRB1*1501-DQBI*0602 haplotype (P = 0.04). Another GWAS for narcolepsy found a GWS association with a SNP in the 3′-untranslated region of purinergic receptor subtype 2Y1 (P2RY11) gene (rs2305795, P = 6.1e-10), an important regulator of immune cell survival, that regulates its own expression in T-lymphocytes and natural killer cells. The associations of the TRCα (rs1154155) and P2RY11 (rs2305795) variants with narcolepsy were replicated in a Chinese sample, while the association with the CPT1B/CHKβ variant (rs5770917) was not. A custom genotyping array for the fine mapping of DNA loci involved in immune function (ImmunoChip) was used to genotype a sample consisting of hypocretin-deficient narcoleptic patients (n = 1,886) and replicated the GWS associations with the HLA loci and the TRCα locus (rs1154155, P < 1e-30). It also identified two novel loci: cathepsin H (CTSH; rs34593439, P = 1.8e-8), a gene involved in degradation of lysosomal proteins, and tumor necrosis factor (ligand) superfamily member 4 (TNFSF4), a cytokine involved in interactions with T-cells (rs7553711, P = 4.1e-8). A GWAS in a Chinese narcolepsy sample also replicated the association with the HLA locus and the TRCα locus (rs1154155, P < 1e-30) and found an association with a wider set of SNPs at the P2RY11A locus (including rs1155170, P = 3.8e-10). This study additionally found three novel loci: the T-cell receptor β locus, TCRβ (rs9648789, P = 3.7e-9); the transcription factor zinc finger gene 365, ZNF365 (rs10995245, P = 1.2e-11); and a SNP between two genes associated with immune function, IL10Rβ-IFNAR1 (rs2252931, P = 2.2e-9). A recent large GWAS for narcolepsy replicated an association with the HLA locus (rs2858884, P = 2e-9) and the TRCα locus (rs1154155, P = 2e-8). Finally, a recent study investigating copy number variants in narcolepsy patients found an association of copy number variants within Parkinson protein 2 with narcolepsy, as well as enrichment of copy number variants in genes involved in immune function. Together, these studies...
highlight the involvement of immune function in the etiology of narcolepsy.

One additional GWAS investigated associations with individual narcolepsy symptoms.10 Although it failed to find any GWAS associations, it did find strong associations of a SNP in **UBXN2B** (rs2859998, \(P=1.3\times10^{-7}\)), a gene involved in Golgi biogenesis, with age of onset of excessive daytime sleepiness and a SNP in **TEAD4** (rs12425451, \(P=2.0\times10^{-7}\)), a transcription factor, with age of cataplexy onset.

**Future directions**

GWAS have had some success in identification of the genetic variation underlying sleep characteristics and sleep disorders; however, additional research is required to validate, refine, and further this approach. GWAS results provide powerful evidence to support the role of a genetic variant in a disorder, but do have limitations. In particular: they often fail to account for all of the heritability associated with a disorder; they rarely find the causal variant, thus requiring functional studies to validate any findings; and due to stringent multiple testing corrections, the statistical cut-offs for GWAS are so high that they can lead to high rates of false negatives.10 In order to fully identify and characterize the genetic factors underlying sleep and sleep disorders, complementary methods are needed to address these shortcomings, as outlined below.

**Missing heritability**

For the majority of complex disorders, the genes specified by GWAS account for a proportion of the underlying genetic heritability. This “missing heritability” may be in part due to such factors as an insufficient power to detect variation with a low effect size on the disorder, epistasis, failure to detect structural variation, such as the role of copy number variations in influencing a trait, and rare genetic variants.111 With the cost of next-generation sequencing rapidly decreasing, a number of techniques, including targeted and exome sequencing, should help to reveal variation that may be missed using a GWAS approach.

**Determining the functional variants**

GWAS rarely reveal the causal genetic variation which is crucial to understanding the biological mechanism underlying these associations. To characterize the functional role that GWAS genes play in the disease mechanism we must turn to other methods, such as targeted resequencing of a GWAS locus followed by validation using cellular and animal models. Resequencing of GWAS loci have revealed rare variants that can be associated with a particular disorder, individually or collectively.70,71 Furthermore, animal models have successfully validated various GWAS findings.20 With the development of genome-wide or next-generation sequencing approaches to determine the genetic variants underlying various sleep-related phenotypes, animal models will likely play a greater role in validating these findings.

**Mining GWAS data**

Ever increasing sample sizes in GWAS have revealed many additional low effect size variants that previously evaded detection, although more could be done to validate the strong but non-GWS significant associations with disease. A number of approaches have been used to further mine subthreshold GWAS loci. GWS SNPs are more likely to affect gene expression67 and lie within DNase I-hypersensitive sites.112 The presence of cis eQTLs has been used successfully to prioritize subthreshold GWAS loci for a range of complex diseases.67,68 The use of data from public databases, such as the information on regulatory elements available in the Encyclopedia of DNA Elements, has led to the identification of novel GWAS loci.113 Furthermore, as GWAS primarily focus on single-gene analysis, they can miss effects of a group of significant variants on a shared pathway. To address this, researchers commonly try to detect enrichment of GO terms within a gene set, commonly investigating all nominally significant associations, to identify shared biological processes or molecular functions underlying a trait or pathology.114,115 In the absence of replicated variants across GWAS, using a gene set enrichment analysis can additionally be used across studies to determine the consistently detected pathways.116

**Summary**

Given the importance of sleep to our well-being, understanding the genetic and environmental factors that underlie sleep and sleep disorders is essential to developing better prevention and treatment for these disorders. GWAS have successfully identified a number of genetic variants associated with sleep characteristics and sleep disorders, but further work need to be done to fully identify and characterize the genetic factors underlying sleep and sleep disorders. Existing GWAS loci may be validated and additional loci discovered through the refinement of phenotypes investigated, strengthening the power available, and by further mining of the subthreshold GWAS significant loci using additional datasets. The use of next-generation sequencing to determine the functional
genetic variants underlying various sleep-related phenotypes, combined with the use of animal models, will be critical in determining the biological mechanism underlying these associations.

Acknowledgments
The author thanks Drs Alison Landman and Alice Gregory for critically reading the manuscript. MJP was supported by the Medical Research Council.

Disclosure
The author reports no conflicts of interest in this work.

References


