THE STUDY AND CHARACTERIZATION OF RNA INSTABILITY MUTATIONS IN CARBAMYL PHOSPHATE SYNTHETASE I (CPSI) THROUGH THE DEVELOPMENT OF A NOVEL BAC-BASED MODEL SYSTEM

By

Angela Michelle Eeds

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Molecular Physiology and Biophysics

August, 2006

Nashville, Tennessee

Approved:

Professor Ron Emeson

Professor Jonathan Haines

Professor Douglas Mortlock

Professor Richard O'Brien

Professor James G. Patton

This work is dedicated to my parents,

Rodger and Janice Payne

ACKNOWLEDGEMENTS

This work was generously supported financially by the Adam Scholarship, the National Urea Cycle Disorders Foundation, RD-CRC, the National Institute of Health (NIH) funded Human Genetics Training Grant (1T32GM062758), as well as additional funding from the NIH (5U54RR019453 and ROIES009915), and the Center for Human Genetic Research within the Vanderbilt University Department of Molecular Physiology and Biophysics. I am grateful to all of the patients and their families whose participation was central for this project.

I am appreciative to all of those with whom I have the opportunity to work during this academic pursuit. There are numerous collaborators, fellow students, and lab members who each played an integral part in this project and my overall graduate experience. I would especially like to thank my mentor, Marshall Summar, who is a great friend and mentor and also provided me with an ideal environment to grow as a scientist. All of the members of my Dissertation Committee were indispensable in giving guidance for my project, especially Douglas Mortlock who graciously chaired this committee.

The biggest inspiration in my academic endeavors has always been my family. I have the utmost admiration for my parents, Rodger and Janice, and my sister Jessica who have been steadfast in their love and encouragement. I am also forever thankful to my best friend and husband, Gary, for his interminable dedication and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A:	adenine residue
AP:	alkaline phosphatase
ATP:	adenosine triphosphate
BAC:	bacterial artificial chromosome
BECC:	BAC+pEHG+CMV+CPSI vector
bp:	base pairs
C:	cytosine residue
CBP:	cap binding protein
cDNA:	complementary deoxyribonucleic acid
CF:	cleavage factor
CMV:	cytomegalovirus
CP:	carbamyl phosphate
CPS1:	carbamyl phosphate synthetase 1
CPS1D:	carbamyl phosphate synthetase 1 deficiency
CPSF:	cleavage and polyadenylation specificity factor
CSF:	cleavage stimulation factor
DNA:	deoxyribonucleic acid
EBNA-1:	Epstein-Barr nuclear antigen 1
EDGE:	environmentally determined gene expression
eIF:	eukaryotic (translation) initiation factor
EJC:	exon junction complex

eRF: eukaryotic (translation) release factor

ESE: exon splicing enhancer

ESS: exon splicing silencer

G: guanine residue

gDNA: genomic deoxyribonucleic acid

GFP: green fluorescent protein

GMP: guanosine monophosphate

HAC: human artificial chromosome

hnRNP: heterogeneous nuclear ribonucleoproteins

hUPF: human up frameshift proteins (the surveillance complex)

ISE: intron splicing enhancer

ISS: intron splicing silencer

kDa: kilodaltons

kb: kilobases

LID: lipofectin/integrin/DNA

MDE: mutation detection enhancement

$MgCl_2$: magnesium chloride

N7: nitrogen residue

NaCl: sodium chloride

NAG: N-acetyl glutamate

NAGS: N-acetyl glutamate synthase

NIH: National Institute of Health

NMD: nonsense-mediated decay

NO:	nitric oxide
OTC:	ornithine transcarbamylase
OriP:	origin of replication
ORNT1:	ornithine/citrulline transporter
PABP:	polyadenosine binding protein
PAGE:	polyacrylamide gel
PAP:	polyadenosine polymerase
PARN:	polyadenosine ribonuclease
PCR:	polymerase chain reaction
PP2A:	protein phosphatase 2A
PTC:	premature termination codon
qRT/PCR:	quantitative reverse transcription and polymerase chain reaction
RNA:	ribonucleic acid
RRM:	RNA recognition motif
RT:	reverse transcription
SDS:	sodium dodecyl sulfate
SMG:	suppressor with morphological effect on genitalia (protein family)
snRNA:	small nuclear ribonucleic acid
snRNP:	small nuclear ribonucleic protein
SR:	serine and arginine rich proteins
SSC:	salt/sodium citrate
SSCP:	single strand conformation polymorphism
SURF:	SMG-1, UPF-1, eRF1, eRF3 protein complex

T:	thymine residue
U:	uracil residue
U1-6:	uridine rich RNAs, components of the spliceosome
µg:	microgram
µl:	microliter
UTR:	untranslated region
VDJ:	variable diversity joining
Wt:	wild-type
YAC:	yeast artificial chromosome

CHAPTER I

INTRODUCTION

Introduction to Carbamyl Phosphate Synthetase I

The Urea Cycle. The urea cycle is the primary mechanism for the body to clear nitrogen produced by protein metabolism (Figure 1.1). This pathway is comprised of five biochemical steps catalyzed by five primary enzymes, as well as a co-factor and several transporters. The first two enzymes of the pathway, carbamyl phosphate synthetase I (CPSI) and ornithine transcarbamylase (OTC) are liver-specific and are localized to the inner mitochondria. Argininosuccinate synthetase, argininosuccinate lyase, and arginase exhibit ubiquitous expression and play roles in arginine homeostasis and nitric oxide synthesis [1]. N-acetyl glutamate synthetase (NAGS) produces NAG, the allosteric co-activator for CPSI (reviewed in [2]). In addition, the ORNT1 ornithine/citrulline transporter [3] and the aspartate carrier, citrin [4], are necessary transporters due to partial localization of the urea cycle in the mitochondria.

Though the primary function of this pathway is nitrogen clearance, the urea cycle is responsible for generating adequate amounts of intermediate products for other metabolic functions. Most notably, the urea cycle intermediate arginine is the precursor for nitric oxide (NO). This biomolecule is important for a variety of biological functions including neurotransmission, vasodilatation for the regulation of blood pressure, and tumoricidal activities (reviewed in [5]). Altered

levels of NO have been implicated in various medical complications and impaired production has been linked to inadequate urea cycle function [6]. Components of the urea cycle also mediate arginine synthesis. This semi-essential amino acid is not only a precursor to NO and urea, but also several peptides with numerous intra-cellular roles including many proteins, creatine, polyamines, proline, glutamine, and agmatine. Additionally, arginine availability can selectively affect the expression of specific genes (reviewed in [7]). Therefore, urea synthesis is not the only function of enzymes in the urea cycle, but the enzymes are essential catalysts for other important biological processes as well.

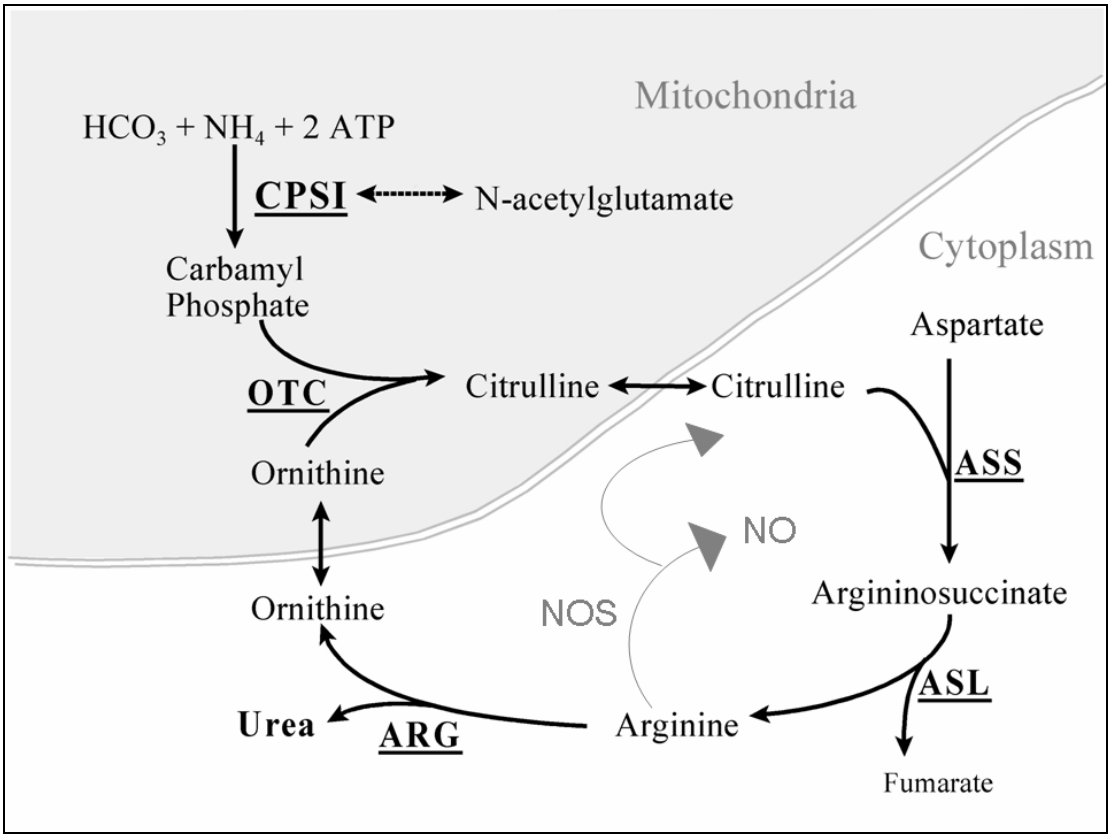


Figure 1.1. The hepatic urea cycle.

CPSI. CPSI is the enzyme that catalyzes the first and rate-determining step of the hepatic urea cycle. CPSI is a 165 kDa protein prior to cleavage of the amino terminus leader peptide sequence. This 38-residue peptide directs localization of the enzyme to the inner mitochondrial membrane and includes eight basic residues and one acidic residue as well as a Pro-Gly sequence four residues before the start of the mature peptide, consistent with other sequences that signal mitochondrial import [8]. Following cleavage of this leader peptide, the mature protein is 160 kDa. For proper activation, CPSI requires the presence of NAG, an allosteric co-activator. NAG is thought to be required for a conformational change in CPSI that exposes the two ATP-binding domains [9]. Activated CPSI converts ammonia and bicarbonate to carbamyl phosphate (CP) with the expenditure of 2 ATP molecules (Figure 1.2) [10-11].

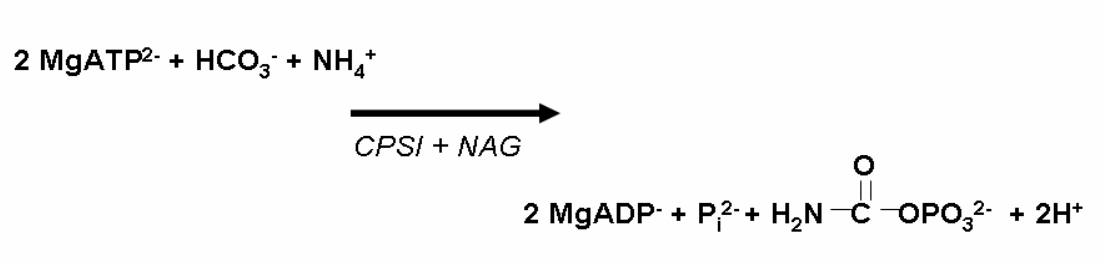


Figure 1.2. The creation of carbamyl phosphate catalyzed by CPSI.

CPSI exhibits high homology between mammalian species. Human CPSI shares 98% amino acid homology with the rat [12]. Sequence conservation between human and yeast or bacteria is lower and demonstrates the probable evolutionary origin of this enzyme. There is substantial evidence that mammalian CPSI arose from the fusion of two separate bacterial subunits, one

that metabolized glutamine and the other responsible for carbamyl phosphate synthesis [8]. Whereas bacteria and yeast have two separate subunits, all chordates studied have the 160 kDa single protein unit, suggesting an early evolutionary fusion of the glutamine processing and CP synthesis subunits. However, the mammalian enzyme is not able to process glutamine due to the loss of a critical residue [13]. Internal sequence homology and the presence of 2 ATP binding domains in the large, prokaryotic CP synthesis subunit suggests an ancient intra-gene duplication event, and this evidence extends into eukaryotic CPSI as well [8,14]. Strong conservation in certain portions of the enzyme between all species studied suggests the location of important functional domains and helped identify the ATP binding sites in rat CPSI [8,15].

CPSI is functionally similar to the other eukaryotic CPS enzymes, though there are notable distinctions. Whereas CPSII can utilize ammonia or glutamine as a nitrogen donor, ammonia is the only nitrogen donor compatible with CPSI. The CP produced from CPSI is used primarily for urea biosynthesis while CPSII provides the CP substrate primarily for pyrimidine nucleotide biosynthesis. Furthermore, CPSII is ubiquitously expressed, functions in the cytoplasm, and does not require NAG for activation [16 and references within]. CPSIII has only been detected in several fish and invertebrate species and catalyzes the formation of urea for osmoregulation and ammonia detoxification [17].

The human *CPSI* gene is located on chromosome 2q35 [18]. It spans approximately 120 kb of genomic DNA and is composed of 38 exons [12]. The cDNA is approximately 5700 bp (GenBank accession number AF154830). Gene

expression is restricted to high amounts in the liver and lower levels in the intestine. Within the hepatic tissue, CPSI is localized in the parenchymal cells surrounding the terminal portal venules [19].

In conjunction with the definition of *CPSI* genomic structure, several polymorphisms were identified [12]. Polymorphisms are differences in DNA sequence, found in both coding and non-coding regions, which are distributed throughout the general population regardless of functional consequences. Some of these polymorphisms likely play minor roles in the proficiency of the urea cycle and may contribute to clinical phenotypes. Specifically, the T1405N polymorphism has been correlated with vascular smooth muscle reactivity [20], the presence of pulmonary hypertension due to decreased arginine and NO production in infants [6], the presence of pulmonary hypertension following congenital heart surgery [21], and the development of hepatic veno-occlusive disease following bone marrow transplant [22]. In the above studies, the patients homozygous for the asparagine (N) genotype of T1405N demonstrated either a decreased incidence of hypertension or a higher production of arginine (and subsequently NO), indicating a functional difference in CPSI with clinical correlations (reviewed in [22]).

In addition to these common variants, the analysis of patients with defective CPSI activity has resulted in the identification of numerous rare variants, or pathogenic mutations in the gene. These genetic changes functionally alter the CPSI enzyme, eliciting a disease phenotype. Though a number of pathogenic mutations have been published previously [23-29], the

majority of mutations found by the Summar lab have not yet been published. These unpublished results include approximately 160 total and 115 unique mutations in the 105 patients diagnosed with CPSI Deficiency screened to date (Marshall Summar, unpublished data).

CPSI Deficiency. CPSI Deficiency (CPSID) is an inborn error of metabolism due to mutations in the *CPSI* gene which is inherited in an autosomal recessive manner. The prevalence of disease is estimated to be about one in 800,000, but the true prevalence is difficult to detect due to rare diagnosis confirmation [30]. CPSID can present clinically in the neonatal period (within 72 hours of life) or as late-onset following an acute trauma. The symptoms of CPSID include impaired mental and physical development and protein intolerance with the hallmark symptom of hyperammonemia, or elevated blood ammonia levels. Normal levels are approximately 35 μM/L in adults and 100 uM/L for infants [31], whereas a CPSID patient can have levels exceeding 3,000 uM/L. These elevated levels cause cerebral edema and may result in irreversible brain damage. MRI documentation illustrates the extreme cerebral sensitivity to hyperammonemia [32]. Rapid reduction of ammonia levels requires invasive measures yet must be performed to avoid long-term cognitive impairment. Treatments include, but are not limited to, the use of nitrogen scavenging drugs (such as sodium benozoate, sodium phenylacetate, and sodium phenylbutyrate), dialysis, and dietary limits on protein intake. Cases of partial CPSI Deficiency have been recorded where the disorder was able to be controlled by drug administration and dietary supplementation [33]. However, the only effective

long-term treatment is a liver transplant, which has rarely been performed given the rapid onset and deterioration of patients with the more severe phenotypes.

The severity of the patient phenotype depends on the specific genetic defect and sometimes environmental factors as well. While there are severe enzyme defects in the neonatal form, partial deficiencies are present in the late-onset form, as determined by measurements of CPSI enzyme activity [23, 33-35]. In the late-onset form, the disease may go undetected until an acute insult such as physical body trauma, sickness, or a rapid change in diet exposes the enzymatic deficiency, usually due to a need for increased nitrogen clearance [36]. Taken together, this demonstrates that under normal physiological conditions, CPSI is not saturated by its substrates and environmental stimuli can be pivotal for disease onset.

The presence of the functional T1405N polymorphism and the variability in CPSID phenotype (particularly the late-onset form) demonstrates that even in single-gene Mendelian disorders, environmental factors affect molecular mechanisms. The most severe genetic defects, such as those resulting in null enzyme activity and neonatal CPSID, cause a disease phenotype irrespective of environmental influences. However, genetic changes such as polymorphisms or those resulting in partial enzyme deficiencies require a certain environmental context to unmask the genetic defect. This trend is embodied in the EDGE, or Environmentally Determined Gene Expression concept developed by the Summar laboratory [22]. The EDGE concept describes a process whereby the effects of a spectrum of genetic variations may, depending on their severity, be

affected by non-genetic factors. The position of CPSI as the rate-limiting step in the first committed step of ureagenesis makes enzyme function particularly sensitive to both genetic and environmental factors.

Taken together, the availability of a large database of CPSID patient mutations (both neonatal and late-onset), the heterogeneity of the rare genetic variations found, the presence of at least one functional polymorphism, and its position as the first and rate-determining enzyme of an important biochemical pathway make CPSI an ideal research subject.

Introduction to RNA processing

The process of gene expression, beginning with DNA transcription, is a complex set of reactions resulting in an RNA copy of a gene which must then be properly translated into a specific protein product. Alterations in genomic DNA can therefore carry into the amino acid sequence of the protein and have minor or profound phenotypic effects. However, sequence variation can also impact the process of gene expression due to the requirement for *cis*-acting sequences in proper RNA processing. Such genomic changes may also have phenotypic consequences.

Though there are several types of RNA involved in gene expression, it is the messenger RNA (mRNA) that is the intermediate step between a gene and its protein product that must contain the appropriate sequence and be properly processed for correct protein production. An mRNA transcript does not exist *in vivo* as a free molecule, but rather is always in a dynamic complex with other

ribonucleic acids and proteins in a structure referred to as an hnRNP, or heterogeneous nuclear ribonucleoprotein. These accessory components of mRNA prevent the formation of secondary structure and participate in mRNA processing.

Following transcription, pre-mRNA must undergo several modifications before transport out of the nucleus for translation of the mature mRNA. Each of these processing events is necessary for proper gene expression. These events include capping, polyadenylation, splicing, and nuclear export. A mature mRNA product is produced by splicing such that non-coding sequences (introns) are removed as the coding sequences (exons) are ligated together. All RNA processing events are thought to be tightly coupled and are all important determinants of RNA stability [37-41].

Capping. As mRNAs are transcribed, a 7-methyl guanosine "cap" is added to the 5' end of the transcript [42]. Important functions of the cap include protection from nucleases, increase in translation efficiency, transport to the cytoplasm, and proper splicing. Capping begins after approximately 30 nucleotides have been synthesized during transcription when the 7-methyl guanosine residue binds to the first nucleoside of the growing transcript via a 5'-5' triphosphate bridge. This modification is functionally conserved in all eukaryotic organisms examined and occurs in three enzymatic steps that begin when phosphorylated RNA polymerase II recruits RNA triphosphatase. This enzyme removes the terminal (γ) phosphate from the pre-mRNA, creating a diphosphate terminus. Next, RNA guanylyl transferase adds the capping GMP to

the terminus. RNA methyl transferase then methylates the N7 in the guanosine residue to complete the cap structure (reviewed in [43]). In part, the cap performs its functions by serving as a binding site for multiple proteins. Notably, the cap binding 80/20 complex is bound to the 5' cap while the mRNA is in the nucleus. However, once transported to the cytoplasm, this complex is replaced by the eukaryotic initiation factor eIF4E, which allows transcript recognition for translation. This process is graphically represented in Figure 1.3.

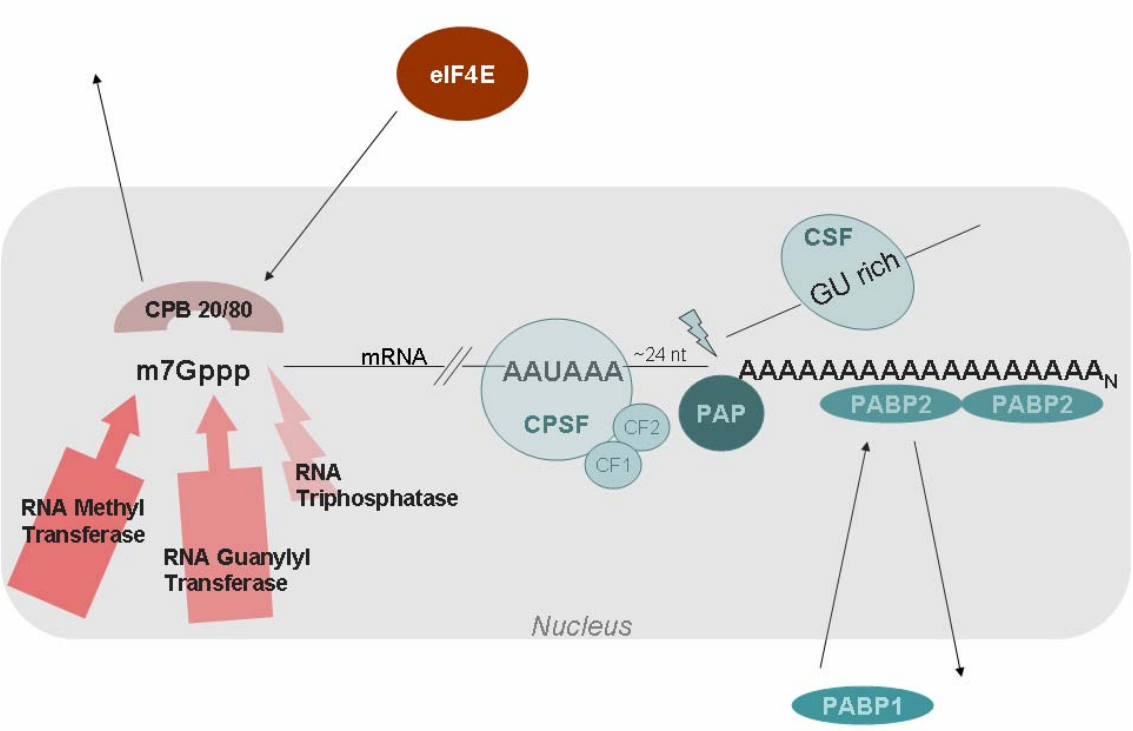


Figure 1.3. Capping and polyadenylation. Capping occurs on the 5' end and cleavage/polyadenylation occurs on the 3' end of mRNA transcripts via several protein components. Capping enzymes are represented in red shades and polyadenylation enzymes are represented in blue shades. CPSF is cleavage and polyadenylation stimulation factor, CF is cleavage factor, CSF is cleavage stimulation factor, PAP is poly(A) polymerase, PABP is poly(A) binding protein.

Polyadenylation. Polyadenylation, the addition of a tail of adenosine residues to the mRNA transcript, occurs following 3' end mRNA cleavage. Cleavage and polyadenylation occur in a tightly coupled set of reactions immediately following pre-mRNA transcription (reviewed in [44]). The cleavage position, also the position of poly(A) addition, is specified by the presence of a tightly conserved AAUAAA sequence positioned approximately 24 nucleotides upstream in the 3' UTR. Mammalian transcripts most often also have a loosely conserved GU-rich sequence immediately upstream of the cleavage position. Cleavage is catalyzed by multiple proteins including cleavage and p(A) specificity factor (CPSF) which recognizes the AAUAAA sequence, cleavage stimulation factor (CSF) which binds downstream in the GU-rich sequence, and cleavage factors I and II (CF I and CF 2) that help stabilize the overall complex (Figure 1.3). Following cleavage, up to 250 adenosine residues are added in a two-step polyadenylation reaction catalyzed by the poly(A) polymerase (PAP) that first slowly adds at least 12 adenosine residues before switching to a more processive synthesis for the addition of ~200 adenosine residues stimulated by the poly(A) binding protein (PABP2) [38,45,46]. While in the nucleus, PAPB2 binds to the poly(A) tail. Once the mRNA relocates to the cytoplasm, PAPB1 replaces PAPB2. The poly(A) tail undergoes constant breakdown in the cytoplasm balanced by consistent rebuilding of the tail. The equilibrium of this dynamic is towards the loss of the poly(A) tail so that it is eventually lost, a signal for mRNA destruction.

Similar to capping, polyadenylation and the associated proteins provide protection to the transcript as well as increasing translation efficiency. Interactions between the 5' cap and 3' poly(A) tail via translation initiation factors (such as eIF4E) that can bind both structures create a closed loop which mediates stability and translation of the mRNA. This closed loop structure has been re-created with recombinant proteins and visualized by atomic force microscopy [47]. PABP association with a transcript was specifically implicated in recruiting these bridging translation initiation factors [46]. Therefore, the cap and poly(A) tail structures are both necessary mediators of downstream mRNA processing and stability that function, at least partially, in a synergistic manner.

Splicing. Splicing is the process of intron removal and exon fusion that occurs via two transesterification reactions as pre-mRNA undergoes processing to become mature mRNA, shown graphically in Figure 1.4. In the first step, the 2' hydroxyl group of the branch point "A" nucleotide within the intron attacks the phosphodiester bond between the 3' end of the first exon and the 5' end of the intron. This reaction frees the first exon and creates a lariat-shaped intron intermediate still attached to the downstream exon. In the second step, the free 3' hydroxyl group on exon 1 attacks the phosphodiester bond between the 3' end of the intron and the 5' end of the next exon. This creates a spliced exon/exon product and releases the intron, retaining the lariat structure. Splicing requires a myriad of both *cis* and *trans* elements for proper splice site recognition and to catalyze the breakage and reformation of phosphodiester bonds.

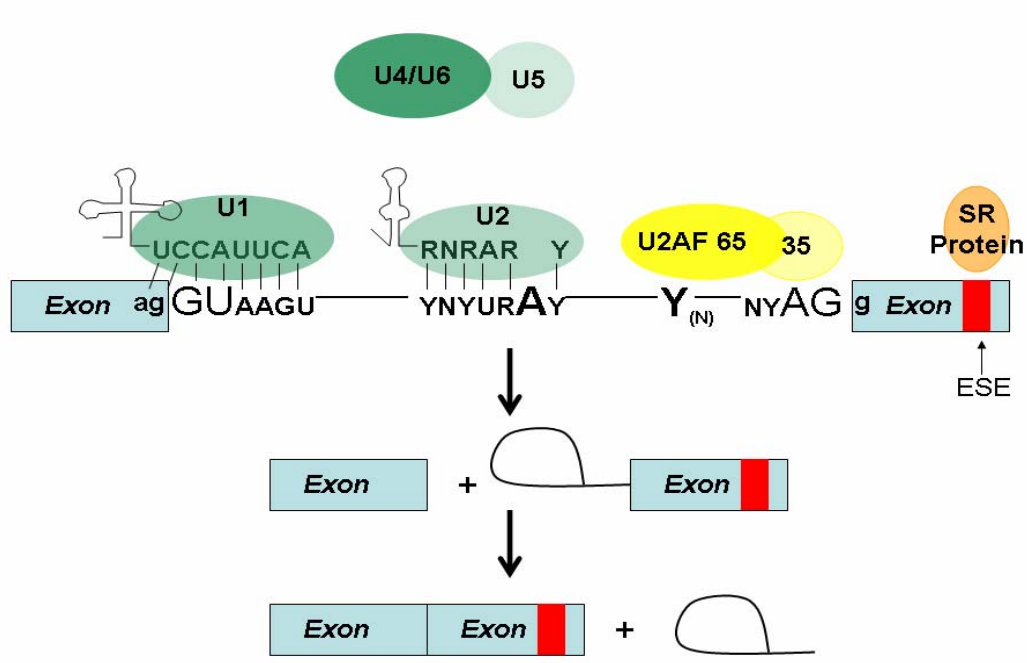


Figure 1.4. The splicing pathway. The conserved sequences necessary for proper splicing are illustrated along with several spliceosome components that recognize these *cis* elements which aid exon ligation and intron removal.

Cis elements include the consensus sequences surrounding the intronic branch point and at the 5' and 3' end of exons and introns. The branch point is most often found within 40 bp upstream of the 3' splice site, almost always the first AG dinucleotide downstream of the branch point [48,49]. The consensus sequences for proper intron and exon definition are ag|GTAAGT for the 5' splice site and $(Y)_n$NYAG|g for the 3' splice site, where exonic nucleotides are in lower case, intronic nucleotides are capitalized, Y represents a pyrimidine (C or U), and n is approximately 8 [50]. These conserved sequences are each recognized by multiple splicing components. A minor intron class represents an exception to the consensus cited above. This minor class, referred to as AT-AC introns, requires a distinct minor spliceosome with components that are functionally analogous to the well-characterized major spliceosome (see below) [51]. For example, disruption of the AT-AC consensus in the tumor suppressor gene, *LKB21*, is one mutation responsible for Peutz-Jeghers syndrome [52].

Because degeneracy is sometimes observed in these consensus sequences and because the splicing motifs may occur at a higher frequency than the number of splicing events, other *cis* elements are often present. These sequences act either as splicing enhancers or silencers and are located in both introns and exons. These intronic and exonic splicing enhancers (ISEs and ESEs) and intronic and exonic silencers (ISSs and ESSs) recruit a number of different SR (serine/arginine) proteins involved in splicing [53,54]. These semi-conserved sequences may be present in any location on a transcript and work in a synergistic manner to increase exon definition and promote the proper exon

splicing order by recruiting certain splicing proteins or repressors [55]. Though consensus sequences are defined for four SR proteins, the degenerate nature of these sequences makes detection difficult [56]. This recognition motif degeneracy may allow binding overlap among the SR proteins, which could be important for eliciting combinatorial effects for specificity that depend on SR protein levels, binding affinities and specific interactions with other proteins (reviewed in [57,58]). This degeneracy may be also be necessary due to the evolutionary selection pressures for genomic variants to maintain both proper splicing sequences and tri-nucleotide codons for proper protein structure [59].

Trans factors required for proper splicing make up the spliceosome, a multi-component complex that is responsible for recognition of the intron/exon boundaries and catalyzes bond breakage and reformation. This dynamic complex includes small nuclear RNAs (snRNAs) as well as multiple proteins. The spliceosome forms in a largely stepwise pathway based on multiple RNA and protein interactions mediated by the conserved RNA sequences and, in most steps, an ATP-dependence. The snRNAs, U1, U2, U4, U5, and U6, together with their associated proteins (snRNPs), specifically bind to distinct positions in the pre-mRNA to mediate recognition and positioning of the 5' and 3' splice sites while also recruiting numerous proteins [60,61]. It is thought that the snRNAs themselves mediate catalysis during splicing (reviewed in [61]). A second group of snRNAs, U11, U12, U4atac and U6atac, mediate splicing in the minor AT-AC intron class in the same manner as the major spliceosome [51].

Serine/arginine rich (SR) proteins are also integral *trans* components of the splicing reaction. These proteins perform numerous functions including binding to RNA (at ESEs for example) as a signal for splicing commitment, recruitment of other splicing components, bridging the 5' and 3' ends of introns, and assisting in exon and intron definition. This protein family is characterized by the presence of one or more RNA recognition motifs (RRMs) and a *trans*-activation domain containing serine/arginine dipeptides (RS domain), responsible for protein/protein interactions. Their activation and subcellular localization is dependent on their phosphorylation levels [53,58].

Estimates indicate that approximately 30% of disease-causing mutations in humans affect splicing [53,54]. It has also been hypothesized that splicing mutations are the most frequent cause of hereditary disease [62]. While many pathogenic splicing mutations disrupt intronic splicing motifs, others are located in the coding region where they can be frameshift, nonsense, missense, or silent mutations. The inclusion of silent and missense mutations in the category of mutations that affect splicing is not intuitive. However, while point and silent mutations are canonically thought to have little or no effect on protein activity, they can have severe effects if located in exon enhancer elements, subsequently affecting RNA integrity [63]. Indeed, several examples of silent and missense mutations that cause splicing errors have been reviewed [53]. Computational approaches support that almost every gene, whether it contains alternative or constitutive exons, contains multiple enhancer and silencer elements [64].

Nuclear export. It is essential that only mature mRNAs serve as the template for polypeptide synthesis during translation. This process occurs in the cytoplasm via ribosomes. Because RNA processing occurs in the nucleus, an important last step before translation is the selective export of only mature mRNA from the nucleus through the nuclear pore complex. Several mechanisms are necessary to prevent pre-mRNAs from reaching the cytoplasm, including the presence of snRNPs that prevent the passage of immature transcripts through the nuclear pore and the association of proteins with a nuclear export signal specifically on fully processed transcripts. These proteins containing a nuclear export signal are components of the EJC. Beyond relocation, nuclear export aids processing of the mRNA transcript. Though some protein constituents remain stably associated with the transcript, other components are strictly cytoplasmic (eIF4E, PABP1) or nuclear (CBP 80/20, PABP2) and it is during or immediately following nuclear export that these dynamic rearrangements take place which are important for translation.

As pre-mRNA is transcribed, capped, polyadenylated, spliced, and exported to the cytoplasm for translation, multiple *cis* and *trans* elements are required. Though the presence and proper function of *trans* elements is essential, it is notable that the sequence of the mRNA itself may have the most profound effects on its own processing. If a genetic change is present or if an improper mRNA sequence is produced, the interactions between additional *trans* components may be able to recognize the transcript as aberrant and elicit degradation through the nonsense-mediated decay pathway.

RNA quality control: nonsense-mediated decay

The nonsense-mediated decay (NMD) pathway is a quality control mechanism that makes mRNA transcripts with premature termination codons (PTCs) sensitive to degradation and thereby prevent them from undergoing translation, which could produce C-terminal truncated proteins with harmful gain-of-function or dominant negative effects. A PTC is a stop codon that resides at least 50 bp upstream of the last exon/exon junction [65] and can result from frameshifts, splicing changes, and nonsense mutations. Genetic mutations that introduce PTCs and therefore result in NMD are implicated as a common cause of disease [66-68], highlighting the importance of this pathway for mediating gene expression and RNA stability.

The evolutionary conservation of this pathway, extensively studied in yeast, nematode, and human cells, suggests its biological importance. Given the conservation seen among NMD components and the complex interplay between other steps of RNA processing, it is likely that the NMD pathway not only functions to degrade aberrant transcripts from genetic mutations, but plays a more central role in gene expression. NMD is also involved in the degradation of transcripts as a means of posttranscriptional control [69], those with naturally occurring PTCs through alternate splicing [70], and unproductive products of VDJ arrangement [71]. The components of this pathway have also been shown to influence translation initiation and termination [72], regulate pseudogenes [73,74], amino acid biosynthesis [75], and are closely associated with splicing [76].

NMD requirements. The currently accepted mechanism for this pathway involves the interactions of multiple proteins (discussed below) as well as active cellular translation and a transcript that, as a pre-mRNA, contained introns. Suppressor tRNAs and various antibiotics such as anisomycin, cyclohexamide, and puromycin inhibit translation by different mechanisms, and treating cells with these compounds in turn allows the stabilization of PTC-containing transcripts (see examples [77-79]). A transcript must have undergone splicing to be NMD-sensitive, as naturally intronless genes and those with the introns removed (such as cDNA constructs) are immune to NMD [80-82]. During splicing, the multi-protein exon junction complex (EJC) is deposited approximately 22 bp upstream of the newly-formed exon/exon junction. One function of this dynamic protein complex is to provide positional information for distinguishing between a PTC, which would be located 5' of the EJC, and the wild type termination codon, which is almost always 3' of the EJC in the terminal exon.

Exon junction complex. The EJC is a dynamic complex of proteins made up of core components as well as auxiliary proteins. An EJC provides essential positional information for distinguishing between legitimate and premature termination codons and recruits the surveillance proteins, which are other obligate protein participants in NMD. Other roles for the EJC include RNA nuclear export and upregulation of mRNA translation, either directly or by recruitment of auxiliary proteins [83-85]. The EJC is approximately 335 kDa and, when bound to RNA, protects 8-10 nucleotides suggesting that most components do not contact the nucleic acid itself but rely on protein-protein interactions [86].

The core components of the EJC are necessary for protein recruitment and anchor onto the RNA in a sequence independent manner. The EJC core remains stably associated with the mRNA, regardless of the overall evolution of EJC composition, to serve as a platform for the multiple post-splicing processing events. Using purified recombinant proteins and synthetic RNA, the necessary and sufficient EJC subunits were recently defined by Ballut et al. as MLN51, MAGOH, Y14, and eIF4AIII+ATP [87]. eIF4AIII was previously identified as a member of the EJC and shown to be required for NMD using siRNA-mediated protein knockdown [88-90]. It is an RNA helicase that binds RNA in a sequence-independent manner, and along with the nucleocytoplasmic shuttling protein, MLN51, comprises the EJC anchor on RNA [87,91]. Stable RNA association is maintained by the Y14/MAGOH heterodimer that inhibits eIF4AIII ATPase activity, an interaction that is in turn stabilized by MLN51 [87]. Prior to the Ballut study, Y14 and MAGOH were defined as core components because their association with the mRNA was not transient; rather they could only be removed by translation [92]. In addition, Y14/MAGOH shuttle and recruit UPFs and other EJC components such as UPF3 and SMG-1 (see “Surveillance proteins” below) [93-95].

Other identified members of the EJC include RNPS1, SRm160, DEK, REF, UAP56, and pinin [86]. RNPS1, an RNA binding protein, co-activates splicing and has been confirmed to have a direct role in NMD as well [96]. SRm160, a serine/arginine-related protein is also a splicing co-activator that is closely associated with DEK, originally identified as an oncoprotein, but shows

specific association with spliced RNAs [97]. REF (also referred to as ALY) interacts with TAP/p15 for mRNA nuclear export (reviewed in [98]). UAP56 is a U2 snRNP auxiliary factor. Pinin, identified by yeast two-hybrid and pull-down assays to interact with RNPS1, is localized in nuclear speckles and participates in splicing [99]. Y14 has been shown to interact specifically with REF, RNPS1, TAP, and MAGOH on spliced RNA [100].

Surveillance proteins. Three conserved proteins make up the core complex of the NMD machinery in all organisms studied. When these surveillance proteins are deleted or silenced, the NMD pathway cannot be activated. The human surveillance proteins, known as up-frameshift proteins, are distinguished as hUPF1, hUPF2, and hUPF3 (has multiple isoforms) [96,101]. Each performs distinct functions in the commitment of a transcript for degradation.

Multiple isoforms of UPF3 encoded by two separate genes were identified based on homology to *Saccharomyces cerevisiae* Upf3 [102]. These isoforms, designated UPF3a and UPF3b, are nucleocytoplasmic shuttling proteins that both contain an RRM-like domain but differ in length by 6 amino acids and were therefore hypothesized to have separable roles in NMD [96]. Indeed, distinct complexes have been identified during NMD that contain the different UPF3 isoforms [103]. Recently, tethering experiments demonstrated that UPF3a is less efficient than UPF3b for inducing NMD [101]. UPF3b is most likely the first surveillance factor to associate with the mRNA [96]. This association occurs during or following splicing, as it specifically interacts with spliced RNAs [93,96].

Protein-protein interactions have been demonstrated between UPF3b and the Y14/MAGOH heterodimer in the EJC [93,104]. In addition, UPF3 plays a role in translation [101] and is at least partially responsible for the recruitment of the second surveillance factor, UPF2, to the mRNA [105].

UPF2 is an anchor protein that interacts with both UPF3 and UPF1. It was identified and characterized based on homology with the *Saccharomyces cerevisiae* UPF2 [96,102,106]. Recent studies suggest that it is a phosphoprotein regulated by various phosphorylation states [107]. Though UPF2 contains nuclear localization signals, a GFP fusion protein revealed its presence in the cytoplasm [106]. Polyclonal antibody staining further refined the location of UPF2 as perinuclear, corroborating evidence that it most likely associates with the mRNA during or immediately following nuclear export [96]. It is recruited to mRNA by UPF3 via a direct protein-protein interaction, as illustrated by the crystal structure of the resulting protein complex [105]. In turn, UPF2 interacts with UPF1 as demonstrated by mutation studies where removing the UPF2 residues analogous to the yeast UPF1-interacting domain prevents NMD [96]. In addition, immunopreciptation assays showed interactions between UPF1 and 2, demonstrating that UPF2 interacts with UPF1 in mammalian cells [102].

UPF1 is the final surveillance protein to interact with an mRNA. It is a shuttling protein with a primarily cytoplasmic localization. It functions as an RNA-dependent ATPase and 5' to 3' helicase and its alternate phosphorylation states are most likely the "triggers" for RNA decay via complex remodeling [103]. Incomplete removal of UPF2, UPF3, and the EJC during translation recruits

UPF1 allowing the assembly of a complete trimeric UPF1, 2, and 3 surveillance complex. UPF1 is recruited to nuclear mRNA as a member of the SURF complex (SMG-1, UPF1, eRF1, and eRF3), most likely just after termination codon recognition [95]. The likely signal that a termination codon is premature is if the SURF complex at the termination codon can interact with a downstream EJC-UPF2-UPF3 complex. This interaction is thought to cause UPF1 phosphorylation by SMG-1, a phosphatidylinositol 3-kinase-related protein kinase [95,108,109]. Phosphorylated UPF1 then induces remodeling of the surveillance complex [95]. Subsequent UPF1 dephosphorylation by protein phosphatase 2A (PP2A), when associated with the human SMG5/7a complex, is also necessary for NMD [103,110]. Tethering experiments suggest that the presence of SMG-7 on a transcript, irrespective of the presence of a PTC or the C-terminal protein interaction domain, targets it for degradation, illustrating a role for dephosphorylation as an NMD trigger [111]. Therefore NMD likely is triggered first by UPF1 phosphorylation to induce remodeling and recruit SMG-7 (in a complex with SMG-5 and PP2A), which then triggers RNA decay and UPF1 dephosphorylation via SMG-5 and PP2A [111]. These observations underscore the importance of UPF1, mediated by its phosphorylation state, in eliciting NMD.

NMD location. The processes of RNA splicing, nuclear export, translation, and NMD surveillance are mechanistically linked in part due to the multiple functions of proteins associated with mRNPs and the numerous steps of mRNP remodeling. However, each of these processes, except NMD, shows distinct sub-cellular localization. The sub-cellular localization of NMD is unclear, but

thought to be “nucleus-associated” in part because some EJC components and the surveillance proteins show nuclear localization (DEK, SRm160), while others are primarily cytoplasmic (UPF1) [83]. Whereas the translational machinery necessary for termination codon recognition is located in the cytoplasm, most evidence suggests that in eukaryotic cells, RNA degradation occurs while associated with the nucleus [80,112]. For example, NMD degradation occurs specifically on transcripts associated with the CBP 80/20 complex which is only bound to nuclear transcripts. Once the nuclear the translation initiation factor eIF4E replaces CPB 80/20 on the mRNA transcript, it is not sensitive to degradation [113]. In support of this observation, an anti-CBP80 antibody, but not an anti-eIF4E antibody, co-immunopurifies with the EJC components RNPS1, Y14, SRm160, REF/Aly, and TAP in both nuclear and cytoplasmic fractions [114]. In mRNP complexes containing UPF2 and UPF3b, CBP80 but not eIF4E is detected by immunopurification suggesting NMD occurs while the transcript is bound by CPB80/20 and that eIF4E binds at the same time as or after transcript surveillance and EJC removal [113,114]. In addition, translation has been observed while transcripts are bound to CPB80/20, consistent with the requirement of translation for NMD [113].

A current model to explain these observations is that transcripts go through a nucleus-associated “pioneer round of translation” (reviewed in [98]). In this model, the RNA undergoes a preliminary scan by the translation machinery while still associated with the surveillance proteins, the EJC, and the nuclear 5’ and 3’ proteins. If no PTCs are located, the RNA is remodeled to the steady-

state translation initiation complex. If a PTC is located, the transcript submits to degradation [98]. In an effort to illustrate distinctions between these two translation events, Chiu et al demonstrated that the pioneer and steady-state translation initiation complexes involve distinct mRNP substrates, despite overlap in their requirement for several translation initiation factors [115]. This nuclear association, along with major components of this pathway, is illustrated in Figure 1.5.

Once NMD is triggered, decapping and deadenylation occur succeeded by 3' to 5' and 5' to 3' exonuclease degradation [116,117]. Though there are two distinct decapping complexes, Dcp1/Dcp2 and DcpS, only the first has been studied in conjunction with NMD-mediated degradation. Whereas Dcp1 is non-catalytic, its presence is necessary for the stimulation of the catalytic Dcp2 enzyme (rev in [43]). When Dcp2, as well as a deadenylase (PARN; poly(A) ribonuclease), and a component of the 3' to 5' exosome (PM/Scl100) were each individually inhibited by siRNA, NMD-containing transcripts were stabilized suggesting the involvement of each in NMD [117]. In addition, UPF1, UPF2, and UPF3 coimmunopurify with each of those proteins as well as other 5′→3′ exonucleases (Rat1 and Xrn1), and other components of the exosome (Rrp4 and Rrp41), suggesting that the surveillance complex is responsible for recruiting decay factors during NMD [117,118]. The relative efficiencies for decapping and deadenylation, however, remain unknown.

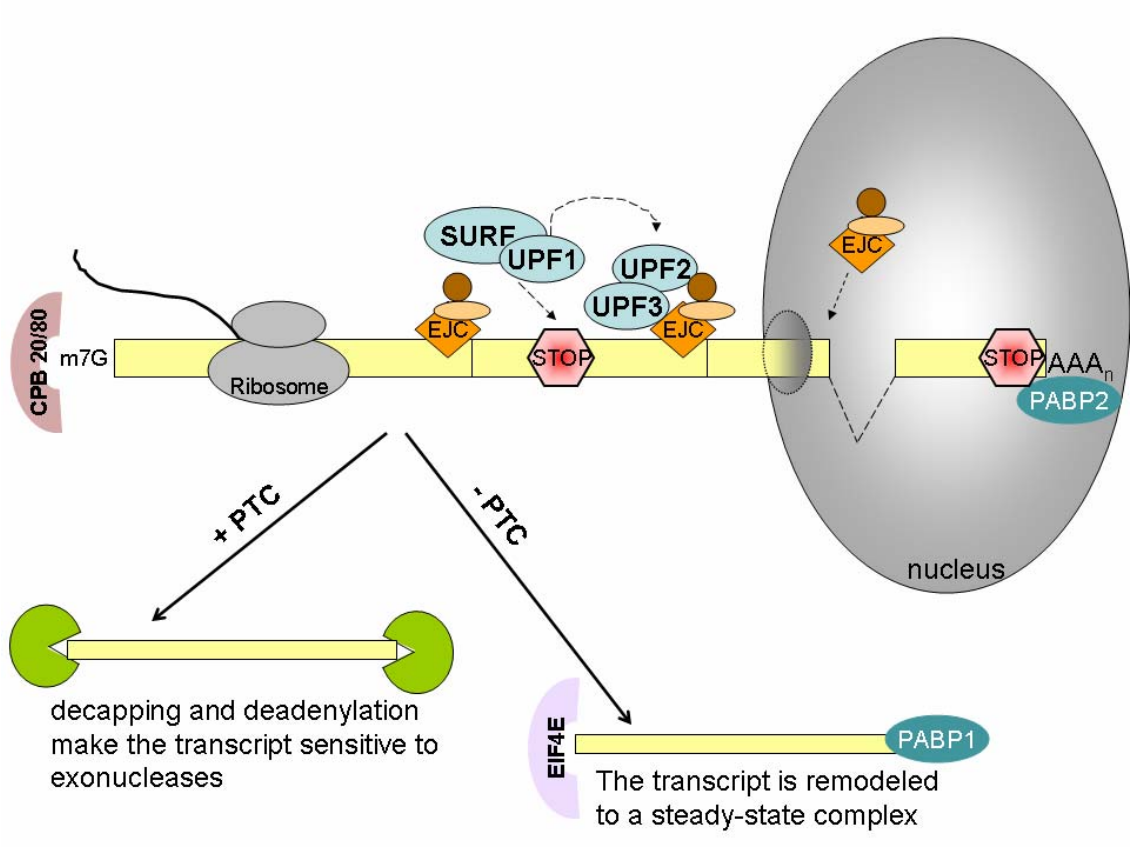


Figure 1.5. Nonsense-mediated decay. This graphical overview of NMD illustrates the link between splicing, nuclear export, and the pioneer round of translation. NMD surveillance occurs on transcripts bound by the CPB complex. The spatial interactions between SURF/UPF1 and the EJC/UPF2/UPF3 complexes identify termination codons and in the presence of a PTC, the complex is degraded following decapping and deadenylation. If no PTC is recognized, the transcript is remodeled to a steady-state complex.

NMD in disease. It has been estimated that up to 30% of disease-associated variants create PTCs through nonsense or frameshift mutations [67]. The presence of NMD has been shown to alter the phenotypic consequences of genetic mutations. Specifically, a mutation that would produce a protein with a dominant negative effect if translated could be changed by NMD to cause a loss of function or haploinsufficiency phenotype.

One physiological example of the phenotypic effects of NMD is the spectrum of disease-causing variants found in the Sox10 transcription factor. Most disease-associated mutations in this gene result in PTC formation. Patients with PTCs that occur upstream of the terminal exon/exon junction have a less severe phenotype because of NMD degradation whereas those with PTCs occurring downstream of the terminal exon/exon junction produce a dominant negative Sox10 protein [119]. A severe phenotype is therefore present because though the protein retains the DNA binding domain, it lacks the protein domain necessary for interaction with other transcription elements and disrupts transcription of target genes.

NMD also shows particular consequences in proteins that bind as multimers, such as the protein components of hemoglobin. Prior to characterization and appreciation of the NMD pathway, phenotypic differences had been described in patients with β-thalassemia. Patients with nonsense codon mutations in non-terminal exons had a less severe phenotype, while nonsense codon mutations in the terminal exon correlated with the most severe phenotype [120]. In addition, mRNA levels were decreased only in patients with

nonsense codons in the non-terminal exons [120]. These results can be explained by the multimeric protein complex that makes up hemoglobin: 2 α- and 2 β- complexes create a tetramer. When incomplete β-globin protein subunits are translated, they are able to aggregate with wild type α- or β-globin proteins. However, the presence of truncated subunits prevents proper function of the holoenzyme. Therefore, the mutant subunits can exert dominant-negative effects by impairing the holoenzyme despite the presence of wild-type subunits. NMD that prevented formation of these truncated subunits would also prevent the formation of these non-functional aggregations. In this manner, NMD can protect heterozygous carriers of genes with a PTC from manifesting a more severe disease phenotype that would result from the expression of truncated proteins.

On the contrary, the activation of NMD may prevent otherwise functional proteins from translation. If the truncated protein does not dimerize with other proteins and active sites for binding substrates are not disrupted, the translation of this truncated transcript could then prevent a disease phenotype. Strategies to inhibit NMD could therefore be useful for diseases such as cystic fibrosis, Duchene's muscular dystrophy, Hurler Syndrome, and X-linked diabetes. These diseases are all associated with PTC forming mutations which would produce at least a partially active protein if translated [68].

Many lines of evidence point to the conclusion that there are many genomic mutations that alter the processing of the RNA transcript rather than the structure and/or function of a protein product. It is becoming increasingly clear that nonsense-mediated decay is a common phenomenon that can change

clinical phenotypes and will greatly impact the interpretation of mutation mechanisms. Disruption of RNA processing and quality control pathways is likely to be highly relevant in the study of severe genetic diseases such as CPSID that are caused by mutations that most often result in severe metabolic defects, suggesting they result in little or no detectable enzyme activity.

Thesis Overview

Hypothesis

The underlying molecular mechanisms causing disease in a substantial fraction of CPSID patients are defects in RNA processing such as splicing mutations and nonsense mediated decay and these mutations can be studied via a novel model using a BAC clone.

Specific Aims

I. Identify and quantify potential RNA instability mutations in patients with CPSID

   A. Mine patient database for patients with an RNA source and complete their cDNA screening

   B. Characterize and quantify the percentage of mutations with evidence for RNA instability

   C. Choose 4-6 mutations to study in detail

II. Devise a model system that can be used to test RNA instability mutations in CPSI

   A. Modify a BAC with the *CPSI* gene to contain E-GFP, Hygromycin resistance, EBNA-1 (pEHG vector)

   B. Place a viral promoter 5' of *CPSI* on BAC to drive expression in non-hepatic cells

   C. Test model system for *CPSI* expression

III. Test putative RNA instability mutations to determine the RNA processing defect

   A. Create patient mutations in BAC model system

   B. Examine mutations to determine RNA processing defect using quantitative RT/PCR, agarose RT/PCR, sequencing, and Northern blots

CHAPTER II

ESTIMATING THE PREVALENCE OF RNA INSTABILITY MUTATIONS IN CPSID

Introduction

A primary objective in understanding CPSI Deficiency is to classify identified mutations according to how each causes the resulting enzyme deficiency because even single-gene, classical Mendelian disorders exhibit alterations in disease severity based on the specific genetic mutation [121,122]. Accordingly, the heterogeneous set of mutations found in CPSID patients likely includes mutations that affect the final protein structure as well as mutations that disrupt normal processing of the RNA transcript. This Chapter describes the quantification of mutations that are likely RNA processing defects and provides evidence for this classification. The data was collected from the CPSID database maintained in the Summar laboratory detailing thorough examination of patient gDNA as well as RNA, when possible. This is the largest known database of CPSID patients, representing almost every known case in the United States and abroad collected over the previous 15 years.

Before the *CPSI* genomic structure was determined, laboratory efforts to identify patient mutations focused on the amplification of RNA from hepatic tissue or established patient fibroblast or lymphoblastoid cell lines. Due to the rare nature of CPSID, the expectation was to find consanguinity and homozygosity as a common cause of disease. However, in a majority of families with no obvious

consanguinity, only one allele could be detected in cDNA screens. Following determination of the gene structure, gDNA mutation detection revealed that these same non-consanguineous patients were actually heterozygous for two mutations. This puzzling “missing allele” observation suggested that some mutations were causing unstable RNA and were therefore not detected in the cDNA sequence. The characterization of the nonsense-mediated decay pathway (NMD) provides a reasonable explanation for the “missing allele” observations.

In addition to observations made through the study of CPSID, there is a growing body of evidence to show that many genetic mutations do not affect the protein product but rather the processing of the transcript, meaning that mature proteins are not always produced [75,123]. As detailed in Chapter I, NMD can change the phenotypic consequences of nonsense mutations by preventing the formation of dominant negative proteins that could result from partial transcripts and therefore may be a significant factor in the mutation pathology and observations of haploinsufficiency for a number of diseases [66]. Multiple publications estimate that up to 30% of genetic disorders and cancers could be the result of PTC-containing transcripts [66,67]. In agreement with these estimates, the CPSID “missing allele” observation suggests that a large number of mutations in the laboratory database affect the RNA transcript.

To understand more about RNA processing, it is beneficial to look for physiological examples of probable RNA processing mutations, as are present in CPSID patients. The study of mutations known to cause a disease phenotype

should be an effective way to identify important elements in the *CPSI* gene as well as to contribute to the overall understanding of RNA processing. In this Chapter, CPSID is used as a model for understanding and predicting the prevalence of PTC-forming mutations responsible for genetic disease. For the patients in the CPISD database for which both an archived RNA and DNA source were available, a significant number of suspected RNA processing mutations were identified. This evidence suggests the importance of accounting for PTC-containing transcripts in all disease mutation screening strategies.

Materials and Methods

Patient selection. For this study, only patients with an available gDNA and RNA source were included. These patients were determined to have CPSID by analysis of hepatic tissue enzyme activity (less than 5% of control), and/or a strong family history of CPSID with convincing clinical and laboratory symptoms of this disease [124]. If no liver sample was available for gDNA and RNA archiving, patient tissue was used to establish lymphoblastoid or fibroblast cell lines. In the instance where patient tissue was unavailable, parent samples were collected for analysis (patient 1 and 18 in Table 1 are CPSID carriers). All patients tested for this study were of the same ethnocultural background and over 100 control chromosomes were examined for each identified genetic variant using a PCR-based strategy. Institutional Review Board approval was obtained.

SSCP. Single strand conformation polymorphism (SSCP) analysis was used as a high throughput method to detect mutations in each exon due to the

sequence-specific folding conformations of single-stranded DNA. PCR was used to amplify all exons and at least 50 bp of the flanking intron sequences in CPSID patients to detect mutations. These reactions were carried out using Platinum SuperMix (Invitrogen) and primer pairs previously published by the laboratory [12]. PCR reactions were then denatured by heating at 95°C for 2 min and quickly transferred to ice for formation of sequence-specific complexes. The complexes were run for 8 to 14 h at 15 w in a non-denaturing MDE (mutation detection enhancement, Cambrex Bio Sciences) gel at 4°C. Gels were silver stained and examined for any conformation shifts compared to wild-type controls. Multiple control samples containing established polymorphisms were used to aid determination of whether variants constitute polymorphisms or mutations specific to CPSID patients. All samples creating conformational shifts were sequenced to determine the exact mutation.

RT/PCR analysis of Patient RNA. For every available patient RNA source, reverse transcription and amplification of the resulting cDNA was performed in four overlapping PCR reactions. Patient RNA was obtained by standard cesium chloride gradient separation or Trizol Reagent (Invitrogen). RNA isolations were from hepatic tissue, if available, or lymphoblastoid or fibroblast cell lines established from the patient. Though these cell types do not express detectable amounts of the CPSI enzyme, sufficient transcription occurs to amplify intact *CPSI* transcripts [25]. Reverse transcription was carried out using 1 µg of total RNA and either an oligo-dT primer or L2894 (5'GTGAGCCCAAGGCATTT3'), an antisense primer from the midpoint of the

CPSI message. Using the RT product as template, PCR reactions were performed with four different primer sets creating four large overlapping fragments spanning the 4600 base coding region (U122 – 5'AAATGACGAGGATTTTGACA3' and L829 -- 5'CTAGCAGGCGGATTACATTG3' amplify exons 1-8; U811 -- 5'AATGTAATCCGCCTGCTAGT3' and L2335 – 5'CAATGAATGCCAATGGGTAG3' amplify exons 8-18; U2229 – 5'CCTTCATCCTACCTCAATGG3' and L3723 – 5'ACCTGCATCTTCAACATGTT3' amplify exons 18-30; U3182 – 5'TGAGCACAGACTTTCATGAG3' and L4674 -- 5'ACATGTGGCTCAGGTTGATT3' amplify exons 25-38. See Figure 2.1. The products amplified in the PCR reaction were then sequenced and mutations located based on comparisons with the known *CPSI* cDNA sequence as well as comparisons with mutations found in the corresponding patient gDNA. Amplification was repeated in all cases to avoid PCR-induced errors.

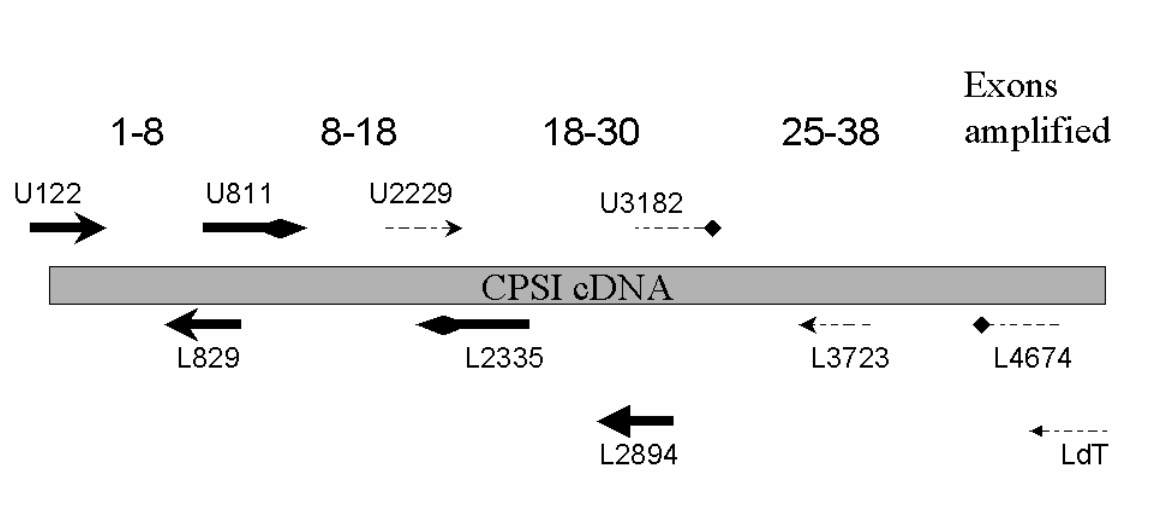


Figure 2.1. Location of RT/PCR primer sets. Primer pairs are denoted by matching arrowheads. Primers used for PCR following RT with L2894 are denoted by a solid tail. Primers used for PCR following RT with oligo LdT are denoted by a dashed tail.

Sequencing and Densitometry. Samples exhibiting shifts in SSCP as well as all RT/PCR products were subjected to dideoxy-sequencing using radio-labeled terminators (Amersham Biosciences). For gDNA, sequencing was always performed in parallel with wild-type controls. The cDNA sequence was routinely examined for the presence of each identified gDNA mutation, intron inclusions, and exon exclusions. To determine the relative representation of each allele in cDNA sequence, densitometry was performed using the QuantityOne platform (BioRad) on high-resolution autoradiograph images.

Northern Blotting. Northern blotting was performed on two patient hepatic samples that had been flash frozen immediately upon extraction (at the time of transplantation or death). Total RNA was obtained through cesium chloride gradient isolation and poly-A enriched RNA was obtained through a subsequent oligo-dT cellulose extraction. A quantity of 10 µg total RNA and 5 µg poly-A RNA was used for gel electrophoresis in a 1.1% agarose/2.2 M formaldehyde gel. The gel was stained with ethidium bromide for photodocumentation and then subsequently transferred to a Hybond N+ membrane (Schleicher and Schuell). CPSI probes were either a 2400 bp fragment from the 3' translated region of the CPSI message or a 700 bp fragment from the 5' translated region, each labeled by nick-translation with α^{32}P dCTP. After probe hybridizations, the membrane was washed in 2X SSC/0.1% SDS at room temperature and in 0.1xSSC/0.1%SDS at 42ºC and 50ºC prior to autoradiography. To control for RNA quality, the same blot was sequentially stripped and re-probed with radio-labeled cDNA probes for α-1-antitrypsin and the intronless HSP-70 gene.

Results

Mutation Detection. In a collection of approximately 100 patients, both DNA and RNA were available for mutation screening from 26 patients, most with the neo-natal onset form of disease (Table 2.1). In these patients, 44 mutations were identified (Table 2.1). Allele heterozygosity for every patient was confirmed in gDNA; in most cases two disease alleles were identified and in cases where a second mutation was not found, the patient was heterozygous at several identified polymorphisms.

RT/PCR. For all patients, RT/PCR was performed and the products were sequenced to determine if identified gDNA changes were also present in the RNA transcript. RT/PCR sequencing revealed that many mutations, though heterozygous and equally represented in gDNA sequence, were missing or unequally represented in patient cDNA sequence (Table 2.1). Allele representation in cDNA sequence was categorized as “homozygous” if there was no visible representation of one of the genomically heterozygous disease alleles in cDNA, “heterozygous” if there was no difference in cDNA expression of both alleles, or “unequally heterozygous” if there was an appreciable difference in intensity between the two alleles, yet both were visible on the sequence (Fig. 2.2). Mutations not present or unequally represented in cDNA suggested a negative effect on the RNA transcript stability.

In 21/26 patients the cDNA was given either an “unequal heterozygous” or “homozygous” score, representing 40% of the total alleles screened in this study. These alleles, listed in the dark gray shaded portion of Table 2.1, were all placed

in the “Allele B” category for the purpose of comparison. To quantify the allele representation on each RT/PCR sequence, densitometry was used to calculate the percent expression of each Allele B mutation when compared to Allele A (set at 100% expression). The genomic alleles not represented or weakly represented in cDNA sequence were primarily nonsense, frameshift, or splicing mutations that cause the formation of a downstream PTC whereas the allele visible on cDNA was typically a missense change. In very few instances, the non-represented sample was a missense or silent mutation (patients 5, 10, and 13). Only in patient 16 were both mutations putative RNA processing defects, confounding the ability to accurately assess the percentage of each allele representation (Table 2.1).

The intronic mutations, each located at the 3’ end of the intron, were identified specifically in CPSID cases (patients 2, 6, and 9). In patient 2, the presence of a faint allele on the cDNA sequence containing a dinucleotide AG insertion at the first two bases of exon 6 confirmed a splicing change. For the other intronic mutations, no specific splicing change was detectable in the cDNA though the patient was homozygous for the alternate allele. A splicing defect was likely for patient 6 because the mutation, located at the last base of the intron, disrupts the conserved 3’ splice site AG dinucleotide. Pathogenicity of the intronic mutations was therefore substantiated by the absence of the mutation in non-CPSID samples, the inability to detect another mutation on that allele, and the disruption of the conserved 3’ splice site sequence.

Table 2.1. CPSID patient mutations.

Pt	Dz Onset	Cell Type	Allele A	Mutation type	Allele B	Mutation Type	Densitometry: B as % of A	cDNA Classification	PTC Position	PTC Exon Location
1	NN*	L	Carrier	---	c.543 C>A	Nonsense	62.6	Unequal Hetero	Y139X	4
2	NN	F	c.2996 T>C	Missense	c.652-3 T>G	Intronic Sub	38.2	Unequal Hetero	+70 bp	6
3	NN	H	c.2271T>A	Missense	c.854 del T	Frameshift	30.9	Homozygous	L243X	8
4	LO	L	N.D.	---	c.854 del T	Frameshift	38.7	Unequal Hetero	L243X	8
5	NN	F	c.2855G>T	Missense	c.1025G>A	Missense	32.7	Unequal Hetero	N.D.	N.D.
6	NN	H	c.2271T>A	Missense	c.1210-1G>T	Intronic Sub	0	Homozygous	N.D.	N.D.
7	LO	L	N.D.	---	c.1318_1319 del AA	Frameshift	21.9	Unequal Hetero	+65 bp	12
8	LO	H	c.4355C>T	Missense	c.1555 C>T	Nonsense	43.2	Unequal Hetero	Q477X	14
9	NN	H	c.3389G>T	Missense	c.1960-8A>G	Intronic Sub	0	Homozygous	N.D.	N.D.
10	NN	L	c.5455T>C	Missense	c.2388C>A	Silent	12.8	Unequal Hetero	N.D.	N.D.
11	NN	F	c.2271T>A	Missense	c.2461 del C	Frameshift	0	Homozygous	+37 bp	19
12	NN	F	N.D.	---	c.3006_3018 del 13	Frameshift	0	Homozygous	+14 bp	24
13	NN	F	c.434 ins GAATGG	Insertion	c.3116C>T	Missense	55.8	Unequal Hetero	N.D.	N.D.
14	NN	L	c.3730T>C	Missense	c.3167 ins G	Frameshift	46.1	Unequal Hetero	+7 bp	25
15	NN	L	c.4255G>A	Missense	c.3308 del A	Frameshift	52.7	Unequal Hetero	+ 111 bp	26
16	NN	F	See row below	---	c.3481_3483 del AA	Frameshift	71.8	Unequal Hetero	+74 bp	28
					c.3907 C>T	Nonsense	17.1	Unequal Hetero	R1261X	32
17	NN	L	c.4450A>G	Missense	c.3907 C>T	Nonsense	52.9	Unequal Hetero	R1261X	32
18	NN*	L	Carrier	---	c.3927_3931 ins 4	Frameshift	48.5	Unequal Hetero	+15 bp	32
19	LO	F	c.4114T>C	Missense	c.4092_4093 ins CC	Frameshift	40.9	Homozygous	+4 bp	33
20	NN	L	c.2536T>C	Missense	N.D.	N.D.	0	Homozygous	N.D.	N.D.
21	NN	F	c.1292T>G	Missense	N.D.	N.D.	0	Homozygous	N.D.	N.D.
22	NN	F	c.2041G>T	Missense	c.4520 ins T	Frameshift	99.2	Heterozygous	+25 bp	38 (terminal)
23	NN	L	c.1189A>G	Missense	c.1289A>G	Missense	95.8	Heterozygous	---	---
24	NN	F	c.2271T>A	Missense	c.2552A>G	Missense	99.9	Heterozygous	---	---
25	NN	H	c.1888G>A	Missense	c.3736G>A	Missense	96.9	Heterozygous	---	---
26	NN	F	c.3067G>A	Missense	N.D.	---	83.0	Heterozygous	---	---

NN = neonatal, LO = late-onset, F = fibroblast, H = hepatic, L = lymphoblastoid, ND = not determined, Gray shading = RNA instability mutation, * denotes child's disease onset

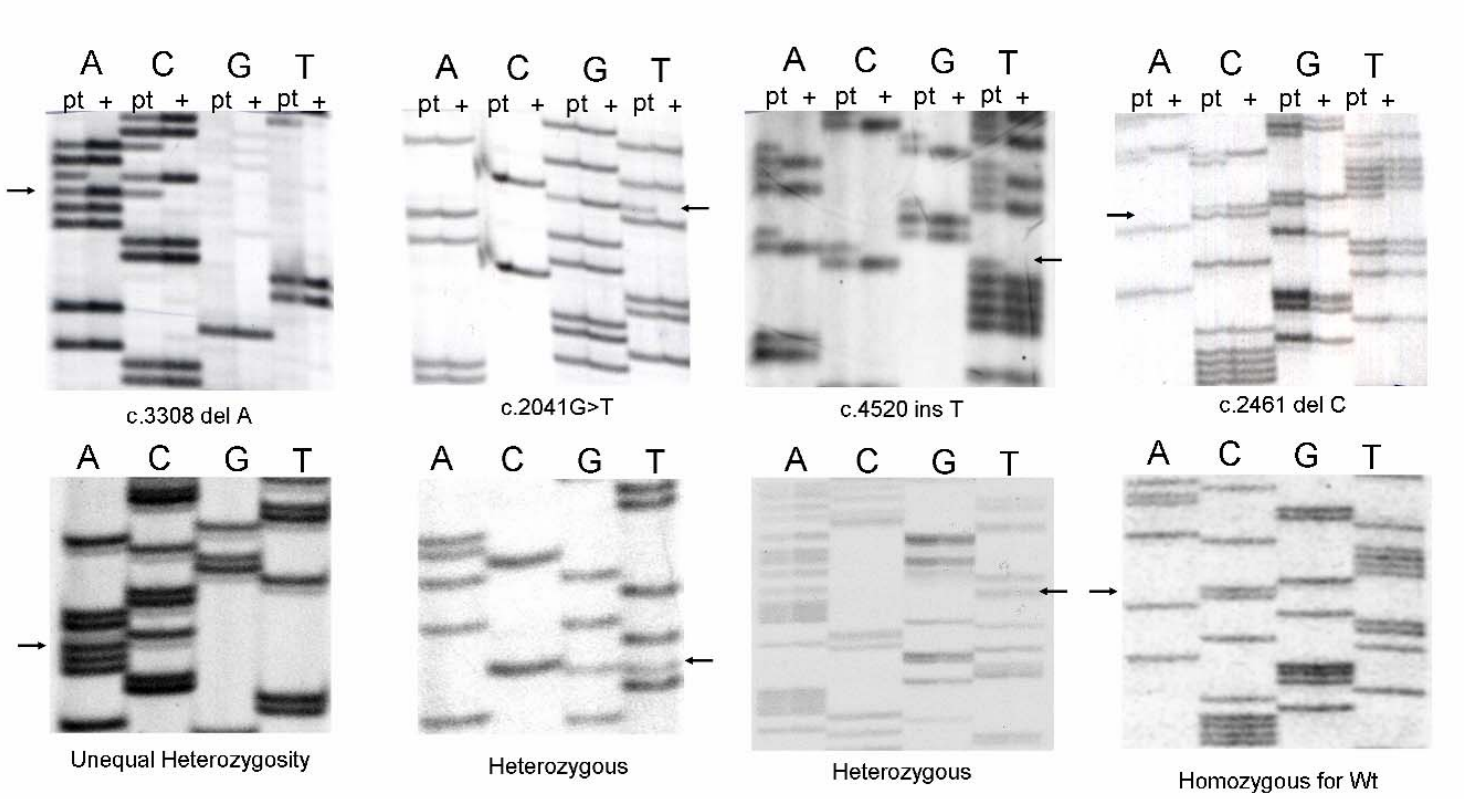


Figure 2.2. Patient gDNA and cDNA sequence depicting differences in cDNA appearance. The top panels show heterozygous genomic DNA sequence with patient DNA (pt) loaded alongside a control DNA sample (+). Arrows identify the mutations. The bottom panels are cDNA sequence showing patient cDNA with arrows identifying the mutation location. A difference in intensity between two visible alleles scores as "unequal heterozygosity." A "heterozygous" score illustrates that there is no difference in cDNA expression of both alleles. When no visible representation of the disease allele is detected, the cDNA is categorized as "homozygous."

In addition to the 19 patients included in the dark gray portion of Table 2.1, there were three patients with disparate gDNA/cDNA appearances. Patients 20 and 21 were confirmed to have heterozygous gDNA based on the identification of one disease allele as well as the presence of alternate alleles for previously-described polymorphisms [12]. Each also was classified as homozygous for the identified disease allele following the cDNA screen, indicating that there is a second, undetermined disease mutation in the allele not represented by cDNA (Table 1, light gray shading). Patient 22 was classified as heterozygous in both the gDNA and the cDNA screen though one disease allele was a frameshift mutation (c.4520 ins T). This frameshift created an aberrant stop codon in the terminal exon, consistent with the inability of aberrant stop codons to elicit NMD when located downstream of the last exon/exon junction. Patients 23-26 were heterozygous for two alleles in both gDNA and cDNA sequences, none of their mutations were frameshifts, nonsense mutations, or apparent splicing defects and would not therefore be expected to affect RNA stability.

There was no observable clustering of mutations towards the beginning or end of the transcript; rather, they extend throughout the entire 4600 bp coding region. There was also no correlation between the cDNA appearance and the tissue type screened, which could have otherwise confounded our conclusions by indicating the possibility that these observations were linked to expression differences between liver, fibroblast, and lymphoblastoid cells. There was also no obvious correlation between the cDNA appearance and the distance of the PTC from the end of the transcript, excluding the last exon.

Northern Blot. Using hepatic RNA obtained from patients 6 and 13, Northern blot analysis was performed to examine CPSI transcripts (Figure 2.3). Patient 6 demonstrated heterozygosity for both the c.2271T>A missense mutation and the c.1210-1G>T intronic substitution mutation in genomic sequence. However, the allele containing the intronic mutation was absent in the patient cDNA sequence. Patient 13 also had heterozygous gDNA but demonstrated unequal heterozygosity in cDNA, as the c.3116C>T missense mutation was underrepresented when compared with c.434ins6 (see Table 2.1). In Figure 2.3A, both the gel stained with ethidium bromide before membrane transfer and then separate hybridizations of HSP-70 and α-1-antitrypsin control probes following transfer demonstrate RNA integrity and equal loading of all samples. This same membrane was then hybridized with a CPSI probe which demonstrated variability among the samples tested. For patient 13 (lane 2), there was a minimal amount of CPSI transcript detected when compared to control livers from a newborn (lane 4) or fetus (lane 1) although equal amounts of RNA were loaded. For patient 6 (lane 3), there was a very small amount of the expected 5.2 kb CPSI transcript and a large continuous smear of smaller RNA fragments. A separate blot, containing an aliquot of the same RNA tested above, was probed with a larger CPSI cDNA fragment that demonstrated degradation from both total and poly-A enriched RNA (Fig. 2.3B). Taken together, these observations indicate evidence of CPSI-specific RNA degradation in both CPSID patients.

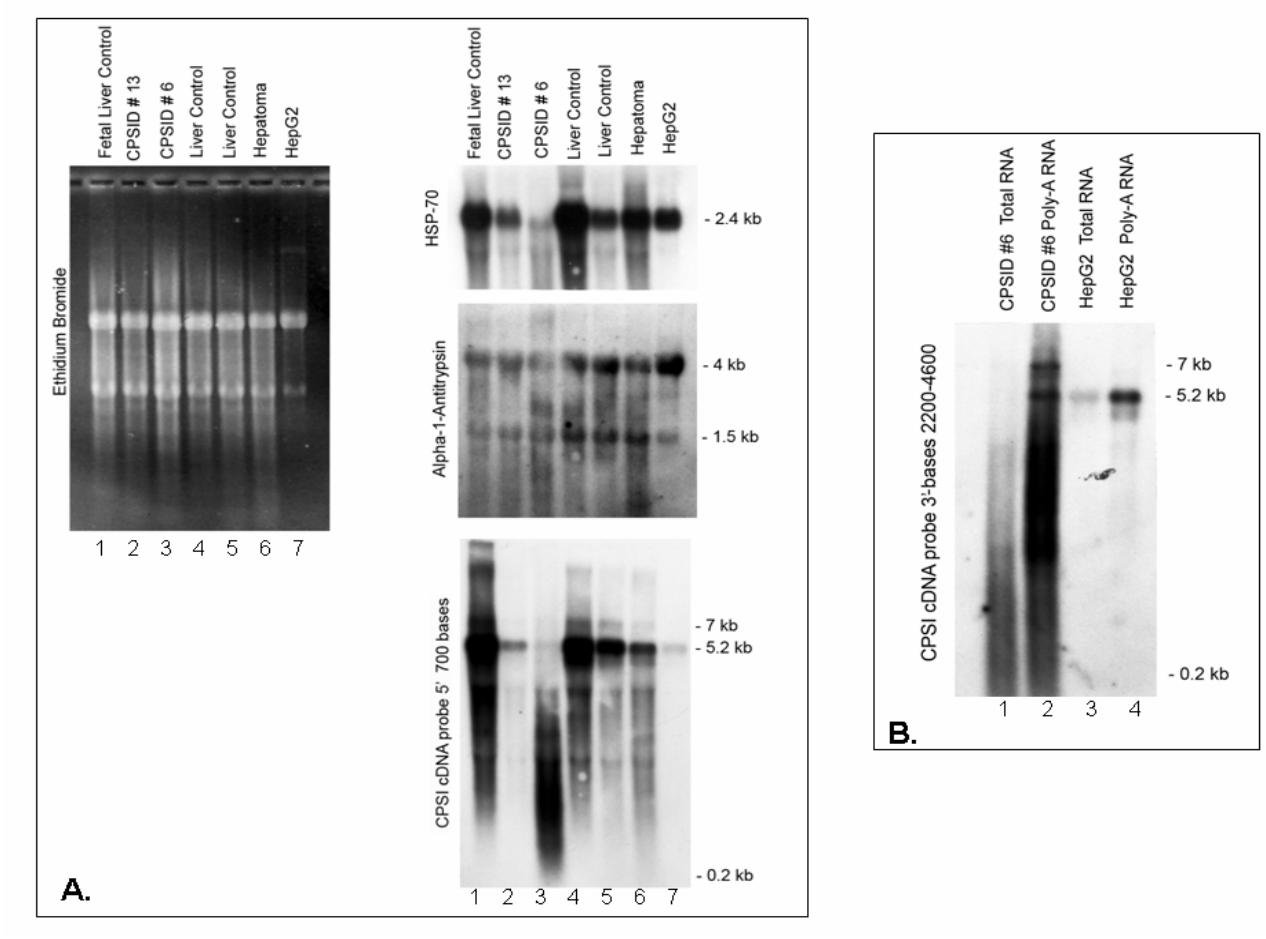


Figure 2.3. Northern Blot of hepatic RNA from patients 6 and 13. A. RNA visualized first by ethidium bromide staining that was then transferred and sequentially probed with HSP70, alpha-1-antitrypsin, and a 5' 700 bp CPSI probe. **B.** HepG2 and patient RNA enriched for polyadenylation from the same preparation as in (A) but on a separate blot was hybridized with a CPSI probe complementary to the 3' end of the transcript.

Discussion

CPSI Deficiency, characterized by a heterogeneous set of mutations, is not only a model for understanding the prevalence of disease-causing mutations that create PTCs, but also an illustration of how the understanding of RNA processing has changed the way mutation pathology should be analyzed. When the Summar laboratory first began the identification of disease causing mutations in CPSID, the absence of the gene structure made it necessary to use patient cDNA, a step that is frequently overlooked in current mutation detection schemes. Though the "missing allele" observation was at first a challenging puzzle, the characterization of the NMD pathway provided clarifications. Following the definition of the CPSI gene structure, the gDNA sequence was examined alongside the cDNA sequence. Screening both not only allowed insight into more precise mutation mechanisms, but also led to an analysis of the considerable number of RNA processing defects found in CPSID patients.

21/26 patients with heterozygous gDNA and a testable RNA source had at least one presumptive RNA processing defect. Factors substantiating this conclusion include the disparities in allelic signal between gDNA and cDNA sequence (Table 2.1) and the CPSI-specific RNA changes seen in patient RNA analyzed by Northern blot (Figure 2.3). There is a direct correlation between mutations resulting in very little or no RT/PCR product and mutations that introduce a PTC (either directly or a point mutation or from a frameshift resulting from an insertion, deletion, or splicing defect). This correlation challenges the possibility of other causes of allelic dropout. Additionally, the identification of

one silent and two missense mutations that were unequally represented in cDNA suggests that they may affect splicing through disruption of an exon splicing enhancer, as described for other genes [123]. Further support for an NMD mechanism was provided by patient 21 because the c.4520 insT mutation, creating a stop codon in the terminal exon, had heterozygous cDNA. This is consistent with previous observations that aberrant stop codons only elicit NMD when located at least 50 bp 5' of the terminal exon/exon junction [65].

Though important in mutation pathology determination, there are a few limitations to studying patient RNA. It is often difficult to obtain patient RNA sources, as evidenced by the laboratory's collection of CPSID patients where only 25% of cases have an archived RNA source. Though this investigation and others have successfully used low-copy transcripts from diagnostically available tissue sources, limitations to this approach have been suggested. Haberle et al [125] showed that patient fibroblast cultures can be manipulated with cyclohexamide to detect NMD on low-level transcripts of NAGS (another hepatic-specific urea cycle gene; see Chapter 1) whereas Bateman et al [126] suggested that relying on collagen X mRNA analysis from a tissue source where collagen X is not expressed possibly results in erroneous conclusions. Such limitations underscore the importance of developing a laboratory model system to allow determination of various disease-causing mutation mechanisms on a uniform background.

The progression of understanding the molecular pathology of CPSID is a chronicle that mirrors the growing understanding of mutation mechanisms in

general. As genetic tools and scientific knowledge have increased over the years, coupled with the description of the *CPSI* genomic structure, clues towards understanding the molecular mechanisms of this disease have progressed. An explanation for the "missing alleles" is that they are an indication of RNA instability, likely the result of degradation through the NMD quality control pathway to ensure that aberrant transcripts are not translated. Mutation discovery using both cDNA and gDNA not only provides a better mutation detection rate, but also affords insight into the molecular mechanism of those mutations. In a screen of 26 CPSID patients, 21 had at least one RNA instability mutation. Out of 52 total alleles, 21 or 40% of the alleles responsible for this disease affect the RNA transcript. As ways become available to test mutations from patients without a RNA source, more mutations of this type will surely be found. Examining RNA instability and NMD in a growing number of diseases, including inborn errors of metabolism, provides evidence that this is a common mechanism leading to genetic deficiencies and highlights the importance of RNA processing defects in molecular pathology.

This Chapter is derived from an article that is also published in *Molecular Genetics and Metabolism* [127].

CHAPTER III

BUILDING THE BECC VECTOR MODEL SYSTEM

Introduction

Examining the functional mechanisms of identified pathogenic alleles in *CPSI* (as well as many other diseases), is often difficult. Frequently, patient tissue is not available for RNA analysis or is not amenable to manipulations necessary for detecting mutation pathology. In addition, many identified *CPSI* mutations are intronic and therefore cannot be tested in cDNA constructs. Because CPSI contains 37 introns and spans over 120 kb of genomic sequence, the use of conventional “minigene” plasmid constructs is also not efficient for screening these intronic mutations. Bacterial artificial chromosomes (BACs) have fewer size constraints than plasmids and allow large genomic inserts, permitting the inclusion of the whole *CPSI* gene on a single construct. Therefore, the goal was to create a unique BAC-based model system that would allow comprehensive, yet efficient, testing of any mutation in *CPSI*, irrespective of type or location. This system should be widely applicable to many experiments geared towards examining both coding and non-coding sequences, and by studying these genetic variants within their wider sequence context, a more accurate determination of the mechanistic effect should be possible. Though it was designed for the study of *CPSI* mutations, its application extends to studies

of other genes, demonstrating its versatility as a platform that decreases the dependence on patient tissue sources.

To create this new expression vector, a BAC clone containing the full *CPSI* gene was modified by two recombineering steps. First, the addition of the pEHG plasmid (see Results) provided elements necessary for stable expression and selection in eukaryotic cells. Second, insertion of a CMV promoter upstream of *CPSI* allowed ubiquitous expression of this hepatic-specific gene. This completed vector, named BECC (Bac+pEHG+CMV+*CPSI*), was then tested for expression in eukaryotic cells. It was necessary to determine conditions that would allow efficient transfection into eukaryotic cells, stable cell line generation, and exogenous gene expression. This Chapter details these BAC manipulations and experiments performed to ensure proper integrity and function of this novel BAC-based model system for studying genetic mutations.

Materials and Methods

BAC recombineering. The pEHG plasmid (see Results) was inserted into the backbone of BAC RP11-349G4 using Cre/*lox*P recombination, as performed previously for other BAC constructs [128]. This method takes advantage of the single *lox*P sequence in the backbone of all BAC vectors and the single *lox*P sequence in the pEHG plasmid. To prokaryotic cells containing the BAC and induced to express the Cre recombinase, the pEHG plasmid was electroporated and recombinant colonies were selected based on both chloramphenicol (gene

located on the BAC) and ampicilin resistance (gene located on the pEHG plasmid).

CMV vector construction and homologous recombination. To insert the human CMV immediate-early promoter (Towne strain) upstream of *CPSI* on the BAC+pEHG vector, the pCMV/Bsd plasmid (Invitrogen) was first modified. A *Kpn*I restriction enzyme site was engineered at position 1937 by site-directed mutagenesis (Stratagene QuickChange Kit). Then, *Kpn*I and *Xho*I were used to create a complementary insertion site for DNA fragment containing the tetracycline resistance gene flanked by FRT sites from plasmid pTet/Frt [129]. A 5' homology arm was then synthesized (Invitrogen) containing 50 bp of homologous sequence to the BAC backbone as well as *Not*I and *Nhe*I overhangs to direct insertion directly 5' of the Tet/FRT sequence. A 3' homology arm was synthesized (Invitrogen) to contain 50 bp of homologous sequence to the *CPSI* gene beginning with the transcription start site of *CPSI* (AF154830) as well as *Rsr*II and *Xma*I overhangs to direct insertion downstream of the CMV promoter. Sequencing of the newly created pCMV/Bsd+Tet+HomArm construct was performed to ensure its integrity. All restriction enzymes used in this protocol were purchased from New England Biolabs. Homologous recombination between the BAC+pEHG construct and the newly created pCMV/Bsd+Tet+HomArm construct was possible because the BAC+pEHG construct had previously been electroporated into the *E. coli* EL250 strain which can be induced for high levels of recombination [130]. A monoclonal colony containing BAC+pEHG was induced for recombination by a temperature shift

from 32ºC to 42ºC followed by electroporation of the constructed pCMV/Bsd+Tet+HomArm vector. Recombinants were selected by growth on LB+tetracycline and confirmed to have undergone homologous recombination by direct sequencing. A monoclonal colony was then treated with arabinose for induction of the Flp recombinase, which will cause removal of the tetracycline resistance gene. Following arabinose induction, colonies were grown on LB-tetracycline plates and replica plated onto LB+tetracycline plates. Colonies growing only on the –tet plates were selected for direct sequencing to confirm removal of the tetracycline cassette. The newly created BAC+pEHG+CMV+*CPSI* construct (BECC) was purified using the NucleoBond BAC Maxi AX500 kit (BDBiosciences).

Pulse field gel electrophoresis (PFGE). PFGE was performed following a 3 h digestion of BAC DNA at 37°C with *Not*I (New England Biolabs) in a 1% agarose 0.5% TAE 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. Gels were stained with ethidium bromide following electrophoresis.

Fingerprinting. This quality control assay was performed by digesting BAC DNA with a *Bam*H1 restriction enzyme (New England Biolabs) for 3 h, then running the products on a 1% agarose 1% TBE gel for 16 h at 35 v. Because the BAC construct contains many *Bam*HI recognition sites, multiple, smaller bands will be present on this gel when compared with a PFGE, creating a unique "fingerprint." Both PFGE and fingerprinting provide verification that no unwanted global deletions or rearrangements occurred during BAC recombineering.

SSCP. Because PFGE and fingerprinting have a low sensitivity in detecting smaller deletions or nucleotide substitutions, single strand conformation polymorphism (SSCP) analysis was used to detect the presence of small mutations in the coding region of *CSPI* on the BAC construct. For this method, all exons as well as the surrounding intronic sequence were amplified in multiple PCR reactions for both the BECC+WtCPSI construct and a wild-type *CPSI* gDNA control. These PCR products were then denatured by incubation at 95°C for 3 min and then quickly shifted to 4°C to induce the formation of single stranded globular molecules that fold based on their specific nucleotide sequence. Any change in sequence causes a change in the globular structure that is detectable as a band shift following electrophoresis in a non-denaturing MDE gel run at 15 w for 10 h. Gels were silver stained to visualize the DNA products. For the BECC+WtCPSI construct, all exons were screened as well as the 5' and 3' untranslated regions in a total of 42 PCR reactions. Direct sequencing was performed instead of SSCP for 12/42 PCR reactions to increase mutation detection rates. In addition, changes between the wtCPSI vector and the control gDNA sample detected by SSCP were analyzed by direct sequencing and all changes were shown to be different alleles of previously reported polymorphisms [12].

Transfections and stable cell line generation. The BECC+WtCPSI construct was transfected into the MRC-5V2 human lung fibroblast immortalized cell line [131] using the LID (Lipofectin, Integrin, DNA) transfection technique [132]. The transfection reaction was set up using 8 µl Lipofectin (Invitrogen), 8

μg Peptide 6, and 4 μg of the BECC constructs in a 2L:2I:1D ratio. Lipofectin was diluted in 400 μl Optimem (Gibco) and incubated 45 min before addition of peptide 6 and DNA, also diluted in 400 μl Optimem. Reactions were incubated 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 drop-wise placement on 5 cm dishes containing MRC-5V2 cells at 60% confluence. Following a 16 h incubation, the transfection media was replaced with DMEM+10%FBS for 72 h. After 72 h, hygromycin selection was added at a concentration of 125 μg/ml. After 2 weeks under hygromycin selection, cells were flow-sorted for GFP expression. All GFP+ cells were cultured under hygromycin selection as a polyclonal cell line.

Sequencing. Fluorescent sequencing was performed by GenHunter Corporation and the Vanderbilt University Medical Center Core Facility using BigDye chemistry from Applied Biosystems. Radioactive di-deoxy sequencing was performed using the Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (USB) according to the manufacturer protocol with 200 ng DNA.

RNA isolation and non-quantitative RT/PCR. To isolate RNA, the RNeasy Midi Kit including the optional RNase free DNase set was used following the manufacturer's protocol (Qiagen). Reverse transcriptions were carried out using 1 μg of total RNA, the CPSI-specific lower primer L2894 (5'GTGAGCCCAAGGCATTT3'), and the AMV-RT enzyme (Promega). The amplification thermoprofile was an initial denaturation of 65°C for 5 min, then a primer-annealing phase at 27°C for 10 min, and an extension phase at 42°C for

45 min. A region of *CPSI* was amplified using these cDNA templates for confirmation that *CPSI* transcription was occurring from the BAC construct. PCR primers used were U122 (5'AAATGACGAGGATTTTGACA3') and L829 (5'CTAGCAGGCGGATTACATTG3'). Products were visualized on a 2% agarose gel. PCR primers spanned multiple exons to eliminate gDNA contamination.

Western Blotting. Cells were lysed by sonication in a 50mM Hepes Buffer with 0.01% Triton X-100. Cell lysates were quantified using the BCA Protein Estimation Kit (Pierce). Following a 5 min denaturation step at 95ºC in SDS sample buffer (375 mM Tris; 4% SDS; 20% glycerol; 10% 2-mercaptoethanol), equal µg quantities of cell lysates were electrophoresed in a 6% PAGE (polyacrylamide) gel for 45 min at 200 v. Proteins were transferred to a nitrocellulose membrane using an electrical tank transfer system at 4ºC for 90 min at 100 v. The membrane was then incubated in blocking buffer (5% nonfat dried milk in 0.01M Tris; 0.5M NaCl; 0.05% Triton X-100; 0.2% Tween; pH 7.4) for 30 min at room temperature. The anti-CPSI (Santa Cruz, sc-10516) primary antibody was added to the blocking buffer for overnight incubation at 4°C at a final dilution of 1:50. The membrane was washed 3 times for 10 min each in blocking buffer before incubation for 1 h in blocking buffer with the bovine anti-goat IgG-AP (Santa Cruz, sc-2351) secondary antibody (1:2000 dilution). The membrane was again washed 2 times in blocking buffer for 10 min each, then 2 times in AP buffer (alkaline phosphatase, 0.1M Tris; 0.1M NaCl; 0.005M $MgCl_2$; pH 9.5) for 10 min each. These steps were all performed at room temperature. The results were visualized through alkaline phosphatase staining. The washed

membrane was incubated in a 1:1 mixture of the Alkaline Phosphatase Conjugate Substrate Kit (Bio-Rad) for 10 min for colorimetric band development.

Results

BECC vector engineering. To create the BECC (Bac+pEHG+CMV+*CPSI*) vector, two modifications were necessary. First, cre/*lox*P recombination was used to retrofit BAC RP11-349G4 with vector pEHG (Figure 3.1) [133]. This vector contains the episomal retention elements (oriP and Nuclear Antigen I or EBNA-1) from Epstein-Barr virus, the hygromycin resistance antibiotic selectable marker, and the visual marker enhancer green fluorescent protein (E-GFP). Figure 3.1B is a pulse-field gel electrophoresis illustrating the insert of one copy of the 11k kb pEHG plasmid following retrofitting. Second, a homologous recombination strategy was utilized to insert the CMV promoter directly upstream of *CPSI* to allow for ubiquitous expression of this hepatic-specific gene. The CMV promoter was targeted for specific insertion using homologous recombination with a Tet-CMV targeting construct flanked by BAC homology segments in recombination-induced EL250 cells (Figure 3.2) [130]. Direct sequencing verified proper insertion of the CMV promoter onto the BAC construct. Following these modifications, the completed BECC construct (Figure 3.3) was 170 kb in size where 150 kb was gDNA from chromosome 2 containing the complete *CPSI* gene, 11 kb was from the pEHG plasmid, and 9 kb was from the BAC vector backbone.

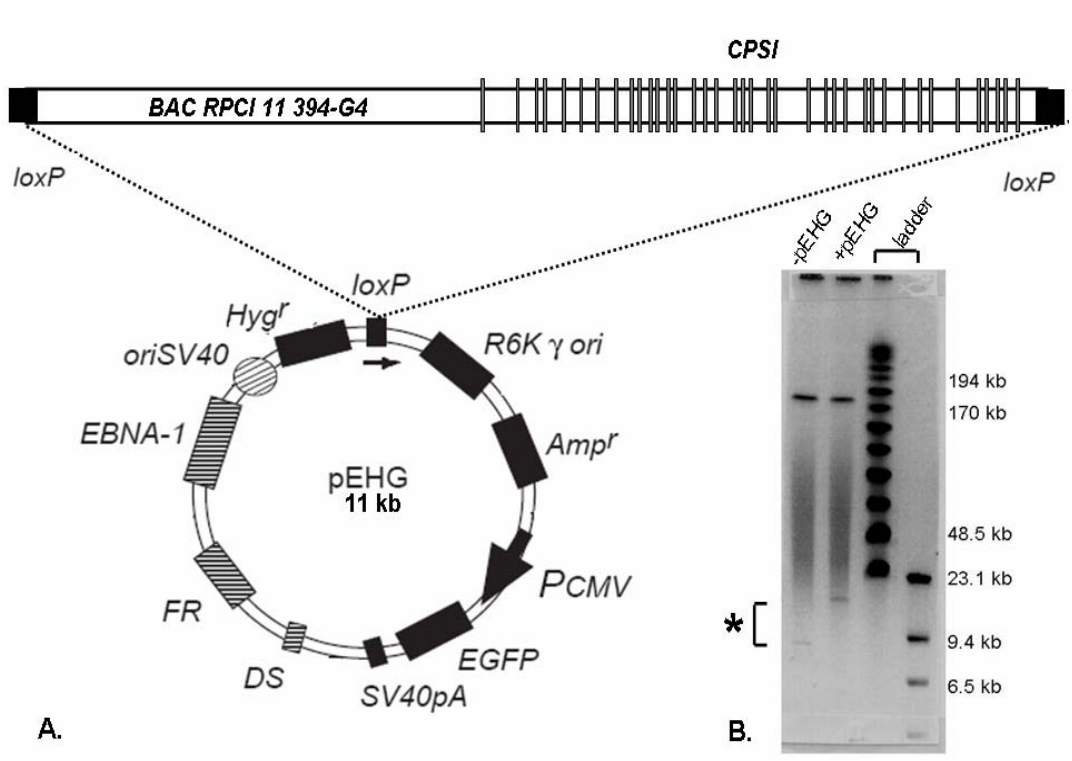


Figure 3.1. Addition of pEHG to the BAC containing CPSI. A. A schematic representation of the pEHG addition to the RPCI11-349G4 BAC construct. Using the *lox*P sites on the pEHG vector and the BAC allowed sequence-specific integration of these two vectors. **B.** A pulse-field gel electrophoresis of *Not*I digested constructs before (-) and after (+) the addition of a single copy of pEHG following cre/*lox*P recombineering in the BAC. The (*) indicates the 11 kb size change in the BAC library vector corresponding to pEHG.

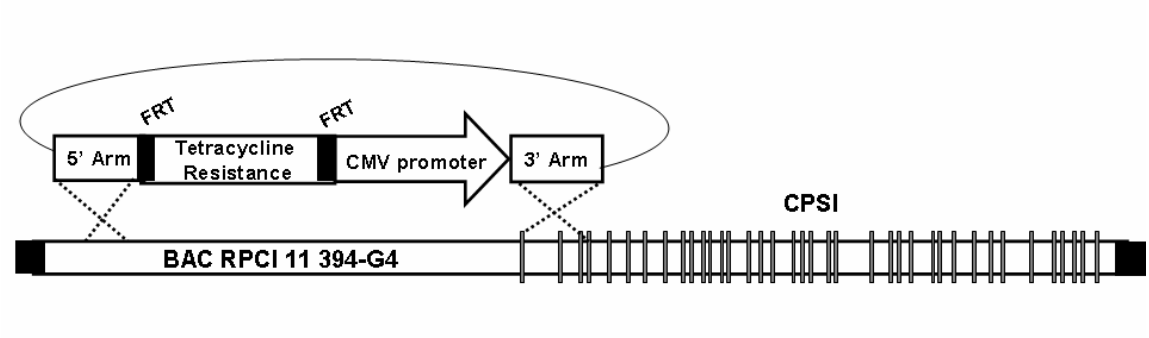


Figure 3.2. Addition of the CMV promoter to the BAC containing CPSI. The homologous recombination vector was used to simultaneously insert the CMV promoter upstream of CPSI and delete the genomic sequence upstream of CPSI. The tetracycline resistance gene flanked by FRT sites was used for positive selection of the recombination event, but was then removed from the final BAC construct.

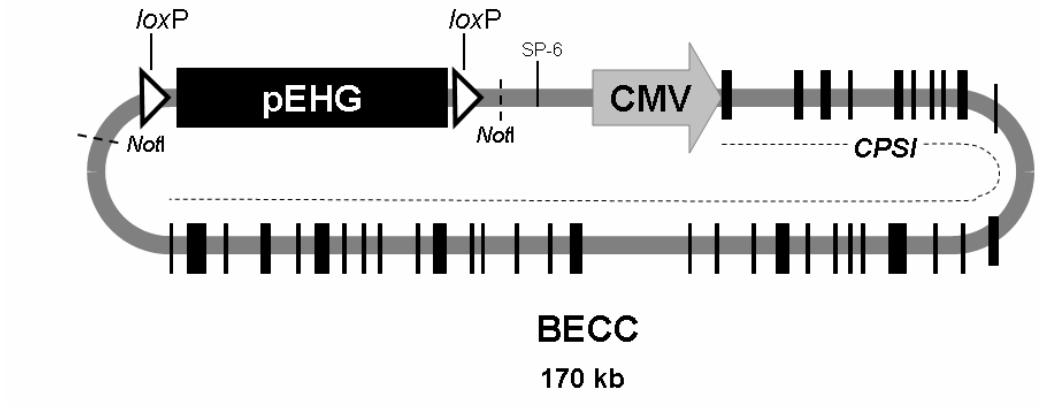


Figure 3.3. The BECC vector.

Numerous quality control assays ensured that the BAC construct did not undergo any unwanted or global rearrangements during the retrofitting and homologous recombination steps. PFGE provided the most general picture of the general integrity of BECC following the addition of the CMV promoter (Figure 3.4A). A fingerprint gel provided confirmation that only expected differences in the construct before and after CMV promoter insertion were present (Figure 3.4B). Direct sequencing of the bases surrounding the CMV promoter provided the most specific verification that the correct modifications were made (Figure 3.4C).

BECC vector expression. Following verification of the integrity of the new BECC+WtCPSI vector, the next step was to examine its behavior in cultured eukaryotic cells. It was transfected into MRC-5V2 cells, an immortalized human lung fibroblast line without measurable CPSI activity [131] using LID transfection (see Materials and Methods) [132]. The wtBECC construct demonstrated stable transfection with an initial efficiency of 20% as determined by visual inspection of GFP expressing cells. Stable transfectants were selected by culturing for two weeks under hygromycin selection (125 μg/ml) followed by GFP+ selection through flow-sorting. During a 1.5 year period, this MRC+WtCPSI polyclonal cell line was passaged 85+ times under hygromycin selection with no measurable loss of GFP expression by repeated visual inspections (Figure 3.5).

Multiple assays were performed next to determine the presence of exogenous *CPSI* transcript and mature protein in the MRC+WtCPSI cell line. A reverse transcription/PCR using RNA extracted from both GFP+ and non-

transfected MRC-5V2 cell lines revealed the presence of *CPSI* transcript only in the MRC+WtCPSI cell line and the hepatic-derived HepG2 cell line (Figure 3.6A). *CPSI* transcript was undetectable in the non-hepatic Cos7 or non-transfected MRC5-V2 cell lines. A Western blot reveals the presence of exogenous CPSI protein specifically in the MRC+Wt*CPSI* GFP+ cell line (Figure 3.6B). Taken together, these results indicate that following transfection, the BECC+WtCPSI construct creates stable transformants that not only express GFP and are resistant to hygromycin, but also produce *CPSI* transcript driven by the CMV promoter and translated into a mature protein.

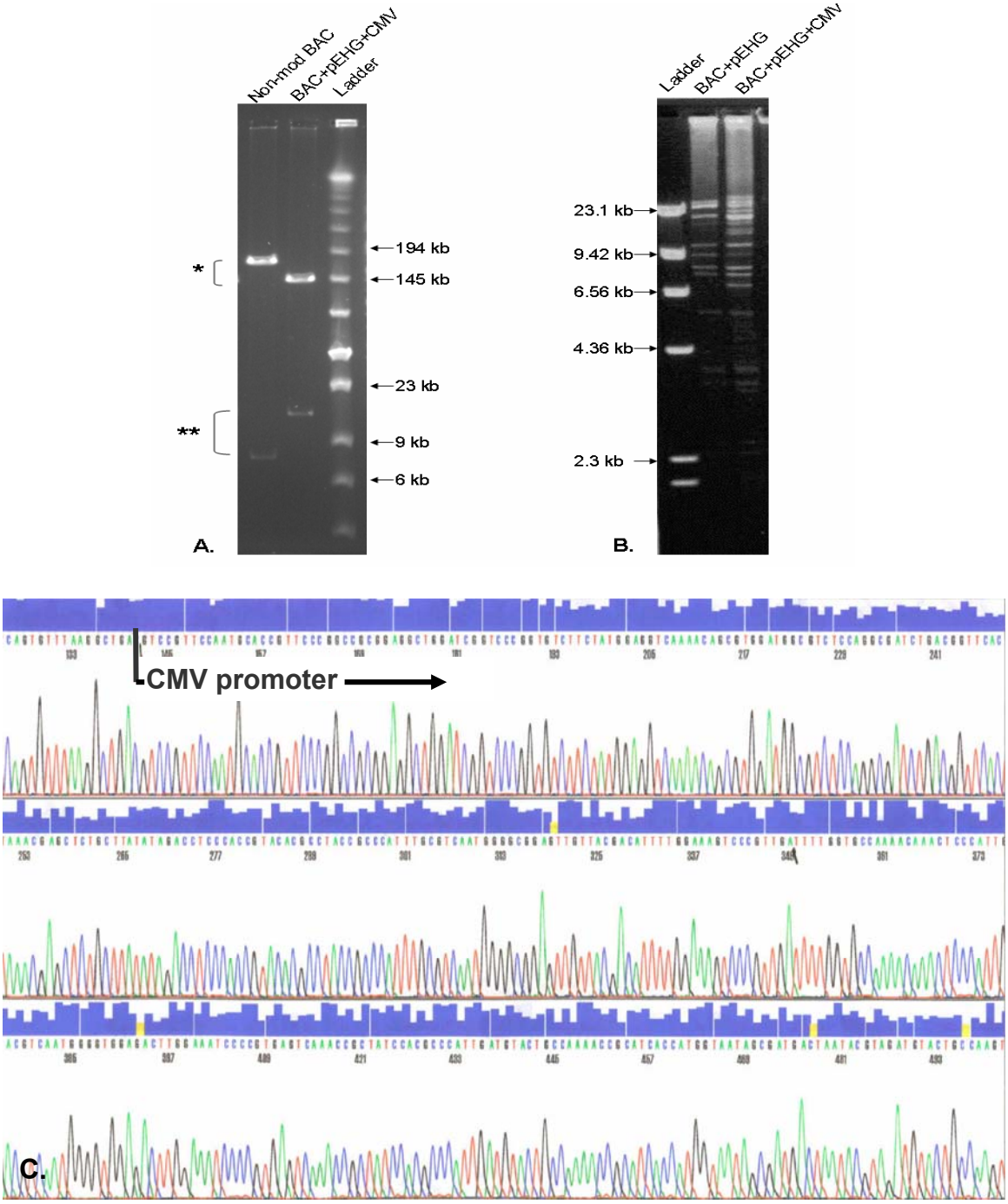


Figure 3.4. Verification of BECC recombineering. A. A PFGE showing the non-modified BAC 349-G4 clone and the BAC following both recombineering steps. The size difference in the upper band (*) is due to the deletion of the sequence upstream of CPSI and the size difference in the lower band (**) is due to the addition of pEHG. **B.** A fingerprinting analysis following *Bam*HI digestion of BAC+pEHG and then following the simultaneous deletion of the sequence upstream of CPSI and addition of CMV. **C.** Direct sequencing verifies the proper insertion of the entire CMV promoter sequence.

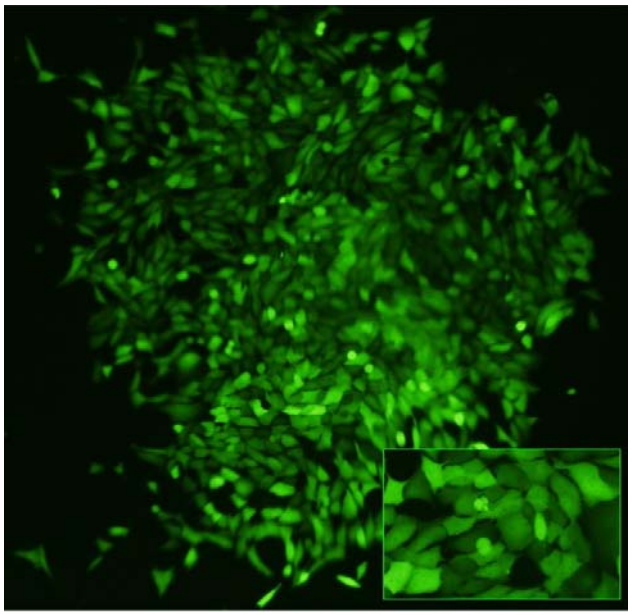


Figure 3.5. GFP expression in transfected MRC5-V2 cells. Following transfection of the wt*CPSI* BECC construct and 2 weeks under hygromycin selection, colonies of 100% GFP expression were visible. The inset is a high magnification of one such colony.

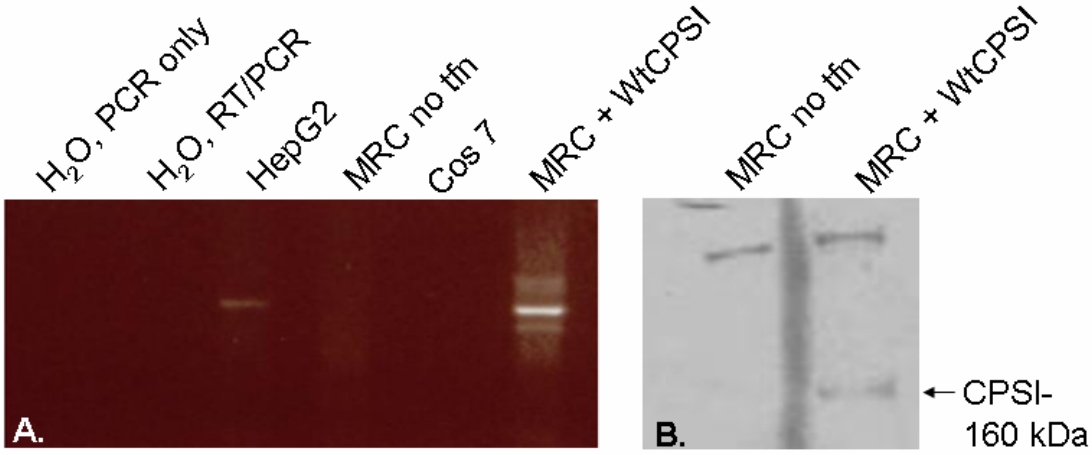


Figure 3.6. Exogenous *CPSI* transcript and protein expression. A. RT/PCR products generated with CPSI-specific primers visualized on a 2% agarose gel reveal that only the liver-derived HepG2 and transfected MRC5-V2 cells contain *CPSI* transcript. **B.** Western blotting shows the presence of CPSI protein only in the transfected MRC cell line. The 200 kDa band detected by the CPSI antibody, provisionally identified as CPSII due to its size and high homology of CPSII to the CPSI antibody recognition sequence, serves as a loading control.

Discussion

As more genes are identified in association with clinical phenotypes, understanding the effects of diverse types of genetic changes has become a difficult, yet vital, undertaking. The complexity of dissecting phenotypes is underscored by the observations that even mutations responsible for rare single-gene disorders such as CPSI Deficiency can result in the same disease phenotype through different mechanisms. To aid CPSID studies and delineate the molecular function of identified genetic variants, the BECC vector was created using a BAC clone and a modular insert. This vector was further modified to allow *CPSI* to exhibit ubiquitous expression under the control of the CMV promoter. This second modification was necessary due to the few available alternatives for cell lines that express *CPSI* using its native regulatory elements and because hepatic cell lines have the confounding factor of expressing endogenous *CPSI*.

In developing this model, the goal was to overcome limitations in classical methods for assessing mutation function. It was important to find a system that would allow the presence of the complete *CPSI* gene on a single construct. Because *CPSI* contains 38 exons and spans 160 kb of gDNA, it was necessary to create a system with a greater efficiency than conventional methods that have significant size restrictions and prevent the mutation from being tested in its complete physiological context. Plasmids containing cDNA sequence are not feasible for assessing mutations that may affect RNA processing. Because introns have been shown to influence multiple mRNA processing steps, the

importance of their inclusion extends beyond testing intronic mutations [134]. Exclusion of introns from a construct prevents the detection of those mutations that may affect splicing or elicit NMD, which has been shown to rely on the presence of introns in mammalian cells [80-82]. “Minigene” construction, though allowing intron inclusion, is not feasible for large genes such as *CPSI* that contain mutations in almost every exon and most introns.

Several BAC characteristics were important in developing this model. BAC constructs can be shuttled between prokaryotic cells for manipulation and eukaryotic cells for expression. In a single modification, a library BAC clone can be retrofitted with pEHG, an 11 kb vector containing all the necessary elements for long-term retention, selection, and tracking of the construct in mammalian culture cells. The resulting modified BAC contains the EBV (Epstein-Barr virus) elements OriP and EBNA-1, which allow for stable transfectants and episomal maintenance of the transfected vector. The EBNA-1 protein has previously been shown to be the only *trans* element necessary for the replication of constructs containing the OriP DNA sequence, to which the protein binds (reviewed in [135]). These two elements have been successfully exploited in the construction of other episomally replicating eukaryotic vectors and gene therapy experiments [136-138].

Other means for transferring and expressing large segments of exogenous DNA in eukaryotic cells have previously been developed, many with an emphasis on gene therapy. These methods include the delivery of BACs using viral [139,140] and bacterial transfer [141], as well as the use of YACs

(yeast artificial chromosomes) [142], and HACs (human artificial chromosomes) [143]. In contrast to generating a HAC vector from a BAC, which requires the addition of ~70 kb of alphoid DNA, creation of the BECC vector only required the addition of the 11 kb pEHG vector. Such gene therapy systems focus on efficient gene packaging and intact delivery rather than the facilitation of functional mutation studies that must be amenable to site-directed mutagenesis. Therefore, the goal was to create a simple non-viral BAC-based expression system that can be manipulated by precise methods of homologous recombination and will allow expression of wild-type and mutant genes from genomic DNA inserts in cultured cells.

The possibilities of using a BAC-based model system like BECC are only limited by the identification of a BAC clone containing the complete gene of interest and an appropriate cell line for vector introduction and expression. However, given the recent improvements in BAC engineering, it is feasible to modify a BAC clone to meet size specifications or join gene segments from multiple BACs together [144]. Furthermore, there are a variety of transfection methods that can be used to introduce large constructs into a spectrum of cell types [132,145,146]. As shown here, promoter swapping permits gene expression in an otherwise non-expressing cell type.

This Chapter outlines the recombineering methods used to create the BECC vector. Selection of constructs with successful cre/*lox*P and homologous recombination events was aided by the presence of selectable markers and numerous quality control assays. Following construction of the BECC vector, this

Chapter outlines the transfection method used to introduce the construct into eukaryotic cells for the creation of a polyclonal, stable cell line, again aided by selectable markers. Because this polyclonal stable cell line produced detectable exogenous *CPSI* transcript and protein, the next goal was to utilize this system to test multiple rare variants identified in the *CPSI* gene of CPSID patients to determine the mechanistic effect of these mutations.

CHAPTER IV

THE STUDY OF CPSID MUTATIONS IN THE BECC VECTOR

Introduction

In conjunction with the development of the BECC *CPSI* expression vector, it was essential to determine if CPSI mutations could be efficiently engineered into BECC and subsequently ascertained for changes in *CPSI* expression. To this end, four putative RNA instability patient mutations were identified with the goals to first determine if the BECC platform could recapitulate the RNA defect originally seen in the patient and second, to determine if it could help elucidate the molecular pathology of mutations without an available patient RNA source. Each chosen mutation was a single base substitution and candidate RNA processing mutation. Two were exonic mutations: c.1893T>G creates a stop codon in place of a tyrosine codon in exon 16 and c.2388C>A was a synonymous change in exon 19 unique to the patient with disease. The other two mutations were intronic and located at the 3' splice acceptor site: c.652-3T>G in intron 5 and c.1210-1G>T in intron 10. It was important to study these mutations not only to ascertain their molecular pathology, but also to validate the BECC vector model system (see Chapter III).

Materials and Methods

Site-directed mutagenesis. To individually introduce each point mutation into the BECC construct,a two-step homologous recombination procedure with positive and negative *galK* selection was used [147]. See Appendix A for a diagram of this two-step recombination procedure and the primer pairs used. The PCR thermoprofile used to amplify *gal*K was an initial denaturation step of 94ºC for 3 min, then 30 cycles of 94ºC for 15 s, 60ºC for 30 s and 70ºC for 1 min, followed by a final extension time of 10 min at 72ºC. To create double-stranded oligos with the desired mutation for homologous recombination with the BAC construct, 20 ug of each single-stranded oligo was added to a final 1x concentration of annealing buffer (0.1M NaCl, 1 mM EDTA pH 8.0, 10 mM Tris pH 7.5) and heated at 95ºC for 5 min, then cooled to room temperature and gel purified. Following the *gal*K mutagenesis procedure, BAC purifications were performed using the NucleoBond BAC Maxi AX500 kit (BDBiosciences).

Verification of BAC engineering. See Chapter III for an explanation of PFGE, fingerprinting, sequencing and transfections. Each of these protocols was performed in conjunction with developing the original BECC vector as well as following the *galK* site-directed mutagenesis to ensure vector quality.

Plasmid construction. To create a “minigene” for the study of c.1210-1G>T, *CPSI* exon 10, intron 10, and exon 11 were amplified in a PCR reaction using primers that contained restriction sites for *Hind*III (5’) and *Eco*RI (3’) as well as an ATG start codon immediately upstream of exon 10 and a TAA stop codon immediately downstream of exon 11 (underlined in primer sequence). UIntron10:

5'CCCGGGAAGCTT**ATG**AGAGGGCAGAATCAGCCTGTTTTG3' and LIntron10: 5'GGGCC**TTA**AGAATGAGTCACAGATAACCGGGGC3'. The PCR reaction was performed with 100 ng of gDNA in EasyStart 50 ul pre-aliquoted PCR reaction tubes (Molecular Bioproducts). The amplification thermoprofile was an initial denaturation at 94ºC for 4 min, then 28 cycles of 94ºC for 1 min, 62ºC for 1 min, 72ºC for 1 min 15 s, and a final extension time of 10 min at 72ºC. Gel purified PCR products (Wizard SV40 PCR Cleanup Kit from Promega) and the pcDNA3.1(+) plasmid (Invitrogen) were digested with *Hind*III and *Eco*RI (NEB) according to the manufacturer's protocol for 2 h at 37°C. Ligations were carried out using the Quick Ligation Kit (NEB) according to the manufacturer's recommendations at 3-fold molar ratio of insert to vector determined to be 30 ng insert : 50 ng vector. Electrocompetent transformation of DH10B cells with 1 µl of the ligation product was performed prior to plating the cells onto LB+Ampicilin (60 ug/ml) for overnight growth at 37°C. Colonies were screened for the proper plasmid insert by overnight growth in 2 ml LB with Ampicilin selection, then plasmids were isolated using a MiniPrep isolation kit (Qiagen) and subsequently screened by a diagnostic restriction digest with *Afl*II or *Sal*I using the same restriction digest protocol as above. Direct sequencing provided verification of proper cloning. Site-directed mutagenesis was then performed using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) according the published protocol to introduce the G>T point mutation at c.1210 which was verified by direct sequencing. Oligos were UGIntron10T-SDM: 5'CTTCCTGTTTCTTATTCCTTTA**T**GGGATTATGCATGAGAGC3' and

LGIntron10T-SDM:

5’CGAAGAAGGGTTTGCTCTCATGCATAATCCC**A**TAAAGGAATA3’.

gDNA isolations and PCR. The Wizard Genomic DNA Purification Kit (Promega) was used to isolate gDNA from all cell lines. To determine the presence of BAC DNA in transfected cell lines, PCR was performed using a forward primer complementary to the CMV promoter, UCMV (5’CCATCCACGCTGTTTTGACCTC3’) and a reverse primer complementary to sequence in the first exon of *CPSI*, L173 (5’CCAGTCTTCAGTGTCCTCA3’). The amplification thermoprofile was an initial denaturation step of 4 min at 95ºC, 40 cycles of 30s at 95ºC, 30s at 67ºC, and 30s at 72ºC, and a final hold at 72ºC for 10 min.

RNA isolation and Reverse Transcription. The RNeasy Midi Kit including the optional RNase free DNase set was used for RNA isolations from generated cell lines following the manufacturer’s protocol (Qiagen). Reverse transcriptions were carried out as described in Chapter III using 2 µg of total RNA for determination of any splicing changes. For quantitative PCR assays, reverse transcription was carried out with 5 µg of total RNA and the TaqMan Reverse Transcription Reagents Kit (Applied Biosystems). The amplification thermoprofile was an initial denaturation of 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 of 95°C for 5 min.

Non-quantitative PCR. To determine the presence of cryptic splice sites resulting from the intronic mutations, PCR to amplify the desired regions of *CPSI*

was performed using the cDNA templates generated from 2 µg of RNA. Following visualization on a 2% agarose gel and prior to sequencing, visible bands were purified using the Wizard PCR Cleanup Kit (Invitrogen). PCR primers were designed at least one exon away from the mutation in question to eliminate gDNA contamination. Amplifications were carried out with the following primer pairs: UBacT344A (5'CTGCTCAGAATCATGACC3')/ L1361 (5'GGCTTCGGTAAGACTGATGT3') for c.1210-1G>T (BECC+c.1210-1G>T-transfected cells), U1119 and LpcDNA3.1 (5'GCAACTAGAAGGCACAGTCGAGGC3') for c.1210-1G>T (minigene-transfected cells), and U393 (5'AGGACAGATTCTCACAATGG3') / L829 (5'CTAGCAGGCGGATTACATTG3') for BECC+c.652-3T>G transfected cells.

RT/PCR subcloning. The RT/PCR products from the BECC+c.1210-1G>T transfected cell line were isolated from the agarose gel using the Wizard SV gel and PCR Clean-up Kit (Promega) according to the manufacturer's protocol and subsequently subcloned into the pPCR-Script Amp SK+ cloning vector for blunt ligation according to the manufacturer's protocol, performing the additional PCR polishing step (Stratagene). This plasmid was then sequenced using a T7 promoter to determine the sequence of the RT/PCR products. Subcloning was performed in order to isolate different splice products of similar size, should they be present.

Quantitative PCR. Relative RNA expression levels were calculated using quantitative PCR on cDNA templates synthesized from the TaqMan Reverse Transcription Kit (Applied Biosystems). These PCR reactions were performed

using TaqMan technology on an automated platform with the ABI PRISM Detection System (Applied Biosystems). To measure *CPSI* expression, the Hs00919484_m1 assay which spans the exon 3-exon 4 junction and the Hs00919480_m1 assay which spans the exon 34-exon 35 junction were purchased (Applied Biosystems). E-GFP served as the endogenous control (Applied Biosystems part # 4331348, Custom Taqman(R) Gene Expression Assay Service). Each probe was subjected to multiple control assays to ensure proper activity (Appendix B). All experiments were performed in triplicate. "No RT" samples were used as a PCR control to correct for gDNA contamination. To perform the ΔΔCt analysis, a calculation for describing relative RNA expression levels, the difference between the Ct of each target *CPSI* and the Ct of the corresponding endogenous control GFP at the 0.200 fluorescence threshold was first calculated. These calculations were then expressed in relation to the "calibrator" (see Appendix C for details).

Northern Blotting. 10 µg of total RNA was used for Northern blotting in a 1.2% agarose formaldehyde gel that was run for 2.5 hours at 3 v/cm. RNA was transferred to a Hybond N+ membrane using the Turboblotter apparatus (both from Schleicher and Schuell) and UV-crosslinked. The membrane was prehybridized in Church buffer (0.25 M Na_2HPO_4, 7% SDS, 50 µg/ml sheared salmon sperm DNA) at 65ºC for 1 h and hybridization occurred overnight at 65ºC with either a CPSI or Cyclophilin probe previously radio-labeled with the Prime-It RmT Random Primer Labeling Kit (Stratagene) and diluted to 1 million counts per min in Church buffer.

Nonsense-mediated decay inhibition. To block nonsense-mediated decay, siRNA-mediated knockdown of UPF2 was performed based on a previous publication [148]. Briefly, cells were 75% confluent in a 10 cm dish. SiQuest transfection reagent (Mirus) was used according to the manufacturer's protocol at a final concentration of 5 µl per ml of DMEM. The UPF2 siRNA (5'-AAGGCTTTTGTCCCAGCCATCTT-3') was used at a final concentration of 70 nM (Dharmacon). In addition, the commercially available siCONTROL RISC-Free kit (RNA induced silencing complex; Dharmacon), was used to show knockdown was specific for UPF2. Cells were treated with siRNA constructs or vehicle alone for 3 consecutive days, as it was determined experimentally to provide the best knock down (Figure 4.1).

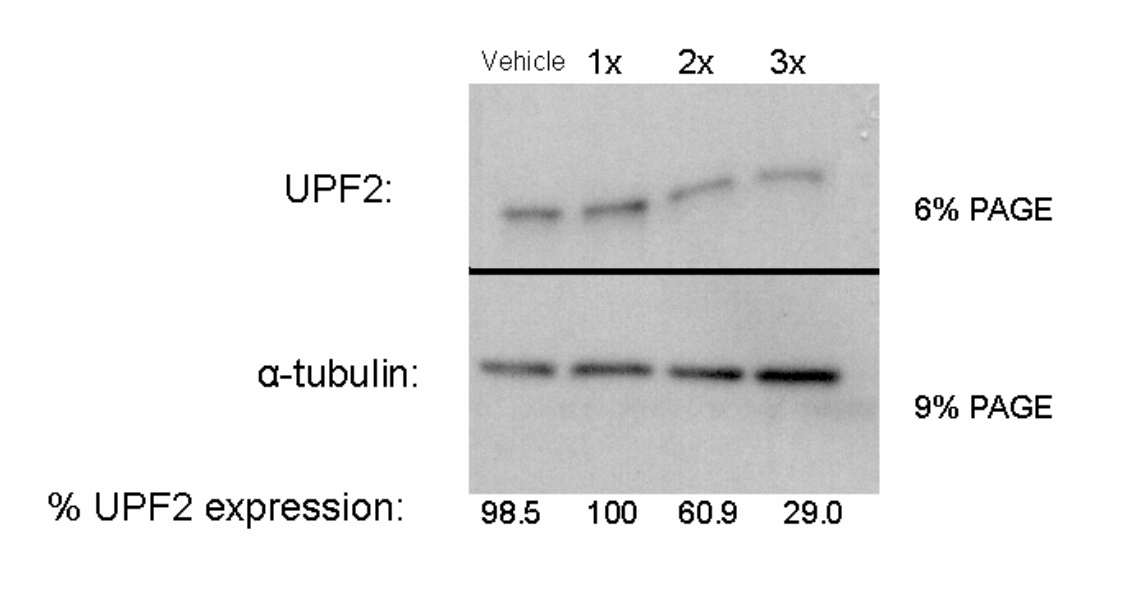


Figure 4.1. Optimal knockdown of UPF2 by 3 consecutive days of siRNA treatment. To determine the maximum knock down of UPF2 possible, cells were treated with transfection reagent only (vehicle) or 1, 2, or 3 consecutive days of siRNA transfections. Equal quantities from the same lysed cell pellet were loaded onto either a 6% or 9% PAGE gel. UPF2 expression was determined based on densitometry calculations compared to the α-tubulin loading control.

Western Blotting. Cells were lysed by sonication in a 50 mM Hepes Buffer with 0.01% TritonX-100. Cell lysates were quantified using the BCA Protein Estimation Kit (Pierce). Following a 5 min denaturation step at 95ºC in SDS sample buffer (375 mM Tris, 4% SDS, 20% glycerol, and 10% 2-mercaptoethanol), equal µg quantities of cell lysates were electrophoresed in a 9% PAGE gel for 45 min at 200 v. Proteins were transferred to a nitrocellulose membrane using the electrical tank transfer system at 4ºC for 90 min at 100 v. The membrane was blocked in blocking buffer (5% nonfat dried milk in 0.01M Tris; 0.5M NaCl; 0.05% Triton X-100; 0.2% Tween; pH 7.4) for 30 min at room temperature. The anti-hUPF2 primary antibody (gift from J. Steitz and J. Patton; 1:2500 dilution) and anti-α-tubulin (Abcam #15246; 1:5000 dilution) primary antibody were added to the blocking buffer for overnight incubation at 4°C with gentle agitation. The membrane was washed 3 times for 10 min each in blocking buffer before incubation for 1 h in blocking buffer with ECL rabbit IgG (Amersham NA934) secondary antibody (1:3000 dilution). The membrane was again washed 3 times in blocking buffer for 10 min. These steps were all performed at room temperature with gentle agitation. The results were visualized through chemiluminescence. The washed membrane was incubated in a 1:1 mixture of the Western Lightning Chemiluminescence Reagent Plus (Perkin Elmer) for 1 min and then subjected to various x-ray film exposures.

Results

Site-directed mutagenesis. Site-directed mutagenesis enabled the introduction of putative RNA-instability mutations identified in CPSID patients into the BECC vector. These point mutations (c.652-3T>G, c.1210-1G>T, c.1893T>G, and c.2388C>A) were made using a recently described and highly efficient two-step homologous recombination strategy with positive and negative *galK* selection [147]. Because this method required both a directed insertion of *galK* and a subsequent removal, multiple assays were performed for selection of proper recombinant colonies after each homologous recombination. Figure 4.2 is an example of these assays where both PCR and fingerprinting revealed a 1 kb change in BECC following *galK* insertion and restoration to wild-type after removal. Following selection of colonies that were determined to have undergone *galK* removal, *Not*I digestion and pulse-field gel electrophoresis showed that no unwanted global rearrangements occurred (Figure 4.3A). Direct sequencing of the region surrounding the desired mutation showed that, in all cases, only the desired point mutations were incorporated (Figure 4.3B). The same transfection and selection procedures used to create the wild-type cell line were used to create stable polyclonal cell lines for each site-directed mutant (see Chapter III). All cell lines demonstrated 100% GFP expression by visual inspection and maintained hygromycin resistance through antibiotic selection.

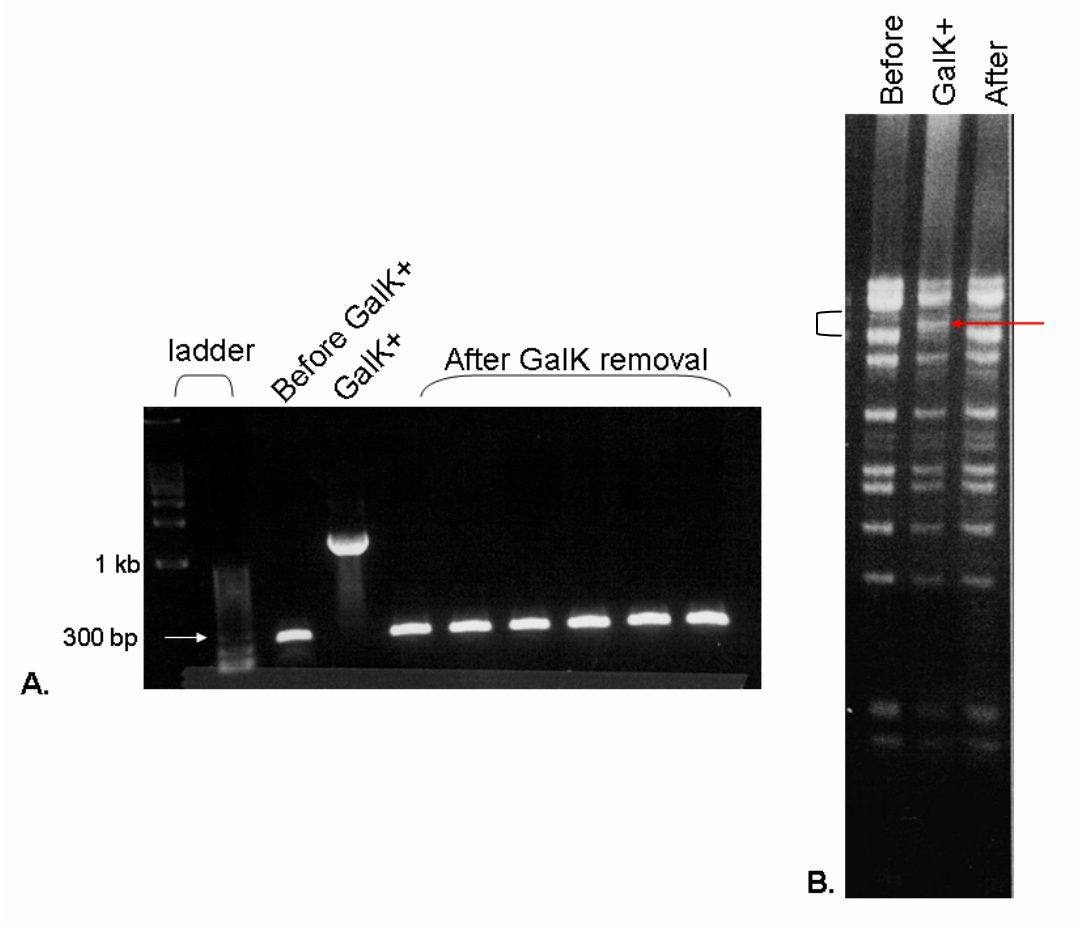


Figure 4.2. PCR and *Bam*HI fingerprints reveal g*al*K insertion and subsequent removal. These assays were performed for each mutation, but only the assays for c.1893T>G are shown here. **A.** A PCR using primers that flank the desired position of *gal*K insertion indicates the 1 kb shift in product size following the first recombination step and the subsequent removal of *gal*K at this position in multiple clones following the second recombination step. **B.** A fingerprint analysis reveals only a single band shift, consistent with the presence of *gal*K, following the first recombination step and the subsequent removal following the second recombination. The band shift, approximately 1 kb, is indicated by the red arrow and bracket.

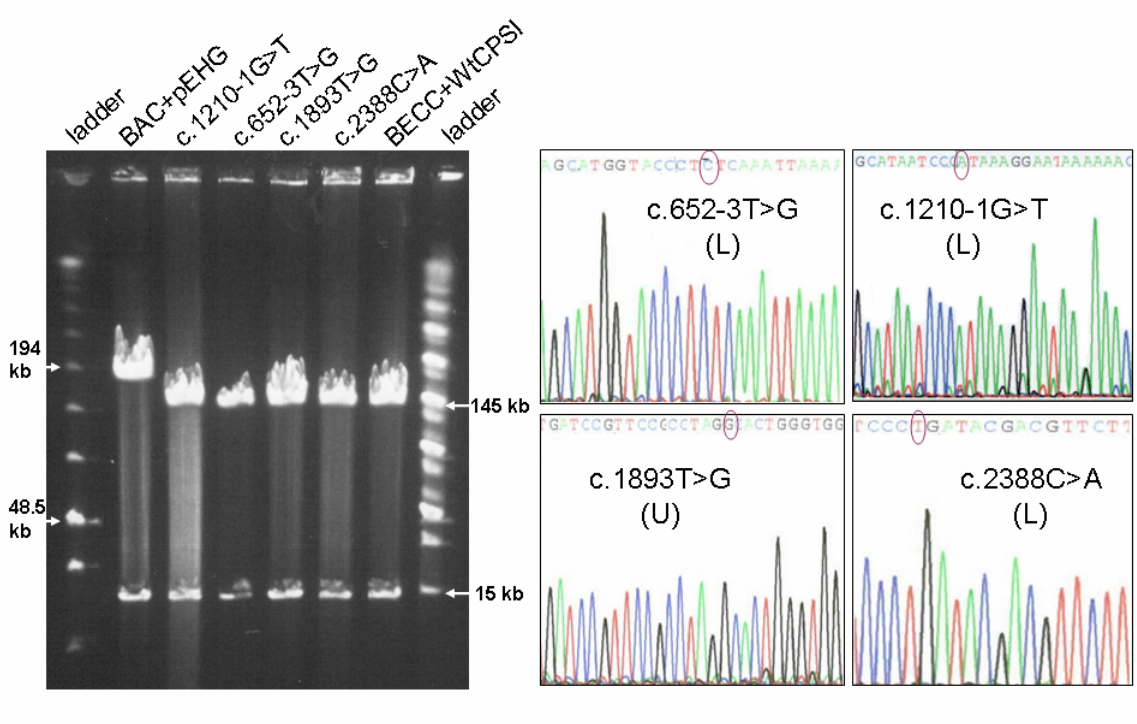


Figure 4.3. PFGE and direct sequencing reveal proper site-directed mutagenesis. All site-directed mutants are identical to BECC+WtCPSI but are ~35 kb shorter than the BAC+pEHG vector before CMV promoter insertion/upstream sequence deletion. Each mutation (circled) is verified by direct sequencing. U or L denotes whether an upper (forward) or lower (reverse) primer was used for sequencing.

To illustrate the presence of BAC-derived *CPSI* DNA in each cell line, PCR was performed to specifically amplify the exogenous *CPSI* from DNA extracted for each cell line. A forward primer was designed from the BAC-specific CMV promoter sequence and the reverse primer from exon 1 of the *CPSI* gene (Figure 4.4A). PCR revealed the presence of exogenous *CPSI* gDNA specifically in the transfected cell lines (Figure 4.4B).

RT/PCR and sequencing of splicing mutations. To determine if the intronic mutations affected splicing and to also identify possible cryptic splice sites, Reverse transcription/PCR was performed on RNA extracts from both the c.652-3T>G and c.1210-1G>T transfected cell lines. Because the reverse transcription protocol used was previously shown to be sensitive enough to pick up very low-level *CPSI* transcripts from a non-hepatic cell line, it was important to distinguish between BAC-derived and endogenous MRC-5V2 *CPSI* transcripts [25].

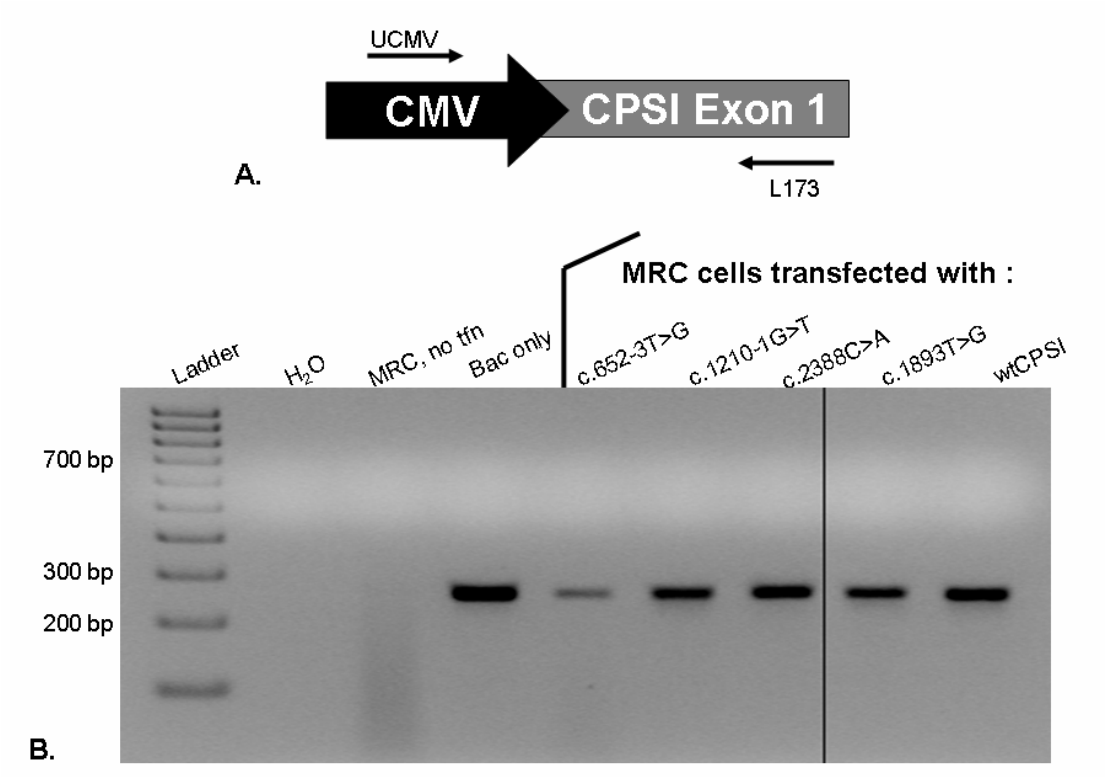


Figure 4.4. PCR of BAC-derived *CPSI* in all transfected cell lines. A. Representation of primer locations for identifying the presence of exogenous CPSI DNA. The forward primer is homologous to the CMV promoter and is specific for BAC-derived CPSI. **B.** 2% Agarose gel of PCR demonstrating the presence of exogenous CPSI gDNA only in the MRC cell lines transfected with the indicated BAC construct.

c.1210-1G>T. To distinguish between low-level exogenous and endogenous transcripts for the study of c.1210-1G>T, it was possible to exploit the presence of a nearby 3-bp polymorphism with alternate alleles present in MRC-5V2-derived (endogenous) and BAC-derived (exogenous) *CPSI*. The previously identified polymorphism, T344A, corresponds to either a nucleic acid sequence of ACC or GCT at position c.1136-1138 [12]. The forward primer used in these PCR reactions (UBacT344A: 5'CTGCTCAGAATCATG**ACC**3') contains homology specifically to the BAC allele at the 3' end (underlined) preventing amplification of any endogenous CPSI transcripts. Comparison of RT/PCR products from the MRC+WtCPSI and MRC+c.1210-1G>T cell lines showed a difference in product size (Figure 4.5A). The c.1210-1 G>T cell line produced a band which was larger than the expected wild-type PCR product, indicating a splicing change in the mutant. Subsequent sub-cloning and sequencing multiple colonies containing the RT/PCR product insert revealed an aberrant splice product that includes 32 bp of intron 10 (Figure 4.5B).

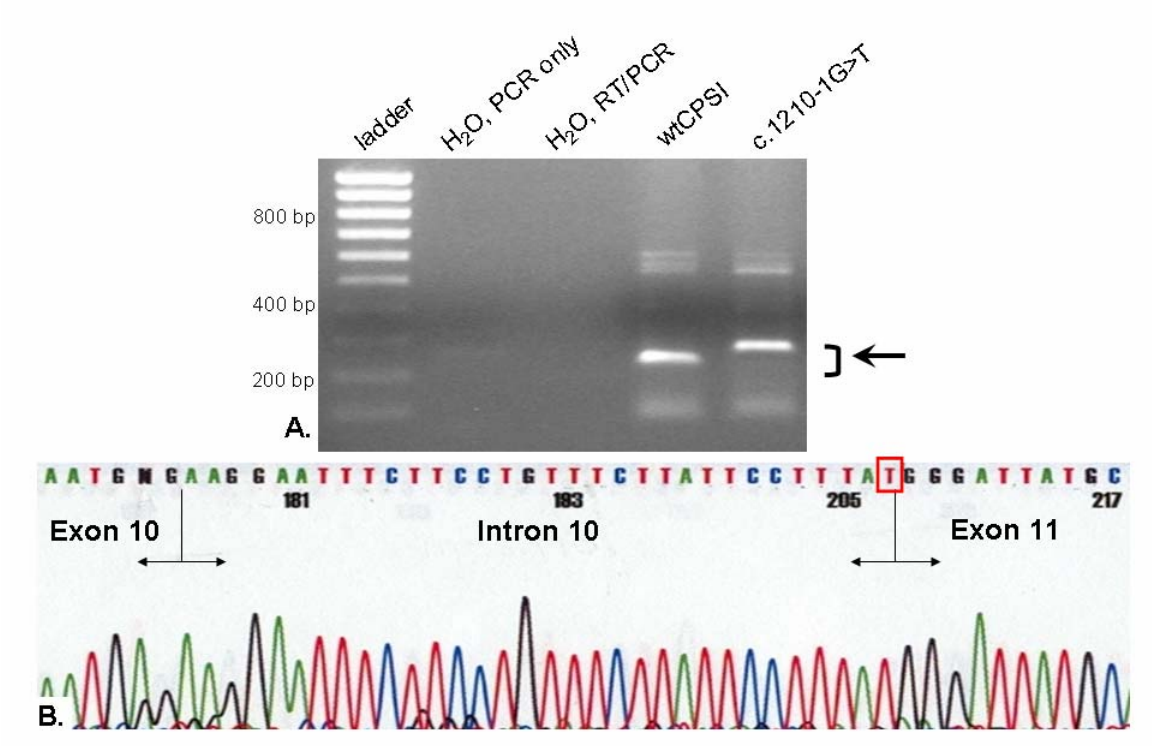


Figure 4.5. RT/PCR products from c.1210-1G>T. A. RT/PCR products from the polyclonal BECC+WtCPSI and BECC+c.1210-1G>T polyclonal cell line on a 2% agarose gel. The size difference between the wild-type and mutant products indicates activation of a cryptic splice site in the mutant cell line. **B.** Fluorescent sequence of the RT/PCR products shown in (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.

This splicing change identified from the BECC+ c.1210-1G>T construct was also identified in a minigene construct with the same mutation. Minigenes with this mutation and the wild-type sequence were built prior to completion of the BECC construct (Figure 4.6A). These minigenes were transiently transfected into both Cos7 and MRC5-V2 cells and subsequent RT/PCR analysis using a CPSI-specific and a minigene-specific LpcDNA3.1 primer revealed an identical splicing change in the mutant minigene (Figure 4.6B), identified by direct sequencing as the same 32 bp insertion of intron 10.

c.652-3T>G. The c.652-3G>T mutation was used to test the BECC system for the ability to recapitulate splicing changes originally identified from patient DNA. Whereas the patient genomic DNA shows a heterozygous mutation at the -3 position of intron 5 (Figure 4.7A), patient cDNA (from a patient-derived fibroblast cell line) shows a heterozygous 2-bp frameshift beginning in exon 6, indicating the mutation activates an aberrant splice site (Figure 4.7B). RT/PCR on the stable cell lines created with the BECC+WtCPSI and BECC+c.652-3T>G constructs verified that the BECC platform reproduces what is observable in the patient data. Because a size change in the c.652-3T>G mutation band was not detectable on an agarose gel (Figure 4.7C) and polymorphic variations could not be exploited to separate endogenous and exogenous transcripts, the RT/PCR product was excised and sequenced radioactively in order to visualize the presence of aberrant transcript. The sequence from the MRC+c.652-3T>G cell line mimics the patient data by revealing the same 2-bp AG dinucleotide insertion, again indicating a splicing alteration in intron 5 (Figure 4.7D).

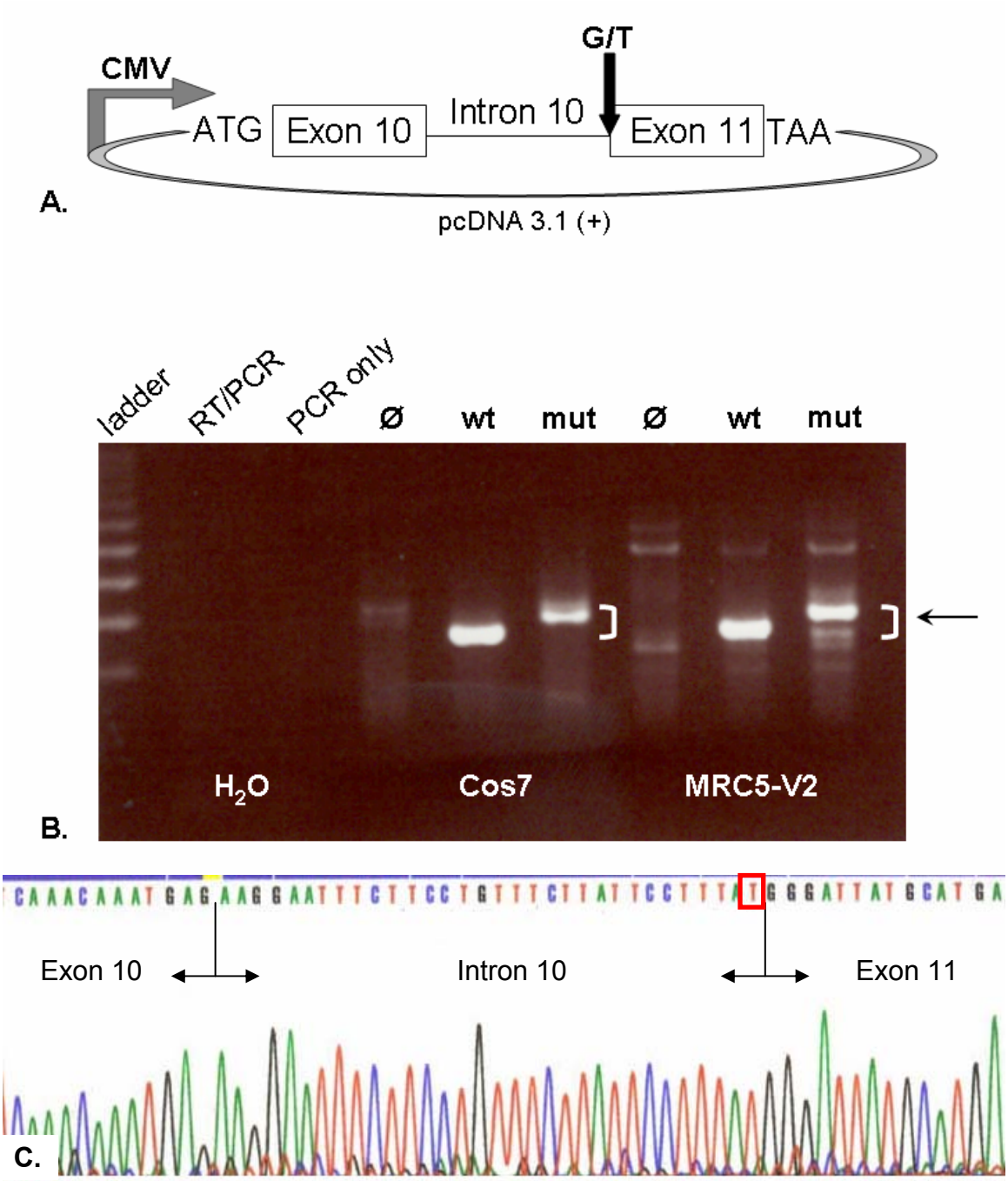


Figure 4.6. Minigene transfections reveal splicing defect from the c.1210-1G>T mutation. A. Schematic representation of the minigene constructed with either a G (wt) or T (mut) at base c.1210. **B.** RT/PCR products on a 2% agarose gel following transient transfections of minigene constructs into Cos7 or MRC5-V2 cells. The brackets highlight the shift seen from the mutant construct that is indicative of a splicing change. Non-transfected cells (Ø) show that the RT/PCR products are specific to the minigene constructs.

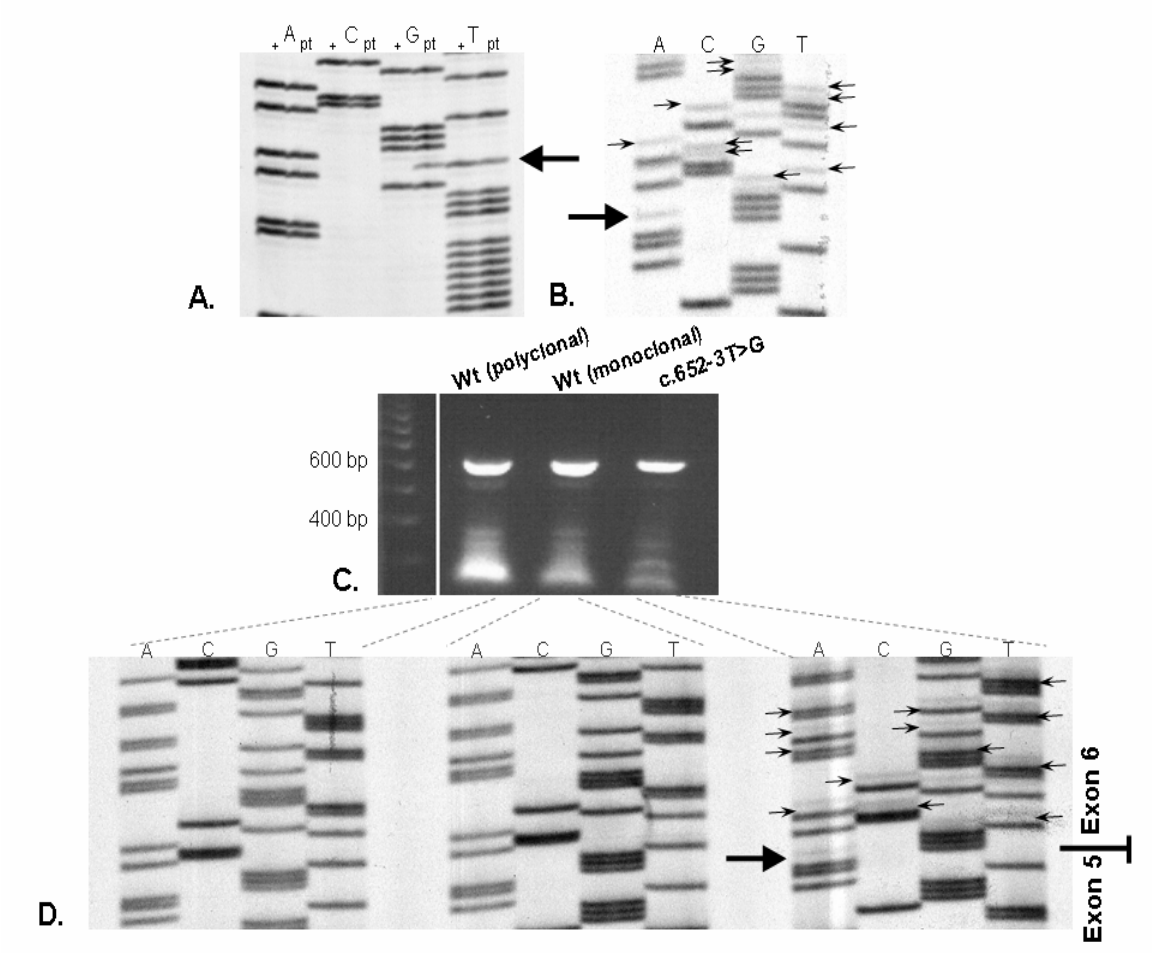


Figure 4.7. RT/PCR products from c.652-3T>G. **A.** Patient gDNA sequence revealing G>T point mutation in intron 5 (arrow). **B.** Patient cDNA revealing a 2-bp insertion (starts at large arrow, small arrows point out other residues in the frameshift) in exon 6 from one allele. **C.** RT/PCRs from the BECC platform. Cell lines tested are a polyclonal wtCPSI, monoclonal wtCPSI, and c.652-3T>G polyclonal cell line on a 2% agarose gel. **D.** Dideoxy-radionucleotide sequence of the RT/PCR products shown in (C) reveals the presence of a 2 bp frameshift specifically in the mutant cell line (starts at large arrow, small arrows point out other residues in the frameshift) that mimics the original patient data.

Quantitative RT/PCR. Because mutations that cause formation of premature termination codons most often elicit NMD and cause degradation of the RNA message, ΔΔCt analysis from quantitative real time RT/PCR reactions was used to determine the relative expression levels of *CPSI* transcript in each transfected cell line. These ΔΔCt calculations describe *CPSI* transcript levels that were first standardized to *GFP* transcript levels and then were expressed in relation to the wild-type "calibrator" sample. Because *CPSI* and *GFP* are present in a 1:1 ratio and both driven by a CMV promoter, GFP provided an ideal endogenous control for standardizing the CPSI data and controlling for differences in BAC copy number in each cell. Three TaqMan probes were used to specifically measure either *CPSI* or *GFP* transcripts in each cell line. One CPSI probe annealed to the transcript at the exon 3-exon 4 junction, upstream of all tested mutations. The other annealed to the exon 34-exon 35 junction, downstream of all tested mutations. All probes were tested extensively to ensure proper function (Appendix B). ΔΔCt analysis revealed that the each mutation resulted in lower expression levels than wild-type, except for c.652-3T>G measured by the exon 3-exon 4 probe (Figure 4.8A). While there was variability in the relative expression differences between each mutant and wild-type sample using the CPSI exon3-exon4 probe, all mutants showed a fairly equal drop off in expression when compared to wild-type using the CPSI exon 34-exon 35 probe. A Northern blot corroborates this data by showing that exogenous *CPSI* is visible only in the polyclonal and a monoclonal cell line harboring the wild-type construct, but none of the site-directed mutants (Figure 4.8B).

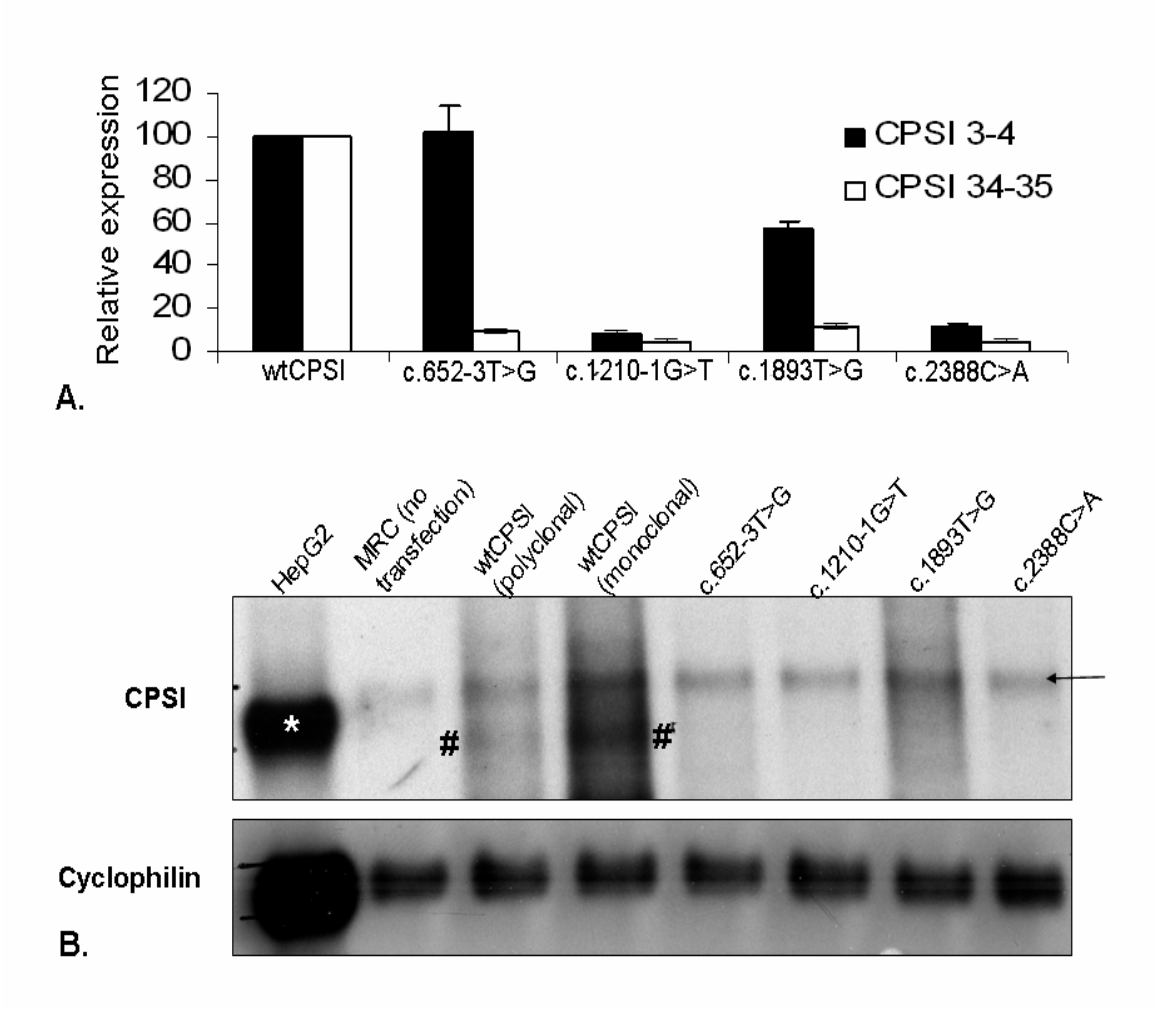


Figure 4.8. ΔΔCt analysis showing relative CPSI RNA expression levels following quantitative RT/PCR. A. Expression levels of CPSI RNA in each mutant cell line relative to Wt, arbitrarily set at 100% expression. Black bars indicate relative CPSI RNA levels measured from the exon 3-exon 4 TaqMan probe and white bars indicate relative CPSI RNA levels measured from the exon 34-exon 35 TaqMan probe. Error bars represent standard error. **B.** A Northern blot of RNA from each transfected cell line. This Northern shows the same decrease in RNA levels in each mutant cell line as in (A). Only the wtCPSI cell lines have a visible band of exogenous, BAC-derived CPSI (between the # signs). Only the HepG2 hepatic cell line shows endogenous CPSI expression (*). An arrow indicates a non-specific 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 MRC5-V2-derived CPSI (shown in the liver-derived HepG2 lane) due to CMV promoter swapping which altered the size of the 5' untranslated region by approximately 100 bp.

Inhibition of nonsense-mediated decay. To determine if the decreased levels of *CPSI* expression seen in the qRT/PCR experiments were due to degradation via the nonsense-mediated decay pathway, this pathway was inhibited using siRNA-mediated UPF2 knockdown, previously identified as a required component of the NMD pathway [96,106,148]. First, the specific knockdown of UPF2 was verified in the MRC+wtCPSI cell line (Figure 4.9A). Using the same conditions, UPF2 was next knocked down in each transfected polyclonal cell line containing the *CPSI* mutations. In all cases, there was a much lower UPF2/α-Tubulin ratio in the siRNA-treated cells than in untreated cells or those treated with vehicle alone or a non-specific siRNA control, again demonstrating specific knockdown of UPF2 (Figure 4.9B).

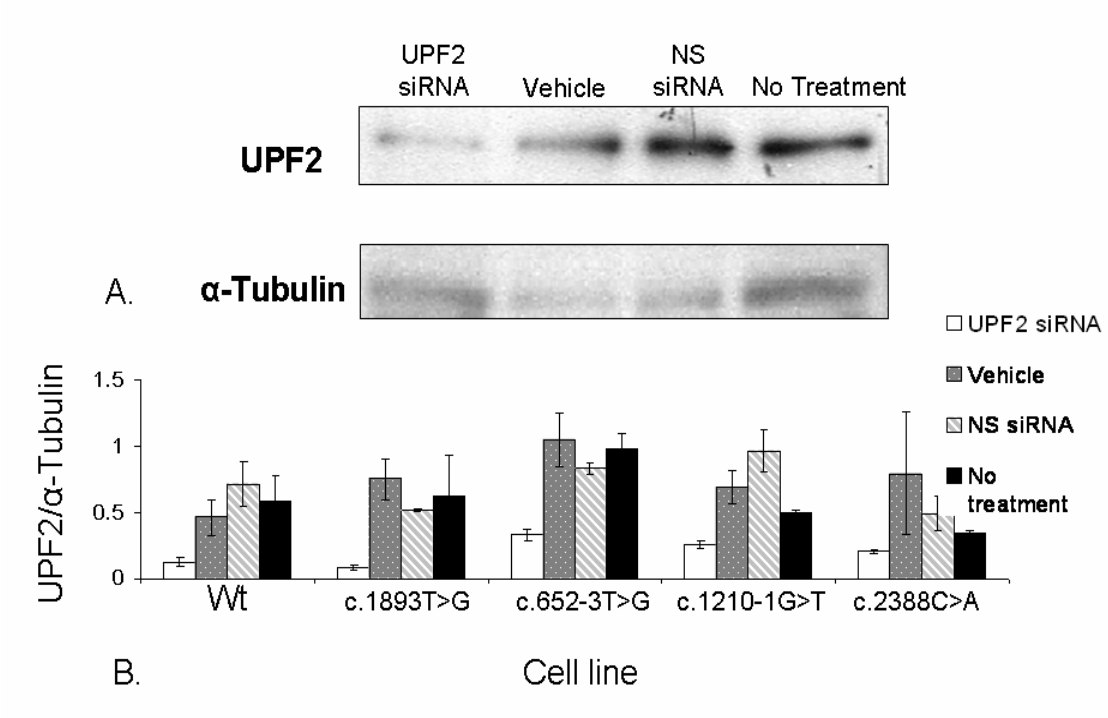


Figure 4.9. siRNA-mediated knockdown of UPF2 in all cell lines. A. A representative Western blot of UPF2 and the loading control, α-tubulin, for the MRC+wtCPSI cell line following each treatment condition: UPF2 siRNA, vehicle alone, non-specific (NS) siRNA, or no treatment. **B.** Average Western blot densitometry quantifications performed in triplicate showing relative levels of UPF2 knockdown when compared with the control treatments, all expressed as UPF2/α-tubulin ratios. The error bars represent standard error.

Quantitative RT/PCR showed that when UPF2 was knocked down in these cell lines, levels of *CPSI* transcript increase when compared to transcript levels from vehicle-only and untreated controls (Figure 4.10). All data was again expressed as a ΔΔCt value, this time representing the fold increase in expression when compared to the calibrator (untreated) sample arbitrarily set at an expression value of 1. These results indicate that the *CPSI* mutations cause RNA instability, at least partially through eliciting the nonsense-mediated decay pathway as knockdown of UPF2 increases the relative expression levels of each mutant *CPSI* transcript when compared with untreated cells. Following UPF2 siRNA treatment, the *CPSI* fold increase observed in every mutant cell line is greater (6- to 230-fold) than the *CPSI* fold increase in the wild-type cell line (3-fold). Although the amount of the relative increase varied depending on the cell line and probe used (possibly due to varying degrees of siRNA-mediated NMD inhibition, or to unique properties of each mutation, or both), these results strongly indicate a role for NMD in degradation of *CPSI* mutant transcripts.

The above data is in contrast to experiments where puromycin was instead used to inhibit NMD. Puromycin inhibits translation and has therefore been previously used to inhibit NMD in other cell lines (example: [78]). However, following multiple concentrations (50, 200, 400, 600, and 800 ug/ml) and multiple time courses for treatment (8 or 14 h), the cells either died, or did not exhibit any change in RNA levels for any transcript tested. This data suggests the use of caution when using harsh treatments that induce widespread cellular changes when only a specific effect is examined or desired.

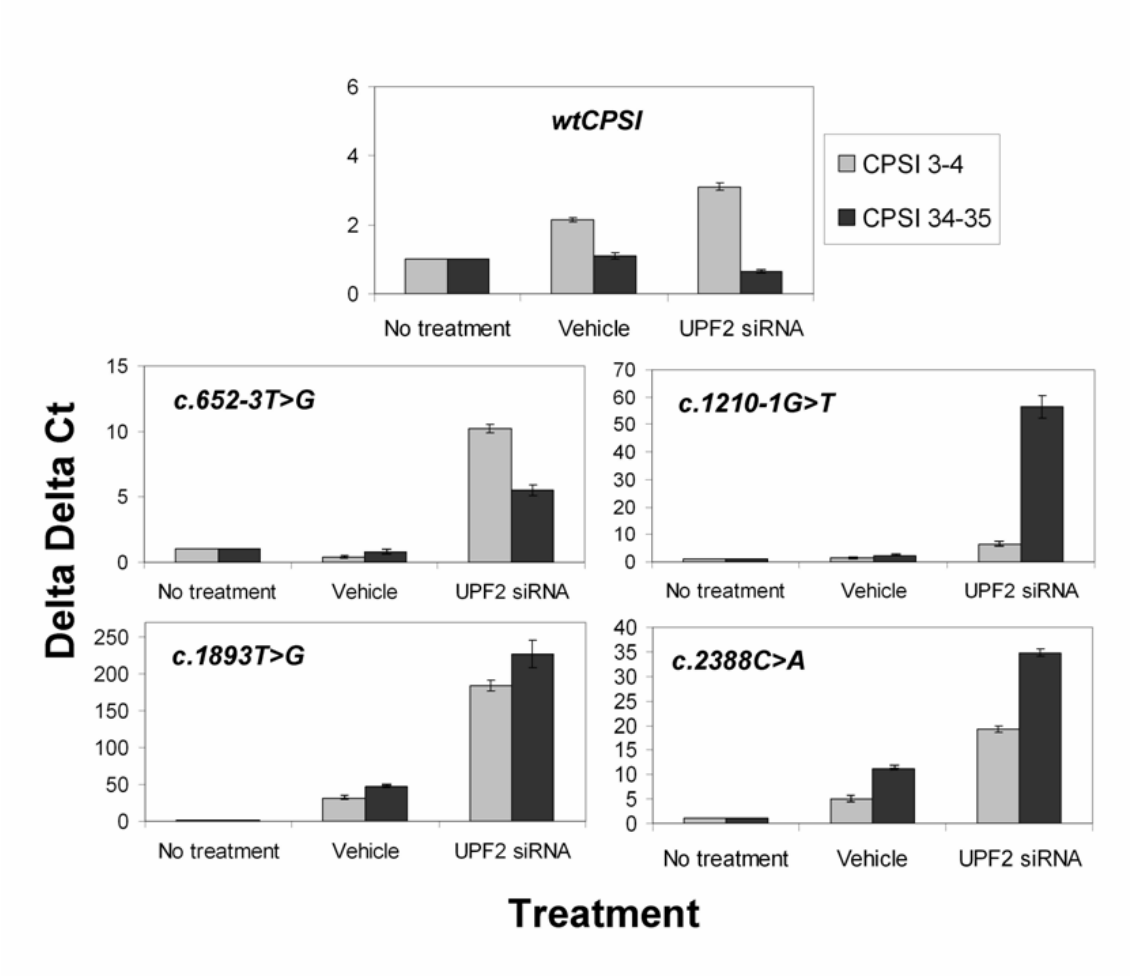


Figure 4.10. ΔΔCt analysis of quantitative RT/PCR data following siRNA-mediated knockdown of UPF2. Each panel graphs relative expression levels (calculated as the delta delta Ct) of CPSI transcripts following UPF2 siRNA or vehicle only treatments in the respective cell line as compared to untreated cells (arbitrarily 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 standard error.

ESE*finder*. To investigate the properties of the sequence around the two exonic mutations, c.1893T>G and c.2388C>A, exons 16 and 19 were searched to identify putative exon splicing enhancers. Inputting both the wild-type and mutant sequence into ESE*finder* revealed changes in ESE consensus sequence scores at both mutations [56]. The c.1893T>G mutation disrupts a SC35 consensus sequence and changes the threshold score (measure of conformity to the consensus sequence) from 2.387 to 0.0671 (Figure 4.11). The c.2388C>A mutation not only disrupts a putative SC35 site by changing the score from 2.242 to 0.9657 but also creates a consensus SRp40 site by changing the threshold from 1.27 to 3.01 (Figure 4.12).

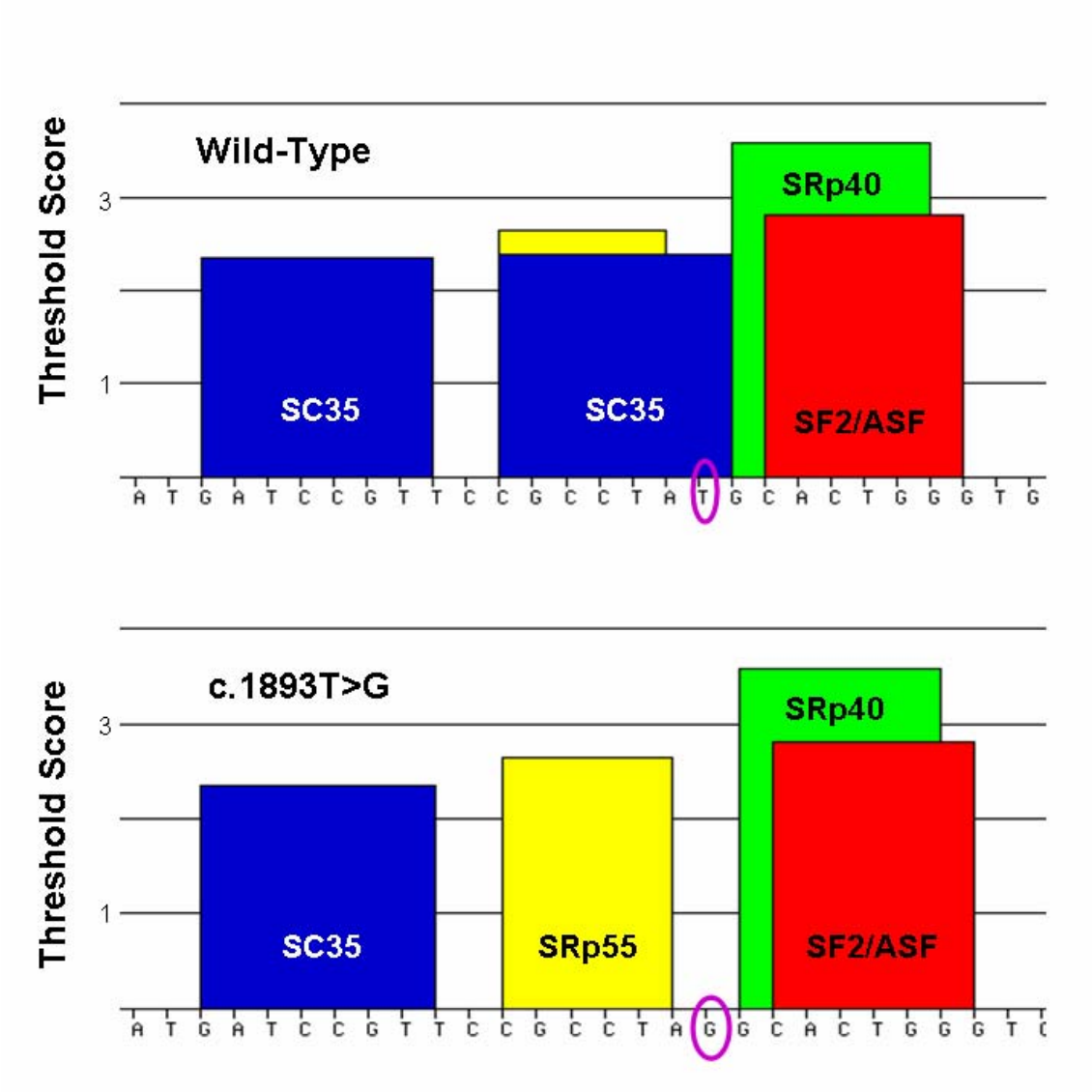


Figure 4.11. ESE*finder* results for wild-type and c.1893T>G. The putative SC35 binding site identified on the wild-type construct (top panel) is destroyed by the mutation (bottom panel). Purple circles denote the wild-type and mutant base pair examined.

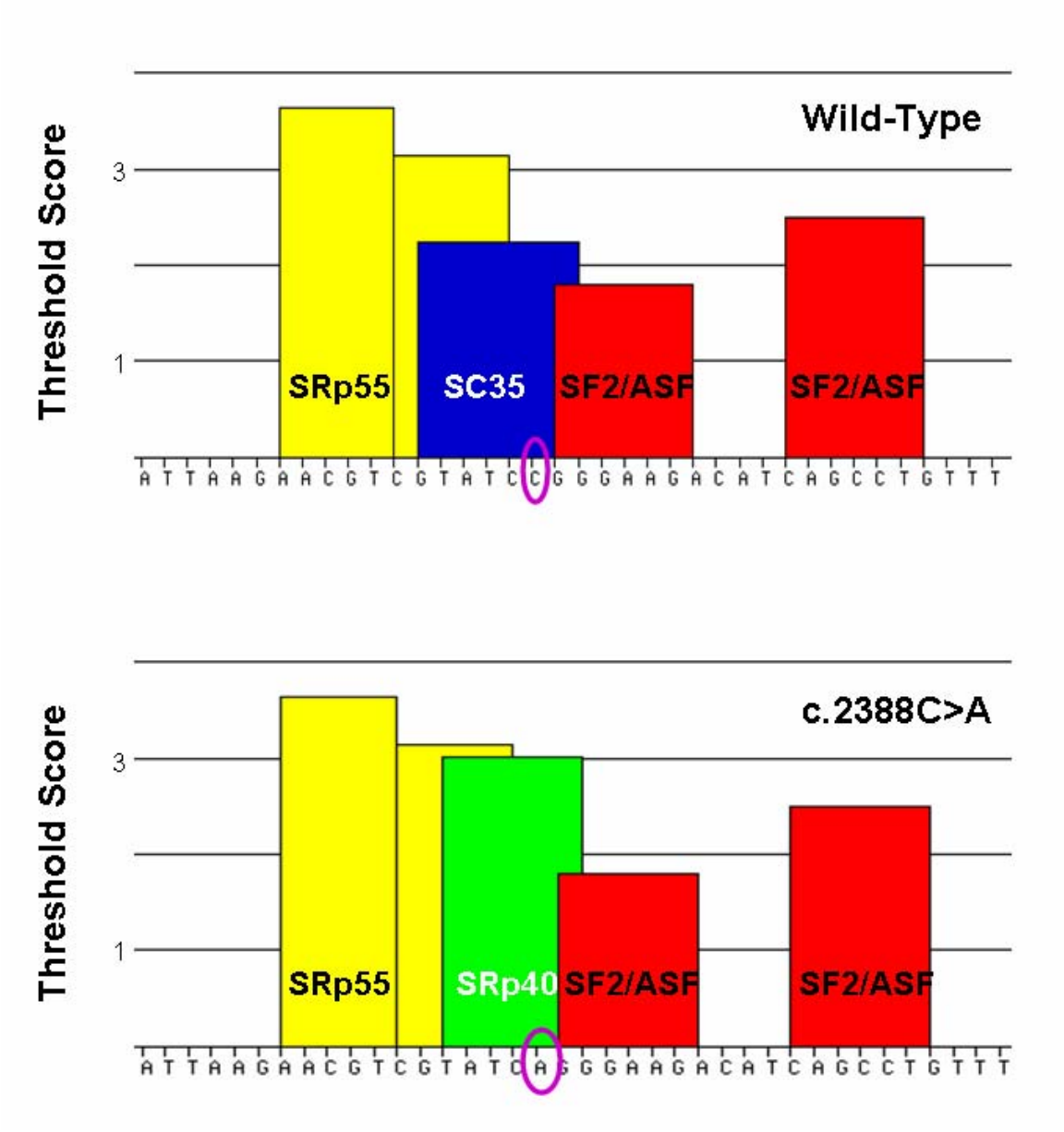


Figure 4.12. ESE*finder* results for wild-type and c.2388C>A. The putative SC35 binding site identified on the wild-type construct (top panel) is destroyed by the mutation but a new putative SRp40 site is detected (bottom panel). Purple circles denote the wild-type and mutant base pair examined.

Discussion

This Chapter outlines data generated from the study of four putative RNA instability *CPSI* mutations (c.652-3T>G, c.1210-1G>T, c.1893T>G, and c.2388C>A) in conjunction with the development of the BECC platform. Re-creating each of the above mutations in the BECC platform had several advantages in addition to validation of the model system. First, as patient samples are usually obtained from individuals who are heterozygous for different mutations, this experimental system enabled isolation of the mutation in question. Second, each mutation could be maintained in a system without the limited shelf-life typically seen for patient tissue samples, which are also often difficult to obtain. Third, this platform enables performing manipulations that are not always possible or reliable in diagnostically available hepatic samples or established patient cell lines with low-level gene transcription.

Beyond supporting the validity of the BECC platform as a novel gene expression system sensitive to experimental manipulations, data presented in this Chapter provides confirmation of the mechanistic effect of each mutation. This data demonstrates that the observed RNA instability for each mutation is at least in part due to transcript degradation through the NMD pathway, as confirmed by knockdown of UPF2 (Figure 4.10). Additionally, splice variants resulting from the two intronic mutations were identified (Figures 4.6 and 4.7).

Quantitative RT/PCR. Following introduction of each mutation into BECC by site-directed mutagenesis, these constructs were transfected into MRC5-V2 cells and the activity of each construct was assayed through quantitative

RT/PCR. The TaqMan 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 to wild-type. In all cases, the 5' CPSI probe detects higher levels of *CPSI* transcript, but these levels are all less than wild-type except c.652-3T>G. The observation that in each case, more mutant transcript was detected with the CPSI exon3-exon 4 probe, which lies upstream of all tested mutations, may indicate degradation primarily from the 3' end. It is also possible that the measurement of *CPSI* in the c.652-3T>G mutant with the exon 3-exon 4 probe is a result of the close proximity of the probe to the location of this mutation rather than an accurate measure of mature transcript. Therefore, an important consideration for quantitative PCR is to measure transcripts with multiple probes being mindful of the distance between the probe and mutation location since the probes only anneal to a small portion of the much larger transcript and could possibly detect this fragmented RNA. This quantitative RT/PCR data, coupled with the analysis in Chapter II, supports that a significant number of pathogenic CPSID mutations result in RNA processing defects.

c.1210-1G>T. While both intronic mutations demonstrate the importance of the AG sequence in proper 3' intron definition, only the c.1210-1G>T mutation disrupts the endogenous AG within the consensus $(Y)_nNYAG|G$ sequence necessary for proper intron definition [50]. This disruption causes the activation of a cryptic splice site, recognizing an improper AG 32 bp upstream as the 3' splice site (Figure 4.5). This same splice site change was identified from minigene constructs with the wild-type and mutant sequence at this location

(Figure 4.6). That the cryptic splice site is upstream of the wild-type AG was slightly unexpected. The competitive scanning model suggests that the first AG downstream of the branch point sequence is the 3' end of the intron with a few exceptions [48,149]. The data shown here may indicate that the wild-type AG is not the first such dinucleotide downstream of the branch point sequence in intron 10. It is possible that there is a more proximal AG that is less favorable due to a location too close to the branch point sequence or because it upstream of the polypyrimidine tract and is therefore not used in the transesterification reaction [149,150]. Though the branch point sequence and downstream polypyrimidine tract are both necessary for proper intron definition [151], the non-canonical positioning of an AG upstream of the polypyrimidine tract does not prevent splicing [149]. Another possible explanation is that alternate branch points or polypyrimidine tracts are recognized in this intron, but this seems less likely given the data that shows the branch point sequence and the polypyrimidine tract are at least sometimes specified independent of 3' splice site definition [48]. However, other data suggests that in some transcripts with weaker polypyrimidine tracts, AG recognition is required for the first step of splicing, making the AG important for polypyrimidine and branch point sequence recognition [152,153]. Figure 4.13 shows the 3' sequence of intron 10 with both the wild-type and cryptic splice sites highlighted alongside the consensus sequence for intron definition. The strongest polypyrimidine tract is highlighted as well as potential branch point sequences, illustrating the complexity of interpreting the observed alternate splice site activation.

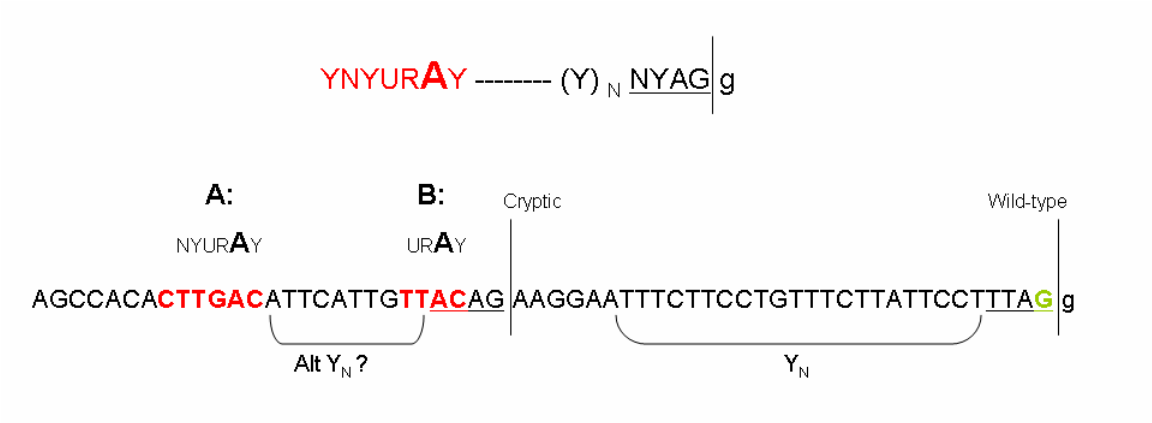


Figure 4.13. Sequence of the 3' end of intron 10 highlighting possible branch point sequences and polypyrimidine tracts. The consensus sequence for 3' intron definition is depicted at the top, and the bottom sequence is intron 10 (in caps) and exon 11 (lower case) with the corresponding consensus elements highlighted. The branch point sequence is depicted in red and two possible sites (A and B) are illustrated for intron 10. In addition, the 3' end of intron 10 is underlined for both the wild-type and cryptic splice sites, both matching the consensus sequence. Y_N highlights the most likely polypyrimidine tract.

c.652-3T>G. The c.652-3T>G mutation creates a premature AG dinucleotide sequence and disrupts the pyrimidine conservation immediately upstream of the wild-type AG. Sequencing of RT/PCR products from wild-type and mutant cell lines revealed transcripts only in the mutant cell line that contained a dinucleotide insertion into exon 6. This insertion resulted from improper recognition of the premature AG as the 3' end of the intron in both patient-derived cells and the BECC-transfected cells (Figure 4.7). This new AG dinucleotide was recognized as the 3' end of intron 5, even though a G residue is directly 5' of the new AG, which has been shown to be the weakest consensus nucleotide for preceding a terminal AG [149]. This intronic mutation again illustrates the importance of the sequential presence of a branch point, polypyrimidine tract and AG dinucleotide in proper 3' intron definition, as the premature AG formed became the first such sequence downstream of the branch point sequence and therefore altered the recognition of the 3' end of intron 5 (Figure 4.14).

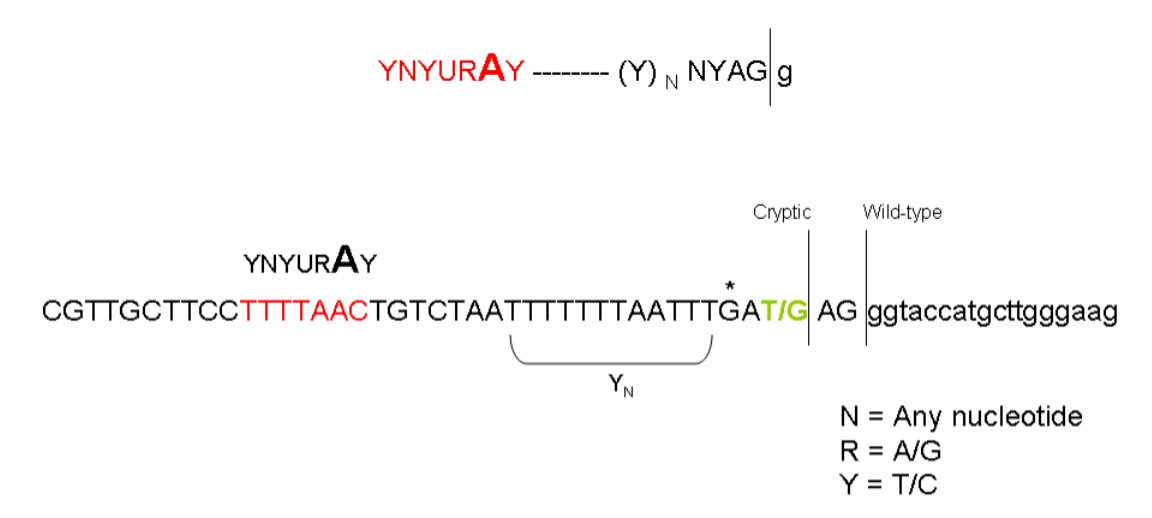


Figure 4.14. Sequence of the 3' end of intron 5 highlighting the probable branch point sequence and polypyrimidine tract. The c.652-3 mutation is shown in green with both the wild-type and cryptic splice sites depicted. The (*) indicates the G residue that does not conform to the consensus for 3' intron definition.

c.1893T>G. 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 ability of NMD to degrade an exogenously expressed gene because the stop codon is formed independent 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. A possible explanation for this observation is that there are separable mechanisms for eliciting NMD based on the type of mutation, as no frameshift occurred to create the PTC. The possibility also exists that there are various cryptic splice sites recognized in transcripts with the intronic mutations which, though undetectable by RT/PCR, alter not only the number of transcripts containing PTCs but also alter the position of the PTC, which may have an effect on recognition by the NMD pathway.

It is possible that the stop codon also alters splicing, should this mutation be located in a *cis*-acting sequence necessary for proper exon definition. To address this possibility, exon 16 was searched using ESE*finder* and disruption of an SC35 consensus sequence containing nucleotide position c.1893 was identified (Figure 4.11). SC35 has previously been shown to increase splicing of exons [154,155]. Therefore, this mutation might not only cause RNA degradation from the presence of the stop codon, but also may negatively affect transcript splicing due to loss of an SC35 binding site. This has previously been shown to occur, as a point mutation resulting in a stop codon (R187X) caused exon skipping due to disruption of an ASF/SF2 consensus site in the mouse modifier

gene, *Scnm1* [156] and a nonsense mutation in the *NDUFS4* gene causes alternative splicing products [157].

c.2388C>A. The BECC platform 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 enzyme function. However, following a thorough genotyping of *CPSI* in this patient, no other genomic changes could be found on the second allele. Coupled with the knowledge that many missense and silent mutations are pathogenic at the RNA processing level if they reside in an exonic splicing enhancer, RNA instability was the proposed mechanistic effect [53,123]. The data presented here supports that hypothesis by showing that transcripts with this silent mutation are increased following the inhibition of NMD (Figure 4.10). This data adds to the increasing evidence that many mutations including silent and missense mutations, while they do not alter protein structure, can still negatively affect the processing of the transcript so that mature proteins are not always produced [53,123].

Additionally, ESE*finder* identified that the c.2388C>A mutation alters putative ESE recognition sites. The mutation not only disrupts the consensus sequence for SC35, but also creates a consensus sequence for SRp40. Because it is most likely that multiple, seemingly redundant splicing signals and enhancers must act in a synergistic manner for proper splicing, this change in ESE consensus sequences is an important consideration in determining the mechanism of c.2388C>A. There is evidence that in many exons, antagonism between splicing activators and repressors exists and sequence variations may

disrupt this equilibrium [158]. Previous studies have specifically indicated the antagonism of SRp40 to SC35 function and the activation of splicing due to SC35 [154,155,159-161]. In addition, the creation of SRp40 consensus sites by a single substitution has previously been indicated as the mutation mechanism in other genetic diseases [162,163]. It is also possible that this mutation affects the recognition site for another component of the spliceosome without a defined consensus sequence.

The data in this Chapter verifies the creation of a novel model system that is useful in determining the mechanism of mutations in CPSID. This system is of particular utility for assaying mutations that affect RNA processing because of the size capabilities that usually allow for the inclusion of a complete gene on one construct. By studying mutations within their wider sequence context, a more accurate determination of the mechanistic effect should be possible. Furthermore, the BECC platform is particularly useful in situations where patient RNA, cells or tissues are not available. It additionally permits experimental manipulations that cannot be done in patient tissue and therefore provides an efficient and reasonable alternative. For example, the data shows that nonsense-mediated decay is a mechanism for the CPSID disease phenotype and corroborates evidence that NMD can be triggered from an exogenously produced transcript, as was shown previously using UPF2-silenced HeLa cells [69]. Therefore, this system is proposed as a relevant eukaryotic cell model for efficiently and effectively studying the role of genetic variations.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

The data outlined in the preceding Chapters tested the hypothesis that the underlying molecular mechanisms causing disease in a substantial fraction of CPSID patients are defects in RNA processing that can be studied following construction of a BAC-based model system. RNA transcripts undergo a series of processing steps involving essential *cis* elements, and transcripts are subject to surveillance and subsequent degradation if premature termination codons are identified. There is increasing evidence that such RNA processing defects comprise a significant portion of disease-causing mutations [66-68], such as is described here for CPSI Deficiency where 40% of disease-causing alleles studied were found to result in RNA instability [127]. The data supports that when studying the relationship between genetic mutations and disease phenotypes, efforts should be taken to detect splicing changes and allelic dropout from nonsense-mediate decay. To this end, a BAC-based platform was developed to accommodate molecular studies on mutations in the large *CPSI* gene. Using this BECC vector overcomes the difficulties associated with study of patient tissue sources that are difficult to obtain and manipulate. Mutational analysis of four CPSID genetic variants (c.652-3T>G, c.1210-1G>T, c.1893T>G, and c.2388C>A) in this novel system showed that two caused splicing errors and

each elicited the nonsense-mediated decay pathway due to the introduction of a PTC.

A primary objective in understanding CPSI Deficiency was to classify the heterogeneous set of identified mutations according to the mechanism conferring the enzyme deficiency. Chapter II describes the quantification of *CPSI* mutations that are likely RNA processing defects and provides evidence for this classification through data collected from the CPSID database. In this Chapter, CPSID was a model for understanding and predicting the prevalence of PTC-forming mutations responsible for genetic disease but only included the study of patients with both an archived RNA and DNA source and therefore focused on RNA processing mutations. With the development of the BECC platform for testing mutations, now all mutations in the database could be tested to determine a functional mutation classification, irrespective of patient RNA availability, allowing a more detailed analysis and understanding of all pathogenic *CPSI* genetic variants.

Chapter III details the construction of the BECC vector and its utility as a model system for studying *CPSI* gene expression. Construction included combining a library BAC clone with the modular pEHG insert and then further modifying the construct to allow *CPSI* to exhibit more ubiquitous expression under the control of the CMV promoter. Following transfections into an immortalized human cell line and creation of stable polyclonal cell lines, exogenous *CPSI* transcript and protein expression was confirmed. The next goal

was to utilize this system to test multiple rare variants identified in the *CPSI* gene of CPSID patients.

Using site-directed mutagenesis, four mutations were introduced into the BECC vector and subsequent analysis verified each affected the stability of the RNA transcript. The *galK*-mediated two-step recombination method proved very efficient for generating point mutations [147]. Future uses include introducing additional mutations into the *CPSI* BECC vector, which is also versatile enough for creating even larger insertions and deletions. There are several multiple base-pair insertions and deletions associated with CPSID; however none was included in the present study. It would be beneficial to test insertions and deletions not created by splicing mutations in the same manner as was performed in Chapter IV to determine if, and to what extent, NMD is involved in their pathogenesis and compare these results to the mutations already studied. The mutations already studied in BECC each provided details on the specific molecular defect, but possibilities exist for further detailed studies to gain more insights into molecular pathology and RNA processing in general.

Both intronic mutations, c.652-3T>G and c.1210-1G>T, disrupted recognition of the AG dinucleotide required for proper intron 3' end definition. An alternate 3' splice site was identified for each from direct sequencing of a major RT/PCR product. The possibility also exists that there are other, more minor transcripts produced which were visibly undetectable in the RT/PCR reactions visualized on agarose gel (Figures 4.5 and 4.6).

The presence of multiple alternate splice products is especially possible from the c.1210-1G>T mutation, as the AG recognized as the 3' end of the intron was 32 bp upstream of the mutation, not downstream. Because the first AG downstream of the branch point usually serves this purpose, the original expectation was to find a splicing change that excluded part of the downstream exon or included the entire intron. This observed splicing change, coupled with the substantial degradation visible from the Northern blot of the patient's hepatic RNA (Figure 2.3), suggests that minor splice products may be present and future studies could be directed towards identifying the presence of other misspliced transcripts.

The study of mutation c.652-3T>G provided a unique strength due to detection of the misspliced product in both the patient RNA and through the BECC model system. Not only did study of this mutation provide verification that this model system works and corroborated the observed splicing change, but its study in BECC also provided an environment for manipulating the NMD pathway to verify at least partial participation in the observed RNA instability of c.652-3T>G.

A mutation not suspected to result in any splicing changes was c.1893T>G. This nonsense mutation created a premature stop codon without altering splicing. The possibility existed, however, that the point mutation occurred in a *cis* element necessary for proper splicing, such as an exon splicing enhancer. Because this possibility has not been directly excluded, it may be beneficial to study this option given that a putative SC35 site was destroyed by

the mutation. However, the increase in relative transcript expression following NMD inhibition suggests a direct effect of the nonsense mutation. To address this, changing the mutation at that position from one that directly creates a nonsense codon to a mutation that still disrupts the observed ESE consensus site would provide insight into the possible presence of alternate splicing products.

The c.2388C>A mutation provided a unique challenge. To determine if this silent mutation was pathogenic, or if it was more likely that this patient harbored an as-yet undetermined mutation, it was important to recapitulate this mutation in a platform that could be manipulated. Indeed, stable cell lines created with the BECC vector containing this mutation showed very low *CPSI* expression when compared to wild-type (Figure 4.8) except under conditions of UPF2-mediated NMD silencing where transcript levels increased 20 to 35-fold over the non-treated cells, implicating degradation via NMD (Figure 4.10). The pathogenicity of this silent mutation indicates its location in an important sequence element for RNA stability. ESE*finder* shows the creation of a consensus recognition site for SRp40 and a disruption of a consensus recognition site for SC35, either or both possibly affect the splicing efficiency of this transcript [56]. Therefore, future BAC manipulation studies will prove useful in further characterizing this region of the *CPSI* gene and its role in RNA stability. Creating the consensus site for SRp40 by nucleotide substitutions at positions within the putative site other than at c.2388 would allow experimental determination of whether it is the creation of this SR protein recognition

sequence that elicits the RNA instability. Also, depleting the SC35 site without creating an SRp40 binding site could provide evidence for the role of SC35 in exon 19 definition. An additional experiment to address the relative functions of each putative SR protein binding site would be to deplete SC35 or SRp40 (either *in vitro* or through siRNA) and look for splicing changes, though these experiments would be likely to cause widespread effects on other genes as well. Further efforts to identify mRNA transcripts with activated cryptic splice sites may also help identify the precise RNA defect caused by this mutation.

A future direction for the BAC-based model system developed here is to show that the same construction can be applied to BACs containing other genes, as one of the advantages to constructing the BECC platform was that it could be applicable to various questions pertaining to the function of genetic changes. It will also be used in other gene expression studies for *CPSI*. The BECC vector will be used to examine effects of other mutations in this gene as well as to determine possible functional effects of polymorphisms identified in this gene, such as the functional polymorphism, T1405N, which has already been incorporated into the vector [6,12]. Also, it is a goal to begin enzyme expression studies and look beyond function of genetic variants at the RNA level.

The study of nonsense-mediated decay is likely to continue at a very rapid pace. A topic warranting further investigation is whether there are separable mechanisms leading to degradation, such as differences based on how the PTC was created (by a mutation directly, or due to a frameshift mutation that may or may not have been induced by a splicing error). One piece of evidence that

alternative NMD pathways may exist in the cell to recognize and destroy transcripts with PTCs is that there are some transcripts that do not require UPF2 for degradation [164]. This evidence is somewhat supported by previous data from another lab that showed UPF3 was able to coimmunoprecipitate UPF1 independent of UPF2 binding [103]. Using the constructs harboring the four *CPSI* mutations studied may provide helpful evidence concerning the alternate NMD pathways, especially since the PTCs resulting from these mutations are created by various mechanisms.

There are many genomic mutations that alter the processing of an RNA transcript rather than the structure and/or function of its protein product, and even mutations in the same gene that result in similar disease phenotypes through different mechanisms. Understanding the function of such diverse sets of genetic changes is an often difficult, yet vital, undertaking. Study of RNA processing mutations is effective not only for identifying important elements in the *CPSI* gene, but also contributes to the overall understanding of RNA processing. The data presented in the preceding Chapters supports the evidence that NMD is a common phenomenon that can alter disease phenotypes and stresses the importance of accounting for PTC-containing transcripts in all disease mutation interpretations and screening strategies.

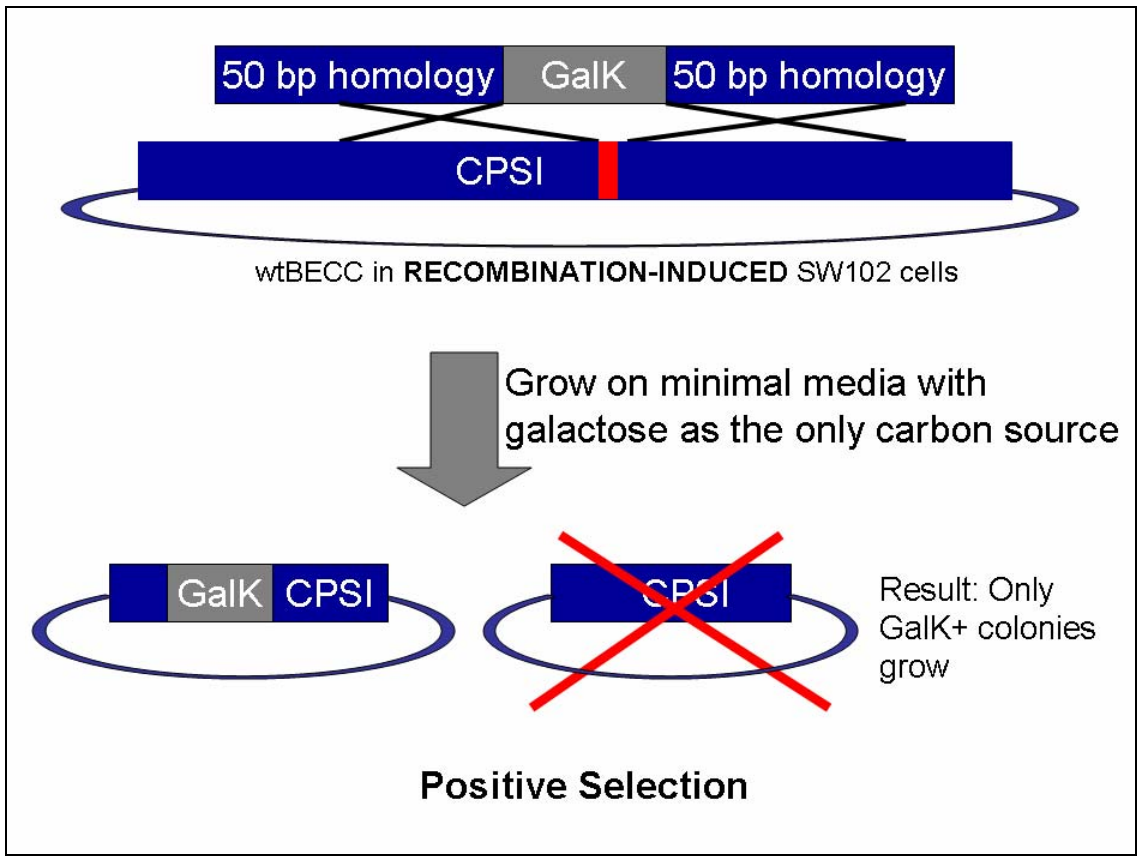
There were three main goals for this project. The first was to perform a survey of all mutations identified in CPSID patients and quantify those that likely elicit RNA instability. The second goal was to create a model system that would allow testing of any identified mutation, in part to determine its effect on RNA

processing. Last, it was important to use this model system to test putative RNA processing mutations to not only verify the system worked, but provide detail on the mechanistic effect of each mutation. The innovation of this platform is that it allows comprehensive yet efficient testing of any mutation in *CPSI*, irrespective of type or location. Though this system was designed for the study of *CPSI* mutations, its versatility will allow it to be applied to other genes and other scientific endeavors. Each of these goals was fulfilled to the extent that it supported the overall hypothesis of this project, that the underlying molecular mechanisms causing disease in a substantial fraction of CPSID patients are defects in RNA processing that can be studied using a novel BAC-based model system.

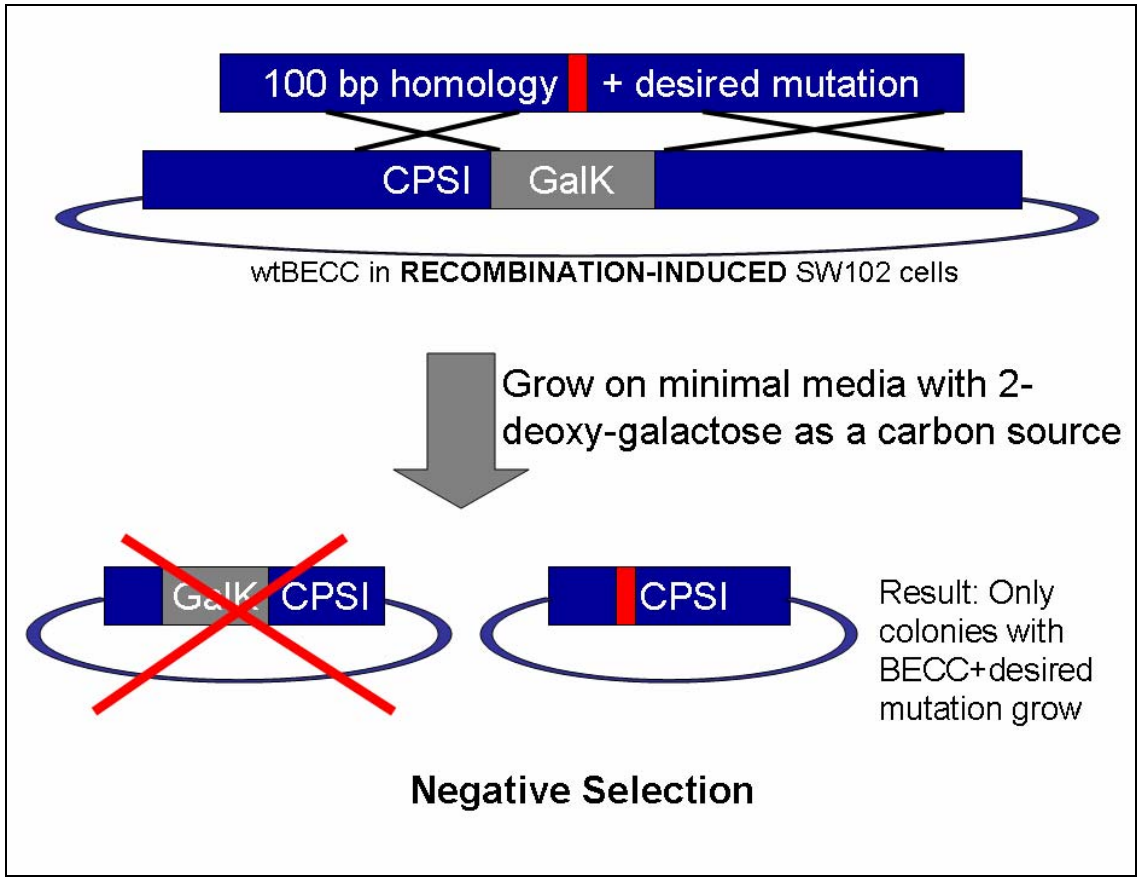
***GalK* mutagenesis protocol.**

The following is a schematic representation of the GalK site-directed mutagenesis protocol performed, as published in [147].

(1) SW102 cells containing BECC+WtCPSI are first grown at 42°C to induce high levels of recombination. Then, a PCR-generated 1.3 kb oligo containing the *galK* gene in the exact location of the desired mutation and 50 bp of homology on either side (from custom primer synthesis, see next section) is purified and electroporated into the SW102 cells. These cells are then spread onto minimal media plates with galactose as the only carbon source, permitting the specific growth of cells that have incorporated *galK*. *GalK* should be located in the position of the desired mutation.



(2) Double-stranded custom oligos containing the desired mutation and 50 bp of homology to the CPSI gene directly flanking the desired mutation are electroporated into SW102 cells selected from the previous homologous recombination step, also induced for high recombination. The cells are grown on minimal media with 2-DOG as the only carbon source, which is toxic in the presence of *galK* (due to phosphorylation that creates a non-metabolized compound). Only cells that have had *galK* replaced by the desired mutation are viable.



GalK mutagenesis primer sequences used.

The oligos ordered from Invitrogen are listed below. For each site-directed mutagenesis, four oligos were designed. Forward and reverse primers were designed for the PCR reaction used in the first homologous recombination step. At the 5' end, the forward primer contained 50 bp of homologous sequence to CPSI directly upstream of the desired mutation and the 3' end contained homology for amplification of the *gal*K sequence (in bold). At the 5' end, the reverse primer contained 50 bp of homologous sequence to CPSI directly downstream of the desired mutation while the 3' end again contained homology for amplification of the *gal*K sequence. Complementary oligos, each 100 bp, were designed for the second homologous recombination step. Each contained the desired mutation (underlined) in the middle and 50 bp on either side were homologous to CPSI at the region surrounding the desired point mutation.

(1) c.652-3T>G

Forward PCR:
5′-CAACACCTTTATCGTTGCTTCCTTTTAACTGTCTAATTTTTTTTAATTTGA
CCTGTTGACAATTAATCATCGGCA-3′

Reverse PCR:
5′-AAAATCCACAGGCTGACCTTCAAATTCAATCTTCCCAAGCATGGTACCCT
TCAGCACTGTCCTGCTCCTT-3'

Homology Arm (sense):
5′-CAACACCTTTATCGTTGCTTCCTTTTAACTGTCTAATTTTTTTTAATTTGA**G**
AGGGTACCATGCTTGGGAAGATTGAATTTGAAGGTCAGCCTGTGGATTT-3′

Homology Arm (rev. comp.):
5′-AAATCCACAGGCTGACCTTCAAATTCAATCTTCCCAAGCATGGTACCCT**C**
TCAAATTAAAAAAATTAGACAGTTAAAAGGAAGCAACGATAAAGGTGTTG-3′

(2) c.1210-1G>T

Forward PCR:
5′-TTGACATTCATTGTTACAGAAGGAATTTCTTCCTGTTTCTTATTCCTTTA
CCTGTTGACAATTAATCATCGGCA-3′

Reverse PCR:
5′-TCTGGGTGGAACTGCACAGCGAAGAAGGGTTTGCTCTCATGCATAATCCC
TCAGCACTGTCCTGCTCCTT-3′

Homology Arm (sense):
5′-GACATTCATTGTTACAGAAGGAATTTCTTCCTGTTTCTTATTCCTTTA**T**
GGGATTATGCATGAGAGCAAACCCTTCTTCGCTGTGCAGTTCCACCCAG-3′

Homology Arm (rev. comp.):
5′-CTGGGTGGAACTGCACAGCGAAGAAGGGTTTGCTCTCATGCATAATCCC**A**
TAAAGGAATAAGAAACAGGAAGAAATTCCTTCTGTAACAATGAATGTC-3′

(3) c.1893T>G

Forward PCR:
5′-CTGAAGGCAGCAGACACCATTGGCTACCCAGTGATGATCCGTTCCGCCTA
CCTGTTGACAATTAATCATCGGCA-3′

Reverse PCR:
5′-ATCAAAGTCTCTCTGTTGGGACAGATGCCTGAGCCTAACCCACCCAGTGC
TCAGCACTGTCCTGCTCCTT-3′

Homology Arm (sense):
5′-CTGAAGGCAGCAGACACCATTGGCTACCCAGTGATGATCCGTTCCGCCTA**G**
GCACTGGGTGGGTTAGGCTCAGGCATCTGTCCCAACAGAGAGACTTTGA-3′

Homology Arm (rev. comp.):
5′-TCAAAGTCTCTCTGTTGGGACAGATGCCTGAGCCTAACCCACCCAGTGC**C**
TAGGCGGAACGGATCATCACTGGGTAGCCAATGGTGTCTGCTGCCTTCAG-3′

(4) c.2388C>A

Forward PCR:
5′-GCAAAGATTGCCCTAGGAATCCCACTTCCAGAAATTAAGAACGTCGTATC
CCTGTTGACAATTAATCATCGGCA-3′

Reverse PCR:
5′-TTGGTGACCATGTAATCCAGGCTAGGTTCAAAACAGGCTGATGTCTTCCC
TCAGCACTGTCCTGCTCCTT-3′

Homology Arm (sense):
5′-GCAAAGATTGCCCTAGGAATCCCACTTCCAGAAATTAAGAACGTCGTATC**A**
GGGAAGACATCAGCCTGTTTTGAACCTAGCCTGGATTACATGGTCACCA-3′

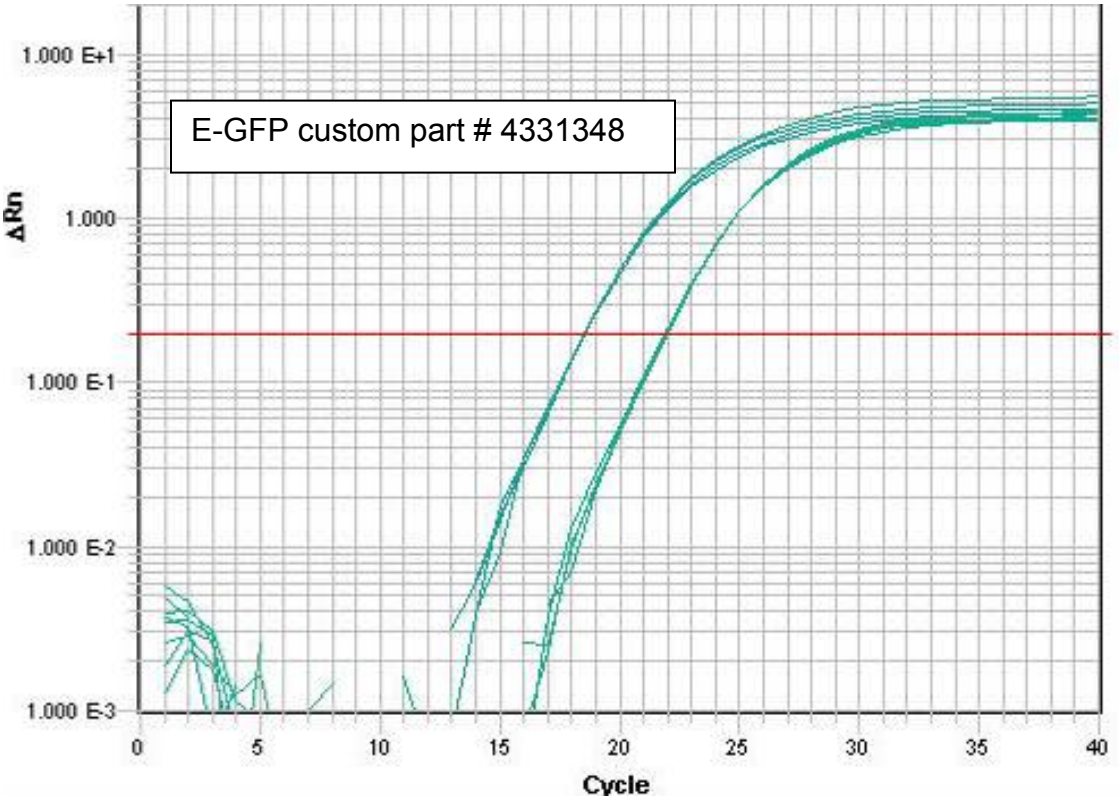
Homology Arm (rev. comp.):
5′-TGGTGACCATGTAATCCAGGCTAGGTTCAAAACAGGCTGATGTCTTCCC**T**
GATACGACGTTCTTAATTTCTGGAAGTGGGATTCCTAGGGCAATCTTTGC-3′

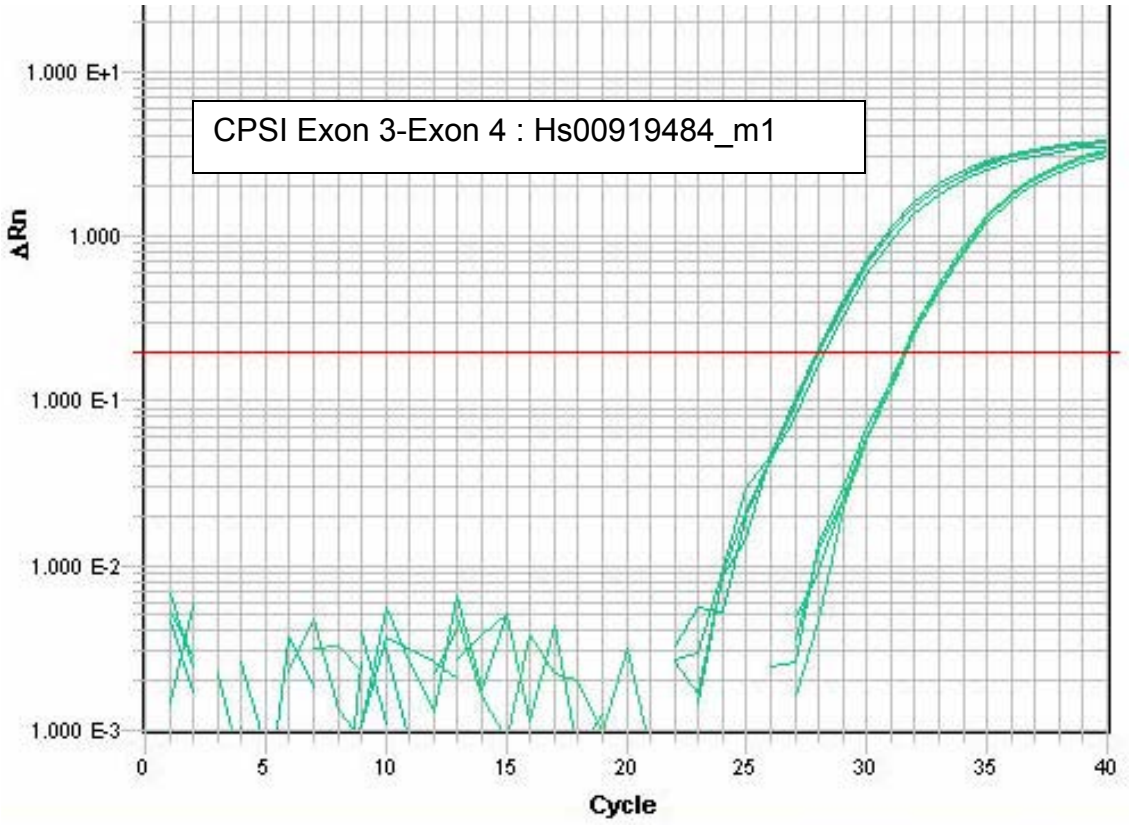
APPENDIX B

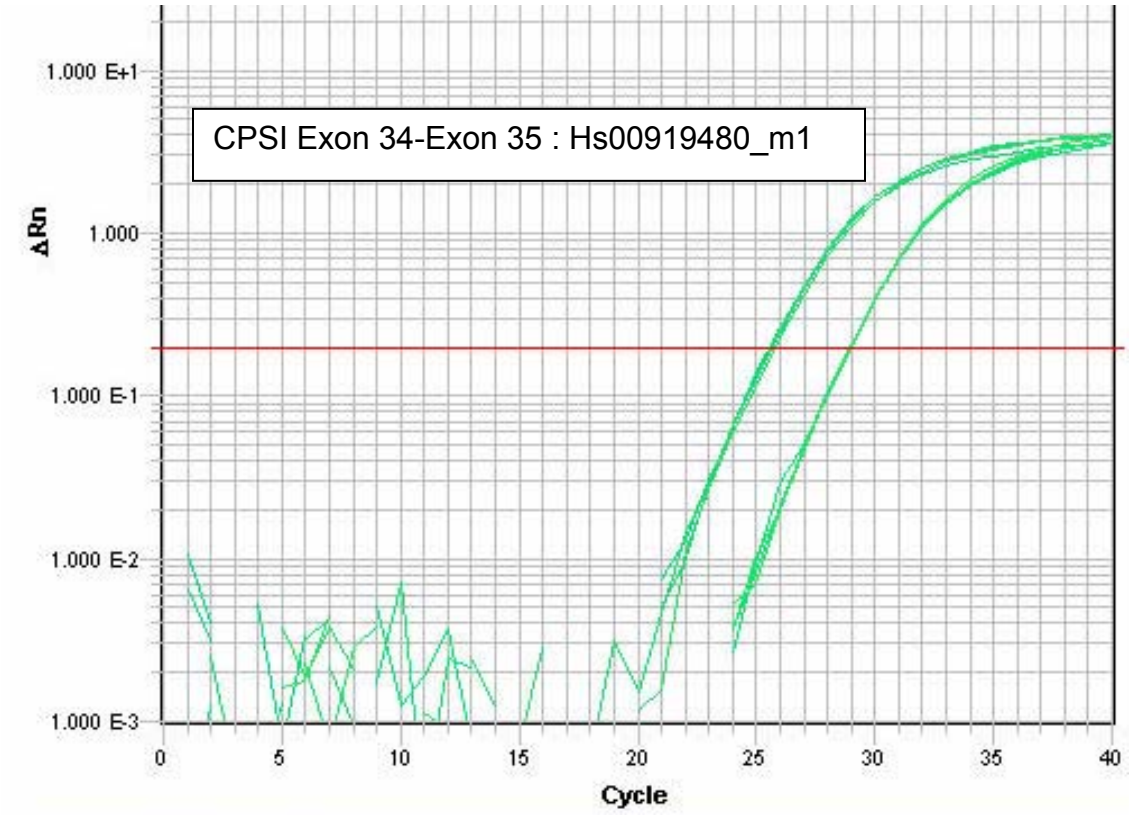
Quantitative RT/PCR assay quality control and methods.

The following images are raw data outputs from the Applied Biosystems PRISM Detection System (automated platform) during the real time RT/PCR reactions performed for Figure 4.7. Because this was the first set of real time RT/PCR experiments performed with the indicated probes, various quality control measures were taken. These controls are outlined below along with a general conception of the experimental protocol and calculations.

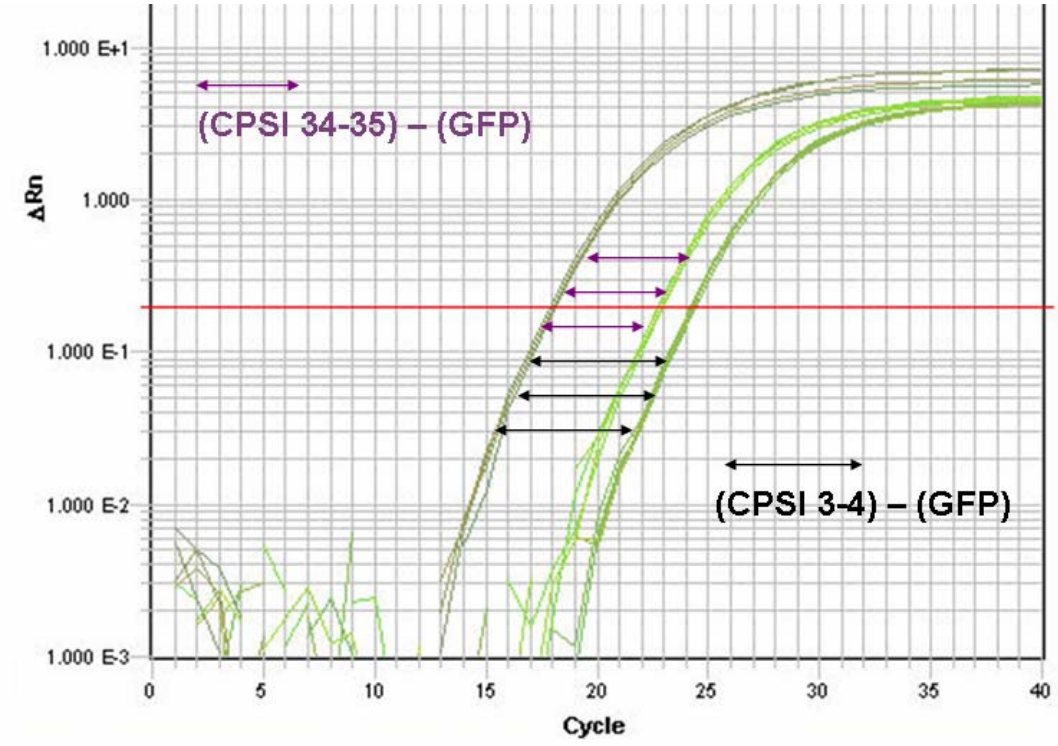
(1) A 10x dilution of the reverse transcription product (the PCR template) will yield a 3.3 cycle difference (X axis) if the indicated reaction is working at the proper efficiency:



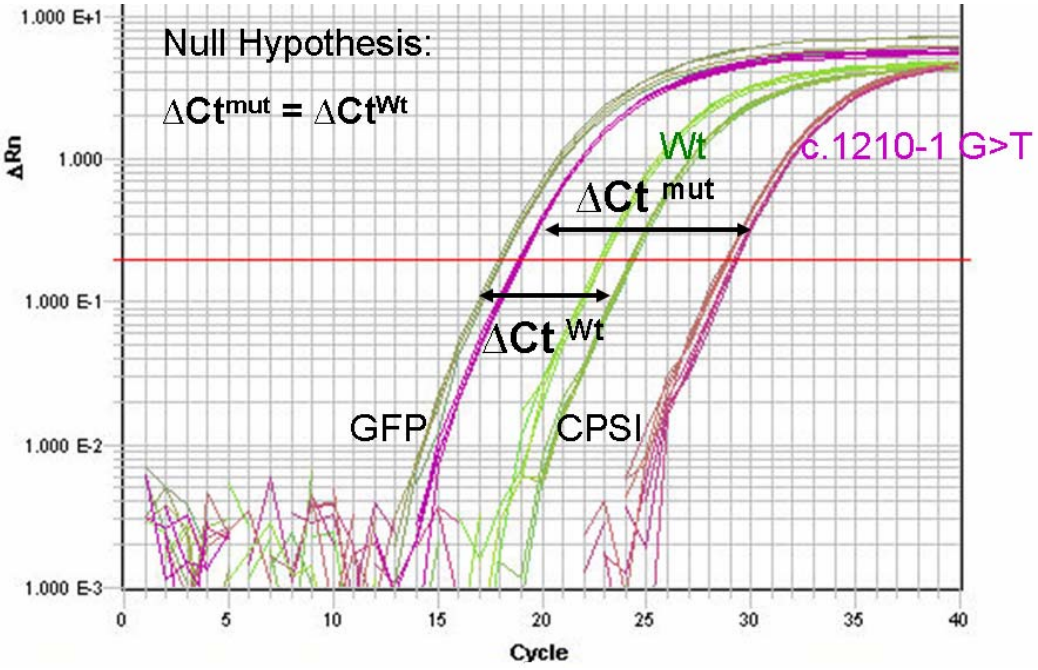

CPSI Exon 3-Exon 4 : Hs00919484_m1
1.000 E+1
1.000
1.000 E-1
1.000 E-2
1.000 E-3
ΔRn
0
5
10
15
20
25
30
35
40
Cycle


CPSI Exon 34-Exon 35 : Hs00919480_m1
1.000 E+1
1.000
1.000 E-1
1.000 E-2
1.000 E-3
ΔRn
0
5
10
15
20
25
30
35
40
Cycle

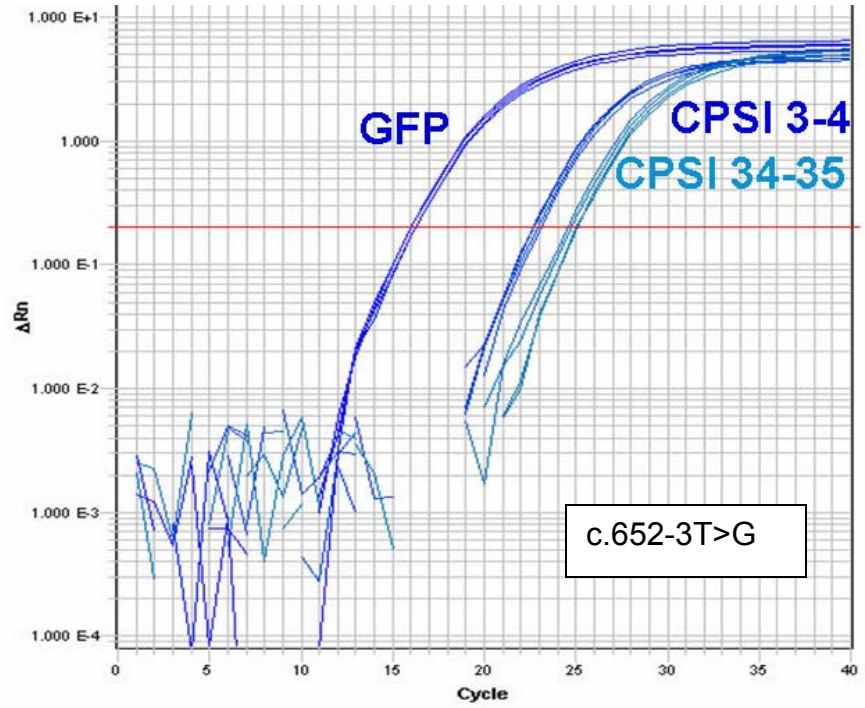
(2) The slope generated by each probe must be the same (indicated by equally sized arrows) for accurate calculations. Because data generated by both CPSI probes were standardized to E-GFP, the slopes generated by each probe were compared.

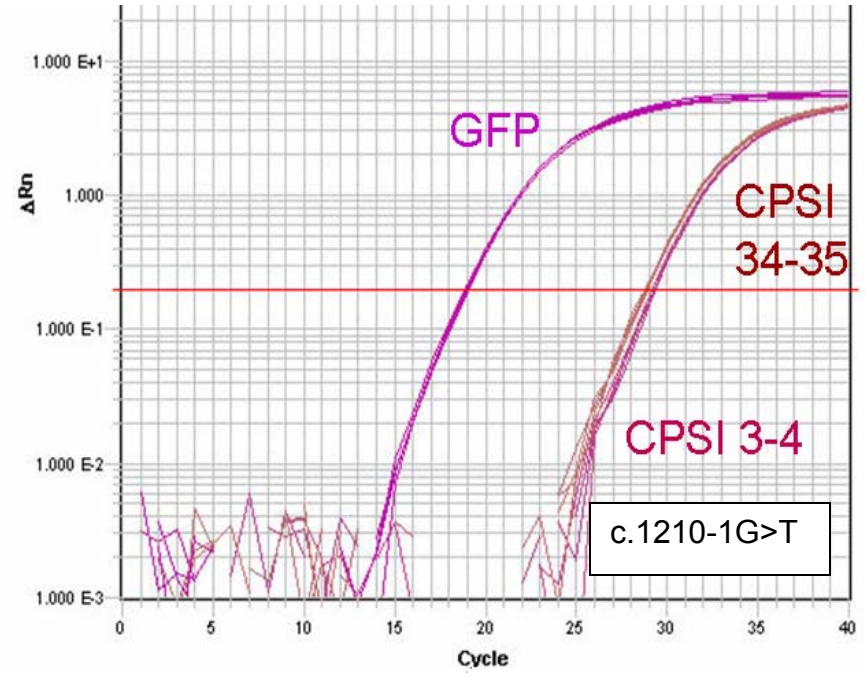


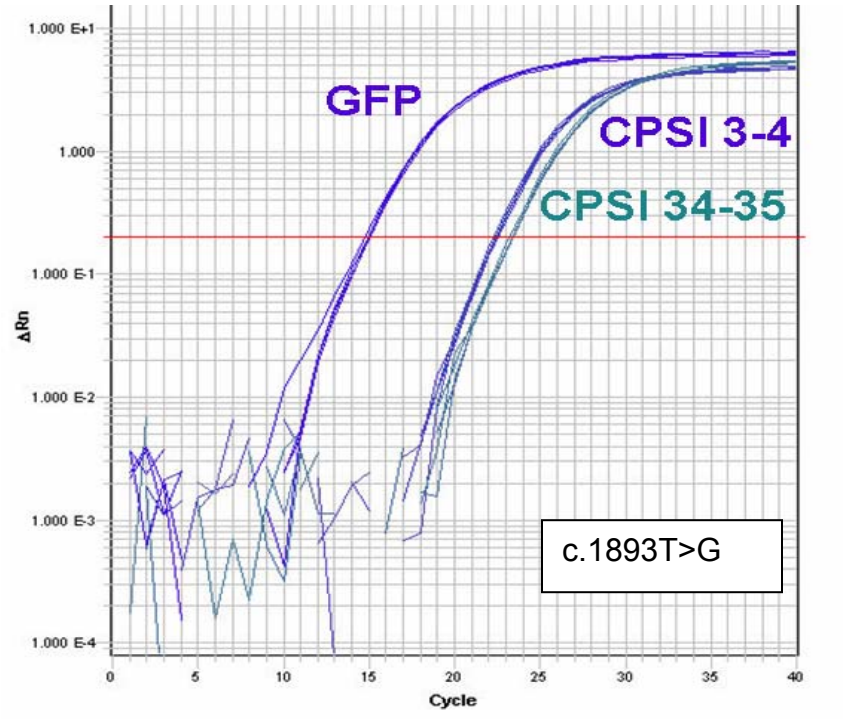
(3) The null hypothesis and necessary calculations are depicted graphically. The difference in Ct (ΔCt) between GFP and either CPSI probe is first calculated to standardize the data. If there is no difference in expression between the wild-type and each mutant cell line, each ΔCt will be equal (null hypothesis).

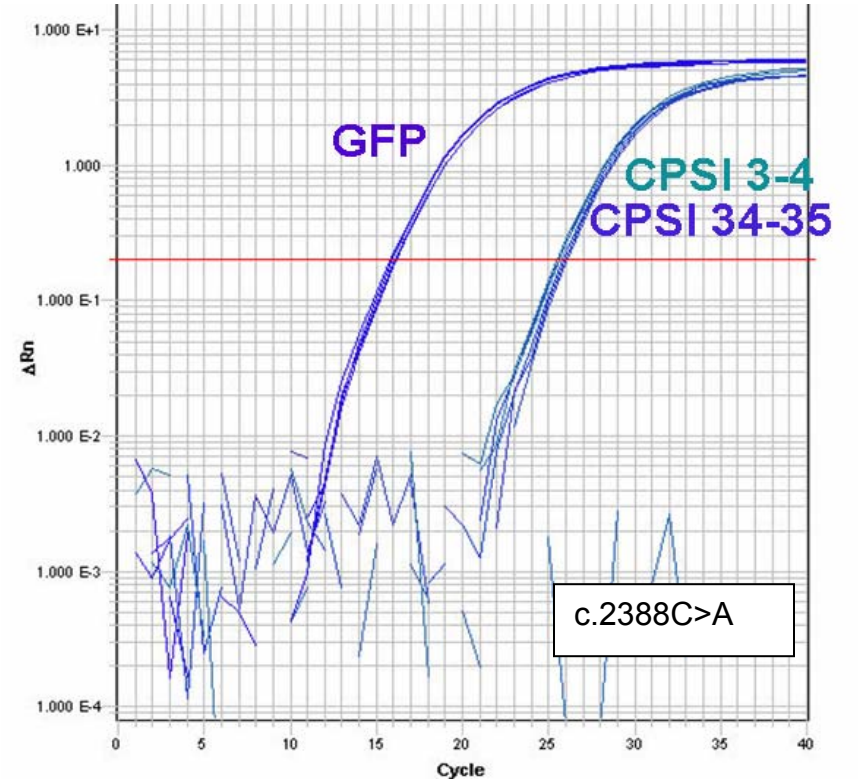


(4) Typical results are as depicted below using each polyclonal cell line cDNA generated with random hexamer primers as the template.






GFP
CPSI 3-4
CPSI 34-35
c.1893T>G
ΔRn
Cycle


GFP
CPSI 3-4
CPSI 34-35
c.2388C>A
ΔRn
Cycle

APPENDIX C

Quantitative RT/PCR formulas and values.

The following pages are the spreadsheet containing the formulas used to perform the quantitative RT/PCR ΔΔCt calculations following siRNA inhibition of UPF2 (see Chapter IV). Following the values calculated are the formula codes for each step. This data is included as a reference and example, as all qRT/PCR experiments were calculated in this manner.

Sample: The name of the sample being measured, each mutation is denoted by a 2 letter abbreviation, M = mock (or vehicle), no RT = control for gDNA contamination, si = samples transfected with siRNA

Detector: The TaqMan probe used

Ct: The cycle time calculated when the fluorescence crossed the threshold

Average Ct: The average of multiple Cts measured for each sample

No RT CF: No reverse transcription correction factor, measures gDNA contamination

CT+CF: The adjustment made to each Ct depending on the correction factor

Delta Ct: The difference between the CPSI and GFP Ct for each sample

Linear conv: Calculation made to convert exponential Ct to a linear value

Calibrator avg: The average linear conversion of the calibrator (no treatment sample) that all other samples will be expressed in relation to

ddCt: Delta Delta Ct value, where each sample is expressed in relation to the calibrator

ddCt avg: The average of each ddCt value for each sample

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
2	CA M	CPSI 3-4	28.25783			29.02028	6.187317	13.72245736		4.961795397	4.960416
3	CA M	CPSI 3-4	28.25957			29.02201	6.189054	13.70594551		4.955825007	
4	CA M	CPSI 3-4	28.2573			29.01974	6.186784	13.72752802		4.963628858	
5	CA M	CPSI 34-35	27.04044			27.04044	4.207483	54.127929		11.01428839	11.1014
6	CA M	CPSI 34-35	26.99159			26.99159	4.158633	55.99209612		11.39362073	
7	CA M	CPSI 34-35	27.05598			27.05598	4.223024	53.54798137		10.89627703	
8	CA M	GFP	16.13833	16.1782073		22.83296					
9	CA M	GFP	16.19029								
10	CA M	GFP	16.206								
11	CA M no RT	CPSI 3-4	40	39.2375567	0.7624433						
12	CA M no RT	CPSI 3-4	40								
13	CA M no RT	CPSI 3-4	37.71267								
14	CA M no RT	CPSI 34-35	40	40	0						
15	CA M no RT	CPSI 34-35	40								
16	CA M no RT	CPSI 34-35	40								
17	CA M no RT	GFP	33.71517	33.3452483	6.6547517						
18	CA M no RT	GFP	33.46709								
19	CA M no RT	GFP	32.85349								
20	CA no si	CPSI 3-4	26.95557			26.95557	8.413432	2.93294961	2.765623381		
21	CA no si	CPSI 3-4	27.1321			27.1321	8.589964	2.595154238			
22	CA no si	CPSI 3-4	27.03868			27.03868	8.496541	2.768766296			
23	CA no si	CPSI 34-35	25.5936			26.20422	7.66208	4.937238317	4.914337367		
24	CA no si	CPSI 34-35	25.56835			26.17897	7.636834	5.024396381			
25	CA no si	CPSI 34-35	25.63987			26.25049	7.708358	4.781377403			
26	CA no si	GFP	16.09406	16.141911		18.54214					
27	CA no si	GFP	16.19801								
28	CA no si	GFP	16.13367								
29	CA no si no RT	CPSI 3-4	40	40	0						
30	CA no si no RT	CPSI 3-4	40								
31	CA no si no RT	CPSI 3-4	40								
32	CA no si no RT	CPSI 34-35	38.16814	39.38938	0.61062						
33	CA no si no RT	CPSI 34-35	40								
34	CA no si no RT	CPSI 34-35	40								
35	CA no si no RT	GFP	40	37.599775	2.400225						
36	CA no si no RT	GFP	36.88326								
37	CA no si no RT	GFP	35.91607								
38	CA si	CPSI 3-4	28.93731			29.1526	4.163743	55.79413682		20.17416298	19.19937
39	CA si	CPSI 3-4	29.10841			29.32369	4.334842	49.55444676		17.91800254	
40	CA si	CPSI 3-4	28.98591			29.20119	4.212339	53.94605731		19.50593045	
41	CA si	CPSI 34-35	27.48784			27.48784	2.498982	176.9014363		35.99700694	34.85002
42	CA si	CPSI 34-35	27.59935			27.59935	2.610499	163.7424925		33.31934304	

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
43	CA si	CPSI 34-35	27.51876			27.51876	2.529902	173.1504049		35.2337237	
44	CA si	GFP	15.54712	15.568475		24.98885					
45	CA si	GFP	15.57975								
46	CA si	GFP	15.57855								
47	CA si no RT	CPSI 3-4	40	39.7847177	0.2152823						
48	CA si no RT	CPSI 3-4	40								
49	CA si no RT	CPSI 3-4	39.35415								
50	CA si no RT	CPSI 34-35	40	40	0						
51	CA si no RT	CPSI 34-35	40								
52	CA si no RT	CPSI 34-35	40								
53	CA si no RT	GFP	30.12634	30.5796223	9.4203777						
54	CA si no RT	GFP	30.75076								
55	CA si no RT	GFP	30.86177								
56	I10 M	CPSI 3-4	27.39708			27.39708	6.732346	9.405069586		1.702099468	1.634618
57	I10 M	CPSI 3-4	27.4601			27.4601	6.795361	9.00311107		1.629354299	
58	I10 M	CPSI 3-4	27.51143			27.51143	6.846692	8.688412611		1.572401177	
59	I10 M	CPSI 34-35	26.30863			26.30863	5.643891	19.99952205		2.440450669	2.337627
60	I10 M	CPSI 34-35	26.45565			26.45565	5.790913	18.06182313		2.204002086	
61	I10 M	CPSI 34-35	26.35185			26.35185	5.687108	19.40930585		2.368429272	
62	I10 M	GFP	17.65021	17.7114343		20.66474					
63	I10 M	GFP	17.76632								
64	I10 M	GFP	17.71778								
65	I10 M no RT	CPSI 3-4	40	40	0						
66	I10 M no RT	CPSI 3-4	40								
67	I10 M no RT	CPSI 3-4	40								
68	I10 M no RT	CPSI 34-35	40	40	0						
69	I10 M no RT	CPSI 34-35	40								
70	I10 M no RT	CPSI 34-35	40								
71	I10 M no RT	GFP	36.26555	37.046697	2.953303						
72	I10 M no RT	GFP	40								
73	I10 M no RT	GFP	34.87455								
74	I10 no si	CPSI 3-4	26.30497			26.30497	7.470569	5.638124301	5.525569899		
75	I10 no si	CPSI 3-4	26.49024			26.49024	7.655847	4.958615239			
76	I10 no si	CPSI 3-4	26.22004			26.22004	7.385646	5.979970157			
77	I10 no si	CPSI 34-35	25.87254			25.87254	7.038143	7.60865414	8.19501181		
78	I10 no si	CPSI 34-35	25.74898			25.74898	6.914583	8.289018511			
79	I10 no si	CPSI 34-35	25.68126			25.68126	6.846866	8.687362778			
80	I10 no si	GFP	18.86411	18.834397		18.8344					
81	I10 no si	GFP	18.82237								
82	I10 no si	GFP	18.81672								
83	I10 no si no RT	CPSI 3-4	40	40	0						

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
84	I10 no si no RT	CPSI 3-4	40								
85	I10 no si no RT	CPSI 3-4	40								
86	I10 no si no RT	CPSI 34-35	40	40	0						
87	I10 no si no RT	CPSI 34-35	40								
88	I10 no si no RT	CPSI 34-35	40								
89	I10 no si no RT	GFP	40	40	0						
90	I10 no si no RT	GFP	40								
91	I10 no si no RT	GFP	40								
92	I10 si	CPSI 3-4	27.14868			31.2575	4.842857	34.84614812		6.30634464	6.253101
93	I10 si	CPSI 3-4	27.24123			31.35005	4.935405	32.68097873		5.914499196	
94	I10 si	CPSI 3-4	27.09654			31.20535	4.79071	36.12872118		6.538460619	
95	I10 si	CPSI 34-35	26.88474			27.66325	1.248605	420.8548553		51.35500291	56.52395
96	I10 si	CPSI 34-35	26.83411			27.61261	1.19797	435.888083		53.18943927	
97	I10 si	CPSI 34-35	26.5442			27.3227	0.908062	532.900343		65.0274039	
98	I10 si	GFP	17.05324	17.0013433		26.41464					
99	I10 si	GFP	17.03269								
100	I10 si	GFP	16.9181								
101	I10 si no RT	CPSI 3-4	35.75824	35.891186	4.108814						
102	I10 si no RT	CPSI 3-4	35.32469								
103	I10 si no RT	CPSI 3-4	36.59063								
104	I10 si no RT	CPSI 34-35	37.66449	39.2214967	0.7785033						
105	I10 si no RT	CPSI 34-35	40								
106	I10 si no RT	CPSI 34-35	40								
107	I10 si no RT	GFP	30.4953	30.5867023	9.4132977						
108	I10 si no RT	GFP	30.63245								
109	I10 si no RT	GFP	30.63236								
110	I5 M	CPSI 3-4	26.45797			27.77712	6.329867	12.43140354		0.369304579	0.360955
111	I5 M	CPSI 3-4	26.538			27.85714	6.409893	11.76061243		0.349377124	
112	I5 M	CPSI 3-4	26.47812			27.79727	6.350016	12.25899057		0.364182639	
113	I5 M	CPSI 34-35	26.88652			26.88652	5.439266	23.04717925		0.835995154	0.797085
114	I5 M	CPSI 34-35	26.99311			26.99311	5.545864	21.40565809		0.776451913	
115	I5 M	CPSI 34-35	26.98874			26.98874	5.541494	21.47059526		0.778807393	
116	I5 M	GFP	18.16435	18.156284		21.44725					
117	I5 M	GFP	18.22585								
118	I5 M	GFP	18.07866								
119	I5 M no RT	CPSI 3-4	39.53389	38.680854	1.319146						
120	I5 M no RT	CPSI 3-4	37.72616								
121	I5 M no RT	CPSI 3-4	38.78252								
122	I5 M no RT	CPSI 34-35	40	40	0						
123	I5 M no RT	CPSI 34-35	40								
124	I5 M no RT	CPSI 34-35	40								

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
125	I5 M no RT	GFP	34.73657	36.709035	3.290965						
126	I5 M no RT	GFP	35.39053								
127	I5 M no RT	GFP	40								
128	I5 no si	CPSI 3-4	25.24652			25.80662	4.858525	34.46976643	33.6616556		
129	I5 no si	CPSI 3-4	25.35049			25.91059	4.962491	32.07313824			
130	I5 no si	CPSI 3-4	25.24768			25.80778	4.859685	34.44206213			
131	I5 no si	CPSI 34-35	25.31529			26.04539	5.097291	29.21207534	27.56855604		
132	I5 no si	CPSI 34-35	25.34956			26.07966	5.131568	28.52620552			
133	I5 no si	CPSI 34-35	25.54181			26.27191	5.323811	24.96738725			
134	I5 no si	GFP	18.44052	18.45936		20.9481					
135	I5 no si	GFP	18.51992								
136	I5 no si	GFP	18.41764								
137	I5 no si no RT	CPSI 3-4	40	39.439902	0.560098						
138	I5 no si no RT	CPSI 3-4	40								
139	I5 no si no RT	CPSI 3-4	38.31971								
140	I5 no si no RT	CPSI 34-35	37.80969	39.2698973	0.7301027						
141	I5 no si no RT	CPSI 34-35	40								
142	I5 no si no RT	CPSI 34-35	40								
143	I5 no si no RT	GFP	40	37.5112637	2.4887363						
144	I5 no si no RT	GFP	35.79591								
145	I5 no si no RT	GFP	36.73789								
146	I5 si	CPSI 3-4	25.87928			25.87928	1.62773	323.5970438		9.613224246	10.18705
147	I5 si	CPSI 3-4	25.78269			25.78269	1.531142	346.0034512		10.27886018	
148	I5 si	CPSI 3-4	25.72894			25.72894	1.477389	359.1382781		10.66906163	
149	I5 si	CPSI 34-35	26.80668			26.80668	2.555132	170.1487353		6.171840668	5.482677
150	I5 si	CPSI 34-35	26.97735			26.97735	2.725796	151.165868		5.483271153	
151	I5 si	CPSI 34-35	27.17148			27.17148	2.919928	132.1338798		4.792919861	
152	I5 si	GFP	17.86009	17.80434		24.25155					
153	I5 si	GFP	17.76037								
154	I5 si	GFP	17.79256								
155	I5 si no RT	CPSI 3-4	40	40	0						
156	I5 si no RT	CPSI 3-4	40								
157	I5 si no RT	CPSI 3-4	40								
158	I5 si no RT	CPSI 34-35	40	40	0						
159	I5 si no RT	CPSI 34-35	40								
160	I5 si no RT	CPSI 34-35	40								
161	I5 si no RT	GFP	33.63929	33.5527877	6.4472123						
162	I5 si no RT	GFP	33.55373								
163	I5 si no RT	GFP	33.46534								
164	TG M	CPSI 3-4	25.21843			25.21843	3.479893	89.62886059		30.91708731	30.9069
165	TG M	CPSI 3-4	25.17748			25.17748	3.438941	92.20949761		31.80726687	

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
166	TG M	CPSI 3-4	25.26205			25.26205	3.523511	86.95960315		29.99633851	
167	TG M	CPSI 34-35	24.85817			24.85817	3.119633	115.0527336		46.76145642	46.80311
168	TG M	CPSI 34-35	24.82698			24.82698	3.088437	117.5676598		47.78361041	
169	TG M	CPSI 34-35	24.88612			24.88612	3.147582	112.8452945		45.86427595	
170	TG M	GFP	15.02686	14.9857895		21.73854					
171	TG M	GFP	14.95589								
172	TG M	GFP	14.97462								
173	TG M no RT	CPSI 3-4	40	40	0						
174	TG M no RT	CPSI 3-4	40								
175	TG M no RT	CPSI 3-4	40								
176	TG M no RT	CPSI 34-35	40	40	0						
177	TG M no RT	CPSI 34-35	40								
178	TG M no RT	CPSI 34-35	40								
179	TG M no RT	GFP	33.01587	33.2472513	6.7527487						
180	TG M no RT	GFP	33.29786								
181	TG M no RT	GFP	33.42802								
182	TG no si	CPSI 3-4	23.98472			23.98472	8.327513	3.112926006	2.899007261		
183	TG no si	CPSI 3-4	24.21031			24.21031	8.553099	2.662322393			
184	TG no si	CPSI 3-4	24.07615			24.07615	8.41894	2.921773382			
185	TG no si	CPSI 34-35	24.28881			24.28881	8.6316	2.521328828	2.460418096		
186	TG no si	CPSI 34-35	24.43029			24.43029	8.773076	2.285814429			
187	TG no si	CPSI 34-35	24.25892			24.25892	8.60171	2.574111033			
188	TG no si	GFP	15.66562	15.657209		15.65721					
189	TG no si	GFP	15.70701								
190	TG no si	GFP	15.59899								
191	TG no si no RT	CPSI 3-4	40	40	0						
192	TG no si no RT	CPSI 3-4	40								
193	TG no si no RT	CPSI 3-4	40								
194	TG no si no RT	CPSI 34-35	40	40	0						
195	TG no si no RT	CPSI 34-35	40								
196	TG no si no RT	CPSI 34-35	40								
197	TG no si no RT	GFP	40	40	0						
198	TG no si no RT	GFP	40								
199	TG no si no RT	GFP	40								
200	TG si	CPSI 3-4	26.04842			26.04842	0.897158	536.943304		185.2162674	184.1025
201	TG si	CPSI 3-4	26.16829			26.16829	1.017023	494.1348338		170.449671	
202	TG si	CPSI 3-4	25.96206			25.96206	0.810801	570.0651319		196.6414985	
203	TG si	CPSI 34-35	25.77958			25.77958	0.628318	646.9300652		262.9350134	227.4306
204	TG si	CPSI 34-35	26.16316			26.16316	1.0119	495.8926223		201.5481121	
205	TG si	CPSI 34-35	26.05122			26.05122	0.899962	535.9007227		217.8088039	
206	TG si	GFP	14.95936	15.0001357		25.15126					

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
207	TG si	GFP	14.95555								
208	TG si	GFP	15.0855								
209	TG si no RT	CPSI 3-4	40	40	0						
210	TG si no RT	CPSI 3-4	40								
211	TG si no RT	CPSI 3-4	40								
212	TG si no RT	CPSI 34-35	40	40	0						
213	TG si no RT	CPSI 34-35	40								
214	TG si no RT	CPSI 34-35	40								
215	TG si no RT	GFP	29.8265	29.848874	10.151126						
216	TG si no RT	GFP	29.84279								
217	TG si no RT	GFP	29.87734								
218	Wt M	CPSI 3-4	23.60367			24.20676	0.682951	622.8898634		2.251422629	2.139149
219	Wt M	CPSI 3-4	23.77016			24.37324	0.849434	555.0024326		2.006044904	
220	Wt M	CPSI 3-4	23.66349			24.26658	0.742771	597.5904509		2.159978422	
221	Wt M	CPSI 34-35	22.1642			22.82597	0.697838	616.4955275		1.024613522	1.106754
222	Wt M	CPSI 34-35	22.12168			22.78345	0.740359	598.5905189		0.994855457	
223	Wt M	CPSI 34-35	22.50851			23.17028	0.353527	782.6685251		1.300792493	
224	Wt M	GFP	16.88582	16.805665		23.52381					
225	Wt M	GFP	16.71724								
226	Wt M	GFP	16.81394								
227	Wt M no RT	CPSI 3-4	38.19074	39.396914	0.603086						
228	Wt M no RT	CPSI 3-4	40								
229	Wt M no RT	CPSI 3-4	40								
230	Wt M no RT	CPSI 34-35	38.01468	39.3382267	0.6617733						
231	Wt M no RT	CPSI 34-35	40								
232	Wt M no RT	CPSI 34-35	40								
233	Wt M no RT	GFP	32.41799	33.281856	6.718144						
234	Wt M no RT	GFP	34.0372								
235	Wt M no RT	GFP	33.39038								
236	Wt no si	CPSI 3-4	22.79945			22.79945	1.765695	294.0850464	276.6650097		
237	Wt no si	CPSI 3-4	22.96254			22.96254	1.928785	262.6503357			
238	Wt no si	CPSI 3-4	22.90541			22.90541	1.871656	273.2596469			
239	Wt no si	CPSI 34-35	20.27426			20.27426	0.75949	590.7049746	601.685918		
240	Wt no si	CPSI 34-35	20.53467			20.53467	0.499083	707.5562089			
241	Wt no si	CPSI 34-35	20.05323			20.05323	0.980521	506.7965704			
242	Wt no si	GFP	16.87989	16.898732		21.03375					
243	Wt no si	GFP	16.93645								
244	Wt no si	GFP	16.87986								
245	Wt no si no RT	CPSI 3-4	40	40	0						
246	Wt no si no RT	CPSI 3-4	40								
247	Wt no si no RT	CPSI 3-4	40								

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
248	Wt no si no RT	CPSI 34-35	40	40	0						
249	Wt no si no RT	CPSI 34-35	40								
250	Wt no si no RT	CPSI 34-35	40								
251	Wt no si no RT	GFP	35.8525	35.8649807	4.1350193						
252	Wt no si no RT	GFP	36.90489								
253	Wt no si no RT	GFP	34.83756								
254	Wt si	CPSI 3-4	22.91344			23.29782	0.186539	878.7112075		3.176083628	3.09636
255	Wt si	CPSI 3-4	23.05441			23.43879	0.327509	796.9112671		2.880419422	
256	Wt si	CPSI 3-4	22.888			23.27238	0.161103	894.3410488		3.232577368	
257	Wt si	CPSI 34-35	21.72753			21.72753	1.383752	383.2208602		0.636911799	0.628733
258	Wt si	CPSI 34-35	21.69313			21.69313	1.418155	374.1905423		0.62190344	
259	Wt si	CPSI 34-35	21.70578			21.70578	1.4055	377.4872966		0.627382635	
260	Wt si	GFP	16.31765	16.3284933		23.11128					
261	Wt si	GFP	16.32774								
262	Wt si	GFP	16.34009								
263	Wt si no RT	CPSI 3-4	38.84685	39.615616	0.384384						
264	Wt si no RT	CPSI 3-4	40								
265	Wt si no RT	CPSI 3-4	40								
266	Wt si no RT	CPSI 34-35	40	40	0						
267	Wt si no RT	CPSI 34-35	40								
268	Wt si no RT	CPSI 34-35	40								
269	Wt si no RT	GFP	33.46902	33.2172133	6.7827867						
270	Wt si no RT	GFP	33.28962								
271	Wt si no RT	GFP	32.893								

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
2	CA M	CPSI 3-4	28.257833			=C2+E11	=ABS(F2-F8)	=(2^-G2)*1000		=H2/I20	=AVERAGE(J2:J4)
3	CA M	CPSI 3-4	28.25957			=C3+E11	=ABS(F3-F8)	=(2^-G3)*1000		=H3/I20	
4	CA M	CPSI 3-4	28.2573			=C4+E11	=ABS(F4-F8)	=(2^-G4)*1000		=H4/I20	
5	CA M	CPSI 34-35	27.040442			=C5+E14	=ABS(F5-F8)	=(2^-G5)*1000		=H5/I23	=AVERAGE(J5:J7)
6	CA M	CPSI 34-35	26.991592			=C6+E14	=ABS(F6-F8)	=(2^-G6)*1000		=H6/I23	
7	CA M	CPSI 34-35	27.055983			=C7+E14	=ABS(F7-F8)	=(2^-G7)*1000		=H7/I23	
8	CA M	GFP	16.138329	=AVERAGE(C8:C10)		=D8+E17					
9	CA M	GFP	16.19029								
10	CA M	GFP	16.206003								
11	CA M no RT	CPSI 3-4	40	=AVERAGE(C11:C13)	=40-D11						
12	CA M no RT	CPSI 3-4	40								
13	CA M no RT	CPSI 3-4	37.71267								
14	CA M no RT	CPSI 34-35	40	=AVERAGE(C14:C16)	=40-D14						
15	CA M no RT	CPSI 34-35	40								
16	CA M no RT	CPSI 34-35	40								
17	CA M no RT	GFP	33.71517	=AVERAGE(C17:C19)	=40-D17						
18	CA M no RT	GFP	33.46709								
19	CA M no RT	GFP	32.853485								
20	CA no si	CPSI 3-4	26.955568			=C20+E29	=ABS(F20-F26)	=(2^-G20)*1000	=AVERAGE(H20:H22)		
21	CA no si	CPSI 3-4	27.1321			=C21+E29	=ABS(F21-F26)	=(2^-G21)*1000			
22	CA no si	CPSI 3-4	27.038677			=C22+E29	=ABS(F22-F26)	=(2^-G22)*1000			
23	CA no si	CPSI 34-35	25.593596			=C23+E32	=ABS(F23-F26)	=(2^-G23)*1000	=AVERAGE(H23:H25)		
24	CA no si	CPSI 34-35	25.56835			=C24+E32	=ABS(F24-F26)	=(2^-G24)*1000			
25	CA no si	CPSI 34-35	25.639874			=C25+E32	=ABS(F25-F26)	=(2^-G25)*1000			
26	CA no si	GFP	16.094057	=AVERAGE(C26:C28)		=D26+E35					
27	CA no si	GFP	16.198006								
28	CA no si	GFP	16.13367								
29	CA no si no RT	CPSI 3-4	40	=AVERAGE(C29:C31)	=40-D29						
30	CA no si no RT	CPSI 3-4	40								
31	CA no si no RT	CPSI 3-4	40								
32	CA no si no RT	CPSI 34-35	38.16814	=AVERAGE(C32:C34)	=40-D32						
33	CA no si no RT	CPSI 34-35	40								
34	CA no si no RT	CPSI 34-35	40								
35	CA no si no RT	GFP	40	=AVERAGE(C35:C37)	=40-D35						
36	CA no si no RT	GFP	36.88326								
37	CA no si no RT	GFP	35.916065								
38	CA si	CPSI 3-4	28.937313			=C38+E47	=ABS(F38-F44)	=(2^-G38)*1000		=H38/I20	=AVERAGE(J38:J40)
39	CA si	CPSI 3-4	29.108412			=C39+E47	=ABS(F39-F44)	=(2^-G39)*1000		=H39/I20	
40	CA si	CPSI 3-4	28.985909			=C40+E47	=ABS(F40-F44)	=(2^-G40)*1000		=H40/I20	
41	CA si	CPSI 34-35	27.487835			=C41+E50	=ABS(F41-F44)	=(2^-G41)*1000		=H41/I23	=AVERAGE(J41:J43)
42	CA si	CPSI 34-35	27.599352			=C42+E50	=ABS(F42-F44)	=(2^-G42)*1000		=H42/I23	
43	CA si	CPSI 34-35	27.518755			=C43+E50	=ABS(F43-F44)	=(2^-G43)*1000		=H43/I23	
44	CA si	GFP	15.547122	=AVERAGE(C44:C46)		=D44+E53					
45	CA si	GFP	15.57975								
46	CA si	GFP	15.578553								
47	CA si no RT	CPSI 3-4	40	=AVERAGE(C47:C49)	=40-D47						
48	CA si no RT	CPSI 3-4	40								
49	CA si no RT	CPSI 3-4	39.354153								
50	CA si no RT	CPSI 34-35	40	=AVERAGE(C50:C52)	=40-D50						

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
51	CA si no RT	CPSI 34-35	40								
52	CA si no RT	CPSI 34-35	40								
53	CA si no RT	GFP	30.126337	=AVERAGE(C53:C55)	=40-D53						
54	CA si no RT	GFP	30.750761								
55	CA si no RT	GFP	30.861769								
56	I10 M	CPSI 3-4	27.397083			=C56+E65	=ABS(F56-F62)	=(2^-G56)*1000		=H56/I74	=AVERAGE(J56:J58)
57	I10 M	CPSI 3-4	27.460098			=C57+E65	=ABS(F57-F62)	=(2^-G57)*1000		=H57/I74	
58	I10 M	CPSI 3-4	27.511429			=C58+E65	=ABS(F58-F62)	=(2^-G58)*1000		=H58/I74	
59	I10 M	CPSI 34-35	26.308628			=C59+E68	=ABS(F59-F62)	=(2^-G59)*1000		=H59/I77	=AVERAGE(J59:J61)
60	I10 M	CPSI 34-35	26.45565			=C60+E68	=ABS(F60-F62)	=(2^-G60)*1000		=H60/I77	
61	I10 M	CPSI 34-35	26.351845			=C61+E68	=ABS(F61-F62)	=(2^-G61)*1000		=H61/I77	
62	I10 M	GFP	17.650208	=AVERAGE(C62:C64)		=D62+E71					
63	I10 M	GFP	17.766315								
64	I10 M	GFP	17.71778								
65	I10 M no RT	CPSI 3-4	40	=AVERAGE(C65:C67)	=40-D65						
66	I10 M no RT	CPSI 3-4	40								
67	I10 M no RT	CPSI 3-4	40								
68	I10 M no RT	CPSI 34-35	40	=AVERAGE(C68:C70)	=40-D68						
69	I10 M no RT	CPSI 34-35	40								
70	I10 M no RT	CPSI 34-35	40								
71	I10 M no RT	GFP	36.265545	=AVERAGE(C71:C73)	=40-D71						
72	I10 M no RT	GFP	40								
73	I10 M no RT	GFP	34.874546								
74	I10 no si	CPSI 3-4	26.304966			=C74+E83	=ABS(F74-F80)	=(2^-G74)*1000	=AVERAGE(H74:H76)		
75	I10 no si	CPSI 3-4	26.490244			=C75+E83	=ABS(F75-F80)	=(2^-G75)*1000			
76	I10 no si	CPSI 3-4	26.220043			=C76+E83	=ABS(F76-F80)	=(2^-G76)*1000			
77	I10 no si	CPSI 34-35	25.87254			=C77+E86	=ABS(F77-F80)	=(2^-G77)*1000	=AVERAGE(H77:H79)		
78	I10 no si	CPSI 34-35	25.74898			=C78+E86	=ABS(F78-F80)	=(2^-G78)*1000			
79	I10 no si	CPSI 34-35	25.681263			=C79+E86	=ABS(F79-F80)	=(2^-G79)*1000			
80	I10 no si	GFP	18.864105	=AVERAGE(C80:C82)		=D80+E89					
81	I10 no si	GFP	18.822365								
82	I10 no si	GFP	18.816721								
83	I10 no si no RT	CPSI 3-4	40	=AVERAGE(C83:C85)	=40-D83						
84	I10 no si no RT	CPSI 3-4	40								
85	I10 no si no RT	CPSI 3-4	40								
86	I10 no si no RT	CPSI 34-35	40	=AVERAGE(C86:C88)	=40-D86						
87	I10 no si no RT	CPSI 34-35	40								
88	I10 no si no RT	CPSI 34-35	40								
89	I10 no si no RT	GFP	40	=AVERAGE(C89:C91)	=40-D89						
90	I10 no si no RT	GFP	40								
91	I10 no si no RT	GFP	40								
92	I10 si	CPSI 3-4	27.148684			=C92+E101	=ABS(F92-F98)	=(2^-G92)*1000		=H92/I74	=AVERAGE(J92:J94)
93	I10 si	CPSI 3-4	27.241232			=C93+E101	=ABS(F93-F98)	=(2^-G93)*1000		=H93/I74	
94	I10 si	CPSI 3-4	27.096537			=C94+E101	=ABS(F94-F98)	=(2^-G94)*1000		=H94/I74	
95	I10 si	CPSI 34-35	26.884743			=C95+E104	=ABS(F95-F98)	=(2^-G95)*1000		=H95/I77	=AVERAGE(J95:J97)
96	I10 si	CPSI 34-35	26.834108			=C96+E104	=ABS(F96-F98)	=(2^-G96)*1000		=H96/I77	
97	I10 si	CPSI 34-35	26.5442			=C97+E104	=ABS(F97-F98)	=(2^-G97)*1000		=H97/I77	
98	I10 si	GFP	17.053242	=AVERAGE(C98:C100)		=D98+E107					
99	I10 si	GFP	17.03269								

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
100	I10 si	GFP	16.918098								
101	I10 si no RT	CPSI 3-4	35.75824	=AVERAGE(C101:C103)	=40-D101						
102	I10 si no RT	CPSI 3-4	35.324688								
103	I10 si no RT	CPSI 3-4	36.59063								
104	I10 si no RT	CPSI 34-35	37.66449	=AVERAGE(C104:C106)	=40-D104						
105	I10 si no RT	CPSI 34-35	40								
106	I10 si no RT	CPSI 34-35	40								
107	I10 si no RT	GFP	30.4953	=AVERAGE(C107:C109)	=40-D107						
108	I10 si no RT	GFP	30.632448								
109	I10 si no RT	GFP	30.632359								
110	I5 M	CPSI 3-4	26.45797			=C110+E119	=ABS(F110-F116)	=(2^-G110)*1000		=H110/I128	=AVERAGE(J110:J112)
111	I5 M	CPSI 3-4	26.537996			=C111+E119	=ABS(F111-F116)	=(2^-G111)*1000		=H111/I128	
112	I5 M	CPSI 3-4	26.478119			=C112+E119	=ABS(F112-F116)	=(2^-G112)*1000		=H112/I128	
113	I5 M	CPSI 34-35	26.886515			=C113+E122	=ABS(F113-F116)	=(2^-G113)*1000		=H113/I131	=AVERAGE(J113:J115)
114	I5 M	CPSI 34-35	26.993113			=C114+E122	=ABS(F114-F116)	=(2^-G114)*1000		=H114/I131	
115	I5 M	CPSI 34-35	26.988743			=C115+E122	=ABS(F115-F116)	=(2^-G115)*1000		=H115/I131	
116	I5 M	GFP	18.16435	=AVERAGE(C116:C118)		=D116+E125					
117	I5 M	GFP	18.225847								
118	I5 M	GFP	18.078655								
119	I5 M no RT	CPSI 3-4	39.533886	=AVERAGE(C119:C121)	=40-D119						
120	I5 M no RT	CPSI 3-4	37.72616								
121	I5 M no RT	CPSI 3-4	38.782516								
122	I5 M no RT	CPSI 34-35	40	=AVERAGE(C122:C124)	=40-D122						
123	I5 M no RT	CPSI 34-35	40								
124	I5 M no RT	CPSI 34-35	40								
125	I5 M no RT	GFP	34.736572	=AVERAGE(C125:C127)	=40-D125						
126	I5 M no RT	GFP	35.390533								
127	I5 M no RT	GFP	40								
128	I5 no si	CPSI 3-4	25.246523			=C128+E137	=ABS(F128-F134)	=(2^-G128)*1000	=AVERAGE(H128:H130)		
129	I5 no si	CPSI 3-4	25.350489			=C129+E137	=ABS(F129-F134)	=(2^-G129)*1000			
130	I5 no si	CPSI 3-4	25.247683			=C130+E137	=ABS(F130-F134)	=(2^-G130)*1000			
131	I5 no si	CPSI 34-35	25.315285			=C131+E140	=ABS(F131-F134)	=(2^-G131)*1000	=AVERAGE(H131:H133)		
132	I5 no si	CPSI 34-35	25.349562			=C132+E140	=ABS(F132-F134)	=(2^-G132)*1000			
133	I5 no si	CPSI 34-35	25.541805			=C133+E140	=ABS(F133-F134)	=(2^-G133)*1000			
134	I5 no si	GFP	18.44052	=AVERAGE(C134:C136)		=D134+E143					
135	I5 no si	GFP	18.519917								
136	I5 no si	GFP	18.417643								
137	I5 no si no RT	CPSI 3-4	40	=AVERAGE(C137:C139)	=40-D137						
138	I5 no si no RT	CPSI 3-4	40								
139	I5 no si no RT	CPSI 3-4	38.319706								
140	I5 no si no RT	CPSI 34-35	37.809692	=AVERAGE(C140:C142)	=40-D140						
141	I5 no si no RT	CPSI 34-35	40								
142	I5 no si no RT	CPSI 34-35	40								
143	I5 no si no RT	GFP	40	=AVERAGE(C143:C145)	=40-D143						
144	I5 no si no RT	GFP	35.795906								
145	I5 no si no RT	GFP	36.737885								
146	I5 si	CPSI 3-4	25.879282			=C146+E155	=ABS(F146-F152)	=(2^-G146)*1000		=H146/I128	=AVERAGE(J146:J148)
147	I5 si	CPSI 3-4	25.782694			=C147+E155	=ABS(F147-F152)	=(2^-G147)*1000		=H147/I128	
148	I5 si	CPSI 3-4	25.728941			=C148+E155	=ABS(F148-F152)	=(2^-G148)*1000		=H148/I128	

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
149	I5 si	CPSI 34-35	26.806684			=C149+E158	=ABS(F149-F152)	=(2^-G149)*1000		=H149/I131	=AVERAGE(J149:J151)
150	I5 si	CPSI 34-35	26.977348			=C150+E158	=ABS(F150-F152)	=(2^-G150)*1000		=H150/I131	
151	I5 si	CPSI 34-35	27.17148			=C151+E158	=ABS(F151-F152)	=(2^-G151)*1000		=H151/I131	
152	I5 si	GFP	17.86009	=AVERAGE(C152:C154)		=D152+E161					
153	I5 si	GFP	17.76037								
154	I5 si	GFP	17.79256								
155	I5 si no RT	CPSI 3-4	40	=AVERAGE(C155:C157)	=40-D155						
156	I5 si no RT	CPSI 3-4	40								
157	I5 si no RT	CPSI 3-4	40								
158	I5 si no RT	CPSI 34-35	40	=AVERAGE(C158:C160)	=40-D158						
159	I5 si no RT	CPSI 34-35	40								
160	I5 si no RT	CPSI 34-35	40								
161	I5 si no RT	GFP	33.63929	=AVERAGE(C161:C163)	=40-D161						
162	I5 si no RT	GFP	33.55373								
163	I5 si no RT	GFP	33.465343								
164	TG M	CPSI 3-4	25.218431			=C164+E173	=ABS(F164-F170)	=(2^-G164)*1000		=H164/I182	=AVERAGE(J164:J166)
165	TG M	CPSI 3-4	25.177479			=C165+E173	=ABS(F165-F170)	=(2^-G165)*1000		=H165/I182	
166	TG M	CPSI 3-4	25.262049			=C166+E173	=ABS(F166-F170)	=(2^-G166)*1000		=H166/I182	
167	TG M	CPSI 34-35	24.858171			=C167+E176	=ABS(F167-F170)	=(2^-G167)*1000		=H167/I185	=AVERAGE(J167:J169)
168	TG M	CPSI 34-35	24.826975			=C168+E176	=ABS(F168-F170)	=(2^-G168)*1000		=H168/I185	
169	TG M	CPSI 34-35	24.88612			=C169+E176	=ABS(F169-F170)	=(2^-G169)*1000		=H169/I185	
170	TG M	GFP	15.026857	=AVERAGE(C170:C172)		=D170+E179					
171	TG M	GFP	14.9558935								
172	TG M	GFP	14.974618								
173	TG M no RT	CPSI 3-4	40	=AVERAGE(C173:C175)	=40-D173						
174	TG M no RT	CPSI 3-4	40								
175	TG M no RT	CPSI 3-4	40								
176	TG M no RT	CPSI 34-35	40	=AVERAGE(C176:C178)	=40-D176						
177	TG M no RT	CPSI 34-35	40								
178	TG M no RT	CPSI 34-35	40								
179	TG M no RT	GFP	33.01587	=AVERAGE(C179:C181)	=40-D179						
180	TG M no RT	GFP	33.29786								
181	TG M no RT	GFP	33.428024								
182	TG no si	CPSI 3-4	23.984722			=C182+E191	=ABS(F182-F188)	=(2^-G182)*1000	=AVERAGE(H182:H184)		
183	TG no si	CPSI 3-4	24.210308			=C183+E191	=ABS(F183-F188)	=(2^-G183)*1000			
184	TG no si	CPSI 3-4	24.076149			=C184+E191	=ABS(F184-F188)	=(2^-G184)*1000			
185	TG no si	CPSI 34-35	24.288809			=C185+E194	=ABS(F185-F188)	=(2^-G185)*1000	=AVERAGE(H185:H187)		
186	TG no si	CPSI 34-35	24.430285			=C186+E194	=ABS(F186-F188)	=(2^-G186)*1000			
187	TG no si	CPSI 34-35	24.258919			=C187+E194	=ABS(F187-F188)	=(2^-G187)*1000			
188	TG no si	GFP	15.665624	=AVERAGE(C188:C190)		=D188+E197					
189	TG no si	GFP	15.70701								
190	TG no si	GFP	15.598993								
191	TG no si no RT	CPSI 3-4	40	=AVERAGE(C191:C193)	=40-D191						
192	TG no si no RT	CPSI 3-4	40								
193	TG no si no RT	CPSI 3-4	40								
194	TG no si no RT	CPSI 34-35	40	=AVERAGE(C194:C196)	=40-D194						
195	TG no si no RT	CPSI 34-35	40								
196	TG no si no RT	CPSI 34-35	40								
197	TG no si no RT	GFP	40	=AVERAGE(C197:C199)	=40-D197						

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
198	TG no si no RT	GFP	40								
199	TG no si no RT	GFP	40								
200	TG si	CPSI 3-4	26.04842			=C200+E209	=ABS(F200-F206)	=(2^-G200)*1000		=H200/I182	=AVERAGE(J200:J202)
201	TG si	CPSI 3-4	26.168285			=C201+E209	=ABS(F201-F206)	=(2^-G201)*1000		=H201/I182	
202	TG si	CPSI 3-4	25.962063			=C202+E209	=ABS(F202-F206)	=(2^-G202)*1000		=H202/I182	
203	TG si	CPSI 34-35	25.77958			=C203+E212	=ABS(F203-F206)	=(2^-G203)*1000		=H203/I185	=AVERAGE(J203:J205)
204	TG si	CPSI 34-35	26.163162			=C204+E212	=ABS(F204-F206)	=(2^-G204)*1000		=H204/I185	
205	TG si	CPSI 34-35	26.051224			=C205+E212	=ABS(F205-F206)	=(2^-G205)*1000		=H205/I185	
206	TG si	GFP	14.959359	=AVERAGE(C206:C208)		=D206+E215					
207	TG si	GFP	14.955553								
208	TG si	GFP	15.085495								
209	TG si no RT	CPSI 3-4	40	=AVERAGE(C209:C211)	=40-D209						
210	TG si no RT	CPSI 3-4	40								
211	TG si no RT	CPSI 3-4	40								
212	TG si no RT	CPSI 34-35	40	=AVERAGE(C212:C214)	=40-D212						
213	TG si no RT	CPSI 34-35	40								
214	TG si no RT	CPSI 34-35	40								
215	TG si no RT	GFP	29.826496	=AVERAGE(C215:C217)	=40-D215						
216	TG si no RT	GFP	29.842789								
217	TG si no RT	GFP	29.877337								
218	Wt M	CPSI 3-4	23.603674			=C218+E227	=ABS(F218-F224)	=(2^-G218)*1000		=H218/I236	=AVERAGE(J218:J220)
219	Wt M	CPSI 3-4	23.770157			=C219+E227	=ABS(F219-F224)	=(2^-G219)*1000		=H219/I236	
220	Wt M	CPSI 3-4	23.663494			=C220+E227	=ABS(F220-F224)	=(2^-G220)*1000		=H220/I236	
221	Wt M	CPSI 34-35	22.164198			=C221+E230	=ABS(F221-F224)	=(2^-G221)*1000		=H221/I239	=AVERAGE(J221:J223)
222	Wt M	CPSI 34-35	22.121677			=C222+E230	=ABS(F222-F224)	=(2^-G222)*1000		=H222/I239	
223	Wt M	CPSI 34-35	22.508509			=C223+E230	=ABS(F223-F224)	=(2^-G223)*1000		=H223/I239	
224	Wt M	GFP	16.885818	=AVERAGE(C224:C226)		=D224+E233					
225	Wt M	GFP	16.717237								
226	Wt M	GFP	16.81394								
227	Wt M no RT	CPSI 3-4	38.190742	=AVERAGE(C227:C229)	=40-D227						
228	Wt M no RT	CPSI 3-4	40								
229	Wt M no RT	CPSI 3-4	40								
230	Wt M no RT	CPSI 34-35	38.01468	=AVERAGE(C230:C232)	=40-D230						
231	Wt M no RT	CPSI 34-35	40								
232	Wt M no RT	CPSI 34-35	40								
233	Wt M no RT	GFP	32.417988	=AVERAGE(C233:C235)	=40-D233						
234	Wt M no RT	GFP	34.0372								
235	Wt M no RT	GFP	33.39038								
236	Wt no si	CPSI 3-4	22.799446			=C236+E245	=ABS(F236-F242)	=(2^-G236)*1000	=AVERAGE(H236:H238)		
237	Wt no si	CPSI 3-4	22.962536			=C237+E245	=ABS(F237-F242)	=(2^-G237)*1000			
238	Wt no si	CPSI 3-4	22.905407			=C238+E245	=ABS(F238-F242)	=(2^-G238)*1000			
239	Wt no si	CPSI 34-35	20.274261			=C239+E248	=ABS(F239-F242)	=(2^-G239)*1000	=AVERAGE(H239:H241)		
240	Wt no si	CPSI 34-35	20.534668			=C240+E248	=ABS(F240-F242)	=(2^-G240)*1000			
241	Wt no si	CPSI 34-35	20.05323			=C241+E248	=ABS(F241-F242)	=(2^-G241)*1000			
242	Wt no si	GFP	16.879885	=AVERAGE(C242:C244)		=D242+E251					
243	Wt no si	GFP	16.936453								
244	Wt no si	GFP	16.879858								
245	Wt no si no RT	CPSI 3-4	40	=AVERAGE(C245:C247)	=40-D245						
246	Wt no si no RT	CPSI 3-4	40								

	A	B	C	D	E	F	G	H	I	J	K
1	**Sample**	**Detector**	**Ct**	**Average Ct**	**No RT CF**	**CT + CF**	**Delta Ct**	**Linear Conversion**	**Calibrator Avg**	**ddCt**	**ddCt Avg**
247	Wt no si no RT	CPSI 3-4	40								
248	Wt no si no RT	CPSI 34-35	40	=AVERAGE(C248:C250)	=40-D248						
249	Wt no si no RT	CPSI 34-35	40								
250	Wt no si no RT	CPSI 34-35	40								
251	Wt no si no RT	GFP	35.852497	=AVERAGE(C251:C253)	=40-D251						
252	Wt no si no RT	GFP	36.90489								
253	Wt no si no RT	GFP	34.837555								
254	Wt si	CPSI 3-4	22.913435			=C254+E263	=ABS(F254-F260)	=(2^-G254)*1000		=H254/I236	=AVERAGE(J254:J256)
255	Wt si	CPSI 3-4	23.054405			=C255+E263	=ABS(F255-F260)	=(2^-G255)*1000		=H255/I236	
256	Wt si	CPSI 3-4	22.887999			=C256+E263	=ABS(F256-F260)	=(2^-G256)*1000		=H256/I236	
257	Wt si	CPSI 34-35	21.727528			=C257+E266	=ABS(F257-F260)	=(2^-G257)*1000		=H257/I239	=AVERAGE(J257:J259)
258	Wt si	CPSI 34-35	21.693125			=C258+E266	=ABS(F258-F260)	=(2^-G258)*1000		=H258/I239	
259	Wt si	CPSI 34-35	21.70578			=C259+E266	=ABS(F259-F260)	=(2^-G259)*1000		=H259/I239	
260	Wt si	GFP	16.317648	=AVERAGE(C260:C262)		=D260+E269					
261	Wt si	GFP	16.32774								
262	Wt si	GFP	16.340092								
263	Wt si no RT	CPSI 3-4	38.846848	=AVERAGE(C263:C265)	=40-D263						
264	Wt si no RT	CPSI 3-4	40								
265	Wt si no RT	CPSI 3-4	40								
266	Wt si no RT	CPSI 34-35	40	=AVERAGE(C266:C268)	=40-D266						
267	Wt si no RT	CPSI 34-35	40								
268	Wt si no RT	CPSI 34-35	40								
269	Wt si no RT	GFP	33.46902	=AVERAGE(C269:C271)	=40-D269						
270	Wt si no RT	GFP	33.28962								
271	Wt si no RT	GFP	32.893								

REFERENCES

1. Scaglia, F., Brunetti-Pierri, N., Kleppe, S., Marini, J., Carter, S., Garlick, P., Jahoor, F., O'Brien, W., and Lee, B. **2004** "Clinical consequences of urea cycle enzyme deficiencies and potential links to arginine and nitric oxide metabolism." *J. Nutr.*, 134: 2775S-2782S.

2. Morizono, H., Caldovic, L., Shi, D., and Tuchman, M. **2004** "Mammalian N-acetylglutamate synthase." *Mol. Genet. Metab.*, 81 Suppl 1: S4-11.

3. Camacho, J. A., Obie, C., Biery, B., Goodman, B. K., Hu, C. A., Almashanu, S., Steel, G., Casey, R., Lambert, M., Mitchell, G. A., and Valle, D. **1999** "Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter." *Nat. Genet.*, 22: 151-158.

4. Kobayashi, K., Sinasac, D. S., Iijima, M., Boright, A. P., Begum, L., Lee, J. R., Yasuda, T., Ikeda, S., Hirano, R., Terazono, H., Crackower, M. A., Kondo, I., Tsui, L. C., Scherer, S. W., and Saheki, T. **1999** "The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein." *Nat. Genet.*, 22: 159-163.

5. Ignarro, L. J. **2002** "Nitric oxide as a unique signaling molecule in the vascular system: a historical overview." *J. Physiol. Pharmacol.*, 53: 503-514.

6. Pearson, D. L., Dawling, S., Walsh, W. F., Haines, J. L., Christman, B. W., Bazyk, A., Scott, N., and Summar, M. L. **2001** "Neonatal pulmonary hypertension--urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function." *N. Engl. J. Med.*, 344: 1832-1838.

7. Morris, S. M., Jr. **2006** "Arginine: beyond protein." *Am. J. Clin. Nutr.*, 83: 508S-512S.

8. Nyunoya, H., Broglie, K. E., Widgren, E. E., and Lusty, C. J. **1985** "Characterization and derivation of the gene coding for mitochondrial carbamyl phosphate synthetase I of rat." *J. Biol. Chem.*, 260: 9346-9356.

9. Rubio, V. **1993** "Structure-function studies in carbamoyl phosphate synthetases." *Biochem. Soc. Trans.*, 21: 198-202.

10. Rubio, V., Britton, H. G., Grisolia, S., Sproat, B. S., and Lowe, G. **1981** "Mechanism of activation of bicarbonate ion by mitochondrial

carbamoyl-phosphate synthetase: formation of enzyme-bound adenosine diphosphate from the adenosine triphosphate that yields inorganic phosphate." *Biochemistry*, 20: 1969-1974.

11. Jackson, M. J., Beaudet, A. L., and O'Brien, W. E. **1986** "Mammalian urea cycle enzymes." *Annu. Rev. Genet.*, 20: 431-464.

12. Summar, M. L., Hall, L. D., Eeds, A. M., Hutcheson, H. B., Kuo, A. N., Willis, A. S., Rubio, V., Arvin, M. K., Schofield, J. P., and Dawson, E. P. **2003** "Characterization of genomic structure and polymorphisms in the human carbamyl phosphate synthetase I gene." *Gene*, 311: 51-57.

13. Nyunoya, H., Broglie, K. E., and Lusty, C. J. **1985** "The gene coding for carbamoyl-phosphate synthetase I was formed by fusion of an ancestral glutaminase gene and a synthetase gene." *Proc. Natl. Acad. Sci.*, 82: 2244-2246.

14. Schofield, J. P. **1993** "Molecular studies on an ancient gene encoding for carbamoyl-phosphate synthetase." *Clin. Sci.*, 84: 119-128.

15. Potter, M. D. and Powers-Lee, S. G. **1992** "Location of the ATP gamma-phosphate-binding sites on rat liver carbamoyl-phosphate synthetase I. Studies with the ATP analog 5'-p-fluorosulfonylbenzoyladenosine." *J. Biol. Chem.*, 267: 2023-2031.

16. Devaney, M. A. and Powers-Lee, S. G. **1984** "Immunological cross-reactivity between carbamyl phosphate synthetases I, II, and III." *J Biol.Chem.*, 259: 703-706.

17. Anderson, P. M. **1980** "Glutamine- and N-acetylglutamate-dependent carbamoyl phosphate synthetase in elasmobranchs." *Science*, 208: 291-293.

18. Summar, M. L., Dasouki, M. J., Schofield, P. J., Krishnamani, M. R., Vnencak-Jones, C., Tuchman, M., Mao, J., and Phillips, J. A., III. **1995** "Physical and linkage mapping of human carbamyl phosphate synthetase I (CPS1) and reassignment from 2p to 2q35." *Cytogenet. Cell Genet.*, 71: 266-267.

19. Moorman, A. F., de Boer, P. A., Das, A. T., Labruyere, W. T., Charles, R., and Lamers, W. H. **1990** "Expression patterns of mRNAs for ammonia-metabolizing enzymes in the developing rat: the ontogenesis of hepatocyte heterogeneity." *Histochem. J.*, 22: 457-468.

20. Summar, M. L., Gainer, J. V., Pretorius, M., Malave, H., Harris, S., Hall, L. D., Weisberg, A., Vaughan, D. E., Christman, B. W., and Brown, N.

J. **2004** "Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function." *Hypertension*, 43: 186-191.

21. Barr, F. E., Beverley, H., VanHook, K., Cermak, E., Christian, K., Drinkwater, D., Dyer, K., Raggio, N. T., Moore, J. H., Christman, B., and Summar, M. **2003** "Effect of cardiopulmonary bypass on urea cycle intermediates and nitric oxide levels after congenital heart surgery." *J. Pediatr.*, 142: 26-30.

22. Summar, M. L., Hall, L., Christman, B., Barr, F., Smith, H., Kallianpur, A., Brown, N., Yadav, M., Willis, A., Eeds, A., Cermak, E., Summar, S., Wilson, A., Arvin, M., Putnam, A., Wills, M., and Cunningham, G. **2004** "Environmentally determined genetic expression: clinical correlates with molecular variants of carbamyl phosphate synthetase I." *Mol. Genet. Metab.*, 81 Suppl 1: S12-S19.

23. Hoshide, R., Matsuura, T., Haraguchi, Y., Endo, F., Yoshinaga, M., and Matsuda, I. **1993** "Carbamyl phosphate synthetase I deficiency. One base substitution in an exon of the CPS I gene causes a 9-basepair deletion due to aberrant splicing." *J. Clin. Invest.*, 91: 1884-1887.

24. Finckh, U., Kohlschutter, A., Schafer, H., Sperhake, K., Colombo, J. P., and Gal, A. **1998** "Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency by identification of a missense mutation in CPS1." *Hum. Mutat.*, 12: 206-211.

25. Summar, M. L. **1998** "Molecular genetic research into carbamoyl-phosphate synthase I: molecular defects and linkage markers." *J. Inherit. Metab. Dis.*, 21 Suppl 1: 30-39.

26. Aoshima, T., Kajita, M., Sekido, Y., Kikuchi, S., Yasuda, I., Saheki, T., Watanabe, K., Shimokata, K., and Niwa, T. **2001** "Novel mutations (H337R and 238-362del) in the CPS1 gene cause carbamoyl phosphate synthetase I deficiency." *Hum. Hered.*, 52: 99-101.

27. Rapp, B., Haberle, J., Linnebank, M., Wermuth, B., Marquardt, T., Harms, E., and Koch, H. G. **2001** "Genetic analysis of carbamoylphosphate synthetase I and ornithine transcarbamylase deficiency using fibroblasts." *Eur. J. Pediatr.*, 160: 283-287.

28. Funghini, S., Donati, M. A., Pasquini, E., Zammarchi, E., and Morrone, A. **2003** "Structural organization of the human carbamyl phosphate synthetase I gene (CPS1) and identification of two novel genetic lesions." *Hum. Mutat.*, 22: 340-341.

29. Haberle, J., Schmidt, E., Pauli, S., Rapp, B., Christensen, E., Wermuth, B., and Koch, H. G. **2003** "Gene structure of human carbamylphosphate synthetase 1 and novel mutations in patients with neonatal onset." *Hum. Mutat.*, 21: 444.

30. Nagata, N., Matsuda, I., and Oyanagi, K. **1991** "Estimated frequency of urea cycle enzymopathies in Japan." *Am. J. Med. Genet.*, 39: 228-229.

31. Tuchman, M. and Yudkoff, M. **1999** "Blood levels of ammonia and nitrogen scavenging amino acids in patients with inherited hyperammonemia." *Mol. Genet. Metab.*, 66: 10-15.

32. Takeoka, M., Soman, T. B., Shih, V. E., Caviness, V. S., Jr., and Krishnamoorthy, K. S. **2001** "Carbamyl phosphate synthetase 1 deficiency: a destructive encephalopathy." *Pediatr. Neurol.*, 24: 193-199.

33. Lo, W. D., Sloan, H. R., Sotos, J. F., and Klinger, R. J. **1993** "Late clinical presentation of partial carbamyl phosphate synthetase I deficiency." *Am. J. Dis. Child*, 147: 267-269.

34. Wong, L. J., Craigen, W. J., and O'Brien, W. E. **1994** "Postpartum coma and death due to carbamoyl-phosphate synthetase I deficiency." *Ann. Intern. Med.*, 120: 216-217.

35. Hoogenraad, N. J., Mitchell, J. D., Don, N. A., Sutherland, T. M., and Mc Leay, A. C. **1980** "Detection of carbamyl phosphate synthetase 1 deficiency using duodenal biopsy samples." *Arch. Dis. Child*, 55: 292-295.

36. Summar, M. L., Barr, F., Dawling, S., Smith, W., Lee, B., Singh, R. H., Rhead, W. J., Sniderman, King L., and Christman, B. W. **2005** "Unmasked adult-onset urea cycle disorders in the critical care setting." *Crit. Care Clin.*, 21: S1-S8.

37. Luo, M. J. and Reed, R. **1999** "Splicing is required for rapid and efficient mRNA export in metazoans." *Proc. Natl. Acad. Sci.*, 96: 14937-14942.

38. Zhao, J., Hyman, L., and Moore, C. **1999** "Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis." *Microbiol. Mol. Biol. Rev.*, 63: 405-445.

39. Gudikote, J. P., Imam, J. S., Garcia, R. F., and Wilkinson, M. F. **2005** "RNA splicing promotes translation and RNA surveillance." *Nat. Struct. Mol. Biol.*, 12: 801-809.

40. Buratowski, S. **2005** "Connections between mRNA 3' end processing and transcription termination." *Curr. Opin. Cell Biol.*, 17: 257-261.

41. Burckin, T., Nagel, R., Mandel-Gutfreund, Y., Shiue, L., Clark, T. A., Chong, J. L., Chang, T. H., Squazzo, S., Hartzog, G., and Ares, M., Jr. **2005** "Exploring functional relationships between components of the gene expression machinery." *Nat. Struct. Mol. Biol.*, 12: 175-182.

42. Shatkin, A. J. **1976** "Capping of eucaryotic mRNAs." *Cell*, 9: 645-653.

43. Gu, M. and Lima, C. D. **2005** "Processing the message: structural insights into capping and decapping mRNA." *Curr. Opin. Struct. Biol.*, 15: 99-106.

44. Edmonds, M. **2002** "A history of poly A sequences: from formation to factors to function." *Prog. Nucleic Acid Res. Mol. Biol.*, 71: 285-389.

45. Wahle, E. and Ruegsegger, U. **1999** "3'-End processing of pre-mRNA in eukaryotes." *FEMS Microbiol. Rev.*, 23: 277-295.

46. Mangus, D. A., Evans, M. C., and Jacobson, A. **2003** "Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression." *Genome Biol.*, 4: 223.

47. Wells, S. E., Hillner, P. E., Vale, R. D., and Sachs, A. B. **1998** "Circularization of mRNA by eukaryotic translation initiation factors." *Mol. Cell*, 2: 135-140.

48. Smith, C. W., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. **1989** "Scanning from an independently specified branch point defines the 3' splice site of mammalian introns." *Nature*, 342: 243-247.

49. Gooding, C., Clark, F., Wollerton, M. C., Grellscheid, S. N., Groom, H., and Smith, C. W. **2006** "A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones." *Genome Biol.*, 7: R1.

50. Zhang, M. Q. **1998** "Statistical features of human exons and their flanking regions." *Hum. Mol. Genet.*, 7: 919-932.

51. Patel, A. A. and Steitz, J. A. **2003** "Splicing double: insights from the second spliceosome." *Nat. Rev. Mol. Cell Biol.*, 4: 960-970.

52. Hastings, M. L., Resta, N., Traum, D., Stella, A., Guanti, G., and Krainer, A. R. **2005** "An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at noncanonical cryptic splice sites." *Nat. Struct. Mol. Biol.*, 12: 54-59.

53. Blencowe, B. J. **2000** "Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases." *Trends Biochem. Sci.*, 25: 106-110.

54. Nissim-Rafinia, M. and Kerem, B. **2002** "Splicing regulation as a potential genetic modifier." *Trends Genet.*, 18: 123-127.

55. Ibrahim, el C., Schaal, T. D., Hertel, K. J., Reed, R., and Maniatis, T. **2005** "Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers." *Proc. Natl. Acad. Sci.*, 102: 5002-5007.

56. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. **2003** "ESEfinder: A web resource to identify exonic splicing enhancers." *Nucleic Acids Res.*, 31: 3568-3571.

57. Graveley, B. R. **2000** "Sorting out the complexity of SR protein functions." *RNA.*, 6: 1197-1211.

58. Hastings, M. L. and Krainer, A. R. **2001** "Pre-mRNA splicing in the new millennium." *Curr. Opin. Cell Biol.*, 13: 302-309.

59. Pagani, F., Raponi, M., and Baralle, F. E. **2005** "Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution." *Proc. Natl. Acad. Sci.*, 102: 6368-6372.

60. Reed, R. **2000** "Mechanisms of fidelity in pre-mRNA splicing." *Curr. Opin. Cell Biol.*, 12: 340-345.

61. Will, C. L. and Luhrmann, R. **2001** "Spliceosomal UsnRNP biogenesis, structure and function." *Curr. Opin. Cell Biol.*, 13: 290-301.

62. Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. **2005** "Are splicing mutations the most frequent cause of hereditary disease?" *FEBS Lett.*, 579: 1900-1903.

63. Gorlov, I. P., Gorlova, O. Y., Frazier, M. L., and Amos, C. I. **2003** "Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers." *Am. J. Hum. Genet.*, 73: 1157-1161.

64. Wang, J., Smith, P. J., Krainer, A. R., and Zhang, M. Q. **2005** "Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes." *Nucleic Acids Res.*, 33: 5053-5062.

65. Nagy, E. and Maquat, L. E. **1998** "A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance." *Trends Biochem. Sci.*, 23: 198-199.

66. Frischmeyer, P. A. and Dietz, H. C. **1999** "Nonsense-mediated mRNA decay in health and disease." *Hum. Mol. Genet.*, 8: 1893-1900.

67. Mendell, J. T. and Dietz, H. C. **2001** "When the message goes awry: disease-producing mutations that influence mRNA content and performance." *Cell*, 107: 411-414.

68. Holbrook, J. A., Neu-Yilik, G., Hentze, M. W., and Kulozik, A. E. **2004** "Nonsense-mediated decay approaches the clinic." *Nat. Genet.*, 36: 801-808.

69. Wittmann, J., Hol, E. M., and Jack, H. M. **2006** "hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay." *Mol. Cell Biol.*, 26: 1272-1287.

70. Lewis, B. P., Green, R. E., and Brenner, S. E. **2003** "Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans." *Proc. Natl. Acad. Sci.*, 100: 189-192.

71. Li, S. and Wilkinson, M. F. **1998** "Nonsense surveillance in lymphocytes?" *Immunity*, 8: 135-141.

72. Wilkinson, M. F. **2005** "A new function for nonsense-mediated mRNA-decay factors." *Trends Genet.*, 21: 143-148.

73. Weischenfeldt, J., Lykke-Andersen, J., and Porse, B. **2005** "Messenger RNA surveillance: neutralizing natural nonsense." *Curr. Biol.*, 15: R559-R562.

74. Mitrovich, Q. M. and Anderson, P. **2005** "mRNA surveillance of expressed pseudogenes in C. elegans." *Curr.Biol.*, 15: 963-967.

75. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F., and Dietz, H. C. **2004** "Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise." *Nat. Genet.*, 36: 1073-1078.

76. Lejeune, F. and Maquat, L. E. **2005** "Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells." *Curr. Opin. Cell Biol.*, 17: 309-315.

77. Belgrader, P., Cheng, J., and Maquat, L. E. **1993** "Evidence to implicate translation by ribosomes in the mechanism by which nonsense codons reduce the nuclear level of human triosephosphate isomerase mRNA." *Proc. Natl. Acad. Sci.*, 90: 482-486.

78. Carter, M. S., Doskow, J., Morris, P., Li, S., Nhim, R. P., Sandstedt, S., and Wilkinson, M. F. **1995** "A regulatory mechanism that detects

premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro." *J. Biol. Chem.*, 270: 28995-29003.

79. Caputi, M., Kendzior, R. J., and Beemon, K. L. **2002** "A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer." *Genes Dev.*, 16: 1754-1759.

80. Zhang, J., Sun, X., Qian, Y., LaDuca, J. P., and Maquat, L. E. **1998** "At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation." *Mol. Cell Biol.*, 18: 5272-5283.

81. Maquat, L. E. and Li, X. **2001** "Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay." *RNA*, 7: 445-456.

82. Brocke, K. S., Neu-Yilik, G., Gehring, N. H., Hentze, M. W., and Kulozik, A. E. **2002** "The human intronless melanocortin 4-receptor gene is NMD insensitive." *Hum. Mol. Genet.*, 11: 331-335.

83. Le, Hir H., Gatfield, D., Izaurralde, E., and Moore, M. J. **2001** "The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay." *EMBO J.*, 20: 4987-4997.

84. Schell, T., Kulozik, A. E., and Hentze, M. W. **2002** "Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway." *Genome Biol.*, 3: REVIEWS1006.

85. Wiegand, H. L., Lu, S., and Cullen, B. R. **2003** "Exon junction complexes mediate the enhancing effect of splicing on mRNA expression." *Proc. Natl. Acad. Sci.*, 100: 11327-11332.

86. Le, Hir H., Izaurralde, E., Maquat, L. E., and Moore, M. J. **2000** "The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions." *EMBO J.*, 19: 6860-6869.

87. Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le, Hir H. **2005** "The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity." *Nat. Struct. Mol. Biol.*, 12: 861-869.

88. Chan, C. C., Dostie, J., Diem, M. D., Feng, W., Mann, M., Rappsilber, J., and Dreyfuss, G. **2004** "eIF4A3 is a novel component of the exon junction complex." *RNA*, 10: 200-209.

89. Ferraiuolo, M. A., Lee, C. S., Ler, L. W., Hsu, J. L., Costa-Mattioli, M., Luo, M. J., Reed, R., and Sonenberg, N. **2004** "A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay." *Proc. Natl. Acad. Sci.*, 101: 4118-4123.

90. Shibuya, T., Tange, T. O., Sonenberg, N., and Moore, M. J. **2004** "eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay." *Nat. Struct. Mol. Biol.*, 11: 346-351.

91. Degot, S., Le, Hir H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M. C., and Tomasetto, C. **2004** "Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module." *J. Biol. Chem.*, 279: 33702-33715.

92. Dostie, J. and Dreyfuss, G. **2002** "Translation is required to remove Y14 from mRNAs in the cytoplasm." *Curr. Biol.*, 12: 1060-1067.

93. Kim, V. N., Kataoka, N., and Dreyfuss, G. **2001** "Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex." *Science*, 293: 1832-1836.

94. Singh, G. and Lykke-Andersen, J. **2003** "New insights into the formation of active nonsense-mediated decay complexes." *Trends Biochem. Sci*, 28: 464-466.

95. Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. **2006** "Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay." *Genes Dev.*, 20: 355-367.

96. Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. **2000** "Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon." *Cell*, 103: 1121-1131.

97. McGarvey, T., Rosonina, E., McCracken, S., Li, Q., Arnaout, R., Mientjes, E., Nickerson, J. A., Awrey, D., Greenblatt, J., Grosveld, G., and Blencowe, B. J. **2000** "The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes." *J. Cell Biol.*, 150: 309-320.

98. Maquat, L. E. **2004** "Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics." *Nat. Rev. Mol. Cell Biol.*, 5: 89-99.

99. Sakashita, E., Tatsumi, S., Werner, D., Endo, H., and Mayeda, A. **2004** "Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo." *Mol. Cell Biol.*, 24: 1174-1187.

100. Kataoka, N., Diem, M. D., Kim, V. N., Yong, J., and Dreyfuss, G. **2001** "Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex." *EMBO J.*, 20: 6424-6433.

101. Kunz, J. B., Neu-Yilik, G., Hentze, M. W., Kulozik, A. E., and Gehring, N. H. **2006** "Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation." *RNA*, In press.

102. Serin, G., Gersappe, A., Black, J. D., Aronoff, R., and Maquat, L. E. **2001** "Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4)." *Mol. Cell Biol.*, 21: 209-223.

103. Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K. R., Grimson, A., Hachiya, T., Hentze, M. W., Anderson, P., and Ohno, S. **2003** "Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7." *Mol. Cell*, 12: 1187-1200.

104. Gehring, N. H., Neu-Yilik, G., Schell, T., Hentze, M. W., and Kulozik, A. E. **2003** "Y14 and hUpf3b form an NMD-activating complex." *Mol. Cell*, 11: 939-949.

105. Kadlec, J., Izaurralde, E., and Cusack, S. **2004** "The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3." *Nat. Struct. Mol. Biol.*, 11: 330-337.

106. Mendell, J. T., Medghalchi, S. M., Lake, R. G., Noensie, E. N., and Dietz, H. C. **2000** "Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes." *Mol. Cell Biol.*, 20: 8944-8957.

107. Wang, W., Cajigas, I. J., Peltz, S. W., Wilkinson, M. F., and Gonzalez, C. I. **2006** "Role for Upf2p Phosphorylation in Saccharomyces cerevisiae Nonsense-Mediated mRNA Decay." *Mol. Cell Biol.*, 26: 3390-3400.

108. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. **2001** "Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay." *Genes Dev.*, 15: 2215-2228.

109. Denning, G., Jamieson, L., Maquat, L. E., Thompson, E. A., and Fields, A. P. **2001** "Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein." *J. Biol. Chem.*, 276: 22709-22714.

110. Chiu, S. Y., Serin, G., Ohara, O., and Maquat, L. E. **2003** "Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1." *RNA*, 9: 77-87.

111. Unterholzner, L. and Izaurralde, E. **2004** "SMG7 acts as a molecular link between mRNA surveillance and mRNA decay." *Mol. Cell*, 16: 587-596.

112. Zhang, J., Sun, X., Qian, Y., and Maquat, L. E. **1998** "Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm." *RNA*, 4: 801-815.

113. Ishigaki, Y., Li, X., Serin, G., and Maquat, L. E. **2001** "Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20." *Cell*, 106: 607-617.

114. Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L. E. **2002** "The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling." *EMBO J.*, 21: 3536-3545.

115. Chiu, S. Y., Lejeune, F., Ranganathan, A. C., and Maquat, L. E. **2004** "The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex." *Genes Dev.*, 18: 745-754.

116. Chen, C. Y. and Shyu, A. B. **2003** "Rapid deadenylation triggered by a nonsense codon precedes decay of the RNA body in a mammalian cytoplasmic nonsense-mediated decay pathway." *Mol. Cell Biol.*, 23: 4805-4813.

117. Lejeune, F., Li, X., and Maquat, L. E. **2003** "Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities." *Mol. Cell*, 12: 675-687.

118. Lykke-Andersen, J. **2002** "Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay." *Mol. Cell Biol.*, 22: 8114-8121.

119. Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S., Wilson, J., Reggin, J. D., Mancias, P., Butler, I. J., Wilkinson, M. F., Wegner, M., and Lupski, J. R. **2004** "Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations." *Nat. Genet.*, 36: 361-369.

120. Hall, G. W. and Thein, S. **1994** "Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia." *Blood*, 83: 2031-2037.

121. Scriver, C. R. and Waters, P. J. **1999** "Monogenic traits are not simple: lessons from phenylketonuria." *Trends Genet.*, 15: 267-272.

122. Dipple, K. M. and McCabe, E. R. **2000** "Modifier genes convert "simple" Mendelian disorders to complex traits." *Mol. Genet. Metab*, 71: 43-50.

123. Cartegni, L., Chew, S. L., and Krainer, A. R. **2002** "Listening to silence and understanding nonsense: exonic mutations that affect splicing." *Nat. Rev. Genet.*, 3: 285-298.

124. Zhang, W., Holzknecht, R. A., Butkowski, R. J., and Tuchman, M. **1990** "Immunochemical analysis of carbamyl phosphate synthetase I and ornithine transcarbamylase deficient livers: elevated N-acetylglutamate level in a liver lacking carbamyl phosphate synthetase protein." *Clin. Invest. Med.*, 13: 183-188.

125. Haberle, J., Denecke, J., Schmidt, E., and Koch, H. G. **2003** "Diagnosis of N-acetylglutamate synthase deficiency by use of cultured fibroblasts and avoidance of nonsense-mediated mRNA decay." *J. Inherit. Metab Dis.*, 26: 601-605.

126. Bateman, J. F., Freddi, S., Nattrass, G., and Savarirayan, R. **2003** "Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage." *Hum. Mol. Genet.*, 12: 217-225.

127. Eeds, A., Hall, L. D., Yadav, M., Willis, A. S., Summar, S., Putnam, A., Barr, F., and Summar, M. L. **2006** "The frequent obsrevation of evidence for nonsense-mediated decay in RNA from patients with carbamyl phosphate synthetase I deficiency." *Mol. Genet. Metab.*, In press.

128. Wade-Martins, R., Smith, E. R., Tyminski, E., Chiocca, E. A., and Saeki, Y. **2001** "An infectious transfer and expression system for genomic DNA loci in human and mouse cells." *Nat. Biotechnol.*, 19: 1067-1070.

129. Mortlock, D. P., Guenther, C., and Kingsley, D. M. **2003** "A general approach for identifying distant regulatory elements applied to the Gdf6 gene." *Genome Res.*, 13: 2069-2081.

130. Lee, E. C., Yu, D., Martinez, de, V, Tessarollo, L., Swing, D. A., Court DL, Jenkins, N. A., and Copeland, N. G. **2001** "A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA." *Genomics*, 73: 56-65.

131. Huschtscha, L. I. and Holliday, R. **1983** "Limited and unlimited growth of SV40-transformed cells from human diploid MRC-5 fibroblasts." *J. Cell Sci.*, 63: 77-99.

132. Hart, S. L., rancibia-Carcamo, C. V., Wolfert, M. A., Mailhos, C., O'Reilly, N. J., Ali, R. R., Coutelle, C., George, A. J., Harbottle, R. P., Knight, A. M., Larkin, D. F., Levinsky, R. J., Seymour, L. W., Thrasher, A. J., and Kinnon, C. **1998** "Lipid-mediated enhancement of transfection by a nonviral integrin-targeting vector." *Hum. Gene Ther.*, 9: 575-585.

133. Wade-Martins, R., Frampton, J., and James, M. R. **1999** "Long-term stability of large insert genomic DNA episomal shuttle vectors in human cells." *Nucleic Acids Res.*, 27: 1674-1682.

134. Le, Hir H., Nott, A., and Moore, M. J. **2003** "How introns influence and enhance eukaryotic gene expression." *Trends Biochem. Sci*, 28: 215-220.

135. Leight, E. R. and Sugden, B. **2000** "EBNA-1: a protein pivotal to latent infection by Epstein-Barr virus." *Rev. Med. Virol.*, 10: 83-100.

136. Yates, J. L., Warren, N., and Sugden, B. **1985** "Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells." *Nature*, 313: 812-815.

137. Belt, P. B., Groeneveld, H., Teubel, W. J., van de, Putte P., and Backendorf, C. **1989** "Construction and properties of an Epstein-Barr-virus-derived cDNA expression vector for human cells." *Gene*, 84: 407-417.

138. Lei, D. C., Kunzelmann, K., Koslowsky, T., Yezzi, M. J., Escobar, L. C., Xu, Z., Ellison, A. R., Rommens, J. M., Tsui, L.-C., Tykocinski, M., and Gruenert, D. C. **1996** "Episomal expression of wild-type CFTR corrects cAMP-dependent chloride transport in respiratory epithelial cells." *Gene Ther.*, 3: 427-436.

139. Wade-Martins, R., Saeki, Y., and Chiocca, E. A. **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.

140. Inoue, R., Moghaddam, K. A., Ranasinghe, M., Saeki, Y., Chiocca, E. A., and 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.

141. Laner, A., Goussard, S., Ramalho, A. S., Schwarz, T., Amaral, M. D., Courvalin, P., Schindelhauer, D., and Grillot-Courvalin, C. **2005** "Bacterial transfer of large functional genomic DNA into human cells." *Gene Ther.*, 12: 1559-1572.

142. Simpson, K. and Huxley, C. **1996** "A shuttle system for transfer of YACs between yeast and mammalian cells." *Nucleic Acids Res.*, 24: 4693-4699.

143. Kotzamanis, G., Cheung, W., Abdulrazzak, H., Perez-Luz, S., Howe, S., Cooke, H., and Huxley, C. **2005** "Construction of human artificial chromosome vectors by recombineering." *Gene*, 351: 29-38.

144. Zhang, X. M. and Huang, J. D. **2003** "Combination of overlapping bacterial artificial chromosomes by a two-step recombinogenic engineering method." *Nucleic Acids Res.*, 31: e81.

145. White, R. E., Wade-Martins, R., Hart, S. L., Frampton, J., Huey, B., sai-Mehta, A., Cerosaletti, K. M., Concannon, P., and James, M. R. **2003** "Functional delivery of large genomic DNA to human cells with a peptide-lipid vector." *J. Gene Med.*, 5: 883-892.

146. Montigny, W. J., Phelps, S. F., Illenye, S., and Heintz, N. H. **2003** "Parameters influencing high-efficiency transfection of bacterial artificial chromosomes into cultured mammalian cells." *Biotechniques*, 35: 796-807.

147. Warming, S., Costantino, N., Court DL, Jenkins, N. A., and Copeland, N. G. **2005** "Simple and highly efficient BAC recombineering using galK selection." *Nucleic Acids Res.*, 33: e36.

148. Mendell, J. T., ap Rhys, C. M., and Dietz, H. C. **2002** "Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts." *Science*, 298: 419-422.

149. Smith, C. W., Chu, T. T., and Nadal-Ginard, B. **1993** "Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns." *Mol. Cell Biol.*, 13: 4939-4952.

150. Chua, K. and Reed, R. **2001** "An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing." *Mol. Cell Biol.*, 21: 1509-1514.

151. Query, C. C., McCaw, P. S., and Sharp, P. A. **1997** "A minimal spliceosomal complex A recognizes the branch site and polypyrimidine tract." *Mol. Cell Biol.*, 17: 2944-2953.

152. Reed, R. **1989** "The organization of 3' splice-site sequences in mammalian introns." *Genes Dev.*, 3: 2113-2123.

153. Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. **1999** "Functional recognition of the 3' splice site AG by the splicing factor U2AF35." *Nature*, 402: 832-835.

154. Gabut, M., Mine, M., Marsac, C., Brivet, M., Tazi, J., and Soret, J. **2005** "The SR protein SC35 is responsible for aberrant splicing of the E1alpha pyruvate dehydrogenase mRNA in a case of mental retardation with lactic acidosis." *Mol. Cell Biol.*, 25: 3286-3294.

155. Crovato, T. E. and Egebjerg, J. **2005** "ASF/SF2 and SC35 regulate the glutamate receptor subunit 2 alternative flip/flop splicing." *FEBS Lett.*, 579: 4138-4144.

156. Buchner, D. A., Trudeau, M., and Meisler, M. H. **2003** "SCNM1, a putative RNA splicing factor that modifies disease severity in mice." *Science*, 301: 967-969.

157. Petruzzella, V., Panelli, D., Torraco, A., Stella, A., and Papa, S. **2005** "Mutations in the NDUFS4 gene of mitochondrial complex I alter stability of the splice variants." *FEBS Lett.*, 579: 3770-3776.

158. Cartegni, L., Hastings, M. L., Calarco, J. A., de, Stanchina E., and Krainer, A. R. **2006** "Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2." *Am. J. Hum. Genet.*, 78: 63-77.

159. Liu, H. X., Zhang, M., and Krainer, A. R. **1998** "Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins." *Genes Dev.*, 12: 1998-2012.

160. Yang, L., Embree, L. J., Tsai, S., and Hickstein, D. D. **1998** "Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing." *J. Biol. Chem.*, 273: 27761-27764.

161. Cowper, A. E., Caceres, J. F., Mayeda, A., and Screaton, G. R. **2001** "Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing." *J. Biol. Chem.*, 276: 48908-48914.

162. Vockley, J., Rogan, P. K., Anderson, B. D., Willard, J., Seelan, R. S., Smith, D. I., and Liu, W. **2000** "Exon skipping in IVD RNA processing in isovaleric acidemia caused by point mutations in the coding region of the IVD gene." *Am. J. Hum. Genet.*, 66: 356-367.

163. Okajima, K., Warman, M. L., Byrne, L. C., and Kerr, D. S. **2006** "Somatic mosaicism in a male with an exon skipping mutation in PDHA1 of the pyruvate dehydrogenase complex results in a milder phenotype." *Mol. Genet. Metab.*, 87: 162-168.

164. Gehring, N. H., Kunz, J. B., Neu-Yilik, G., Breit, S., Viegas, M. H., Hentze, M. W., and Kulozik, A. E. **2005** "Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements." *Mol. Cell*, 20: 65-75.