The Mouse *Clock* Locus: Sequence and Comparative Analysis of 204 Kb from Mouse Chromosome 5

Lisa D. Wilsbacher, Ashvin M. Sangoram, Marina P. Antoch, and Joseph S. Takahashi[1]

Howard Hughes Medical Institute, Department of Neurobiology and Physiology, Northwestern University, Evanston, Illinois 60208, USA

The *Clock* gene encodes a basic helix-loop-helix (bHLH)–PAS transcription factor that regulates circadian rhythms in mice. We previously cloned *Clock* in mouse and human using a battery of behavioral and molecular techniques, including shotgun sequencing of two bacterial artificial chromosome (BAC) clones. Here we report the finished sequence of a 204-kb region from mouse chromosome 5. This region contains the complete loci for the *Clock* and *Tpardl* (*pFT27*) genes, as well as the 3′ partial locus of the *Neuromedin U* gene; sequence analysis also suggests the presence of two previously unidentified genes. In addition, we provide a comparative genomic sequence analysis with the syntenic region from human chromosome 4. Finally, a new BAC transgenic line indicates that the genomic region that is sufficient for rescue of the *Clock* mutant phenotype is no greater than 120 kb and tightly flanks the 3′ end of the *Clock* gene.

[The sequence data reported in this paper have been submitted to the GenBank data library under accession no. AF146793.]

The last three years have witnessed remarkable progress in our understanding of the molecular basis of circadian rhythms (Wilsbacher and Takahashi 1998; Dunlap 1999; Strayer and Kay 1999; King and Takahashi 2000). In most if not all living systems, a conserved transcription-translation-based negative autoregulatory feedback loop appears to form the basis for circadian mechanisms. In mammals, the transcription factors CLOCK and BMAL (also known as MOP3) activate expression of the *Period* (*Per*) and *Cryptochrome* (*Cry*) genes; the steady-state mRNA levels of these genes express circadian rhythms of abundance in the suprachiasmatic nucleus (SCN), which is the site of the master circadian pacemaker in mammals (Gekakis et al. 1998; Hogenesch et al. 1998; King and Takahashi 2000; Shearman et al. 2000). Through mechanisms that are still being defined, the PER and CRY proteins translocate to the nucleus and inhibit the activity of CLOCK and BMAL acting upon the promoter regions of the *Per* and *Cry* genes (Gekakis et al. 1998; Sangoram et al. 1998; Griffin et al. 1999; Kume et al. 1999). Finally, turnover of the inhibitory complex leads to the release of CLOCK/BMAL inhibition, and CLOCK/BMAL-driven transcription begins anew.

Because transcriptional control plays such an important role in the circadian model, it will be essential to define the regulatory elements that control circadian gene expression. Genomic sequence of circadian loci will provide extremely useful information for the study of transcriptional regulation. For example, in the *Drosophila period* (*dper*) gene, an E box element (5′-CACGTG-3′) lies within a 69-bp enhancer that is required for high-amplitude cycling of the *per* transcript (Hao et al. 1997). The *dper* E box element and those found in the promoter regions of mouse *Period1* (*mPer1*), mouse *vasopressin*, and *Drosophila timeless* (*dtim*) are known to mediate CLOCK/BMAL-activated transcription of those genes (Darlington et al. 1998; Gekakis et al. 1998; Jin et al. 1999; Ripperger et al. 2000). Genomic sequence also provides an important tool for circadian gene discovery, as shown with the identification of human *Per1* (Sun et al. 1997). Here we report the complete sequence of two overlapping mouse BAC clones from mouse chromosome 5 and describe the structures of the *Clock* locus and four other genes within this 204-kb interval. In addition, we provide a comparative genomic sequence analysis with the syntenic region from human chromosome 4. Finally, we present new BAC transgenic rescue results that reduce the genomic interval that is necessary for proper *Clock* expression.

RESULTS

Shotgun Sequencing and Finishing of Two Bacterial Artificial Chromosome Clones

Twelve bacterial artificial chromosome (BAC) clones were isolated in a library screen for clones spanning the

*Clock* nonrecombinant region (King et al. 1997a). Based on genetic and physical mapping near the *Clock* locus, the 140-kb clone BAC 54 and the 160-kb clone BAC 52 were chosen for sequencing (King et al. 1997a; King et al. 1997b). These two BACs were also selected for transgenic rescue experiments (Antoch et al. 1997), as was BAC 49 (described in this paper). Previous results showed that the full-length BAC 54 clone rescued the phenotypic effects of the *Clock* mutation (Fig. 1; Antoch et al. 1997).

We used the shotgun sequencing approach (Sulston et al. 1992; Ansari-Lari et al. 1998) to obtain genomic sequence near the *Clock* locus. Specifically, we adapted the high-throughput shotgun sequencing protocols available at the Washington University Genome Sequencing Center (http://genome.wustl.edu/gsc/index.shtml) and the Whitehead Institute for Biomedical Research (http://www-genome.wi.mit.edu/) for use in an individual laboratory. BAC 52 and BAC 54 were subcloned into M13 shotgun libraries; approximately 1200 BAC 52 and 1400 BAC 54 M13 subclones were purified and sequenced. This first phase of sequencing generated many small contigs (2 kb–30 kb), most of which were joined via reverse strand sequencing of M13 clones at the ends of the contigs. The prefinished project consisted of five contigs (10 kb–70 kb); gaps were filled using PCR and primer walking. The complete, finished sequence encompasses a region of 204

kb in the midportion of mouse chromosome 5 (Figs. 1, 2). Figure 2, a percentage identity plot (PIP; see below) of the mouse sequence against two unfinished human BAC clones (GenBank accession nos. AC069200 and AC024243), displays the organization of genes and their exons, repetitive elements, CpG islands, predicted promoter elements, pseudogenes, and genetic markers within the region. The complete loci of two previously identified genes, *Clock* (GenBank accession no. AF000998) and *Tpardl* (*pFT27*; GenBank accession no. M23568), map to this region. The 3′ end of another previously identified gene, *Neuromedin U* (Nmu) precursor (GenBank AF203444), maps to the telomeric end of BAC 54. Two Unigene clusters lie within this interval, one at the centromeric end of BAC 52 and the other between *Nmu* and *Clock*. Finally, pseudogenes for platelet-activating factor acetylhydrolase Ib alpha subunit (*Pafah1b1-ps2*), adenopolyposis coli (*APC*), and ubiquitin-conjugating enzyme homolog (*bendless*) exist in the 204-kb region (Joslyn et al. 1991; Su et al. 1992; Yamaguchi et al. 1996; Peterfy et al. 1998). The *Pafah1b1-ps2* and *bendless* pseudogenes are full-length, while the *APC* pseudogene represents 1240 bases near the 3′ end of the functional *APC* gene. All three pseudogenes appear to have arisen via fusion with a member of the long-terminal-repeat retrotransposon family followed by insertion at this locus of study (Miki et al. 1992; Peterfy et al. 1998).
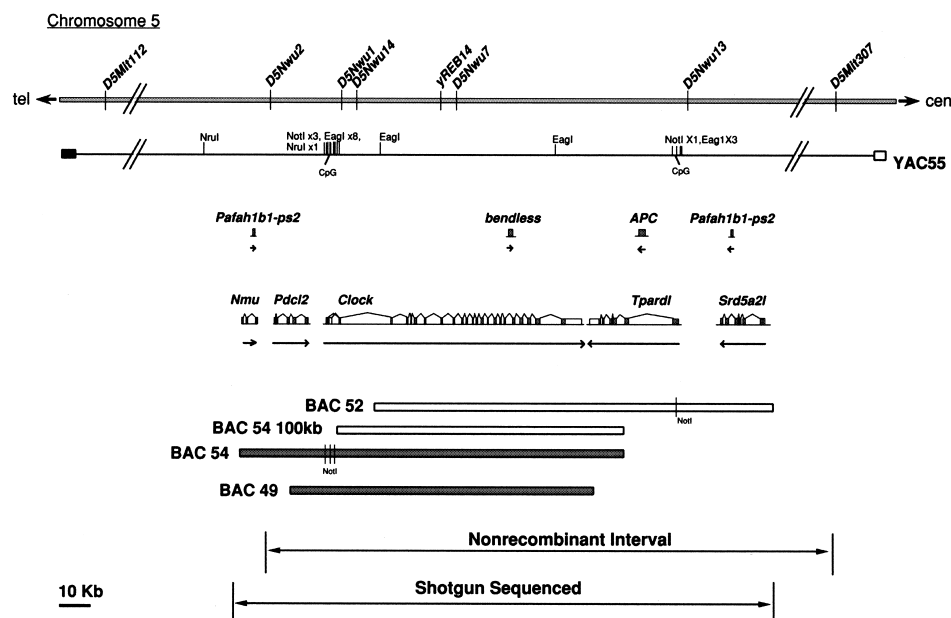
Figure 1 Physical map of the 204 kb flanking the *Clock* locus. The *Clock* locus lies at the midportion of mouse chromosome 5. The genomic locus is updated from Antoch et al. (1997); genetic markers and restriction sites are indicated on the chromosomal and YAC55 maps. The positions of pseudogenes *Pafah1b1-ps2* (2 copies), *bendless*, and *APC* are shown relative to *Nmu*, *Pdcl2*, *Clock*, *Tpardl*, and *Srd5a2l*. The bacterial artificial chromosome (BAC) clones that rescue the *Clock* mutation, BAC54 and BAC49, are shaded, whereas the nonrescuing BAC clones 52 and 54 (100 kb) are shown as open boxes. BAC 54 and BAC 52 were sequenced to generate the data presented here.
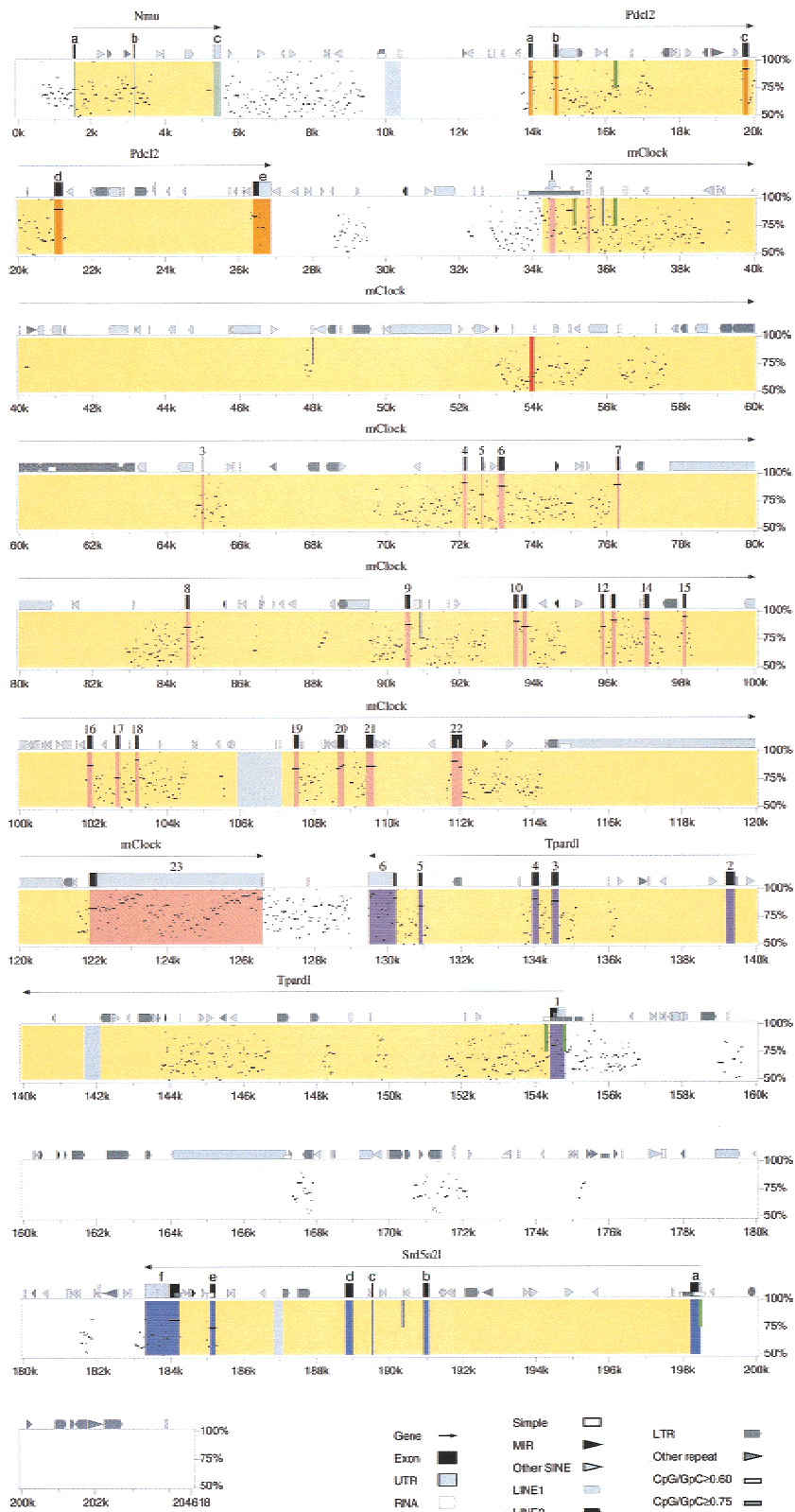
Figure 2 Percentage identity plot (PIP) of 204-kb mouse sequence against unfinished human sequence. The PIP plots the mouse genomic sequence on the *x*-axis and the percentage identity to the human sequence on the *y*-axis. Annotation of the mouse sequence is indicated along the top of each plot: gene name, gene orientation, exons, repetitive elements, and CpG islands. Gene regions are shaded in yellow, and the exons of each gene are marked as full-length bands in corresponding colors: *Nmu*, light green; *Pdcl2*, orange; *Clock*, pink; *Tpardl*, purple; and *Srd5a2l*, blue. Human *Clock* exon 1 is shown in red at 54.0 kb. Pseudogenes are shown as light gray full-length bands: *Pafah1b1-ps2*, 10 kb; *bendless*, 106 kb; *APC*, 142 kb; and *Pafah1b1-ps2*, 187 kb. Predicted promoter elements are shown as dark green half-length bands: promoter 1 (P1), 16.2 kb; P2, 35.0 kb; P3, 36.2 kb; P4, 154.3 kb; P5, 154.8 kb; and P6, 198.5 kb. SSLP markers are shown as dark gray half-bands: D5Nwu 1, 35.9 kb; D5Nwu14, 48.0 kb; D5Nwu7, 90.9 kb; and D5Nwu13, 190.4 kb. See key for annotation of exons, UTRs, repetitive elements, and CpG islands.

Neuromedin U Locus

BLASTN homology revealed the presence of the three 3′-most exons of mouse *Nmu* precursor mRNA (GenBank accession no. AF203444). Exon/intron boundary locations and exon and intron lengths are listed in Table 1. *Nmu* codes for the *Nmu* precursor peptide and is orthologous to rat *Nmu* (Lo et al. 1992). The three mouse *Nmu* exons (named exon a–c) lie at the telomeric end of BAC 54 and encompass bases 506–827 of the gene (Fig. 2). A consensus polyadenylation signal occurs 22 bases upstream of the 3′ end. Neither GRAIL nor GenScan gene prediction programs detected any *Nmu* exons.

A region with BLASTX homology to F_O membrane domain ATP synthase subunit g from cow, mouse, and human (GenBank accession nos. AAB31108, CAA76699, and AAC61597, respectively) was detected within intron b of *Nmu* (data not shown; Collinson et al. 1994). Due to low similarity and the lack of ESTs with which to characterize the area, we did not further investigate this similarity.

Unigene Cluster: Phosducin-Like-2 (Pdcl2) Locus

A PowerBLAST (Zhang and Madden 1997) search revealed a family of >50 mouse RIKEN mouse EST clones immediately centromeric of *Nmu*. The ESTs showed a clear exon/intron structure of 5 exons that spanned 13 kb; the 5′ end of the alignment arose precisely at an intron/exon junction (Fig. 2). A 717-bp open reading frame, devoid of an initiating methionine, lies at the 5′ end of the 1038-bp contiguous alignment of the EST family. The putative translation displays similarity to human Phosducin-Like Protein (PhLP; GenBank accession no. AAD01930). Due to this similarity, we named this putative gene *Phosducin-like protein similar 1* (*Pdcl2*). Because we did not determine the 5′ end of this gene, we annotate it as exon a to exon e. At the 3′ end, a polyadenylation signal occurs at bases 1020–1027. GRAIL detected exon b with inaccuracies at the 5′ and 3′ ends and accurately predicted exons c and d. GenScan accurately predicted all five exons, as well as a potential 87-bp initiator exon 1682 bases upstream of exon a.

Clock Locus

The *Clock* gene spans 90 kb and contains 23 conventional exons, as well as one alternatively spliced exon (exon 1a; in Fig. 2, exon 1a is not included; King et al. 1997a). Exon/intron boundary locations, exon lengths, and intron lengths are listed in Table 1. The ATG initiator codon occurs in exon 4 and agrees well with the Kozak consensus sequence (Kozak 1996; King et al. 1997a). The region encoding the basic helix-loop-helix regulatory motif spans exons 5 and 6; the PAS A-encoding region lies in exons 7, 8, and 9; the PAS B-encoding region lies in exons 12 and 13; and the region coding for a glutamine rich domain is found in exons 19–23. GRAIL and GenScan differed remarkably in their abilities to detect *Clock* exons (Table 1). GRAIL identified only 13 of the 24 exons; of the 13 *Clock* exons GRAIL found, 7 were assigned an incorrect splice site. In contrast, GenScan detected 17 *Clock* exons, 16 of which were predicted with accurate splice sites. Exons 1a, 1b, 2, 8, and 10, which range in size from 70 bp to 181 bp, were not detected by either program. Single nucleotide polymorphisms at 34.5 kb and 124.0 kb were detected between C57BL/6J, the strain used in circadian behavior experiments and that carries the *Clock* mutation (Jackson Laboratory Stock no. 002923), and 129S1/Sv- + ^{P} + ^{Tyr} - ^{c} *Mgf*$^{Sl-J/}$ + (Jackson Laboratory Stock no. 00090), the 129 strain from which the CJ7 ES cell line was derived and used for the Research Genetics BAC library construction (Swiatek and Gridley 1993). Finally, 12 simple sequence repeats were detected in the *Clock* locus, some of which are known to be polymorphic between mouse strains (dark gray half-length bands in Fig. 2; GenBank accession nos. AF146780, AF146792, and AF146786).

We previously reported the presence of two *Clock* transcripts (8 kb and 11 kb) in all tissues that express *Clock* (King et al. 1997a). The smaller transcript was found to be 7490 bp in length (Genbank accession no. AF000998; King et al. 1997a), but the longer transcript was not detected in any of the mouse cDNA libraries we screened. A BLASTN search of the genomic sequence against the EST database (Altschul et al. 1990) revealed several expressed sequence tags (ESTs) just downstream of the 3′ end of the 7490 bp transcript. These ESTs, most of which are 3′ reads in the correct orientation with respect to *Clock*, may represent the terminal end of the larger *Clock* transcript. However, the EST sequences are not contiguous with the known *Clock* transcript. One family of ESTs (960 bp in length) begins 37 bases downstream from *Clock*. This family appears to be primed off of a poly(A) tract in the genomic sequence and likely does not represent a true expressed sequence. A second EST family of 710 bp arises 370 bases downstream from the first family, and this cluster contains a polyadenylation signal in an appropriate position for an expressed sequence (EST family GenBank accession nos. AA607824, AW492507, AA164040, AI843957, AI451178, and AV367325). If this 710-bp EST family does indeed represent *Clock*, it predicts a 9.6-kb transcript.

Tpardl Locus

Tpardl, also known as *pFT27*, encodes a putative transmembrane protein of unknown function (Fig. 2; Akagi et al. 1988). The 5′ UTR, complete coding sequence, and 3′ UTR comprise six exons over 25 kb. Exon/intron

boundary locations, exon lengths, and intron lengths are listed in Table 1. The 3′ end of *Tpardl* lies 2.9 kb away from the *Clock* polyadenylation signal which

generates the 7490-bp transcript and merely 624 bp away from the 710 bp EST family downstream from *Clock*. GenScan detected all six exons; exons 1–5 were

Table 1. Gene Exon/Intron Characteristics

Gene	Exon: Length (bp)	cDNA position (bp)	Detected by GRAIL?	Detected by GenScan	Intron: Length (bp)	Splice acceptor	Splice donor
Nmu							
exon a	54	541–594	No	No	1575	gtcatccttatttcttccagGCAGAATAC	TTATTCAGGgtacgtaca
exon b	40	595–634	No	No	2145	tctaatttattatttcgcagCCACGCAAC	TTTAAAATGgtatgtgct
exon c	182	635–816	No	No	N.A.	ttttttttttcttttttgcagGATGTGAAA	3′UTR after stop
Pdcl2							
exon a	121	(1–121)	No	Yes	596	tttttgggggatgggggtagGATCCAAAC	AGGCCATGGgtagtgcg
exon b	91	(122–212)	− both ends	Yes	5024	acctaagatttctgttttagTTAAACCAT	AATATATAGgtacagcac
exon c	144	(213–356)	Yes	Yes	1117	aatattttgttttacttagAGAAAAGCG	TAGATCAAGgtaattatc
exon d	209	(357–565)	Yes	Yes	5201	actgtgtattctctctttagTGTCCCAAT	AGCTAGAAGgtaatggtg
exon e	473	(566–1038)	No	Yes	N.A.	taggtttttatctattgtagAACTTGAAT	3′UTR after stop
Clock							
exon 1a	(70)	(1–70)	No	No	(1152)	5′UTR	GCGCTCCCGgtgagtgcg
exon 1b	181	1–181	No	No	865	5′UTR	TCAGGCACGgtgaggacg
exon 2	79	182–260	No	No	29386	tctggtgtttctattgcagTGAAAGAAA	ACCAGCAAGgtaatttcc
exon 3	85	261–435	− 3′ end	No	7069	cttttgtttttttaaaacagAGTTCTGAT	GTGAAAGAGgtaaaggcg
exon 4	90	436–495	Yes	− 5′ end	383	atgtttctttctcacaagGAGAAGTAC	TGTTGACAGgtatgtttt
exon 5	60	436–495	+5′ end	Yes	417	ctctgtctttctcttgtagAGATGACAG	AGCAAAAAGgtagttagc
exon 6	149	496–644	+3′ end	Yes	3025	taatttctttttcttcatagAGTATCTAG	AACATAAAGgtaaagtgc
exon 7	92	645–736	Yes	Yes	8211	acgtgtcaatctgtttacagAGACCACTG	ATGTTAGAGgtatgttca
exon 8	90	737–826	No	No	5903	accattatgtttaatttcagGCTCTTGAT	CATTTACCAgtaagtatg
exon 9	121	827–947	Yes	No	2820	ttttttttttttattttcagTCTGATCTT	ACTTAAAATgtaagtagg
exon 10	114	948–1061	No	No	107	cttttttatcacttattccagCAAAAAATC	TAACCAGTGgtaagttaa
exon 11	119	1062–1180	No	Yes	2034	atgtctccttgctgttttagTATCAACTT	TTCATCAAGgtatgcttc
exon 12	83	1181–1263	No	Yes	190	acttgttaatttgtttgtagGAAATGTGT	AGATCACAGgtaacatta
exon 13	107	1264–1370	Yes	Yes	764	attattactgtataatttagGGCACCACC	ACGAGCACTgtaagtagc
exon 14	148	1371–1518	No	Yes	923	ttttattttttttatttttagTAATGCAAT	TGTAGTAAGgtaataact
exon 15	76	1519–1594	Yes	Yes	3709	ttggttctttccatttgtagTTATTGCAG	GCTGACAAAgtatgtttc
exon 16	142	1595–1736	Yes	Yes	638	tgttcctcttatctccttagAGCCAAGAT	ACCCTTCCTgtgtgagtgcc
exon 17	101	1737–1837	No	Yes	426	tctctgttgactgtctttagCCACACCGA	AGCAGTCAGgtacgcctt
exon 18	90	1838–1927	No	Yes	4217	atcttttattttgcttctagTCCATAAAC	ATGTCACAGgtatttttg
exon 19	153	1928–2080	− 5′ end	Yes	1045	ctttccatgtgctgcttcagTTTCAGTTT	GGGCTACAGgtaacttat
exon 20	206	2081–2481	− 3′ end	Yes	601	tgtgatctttgttttcaaagATGTTTTTG	TCAACTCAGgtaattgac
exon 21	206	2276–2481	− 5′ end	Yes	2105	ttccatacgatctttctagCAGAGTCAA	ACAGATAAGgtagttgtc
exon 22	295	2483–2776	+3′ end	Yes	9847	tattttgttttctctcacagATTTTCTCA	TTCTTACAGgtaaccccc
exon 23	4702	2777–7478	No	Yes	N.A.	atcatccttttttgtttttagACATCTAGG	3′UTR after stop
Tpardl							
exon 1	402	2–405	− 5′ end	Yes	14988	5′UTR	CGGGCCGAGgtgagagcc
exon 2	226	406–631	Yes	Yes	4536	atttggttttgattttcagAAAGGATTG	GCTTGTCGGgtaagggca
exon 3	176	632–807	Yes	Yes	453	tgttggcatttccttacagtTTTTTGTTTG	GATGAAGAAgtaagttgt
exon 4	180	808–987	No	Yes	2989	gtattttcatttgtttcagTTCCAACGA	GCTAGAGAGgtgagtgta
exon 5	106	988–1093	No	Yes	606	caagctacatctgcttccagGACCCTTAT	TCCGAACTGgtaaggtct
exon 6	735	1094–1828	No	− 3′ end	N.A.	cttctctctctctttccagTGACAATCA	3′ UTR after stop
Srd5a2l							
exon a	(278)	(1–278)	− 5′ end	− 5′ end	7139	N.D.	CCCCAAGAGgtaacgtcg
exon b	143	(279–421)	Yes	Yes	1363	aactcttcctccttctgcagGTACTTTTC	AGTTCCAAGgtaatttgg
exon c	36	(422–457)	Yes	Yes	519	cttgcttttcctttggcagCCCTGGAGA	GGATGCCAGgtaagtgct
exon d	198	(458–655)	− 5′ end	Yes	3578	ccttctctcctgcattgcagCGGCTGAGC	ACAAGAATGgtaagtggc
exon e	135	(656–710)	No	Yes	858	cagaactcctattcccgtagTGTATGTTC	ACAAAAAGgtgaggtga
exon f	936	(791–1726)	Yes	− 3′ end	N.A.	cgatcgcccttccgtttcagGTGTGGTCA	3′ UTR after stop

Clock exon 1a does not appear in the Genbank cDNA entry (accession no. AF000998), hence the parentheses around its values. *Nmu*, *Pdcl2* and *Srd5a2l* exons are assigned by letters rather than numbers, as the 5′ ends of these genes have not been experimentally mapped; exon position values are in parentheses for the same reason. Intron sequence is noted in lowercase, exon sequence in uppercase. Under columns for Detected by GRAIL and Detected by GenScan, "−" and "+" denote that the predicted exon was underestimated or overestimated, respectively, at the specified end. N.A., not applicable. N.D., not determined.

predicted accurately, but the 3′ end of exon 6 was underestimated.

Unigene Cluster: *Steroid 5–alpha Reductase Type 2–Like (Srd5a2l)* Locus

The PowerBLAST search revealed a second family of >50 mouse and human ESTs centromeric of *Clock*. When aligned to the genomic sequence, the mouse ESTs represent a gene that spans 15 kb and includes six exons (Fig. 2). The 1746-bp contiguous EST alignment contains a large open reading frame that encodes a putative protein of at least 330 amino acids. This hypothetical protein shares similarity with *Caenorhabditis elegans* hypothetical protein Z71178 (GenBank accession no. CAA94885), as well as human and rat 3-oxo-5-alpha-steroid 4-dehydrogenase 2 (GenBank accession nos. P31213 and P31214, respectively; Andersson et al. 1991; Normington and Russell 1992; Wilson et al. 1994). Due to these protein similarities, we named this gene *Steroid 5-alpha reductase type 2–like (Srd5a2l)*. Sequencing of the 3′-most EST clone (IMAGE 949802, GenBank accession no. AA537506) revealed two polyadenylation signals, one at bases 1483–1490 and the other at bases 1706–1713 of the contiguous EST alignment. IMAGE clone 949802 was used to probe a Clontech mouse multiple tissue Northern blot (data not shown). Transcripts of ~1.6 kb and ~1.8 kb in length were detected, which confirmed the existence of *Srd5a2l* and is consistent with the two polyadenylation sites. *Srd5a2l* is highly expressed in liver and kidney; moderately expressed in heart, spleen, and lung; minimally expressed in brain and testis; and not expressed in skeletal muscle. Although the transcript sizes agree reasonably well with the length of the contiguous EST alignment, no conclusive data confirms the 5′ end of *Srd5a2l*. Therefore, we consider this gene partial and annotate the gene as exon a to exon f. GenScan detected all of the proposed exons except exon e, which GRAIL accurately predicted.

Comparative Genomic Sequence Analysis

Previously, the human *CLOCK* exon structure was determined via sequencing of a human BAC using *CLOCK* cDNA-derived primers (Steeves et al. 1999). The exon structure of mouse *Clock* and human *CLOCK* is perfectly conserved between mouse coding exons 4–23 and human coding exons 3–22 (human exons were named arbitrarily, as the BAC from which they were sequenced does not contain the 5′ end of the gene; Steeves et al. 1999). Human *CLOCK* 5′ UTR exons 1 and 2 do not appear to be conserved with mouse *Clock* 5′ UTR exons 1, 2, or 3 (data not shown). We searched the HTGS database with the human *CLOCK* (GenBank accession no. AB002332; Nagase et al. 1997) and the mouse *Pdcl2* cDNA sequences and identified two nonoverlapping human chromosome 4 BAC clones.

Both clones are in HGTS Phase 1 of sequencing at the Washington University Genome Sequencing Center. BAC RP11–528I4 (GenBank accession no. AC024243), which contains sequence similarities to *Pdcl2*, contains 164070 bp in 18 unordered pieces. BAC clone RP11–177J6 (GenBank accession no. AC069200), which contains the human *CLOCK* transcription unit, contains 148205 bp in 19 unordered pieces. All *CLOCK* exons map with 100% identity to RP11–177J6. Of note, human exons 1, 2, and 3 align within a single contig; as in mouse, a large intron (20.4 kb) lies between human exons 1 and 2. Human exons 11–22 align within the largest contig, while exons 4–10 are interspersed among four small contigs.

We used PipMaker (http://nog.cse.psu.edu/pipmaker) to create a PIP for comparison of the finished 204-kb mouse sequence to the unfinished human sequences (Fig. 2; Schwartz et al. 2000). The PIP indicates extensive similarity between the human RP11–528I4 (*Pdcl2*-containing) sequence and the 0 kb–40 kb subregion of mouse sequence, as well as a small region of similarity at 48 kb of the mouse sequence. The region at 48 kb has no significant similarity to sequences in the BLAST nonredundant nucleotide database. The PIP also displays similarity between the human RP11–177J6 (*CLOCK*-containing) sequence and the 53.0 kb–185.2 kb subregion of the mouse sequence. The PIP, therefore, suggests that the gap between the two human clones corresponds to an ~13-kb region in mouse (40 kb–53 kb), all of which lies in the 29-kb *mClock* intron 2. The coding exons of all genes in the 204-kb mouse sequence show >75% identity with the human sequence. Significant similarity exists within the unmasked portions of introns as well (mouse repetitive elements were masked using RepeatMasker; see Methods) and may represent conserved regulatory regions.

Promoter Analysis

Transgenic Rescue of the Clock *Mutant Phenotype Using BAC 49*

Previously, we showed that the 140-kb transgene BAC 54 could rescue the phenotypic effects of the *Clock* mutation (Antoch et al. 1997). BAC 54 contains the entire *Clock* gene, 34 kb of upstream sequence and 12 kb of downstream sequence. The 120-kb BAC 49 flanks the *Clock* gene more tightly than BAC 54, as it contains the complete *Clock* coding region, 25 kb of upstream sequence and only 4 kb of downstream sequence. Given this information, we generated a transgenic line via pronuclear injection of BAC 49 to confirm functionally the extent of the *Clock* gene regulatory region.

Due to the breeding scheme, all BAC 49 transgenic founder progeny were heterozygous for the *Clock* mutation. Representative activity records show that BAC 49 transgene-positive progeny show wild-type pheno-

typic effects: The free-running period is shorter than 24 h (mean = 23.51 ± 0.052 h), and a saturating light pulse administered at approximately circadian time 17 results in a 3.77 ± 0.453-h phase delay (Fig. 3). In contrast, nontransgenic progeny from this founder express all the phenotypic characteristics of *Clock/+* animals, including longer circadian period (mean = 24.04 ± 0.06 h) and 5.645 ± 0.706-h phase delay in response to a light pulse at circadian time 17 (Fig. 3). These results indicate that the genomic interval that is
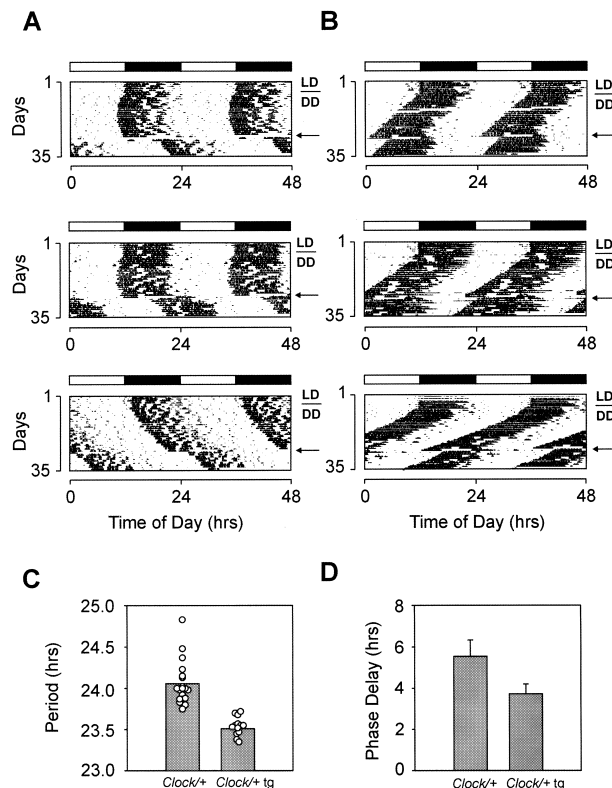
Figure 3 The BAC 49 transgene rescues *Clock/+* mutant phenotype. Wheel-running activity records and period estimates of *Clock/+* F1 transgenic mice and their nontransgenic littermates are shown. A total of 19 control and 15 transgenic animals were tested for their circadian phenotype. All animals were maintained on an LD12:12 cycle for 1 wk (LD) and were transferred into constant darkness (DD) on the day indicated by the bar to the right of each actogram. All animals received a saturating 6-h light-pulse at CT17 after a 21-d exposure to DD (indicated by an arrow). (*A*) Activity records of three individual *Clock/+* mice without the BAC 49 transgene. Light pulse results in an average phase delay of 5.645 ± 0.706 h. (*B*) Activity records of three individual *Clock/+* mice with the BAC 49 transgene. Light pulse results in an average phase delay of 3.77 ± 0.453 h. (*C*) Period estimate in *Clock/+* nontransgenic mice and *Clock/+* transgenic (tg) mice. Average period length in *Clock/+* mice is 24.04 ± 0.06 h ($n = 19$), while average period length in *Clock/+* tg mice is 23.51 ± 0.052 h ($n = 15$). The difference in average period is significant by *t*-test analysis ($P < 0.00001$). (*D*) Phase delay response to a light pulse in *Clock/+* nontransgenic mice and *Clock/+* transgenic (tg) mice. Average phase delay in *Clock/+* mice is 5.645 ± 0.706 h ($n = 10$), while average phase delay in *Clock/+* tg mice is 3.77 ± 0.453 h ($n = 11$). The difference in average phase delay is significant by *t*-test analysis ($P < 0.03$).

sufficient for *Clock* expression in the circadian context is no larger than 120 kb. Most, if not all, of the circadian regulatory elements are likely to be found either in the 25-kb upstream region or within the exon/intron structure, as the downstream region encompasses only 4 kb and contains the 3′ end of the *Tpardl* gene.

Promoter Prediction Programs

To facilitate the analysis of regulatory regions, particularly those of *Clock*, we utilized the TSSG promoter prediction program and MatInspector transcription factor binding site analysis program (Quandt et al. 1995). Taking the error rate of TSSG into account (at 50%–55% true promoter recognition, one false-positive promoter per 5000 bp will be found on average), we disregarded predicted promoters with a score less than 5.0. Thirteen promoters were predicted throughout the 204-kb region, six of which had a score >7.5. Of these six predicted promoters, promoter 1 (P1) lies within *Pdcl2*, P2 and P3 flank exon 2 of the *Clock* gene, P4 and P5 flank exon 1 of the *Tpardl* gene, and P6 lies just centromeric of *Srd5a2l* (dark green half-length bands in Fig. 2). Groups of lower-scoring promoters flank the beginning of the *Clock* and *Tpardl* genes (data not shown), and, in general, high scores occurred near regions of high GC content in the genomic DNA.

MatInspector identified several interesting binding sites in the regions upstream of *Clock*, *Tpardl*, and *Srd5a2l*. The *Clock* 10-kb upstream region contains four E boxes (two 5′-CACGTG-3′ sequences and two 5′-CAGCTG-3′ sequences), three activator protein 1 (AP-1) sites, seven consensus GATA sites, six fork-head transcription factor (HNF3β) sites, three NF-κB sites, and eight core-consensus Sp1 sites (5′-GGGCGG-3′). To date, all E boxes that confer circadian rhythm regulation have the 5′-CACGTG-3′ sequence; whether additional bases outside of the E box are required for circadian rhythm specificity is unknown (Gekakis et al. 1995; Darlington et al. 1998; Hogenesch et al. 1998; Sangoram et al. 1998; Jin et al. 1999). A 7-kb region comprising both predicted promoters of *Tpardl* includes four E boxes (three subtypes of the 5′-CANNGT-3′ consensus), two fork-head transcription factor sites, a cyclic AMP response element binding (CREB) site, and four Sp1 sites near the P4 promoter. An AP-1 site and nuclear factor Y (NFY) site lie near the P5 promoter. The 3.1-kb region upstream of *Srd5a2l* contains 6 E boxes (5′-CACGTG-3′ and 5′-CAGCTG-3′), 4 Sp1 sites, an AP-1 site, and an NFY/CAAT binding site. We interpret all promoter and binding-site predictions with caution, as no prediction program is perfect; however, these results will aid in the design of experiments concerning the regulation of these genes.

DISCUSSION

The results of the completed yeast, *C. elegans* and *Dro-*

sophila genome sequencing projects, as well as the growing human and mouse genome sequencing projects, will generate a genomic structure/function framework on which gene expression models can be built across eukaryotic families. The field of circadian rhythms will certainly benefit from the complete sequencing of the mouse and *Drosophila* genomes, as these organisms constitute two of the four well-understood, genetically amenable circadian model systems. The current circadian model indicates that the regulation of circadian gene expression is a key component of rhythm generation. The genomic sequence of circadian loci will provide much information toward understanding this regulation. To this end, we sequenced two BAC clones that encompass the mouse circadian *Clock* locus. In addition to *Clock*, this 204-kb region contains the *Tpardl* (*pFT27*) gene, the 3′ end of the *Nmu* gene, and two Unigene clusters.

Each Unigene cluster shows similarity to known proteins. The PDCL2 putative product encodes a protein of at least 239 amino acids. The conceptual translation shows similarity throughout the coding sequence to predicted proteins from *Drosophila melanogaster* (GenBank accession no. AAF49974.1), *Schizosaccharomyces pombe* (GenBank accession nos. CAB39851.2 and T40100), *Arabidopsis thaliana* (GenBank accession no. CAB87763.1), and *Saccharomyces cerevisiae* (GenBank accession no. S67183). Lower similarity is found within PDCL2 putative amino acids 35–179 to the carboxy-terminal portion of several orthologs of Phosducin-Like Protein (PHLP), including human (GenBank accession no. AAC78849.1) and rat (long and short forms; GenBank accession nos. AAB00333.1 and AAB00334.1, respectively). In the retina, phosducin down-regulates the phototransduction cascade by binding the G-protein $\beta\gamma$ subunit and preventing reaggregation with the $G\alpha$ transducin subunit (Lee et al. 1992). Rat phosducin-like protein potentially regulates G-protein function, and the PhLP1 phosducin family isoform was shown to bind $G\beta\gamma$ (Miles et al. 1993; Craft et al. 1998). The exclusively testicular expression of *Pdcl2* suggests a possible G-protein regulator in testicular function. The SRD5A2L putative product encodes a protein of at least 331 amino acids; this conceptual translation shows highest similarity throughout the coding region with unknown proteins in *Arabidopsis thaliana* (GenBank accession nos. AC007195_5 and AC010926_21), CG7840 gene product in *Drosophila melanogaster* (GenBank accession nos. AAF52635.1), and hypothetical protein B0024.13 in *C. elegans* (GenBank accession no. T18648) with BLASTX-nr E values from 8^{e-20} to 2^{e-14} (Altschul et al. 1990). SRD5A2L putative amino acids 238–349 shows lower similarity with members of the 3-oxo-5 alpha-steroid 4-dehydrogenase type 2 (steroid 5-alpha-reductase 2) family, with a BLASTX-nr E value of 5^{e-10}

to the human ortholog (GenBank accession no. P31213; Labrie et al. 1992). Steroid 5-alpha-reductase 2 is a critical enzyme in androgen physiology and is expressed both in androgen-responsive tissues such as prostate as well as in hair follicles, cerebral cortex, hepatocytes, bile duct cells, and fat cells (Andersson et al. 1991; Eicheler et al. 1994). *Srd5a2l* is widely expressed, with highest expression in liver and kidney (Fig. 3), and its function remains to be determined.

The function of *Tpardl* remains unknown, even 11 years after it was cloned from the pluripotent mouse F9 teratocarcinoma cell line (Akagi et al. 1988). Its gene product encodes a 323–amino acid protein with six putative transmembrane domains, and the gene is down-regulated following treatment with retinoic acid, cAMP, or the phorbol ester tetradecanoyl phorbol acetate (TPA) (Miyata et al. 1987; Akagi et al. 1988). A CpG island exists upstream of the *Tpardl* gene, and the TSSG program predicted two promoters near this island. The MatInspector binding-site prediction program (Quandt et al. 1995) identified binding sites for E box binding proteins, fork-head transcription factors, NFY, AP-1, and CREB. In fact, this gene contains the only consensus CRE of any of the genes reported here; it may play a role in the regulatory effects of cAMP on *Tpardl*. *Tpardl* is a candidate for the tumorigenic effects of TPA (Akagi et al. 1988); knowledge of the *Tpardl* promoter sequence may provide information about the relationship between TPA, *Tpardl* down-regulation, and tumorigenesis.

In contrast to the other genes, the function of *Clock* is relatively well understood. The *Clock* mutation was discovered in mice through a mutagenesis screen for altered locomotor activity (Vitaterna et al. 1994). In constant darkness, wild-type mice (strain C57BL/6J) typically complete one circadian cycle with a period ranging from 23.3 to 23.8 h (Vitaterna et al. 1994). In contrast, heterozygous *Clock* mutant mice show an ~1-h increase in period length over wild-type, and homozygous *Clock* mutants show an ~4-h increase in period length, followed by a loss of rhythmicity (Vitaterna et al. 1994). Therefore, the *Clock* mutation affects both period length and stability (Vitaterna et al. 1994). The *Clock* gene was cloned using a combination of techniques: (1) rescue of the *Clock* mutant phenotype by transgenic BAC clones, (2) transcription unit analysis via cDNA subtraction and an SCN cDNA library screen, (3) elucidation of gene structure using genomic survey sequencing data, and (4) mutation detection through comparison of genomic sequence from *Clock* mutant animals and wild-type animals (Antoch et al. 1997; King et al. 1997a). *Clock* encodes a bHLH-PAS transcription factor that is a positive activator within the circadian autoregulatory negative feedback loop (King and Takahashi 2000). However, little information exists about the regulation of *Clock* itself. The

transcriptional regulation of *Clock* may not seem important initially; only the circadian activity of the CLOCK protein, which appears to be regulated by the PER, CRY, and possibly TIM proteins, is required for proper rhythm generation (Gekakis et al. 1998; Sangoram et al. 1998; Griffin et al. 1999; Kume et al. 1999). Nonetheless, the regulated expression of *Clock* should still play a critical role in the circadian cycle. For example, the *Clock* expression profile may be different in the SCN, as compared to the rest of the body. Indeed, although no rhythm in *mClock* expression has been reported in the mouse SCN, a low-amplitude *rClock* rhythm was detected in rat retina (Namihira et al. 1999).

The organization of *mClock* exons 1 through 3 is intriguing: exons 1a, 1b, and 2 fall within a 1.5-kb region and then a 29-kb intron precedes exon 3. Excluding repeat elements, this intron shows no similarity to the nonredundant nucleotide, nonredundant protein, and EST databases. No promoter elements were predicted within the repeat-masked region, and the interval contains a binding element profile similar to a nonpromoter region (i.e., a cDNA). We cannot rule out an enhancer function within the intron; even though the nonrescuing BAC clones contain this intron (Fig. 1), it is possible that position effects could mask a regulatory element near the 5′ end of the intron.

The transgenic rescue experiments with BAC 49 reduce the region that is sufficient for proper *Clock* expression in the circadian context from 140 kb to 120 kb. BAC 54 is 135 kb in length; 35 kb of potential promoter sequence and 12 kb of potential 3′ enhancer sequence exist in this construct. BAC 49 contains 25 kb of upstream sequence and only 2–4 kb of downstream sequence (depending on the identity of the 3′ EST family flanking the 3′ end of the known *Clock* gene). The presence of the 3′ end of *Tpardl* within 3 kb of the *Clock* gene argues against a 3′ enhancer sequence in either BAC 49 or BAC 54. Together, these data indicate that most if not all the circadian regulatory elements lie within the *Clock* locus itself and a 25-kb upstream interval.

The human *Clock* transcript shares a high level of structural homology with the mouse *Clock* transcript: From the initiator ATG-containing exon in mouse and in human, each exon/intron boundary is identical between the two species (Steeves et al. 1999). Such conservation of gene structure appears to be a common theme (Hardison et al. 1997; Ansari-Lari et al. 1998; Blechschmidt et al. 1999). Indeed, significant similarity was detected between the mouse sequence and the unfinished human chromosome 4–derived BAC sequence. Three regions of similarity stand out: region A, 28.5 kb–29.6 kb; region B, 32.3 kb–38.7 kb; and region C, 126.7 kb–129 kb (Fig. 2). Region A contains se-

quence present in BAC 49 but absent from the nonrescuing BAC 54 100-kb clone (Fig. 1). This area could represent a *Clock* promoter region that is required for proper expression of the *Clock* gene. In addition, Mat-Inspector detected putative AP-1 and E box binding sites within the 27 kb–34 kb region of the mouse sequence. The AP-1 sites may represent targets for the input pathways that mediate entrainment of circadian rhythms (Kornhauser et al. 1992; Takeuchi et al. 1993). The presence of E box elements in the *Clock* promoter may suggest, but does not dictate, a circadian regulation of *Clock* expression. In region B, the 32.3–34.0-kb portion is excluded from the BAC 54 100-kb construct, but the remainder of this 6.4-kb region is included in the nonrescuing clone. Therefore, region B may be critically important for appropriate expression of *Clock* as a complete, uninterrupted unit. Recent advances in BAC engineering should make it possible to identify any regulatory elements via deletion scanning mutagenesis through regions A and B (Jessen et al. 1998). Region C represents the putative 3′ UTR of the ~10-kb mouse *Clock* transcript described above and provides additional evidence in support of our hypothesis concerning the long *Clock* transcript. Finally, the PIP shows differences between the mouse and human genomic regions, particularly in the large *Clock* intron from each species. Human *CLOCK* exon 1, which shows no similarity to the *mClock* 5′ UTR exons, aligns at 54.0 kb of the mouse sequence (61% identity; red band in Fig. 2), *hCLOCK* exon 2 does not align with the mouse sequence, and *hCLOCK* exon 3 aligns with *mClock* exon 4 at 72.1 kb (93% identity). The mouse-human alignment suggests an 18.1-kb distance between *hCLOCK* exons 1 and 3; however, the *hCLOCK* alignment to human genomic sequence clearly shows a 27-kb distance between human exons 1 and 3 (data not shown). These analyses suggest that rearrangements have occurred in human genomic sequence relative to mouse genomic sequence.

We anticipate that comparative sequence analyses will be of great use in the study of human circadian diseases such as delayed sleep phase syndrome and advanced phase sleep syndrome (Weitzman et al. 1981; Jones et al. 1999). Furthermore, the genes that make up the core circadian autoregulatory feedback loop appear to be well-conserved between *Drosophila* and mammals (Wilsbacher and Takahashi 1998; Dunlap 1999; King and Takahashi 2000). Combining the forward genetic approach of gene discovery with the resources of genomic sequence will lead to rapid, parallel identification of novel circadian genes among many organisms. These genetic tools will undoubtedly accelerate our understanding of the circadian clock, including input pathways to the clock, output pathways from the clock, and the core circadian oscillator itself.

METHODS

Shotgun Sequencing and Finishing of BAC 52 and BAC 54

BAC DNA Isolation and Purification

BAC DNA was isolated from 250 mL cultures via an alkaline lysis method using 6 mL of Solution I (50 mM glucose; 10 mM EDTA at pH 8.0; 25 mM Tris at pH 8.0), 8 mL Solution II (1% SDS, 0.2 N NaOH), and 6 mL Solution III (60 mL of 5M potassium acetate, 11.5 mL of glacial acetic acid, 28.5 mL water). After addition of Solution II, the cultures were mixed gently and incubated on ice for 7.5 min; after addition of Solution III, the lysates were gently mixed and cleared of cellular debris by two 25-min centrifugations at $14,000 \times g$, 4°C. The supernatant was precipitated with 0.7 volumes of isopropanol and centrifuged at $2000 \times g$ for 10 min at room temperature. The air-dried DNA pellet was resuspended in 2.8 mL TE at pH 8.0 that contained 10 µg/µL RNase A, and the solution was incubated for 30 min at 37°C. BAC DNA was purified via double cesium chloride centrifugation; 3.25 g of CsCl (Gallard-Schlesinger Industries) and 280 µL of 10 mg/mL ethidium bromide (EtBr) were added to the resuspended BAC DNA. This solution was centrifuged at $3000 \times g$ for 10 min at room temperature; the density of the clear bottom layer was adjusted to 1.55 g/mL if necessary. DNA was loaded into a 3.9-mL Quick-Seal polyallomer tube (Beckman) and centrifuged in an Optima TLX ultracentrifuge (Beckman) at 80,000 rpm for 16 h at room temperature in a TLN100 rotor. The lower (plasmid) band was pulled using an 18-gauge needle, transferred to a new Quick-Seal tube, filled with 1.55 mg/mL CsCl in TE, and centrifuged at 100,000 rpm for 4 h at room temperature. This second CsCl centrifugation step was required to purify the BAC DNA from contaminating *Escherchia coli* DNA. The lower band from the second spin was pulled, diluted with 3 volumes of water, and EtBr was extracted three times with butanol. DNA was precipitated with 2 volumes of ethanol and stored overnight at −20°C to maximize BAC DNA recovery. Following centrifugation and a 70% ethanol wash, the BAC DNA pellet was air-dried and resuspended in 10–20 µL TE. Concentration was determined using fluorometry. This prep typically yields about 10–20 µg of supercoiled BAC DNA.

M13 Shotgun Library Construction

For most steps, M13 vector preparation and construction of M13 libraries were completed according to the Washington University Genome Sequencing Center protocols (available at http://genome.wustl.edu/gsc/Protocols/protocols.shtml).

M13 Vector Preparation

M13mp19 RFI (Pharmacia) was digested with *Sma*I (NEB) and then precipitated using 1/50 volume of 5 M NaCl and 2.5 volumes of 95% ethanol. The resultant DNA pellet was treated with 5 units/µg DNA of calf intestinal alkaline phosphatase for 1 h at 37°C. M13 DNA was phenol extracted once and then purified using a two-gel procedure: DNA was run on a 0.8% agarose gel in $1 \times$ TAE for 4 h at 60 volts and then the linear M13 band was excised and run again in a new 0.8% low-melting-point agarose gel for 3 h at 60 volts. The linear M13 band was gel-purified and resuspended in water.

Size Selection of BAC DNA

BAC DNA was sonicated using a Heat Systems Ultrasonics sonicator (model W-225R) with a chilled-cup horn probe. 2.5 µg BAC DNA was diluted in $1 \times$ mung bean nuclease buffer (Pharmacia) to a final volume of 60 µL and sonicated for 7–8 sec at power level 6. DNA fragments were blunt-ended using 1 unit mung bean nuclease (Pharmacia) for 10 min at 30°C. Proteins were phenol-chloroform-isoamyl extracted and DNA was precipitated. The resultant DNA pellet was resuspended in 10 µL of water. Size selection was performed by running the DNA on a 0.8% SeaPlaque GTG agarose (FMC) gel in $1 \times$ TAE (no ethidium bromide) for 3.5 h at 60 volts. Size marker lanes run on the same gel were cut out and stained; these markers were used to estimate and excise the 1.3 kb–1.7 kb region in the unstained sheared BAC DNA lane. The size-selected DNA was gel-purified using β-agarase (NEB) according to the quick-equilibration protocol, and 2 units β-agarase per 100 mg of agarose were used at 42°C in an overnight digestion. DNA was precipitated and resuspended in 7 µL of water. DNA concentration was measured using a Hoefer TKO 100 fluorometer (Pharmacia).

Ligation and Transformation

Twenty nanograms of prepared vector and 100–200 ng sheared, size-selected BAC DNA (in 1–3 µL) were added to 1 µL of $10 \times$ blunt-end ligation buffer (0.5M Tris-Cl, 0.1M MgCl2, 0.1M DTT, 0.5mM ATP, 0.5 mg/ml BSA at pH 7.8) and 0.3 µL of T4 high concentration ligase (New England Biolabs) in a final volume of 8 µL. The ligation reaction was incubated overnight at 15°C. For transformation of the M13 library, XL1-Blue MRF′ cells were grown in YT broth +12.5 µg/mL tetracycline to O.D. 600 = 0.8. Top YT agar was freshly prepared, aliquoted into 6-mL samples, maintained at 45°C in a water bath, and supplemented with 50 µL of 25 mg/mL IPTG and 50 µL of 25 mg/mL of X-gal. The ligation reaction was diluted 25-fold in water, and 2 µL were electroporated into 25 µL of electrocompetent XL1-Blue MRF′. Electroporator (BTX 600) settings were voltage = 1.7 kV and resistance = 129 ohm. The electroporated cells were mixed with 400 µL of pre-warmed SOC and 100 µL of XL1-Blue MRF′ lawn cells. This solution was mixed with the top YT and plated onto pre-warmed YT +12.5 µg/mL tetracycline plates. Plates were incubated for 12–16 h at 37°C. Individual M13 plaques were picked with a sterile toothpick, eluted into 400 µL of SM buffer and stored at 4°C.

Ninety–six–Well M13 Template Purification

Preparation of single-stranded M13 DNA was performed using the Solid Phase Reversible Immobilization (SPRI) technique for M13 (Hawkins et al. 1994). All steps were carried out in 96-tube format using Fisher-racked 1.1 mL tubes (P/N 1184294); 650 µL of XL1 Blue MRF′ starter culture was inoculated with 40 µL of the M13-containing SM stock and grown for 16 h with shaking at 37°C. Cultures were centrifuged at $2000 \times g$ for 15 min; 450 µL of lysate was added to 45 µL of 7.5% SDS and incubated for 10 min in an 85°C water bath to lyse the phage coats. Magnetic beads (Bangs Estaphor number MN0009200CN) were washed three times in 0.5 M EDTA at pH 8.0 and resuspended at 20 µg/µL in the same EDTA buffer. DNA was hybridized to 50 µL (1 mg) of washed beads in 545 µL of hybridization solution (26% PEG 8000/20mM MgCl2) for 30 min at room temperature. Because the tubes were full to capacity, they were centrifuged for 3 min at $1000 \times g$ to collect the beads initially. The supernatant was discarded by inverting the tube box and magnetic stand (Promega) together so as to keep the beads and DNA in the tubes. The beads were washed twice in 200 µL of 70% ethanol and collected using the magnetic stand. The beads were air-dried for 10 to 20 min and then DNA was eluted in 30 µL of water for 2 min. Using

this protocol, 192 M13 templates could be prepared in about 5 h with an average yield of 1 µg of sequence-ready DNA. A similar protocol is currently available at http://www-seq. wi.mit.edu/protocols/M13SPRI.html.

Sequencing of M13 Templates

M13 sequencing reactions were run in an ABI PRISM Catalyst Turbo 877 Molecular Biology LabStation using the Dye Primer Cycle Sequencing Ready Reaction-21 M13 Kit (Perkin Elmer) at half-chemistry concentrations: 2.0 µL of A and C premixes and 4.0 µL of G and T premixes were used for each reaction; 1.5 µL and 3.0 µL of template were delivered to the A/C and G/T reactions, respectively. Thermocycling conditions were modified as recommended for use of FS Dye Primer chemistry with single-stranded template, and 15 µg glycogen (Boehringer Mannheim) per reaction was used as carrier after precipitation. Sequences were run on a 4% Long Ranger acrylamide gel (FMC) using an ABI PRISM 377 DNA Sequencer. Electrophoresis run time was 7.5 h, and read length was truncated at 650 bases due to peak expansion and subsequent base-calling errors past ~650 bases. If necessary, ends were trimmed using Sequencher 3.0 (Gene Codes, Inc.) with the following criteria: (1) at the 5′ end, sequence was trimmed until the first 25 bases contained no more than three consecutive ambiguities; (2) 100 bases after the 5′ trim, sequence was trimmed after the first 25 bases that contained more than three consecutive ambiguities; and (3) sequence was trimmed from the 3′ end until the last 25 bases contained no more than three consecutive ambiguities.

Reverse Sequencing

Double-stranded M13 DNA was generated using PCR with M13 Universal and M13 reverse primers on 0.5 µL of the M13 SM stock as template; cycling conditions were 94°C for five minutes, 26 cycles of 94°C for one minute, 55°C for one minutes, and 72°C for one minute; 72°C for 10 minutes; and 4°C hold. PCR products were purified from mineral oil by pipeting the reaction onto a strip of Parafilm (3M), rolling the droplet along 5–10 cm of Parafilm, and collecting the aqueous droplet. Primers and unincorporated nucleotides were removed from the PCR products using the PCR SPRI technique (T. Hawkins, unpubl.). For a 50 µL PCR, 10 µL of washed magnetic beads (see "High throughput M13 prep" above) and 50 µL of PCR hybridization buffer (2.5M NaCl/20% PEG8000) were added to the PCR product, mixed, and incubated for 10 min at room temperature. Beads were magnetically collected, the supernatant was discarded, and the beads were washed two times with 150 µL of 70% ethanol. Beads were air-dried for 10 min and then resuspended in 20 µL of 10 mM Tris-Acetate at pH 7.8 for 5 min. Reverse-sequencing reactions were performed on 50 ng purified PCR product using the Catalyst 877 Turbo robot with Dye Primer M13 Reverse Ready Reaction Mix (Perkin Elmer) as for the M13 forward shotgun samples, and products were run on an ABI 377. The PCR purification protocol is currently available at http://www-seq. wi.mit.edu/protocols/PCRSPRI.html.

Direct BAC Sequencing

BAC DNA was used as sequencing template for finishing purposes. Template was prepared either by CsCl centrifugation or using the Qiagen Midi Plasmid Prep Kit (Qiagen). One to 1.5 µg BAC DNA, 20 pmole primer, and 8 µL Big Dye Terminator Ready Reaction Mix were used in 20-µL reactions. Reactions were run in a Perkin Elmer 9700 thermal cycler using condi-

tions of 95° for 5 minutes; 34 cycles of 95°C for 30 seconds, 56°C for 1 minute, 60°C for 4 minutes; and 4°C hold.

Gap PCR on BAC Template

The Stratagene TaqPlus Long PCR system was used to generate PCR products across four large (3.5 kb, 5.8 kb, 6.5 kb, 7.8 kb) gapsin genomic sequence. Primers were designed from unique genomic sequence flanking the gaps. The TaqPlus low-salt buffer was used in the reaction that produced the 6.8-kb product, while the high-salt buffer was used in all other reactions. The reactions for the three largest products used thermal cycling conditions of 94°C for 5 minutes; 16 cycles of 94°C for 15 seconds, 60°C for 1 minute, and 72°C for 10 minutes; 12 cycles of 94°C for 15 seconds, 60°C for 1 minute, and 72°C for 10 minutes with an increasing increment of 15 seconds per cycle; 72°C for 10 minutes, and 4°C hold. The reaction conditions for the 3.5-kb product were identical except for a 6-min extension time. For the 6.5-kb and 6.8-kb products, ten independent reactions were performed, pooled, and sequenced; for the 3.5 kb and 5.8 kb products, three independent reactions were performed and sequenced individually (to resolve Taq polymerase incorporation errors).

A relatively short (250-bp) gap was amplified using AmpliTaq polymerase (Perkin Elmer) with cycling conditions of 94°C for 2 minutes; 26 cycles of 94°C for 30 seconds, 55°C for 1 minute, and 72°C for 1 minute; 72°C for 20 minutes, and 4°C hold. Four independent reactions were performed and sequenced separately. All gap PCR products were sequenced by primer walking across both strands for the entire length of the product. Reactions were run using the Big Dye Terminator Ready Reaction Mix (Perkin Elmer).

Analysis of Genomic Sequences

All alignments were performed using Sequencher 3.0 (Gene Codes Corp.). Individual M13 sequences were BLAST searched using BLASTN against the nr nucleotide and dbEST databases, BLASTX against the nr protein database, and TBLASTX against the nr protein database (Altschul et al. 1990). For the complete, finished BAC 54 and BAC 52 sequence, repeat elements were masked using RepeatMasker against the rodent Repbase database (A.F.A. Smit and P. Green, unpubl.; available at http://www.genome.washington.edu/UWGC/analysistools/repeatmask.htm). Masked output was BLAST searched in PowerBLAST (Zhang and Madden 1997) against the nr nucleotide, EST, and STS databases for BLASTN (parameters of M = 1, N = −3, S = 40, S2 = 40) and the nr protein database for BLASTX (parameters of S = 90, S2 = 90, -filter = seg). Masked sequence was also searched against GRAIL (Xu et al. 1994) and GenScan (Burge and Karlin 1997) for exon prediction. Promoter analysis was performed using TSSG (V.V. Solovyev et al., unpubl.; available at http://dot.imgen.bcm.tmc.edu:9331/seq-search/gene-search.html) and MatInspector (Quandt et al. 1995). The PIP was created using Advanced PipMaker (http://nog.cse.psu.edu/pipmaker; Schwartz et al. 2000) using the "Search both strands" and "Show all matches" options with the following files: sequence 1, GenBank accession no. AF146793 (mouse genomic); sequence 2, GenBank accession no. AC069200 and AC024243 (human genomic); mask file, RepeatMasker against the rodent Repbase database (as above); exon file, manually generated; first sequence underlay file, manually generated.

Generation of BAC Transgenic Mice

Circular BAC DNA was purified using the CsCl method de-

scribed above, resuspended in injection buffer (10 mM Tris-HCl at pH 7.5, 0.1 mM EDTA, 100 mM NaCl, 30 µM spermine, 70 µM spermidine) to a concentration of 1 ng/µL, and injected into the oocytes of wild-type CD1 females. Injections were done at the Northwestern University/CMIER Transgenic facility as previously described (Hogan et al. 1994). Transgenic mice were identified using both PCR and Southern blot analysis of genomic DNA prepared from tail biopsies as described in Antoch et al. (1997). A 240-bp fragment corresponding to the Sp6-end of BAC vector-insert junction was amplified using primers 5′-GACCATGATTACGCCAAGCTATTTAG and 5′-ACTGTAGCAGTCATCACCTTAACCC. Out of 14 mice tested, two were positive for the transgene by PCR analysis and only one (no. 5633) by both PCR and Southern hybridization. This transgenic founder (female) was transferred to the Northwestern University Center for Experimental Animal Resources and crossed to a *Clock/Clock* (BALB/cJ X C57BL/6J) F3 male. The progeny (all heterozygous for the *Clock* mutation) were analyzed for transgene presence using PCR. Both transgenic animals and their nontransgenic littermates were analyzed for their locomotor activity.

Locomotor Activity Rhythm Monitoring and Analysis

Mice were housed in LD 12:12 (lights on at 0500, Central Standard Time) and behaviorally tested after reaching 8 wk of age. Locomotor activity was performed as described in Vitaterna et al. (1994). Mice were individually housed in the running-wheel cages for 1 wk in LD 12:12 and then transferred to constant darkness for 3 wk. To confirm *Clock*/+ phenotype, a 6-h saturating light pulse at circadian time 17 was given at day 21 in constant darkness. Activity was monitored with an on-line PC computer system (Stanford Software Systems, Chronobiology Kit).

Period calculations were performed for a 20-d interval in constant darkness (1–20 d) by χ^2 periodogram analysis as previously described (Antoch et al. 1997) using ClockLab Software (Circadian Systems). Statistical analyses were performed using one-way Generalized Linear Models Analysis of Variance using Number Cruncher Statistical Systems (NCSS 97).

REFERENCES

Akagi, J.M., Nomiyama, H., Setoyama, C., and Akagi, M. 1988. Messenger RNA expressed in mouse teratocarcinoma stem cells and down-regulated by a tumor-promoting phorbol ester codes for a novel transmembrane protein. *Biochem. Biophys. Res. Commun.* **157:** 548–557.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215:** 403–410.

Andersson, S., Berman, D.M., Jenkins, E.P., and Russel, D.W. 1991. Deletion of steroid 5 alpha-reductase 2 gene in male pseudohermaphroditism. *Nature* **354:** 159–161.

Ansari-Lari, M.A., Oeltjen, J.C., Schwartz, S., Zhang, Z., Munzy, D.M., Lu, J., Gorrell, J.H., Chinault, A.C., Belmont, J.W., Miller, W., et al. 1998. Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res.* **8:** 29–40.

Antoch, M.P., Song, E-J., Chang, A.-M., Vitaterna, M.H., Zhao, Y., Wilsbacher, L.D., Sangoram, A.M., King, D.P., Pinto, L.H., and Takahashi, J.S. 1997. Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. *Cell* **89:** 655–667.

Blechschmidt, K., Schweiger, M., Wertz, K., Poulson, R., Christensen, H.M., Rosenthal, A., Lehrach, H., and Yaspo, M.L. 1999. The mouse aire gene: Comparative genomic sequencing, gene organization, and expression. *Genome Res.* **9:** 158–166.

Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268:** 78–94.

Collinson, I.R., Runswick, M.J., Buchanan, S.K., Fearnley, I.M., Skehel, J.M., van Raaij, M.J., Griffiths, D.E., and Walker, J.E. 1994. Fo membrane domain of ATP synthase from bovine heart mitochondria: Purification, subunit composition, and reconstitution with F1-ATPase. *Biochemistry* **33:** 7971–7978.

Craft, C.M., Xu, J., Slepak, V.Z., Zhan-Poe, X., Zhu, X., Brown, B., and Lolley, R.N. 1998. PhLPs and PhLOPs in the phosducin family of G beta-gamma binding proteins. *Biochemistry* **37:** 15758–15772.

Darlington, T.K., Wager-Smith, K., Ceriani, M.F., Staknis, D., Gekakis, N., Steeves, T.D.L., Wietz, C.J., Takahashi, J.S., and Kay, S.A. 1998. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim. Science* **280:** 1599–1603.

Dunlap, J.C. 1999. Molecular bases for circadian clocks. *Cell* **96:** 271–290.

Eicheler, W., Tuohimaa, P., Vilja, P., Adermann, K., Forssmann, W.G., and Aumuller, G. 1994. Immunocytochemical localization of human 5 alpha-reductase 2 with polyclonal antibodies in androgen target and non-target human tissues. *J. Histochem. Cytochem.* **42:** 667–675.

Gekakis, N., Saez, L., Delahaye-Brown, A.-M., Myers, M.P., Sehgal, A., Young, M.W., and Weitz, C.J. 1995. Isolation of *timeless* by PER protein interaction: Defective interaction between *timeless* protein and long-period mutant PERL. *Science* **270:** 811–815.

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D P., Takahashi, J.S., and Weitz, C.J. 1998. Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280:** 1564–1569.

Griffin, E.A.J., Staknis, D., and Weitz, C.J. 1999. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* **286:** 768–771.

Hao, H., Allen, D.L., and Hardin, P.E. 1997. A circadian enhancer mediates PER-dependent mRNA cycling in *Drosophila melanogaster. Mol. Cell Biol.* **17:** 3687–3693.

Hardison, R.C., Oeltjen, J., and Miller, W. 1997. Long human-mouse sequence alignments reveal novel regulatory elements: A reason to sequence the mouse genome. *Genome Res.* **7:** 959–966.

Hawkins, T.L., O'Conner-Morin, T., Roy, A., and Santillan, C. 1994. DNA purification and isolation using a solid-phase. *Nucleic Acids Res.* **22:** 4543–4544.

Hogan, B., Beddington, R., Constantini, F., and Lacey, E. 1994. Manipulating the Mouse Embryo, A Laboratory Manual, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hogenesch, J.B., Gu, Y.-Z., Jain, S., and Bradfield, C.A. 1998. The basic helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proc. Natl. Acad. Sci.* **95:** 5474–5479.

Jessen, J.R., Meng, A., McFarlane, R.J., Paw, B.H., Zon, L.I., Smith, G.R., and Lin, S. 1998. Modification of bacterial artificial chromosomes through chi-stimulated homologous recombination and its application in zebrafish transgenesis. *Proc. Natl. Acad. Sci.* **95:** 5121–5126.

Jin, X., Shearman, L.P., Weaver, D.R., Zylka, M.J., de Vries, G.J., and Reppert, S.M. 1999. A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell* **96:** 57–68.

Jones, C.R., Campbell, S.S., Zone, S.E., Cooper, F., DeSano, A., Murphy, P.J., Jones, B., Czajkowski, L., and Ptacek, L.J. 1999. Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nat. Med.* **5:** 1062–1065.

Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., and Robertson, M. 1991. Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell* **66:** 601–613.

King, D.P. and Takahashi, J.S. 2000. Molecular Genetics of Circadian Rhythms in Mammals. *Annu. Rev. Neurosci.* **23:** 713–742.

King, D.P., Zhao, Y., Sangoram, A.M., Wilsbacher, L.D., Tanaka, M., Antoch, M.P., Steeves, T.D.L., Vitaterna, M.H., Kornhauser, J.M., Lowrey, P.L.,et al. 1997a. Positional cloning of the mouse circadian *Clock* gene. *Cell* **89:** 641–653.

King, D.P., Vitaterna, M.H., Chang, A.M., Dove, W.F., Pinto, L.H., Turek, F.W., and Takahashi, J.S. 1997b. The mouse *Clock* mutation behaves as an antimorph and maps within the *W*19H deletion, distal of *Kit*. *Genetics* **146:** 1049–1060.

Kornhauser, J.M., Nelson, D.E., Mayo, K.E., and Takahashi, J.S. 1992. Regulation of jun-B messenger RNA and AP-1 activity by light and a circadian clock. *Science* **255:** 1581–1584.

Kozak, M. 1996. Interpreting cDNA sequences: Some insights from studies on translation. *Mammal. Gen.* **7:** 563–574.

Kume, K., Zylka, M.J., Sriram, S., Shearman, L.P., Weaver, D.R., Jin, X., Maywood, E.S., Hastings, M.H., and Reppert, S.M. 1999. mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98:** 193–205.

Labrie, F., Sugimoto, Y., Luu-The, V., Simard, J., Lachance, Y., Bachvarov, D., Leblanc, G., Durocher, F., and Paquet, N. 1992. Structure of human type II 5 alpha-reductase gene. *Endocrinology* **131:** 1571–1573.

Lee, R.H., Ting, T.D., Lieberman, B.S., Tobias, D.E., Lolley, R.N., and Ho, Y.K. 1992. Regulation of retinal cGMP cascade by phosducin in bovine rod photoreceptor cells. Interaction of phosducin and transducin. *J. Biol. Chem.* **267:** 25104–25112.

Lo, G., Legon, S., Austin, C., Wallis, S., Wang, Z., and Bloom, S.R. 1992. Characterization of complementary DNA encoding the rat neuromedin U precursor. *Mol. Endocrinol.* **6:** 1538–1544.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. 1992. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52:** 643–645.

Miles, M.F., Barhite, S., Sganga, M., and Elliot, M. 1993. Phosducin-like protein: An ethanol-responsive potential modulator of guanine nucleotide-bnding protein function. *Proc. Natl. Acad. Sci.* **90:** 10831–10835.

Miyata, A., Akagi, J., Fukuda, M., Setoyama, C., Shimada, K., and Akagi, M. 1987. Isolation of complementary DNA clones for genes exhibiting reduced expression after treatment of mouse teratocarcinoma stem cells with a tumor-promoting phorbol ester. *J. Natl. Cancer Inst.* **79:** 881–890.

Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N., and Ohara, O. 1997. Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res.* **4:** 141–150.

Namihira, M., Honma, S., Abe, H., Tanahashi, Y., Ikeda, M., and Honma, K. 1999. Circadian rhythms and light responsiveness of mammalian clock gene, Clock and BMAL1, transcripts in the rat retina. *Neurosci. Lett.* **271:** 1–4.

Normington, K. and Russell, D.W. 1992. Tissue distribution and kinetic characteristics of rat steroid 5 alpha-reductase isozymes. Evidence for distinct physiological functions. *J. Biol. Chem.* **267:** 19548–19554.

Peterfy, M., Gyuris, T., and Antonio, L. 1998. Characterization and chromosomal mapping of two pseudogenes of the mouse Pafaha/Lis1 gene: Retrointegration hotspots in the mouse genome. *Gene* **216:** 225–231.

Quandt, K., Frech, K., Karas, H., Wingender, E., and Werner, T. 1995. MatInd and MatInspector: New, fast, and versatile tools for detection of consensus matches in nucleotide sequence data.

Nucleic Acids Res. **23:** 4878–4884.

Ripperger, J.A., Shearman, L.P., Reppert, S.M., and Schibler, U. 2000. CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes & Dev.* **14:** 679–689.

Sangoram, A. M., Saez, L., Antoch, M. P., Gekakis, N., Staknis, D., Whiteley, A., Fruechte, E. M., Vitaterna, M. H., Shimomura, K., King, D. P., et al. 1998. Mammalian circadian autoregulatory loop: A timeless ortholog and mPer1 interact and negatively regulate CLOCK-BMAL1-induced transcription. *Neuron* **21:** 1101–1113.

Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. 2000. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* **10:** 577–586.

Shearman, L.P., Sriram, S., Weaver, D.R., Maywood, E., Chaves, I., Zheng, B., Kume, K., Lee, C.C., van der Horst, G.T.J., Hastings, M., et al. 2000. Interacting molecular loops in the mammalian circadian clock. *Science* **288:** 1013–1019.

Steeves, T.D., King, D.P., Zhao, Y., Sangoram, A.M., Du, F., Bowcock, A.M., Moore, R.Y., and Takahashi, J.S. 1999. Molecular Cloning and Characterization of the Human CLOCK Gene: Expression in the Suprachiasmatic Nuclei. *Genomics* **57:** 189–200.

Strayer, C.A. and Kay, S.A. 1999. The ins and outs of circadian regulated gene expression. *Curr. Opin. Plant Biol.* **2:** 114–120.

Su, L.-K., Kinzler, K.W., Vogelstein, B., Preisinger, A.C., Moser, A.R., Luongo, C., Gould, K.A., and Dove, W.F. 1992. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256:** 668–670.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., et al. 1992. The *C. elegans* genome sequencing project: A beginning. *Nature* **356:** 37–41.

Sun, Z.S., Albrecht, U., Zhuchenko, O., Bailey, J., Eichele, G., and Lee, C.C. 1997. *RIGUI*, a putative mammalian ortholog of the Drospholia *period* gene. *Cell* **90:** 1003–1011.

Swiatek, P.J. and Gridley, T. 1993. Perinatal lethality and defects in hindbrain development in mice homozygous for a targeted mutation of the zinc finger gene Krox20. *Genes Dev.* **7:** 2071–2084.

Takeuchi, J., Shannon, W., Aronin, N., and Schwartz, W.J. 1993. Compositional changes of AP-1 DNA-binding proteins are regulated by light in a mammalian circadian clock. *Neuron* **11:** 825–836.

Vitaterna, M.H., King, D.P., Chang, A.-M., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., and Takahashi, J.S. 1994. Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science* **264:** 719–725.

Weitzman, E.D., Czeisler, C.A., Coleman, R.M., Spielman, A.J., Zimmerman, J.C., Dement, W., Richardson, G., and Pollak, C.P. 1981. Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. *Arch. Gen. Psychiatry* **38:** 737–746.

Wilsbacher, L.D. and Takahashi, J.S. 1998. Circadian rhythms: Molecular basis of the clock. *Curr. Opin. Genet. Dev.* **8:** 595–602.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368:** 32–38.

Xu, Y., Mural, R., Shah, M., and Uberbacher, E. 1994. Recognizing exons in genomic sequence using GRAIL II. *Genet. Eng.* **16:** 241–253.

Yamaguchi, T., Kim, N.S., Sekine, S., Seino, H., Osaka, F., Yamao, F., and Kato, S. 1996. Cloning and expression of cDNA encoding a human ubiquitin-conjugating enzyme similar to the Drosophila *bendless* gene product. *J. Biochem.* **120:** 494–497.

Zhang, J. and Madden, T.L. 1997. PowerBLAST: A new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res.* **7:** 649–656.

WWW RESOURCES

dot.imgen.bcm.tmc.edu:9331/seq-search.html.
genome.wustl.edu/gsc/index.shtml.
genome.wustl.edu/Protocols/protocols.shtml.
nog.cse.psu.edu/pipmaker.
www.genome.washington.edu/UWGC/analysistools/repeatmask.htm.
www-genome.wi.mit.edu/.
www-seq.wi.mit.edu/protocols/M13SPRI.html.
www-seq.wi.mit.edu/protocols/PCRSPRI.html.