ARTICLE

Assessing the Functional Characteristics of Synonymous and Nonsynonymous Mutation Candidates by Use of Large DNA Constructs

A. M. Eeds, D. Mortlock, R. Wade-Martins, and M. L. Summar

As we identify more and more genetic changes, either through mutation studies or population screens, we need powerful tools to study their potential molecular effects. With these tools, we can begin to understand the contributions of genetic variations to the wide range of human phenotypes. We used our catalogue of molecular changes in patients with carbamyl phosphate synthetase I (CPSI) deficiency to develop such a system for use in eukaryotic cells. We developed the tools and methods for rapidly modifying bacterial artificial chromosomes (BACs) for eukaryotic episomal replication, marker expression, and selection and then applied this protocol to a BAC containing the entire *CPSI* gene. Although this *CPSI* BAC construct was suitable for studying nonsynonymous mutations, potential splicing defects, and promoter variations, our focus was on studying potential splicing and RNA-processing defects to validate this system. In this article, we describe the construction of this system and subsequently examine the mechanism of four putative splicing mutations in patients deficient in CPSI. Using this model, we also demonstrate the reversible role of nonsense-mediated decay in all four mutations, using small interfering RNA knockdown of hUPF2. Furthermore, we were able to locate cryptic splicing sites for the two intronic mutations. This BAC-based system permits expression studies in the absence of patient RNA or tissues with relevant gene expression and provides experimental flexibility not available in genomic DNA or plasmid constructs. Our splicing and RNA degradation data demonstrate the advantages of using whole-gene constructs to study the effects of sequence variation on gene expression and function.

Identifying genetic contributions to a disease phenotype requires an understanding of the molecular pathology of mutations. Genetic variants can alter protein function through amino acid substitutions and promoter function and sensitivity through alterations in binding sites. Additionally, changing intronic and exonic sequences may result in splicing or RNA-processing defects.[1,2] Such RNA-processing defects often trigger the nonsense-mediated decay (NMD) pathway, a surveillance mechanism that degrades mRNA transcripts encoding a premature termination codon (PTC).[3] It is estimated that NMD accounts for as many as 30% of disease alleles.[4,5] When molecular variation in a gene is under investigation, it is imperative but often difficult to determine possible functionality and biological relevance. Specifically, RNA-processing defects are often difficult to study without the development of elaborate artificial constructs.

We recently examined the impact of RNA-processing defects that trigger NMD in patients with carbamyl phosphate synthetase I deficiency (CPSID [MIM #237300]).[6] CPSID is an autosomal recessive disease characterized by hyperammonemia and is the result of defects in carbamyl phosphate synthetase I (CPSI), the enzyme required to catalyze the first and rate-determining step of the hepatic urea cycle. We found RNA-based physical evidence that ~40% of the novel mutations resulting in CPSID created RNA-processing defects that trigger NMD.[6] Unfortunately, the unavailability of cells that expressed the *CPSI* gene in both adequate quantity and quality hampered our study. Consequently, we were forced to eliminate almost 2/3 of the patients for whom only genomic DNA (gDNA) was available. Examining the functional mechanisms of pathogenic alleles in many diseases such as CPSID presents a formidable challenge due to the frequent lack of patient tissue for RNA analysis, as well as to the large size of the gene, which has 38 exons that span >120 kb of gDNA.[7] Developing the means to explore these functional mechanisms is absolutely crucial for determining both the potential pathologic effects of rare variations and the milder functional variation seen in common polymorphisms. This problem is not unique to the study of *CPSI* mutations; indeed, it will certainly warrant considerable attention as more genetic changes become associated with a variety of common disease states.

To facilitate functional genetics studies, we developed a versatile BAC-based model system to test the effects of a diverse set of intronic and exonic mutations in a *CPSI* whole-gene construct. Essential elements of this system included normal BAC replication in bacteria, episomal eukaryotic replication, antibiotic selection in eukaryotic cells, and the expression of green fluorescent protein (GFP). Each element was incorporated into a vector that

From the Program in Translational Genetics, Center for Human Genetic Research (A.M.E.; D.M.; M.L.S.), and Department of Pediatrics (M.L.S.), Vanderbilt University Medical Center, Nashville, TN; and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom (R.W.-M.)

Received October 24, 2006; accepted for publication January 30, 2007; electronically published March 8, 2007.

Address for correspondence and reprints: Dr. Marshall Summar, Program in Translational Genetics, Center for Human Genetic Research, Vanderbilt University Medical Center, 1175 Medical Research Building IV, Nashville, TN 37232-0007. E-mail: marshall.summar@vanderbilt.edu

Am. J. Hum. Genet. 2007;80:740–750. 0002-9297/2007/8004-0015$15.00
DOI: 10.1086/513287

is easily exported to other BAC constructs. This BAC-based expression system, coupled with other recent BAC engineering and mutagenesis improvements,[8,9] provides an efficient way to test genetic variants, irrespective of type, location, or gene size. This expression system is widely applicable to many experiments that examine both coding and noncoding sequences, and it is particularly useful for assaying mutations that affect RNA processing because of the large size of genomic inserts allowed by BACs. Additionally, studying mutations within their wider sequence context should enable us to determine their mechanistic effect with greater accuracy. Here, we describe the construction and testing of this system and demonstrate its utility as an effective eukaryotic cell model in the study of the role of genetic variations with both known and unknown molecular effects identified in patients with CPSID.

Material and Methods

Plasmid and BAC Engineering

We inserted the pEHG plasmid into the backbone of BAC RP11-349G4, using Cre/*lox*P recombination, as described elsewhere.[10] To provide ubiquitous expression of the normally hepatically expressed CPSI in our test cells (MRC-5V2 cells),[11] we inserted the human cytomegalovirus (CMV) immediate-early promoter (Towne strain) upstream of *CPSI*. We modified pCMV/Bsd (Invitrogen) by inserting a *Kpn*I site at position 1937 by site-directed mutagenesis (Stratagene QuickChange). Then, *Kpn*I and *Xho*I were used to create a complementary insertion site for an oligo containing the tetracycline resistance cassette flanked by FRT sites from plasmid pTet/FRT.[12] A 5′ homology arm (Invitrogen) containing 50 bp of sequence homologous to the BAC backbone was then synthesized with *Not*I and *Nhe*I overhangs to direct insertion of the homology arm immediately 5′ of the Tet/FRT sequence. A 3′ homology arm (Invitrogen) containing 50 bp of sequence homologous to the *CPSI* gene was synthesized, beginning with the transcription start site of *CPSI* (AF154830), with *Rsr*II and *Xma*I overhangs to direct insertion of the homology arm downstream of the CMV promoter. The BAC+pEHG construct was transformed into EL250 *Escherichia coli* cells, and a monoclonal colony was induced for recombination by a temperature shift from 32°C to 42°C before electroporation of the constructed CMV vector.[13] Recombinants were selected with tetracycline, after which the tetracycline cassette was removed by arabinose induction of Flp recombinase. Site-directed mutagenesis to introduce point mutations into the wild-type (wt) BECC (BAC+pEHG+CMV+CPSI) construct was performed using double homologous recombination with *galK* selection after transfection of the wtBECC construct into SW102 cells.[9] The first homologous recombination step is to add the *galK* gene at the position of the desired mutation, with the use of positive selection for colonies able to metabolize galactose. The second homologous recombination step is to replace the galK cassette with the desired mutation sequence, with the use of negative selection against colonies that retain galK and therefore toxically metabolize 2-deoxygalactose, which produces a toxin in the presence of galK expression. All restriction enzymes were purchased from New England Biolabs. BAC purifications were performed using the NucleoBond BAC Maxi AX500 kit (BDBiosciences).

Verification of BAC Engineering

Pulse-field gel electrophoresis (PFGE) and fingerprinting revealed no unwanted or global rearrangements of the vector during the retrofitting or homologous recombination steps. PFGE was performed after a 3-h digestion of BAC DNA with *Not*I (New England Biolabs) in a 1% agarose 0.5% tris-acetate EDTA gel run at 6 V for 16 h, with an initial switching time of 0.2 s ramped to a final switching time of 22 s. Fingerprinting was performed by digesting BAC DNA with *Bam*HI (New England Biolabs) for 3 h and then running the samples on a 1% agarose 1% tris-borate EDTA gel for 16 h at 35 V. Direct sequencing was also performed to verify proper insertion of the CMV promoter upstream of *CPSI*. Single-strand conformation polymorphism (SSCP) analysis for mutation detection in the coding region of *CSPI* revealed that no unwanted rearrangements or mutations occurred in the coding region of the gene during construct manipulations with the use of a previously published protocol and primer pairs.[7,14] All exons, as well as the 5′ and 3′ UTRs, were screened in a total of 42 PCRs. To increase mutation detection rates, direct sequencing was performed instead of SSCP for 12 of the 42 PCRs. In addition, changes between the BECC construct and a control DNA sample detected by SSCP were analyzed by direct sequencing, and all changes were shown to be different alleles of polymorphisms reported elsewhere.[7]

Sequencing

Fluorescent sequencing was performed by GenHunter Corporation and the Vanderbilt University Medical Center Core Facility, with use of BigDye chemistry from Applied Biosystems. We performed radioactive dideoxynucleotide sequencing, using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (USB) in accordance with the manufacturer's protocol, with 200 ng DNA.

Transfections and Stable Cell-Line Generation

Transfections were performed using the LID (Lipofectin:Integrin:DNA) technique[8] with peptide 6. We used 8 μl lipofectin (Invitrogen), 8 μg peptide 6, and 4 μg of each BAC construct in a 2L:2I:1D ratio. Lipofectin was diluted in 400 μl Optimem (Gibco) and was incubated 45 min before the addition of peptide 6 and DNA, which was also diluted in 400 μl Optimem. Reactions were incubated for 10 min, during which time all cells were washed with 3 ml Optimem (no serum). All transfection reactions were increased to a 4-ml total volume with Optimem before dropwise placement on 5-cm dishes containing cells at 60% confluence. After a 16-h incubation, the transfection media were replaced with Dulbecco's modified Eagle medium (DMEM) and 10% fetal bovine serum for 72 h. After 72 h, hygromycin selection was added at a concentration of 125 μg/ml. After 2 wk under hygromycin selection, cells were flow-sorted for GFP expression. All GFP-positive cells were cultured under hygromycin selection as polyclonal cell lines.

Western Blotting

Cells were lysed by sonication and were quantified using the BCA Protein Estimation Kit (Pierce). After a 5-min denaturation step, equal microgram quantities of cell lysates were electrophoresed in a 6% polyacrylamide gel for 45 min at 200 V. Proteins were transferred to a nitrocellulose membrane. The anti-CPSI (product

number sc-10516 [Santa Cruz]) primary antibody was incubated overnight at 4°C at a final dilution of 1:50. The bovine antigoat immunoglobulin G (IgG)–alkaline phosphatase (AP) (product number sc-2351 [Santa Cruz]) secondary antibody was incubated (1:2,000 dilution) at room temperature before alkaline phosphate staining (Bio-Rad).

gDNA Isolation and PCR

We used the Wizard Genomic DNA Purification Kit (Promega) to isolate DNA from all cell lines. To determine the presence of BAC DNA in transfected cell lines, we performed PCR with an upper primer complementary to the CMV promoter—UCMV (5′-CCATCCACGCTGTTTTGACCTC-3′)—and with a lower primer complementary to the sequence in the first exon of *CPSI*—L173 (5′-CCAGTCTTCAGTGTCCTCA-3′). The amplification thermoprofile was an initial denaturation step of 4 min at 95°C, 40 cycles of 30 s at 95°C, 30 s at 67°C, and 30 s at 72°C, and a final hold at 72°C for 10 min.

RNA Isolation, RT-PCR, and Northern Blotting

To isolate RNA, we used the RNeasy Midi Kit, including the optional RNase-free DNase set, following the manufacturer's protocol (Qiagen). Reverse transcriptions were performed using 2 μg or 5 μg of total RNA and the TaqMan Reverse Transcription Reagents Kit (Applied Biosystems). The amplification thermoprofile was an initial denaturation at 65°C for 5 min, then a primer-annealing phase at 27°C for 10 min, an extension phase at 42°C for 45 min, and an enzyme denaturation step at 95°C for 5 min. We performed PCR, using the cDNA templates produced from 2 μg RNA to amplify desired regions of *CPSI,* for the determination of splicing changes from the intronic mutations. After visualization on a 2% agarose gel, bands were purified using the Wizard PCR Cleanup Kit (Invitrogen) and then were sequenced. PCR primers were designed at least one exon away from the mutation in question, to eliminate gDNA contamination. We amplified cDNA, using UBacT344A (5′-CTGCTCAGAATCATGACC-3′) and L1361 (5′-GGCTTCGGTAAGACTGATGT-3′) for c.1210-1G→T and U393 (5′-AGGACAGATTCTCACAATGG-3′) and L829 (5′-CTAGCAGGCGGATTACATTG-3′) for c.652-3T→G. For northern blotting, 10 μg of total RNA was used in a 1.2% agarose formaldehyde gel that was run for 2.5 h at 3 V/cm. RNA was transferred to a Hybond N+ membrane by use of the Turboblotter apparatus (Schleicher and Schuell) and was UV crosslinked. The membrane was prehybridized in Church buffer (0.25 M Na_2HPO_4, 7% SDS, and 50 μg/ml sheared salmon sperm DNA) at 65°C for 1 h, and hybridization occurred overnight at 65°C with either a CPSI or cyclophilin probe that was previously radiolabeled with the Prime-It RmT Random Primer Labeling Kit (Stratagene) and diluted to 1 million counts per min.

Quantitative PCR

We calculated relative RNA expression levels, using quantitative PCR with cDNA templates synthesized from 5 μg RNA. We performed these reactions using TaqMan (Applied Biosystems) technology on an automated platform with the ABI PRISM Detection System. To measure *CPSI* expression, we used the Hs00919484_m1 assay that spans the exon 3–exon 4 junction and the Hs00919480_m1 assay that spans the exon 34–exon 35 junction. We used an E-GFP probe as the endogenous control (part number 4331348 [Custom Taqman Gene Expression Assay Service]). All experiments were performed in triplicate. "No RT" samples were used as a PCR control measure for gDNA contamination. We performed ΔΔCt analysis, a description of relative RNA expression levels, by first calculating the difference between the average cycle time (Ct) value of each target *CPSI* sample and the average Ct value of the corresponding endogenous *GFP* sample at the 0.200 fluorescence threshold. Standardizing the *CPSI* Ct values to the *GFP* transcripts serves as a control, since GFP expression remains constant, and can also account for variation in BAC copy number, since both genes are present in a 1:1 ratio and are driven by a CMV promoter. These calculations were then expressed in relation to the calibrator (wild-type cell line in fig. 5*A* and untreated cell lines in fig. 6), which was arbitrarily set at an expression level of 100% (fig. 5*A*), or onefold (fig. 6).

NMD Inhibition and Western Blotting

We used small interfering RNA (siRNA) to knock down UPF2 and to therefore block NMD, as performed elsewhere.[15] Cells were 75% confluent in a 10-cm dish and were treated for 3 consecutive d. SiQuest transfection reagent was used according to the manufacturer's protocol, at a final concentration of 5 μl/ml of DMEM (Mirus). siRNAs were used at a final concentration of 70 nM (Dharmacon). The siRNA sequence for UPF2 was published elsewhere,[15] and the commercially available siCONTROL RISC-Free kit (Dharmacon) used to show knockdown was specific for UPF2, since this siRNA sequence does not target any known human genes. Western blotting was performed using anti-hUPF2, anti-α-tubulin (Abcam product number ab15246), and anti-CPSI (product number sc-10516 [Santa Cruz]) primary antibodies, as well as the electrochemiluminescent rabbit IgG (Amersham NA934) and bovine anti-goat IgG-AP (product number sc-2351 [Santa Cruz]) secondary antibodies. Western blot densitometry quantification was performed on blots repeated in triplicate, with use of the QuantityOne software (BioRad).

Results

Patient Mutations

Here, we have identified the mutation mechanisms of two exonic and two intronic mutations originally identified by genomic mutation screens of patients with CPSID.[6,16] The c.1893T→G substitution is a nonsense mutation located in exon 16, and the c.2388C→A mutation is a synonymous change in exon 19. Initially, we did not believe that this silent mutation was pathogenic. However, we were unable to identify another mutation on this patient allele. The two intronic substitutions, c.652-3T→G in intron 5 and c.1210-1G→T in intron 10, are both located at the 3′ splice acceptor site. Each mutation was unique to the patient with disease, since no others were detected in a screen of >200 additional chromosomes.

Construction and Verification of the BAC Model System

To study these and other *CPSI* mutations in a full-gene context, we modified a BAC construct to create a platform for mutation testing. Figure 1*A* shows the construction of this BECC vector. First, we used Cre/*lox*P recombination to retrofit BAC RP11-349G4 (containing the full *CPSI* gene) with vector pEHG. This vector contains episomal reten-

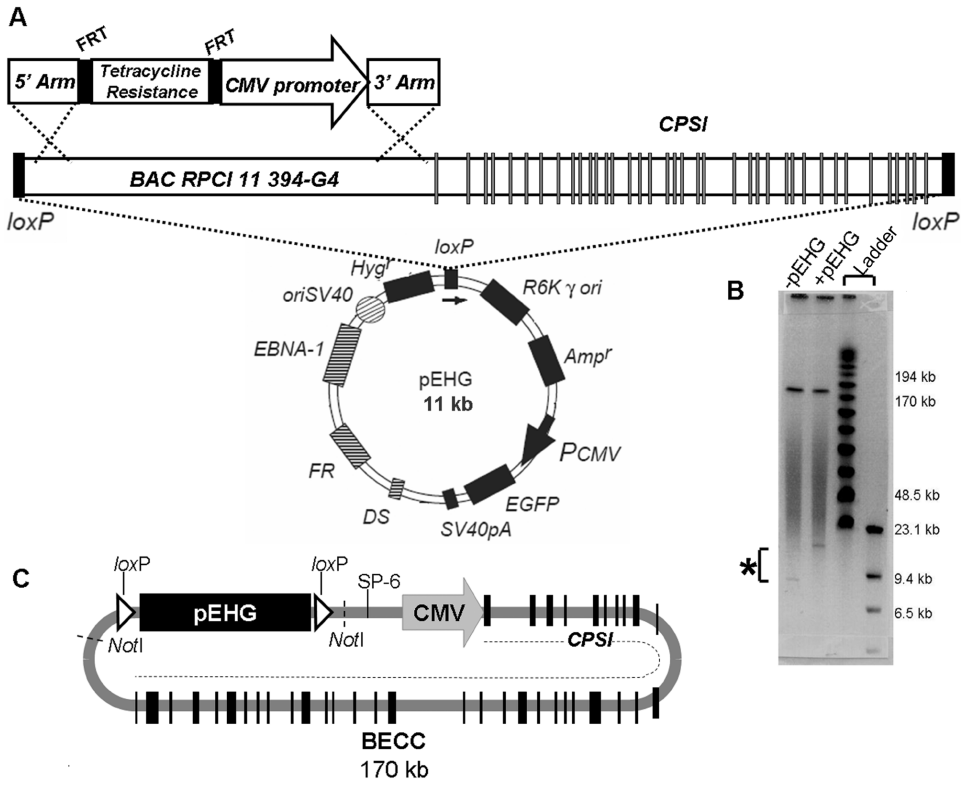

Figure 1. BECC model system. *A,* Schematic representation of the modifications to the RPCI11-349G4 BAC construct. The homologous recombination vector was used to simultaneously insert the CMV promoter upstream of *CPSI* and to delete the genomic sequence upstream of *CPSI*. The tetracycline resistance gene flanked by FRT sites was used for positive selection of the recombination event but was then removed from the final BAC construct. Use of the *lox*P sites on the pEHG vector and the BAC allowed sequence-specific integration of these two vectors. *B,* PFGE of *Not*I-digested constructs before (−) and after (+) the addition of a single copy of pEHG after Cre/*lox*P recombineering in the BAC. Asterisk (*) indicates the 11-kb size change in the BAC library vector corresponding to pEHG. *C,* Schematic of the 170-kb BECC vector after both modifications.

tion elements from Epstein-Barr virus (EBV)—specifically, the latent origin of replication (OriP) and the EBV nuclear antigen 1 (EBNA-1)—as well as the hygromycin resistance antibiotic marker and a constitutive GFP reporter (E-GFP) driven by the CMV promoter.[17,18] Figure 1*B* shows a PFGE illustrating the insertion of one copy of this 11-kb vector. Second, homologous recombination allowed simultaneous deletion of the sequence 5′ of *CPSI* and insertion of the CMV promoter directly upstream of the first exon, thereby producing constitutive expression of this hepatic-specific gene. The CMV promoter was inserted using a Tet-CMV targeting construct flanked by BAC homology arms (fig. 1*A*). Multiple assays verified the integrity of this construct (see the "Materials and Methods" section). The completed wild-type construct (wtBECC) is 170 kb in size, of which 150 kb is gDNA from chromosome 2 containing the complete *CPSI* gene and 11 kb is from pEHG (fig. 1*C*).

We next performed site-directed mutagenesis to introduce the c.652-3T→G, c.1210-1G→T, c.1893T→G, and c.2388C→A mutations into the BECC vector. These mutations were introduced using two-step homologous recombination with positive and negative *galK* selection.[9] Various control assays, including PFGE and sequencing, verified that only the desired point mutations were incorporated (data not shown).

We then created polyclonal, stable cell lines for each of the BECC constructs and examined expression of the exogenous genes in cultured cells. After transfection into the MRC-5V2 immortalized human lung fibroblast cell line,[8,11] we obtained initial 20% transfection efficiencies as determined by visual inspection of GFP-expressing cells. GFP-positive cells demonstrating stable transfection were selected by collection through flow-sorting followed by continuous culture under hygromycin selection. The MRC+wt*CPSI* polyclonal cell line was passaged >85 times during a 1.5-year period under continuous hygromycin selection with no visual loss of GFP expression. To illustrate the presence of BAC-derived *CPSI* DNA in each cell line, we specifically amplified exogenous *CPSI* DNA, using the BAC-specific CMV promoter sequence and *CPSI* exon 1 sequence (fig. 2*A*). PCR revealed the presence of BAC-derived *CPSI* DNA only in the transfected cell lines (fig. 2*B*). Furthermore, we performed western blotting to verify the presence of exogenous CPSI protein specifically in the

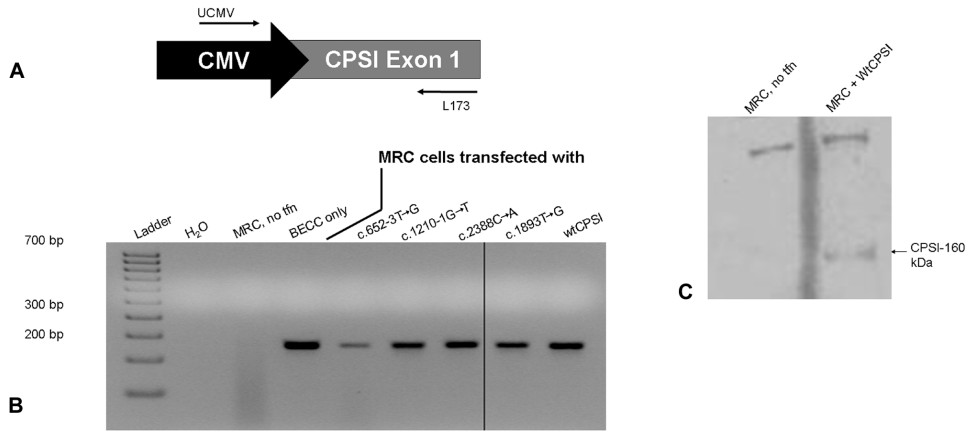

Figure 2. PCR demonstrating exogenous *CPSI* DNA, specifically in transfected cells. *A,* Representation of primer locations for identifying the presence of exogenous *CPSI* DNA. The forward primer is homologous to the CMV promoter and is specific for BAC-derived CPSI. *B,* PCR on a 2% agarose gel, demonstrating the presence of exogenous *CPSI* gDNA only in the MRC cell lines transfected with the indicated BAC construct. no tfn = No transfection. *C,* Western blot that uses a CPSI-specific antibody, showing the presence of CPSI protein specifically in the MRC cells transfected with the wtCPSI construct.

MRC+wtBECC cell line. Our results showed that, after transfection, not only did the BECC constructs create stable transformants that expressed GFP and were resistant to hygromycin, but the wild-type construct also produced CPSI (fig. 2*C*).

RT-PCR and Sequencing of Splicing Mutations

To determine the effect of both intronic mutations on splicing, we performed RT-PCR on RNA isolated from the c.652-3T→G, c.1210-1G→T, and wild-type transfected cell lines. Because the reverse transcription protocol that we used was previously proved sensitive enough to pick up low-level *CPSI* transcripts from a nonhepatic cell line,[16] it was important to distinguish between BAC-derived and low-level endogenous MRC-5V2 *CPSI* transcripts.

c.652-3T→G.—We first illustrate that the BECC model system mimics the original patient sequence data for c.652-3T→G. Whereas patient gDNA showed a heterozygous mutation at the −3 position of intron 5 (fig. 3*A*), patient cDNA (from a fibroblast cell line) showed a heterozygous 2-bp frameshift beginning in exon 6, indicating that the mutation activated an aberrant splice site (fig. 3*B* and 3*E*). To verify that the BECC platform reproduced what we observed directly from the patient, we performed RT-PCR on the MRC+wt*CPSI* and MRC+c.652-3T→G cell lines. Because a size difference indicating a splicing change from the c.652-3T→G mutation was not detectable on an agarose gel (fig. 3*C*), the RT-PCR products were radioactively sequenced to visualize the presence of aberrant transcript and to separate endogenous and exogenous transcripts. The sequence from the MRC+c.652-3T→G cell line mimicked the patient data by revealing the same 2-bp AG dinucleotide splicing alteration (fig. 3*D* and 3*E*). This comparison of data gathered from a patient cell line and from the BECC-transfected cell lines validated the use of BECC for studies on additional mutations, including those that affect RNA processing.

c.1210-1G→T.—To aid RT-PCR analysis of the c.1210-1G→T mutation and to distinguish between exogenous and endogenous *CPSI* transcripts, we were able to exploit the presence of a nearby 3-bp polymorphism with alternate alleles present in MRC-5V2–derived and BAC-derived *CPSI*. This polymorphism, T344A, corresponds to either nucleic ACC or GCT at position c.1136–1138.[7] The forward primer used in these PCRs contained homology specific to the BAC allele at the 3′ end, preventing amplification of any endogenous *CPSI* transcripts. Comparison of RT-PCR products from the MRC+wt*CPSI* and MRC+c.1210-1G→T cell lines showed a difference in product size, with a larger band from the MRC+c.1210-1G→T cell line, indicating a splicing change (fig. 4*A*). Subsequent sequencing revealed the inclusion of 32 bp from the 3′ end of intron 10 in the spliced transcript (fig. 4*B*).

Quantitative RT-PCR

To examine the effect of each mutation on RNA stability, we next performed ΔΔCt analysis from quantitative RT-PCR (qRT-PCR) assays, to determine the relative expression levels of *CPSI* transcript in each transfected cell line. These ΔΔCt calculations describe *CPSI* transcript levels that have been standardized to *GFP* transcript levels (both are present in a 1:1 ratio, and both are expressed from a CMV promoter) and then expressed in relation to the wild-type "calibrator" sample. We used three probes to measure either *CPSI* or *GFP* transcripts in each transfected cell line. One CPSI probe annealed to the transcript at the exon 3–exon 4 junction, upstream of all tested mutations. The other spanned the exon 34–exon 35 junction, down-

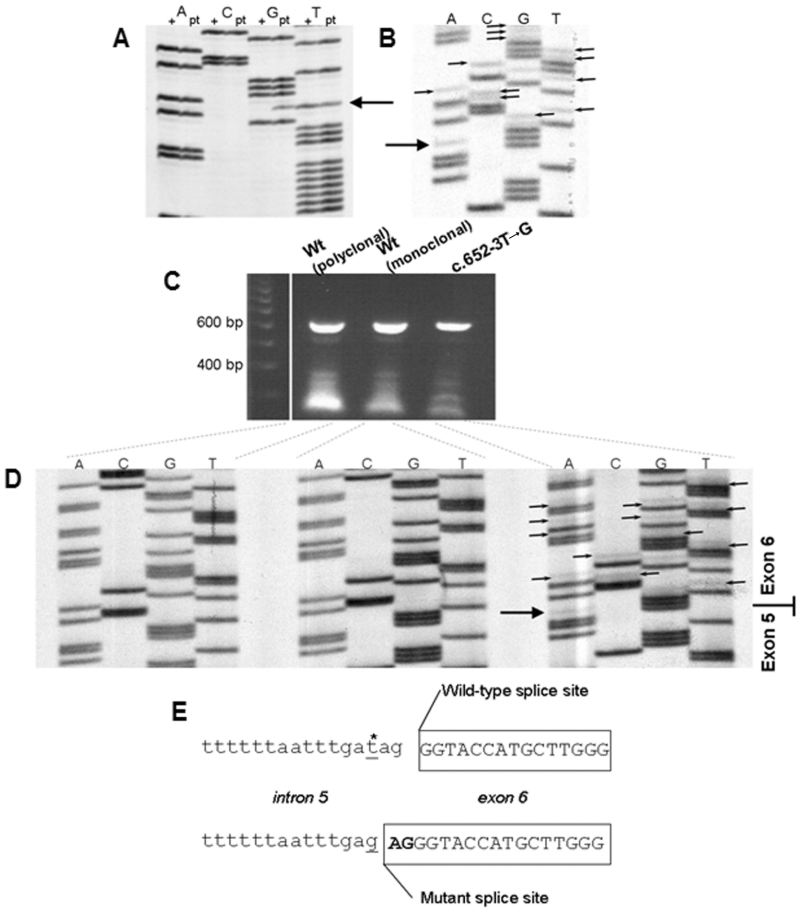

Figure 3. RT-PCR products from c.652-3T→G. *A,* Patient gDNA sequence revealing a G→T point mutation in intron 5 (*arrow*). *B,* Patient cDNA revealing a 2-bp insertion in exon 6 from one allele. The large arrow indicates the start of the insertion, and the small arrows indicate other residues in the frameshift. *C,* RT-PCRs from the BECC platform. Polyclonal wtCPSI, monoclonal wtCPSI, and c.652-3T→G polyclonal cell lines were tested on a 2% agarose gel. *D,* Dideoxyradionucleotide sequence of the RT-PCR products shown in panel A reveals the presence of a 2-bp frameshift specifically in the mutant cell line, which mimics the original patient data. The large arrow indicates the start of the frameshift, and the small arrows indicate other residues in the frameshift. *E,* Schematic of the splice site change caused by the T→G mutation (*underlined nucleotides*). The wild-type sequence is identified by an asterisk (*), the intronic sequence is in lowercase letters, and the exonic sequence is boxed and in uppercase letters.

stream of all tested mutations. ΔΔCt analysis revealed that, for each mutation, *CPSI* expression levels were much lower than wild type, except for c.652-3T→G measured by the exon 3–exon 4 probe (fig. 5*A*, black bars). The relative expression differences between each mutant and each wild-type sample, with use of the CPSI exon 3–exon 4 probe, varied significantly. However, all mutants demonstrated a similar drop in relative expression levels with use of the CPSI exon 34–exon 35 probe (fig. 5*A*, white bars). A northern blot corroborated these data by showing that exogenous *CPSI* was present in both the wild-type polyclonal and monoclonal cell lines but not in any of the mutants (fig. 5*B*). Taken together, these data suggest that each mutation causes a decrease in *CPSI* RNA levels. It also suggests that degradation of RNA by this mechanism may slant toward the 3′ end of the transcript.

Inhibition of NMD

To determine if the decreased *CPSI* expression was due to degradation via NMD, we used siRNA to inhibit UPF2, which has been shown elsewhere to be sufficient to prevent NMD.[15,19,20] We demonstrated the siRNA-mediated specific knockdown of UPF2 in each transfected cell line by a western blot analyzed with densitometry; in all cases, the UPF2/α-tubulin ratio is considerably lower in the UPF2 siRNA-treated cells than in either untreated cells or those treated with vehicle alone or a nonspecific siRNA control.

qRT-PCR after siRNA-mediated inhibition of NMD showed increased levels of *CPSI* transcript when compared with vehicle-only and untreated controls (fig. 6). All data were again expressed as a ΔΔCt, representing a fold increase in expression when compared with the calibrator (untreated) sample set at an expression value of 1. After

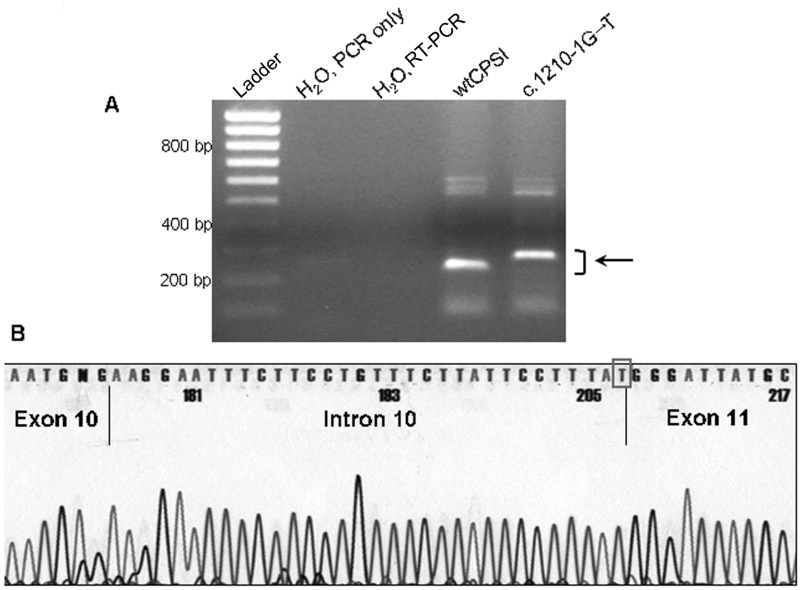

Figure 4. RT-PCR products from c.1210-1G→T. *A,* RT-PCR products from a polyclonal wtCPSI and c.1210-1G→T polyclonal cell line on a 2% agarose gel. The size difference between the wild-type and mutant products indicates activation of a cryptic splice site in the mutant cell line. *B,* Fluorescent sequence of the RT-PCR products shown in panel A reveals the insertion of 32 bp of intron 10 into the spliced message specifically in the mutant cell line. The box highlights the location of the c.1210-1G→T mutation.

UPF2 siRNA treatment, the *CPSI* fold increase observed in every mutant cell line was greater (6-fold to 230-fold) than the *CPSI* fold increase in the wild-type cell line (3-fold). These results indicate that the *CPSI* mutations cause RNA instability—at least partially—by eliciting the NMD pathway, since knockdown of UPF2 increases the relative expression levels of each mutant *CPSI* transcript when compared with untreated cells. Although the relative increase varied depending on the cell line and probe used (possibly because of varying degrees of siRNA-mediated NMD inhibition or unique properties of each mutation, or both), these results strongly indicate a critical role for NMD in the degradation of these *CPSI* mutant transcripts.

Discussion

Understanding the effects of genetic variants associated with clinical phenotypes is a difficult yet vital undertaking. To this end, we created a novel system as a platform for determining the molecular function of any genomic change in *CPSI* and subsequently used this model to determine the functional significance of identified exonic and intronic mutations. To a BAC clone, we added a modular insert to create a eukaryotic expression vector, modified *CPSI* for ubiquitous expression under the control of the CMV promoter, and introduced specific mutations. In this article, we show—to our knowledge, for the first time—the effect of each of these mutations on the CPSI transcript, by demonstrating that each mutation causes transcript instability and degradation through NMD. In addition, our novel studies show the activation of aberrant splice sites for the two intronic mutations tested.

The BECC model system, coupled with other recent BAC engineering improvements,[8,9] provides an efficient way to test a heterogeneous set of mutations from any gene in eukaryotic cells irrespective of mutation type, location, or gene size. With a single modification, any BAC clone can be retrofitted with pEHG, which contains the necessary elements for long-term retention, selection, and construct tracking in mammalian culture cells.[17,18] The EBV elements located on pEHG, OriP, and *EBNA-1* have been shown elsewhere to be the only components necessary for the episomal replication of constructs containing them, and they have been successfully exploited in the construction of other episomally replicating eukaryotic vectors.[17,21–24]

This model successfully overcomes limitations of classical methods for assessing mutation function. It was important to test mutations in the physiological context of the complete *CPSI* gene on a single construct, but the large size of *CPSI* prevented the use of conventional methods, which have significant size restrictions. In addition, because many *CPSI* mutations are intronic or are suspected to affect splicing, the inclusion of introns was imperative.[7] "Minigene" construction, although it allows intron inclusion, is not feasible for large genes such as *CPSI*. Other means for transferring and expressing large segments of exogenous DNA in eukaryotic cells have been developed, such as the use of BACs with viral delivery[25,26] or bacterial transfer,[27] yeast artificial chromosomes,[28] and human artificial chromosomes.[29] However, we desired to design a

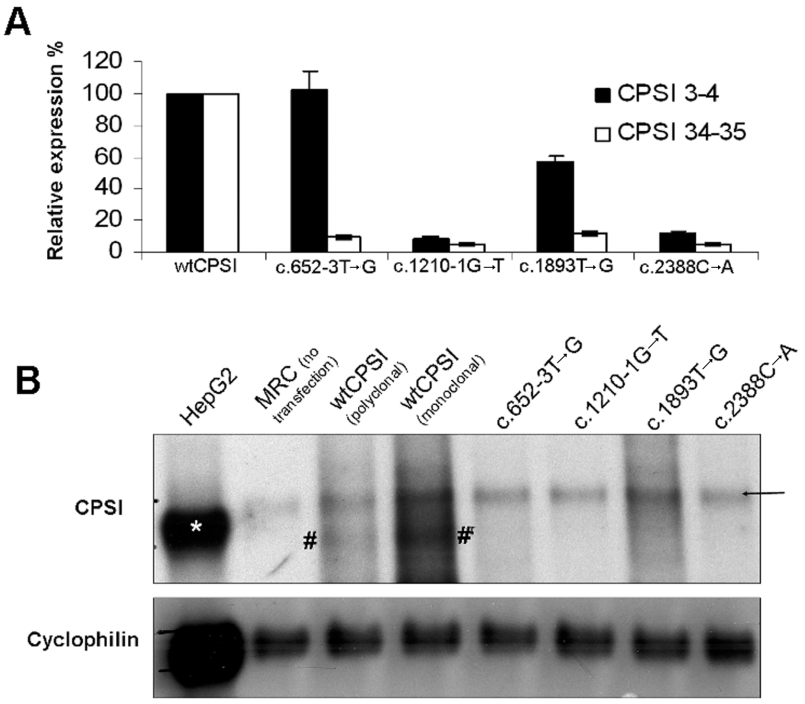

Figure 5. ΔΔCt analysis showing relative CPSI RNA expression levels after qRT-PCR. *A,* Expression levels of CPSI RNA in each mutant cell line relative to wild type, set at 100% expression. Black bars indicate relative CPSI RNA levels measured from the exon 3–exon 4 TaqMan probe, and white bars indicate relative CPSI RNA levels measured from the exon 34–exon 35 TaqMan probe. SE is represented. *B,* Northern blot of RNA from each transfected cell line, showing the same decrease in RNA levels in each mutant cell line as in panel A. Only the wtCPSI cell lines have a visible band of exogenous, BAC-derived CPSI (between the number [#] signs). Only the HepG2 hepatic cell line shows endogenous CPSI expression (marked by an asterisk [*]). An arrow indicates a nonspecific band detected by the CPSI probe. A cyclophilin probe was used as a loading control. The BAC-derived CPSI transcript has an expected difference in size from endogenous CPSI (shown in the HepG2 lane) because of CMV promoter swapping that altered the size of the 5′ UTR by ~100 bp.

nonviral, versatile expression system that requires only the addition of an 11-kb vector, to ensure stable transfections, and that is capable of undergoing precise manipulations through homologous recombination. The BAC-based model system is limited only by the identification of a BAC containing the complete gene of interest and an appropriate cell line for vector expression. However, given the recent improvements in BAC engineering, it is entirely feasible to modify a BAC clone to meet size specifications or to join gene segments from multiple BACs together.[30] Furthermore, various transfection methods can be used to introduce large constructs in a spectrum of cell types.[8,31,32] As shown here, promoter swapping permits gene expression in an otherwise nonexpressing cell type.

We used the BECC vector to examine the mechanistic effect of four *CPSI* mutations—c.652-3T→G, c.1201-1G→T, c.1893T→G, and c.2388C→A—that were each thought to affect transcript stability. Introducing each mutation in the BECC platform had several advantages beyond validating the model system. First, since patient samples are usually obtained from individuals who are heterozygous for different mutations, we were able to isolate the mutation in question for study in an experimental system. Second, these mutations could be maintained in a system without the limited shelf-life typically seen for patient tissue samples, which may be difficult to obtain. Third, we were able to perform follow-up manipulations that are not always possible or reliable in hepatic samples or patient cell lines.

As shown in figure 5, the qRT-PCR probe specific for the 3′ end of the transcript shows relatively equal large decreases in detectable *CPSI* message in all mutant cell lines when compared with wild type. In all cases, the 5′ CPSI probe detects higher levels of *CPSI* transcript, but these levels are all less than those in wild type except c.652-3T→G. In each case, the observation that more mutant transcript was detected with the CPSI exon 3–exon 4 probe, which lies upstream of all tested mutations, may indicate degradation primarily from the 3′ end. Although it is possible that there are alternate splicing or polyadenylation isoforms that recognize the exon 3–exon 4 but not the exon 34–exon 35 probe, no such isoforms have been described for *CPSI*. Ongoing and extensive testing performed in our laboratory with a variety of techniques has been unable to identify alternate isoforms of *CPSI* that exclude the exon 34–exon 35 junction. Our qRT-PCR data, coupled with a survey of all mutations known to create PTCs in patients with CPSID,[6] support the hypothesis that

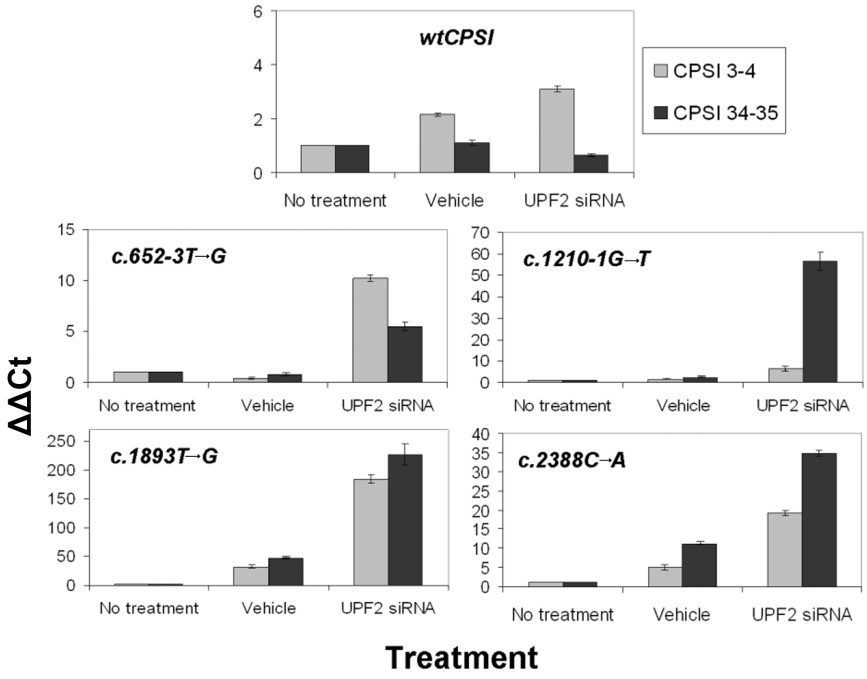

Figure 6. ΔΔCt analysis of qRT-PCR data after siRNA-mediated knockdown of UPF2. Each panel graphs the relative expression levels of CPSI transcripts after UPF2 siRNA or vehicle-only treatments in the respective cell line as compared with untreated cells (set at a value of 1). Gray bars indicate relative levels measured from the exon 3–exon 4 TaqMan probe, and black bars indicate relative levels measured from the exon 34–exon 35 TaqMan probe. Error bars represent SE.

a significant number of pathogenic *CPSID* mutations result in RNA-processing defects.

Specific Mutation Observations

Both intronic mutations, c.652-3T→G and c.1210-1G→T, affect the splice acceptor site at the 3′ end of the intron by disrupting the 3′ splice site consensus $(Y)_nNYAG|G$ sequence.[33] For both mutations, sequencing of RT-PCR products revealed the recognition of an alternate AG dinucleotide as the 3′ splice site. The c.652-3T→G mutation creates a premature AG dinucleotide sequence that was recognized as the 3′ end of the intron in both the patient and the transfected MRC5-V2 cell lines (fig. 3). The c.1210-1G→T mutation abolished the endogenous AG dinucleotide sequence at the 3′ end of the intron, causing use of an upstream AG (fig. 4). Both mutations result in a splicing change that causes a frameshift, thereby illustrating the importance of the AG dinucleotide in proper 3′ intron definition.

The c.1893T→G mutation creates a TAG stop codon in place of tyrosine at amino acid position 590. This mutation provides the most direct measure of the presence of NMD, since the stop codon is formed independently of any cryptic splicing events. When NMD was inhibited through siRNA-mediated knockdown of UPF2, the transfected cell line containing this mutant construct had the largest fold increase in *CPSI* transcript (fig. 6). A possible explanation for this observation is that there are separable mechanisms for eliciting NMD based on the type of mutation, since no frameshift occurred to create this PTC. It is also possible that various cryptic splice sites are recognized in transcripts with intronic mutations which, although undetectable by RT-PCR, alter not only the number of transcripts containing PTCs but also the position of the PTC, which may have an effect on recognition by the NMD pathway.

Finally, the BECC system was particularly useful for determining the mechanistic effect of the c.2388C→A mutation. This synonymous change (S755S) in exon 19 was not originally thought to cause loss of function. However, after a thorough genotyping of *CPSI* in this patient, no other genomic changes could be found on this allele. This result, coupled with the knowledge that many missense and silent mutations are pathogenic at the RNA level if they reside in an exonic splicing enhancer,[34,35] leads us to propose that c.2388C→A either disrupts a splicing enhancer or creates a splicing silencer binding motif. As shown in figures 5 and 6, this mutation resulted in a large drop in transcript that was restored by NMD inhibition. For this silent mutation to be sensitive to NMD, it must create a PTC, which would be achieved if exon 19 were skipped or if a cryptic splice site were activated that caused a frameshift due to partial exon inclusion. These data add to the increasing evidence that, although they do not alter actual protein structure, many mutations, such as silent and missense mutations, negatively affect the processing

of the transcript.[35] This particular mutation underscores the utility of the BECC system for testing the functional effects of seemingly innocuous genetic variants. This system can be used to test elements thought to be important in processing exon splice enhancers.

Summary

Our data indicate that RNA-processing defects must always be considered when determining the functional significance of genetic changes because of the numerous ways this pathway can be engaged at the sequence level. Accounting for them in patient mutation screens is imperative yet difficult, because patient RNA samples are not always available or able to be manipulated for the detection of splicing changes and allelic dropout from NMD. To facilitate such studies, we report in this article the creation of a novel model system to determine the mechanism of mutations in *CPSI* that can be applied to other genes as well. Whereas this model system can be employed to test any sequence change, including those in regulatory sequences, its size capabilities make it particularly useful for assaying mutations that affect RNA processing. This system not only permits expression studies in the absence of patient samples but also allows experimental manipulations that could not otherwise be performed, such as inhibiting NMD to demonstrate its role in *CPSI* mutation pathology. We will continue to use this model to examine effects of other mutations in *CPSI,* as well as to determine possible functional effects of identified polymorphisms, since this BAC-based platform will surely prove helpful as we strive to unravel the many mysteries of functional genetic changes.

Acknowledgments

We thank L. Hall, A. Putnam, R. Price, Natasha Strauss, the Vanderbilt University Medical Center DNA Resources Core, Sequencing Core, and GenHunter Corporation, for invaluable technical assistance. We thank J. Patton, who, through collaboration with J. Steitz, shared the hUPF2 antibody; S. Hart, for sharing peptide 6; and N. Copeland, for the SW102 and EL250 *E. coli* strains. We also thank J. Shelton for assistance in manuscript preparation. We acknowledge the patients and their families for their consent. Partial funding for this project came from the Adam Scholarship. A.M.E. and M.L.S. are supported by the Rare Disease Clinical Research Center Urea Cycle grant RR19453-01 and the Genetics Training grant 1T32GM062758. D.P.M. is supported by National Institutes of Health grants 1RO1HD47880-01 and 1RO3N5051695-01. R.W.-M. is supported by a Wellcome Trust Research Career Development Fellowship. We have no conflicts of interest to disclose.

Web Resource

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for CPSID)

References

1. Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419–437
2. Nissim-Rafinia M, Kerem B (2005) The splicing machinery is a genetic modifier of disease severity. Trends Genet 21:480–483
3. Maquat LE (2005) Nonsense-mediated mRNA decay in mammals. J Cell Sci 118:1773–1776
4. Frischmeyer PA, Dietz HC (1999) Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet 8:1893–1900
5. Mendell JT, Dietz HC (2001) When the message goes awry: disease-producing mutations that influence mRNA content and performance. Cell 107:411–414
6. Eeds A, Hall LD, Yadav M, Willis AS, Summar S, Putnam A, Barr F, Summar ML (2006) The frequent observation of evidence for nonsense-mediated decay in RNA from patients with carbamyl phosphate synthetase I deficiency. Mol Genet Metab 89:80–86
7. Summar ML, Hall LD, Eeds AM, Hutcheson HB, Kuo AN, Willis AS, Rubio V, Arvin MK, Schofield JP, Dawson EP (2003) Characterization of genomic structure and polymorphisms in the human carbamyl phosphate synthetase I gene. Gene 311:51–57
8. Hart SL, Rancibia-Carcamo CV, Wolfert MA, Mailhos C, O'Reilly NJ, Ali RR, Coutelle C, George AJ, Harbottle RP, Knight AM, et al (1998) Lipid-mediated enhancement of transfection by a nonviral integrin-targeting vector. Hum Gene Ther 9:575–585
9. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33:e36
10. Wade-Martins R, Smith ER, Tyminski E, Chiocca EA, Saeki Y (2001) An infectious transfer and expression system for genomic DNA loci in human and mouse cells. Nat Biotechnol 19:1067–1070
11. Huschtscha LI, Holliday R (1983) Limited and unlimited growth of SV40-transformed cells from human diploid MRC-5 fibroblasts. J Cell Sci 63:77–99
12. Mortlock DP, Guenther C, Kingsley DM (2003) A general approach for identifying distant regulatory elements applied to the *Gdf6* gene. Genome Res 13:2069–2081
13. Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73:56–65
14. Willis AS, Freeman ML, Summar SR, Barr FE, Williams SM, Dawson E, Summar ML (2003) Ethnic diversity in a critical gene responsible for glutathione synthesis. Free Radic Biol Med 34:72–76
15. Mendell JT, ap Rhys CM, Dietz HC (2002) Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. Science 298:419–422
16. Summar ML (1998) Molecular genetic research into carbamoyl-phosphate synthase I: molecular defects and linkage markers. J Inherit Metab Dis Suppl 1 21:30–39
17. Wade-Martins R, Frampton J, James MR (1999) Long-term stability of large insert genomic DNA episomal shuttle vectors in human cells. Nucleic Acids Res 27:1674–1682
18. Wade-Martins R, White RE, Kimura H, Cook PR, James MR (2000) Stable correction of a genetic deficiency in human cells

by an episome carrying a 115 kb genomic transgene. Nat Biotechnol 18:1311–1314
19. Mendell JT, Medghalchi SM, Lake RG, Noensie EN, Dietz HC (2000) Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes. Mol Cell Biol 20:8944–8957
20. Lykke-Andersen J, Shu MD, Steitz JA (2000) Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell 103:1121–1131
21. Yates JL, Warren N, Sugden B (1985) Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. Nature 313:812–815
22. Belt PB, Groeneveld H, Teubel WJ, van de Putte P, Backendorf C (1989) Construction and properties of an Epstein-Barr-virus-derived cDNA expression vector for human cells. Gene 84:407–417
23. Lei DC, Kunzelmann K, Koslowsky T, Yezzi MJ, Escobar LC, Xu Z, Ellison AR, Rommens JM, Tsui LC, Tykocinski M, et al (1996) Episomal expression of wild-type CFTR corrects cAMP-dependent chloride transport in respiratory epithelial cells. Gene Ther 3:427–436
24. Leight ER, Sugden B (2000) EBNA-1: a protein pivotal to latent infection by Epstein-Barr virus. Rev Med Virol 10:83–100
25. Wade-Martins R, Saeki Y, Chiocca EA (2003) Infectious delivery of a 135-kb LDLR genomic locus leads to regulated complementation of low-density lipoprotein receptor deficiency in human cells. Mol Ther 7:604–612
26. Inoue R, Moghaddam KA, Ranasinghe M, Saeki Y, Chiocca EA, Wade-Martins R (2004) Infectious delivery of the 132 kb CDKN2A/CDKN2B genomic DNA region results in correctly spliced gene expression and growth suppression in glioma cells. Gene Ther 11:1195–1204
27. Laner A, Goussard S, Ramalho AS, Schwarz T, Amaral MD, Courvalin P, Schindelhauer D, Grillot-Courvalin C (2005) Bacterial transfer of large functional genomic DNA into human cells. Gene Ther 12:1559–1572
28. Simpson K, Huxley C (1996) A shuttle system for transfer of YACs between yeast and mammalian cells. Nucleic Acids Res 24:4693–4699
29. Kotzamanis G, Cheung W, Abdulrazzak H, Perez-Luz S, Howe S, Cooke H, Huxley C (2005) Construction of human artificial chromosome vectors by recombineering. Gene 351:29–38
30. Zhang XM, Huang JD (2003) Combination of overlapping bacterial artificial chromosomes by a two-step recombinogenic engineering method. Nucleic Acids Res 31:e81
31. White RE, Wade-Martins R, Hart SL, Frampton J, Huey B, Sai-Mehta A, Cerosaletti KM, Concannon P, James MR (2003) Functional delivery of large genomic DNA to human cells with a peptide-lipid vector. J Gene Med 5:883–892
32. Montigny WJ, Phelps SF, Illenye S, Heintz NH (2003) Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells. Biotechniques 35:796–807
33. Zhang MQ (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919–932
34. Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25:106–110
35. Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298