

HUMAN CARBOXYLESTERASE 2 SPLICE VARIANTS:  
EXPRESSION, ACTIVITY, AND ROLE IN THE  
METABOLISM OF IRINOTECAN AND CAPECITABINE

Marissa Ann Schiel

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the Department of Biochemistry and Molecular Biology,  
Indiana University

February 2009

Accepted by the Faculty of Indiana University, in partial  
fulfillment of the requirements for the degree of Doctor of Philosophy.

---

William F. Bosron, Ph.D., Chair

---

E. Gabriela Chiorean, M.D.

Doctoral Committee

---

David A. Flockhart, M.D., Ph.D.

---

Maureen A. Harrington, Ph.D.

August 28, 2008

---

Sonal P. Sanghani, Ph.D.

*To my Poppa,  
Who encouraged me to finish and do my best.*

## ACKNOWLEDGEMENTS

I am sincerely thankful for all the help and encouragement I have received while pursuing my graduate education. I would like to gratefully acknowledge the following individuals:

- Dr. William Bosron for his passion for both science and education. I met Dr. Bosron on my very first visit to IUSM, and I was beyond pleased when I found a place in his lab. His wisdom and generosity truly enhanced my graduate experience.
- Dr. Sonal Sanghani for her knowledge and guidance during every day of this journey. I am grateful that she is both my mentor and my friend.
- Dr. Maureen Harrington, committee member and co-director of the MSTP. Her mentorship has been invaluable as I pursued both my graduate and medical studies. With her guidance, I happily did a rotation and found a home in the Bosron lab. She and Dr. Sanghani are my role models for being a strong female scientist.
- Dr. David Flockhart for the wisdom and thoughtfulness he brought to each committee meeting. His questions were encouraging and thought-provoking, and they always led to a step forward in my research.
- Dr. Gabi Chiorean, my translation research mentor, for her kindness and enthusiasm during our collaboration on the HOG GI03-53 project. I admire her passion for medicine and clinical research.
- Dr. Paresh Sanghani for his support in the lab, for sharing his knowledge of protein biochemistry, and for completing the circular dichroism studies.

- Wilhelmina Davis, my lab mate, for her assistance with “all things protein,” especially protein purification and westerns. I am truly thankful for her encouragement and friendship.
- Sharry Fears, my lab mate, lunch buddy, and fellow Big Ten supporter, for her work on the sub-cellular localization studies. I am grateful for all of her help, support, and friendship.
- Scheri-lyn Green for all of her work on PCR and cloning. I look forward to working with her in the future as a physician colleague.
- Lan Min Zhai for her assistance with cloning, protein purification, and cell culture and for her ability to always make me smile.
- Susan Perkins from the Indiana University Cancer Center for performing kurtosis analysis on the tissue sample data.
- All of the friends I made on third floor of the BRTC including Darlene Lambert, Jack Arthur, Bradley Poteat, Alice Nakatsuka, Oun Kiev, Amy Dietrich, Pam Kelley, and the members of the Goebel and Harris labs. I appreciate the knowledge, advice, humor and commiserating we have all shared.
- Dr. Wade Clapp, Jan Receveur, and my fellow combined degree students for their friendship, support and advice during this seven year journey.
- Dr. Mike Zimmer and Dr. Hendrick Szurmant whose enthusiasm for science while graduate students at the University of Illinois inspired me to pursue a graduate degree in research.

- My extended family, otherwise known as my entourage, my grandparents June Collins and Zvonimir and Maria Jugovic; my aunts and uncles Bob and Linda Reiff and John and Cheryl Jugovic; and my cousins Erin Dunivan and Kristin and Scott Petherick. Your love and support throughout my life and education has been and continues to be incredible.
- My brother Robbie Collins for challenging me and supporting me in ways only a sibling could. I am grateful that you are both my brother and my friend.
- My parents Bob and Mary Ann Collins for always loving me, supporting me and inspiring me to do my best. I am truly blessed to have such remarkable parents.
- My husband Zack Schiel for sharing with me in both the joys and frustrations of this adventure. I am genuinely grateful for his boundless love and support without which I would not have happily made it this far.

## ABSTRACT

Marissa Ann Schiel

### Human Carboxylesterase 2 Splice Variants: Expression, Activity, and Role in the Metabolism of Irinotecan and Capecitabine

Carboxylesterases (CES) are enzymes that metabolize a wide variety of compounds including esters, thioesters, carbamates, and amides. In humans there are three known carboxylesterase genes CES1, CES2, and CES3. Irinotecan (CPT-11) and capecitabine are important chemotherapeutic prodrugs that are used for the treatment of colorectal cancer. Of the three CES isoenzymes, CES2 has the highest catalytic efficiency for irinotecan activation. There is large inter-individual variation in response to treatment with irinotecan. Life-threatening late-onset diarrhea has been reported in approximately 13% of patients receiving irinotecan. Several studies have reported single nucleotide polymorphisms (SNPs) for the CES2 gene. However, there has been no consensus on the effect of different CES2 SNPs and their relationship to CES2 RNA expression or irinotecan hydrolase activity. Three CES2 mRNA transcripts of approximately 2kb, 3kb, and 4kb have been identified by multi-tissue northern analysis. The expressed sequence tag (EST) database indicates that CES2 undergoes several splicing events that could generate up to six potential proteins. Four of the proteins CES2, CES2<sup>Δ458-473</sup>, CES2<sup>+64</sup>, CES2<sup>Δ1-93</sup> were studied to characterize their expression and activity. Multi-tissue northern analysis revealed that CES2<sup>+64</sup> corresponds to the 4kb and 3kb transcripts while CES2<sup>Δ1-93</sup> is located only in the 4 kb transcript. CES2<sup>Δ458-473</sup> is an inactive splice variant

that accounts for approximately 6% of the CES2 transcripts in normal and tumor colon tissue. There is large inter-individual variation in CES2 expression in both tumor and normal colon samples. Characterization of CES2<sup>+64</sup> identified the protein as normal CES2 indicating that the signal peptide is recognized in spite of the additional 64 amino acids at the N-terminus. Sub-cellular localization studies revealed that CES2 and CES2<sup>+64</sup> localize to the ER, and CES2<sup>Δ1-93</sup> localizes to the cytoplasm. To date CES2 SNP data has not provided any explanation for the high inter-individual variability in response to irinotecan treatment. Multi-tissue northern blots indicate that CES2 is expressed in a tissue specific manner. We have identified the CES2 variants which correspond to each mRNA transcript. This information will be critical to defining the role of CES2 variants in the different tissues.

William F. Bosron, Ph.D.



## TABLE OF CONTENTS

List of Tables .....	xi
List of Figures .....	xii
List of Abbreviations .....	xiv

### INTRODUCTION

I. Carboxylesterase genes and enzyme functions.....	2
II. CES2 structure and polymorphisms.....	6
III. Gene splicing .....	10
IV. Colorectal cancer .....	12
V. Irinotecan .....	14
VI. Capecitabine.....	18
VII. Research objectives.....	21

### METHODS

I. Materials .....	22
II. Tissue-specific expression of <i>CES2</i> splice variants.....	23
III. Analysis of <i>CES2</i> and <i>CES2</i> <sup>Δ458-473</sup> in paired tumor and normal colon samples.....	25
IV. Characterization of the <i>CES2</i> <sup>Δ458-473</sup> variant .....	29
V. Characterization of the <i>CES2</i> <sup>+64</sup> variant .....	31
VI. Sub-cellular localization of <i>CES2</i> variants .....	38
VII. The role of <i>CES2</i> , <i>CES1</i> , <i>TOPO I</i> , <i>TP</i> , <i>TS</i> , <i>DPD</i> , $\beta$ - <i>GUS</i> , and <i>UGT1A1</i> in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine.....	41

## RESULTS

I.	Tissue-specific expression of <i>CES2</i> splice variants .....	47
II.	Analysis of <i>CES2</i> and <i>CES2</i> <sup>Δ458-473</sup> in paired tumor and normal colon samples .....	48
III.	Characterization of the <i>CES2</i> <sup>Δ458-473</sup> variant .....	58
IV.	Characterization of the <i>CES2</i> <sup>+64</sup> variant .....	63
V.	Sub-cellular localization of <i>CES2</i> variants .....	76
VI.	The role of <i>CES2</i> , <i>CES1</i> , <i>TOPO I</i> , <i>TP</i> , <i>TS</i> , <i>DPD</i> , $\beta$ - <i>GUS</i> , and <i>UGT1A1</i> in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine.....	78

## DISCUSSION

I.	Characterization of <i>CES2</i> splice variants .....	86
II.	The role of <i>CES2</i> , <i>CES1</i> , <i>TOPO I</i> , <i>TP</i> , <i>TS</i> , <i>DPD</i> , $\beta$ - <i>GUS</i> , and <i>UGT1A1</i> in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine.....	96
III.	Summary .....	100

REFERENCES .....	102
------------------	-----

## CURRICULUM VITAE

## LIST OF TABLES

1. Human carboxylesterase gene family	3
2. Nonsynonymous coding SNPs reported for <i>CES2</i>	9
3. Colon tumor samples	25
4. Forward (F) and reverse (R) primers for real-time PCR	44
5. Plasmids used for standard curves in real-timePCR	45
6. Expression and activity data for paired tumor (T) and normal (N) colon tissue samples	55
7. $CES2^{+64}$ protein purification yield	68
8. N-terminal sequencing results for $CES2^{+64}$	75
9. Gene expression data for HOG GI03-053 rectal samples	80

## LIST OF FIGURES

1. Multi-tissue Northern blot analysis of human carboxylesterases	4
2. <i>CES2</i> gene structure	7
3. <i>CES2</i> variant proteins	8
4. Irinotecan (CPT-11) metabolism	15
5. Capecitabine metabolism	19
6. Multiple Tissue Northern (MTN) blot analysis	24
7. Strategy for cloning the <i>pEGFP-CES2<sup>+64</sup></i> construct	39
8. Outline of the strategy for rectal samples collected for the GI03-53 study	42
9. Northern analysis of <i>CES2<sup>Δ1-93</sup></i> and <i>CES2<sup>+64</sup></i>	47
10. Alternative splicing in exon 10	49
11. Real-time PCR standard curve for <i>CES2<sup>Δ458-473</sup></i>	50
12. Melt curve analysis for <i>CES2<sup>Δ458-473</sup></i> real-time PCR products	50
13. Reproducibility of real-time PCR methods	51
14. Expression of <i>CES2</i> and <i>CES2<sup>Δ458-473</sup></i> in 10 paired tumor and normal colon samples	53
15. Non-denaturing polyacrylamide activity gel for paired colon tissue samples	54
16. Correlation of <i>CES2</i> expression with carboxylesterase activity in colon tissue	57
17. Characterizations of recombinant <i>CES2<sup>Δ458-473</sup></i> and <i>CES2</i> proteins	60
18. CPT-11 hydrolysis by <i>CES2</i> and <i>CES2<sup>Δ458-473</sup></i>	62
19. PCR analysis of viral DNA for selection of a <i>CES2<sup>+64</sup></i> virus	65

20. SDS-PAGE analysis of the purification of recombinant CES2 <sup>+64</sup> protein from the media of <i>Sf9</i> insect cells	66
21. SDS-PAGE analysis of the purification of recombinant CES2 <sup>+64</sup> proteins from <i>Sf9</i> insect cells	67
22. Activity (A) and western blot (B) analysis of CES2 <sup>+64</sup>	70
23. Coomassie blue staining of CES2 <sup>+64</sup> on a non-denaturing polyacrylamide gel	70
24. SDS-PAGE analysis of recombinant CES2 <sup>+64</sup> proteins	72
25. Western blot analysis of recombinant CES2 <sup>+64</sup> proteins	72
26. GNA glycosylation staining of recombinant CES2 <sup>+64</sup> protein	73
27. PVDF membranes with CES2 <sup>+64</sup> protein bands for N-terminal sequencing	75
28. Localization of CES2 variant-GFP constructs in HCT-15 cells	77
29. Summary of the protocol for the HOG GI03-053 rectal tissue samples	79
30. <i>UGT1A1</i> sequencing chromatograms	81
31. Expression profiles of HOG GI03-53 complete responders (pCR) and non-complete responders (pNCR)	83
32. Comparison between complete responders (pCR) and non-complete responders (pNCR) with respect to the expression of <i>TP</i> , <i>TS</i> , <i>TOPO I</i> , and <i>CES1</i>	84

## LIST OF ABBREVIATIONS

4-MU	4-methylumbelliferone
4-MUA	4- methylumbelliferyl acetate
5'-DFCR	5'-deoxy-5-fluorocytidine
5'-DFUR	5'-deoxy-5-fluorouridine
5-FU	5-floururacil
Ala	alanine
APC	7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin
Asn	asparagine
$\beta$ -GUS	$\beta$ -glucuronidase
CD	circular dichroism
CES	carboxylesterase
ConA	concanavalin
CNV	copy number variant
CPT-11	Irinotecan, 7-ethyl-10-[4-(1-piperidino)-1-piperidino] carbonyloxycamptothecin
CYP	cytochrome P450
DIG	Digoxigenin
DPD	dihydropyrimidine dehydrogenase
DSA	Datura stramonium agglutinin
ER	Endoplasmic reticulum
EST	Expressed sequence tag
FdUMP	5-fluoro-2'-deoxyuridine 5'-monophosphate

GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green florescence protein
Glu	glutamate
GLY	glycosylation sites
GNA	<i>Galanthus nivalis</i> agglutinin
HCT-15	Human colon adenocarcinoma cell line
His	histidine
HOG	Hoosier Oncology Group
MD	moderately differentiated
MTN	multi-tissue northern
N	normal
NBT/ X-phosphate	4-nitro blue tetrazolium chloride/5-bromo-4-chloro-3-indolyl-phosphate
NCBI	National Center for Biotechnology Information
NPC	7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin
N-X-S/T-(P)	Asparagine-any amino acid-serine/threonine-(phosphorylated)
P1	Passage 1
P2	Passage 2
P3	Passage 3
PCR	polymerase chain reaction
pCR	pathologic complete responder
PAGE	polyacrylamide gel electrophoresis
PD	poorly differentiated
PGAP	Pyroglutamate aminopeptidase

pNCR	pathologic non-complete responder
Pro	Proline
PVDF	Polyvinylidene fluoride
SDS	sodium dodecyl sulfate
Ser	serine
SN-38	7-Ethyl-10-hydroxycamptothecin
SN-38G	7-Ethyl-10-hydroxycamptothecin glucoronide
SNA	Sambucus nigra agglutinin
SNP	single nucleotide polymorphism
SSC	sodium chloride-sodium citrate
T	tumor
TOPO I	Topoisomerase I
TP	thymidine phosphorylase
TS	thymidylate synthase
UGT	UDP-glucuronosyltransferases



## INTRODUCTION

Carboxylesterases (CES) are enzymes that metabolize a wide variety of compounds including esters, thioesters, carbamates, and amides. In humans there are three known carboxylesterase genes *CES1*, *CES2*, and *CES3*. Of the three, *CES2* has the highest catalytic efficiency with regards to irinotecan metabolism. *CES2* as well as *CES1* also contribute to the metabolism of capecitabine. Both irinotecan (CPT-11) and capecitabine are important chemotherapeutics for the treatment of colorectal cancer. There is large inter-individual variation in response to treatment with irinotecan. Life-threatening late-onset diarrhea has been reported in about 13% of patients receiving irinotecan. Several studies have reported single nucleotide polymorphisms (SNPs) for the *CES2* gene. However, there has been no consensus on the effect of different *CES2* SNPs and their relationship to *CES2* RNA expression or irinotecan hydrolase activity. The expressed sequence tag (EST) database indicates that *CES2* undergoes several splicing events that could lead to six potential proteins. It is essential to study the pharmacodynamics and pharmacokinetics of these drugs in order to improve treatment outcomes and limit side effects. It is our hypothesis that inter-individual variation in response to irinotecan and capecitabine therapy, used for the treatment of colorectal cancer, may be attributed to the expression levels and activities of the *CES2* splice variants. Only one of the six potential proteins, wild-type *CES2*, has been studied to a significant degree. The goal of this research is to understand the expression patterns and activity of the *CES2* splice variants and to study factors that are responsible for the inter-individual variation in response to irinotecan and capecitabine treatment.

## **I. Carboxylesterase genes and enzyme functions**

Carboxylesterases (CES) (E.C.3.1.1.1) are  $\alpha/\beta$ -hydrolase fold proteins belonging to the serine esterase superfamily (Aldridge, 1993). Members of the superfamily of  $\alpha/\beta$  hydrolases are described at the ESTHER database (Hotelier et al., 2004).

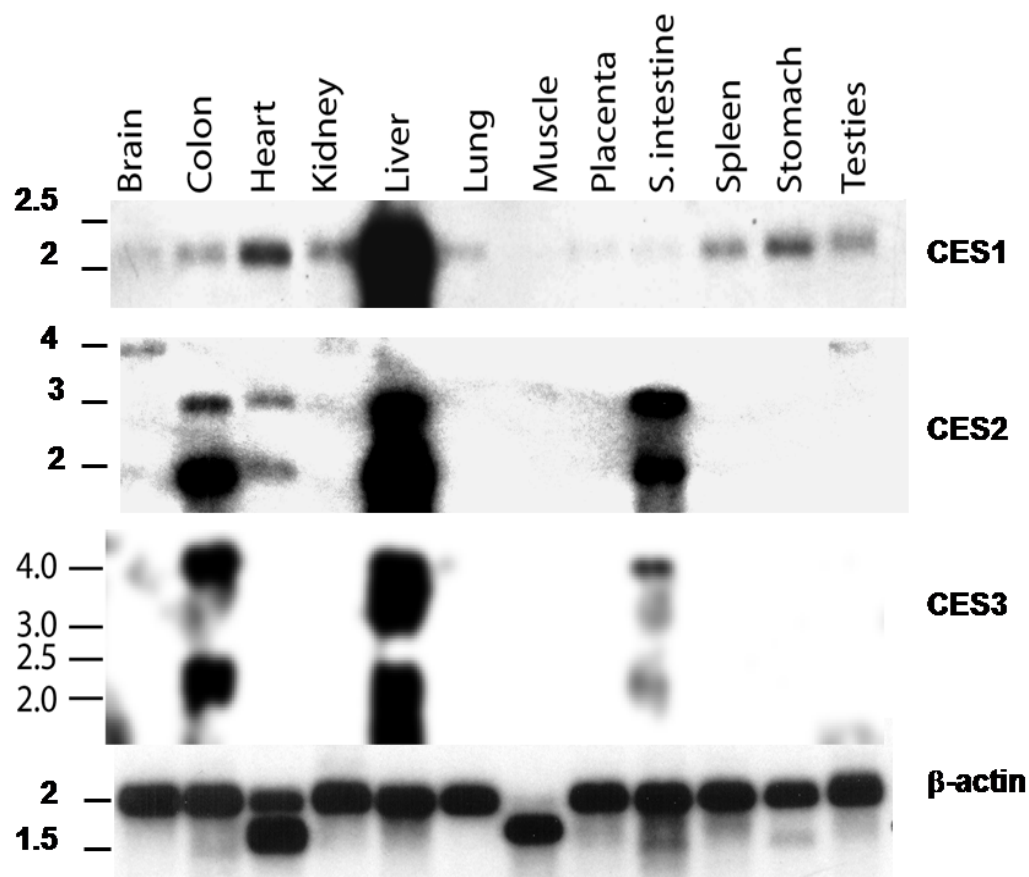
Carboxylesterases catalyze the hydrolysis of esters, thioesters, carbamates, and amides. Endogenous substrates of carboxylesterases include short and long chain acyl-glycerols, long chain acylcarnitines, and long-chain acyl CoA esters. A significant physiological role of carboxylesterases is the detoxification of exogenous compounds as well as the activation of prodrugs (Satoh and Hosokawa, 1998). Catalyzing phase I hydrolysis reactions, carboxylesterases can increase the polarity of an exogenous substrate thus enhancing its elimination. Exogenous substrates of carboxylesterases include angiotensin-converting enzyme inhibitors, salicylates, haloperidol, cocaine, heroin, and the chemotherapeutics irinotecan and capecitabine (Satoh and Hosokawa, 1998). Due to their broad substrate specificity and ability to function as esterases or lipases, it became increasingly difficult to classify carboxylesterases by substrate type. Satoh and Hosokawa (1998) proposed a novel classification system that organized the carboxylesterases into four main classes based on sequence similarity. More recently a fifth class of carboxylesterases has been identified that differs in structure from the other four families (Satoh and Hosokawa, 2006).

In humans, there are five carboxylesterase classes recognized by the Human Gene Organization Nomenclature Committee (Eyre et al., 2006) (Table 1). The three major carboxylesterase genes CES1, CES2, and CES3 each belong to a different class (Satoh and Hosokawa, 1998). CES1 is a 180kDa trimer, while CES2 and CES3 are

60kDa monomers. There is approximately 48% sequence homology between CES1 and CES2. CES3 shares approximately 40% sequence homology with both CES1A1 and CES2 (Sanghani et al., 2004). CES1 is ubiquitously expressed, and CES2 is mainly found in the liver and intestines (Quinney et al., 2005; Satoh et al., 2002; Wu et al., 2003). CES3 has a similar tissue distribution pattern to that of CES2 (Sanghani et al., 2004) (Figure 1). However, the amount of *CES3* transcript in the colon is significantly less than that of *CES2* (Sanghani et al., 2003).

<b>HUGO nomenclature</b>	<b>GeneID</b>	<b>Genbank Accession number</b>	<b>Gene Type</b>	<b>Aliases</b>
CES1	1066	NM_001025195 NM_001025194	Protein coding	hCE1, CEH, PCE-1
CES2	8824	NM_003869 NM_198061	Protein coding	hCE2, iCE, PCE-2
CES3 <sup>a</sup>	23491	NM_024922	Protein coding	
CES4	51716	NR_003276	Pseudo	
CES7	221223	NM_145024	Protein coding	CAUXIN, CES5

**Table 1. Human carboxylesterase gene family**  
((Sanghani et al., 2008, accepted for publication))



**Figure 1. Multi-tissue Northern blot analysis of human carboxylesterases:**

Distribution of carboxylesterases in human tissues was examined using a multi-tissue Northern Blot purchased from Origene Technologies (Rockville, MD). Specific cDNA probes were developed for *CES1*, *CES2*, and *CES3*.  $\beta$ -actin was probed as a loading control. Exposure time varied from 12 hours (*CES1*) to 8 days (*CES3*). (From Quinney, 2004))

The majority of mammalian carboxylesterases are glycosylated ER proteins containing ER signal peptide and retention sequences at the N-terminus and C-terminus, respectively (Satoh and Hosokawa, 1998). However, Takagi et al. (1988) and Long et al. (1988) have reported sequences that encode for secretory carboxylesterases. The ER signal sequence generally is comprised of 17-20 hydrophobic amino acids with a bulky aromatic residue preceded by a small neutral residue immediately followed by the cleavage site (von, 1983). The C-terminal HXEL consensus sequence (Robbi and Beaufay, 1991) interacts with the KDEL receptor in the ER lumen (Satoh and Hosokawa, 1998). N-linked glycosylation motifs N-X-S/T-(P) are also conserved among the carboxylesterases. Kroetz et al. (1993) provided data indicating that glycosylation may be necessary for optimal esterase activity. Many carboxylesterases also contain four cysteine residues that are involved in disulfide bonds. The carboxylesterase conserved catalytic site is comprised of a triad of the amino acid residues serine (Ser), glutamate (Glu), and histidine (His) (Cyglar et al., 1993; Hosokawa, 2008; Satoh and Hosokawa, 1998). Carboxylesterases hydrolyze substrates by a two-step, ping-pong catalytic mechanism. Histidine acts as a base to remove a proton from serine. The serine-O<sup>-</sup> nucleophile is free to attack the carbonyl group of the substrate forming a tetrahedral intermediate. Two conserved glycine residues form an oxyanion hole which serves to stabilize the tetrahedral intermediate. The ester bond breaks and acyl-enzyme complex forms. Glutamate and histidine pair within the catalytic triad to stabilize and orient the structure. Histidine donates a proton to the alcohol leaving group. A water molecule acts as a nucleophile and attacks the acyl-enzyme intermediate producing a second tetrahedral

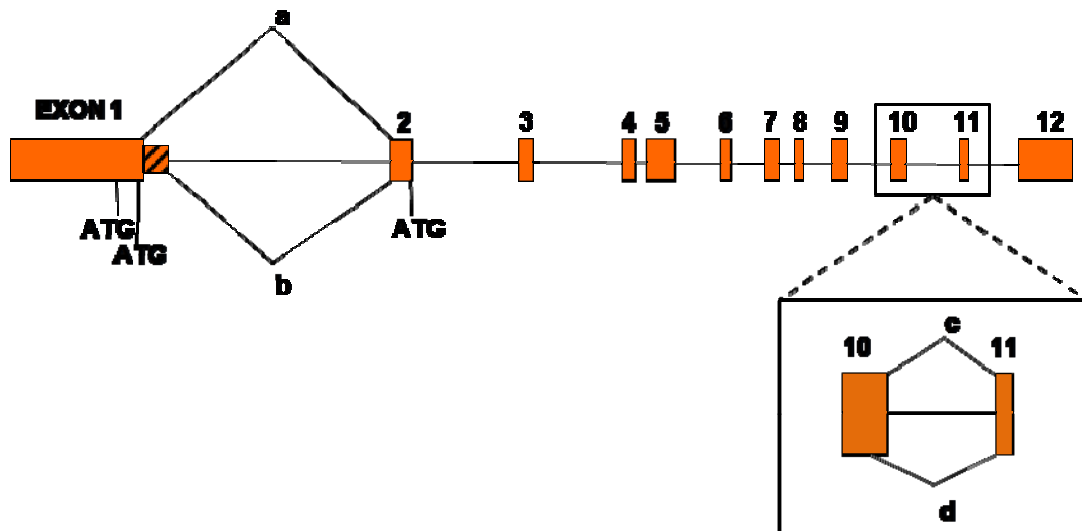
intermediate. The carboxylic acid product is eliminated, and the enzyme catalytic site is reconstituted.

## **II. CES2 structure and polymorphisms**

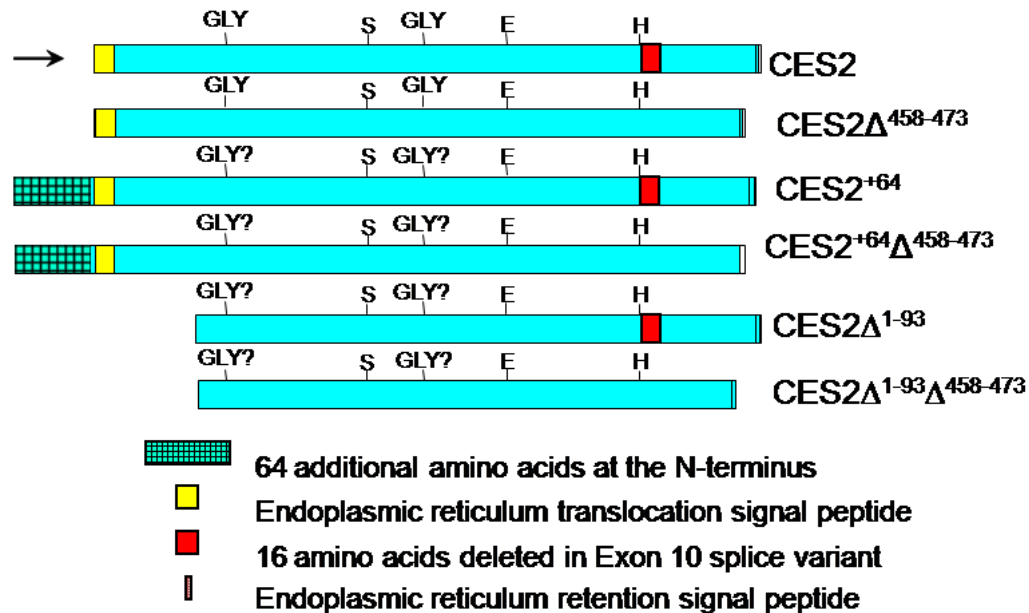
Pindel et al. (1997) reported the cloning of a 533-amino acid mature human liver h-CE2 (old nomenclature for CES2) which displayed 73% homology to rabbit liver CES2 and 67% homology to hamster AT51p. CES2 is a 60kDa serine ester hydrolase (carboxylesterase) with Ser<sup>228</sup>, Glu<sup>345</sup>, and His<sup>457</sup> forming its catalytic triad. There is an ER signal peptide within the first 27 N-terminal amino acid residues. The C-terminus has the ER retention sequence HTEL (Pindel et al., 1997; Robbi and Beaufay, 1991). Two N-linked glycosylation sites are located at residues Asn<sup>111</sup> and Asn<sup>276</sup> (Schwer et al., 1997).

The *CES2* gene is situated on chromosome 16, spans 10 kb, and includes 12 exons. Eleven splice variants encoding ten proteins are reported for *CES2* in the AceView database (Thierry-Mieg and Thierry-Mieg, 2006). The AceView database “provides a strictly cDNA-supported view of the human transcriptome and the genes by summarizing all quality-filtered human cDNA data from GenBank, dbEST and the RefSeq” (Thierry-Mieg and Thierry-Mieg, 2006). The expressed sequence tag (EST) database (Unigene) indicates that *CES2* includes two in-frame ATGs in exon 1 and potential alternative splicing sites in exon 1 and exon 10 (Figure 2). Combinations of these splicing events could lead to six potential CES2 protein splice variants (Figure 3). Only one of the six proteins, wild-type CES2 (indicated with an arrow in Figure 3), has been studied to a significant degree. It is believed that CES2 is expressed when translation begins at the second ATG in exon 1; it is surrounded by the Kozak consensus

sequence (Schwer et al., 1997). If translation were to begin at the first ATG, 64 amino acids would be added to the N-terminus creating CES2<sup>+64</sup>. In one EST clone (gi:21754794), alternative splicing adds 270 nucleotides from intron 1 to the mRNA. When this happens, the first in-frame ATG is found in exon 2 and the resulting protein CES2<sup>Δ1-93</sup> lacks the 93 N-terminal amino acids. The various events in exon 1 could potentially be paired with alternative splicing events in exon 10 yielding CES2<sup>Δ458-473</sup>, CES2<sup>+64Δ458-473</sup>, and CES2<sup>Δ1-93Δ458-473</sup>. Alternative splicing in exon 10 (Figure 3d) eliminates the 16 amino acids immediately following the active site histidine.



**Figure 2. *CES2* gene structure:** The *CES2* gene covers approximately 10 kb and is comprised of 12 exons. There are two in frame ATGs found in exon 1. Alternative splicing in exon 1 (b) adds 270 nucleotides to the message, and the first available in frame ATG is then found in exon 2. Alternative splicing in exon 10 (d) results in a 48 nucleotide deletion in the message. These splicing events may yield 6 potential proteins found in Figure 3.



**Figure 3. CES2 variant proteins:** Based on the splicing shown in Figure 2 there are six potential CES2 proteins. The protein marked by the arrow is the most commonly studied isoform of CES2. Serine (S), Glutamate (E), and Histidine (H) are the three catalytic site residues. Alternative splicing in exon 10 removes the 16 amino acid residues immediately following the catalytic residue, Histidine. CES2 is a glycoprotein with the glycosylation sites marked GLY.

Northern blot analysis of *CES2* reveals three transcripts (Satoh et al., 2002; Schwer et al., 1997; Wu et al., 2003) of approximately 2 kb, 3 kb, and 4.2 kb in length. The expression pattern and intensity varies between the different tissue types (Figure 1). The 2 kb and 3 kb transcripts are largely expressed in the liver, colon, and small intestine and to a lesser degree in the heart. The approximately 4 kb transcript is located in the brain, kidney, and testes. The multiple transcripts may arise from splice variants or the use of alternate promoters (Satoh et al., 2002; Wu et al., 2003).



Seventy-two single nucleotide polymorphisms (SNPs) for *CES2* have been identified in the NCBI SNP database. There are seven SNPs reported in the coding region of which four are nonsynonymous. Different labs have reported single nucleotide polymorphisms for the *CES2* gene (Charasson et al., 2004; Kim et al., 2003; Kubo et al., 2005; Marsh et al., 2004; Wu et al., 2004). Marsh et al. (2004) found no correlation between SNPs and *CES2* mRNA expression in normal tissues, but the intronic SNP IVS10-88 was associated with decreased *CES2* mRNA levels in colorectal tumors. Studying the Japanese population, Kim et al. (2003) found no such association but did find R34W to have decreased enzymatic activity. As a follow-up study, Kubo et al. (2005) reported that two nonsynonymous SNPs, C100T (R34W) and G424A (V124M), resulted in decreased carboxylesterase activity in spite of increased protein expression. Also, the intronic SNP IVS8-2A resulted in truncated proteins. Charasson et al. (2004) did not identify any SNPs that had significant influence on mRNA expression or CES activity. These studies indicate that there is confusion as to the role of SNPs in inter-individual variation in response to irinotecan.

	<b>Nucleotide</b>	<b>Amino acid</b>	<b>SNP number</b>
1	C1406T	R136ter	rs28382815
2	G1685A	A229T	rs11568312
3	G1809A	R270H	rs8192924
4	G1937A	G313R	rs10852434

**Table 2. Nonsynonymous coding SNPs reported for *CES2*:** The amino acid numbers include an additional 64 amino acids at the N-terminal of *CES2*. Ter = termination.

### III. Gene splicing

For a gene to be expressed, the genomic code must be transcribed into RNA which is then translated into protein. A key element of this process is the splicing of pre-mRNA into mature mRNA. Splicing is the process whereby non-coding intronic sequences are removed from the pre-mRNA and the exons are re-ligated to form the mature mRNA molecule (Nilsen, 2003). On average greater than 90% of pre-mRNA sequence is spliced out to form mature mRNA (Stamm et al., 2005). Four key sequences defining splice sites are contained within the introns. These sequences are the 5' and 3' splice sites, the branch point region, and the polypyrimidine tract (Matlin et al., 2005; Stamm et al., 2005). In 99% of all introns the first and last dinucleotides are GT and AG which comprise the 5' and 3' splice sites, respectively (Venables, 2004). Splicing is assisted by the spliceosome, a macromolecular ribonucleoprotein complex, whose assembly is guided by the conserved splice sites (Matlin et al., 2005). The strength of a splice site is determined by other *cis*-elements including intronic and exonic enhancers and silencers. The sequence variability, location, and number of *cis*-elements play a large role in determining how the RNA will be spliced. Protein *trans*-acting factors associate with the *cis*-elements and affect the assembly of the spliceosome to further influence the outcome of splicing (Matlin et al., 2005; Mauritz et al., 2007; Stamm et al., 2005).

In 1958 George Beadle and Edward Tatum won the noble prize for their “one gene one enzyme” theory which held that one gene was responsible for the production of one enzyme in a metabolic pathway (Singer and Berg, 2004). Over time this was modified to the “one gene one polypeptide” theory to account for non-enzymatic proteins as well as proteins composed of multiple polypeptides. It was originally believed that the

human genome was comprised of over 100,000 genes. Completion of the human genome project determined that fewer than 30,000 genes formed the genome, but the human proteome is estimated to contain over 90,000 proteins (Ast, 2004; Xing, 2007). As scientists have discovered far fewer human genes than once believed, attention has turned to alternative splicing as a means of generating complex proteomes (Matlin et al., 2005; Stamm et al., 2005). Alternative splicing refers to the different ways in which the exons and introns of a single pre-mRNA can be spliced together to yield several to many different mRNA transcripts ultimately leading to the production of multiple polypeptides from one gene (Venables, 2004). There are five main types of alternative splicing: exon skipping, alternative 5' splice sites, alternative 3' splice sites, intron retention, and mutually exclusive exons (Ast, 2004). Constitutive and alternative splicing are regulated by the *cis*- and *trans*-elements previously discussed. There are two main theories to explain alternative splicing. Mutations in splice site sequences can lead to the use of weaker, alternative splice sites. Alternative splicing can also be influenced by the types, combinations, and concentrations of *trans*-acting elements present (Ast, 2004; Matlin et al., 2005; Venables, 2006).

Expressed sequence tags (ESTs) which are fragments of mature mRNA are useful in the identification of alternatively spliced transcripts. ESTs as well as full length mRNA sequences can be aligned with genomic DNA to determine the exon/intron boundaries (Xing, 2007). It is estimated that anywhere from 40 to 80% of genes are subject to alternative splicing (Matlin et al., 2005). Alternative splicing has the capacity to influence mRNA transcript levels as well as protein binding properties, enzymatic activity, intracellular location, and transcript stability (Stamm et al., 2005). Splice

variants can also be expressed in cellular or tissue specific manners. At least 15% of genetic diseases are attributable to mutations in sites that influence splicing (Matlin et al., 2005). Many studies have found that alternative splicing is associated with some cancers (Venables, 2004; 2006). It is possible that CES2 splice variants could be expressed in a tumor specific manner and therefore be responsible for the inter-individual variation that is seen in response to irinotecan and capecitabine treatment.

#### **IV. Colorectal cancer**

Colorectal cancer is the third leading cause of new cancer cases and cancer deaths in both men and women. Overall, it is the fourth most frequently diagnosed cancer in the United States and is the second leading cause of all cancer-related deaths, following only lung cancer. Colorectal cancer accounts for approximately 10% of all cancer-related deaths (American Cancer Society, 2007). Depending on the grade and stage of the colorectal cancer, surgery, radiation and/or chemotherapy may be used for treatment. If colorectal cancer is detected at an early, localized stage, surgery can successfully cure the disease with a five-year survival rate of 90%. Unfortunately, only 39% of cases are detected at such a stage. If the tumor has spread then chemotherapy alone or in combination with radiation is given as adjuvant treatment. Regionally advanced colorectal cancer has a 5-year survival rate of 68% decreasing to 10% with the involvement of distant metastases. Therefore, the availability of safe and efficacious chemotherapeutics is essential. Over the past 50 years, 5-fluorouracil (5-FU) has been the most commonly used chemotherapy for the treatment of colon cancer. Two more recently approved drugs for use in the treatment of colorectal cancer are capecitabine, an

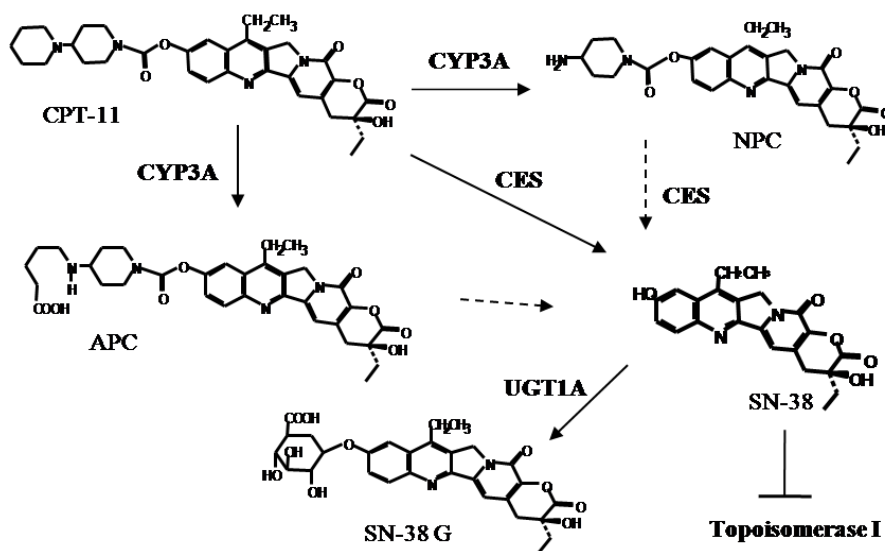
oral prodrug of 5-FU, and irinotecan. Both are carbamate prodrugs which are activated *in vivo* by carboxylesterases.

Studies have shown irinotecan to be effective in the treatment of gastric, esophageal, and colorectal cancers. Irinotecan has demonstrated its effectiveness as second-line therapy for colorectal tumors failing to respond to bolus 5-FU (Cunningham et al., 1998; Rougier et al., 1998). More recent studies have shown that the combination of irinotecan/5-FU/leucovorin as first line chemotherapy is more effective at treating metastatic colorectal cancer than 5-FU/leucovorin alone (Douillard et al., 2000; Saltz et al., 2000). Capecitabine can successfully replace 5-FU in combination therapy with irinotecan. Capecitabine does not significantly alter the pharmacokinetics of irinotecan even though both enzymes are metabolized by carboxylesterases (Czejka et al., 2005). There is large inter-individual variation in response to treatment with irinotecan and capecitabine. Addition of irinotecan to treatment regimens increases the likelihood of deleterious side effects. The most common adverse effects associated with irinotecan are late-onset diarrhea and neutropenia. Life-threatening Grade 4 diarrhea has been seen in 13% of patients treated with irinotecan (Saltz et al., 2000). The most common toxicities associated with capecitabine are anemia, diarrhea, and hand-foot syndrome (Walko and Lindley, 2005). Inter-individual variation in both therapeutic response and adverse effects are likely due to a complex combination of pharmacodynamic, pharmacokinetic, and pharmacogenomic factors. Variations in expression or activity of the genes contributing to irinotecan and capecitabine metabolism have been noted to affect drug distribution, metabolism, and elimination. Methods used to study these factors include real-time PCR, enzyme activity assays, immunohistochemistry, and gene sequencing. It

will be important to understand all aspects of the irinotecan and capecitabine pathways in order to make informed predictions on treatment outcome.

## **V. Irinotecan**

Irinotecan is an intravenously-dosed water-soluble derivative of camptothecin (Chabot, 1996). Irinotecan is metabolized by carboxylesterases and CYP enzymes found in the liver (Figure 4). Carboxylesterases in the liver convert irinotecan to SN-38, a potent topoisomerase I inhibitor. Irinotecan binds and stabilizes the “cleavable” topoisomerase I - DNA complexes. When the topoisomerase I - DNA - irinotecan complex meets the advancing replication fork double stranded breaks occur in the DNA leading to replication arrest and cell death (Liu et al., 2000). CYP3A can oxidize irinotecan to NPC or APC (Dodds et al., 1998; Haaz et al., 1998; Sanghani et al., 2004). NPC, and to a much lesser extent APC, can then be converted by carboxylesterases to SN-38. SN-38 has shown to be 1000 times more cytotoxic than irinotecan (Humerickhouse et al., 2000; Xu et al., 2002). SN-38 is converted to its inactive form SN-38G through glucuronidation. UDP-glucuronosyltransferase (UGT) 1A1, and possibly other members of the UGT1A family including UGT1A7 and UGT1A9, are responsible for the inactivation (Hanioka et al., 2001; Lankisch et al., 2005). SN-38G can be converted back to SN-38 by endogenous  $\beta$ -glucuronidases ( $\beta$ -GUS) in the liver as well as by bacterial  $\beta$ -glucuronidases found in the gut flora.



**Figure 4. Irinotecan (CPT-11) metabolism:** CPT-11 can be oxidized to NPC or APC by CYP3A4. CPT-11 is also metabolized by CES2 to the more active SN-38. SN-38 is a potent topoisomerase I inhibitor. UGT1A glucuronidates SN-38 to its inactive form SN-38G.

There is significant inter-individual variation in response to treatment with irinotecan (Canal et al., 1996; Couteau et al., 2000; Gupta et al., 1997). Many labs have studied the importance of the various irinotecan associated enzymes: CYP3A4 (Hanioka et al., 2002; Mathijssen et al., 2004; Sai et al., 2001; Santos et al., 2000), UGT1A (Carlini et al., 2005; Gagne et al., 2002; Hanioka et al., 2001; Innocenti et al., 2004; Jinno et al., 2003; Lankisch et al., 2005; Tukey et al., 2002),  $\beta$ -GUS (Kehrer et al., 2000; Takasuna et al., 1996; Tobin et al., 2006), TOPO I (Guichard et al., 1999; Jansen et al., 1997; Pavillard et al., 2004; Sanghani et al., 2003). Studies by our laboratory and others have found that CES2 may contribute to variations in response to irinotecan (Sanghani et al., 2003; Xie et al., 2002; Xu et al., 2002).

CES2 has higher affinity for irinotecan and a 100-fold greater catalytic efficiency than CES1 with respect to irinotecan metabolism (Humerickhouse et al., 2000; Sanghani et al., 2004). CES2 has a 2000-fold greater catalytic efficiency than CES3 (Sanghani et al., 2004). When treated with irinotecan, cells over-expressing CES2 have shown more cytotoxicity than cells over-expressing CES1 (Wu et al., 2002). Due to its increased affinity and activity for irinotecan, CES2 is believed to be the key enzyme *in vivo* that activates irinotecan. In patients, large inter-individual variation in response to irinotecan treatment has been demonstrated. The idea has emerged that intra-tumoral carboxylesterase activation of irinotecan may be more important than the production of SN-38 by liver carboxylesterase (de Jong et al., 2006). SN-38 levels have been shown to correlate with systemic toxicity but not with therapeutic effect (de Jong et al., 2007). There is a higher response rate in solid tumors to irinotecan as compared to other camptothecins. This also suggests a role for local, intracellular activation of irinotecan,



because the other camptothecins are not activated by carboxylesterase (Ratain, 2000). A previous study done in our laboratory found a 23-fold variation in the expression of *CES2* among 24 colorectal tumor samples (Sanghani et al., 2003). To date there has been no consensus on the role of SNPs in this inter-individual variation. In addition to wide variation in normal *CES2* expression, we propose that expression levels and activity of *CES2* splice variants may contribute to the inter-individual variation.

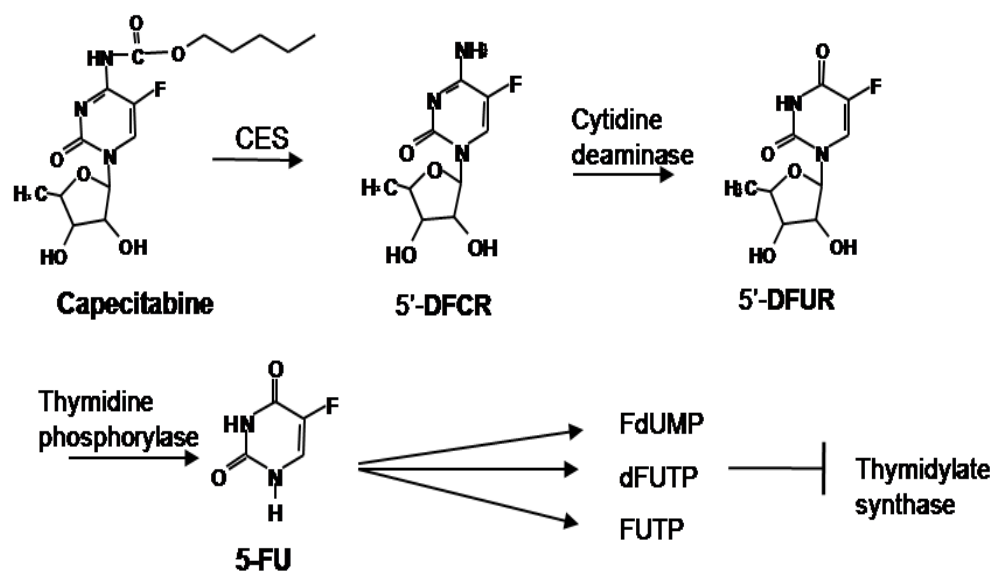
While carboxylesterase is important for the activation of irinotecan to SN-38, other factors may also contribute to the inter-individual variation in response to irinotecan. Topoisomerase I activity, but not expression level, has been shown to correlate with irinotecan sensitivity in human colon cancer cell lines (Jansen et al., 1997). However, a different study using tissue samples suggested that both topoisomerase I activity and expression were the best predictors of response to irinotecan (Pavillard et al., 2004). Guichard et al. (1999) showed a wide range in both topoisomerase I and carboxylesterase activities in colorectal carcinomas. SN-38 undergoes glucuronidation by UGTs found primarily in the liver. The wild-type *UGT1A1* promoter has six TA nucleotide repeats in the TATA box, (TA)<sub>6</sub>TAA. The *UGT1A1*\*28 allele has seven TA repeats (TA)<sub>7</sub>TAA which results in a 70% decrease in activity of the *UGT1A1* promoter (Ramchandani et al., 2007). Individuals homozygous for *UGT1A1*\*28 are the most affected. A correlation between the *UGT1A1*\*28 genotype and neutropenia has been demonstrated, but the correlation with diarrhea is less clear (Rouits et al., 2004).  $\beta$ -GUS is expressed by both the body and the gut flora of the colon. Takasuna et al. (1996) found  $\beta$ -GUS activity, but not carboxylesterase, to correlate with histological damage of the intestines. Further, antibiotic inhibition of flora  $\beta$ -GUS decreased diarrhea. Other studies

have indicated that cellular transport proteins such as *P*-glycoprotein and canalicular multispecific organic anion transporter contribute to the inter-individual variation in response to treatment (de Jong et al., 2007; Mathijssen et al., 2001).

## **VI. Capecitabine**

Capecitabine is an oral prodrug of the anti-metabolite 5-FU (Figure 5) that was designed to limit the toxicities associated with 5-FU (Miwa et al., 1998). Studies indicate that capecitabine can successfully replace 5-FU in chemotherapeutic regimens (Hoff et al., 2001; Twelves et al., 2001). For capecitabine to be activated to 5-FU, it must undergo three metabolic processes. Carboxylesterases in the liver convert 5-FU to 5'-DFCR which is then deaminated to 5'-DFUR by cytidine deaminase. CES1A1 and CES2 have similar catalytic efficiencies for capecitabine hydrolysis (Quinney et al., 2005). Thymidine phosphorylase (TP) converts 5'-DFUR to the active drug 5-FU. TP has shown to be expressed at higher levels in many tumor tissues. Over-expression of thymidine phosphorylase in tumor tissues allows for greater local conversion thus increasing the concentration of 5-FU in tumor tissues (Budman et al., 1998; Miwa et al., 1998; Schuller et al., 2000). The local conversion to 5-FU by TP is responsible for the decrease in toxicity. 5-FU metabolites can be incorporated into active nucleotide metabolites. The metabolite FdUMP targets thymidylate synthase (TS) thus lowering the production of thymidine. Both RNA and DNA production are adversely affected by metabolites of 5-FU. The enzyme dihydropyrimidine dehydrogenase (DPD) inactivates 5-FU. Patients whose tumors express low levels of TP, TS, and DPD are more likely to respond to 5-FU based therapy (Ichikawa et al., 2003; Salonga et al., 2000). However,

low levels of TS and DPD have also been associated with increased toxicity (Walko and Lindley, 2005).



**Figure 5. Capecitabine metabolism:** Capecitabine is an oral prodrug of 5-FU. Capecitabine is converted to 5'-deoxy-5-fluorocytidine (5'-DFCR) by carboxylesterases. Cytidine deaminase converts 5'-DFCR to 5'-deoxy-5-fluorouridine (5'-DFUR). Thymidine phosphorylase which is over-expressed in some cancers converts 5'-DFUR to the active 5-FU. The metabolites FdUMP, dFUTP, and FUTP interfere with DNA and RNA synthesis.

The studies of both irinotecan and capecitabine indicate complex pharmacologic profiles. Factors that predict toxicity may or may not be the same factors that predict therapeutic outcome. It will be important to further elucidate the pharmacogenomic profiles of each enzyme or drug transporter involved in the metabolism of these drugs. The data should then be used together to tailor chemotherapy for the best outcomes.

## **VII. Research objectives**

It is our hypothesis that the expression levels and activities of the CES2 splice variants will correlate with irinotecan and capecitabine activation in tumor and normal tissue. We expect to characterize the expression and activity of the carboxylesterase 2 gene and its splice variants (Aim #1 and #2). We also plan to examine the expression levels of other enzymes in the metabolic pathways of irinotecan and capecitabine (Aim #3). We expect that our research will lead to better understanding of chemotherapy by aiding clinicians in identifying patients that will benefit the most while enduring the fewest number of side effects. While this project focuses mainly on CES2 in relationship to colorectal cancer, its findings may be applicable to other cancers whose treatments include irinotecan or capecitabine.

Aim #1: Understand the expression patterns of the CES2 splice variants

- a. Tissue-specific expression
- b. Variant expression and activity in colorectal tissue

Aim #2: Characterization of CES2 variant proteins

- a. Cloning and expression of CES2 splice variants
- b. Activity of recombinant CES2 variant proteins
- c. Sub-cellular localization of CES2 variants

Aim #3: Evaluate the role of CES2 variants and other enzymes in the inter-individual variation in response to treatment of colorectal cancer with irinotecan and capecitabine.

## METHODS

### I. Materials

The Human 12-lane Multiple Tissue Northern (MTN) blot was purchased from Clontech (Palo Alto, CA) by Dr. Eileen Dolan. All radioactive nucleotides were from Perkin Elmer (Waltham, MA). The Random Primed DNA Labeling Kit and G-50 Quick Spin columns were from Roche Diagnostics (Indianapolis, IN). The Ultrahyb Solution was ordered from Ambion (Austin, TX). The RNeasy Plus Mini Kits, QIAshredders, the Allprep DNA/RNA kits, and QIAquick PCR Purification Kit were from Qiagen (Valencia, CA). Disposable mortars and pestles were purchased from Kontes. The GeneAmp Gold RNA PCR kits and SYBR Green kits were purchased from Applied Biosystems (Foster City, CA). All primers were ordered from Integrated DNA Technologies (Coralville, IA). The Zero Blunt TOPO PCR cloning kit was from Invitrogen (Carlsbad, CA). The *Sf9* insect cell and media was also from Invitrogen (Carlsbad, CA). Irinotecan was a gift from Dr. Patrick McGovern, Pharmacia-Upjohn Corporation (Peapack, NJ). Oasis HLB solid phase columns were from Waters (Milford, MA). Calf intestine alkaline phosphatase was from New England Biolabs (Ipswich, MA). *Pfu* Pyroglutamate Aminopeptidase was purchased from Takara (Otsu, Shiga, Japan). Protein standards and DEAE Affi-Gel blue were from Bio-Rad Labs (Hercules, CA). Baculogold DNA and the transfer vector were from BD-Pharmingen (San Diego, CA). Vectashield was from Vector Laboratories (Burlingame, CA). General chemicals and supplies were ordered from Sigma Chemical (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA).

## II. Tissue-specific expression of *CES2* splice variants

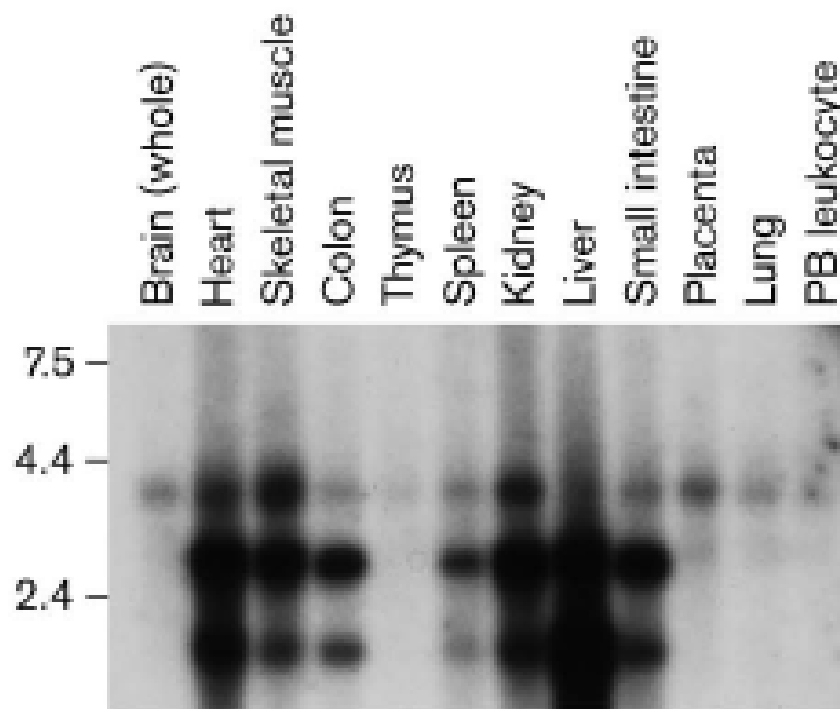
### Northern analysis

A Human 12-lane Multiple Tissue Northern (MTN) blot (Clontech) with normal tissue (Figure 6) (Wu et al., 2003) was probed with cDNA probes specific for *CES2*<sup>+64</sup> and *CES2*<sup>Δ1-93</sup>. *CES2*<sup>+64</sup> had been cloned into the pVL1392 vector (described in Methods Section V). To construct the *CES2*<sup>+64</sup> cDNA probe, approximately 40 μg of the *CES2*<sup>+64</sup>-*pVL1392* vector was digested with Bgl II and Xma I. The digested vector was electrophoresed on a 1% agarose gel. An approximately 130 bp fragment was excised and gel purified (Qiagen). To construct the *CES2*<sup>Δ1-93</sup> probe, the GeneAmp Gold RNA PCR kit (Applied Biosystems) was used to amplify the cDNA. The primers 5'-GACAGGGACCGGGCTCAGATCT-3' (sense) and 5'-TGTACTCCGCTGGTTCC TTGCC-3' (antisense) amplified a 210 bp from intron 1 of *CES2*. The reaction contained colon tumor cDNA as the template and 0.3 μM of each primer. The reaction conditions were 95°C for 10 minutes; 35 cycles of 95°C for 30 sec, 63°C for 30 sec, and 72°C for 1 min, followed by 72°C for 7 minutes. The PCR product was electrophoresed on a 1% agarose gel. A 210 bp product was excised and gel purified (Qiagen).

For both *CES2*<sup>+64</sup> and *CES2*<sup>Δ1-93</sup>, approximately 15 ng of probe were labeled with [ $\alpha$ -<sup>32</sup>P]-CTP using the Random Primed DNA Labeling Kit (Roche Diagnostics).

Unincorporated radionucleotides were separated from the labeled probe with the G-50 Quick Spin columns (Roche). Labeled probe was mixed with 50 μl of sonicated salmon sperm, denatured and added to the blot which had been pre-hybridized for approximately 75 min at 42°C with Ultrahyb Solution (Ambion). The blot was hybridized overnight.

The blot was then washed two times for 15 minutes at room temperature with 2x SSC containing 0.1% SDS followed by two 30 minute wash at 58°C and one 10 minute wash at 60°C with 0.1x SSC containing 0.1% SDS. The blot was exposed to a phosphor-storage screen at room temperature for 4-5 days.



**Figure 6. Multiple Tissue Northern (MTN) blot analysis**  
(Wu et al., 2003)



### III. Analysis of CES2 and CES2<sup>Δ458-473</sup> in paired tumor and normal colon samples

#### Colon tissue

The collection and use of these samples was approved by the Indiana University Institutional Review Board. The tumors were primary colon carcinomas of various grades and stages from all segments of the colon (Table 3). Normal colon specimens were paired with each tumor sample. All specimens were surgery samples that were reviewed by a pathologist. All samples were immediately frozen in liquid nitrogen and maintained at -70°C until use.

Sample	Site	Diagnosis	Stage
T1	Sigmoid colon	Adenocarcinoma	
T2	Right/transverse colon	PD Mucinous Adenocarcinoma	
T3	Sigmoid colon	MD Adenocarcinoma	pT3, N0, MX
T4	Colon	MD Adenocarcinoma	pT3, N0, MX
T5	Colon	MD Adenocarcinoma	pT3, N2, MX
T6	Sigmoid colon	MD Adenocarcinoma	pT3, N1, MX
T7	Transverse colon	PD Adenocarcinoma	pT4, N1, MX
T8	Cecum	MD Adenocarcinoma	pT3, N2, MX
T9	Right colon	MD Adenocarcinoma	
T10	Sigmoid colon	MD Adenocarcinoma	pT2, N0, MX

**Table 3. Colon tumor samples:** Each tumor sample was paired with a normal colon sample originating from a segment of the colon. MD and PD represent moderately- and poorly- differentiated, respectively.

### RNA isolation and reverse transcription

Tissues, ~5-30 mg, were disrupted in RLT Buffer (Qiagen) using a disposable mortar and pestle (Kontes). Disruption was followed by homogenization via QIAshredder columns (Qiagen). Total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen). RNA was quantified using a ND-1000 Spectrophotometer (Nanodrop). RNA quality was assessed by running 500 ng of total RNA on a 1% agarose gel. The GeneAmp Gold RNA PCR kit (Applied Biosystems) was used for reverse transcription with 1µg of total RNA added to a 50-µl reaction. The reaction components were 2.5 mM MgCl<sub>2</sub>, 250 µM of each deoxynucleotide triphosphate, 1.25 µM oligodeoxythymidylic acid primer, 0.5 U/µL RNase inhibitor, and 0.75 U/µL MultiScribe reverse transcriptase. The reaction conditions were room temperature for 10 min, 42°C for 60 min, 68°C for 10 min, and 95°C for 5 min. All reverse transcription reactions were verified by PCR using the GeneAmp Gold RNA PCR kit (Applied Biosystems) and primers for *GAPDH*. The *GAPDH* primers were 5'-GACCACAGTCCATGCCATCACT-3' (sense) and 5'-TCCACCACCCTGTTGCTGTAG-3' (antisense). A 20 µL PCR reaction contained 1x buffer, 1.75mM MgCl<sub>2</sub>, 0.8mM of each deoxynucleotide triphosphate, 0.375 µM of each primer, 1 µl of reverse transcription reaction, and 0.05U/ µL AmpliTaq Gold. The reaction conditions were 95°C for 10 min; 35 cycles of 95°C for 30 s, 63°C for 30 s, and 72°C for 1 min, followed by 72°C for 5 min. The *GAPDH* product size was ~450 bp.

### Real-time PCR

This real-time PCR protocol has been modified from that previously published (Sanghani et al., 2003) to distinguish between *CES2* and *CES2*<sup>Δ458-473</sup> variants. Variant-specific forward primers were designed for *CES2*, 5'-CCATGGTGATGAGCTTCCTT

TTGT-3', and *CES2*<sup>Δ458-473</sup>, 5'-AGGCAGACCATGTTAAATTCACTGA-3'. The reverse primer 5'-AGGTATTGCTCCTCCTGGTCGAA-3' was common to both transcripts. The expected lengths of the PCR products were 186 bp and 145 bp for *CES2* and *CES2*<sup>Δ458-473</sup>, respectively. The PCR conditions were 3 mM Mg<sup>+2</sup>, 0.5 μM of each primer, 0.2 mM of deoxynucleotide triphosphates using the SYBR Green kit (Applied Biosystems). The cDNA equivalent to 20 ng of RNA was added to each 25 μl PCR reaction. PCR cycling conditions were 50°C for 2 min, 95°C for 10 min and 40 cycles of 95°C for 30 s, 65°C for 30 s, and 72°C for 1 min on an Eppendorf Mastercycler EP instrument. The uniformity of the PCR products was determined using the melt curve function. A standard curve for each gene was generated using clones that were constructed in our laboratory. The copy numbers of each variant present in the tissues were determined by comparison to the appropriate standard curve. All cDNA samples were analyzed in triplicate for each gene.

#### Tissue lysate preparation

Approximately 10-60 mg of colon tissue were homogenized in 100-200 μl of 25 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5, plus 0.05% Triton X-100, leupeptin (10 μg/mL), pepstatin (1 μg/mL) and DNase (0.5 μg/mL) using motorized pestles with microfuge tubes (Kontes). Disrupted samples were incubated on ice for approximately 20 minutes before a final round of homogenization. The lysates were centrifuged at 40,000 *x g* for 15-20 min at 4°C. The supernatant was collected for use in activity assays. Protein content was quantified via the Bio-Rad Protein Assay with bovine serum albumin used for a standard curve.

#### 4-Methylumbelliferyl acetate hydrolysis assay

Brzenzinski et al. (1997) have previously described the use of 4-methylumbelliferyl acetate (4-MUA) in spectrophotometric assays to measure non-specific carboxylesterase activity. Ten microliters of lysate were incubated with 0.5 mM 4-methylumbelliferyl acetate at 37°C in a buffer containing 90 mM potassium phosphate buffer and 40 mM KCl (pH 7.4). The rate of formation of 4-methylumbelliferone (4-MU) was monitored at 350 nm. 4-methylumbelliferone has an extinction coefficient of 12.2 mM<sup>-1</sup> cm<sup>-1</sup>. Specific activity was calculated using the results of the 4-MUA assay and the Bio-Rad protein assay. Units of specific activity were in μmol mg<sup>-1</sup> min<sup>-1</sup>. Each lysate sample was run in duplicate.

#### Analytical non-denaturing polyacrylamide gel electrophoresis

The colon lysates were run on analytical discontinuous non-denaturing PAGE as previously described by Dean et al. (1995). Approximately 50 to 100 μg of lysates were run, as well as 0.8 μg purified CES2 and 1.0 μg CES1 as controls. To detect carboxylesterase activity, the gels were incubated with a solution of 1 mM 4-MUA in 100 mM potassium phosphate buffer pH 6.0 for 10 minutes. The product 4-MU was detected under UV light, and gel images were captured using the Gel Doc 1000 system (BioRad). Densitometry analysis was done using Quantity ONE version 1 (BioRad). To account for differences between gels, the band density values were normalized to the CES2 density on the corresponding gel.

#### Statistical analysis

Skewness, kurtosis, and overall normality of copy number, 4-MUA hydrolase activity and band density data were assessed using the methods of D'Agostino et al.

(1990). The copy numbers for each variant were independently subjected to linear regression analysis against both 4-MUA hydrolase activity and band density using JMP 4.0 (SAS Inst Inc., Cary, NC). The data were considered significant if  $p < 0.05$ . Susan Perkins from the Indiana University Cancer Center performed the kurtosis analysis.

#### **IV. Characterization of the CES2<sup>Δ458-473</sup> variant**

##### Cloning, expression, and purification of the CES2<sup>Δ458-473</sup> variant

The CES2<sup>Δ458-473</sup> variant was cloned by Dr. Sonal Sanghani and virus production was completed by Scheri-lyn Green. Briefly, a partial EST clone for CES2<sup>Δ458-473</sup> was obtained from ATCC (Image Clone ID 6195662). The CES2<sup>Δ458-473</sup> EST clone and a CES2 clone in the pVL1392 vector were both digested with Xba I and Sma I restriction enzymes. The Xba I/Sma I CES2<sup>Δ458-473</sup> fragment was ligated into the digested CES2-pVL1392 vector. The new CES2<sup>Δ458-473</sup>-pVL1392 clone was sequenced to confirm the presence of the 48 bp deletion.

The CES2<sup>Δ458-473</sup> protein was expressed using the baculovirus system (BD Biosciences Pharmingen) as previously reported (Sanghani et al., 2004; Sun et al., 2004). The CES2<sup>Δ458-473</sup>-pVL1392 vector was co-transfected with linearized Baculogold DNA (BD-Pharmingen) into  $2.5 \times 10^6$  Sf9 cells following the manufacturer's protocol. The CES2<sup>Δ458-473</sup> recombinant virus was collected on day 8, post-transfection. Sf9 insect cells were infected with CES2<sup>Δ458-473</sup> recombinant virus. After 70 hours, the cells were collected by centrifuging at  $500 \times g$  for 5 min. The cell pellet was sonicated in 30 ml of 20 mM Tris buffer, pH 7.4, containing 1 mM benzamidine, 1 mM dithiothreitol, 1 μM leupeptin and 0.1% Triton X-100. The cell lysate was ultracentrifuged at  $100,000 \times g$  for 45 min. The recombinant CES2<sup>Δ458-473</sup> variant was purified by a two step purification

protocol (Sanghani et al., 2004) using concanavalin A affinity chromatography and preparative non-denaturing PAGE (Sun et al., 2004). A protein peak was detected at the same position as CES2; however, it did not have any activity for the substrate 4-methylumbelliferyl acetate. Fractions from this peak were concentrated and loaded onto an 8.5 cm 6% non-denaturing preparative polyacrylamide gel as described earlier (Sun et al., 2004). A protein peak eluting similarly to CES2 was detected, and the eluted fractions corresponding to protein peak were pooled and concentrated.

The purified recombinant CES2<sup>Δ458-473</sup> protein was analyzed by SDS-PAGE, western blot analysis, non-denaturing gel electrophoresis, mass spectroscopy, and circular dichroism to assess its identity, purity, and physical properties. Antibodies for the western blotting were raised in rabbit to the deglycosylated human CES2 protein. The antibody was purified from 8 ml of serum using the DEAE Affi-Gel blue (Bio-Rad Labs) according to the manufacturer's protocol. The purified anti-CES2 antibody was diluted 5000-fold, and the secondary antibody was HRP-conjugated. Circular dichroism (CD) spectra for the proteins were determined using a J-720 spectropolarimeter (Jasco). The spectra were then analyzed using the CDPro software developed by Sreerama and Woody (Sreerama and Woody, 1993; Sreerama and Woody, 2004) Dr. Sonal Sanghani and Wilhelmina Davis purified the protein. Wilhelmina Davis completed the gels, protein assays, and activity assays. Dr. Paresh Sanghani completed the CD analysis.

#### CPT-11 hydrolysis assay

The CPT-11 hydrolysis assay has previously been described by Sanghani et al. (2004). In 250 µl reactions, 55µM CPT-11 was incubated with varying amounts of CES2 (0-30 µg) and CES2<sup>Δ458-473</sup> (0-50 µg) in 50 mM HEPES buffer, pH 7.4, including 10%

ethylene glycol at 37°C for 30 minutes and two hours, respectively. The reactions were halted by the addition of 670 µl of 0.1 M HCl. Fifteen µl of 0.14 mM camptothecin was added as an internal standard to the reactions. Samples were extracted on Oasis HLB solid phase columns (Waters) and were prepared as previously described (Sanghani et al., 2004). The samples were injected onto a 5 µm C18, 150 x 4.6 mm Luna column (Phenomenex). The mobile phase consisted of 28.5% acetonitrile in 0.1M KH<sub>2</sub>PO<sub>4</sub>, pH 4.0, with 3mM heptane sulfonic acid. The samples were eluted at a flow rate of 1 ml/min and then detected via fluorescence (excitation=375nm, emission=560nm). SN-38 samples (0-5 µM) were treated in a similar manner and used to create a standard curve. The ratios of area under the curve for SN-38 and camptothecin were plotted against concentration (GraphPad Prism 4.0). The unknown concentrations were calculated from the standard curve.

## **V. Characterization of the CES2<sup>+64</sup> variant**

### Cloning of the CES2<sup>+64</sup> variant

The forward 5'-ATCAGATCT**ATG**ACTGCTCAGTCCCGCTCT-3' and reverse 5'-GACCTAGAGGTGGCTTGGCAAAT-3' primers were designed to amplify approximately 400 bp corresponding to the N-terminal sequence for CES2<sup>+64</sup>. The forward primer added a Bgl II site (underlined) and contained the ATG (bold) translation start site for CES2<sup>+64</sup>. The reaction contained 0.6 µM of each primer, 0.2 mM of each nucleotide, 1 µl of human liver cDNA, 1x pfu native buffer, and pfu native enzyme. The reaction conditions were 95°C for 4 min, 5 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1.5 min, followed by 35 cycles of 95°C for 30 s and 72°C for 1.5 min. This was

followed by a final extension time of 5 min at 72°C. The PCR product was run on a 1% gel to verify the size. Scheri-lyn Green completed the PCR for *CES2*<sup>+64</sup> cloning.

The remaining PCR product was purified using the QIAquick PCR Purification Kit. The Zero Blunt TOPO PCR cloning kit (Invitrogen) was used as per the given protocol to clone the purified PCR product into the pCR-Blunt II-TOPO vector (Invitrogen). Ligation was followed by transformation into One Shot TOP10 chemically-competent *E. coli* (Invitrogen). The transformed bacteria were plated on LB plates containing 50 µg/ml kanamycin. Four colonies were selected for overnight cultures, 5ml each with 50 µg/ml kanamycin. Plasmids were isolated by miniprep (Qiagen) followed by digestion with Eco RI. The digests were analyzed by 1% agarose gel, and all four clones yielded the expected digestion pattern.

One and a half ug of both the <sup>+64</sup> insert in the pCR-Blunt II-TOPO vector and *CES2* in the pVL1392 baculovirus transfer vector were subject to overnight digestion with Bgl II and Aar I at 37°C. The digestions were analyzed on a 1% agarose gel. A 300 bp band from the <sup>+64</sup>-pCR Blunt II-TOPO clone was gel-purified using the QIAquick Gel Extraction Kit. The linearized *CES2-pVL1392* vector was incubated with 1 µL of calf intestine alkaline phosphatase (NEB) at 37°C for one hour after which it was purified using the QIAquick PCR purification kit. The 300 bp insert (2.1 µl) and the dephosphorylated vector (2.2 µl) were incubated with Fast-link DNA ligase (Fastlink, Epicenter) for 15 minutes at room temperature. The ligation reaction was transformed into TOP10 *E. coli* which were plated on a LB plate containing 50 µg/mL ampicillin. Two colonies were amplified in 5 ml overnight cultures with 50 µg/mL ampicillin. From



the overnight cultures, 150 µl were used to inoculate 100 ml overnight cultures. Plasmids were isolated by midiprep (Qiagen). Restriction digests of the clones with Bgl II and Aar I yielded the expected digestion pattern. These clones were submitted for sequencing to the Biochemistry DNA Sequencing Core. The forward cloning primer and the reverse primer 5'-TGCAAATCGCAGCGGACCTAGA-3' were used for sequencing. Sequences of the clones were aligned with the expected *CES2*<sup>+64</sup> sequence using the DS gene software.

#### Production of the recombinant *CES2*<sup>+64</sup> virus

Two and one-half µg of the *CES2*<sup>+64</sup>-*pVL1392* vector were co-transfected with 0.5 µg of linearized Baculogold viral DNA (BD-Pharmingen) into  $2.5 \times 10^6$  *Sf9* insect cells following the manufacturer's protocol. One week post-transfection, the media containing the recombinant *CES2*<sup>+64</sup> virus was collected.

Plaque assays were performed to obtain a more virulent virus. Approximately  $6 \times 10^6$  *Sf9* insect cells were plated per 100-mm dish. The cells were infected with 1 ml viral dilutions ( $10^{-4}$ ,  $10^{-5}$ ,  $10^{-6}$ , and  $10^{-7}$ ), each in duplicate. After one hour, the media was removed and the cells were overlaid with a 1% agarose solution in TNM media plus 1x penicillin/streptomycin. The plates were incubated at 27°C for about 2 weeks until plaques were visualized. Eleven plaques were selected, and plaque plugs were collected through sterile pipetting. The plugs were added directly to separate wells of a 12-well plate seeded with *Sf9* cells at a density of  $3 \times 10^5$  cells/well. One well was left uninfected as a control. The plate was placed at 27°C for 12 days after which the media from each well was collected separately as passage 1 (P1).

Viral DNA for each sample was isolated from 750 µl of P1 stock. PCR was completed to confirm the presence of CES2<sup>+64</sup> viral DNA. PCR reactions were set up using the GeneAmp RNA PCR core kit (Applied Biosystems). Each twenty-five µl reaction contained 1x buffer, 2 mM MgCl<sub>2</sub>, 0.2 mM of each deoxynucleotide triphosphate, 0.3 µM of each primer (same as those used for sequencing), 12.5 U AmpliTaq, and 1 µl template DNA. For positive and negative DNA template controls, CES2<sup>+64</sup>-*pVL1392* and CES2-*pVL1392* were used, respectively. The reaction conditions were 95°C for 4 min, 5 cycles of 95°C for 30 s, 60°C for 30 s, and 72°C for 1.5 min, followed by 35 cycles of 95°C for 30 s and 72°C for 1.5 min. The PCR products were electrophoresed on an agarose gel. If the template DNA contained the sequence for CES2<sup>+64</sup>, the expected PCR product size was approximately 430 bp. Nine of the eleven viruses yielded the correct band. One recombinant virus containing the CES2<sup>+64</sup> DNA was selected for future use in the expression of CES2<sup>+64</sup>. One hundred µl of the selected P1 virus was used to infect 6x10<sup>6</sup> cells *Sf9* insect cells in T75 flasks. The cells were incubated at 27°C for 9 days. The media was collected for the passage 2 (P2) virus. A passage 3 (P3) viral stock was made by infecting 400 ml of *Sf9* cells seeded at 2x10<sup>6</sup> cells/ml with 8 ml of P2. P3 was collected after 9 days.

#### Purification of the CES2<sup>+64</sup> variant protein

The CES2<sup>+64</sup> variant protein was expressed and purified in the Protein Expression Core by Lan Min Zhai. A one liter culture of *Sf9* cells (~2x10<sup>6</sup> cells/ml) was each infected with 20 mL of P3 virus. Approximately 65 hours post-infection, the cultures were combined and centrifuged for 5 minutes at 500 *x g*. The media was collected and the cell pellet was stored for later use. The recombinant CES2<sup>+64</sup> protein was tagged with

a C-terminal His-tag, and nickel affinity resin was used in the purification procedures. The protein was eluted from the nickel resin with 60 mM imidazole. Protein was purified from the media fraction and the cell pellet, separately. The purified protein was buffered exchanged, and the final buffer was 20 mM Tris, pH 7.5 plus 1 mM benzamidine and 10% ethylene glycol. The amount of protein was quantified using the Protein the Bio-Rad Protein Assay with bovine serum albumin used for a standard curve. The purity of the protein, from the media and the cells, was analyzed by SDS-PAGE followed by Coomassie Blue staining.

#### 4-Methylumbelliferyl Acetate Hydrolysis Assay

Carboxylesterase activity was determined by 4-MUA hydrolysis as discussed previously. Briefly, 5 to 10  $\mu$ l of recombinant CES2<sup>+64</sup> protein, isolated from the media and cells, was incubated with 0.5 mM 4-methylumbelliferyl acetate in 90 mM potassium phosphate buffer with 40 mM KCl (pH 7.4) at 37°C. The rate of formation of 4-MU was monitored at 350 nm. Specific activity ( $\mu$ mol mg<sup>-1</sup> min<sup>-1</sup>) was calculated from the results of the 4-MUA assay and the Bio-Rad protein assay. Samples were analyzed in duplicate.

#### Analytical non-denaturing polyacrylamide gel electrophoresis

Purified CES2<sup>+64</sup> proteins were analyzed by analytic non-denaturing polyacrylamide gel electrophoresis as previously described by Dean et. al. (1995). Purified CES2<sup>+64</sup> protein (0.1-0.2  $\mu$ g) and CES2 protein (0.1  $\mu$ g) were loaded onto an 8% non-denaturing polyacrylamide gel. The samples were electrophoresed for 2 hours at 4°C after which the gel was incubated with 1 mM 4-MUA in 100mM potassium phosphate buffer pH 6.0 for 10 minutes. The Gel Doc 1000 system (BioRad) was used to capture the gel image. After imaging the gel was denatured and transferred to an

Immobilon-P membrane for western blotting. The western blotting procedure was the same as described below for the SDS-polyacrylamide gels. A separate non-denaturing PAGE was run followed by Coomassie blue staining to visualize the proteins.

#### SDS-Polyacrylamide gel electrophoresis and western blot analysis

Samples of CES2<sup>+64</sup> protein, purified from media and cells, were electrophoresed long 8% SDS polyacrylamide gels at 100 to 150 volts. To visualize the proteins, one gel was stained with Coomassie blue. The proteins from another gel were transferred onto Immobilon-P membrane through semi-dry transfer for one hour. The western blotting protocol was similar to that described in Methods Section II.

#### Glycosylation analysis of CES2<sup>+64</sup>

CES2<sup>+64</sup> protein, purified from the media and cells, was electrophoresed on an 8% non-denaturing polyacrylamide gel at 15 mA for 2 hours at 4°C. Purified CES2 was loaded on the gel as a control. The gel was denatured for 30 min in a 1% SDS solution with 20% MeOH and 2% β-ME. The proteins were transferred onto Immobilon-P membranes (Millipore) by semi-dry transfer for one hour. Transfer was detected with Ponceau S. The DIG Glycan Differentiation Kit and protocol (Roche) were used to determine the glycosylation state of the proteins. The membrane was blocked overnight at 4°C. All of the remaining steps were completed at room temperature. The membrane was washed two times with Tris buffered saline (TBS) and one time with TBS plus 1 mM MgCl<sub>2</sub>, 1 mM MnCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, pH 7.5 (Buffer 1) for 10 minutes each. The blot was incubated with a 1:1000 dilution of digoxigenin-labeled Galanthus nivalis agglutinin (GNA) in Buffer 1 for one hour. The incubation was followed by three 10 minute washes with TBS. The blot was then incubated with a 1:1000 dilution of anti-digoxigenin-AP in

TBS. After one hour, the blot was washed 3 times for 10 minutes with TBS.

Glycosylation was visualized with NBT/X-phosphate in 0.1 M Tris-HCl, 0.05 M MgCl<sub>2</sub>, 0.1 M NaCl, pH 9.5.

#### N-terminal Sequencing

Approximately 11 µg of CES2<sup>+64</sup> purified from media and 57 µg of CES2<sup>+64</sup> purified from cells were electrophoresed on 8% SDS-polyacrylamide and non-denaturing polyacrylamide gels, respectively. The non-denaturing gel was incubated with denaturing solution for 30 minutes followed by washes with water. Proteins from both gels were transferred onto Immobilon-P membranes through semi-dry transfer. The blots were placed in Coomassie (0.1% Brilliant Blue, 50% methanol, 10% acetic acid) for 10 minutes. The blots were destained with 50% methanol plus 10% acetic acid for 10 minutes followed by 10 quick rinses with water. The blots were stored at room temperature until further use.

Pham et al. (2005) described an on-membrane enzymatic cleavage method for use prior to N-terminal sequencing. Four desired protein bands were excised from the blots and placed into microfuge tubes. The excised bands of membrane were wet with methanol and rinsed ten times with water. The bands were blocked in a solution of 0.5% Zwittergent 3-16 (Sigma) in 0.1 M Tris, pH8 for 5 minutes, shaking at room temperature. The bands were then placed into fresh microfuge tubes and were washed 10 times with water. Each band was incubated with 1.97 mU *Pfu* Pyroglutamate Aminopeptidase (Takara) in 50 mM sodium phosphate with 10 mM DTT and 1 mM EDTA at 75°C for 2 hours. The bands were rinsed 10 times with water and allowed to dry. N-terminal sequencing was completed by Cambride Peptides (Birmingham, West Midlands, UK)

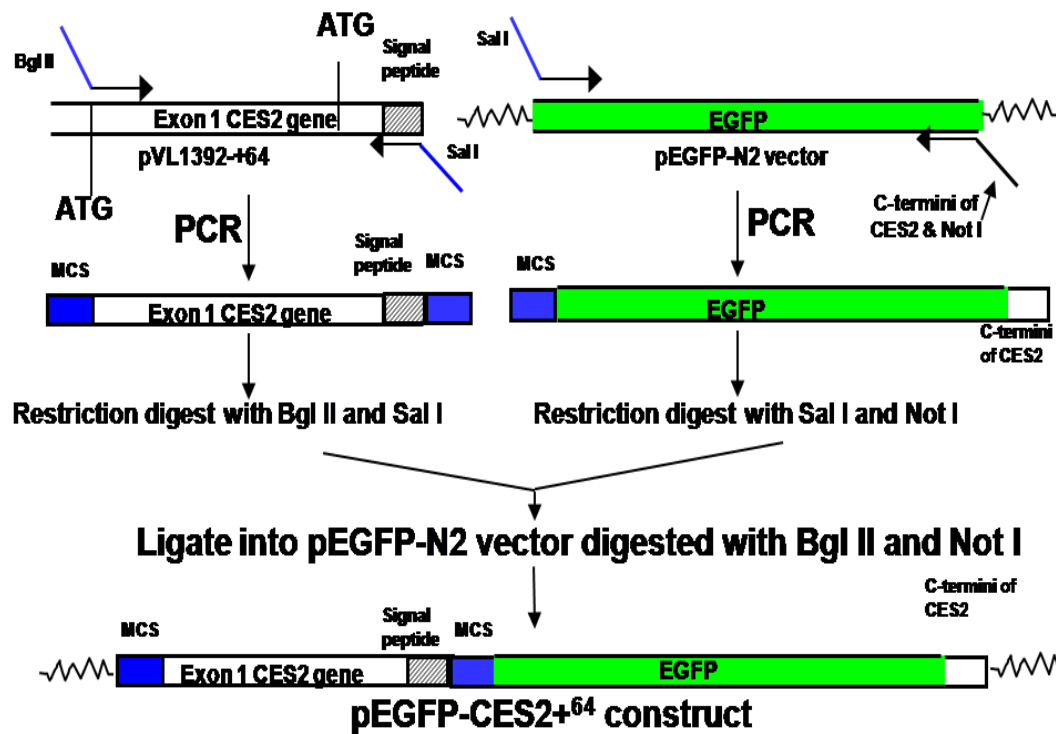
using the Edman degradation process. Each band was processed for five rounds of sequencing.

## VI. Sub-cellular localization of CES2 variants

### Cloning of *pEGFP-CES2<sup>+64</sup>*

Figure 7 outlines the cloning strategy for *pEGFP-CES2<sup>+64</sup>*. A chimera was designed to include the N-terminal portion of the variant including both ATGs and the signal peptide, EGFP, and the C-terminal ER retention signal of CES2. The N-terminal sequencing region of *CES2<sup>+64</sup>* was amplified using the forward primer 5'-ATCAGATCT *ATGACTGCTCAGTCCCG*-3' and the reverse primer 5'-GTCGACTCCGGATGGGACT *GGCTGAGTC*-3'. The *CES2<sup>+64</sup>* sequence is shown in italics. Bgl II and Sal I sites (underlined) have been added to the forward and reverse primers, respectively. The forward primer was previously used in cloning *CES2<sup>+64</sup>* into the pVL1392 vector. This clone was used as template for the N-terminal amplification. *EGFP* was amplified using the forward primer 5'-AGCTCAAGCTTCTGAATTC TGCAGTCGACGG-3' and the reverse primer 5'-GCGGCCGCCTACAGCTCTGTGTGCTTGTAC-3' and pEGFP-N2 as the template. The forward primer contains a Sal I site (underlined). The reverse primer introduced a Not I site (underlined) and the HTEL C-term of CES2 (italicized). Both of the PCR products were cloned into an *E. coli* replication vector by blunt-end ligation followed by transformation into TOP10 *E. coli*. Colonies were selected for overnight culture. The recombinant vectors were isolated by Qiagen mini-prep and then digested with the appropriate enzymes. pEGFP-N2 was digested with Bgl II and Not I followed by dephosphorylation. The three pieces (N-term *CES2<sup>+64</sup>*, EGFP+HTEL, and the linearized pEGFP-N2 vector) were then ligated and transformed into TOP 10 *E. coli*.

Clones were selected and grown for mini-preps. The DNA construct was then confirmed by sequencing.



**Figure 7. Strategy for cloning the *pEGFP-CES2<sup>+64</sup>* construct:** The N- and C-termini of *CES2<sup>+64</sup>* are added to the N- and C-termini of EGFP, respectively.

### Cloning of *pEGFP-CES2*

A *pEGFP-CES2* was constructed in a similar manner as *pEGFP-CES2*<sup>+64</sup>. Briefly, the N-terminal coding sequence of *CES2* was amplified with the forward primer 5'-AGATCT***ATGCGGCTGCACAGACTTCGTGCGCGGCTG***-3' and the same reverse primer used to amplify the N-terminus of *CES2*<sup>+64</sup>. The forward primer which was used to clone *CES2* into the pVL1392 vector contains a Bgl II restriction enzyme cut site (underlined) and the *CES2* ATG and sequence (bold, italicized). The PCR product was then blunt-end ligated and digested as stated above. The PCR for the EGFP-HTEL construct did not have to be repeated. The *pEGFP-CES2*<sup>+64</sup> clone was digested with Bgl II and Sal I as was the *CES2* N-term in the TOPO vector. The *CES2* N-term sequence was ligated into the digested pEGFP-HTEL vector to make the final construct *pEGFP-CES2*.

### Cloning of *pEGFP- CES2* <sup>$\Delta 1-93$</sup>

Because the N-terminal sequence of *CES2* <sup>$\Delta 1-93$</sup>  is not unique to this protein and is not believed to contain any signal sequences, the entire *CES2* <sup>$\Delta 1-93$</sup>  sequence needed to be cloned into the pEGFP vector. The forward primer which was used to clone *CES2* <sup>$\Delta 1-93$</sup>  into the pVL1392 vector was 5'-TGGAAAGATCT***ATGTGTCTACAGGACC***-3' and the reverse primer was 5'-GTCGACTTCTCTCTTCAGGCTCCTCG-3'. The forward primer contained a Bgl II site (underlined), and the reverse primer introduced a Sal I site (underlined). *pEGFP- CES2* <sup>$\Delta 1-93$</sup>  was made in a similar manner to *pEGFP- CES2* (above). The construction of all *pEGFP-CES2 variant* plasmids was completed by Lan Min Zhai in the Protein Expression Core.

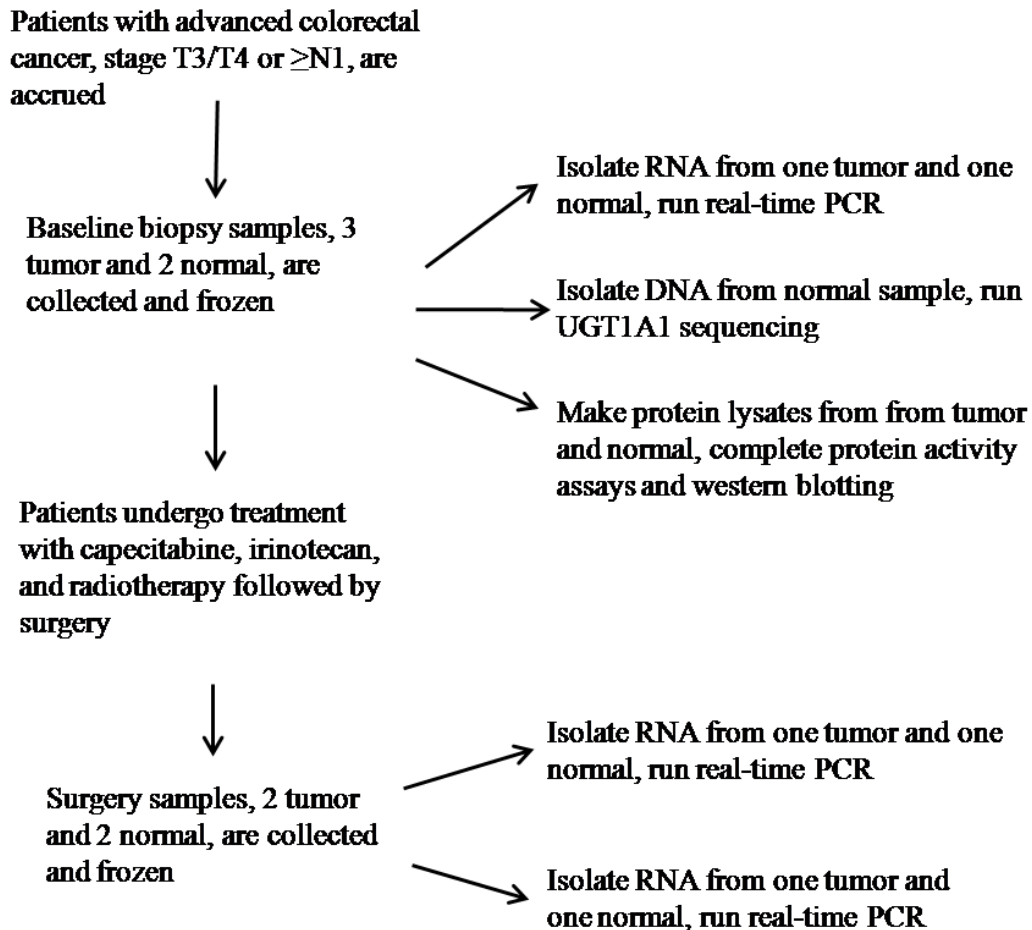


### Transfection of GFP constructs

HCT-15 cells were seeded at  $\sim 3 \times 10^5$  cells per 2 ml on poly-L-lysine treated cover slips placed in 35mm dishes. The cells were incubated at 37°C with 5% CO<sub>2</sub> for one to two days. Transfections were completed using the FuGene6 reagent (Roche) in a 3:1 ratio with 2 µg of DNA. Six µl of FuGene were added to a volume of Opti-MEM media (Invitrogen) that yielded a final volume of 100 µl after the addition of DNA. The FuGene-media mixture was incubated at room temperature for 5 min. A volume of DNA equivalent to 2 µg was added, and the new mixture was incubated for 20 min at room temperature. After removing the media, the transfection solution was added to the cells on the cover slip. The cells were incubated at 37°C, 5% CO<sub>2</sub> for approximately 5-10 minutes after which 2 mls of fresh RPMI (Cambrex) media with 10% FBS and 1% pen/strep was added. The production and localization of the chimeric GFP proteins was followed over a 24-48 hour period using the Axiovert 200 microscope (Zeiss). The GFP transfections and the immunostaining assays were completed by Sharry Fears.

### **VII. The role of CES2, CES1, TOPO I, TP, TS, DPD, β-GUS, and UGT1A1 in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine**

Patients were evaluated and treated as per the HOG GI03-53 protocol (outlined briefly in Figure 8). Pre-treatment biopsy samples and post-treatment surgery samples of normal and tumor tissue were collected and immediately frozen. All specimens were reviewed by a pathologist. The biopsy samples ranged from  $\sim 2$  to 20 mg while the surgery samples were significantly larger. RNA and genomic DNA were isolated from eleven biopsy samples for analysis. RNA was collected from eleven biopsy samples in



**Figure 8. Outline of the strategy for rectal samples collected for the GI03-53 study**

the manner described in Methods section III. Genomic DNA as well as the total RNA was collected from either a biopsy sample or corresponding surgery sample for each of the eleven patients using the Allprep DNA/RNA kit (Qiagen). All tissue samples were stored at -70°C.

#### Clinical methods and data collection

All clinical data were collected by other members of the HOG GI03-53 team. Clinical endpoints included tumor response and toxicity. After receiving their course of chemotherapy, patients were noted as having a pathological complete response (pCR) or a pathological non-complete response (pNCR).

#### Reverse transcription and real-time PCR

Methods similar to those described in Methods section III were employed. Reaction volumes at times differed from Methods section III; however, concentrations remained the same. RNA quality was not assessed by gel electrophoresis. Real-time PCR was completed for *CES2*, *CES1*, topoisomerase I (*TOPO I*), thymidylate synthase (*TS*), thymidine phosphorylase (*TP*),  $\beta$ -glucuronidase ( $\beta$ -*GUS*), and dihydropyrimidine dehydrogenase (*DPD*). Clones were constructed for each of these genes for the use in creating standard curves. The primers used for cloning (excluding *CES1* and *CES2*) and real-time PCR are listed in Table 4. All PCR products were checked by sequencing. The product sizes and plasmids are listed in Table 5.

Gene	Primers	Conc. primer (μM)	Melting time at 95°C (sec)	Annealing time at 65°C (sec)	Extention time at 72°C (sec)
<i>CES2</i>	F 5'-CCATGGTGATGAGCTTCCTTTTGT-3' R 5'-AGGTATTGCTCCTCCTGGTCGAA-3'	0.5	30	30	60
<i>CES1</i>	F 5'- AGAGGAGCTCTTGGAGACGACAT-3' R 5'- ACTCCTGCTTGTTAATCCGACC-3'	0.2	30	30	60
<i>TOPO1</i>	F 5'-CGTTCTACCAGGCAAATTCAGTGT-3' R 5'-TGAAATGGGAGAGAGGGAAGGGA-3'	0.3	20	15	40
<i>β-GUS</i>	F 5'-TCAACAAGCATGAGGATGCGGAC-3' R 5'-TACGCACCACTTCTTCCATCACC-3'	0.3	30	30	60
<i>TP</i>	F 5'-AATGTCATCCAGAGCCCAGAGCA-3' R 5'-GAACTTAACGTCCACCACCAGAG-3'	0.5	30	30	60
<i>TS</i>	F 5'-TTTACCTGAATCACATCGAGCCAC-3' R 5'-GACTGACAATATCCTTCAAGCTCC-3'	0.5	30	30	20
<i>DPD</i>	F 5'-GGTCTTCAGTTTCTCCATAGTGGT-3' R 5'-GACTCTGTCCATCCCAGTCTTGT-3'	0.5	20	20	45

**Table 4. Forward (F) and reverse (R) primers for real-time PCR:** Reaction conditions were 50°C for 2 min, 95°C for 10 min, followed by 40 cycles PCR with the temperature and times listed in the table.

Gene	Vector	Insert size (bp)	Vector size (bp)	Total size (bp)
<i>CES2</i>	pVL1392	1680	9639	11319
<i>CES1</i>	pCR Blunt TOPO II	1707	3519	5226
<i>TOPO I</i>	pCR4-TOPO	208	3957	4165
<i><math>\beta</math>-GUS</i>	pCR Blunt TOPO II	264	3519	3783
<i>TP</i>	pCR Blunt TOPO II	216	3519	3735
<i>TS</i>	pCR Blunt TOPO II	209	3519	3728
<i>DPD</i>	pCR Blunt TOPO II	157	3519	3676

**Table 5. Plasmids used for standard curves in real-time PCR:** Reported insert size corresponds to PCR product size except in the cases of *CES2* and *CES1*. Expected product size for *CES2* and *CES1* are 0.15 kb and 0.19 kb, respectively.

### UGT1A1 sequencing

An approximately 255 bp region surrounding the promoter region of the *UGT1A1* gene was amplified by PCR. As per Monaghan et al. (1996), the forward primer was 5'-AAGTGAAGTCCCTGCTACCTT-3' and the reverse primer was 5'-CCACTGGGATCAACAGTATCT-3'. Five hundred ng of genomic DNA from either a normal biopsy or surgery sample were used as the template in a 50 µl PCR reaction. The reactions contained 1x buffer, 1.75mM MgCl<sub>2</sub>, 0.8mM of each deoxynucleotide triphosphate, 0.25 µM of each primer, and 0.05U/ µL AmpliTaq Gold (Applied Biosystems). The PCR conditions were 95°C for 10 minutes followed by 30 cycles of 95°C for 30 s, 58°C for 40 s, and 72°C for 40 s (Monaghan et al., 1996). The PCR products were purified by electrophoresis on 2% agarose (Sigma) gels followed by extraction using the QIAquick Gel Extraction Kit. The purified products were sequenced using the forward primer, and the chromatograms were analyzed for TA repeats in the TATA box region of the promoter.

### Clinical data and statistical analysis

All clinical data were collected by members of the HOG GI03-53 team. Clinical endpoints included tumor response and toxicity. Correlation analysis of gene expression with clinical outcome was completed by statisticians involved with the HOG GI03-53 project. The data were considered significant if  $p < 0.05$ .

## RESULTS

### I. Tissue-specific expression of *CES2* splice variants

#### Northern analysis

A multi-tissue northern blot was probed with cDNA specific for *CES2*<sup>+64</sup> and *CES2*<sup>Δ1-93</sup> (Figure 9). Wu et al. (2003) previously analyzed the blot for expression of *CES2*. The *CES2*<sup>+64</sup> sequence was identified in the 4.1 kb and 2.8 kb transcripts. A < 1.7 kb transcript was also present in the liver and heart. The *CES2*<sup>Δ1-93</sup> sequence was localized to the approximately 4 kb transcript.

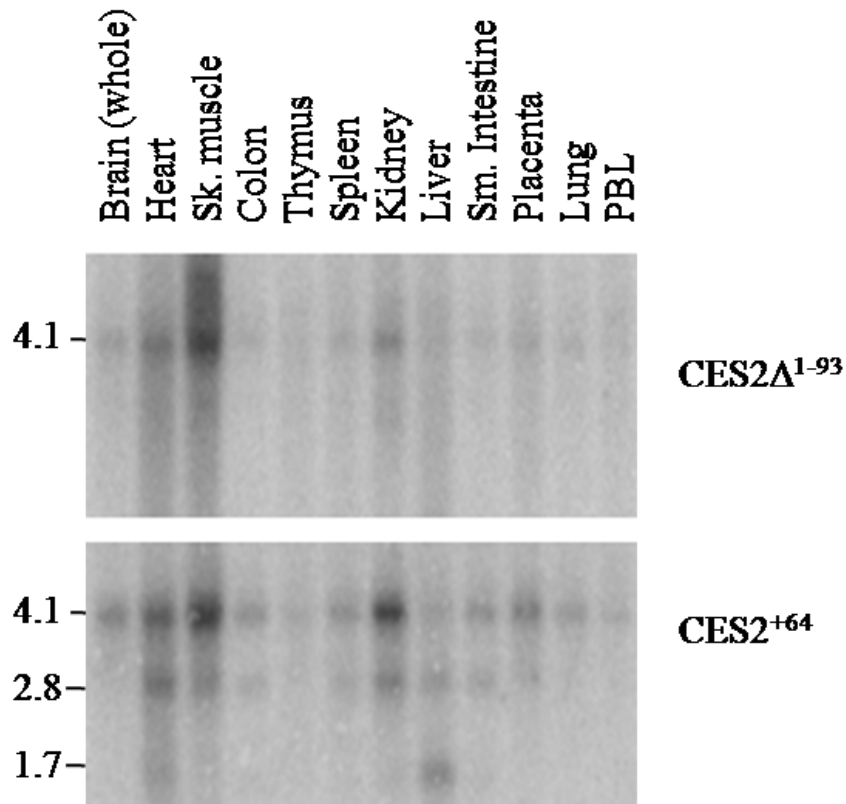


Figure 9. Northern analysis of *CES2*<sup>Δ1-93</sup> and *CES2*<sup>+64</sup>

## II. Analysis of *CES2* and *CES2*<sup>Δ458-473</sup> in paired tumor and normal colon samples

### Expression of *CES2* and *CES2*<sup>Δ458-473</sup> in human colon samples

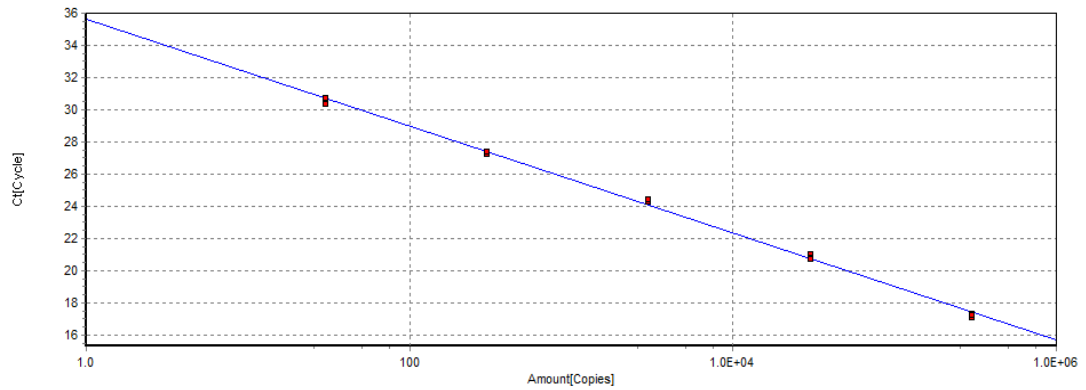
Real-time PCR was used to quantify the expression levels of the *CES2* and *CES2*<sup>Δ458-473</sup> transcripts in 10 paired tumor and normal colon samples. Forward primers (shown in bold in Figure 10) were designed to be variant-specific for *CES2* and *CES2*<sup>Δ458-473</sup>. Analysis of RNA quality by agarose gel electrophoresis showed little to no degradation. Standard curves were created for both variants using DNA clones in the pVL1392 vector which were constructed by our laboratory. Amplification of each standard curve was linear over a 100,000-fold range ( $r^2=0.99$ ) with respect to copy number (Figure 11). The amplification efficiency, as calculated by the eppendorf real-time PCR software, was  $\geq 0.9$  for each of the variants. The uniformity of the PCR products was determined using melting curve analysis (Figure 12). All cDNA samples were analyzed in triplicate with the standard deviation generally  $< 12\%$  for both *CES2* and *CES2*<sup>Δ458-473</sup>.

The reproducibility of all the steps involved in real-time PCR was studied in control experiments. In one experiment, RNA was isolated from five different pieces of the same tumor sample. It was determined that  $< 1.6$ -fold variation in the *CES2* copy number could be attributed to the entire method from tissue disruption through real-time PCR (Figure 13A). In another experiment, one sample RNA was reverse transcribed in quintuplicate. It was determined that  $< 1.3$ -fold variation in copy number could be attributed to reverse transcription (Figure 13B).

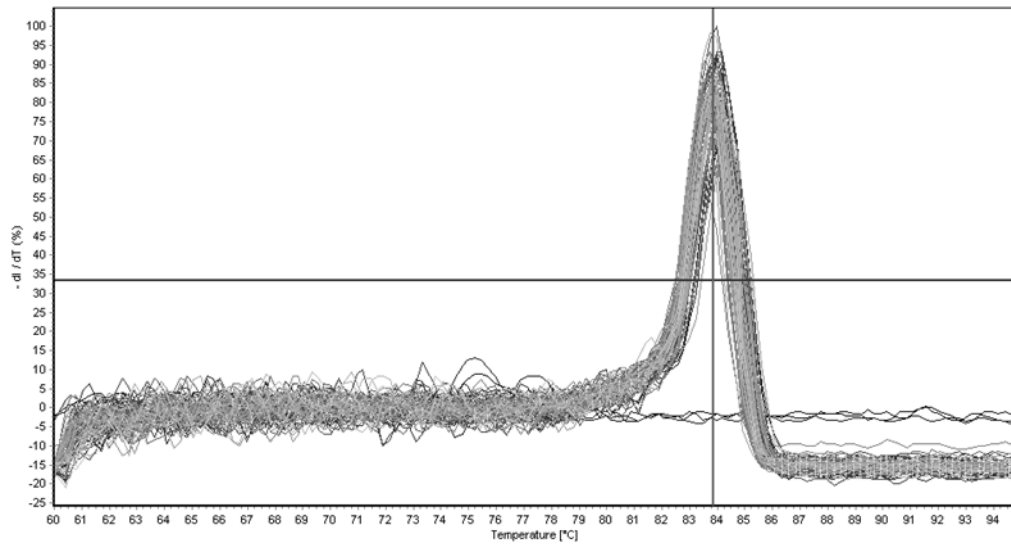




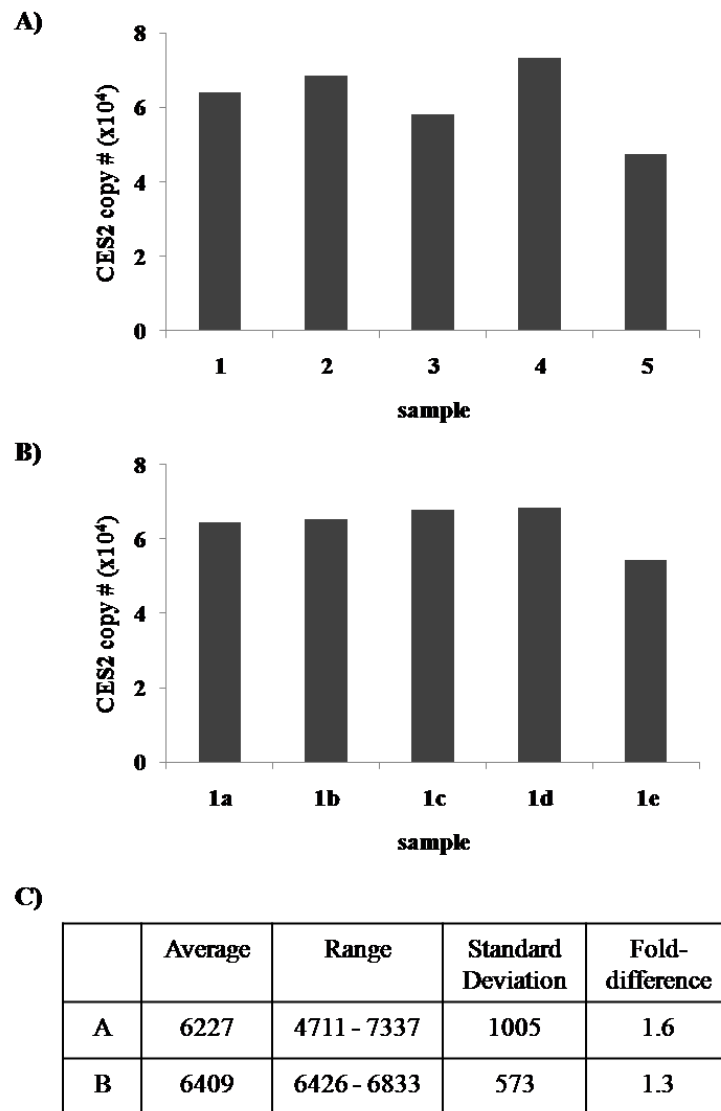
**Figure 10. Alternative splicing in exon 10:** Alternative splicing in exon 10: (A) shows a partial DNA sequence for *CES2* gene. The exon 10 sequence is shaded. Alternative use of 5' splice site in exon 10 results in *CES2*<sup>Δ458-473</sup> variant which has a 48 bp deletion (boxed) in comparison to *CES2*. Variant specific-primer sequences are shown in bold. (B) The 16 amino acid deletion immediately following the active site H457 (bold) in the *CES2*<sup>Δ458-473</sup> protein results from the 48 bp deletion in its transcript shown in Panel A. Unique peptides that were identified for *CES2* and *CES2*<sup>Δ458-473</sup> by LC-MS/MS are underlined (Schiel et al., 2007).



**Figure 11. Real-time PCR standard curve for *CES2*<sup>Δ458-473</sup>:** The standard curve for *CES2*<sup>Δ458-473</sup> was linear over a 100,000-fold range ( $r^2=0.998$ ). The efficiency of the PCR was 1.00. (The x-axis and y-axis are copy number and cycle number, respectively.)



**Figure 12. Melt curve analysis for *CES2*<sup>Δ458-473</sup> real-time PCR products:** All real-time PCR products for *CES2*<sup>Δ458-473</sup> had an average melting temperature of 83.90°C indicating a single uniform product. The three lines with no peaks represent the no template control samples. (The x-axis and y-axis are temperature in °C and  $-dI/dT$  (%), respectively.)

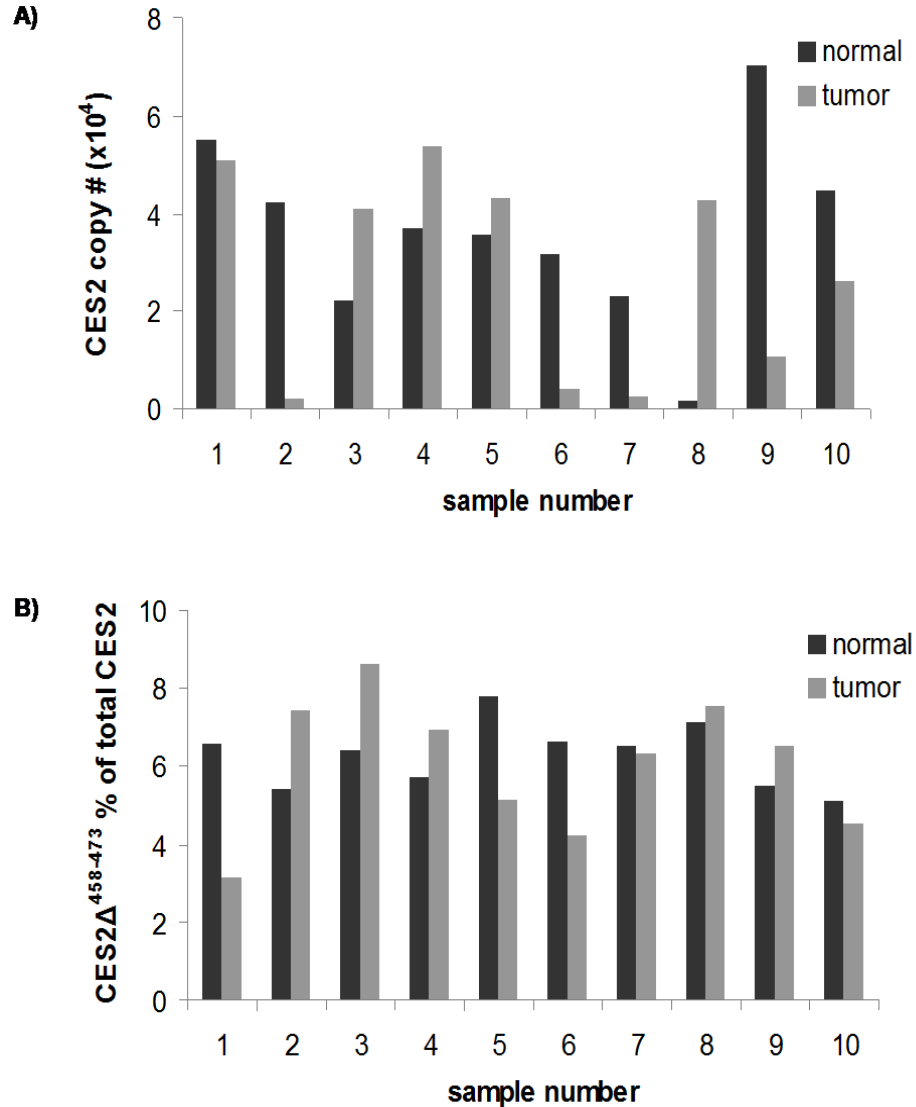


**Figure 13. Reproducibility of real-time PCR methods:** The reproducibility of all the steps involved in real-time PCR was analyzed. (A) RNA from sample 1 was reverse transcribed in quintuplicate. Less than a 1.3-fold variation in copy number was attributed to reverse transcription alone. (B) RNA was isolated from five different pieces (1-5) of the same esophageal tumor sample. A <1.6-fold variation in the *CES2* copy numbers was attributed to the entire method from tissue disruption through real-time PCR. (C) Statistical summary of panels (A) and (B).

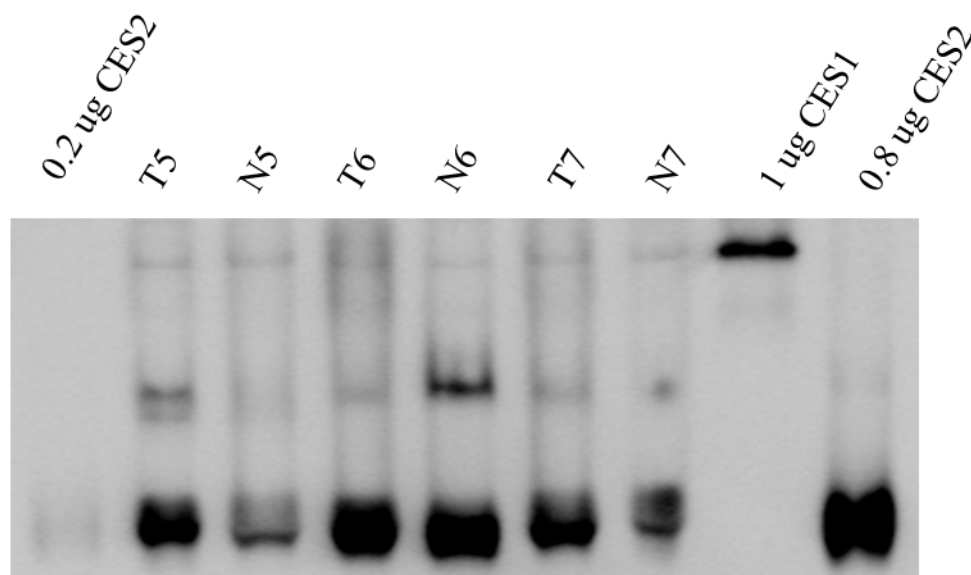
*CES2* and *CES2*<sup>Δ458-473</sup> transcripts were found in all tumor and normal colon samples. There was a 45-fold variation in expression of *CES2* transcript in normal tissue and a 27-fold variation in tumor tissue. Among four of the paired samples, *CES2* was expressed in greater abundance in tumor as compared to normal tissue (Figure 14A). No significant linear correlation ( $R^2=0.025$ ) was found in *CES2* transcript expression between the paired tumor and normal tissue samples. *CES2*<sup>Δ458-473</sup> varied by 34-fold in normal tissue and 25-fold in the tumor tissues. The percentage of *CES2*<sup>Δ458-473</sup> transcript in colon tissue accounts for approximately 6% (SD < 1.3%) of the total *CES2* transcripts (Figure 14B).

#### Carboxylesterase activity in colon tissue lysates

The carboxylesterase activities of the colon tissue lysates were quantified using the 4-methylumbelliferyl acetate hydrolysis (4-MUA) assay and band densitometry analysis as described in the methods. All samples showed non-specific carboxylesterase activity with 4-MUA as substrate (Table 6). The range for normal samples was 2.5 to 9.8  $\mu\text{mol mg}^{-1} \text{h}^{-1}$  with a 4-fold variation, while the range for the tumor samples was 2.0-7.4  $\mu\text{mol mg}^{-1} \text{h}^{-1}$  with a 3.7-fold variation. Non-denaturing PAGE was used to separate the carboxylesterases in the samples and purified CES2 protein was loaded on each gel as a control. The gel for paired colon tissue samples 5 through 7 is shown in Figure 15. The band density attributable to CES2 was determined for each sample. All samples demonstrated activity for 4-MUA in the non-denaturing PAGE assay (Table 6). There was a 24-fold variation in normalized band density among the normal colon samples and an 8-fold variation among the tumor samples.



**Figure 14. Expression of *CES2* and *CES2*<sup>458-473</sup> in 10 paired tumor and normal colon samples:** (A) Real-time PCR was used to determine the amount of *CES2* transcripts in 10 paired tumor and normal colon samples (Methods III). A 45-fold variation in expression of *CES2* was found among the normal samples, and a 27-fold variation was found for the tumor samples. There was no disease-specific expression pattern. (B) *CES2*<sup>458-473</sup> was expressed in all of the colon tissues samples. *CES2*<sup>458-473</sup> accounts for ~6% of total *CES2* transcript in all of the colon samples (Schiel et al., 2007).



**Figure 15. Non-denaturing polyacrylamide active gel for paired colon tissue samples:** Colon tissue lysates from paired normal (N) and tumor (T) tissue, as well as control samples of CES1 and CES2, were electrophoresed on a non-denaturing gel. The gel was incubated with 4-MUA and the product 4-MU was detected under UV light. Band densinometry analysis was completed and reported in Table 6. The middle band has been previously observed but is not yet identified.

sample	4-MUA activity	band density	sample	4-MUA activity	band density
N1	7.1	0.64	T1	5.8	0.65
N2	6.6	0.74	T2	3.1	0.12
N3	4.1	0.31	T3	5.5	0.51
N4	8.0	0.78	T4	7.4	0.71
N5	3.8	0.32	T5	7.2	0.57
N6	9.8	0.55	T6	3.2	0.14
N7	6.2	0.44	T7	2.0	0.09
N8	2.5	0.03	T8	7.4	0.53
N9	8.6	0.50	T9	2.2	0.12
N10	6.7	0.44	T10	4.4	0.33
Ave	6.3	0.48	Ave	4.8	0.38
SD	2.2	0.22	SD	2.1	0.24

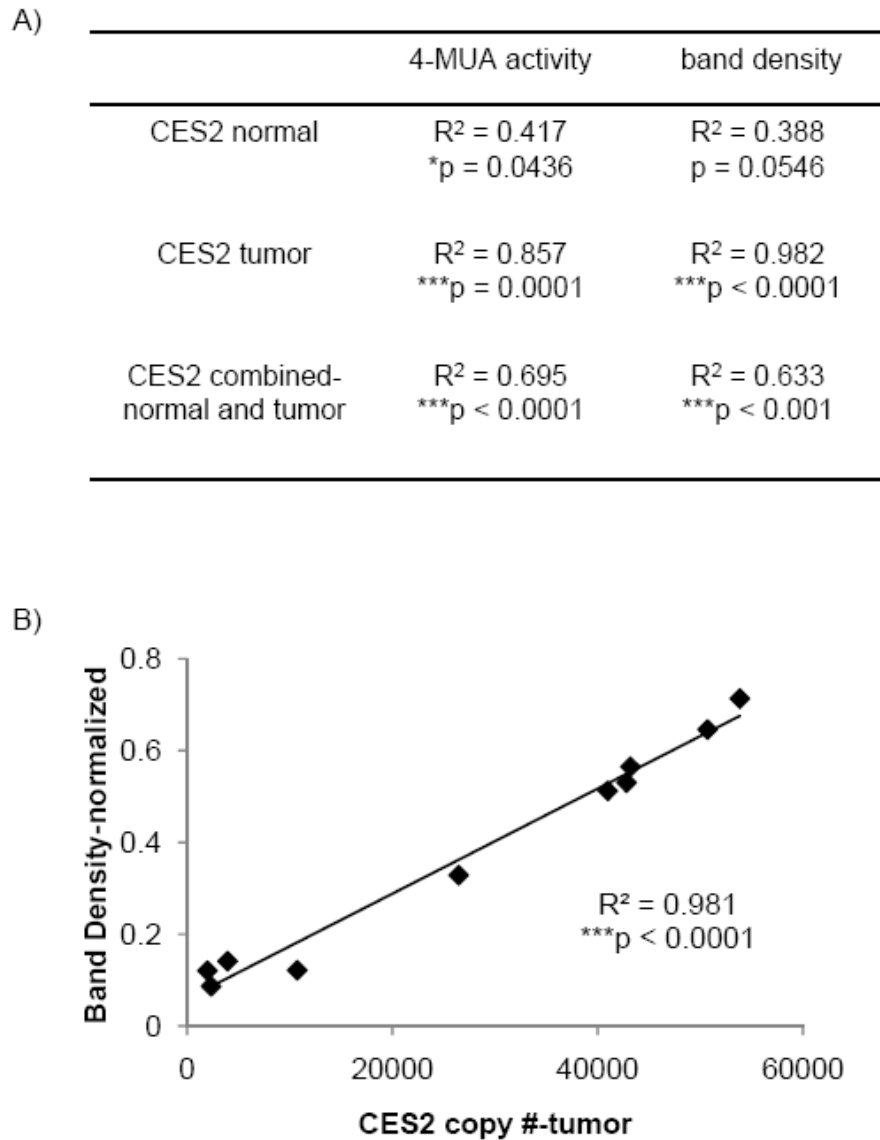
**Table 6. Expression and activity data for paired tumor (T) and normal (N) colon tissue samples:** Ten microliters of colon tissue lysate were incubated with 0.5 mM 4-MUA in 90 mM potassium phosphate containing 40 mM potassium chloride buffer (pH 7.4) at 37°C. Specific activity for 4-MUA was calculated in  $\mu\text{mol mg}^{-1} \text{h}^{-1}$ . CES2 band density was determined by running 50-100  $\mu\text{g}$  of colon tissues lysates on non-denaturing PAGE. The gels were exposed to 1 mM 4-MUA in 100 mM potassium phosphate buffer (pH 6.0) for 10 minutes at room temperature. Band density was measured by fluorescence. Band densities were normalized to CES2 controls in order to account for variations between gels. Averages (Ave) and standard deviations (SD) are included (Schiel et al., 2007).

### Statistical analysis

There was evidence of kurtosis (data not shown) for both the *CES2* copy numbers (p-value=0.03) and band densities (p-value=0.04) of the tumor samples. Kurtosis is often an indication of bimodality (Darlington R.B., 2007). All other measurements in tumor and all measurements in normal tissue appeared to be normally distributed. *CES2* copy numbers, determined by real-time PCR, were subjected to linear regression analysis (JMP 4.0) with 4-MUA hydrolase activity and normalized band density (Figure 16).

Correlations of *CES2* expression with 4-MUA hydrolase activity and band density are considered significant if  $p \leq 0.05$  (Takahashi et al., 2000). *CES2* copy numbers, in tumor samples, positively and significantly ( $p \leq 0.0001$ ) correlated with both 4-MUA hydrolase activity ( $r^2 = 0.857$ ) and band density ( $r^2 = 0.982$ ). For the normal samples, *CES2* copy numbers significantly correlated with 4-MUA hydrolase activity ( $p < 0.0436$ ), but correlation with band density ( $p < 0.0546$ ) was just outside the range of significance. Data from the normal and tumor samples were analyzed collectively, and significant correlation existed between *CES2* copy numbers and both 4-MUA ( $p < 0.0001$ ) and band density ( $p < 0.001$ ).





**Figure 16. Correlation of *CES2* expression with carboxylesterase activity in colon tissue:** (A) Real-time PCR data for *CES2* was subject to linear regression (JMP 4.0) with 4-MUA hydrolase activity ( $\mu\text{mol mg}^{-1} \text{h}^{-1}$ ) and normalized band density from corresponding samples. The correlation coefficients and P values are reported. Data are considered significant if  $p \leq 0.05$ . (B) Scatter plot showing significant, positive correlation between *CES2* copy number and band density in tumor samples (Schiel et al., 2007).

### III. Characterization of the CES2<sup>Δ458-473</sup> variant

CES2 is a serine ester hydrolase (carboxylesterase) with Ser<sup>228</sup>, Glu<sup>345</sup> and His<sup>457</sup> forming its catalytic triad. CES2 is the major carboxylesterase responsible for activation of CPT-11 (Khanna et al., 2000; Humerickhouse et al., 2000; Sanghani et al., 2004). A variant DNA sequence for *CES2* was reported in the EST database. This variant transcript has a 48 bp deletion (Figure 10A) that gives rise to a protein with a sixteen amino acid deletion immediately following the active site His<sup>457</sup> residue (Figure 10B). Use of the SPIDEY software (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html>) to align the *CES2* (GI:37622884) and *CES2*<sup>Δ458-473</sup> (GI:37622886) mRNA transcripts with the *CES2* genomic clone (GI:56788327) demonstrates that the exon 10 sequence contains good 5' splice sites for both of the transcripts. Hence, the two transcripts are created by the use of alternate 5' donor splice sites in exon 10. This evidence is further supported by the identification of 14 EST sequences in the human EST database as determined by blasting the database with a 25-nucleotide sequence unique to *CES2*<sup>Δ458-473</sup> (shown in bold in Figure 10).

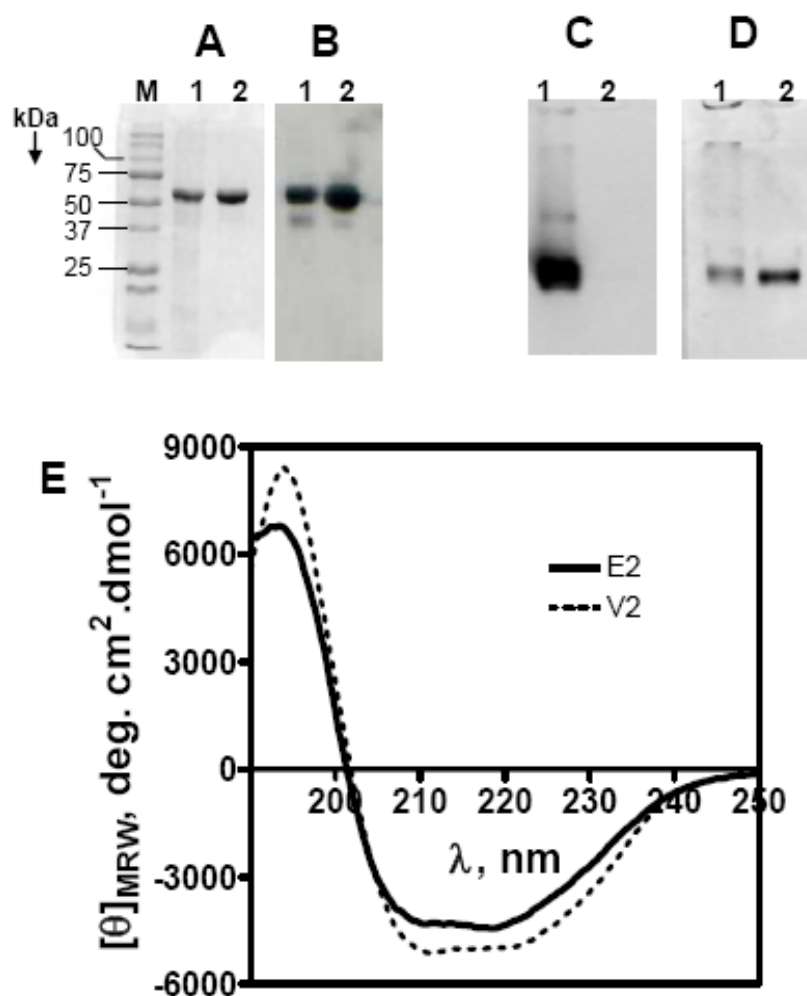
#### Expression and purification of the CES2<sup>Δ458-473</sup> protein variant

A partial EST clone for *CES2*<sup>Δ458-473</sup> was obtained from ATCC (Image Clone ID 6195662). In the *CES2*<sup>Δ458-473</sup> message, the final 48bp of exon 10 are removed by alternative splicing. The partial *CES2*<sup>Δ458-473</sup> clone and *CES2* in the pVL1392 vector were digested with Xba I and Sma I. The *CES2*<sup>Δ458-473</sup> Xba I/Sma I digestion fragment was ligated into the digested *pVL139-CES2* vector introducing the 48 bp deletion. Restriction digests and sequencing confirmed the selection of the correct clones. A recombinant

baculovirus containing *CES2*<sup>Δ458-473</sup> cDNA was used to express the recombinant *CES2*<sup>Δ458-473</sup> protein in *Sf9* insect cells. *CES2*<sup>Δ458-473</sup> was purified in a two-step protocol using concanavalin A (conA) affinity columns followed by preparative non-denaturing PAGE (Sun et al., 2004). Due to its lack of 4-MUA hydrolase activity, during each purification step the protein elution profile of *CES2*<sup>Δ458-473</sup> was compared with *CES2* to identify the protein peak of interest. The protein elution profile for *CES2*<sup>Δ458-473</sup> was similar to *CES2* for both purification steps.

#### Analysis of the *CES2*<sup>Δ458-473</sup> protein

The purified *CES2* and *CES2*<sup>Δ458-473</sup> proteins were analyzed by SDS-PAGE (Figure 17A), western analysis (Figure 17B), non-denaturing polyacrylamide gel electrophoresis (Figure 17C and 5D), circular dichroism (CD) (Figure 17) and mass spectrometry to assess their identity, purity, and physical properties. The calculated molecular weight of *CES2*<sup>Δ458-473</sup> protein from the SDS-PAGE gel (Figure 17A) is 59 kDa and the expected molecular weight is 57.3 kDa before glycosylation. In the western blot analysis, both *CES2* and *CES2*<sup>Δ458-473</sup> were identified with an anti-*CES2* antibody (Figure 17B). Activity staining of the non-denaturing polyacrylamide gel showed that the *CES2*<sup>Δ458-473</sup> protein lacks carboxylesterase activity for 4-MUA (Figure 17C). Coomassie blue stain of the same gel shows the protein in the *CES2*<sup>Δ458-473</sup> lane (Lane 2, Figure 17D) migrating to the same place as *CES2* (Lane 1, Figure 17D). The identity of the recombinant *CES2*<sup>Δ458-473</sup> protein was confirmed by mass spectrometry where

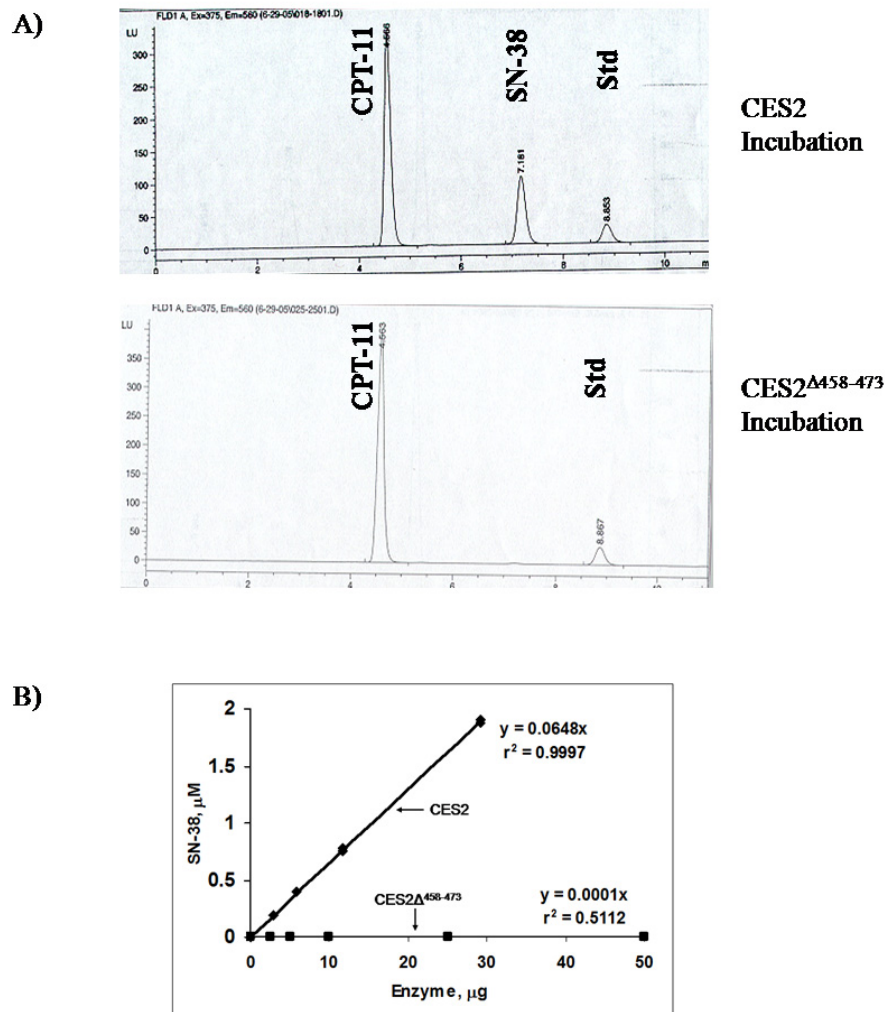


**Figure 17. Characterizations of recombinant CES2<sup>Δ458-473</sup> and CES2 proteins:** Lane 1: Two  $\mu$ g of CES2 and Lane 2: Two  $\mu$ g of CES2<sup>Δ458-473</sup> (A) SDS-PAGE stained with Coomassie blue (B) Western blot of the same gel using CES2 antibody (C) Non-denaturing PAGE stained for activity with 4-MUA (D) Non-denaturing PAGE from panel C stained for protein with Coomassie blue (E) CD spectrum of CES2 (E2) and CES2<sup>Δ458-473</sup> (V2) performed under identical conditions (Schiel et al., 2007).

a tryptic peptide with one miscleavage, ADHVKFTEEEEQLSR, specific for CES2<sup>Δ458-473</sup> (underlined in Figure 17B) was identified. Folding of CES2<sup>Δ458-473</sup> protein was evaluated by circular dichroism. The mean residue molar ellipticity was plotted as a function of wavelength (Figure 17E). The secondary structure composition analysis was done with CDpro software. Results from the CONTIN/LL method showed that the compositions of CES2 and CES2<sup>Δ458-473</sup> were 15 % α-helical, 34 % β-strand, 22 % turn, 29 % unordered and 20 % α-helix, 29 % β-strand, 22 % turn, 29 % unordered, respectively.

#### CPT-11 hydrolase assay

CES2<sup>Δ458-473</sup> protein (0-50 μg) was incubated with 55 μM CPT-11 for 2 h to determine its CPT-11 hydrolase activity. Under these conditions, no significant SN-38 (< 5 nM) peak was detected in any of the CES2<sup>Δ458-473</sup> samples after 2 hours. In comparison, 1.98 μM of SN-38 was detected when CES2 (30 μg) was incubated with the same concentration of CPT-11 for just 30 min (Figure 18).



**Figure 18. CPT-11 hydrolysis by CES2 and CES2<sup>Δ458-473</sup>:** Varying amounts of purified, recombinant CES2 and CES2<sup>Δ458-473</sup> proteins were incubated with 55 $\mu\text{M}$  CPT-11. The reactions were carried out in 250  $\mu\text{l}$  volume at 37°C. CES2 samples were incubated for 30 min and CES2<sup>Δ458-473</sup> samples for 2h. SN-38 formation was quantified by HPLC as described in material and methods. (A) HPLC chromatograms are shown. (B) SN-38 concentration was plotted as a function of protein concentration (Schiel et al., 2007).

#### IV. Characterization of the CES2<sup>+64</sup> variant

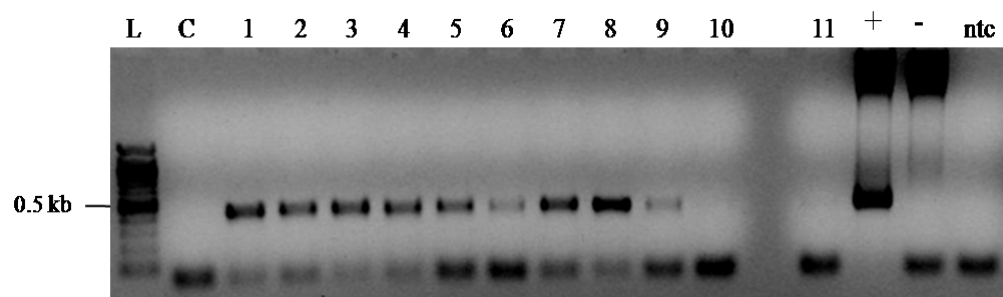
##### Cloning, expression, and purification of the CES2<sup>+64</sup> variant

The CES2<sup>+64</sup> N-terminal coding sequence was amplified by PCR using human liver cDNA (BD Biosciences Clontech) as the template. The forward primer introduced a Bgl II site immediately before the first ATG of CES2<sup>+64</sup>, and the PCR product included a natural Aar I site. The 400 bp PCR product was purified and cloned into the pCR Blunt II-TOPO vector (Invitrogen). The recombinant vector was transformed into TOP10 *E. coli* (Invitrogen). Positively transformed colonies were selected for on LB plates with 50µg/ml kanamycin. To amplify the recombinant vector, four colonies were cultured overnight. Recombinant vectors were isolated by miniprep (Qiagen) and digested with EcoRI. All digested clones yielded the expected 400 bp band on a 1% agarose gel.

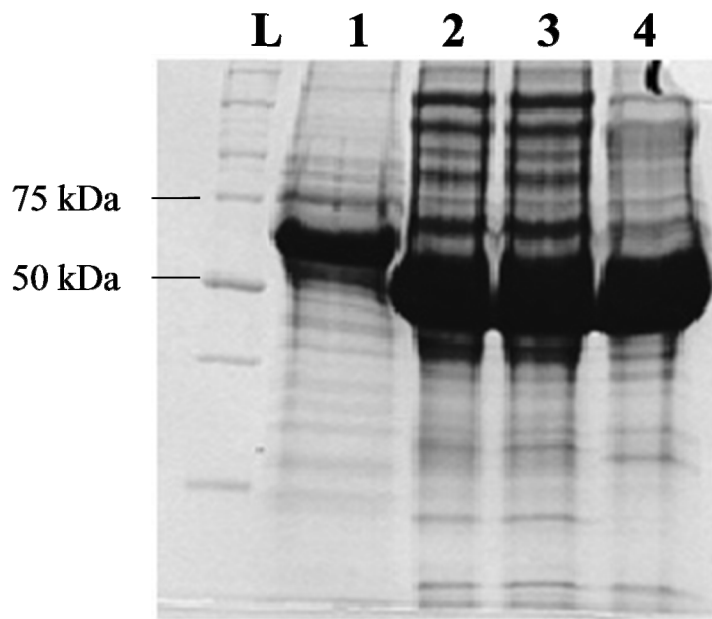
A recombinant CES2-*pVL1392* vector had previously been constructed in our laboratory. CES2 with a C-terminal His-tag was cloned into pVL1392 using an N-terminal Bgl II site and a C-terminal Xba I site. An Aar I restriction digest site is located 108 bp downstream from the ATG start site of CES2. To construct the full length CES2<sup>+64</sup> coding sequence, the CES2-*pVL1392* vector and the 400 bp product in the pCR Blunt II-TOPO vector were digested with Bgl II and Aar I. A 300 bp fragment from the recombinant <sup>+64</sup>-pCR Blunt II-TOPO vector was gel purified and sub-cloned into the digested CES2-*pVL1392* vector to form the recombinant CES2<sup>+64</sup>-*pVL1392* vector. The recombinant vector was transformed into TOP10 *E. coli* (Invitrogen), amplified, and was isolated by midiprep (Qiagen). Sequencing, in both directions, confirmed the inclusion of the N-terminal coding sequence for CES2<sup>+64</sup>.

$CES2^{+64}$  in the pVL1392 baculovirus transfer vector was co-transfected with linearized baculogold viral DNA into *Sf9* insect cells to form recombinant  $CES2^{+64}$  virus. Plaque assays were completed to identify a virus of increase virulence. Nine of the eleven analyzed plaques contained DNA encoding for  $CES2^{+64}$  (Figure 19). The viral DNA that yielded the strongest PCR band on agarose gel electrophoresis was chosen for  $CES2^{+64}$  expression. *Sf9* cells were infected with the selected  $CES2^{+64}$  virus.  $CES2^{+64}$  was purified separately from both the media and the cells in a one-step procedure using nickel resin. The  $CES2^{+64}$  protein in the media eluted from the nickel resin in the 60 mM imidazole and 100 mM imidazole fractions.  $CES2^{+64}$  protein purified from the cells was collected in the 60 mM imidazole fractions. Fractions were pooled and concentrated resulting in 3 final samples 1)  $CES2^{+64}_{\text{media 60}}$ , 2)  $CES2^{+64}_{\text{media 100}}$ , 3)  $CES2^{+64}_{\text{cells}}$ . Eighty-nine percent of the recovered activity appeared in the media and 61% was in the 60 mM imidazole fraction (Table 7). The purification of the recombinant proteins was analyzed by SDS-PAGE (Figure 21 and Figure 22). The protein purified from both the media and the cells have not been purified to homogeneity. Table 7 details the protein concentration, specific activity and protein yield and activity for each of the three samples. Three times more protein was purified from the media than from the cells.

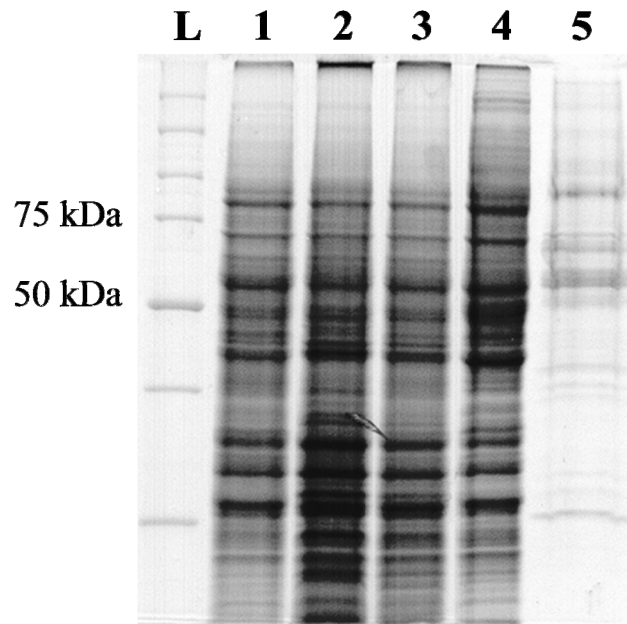




**Figure 19. PCR analysis of viral DNA for selection of a CES2<sup>+</sup> virus:**  
 PCR was used to identify which P1 viral stock contained recombinant virus with CES2<sup>+</sup> DNA. L=100 bp ladder. The templates were 'C' DNA isolated from media of non-infected cells, '1-11' = viral DNA representing the eleven P1 stocks, '+' = CES2<sup>+</sup>-pVL1392, '-' = CES2-pVL1392, ntc = no template control. All negative controls (C, -, and ntc) were clean. The P1 stock from sample 8 was chosen for creating the P2 viral stock.



**Figure 20. SDS-PAGE analysis of the purification of recombinant CES2<sup>+64</sup> protein from the media of *Sf9* insect cells:** CES2<sup>+64</sup> protein was purified from the media (1) of a one liter culture of infected *Sf9* insect cells. Nickel resin was used to purify the recombinant CES2<sup>+64</sup> which had a C-terminal His-tag. Protein was eluted from the nickel resin with 60 mM imidazole (2, 3) and 100 mM imidazole (4). The gel was stained for protein with Coomassie blue.



**Figure 21. SDS-PAGE analysis of the purification of recombinant CES2<sup>+64</sup> protein from *Sf9* insect cells:** CES2<sup>+64</sup> was purified from *Sf9* insect cells infected with recombinant baculovirus containing the recombinant CES2<sup>+64</sup> cDNA. The cells were sonicated followed by centrifugation at 35,000 x *g* for 30 minutes. The supernatant (1) was collected and separated from the pellet (2). A small dense layer (3) formed between the supernatant and the pellet. The supernatant was incubated with nickel resin, and after one hour the flow-through (4) was collected. The purified protein (5) was eluted from the resin with 60 mM imidazole. The gel was stained with Coomassie blue.

Sample	Concentration (mg/ml)	Volume (ml)	Total protein (mg)	Specific activity
CES2 <sup>+64</sup> <sub>media 60</sub>	1.12	2	2.2	51.5
CES2 <sup>+64</sup> <sub>media 100</sub>	1.04	1	1.0	52.6
CES2 <sup>+64</sup> <sub>cells 60</sub>	0.36	2.9	1.0	17.6

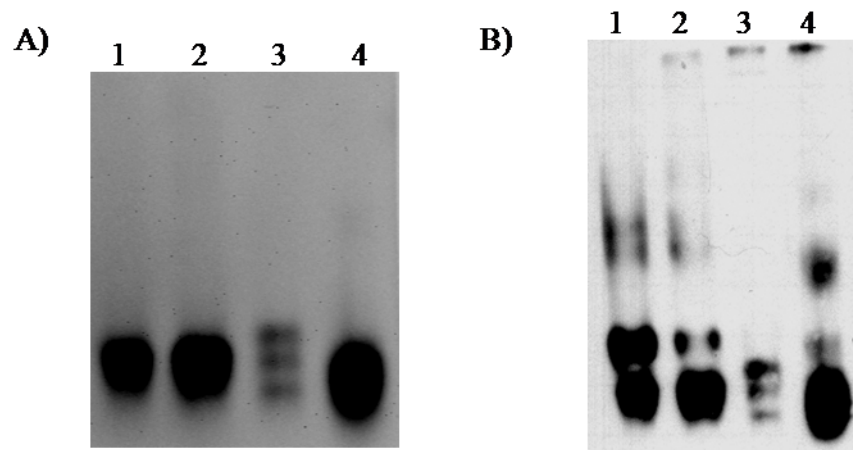
**Table 7. CES2<sup>+64</sup> protein purification yield:** CES2<sup>+64</sup> protein was purified separately from the media and cells of a 1 liter preparation of *Sf9* insect cells. The subscripts following CES2<sup>+64</sup> indicate the sample from which the protein was purified (media or cells) and the concentration of imidazole (60 or 100mM) used to elute the protein from the nickel resin. Protein concentrations were determined by the BioRad protein assay. A total of 3.2 mg of protein was purified from the media as compared to 1 mg of protein purified from the cell pellet. The specific activity and total activity are listed in  $\mu\text{mol min}^{-1} \text{mg}^{-1}$  and  $\mu\text{mol min}^{-1}$ , respectively.

#### 4-Methylumbelliferyl acetate hydrolysis assay

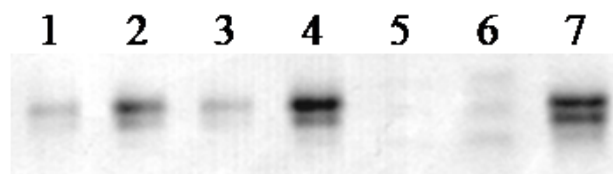
The carboxylesterase activities of the purified CES2<sup>+64</sup> samples were determined by the 4-MUA hydrolysis assay and protein concentration. The average activities from CES2<sup>+64</sup><sub>media 60</sub> and CES2<sup>+64</sup><sub>media 100</sub> were 51.5 and 52.6  $\mu\text{mol min}^{-1} \text{mg}^{-1}$ , respectively. The 4-MUA hydrolysis activity of CES2<sup>+64</sup><sub>cells</sub> was 17.6  $\mu\text{mol min}^{-1} \text{mg}^{-1}$  (Table 7).

#### Analytical non-denaturing polyacrylamide gel electrophoresis and western blot analysis

The CES2<sup>+64</sup> samples and purified recombinant CES2 were electrophoresed on an 8% non-denaturing polyacrylamide gel and stained for activity with 4-MUA (Figure 22A). The CES2<sup>+64</sup><sub>media</sub> samples had only one activity band each. The bands were similar in position and intensity to CES2. The stains were not quantitative due to over-saturation of the bands so recoveries in samples could not be compared from this data. CES2<sup>+64</sup><sub>cells</sub> showed three distinct activity bands. All three active bands of CES2<sup>+64</sup><sub>cells</sub> and the active bands from both CES2<sup>+64</sup><sub>media</sub> samples were identified by the CES2 antibody (Figure 22B). Several other bands lacking 4-MUA activity were also detected in the CES2<sup>+64</sup><sub>media</sub> and CES2 samples, especially in lanes 1 and 4 of Figure 22B. The identity of these inactive bands is not known but they could be truncated or non-glycosylated forms of CES2 or insect cell proteins. CES2<sup>+64</sup> samples were electrophoresed on a separate 8% non-denaturing polyacrylamide gel and stained for protein with Coomassie blue (Figure 23). The protein banding pattern for the CES2<sup>+64</sup><sub>media</sub> samples resembled that of CES2. CES2<sup>+64</sup><sub>cells</sub> migrates to the same vicinity as CES2; however, there were three distinct bands seen both by activity stain and western blotting.



**Figure 22. Activity (A) and western blot (B) analysis of CES2<sup>+64</sup>:** Proteins were electrophoresed on an 8% non-denaturing gel and stained for activity. The same gel was subsequently used for western blot analysis. Lane 1) 0.1  $\mu$ g CES2<sup>+64</sup><sub>media 60</sub>, Lane 2) 0.1  $\mu$ g CES2<sup>+64</sup><sub>media 100</sub>, Lane 3) 0.2  $\mu$ g CES2<sup>+64</sup><sub>cells 60</sub>, and Lane 4) 0.1  $\mu$ g CES2.



**Figure 23. Coomassie blue staining of CES2<sup>+64</sup> on a non-denaturing polyacrylamide gel:** Lanes 1) and 2) 2  $\mu$ g and 5  $\mu$ g CES2<sup>+64</sup><sub>media 60</sub>, respectively; Lanes 3) and 4) 2  $\mu$ g and 5  $\mu$ g CES2<sup>+64</sup><sub>media 100</sub>, respectively; Lanes 5) and 6) 2  $\mu$ g and 5  $\mu$ g CES2<sup>+64</sup><sub>cells 60</sub>, respectively; and Lane 7) 2  $\mu$ g CES2.

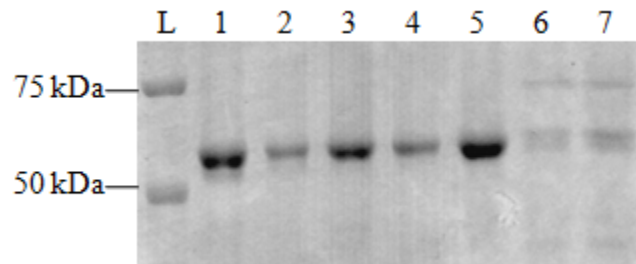
### SDS-polyacrylamide gel electrophoresis and western blot analysis

The CES2<sup>+64</sup> purified proteins were electrophoresed on an 8% polyacrylamide gel followed by either Coomassie blue staining or western analysis (Figure 24 and Figure 25). Based on sequence analysis, CES2<sup>+64</sup> was predicted to be approximately 69 kDa. The CES2<sup>+64</sup><sub>media</sub> samples migrated similarly to CES2 and were approximately 60 kDa. These bands were identified by the CES2 antibody. Coomassie staining revealed that CES2<sup>+64</sup><sub>cells</sub> was not purified to homogeneity, and there appeared to be two or three poorly separated bands that migrated similarly to CES2. On western analysis, only one band was detected in the same position as CES2. Western analysis identified several larger bands for CES2<sup>+64</sup><sub>cells</sub>; however, these bands produced a weaker signal than the 60 kDa band.

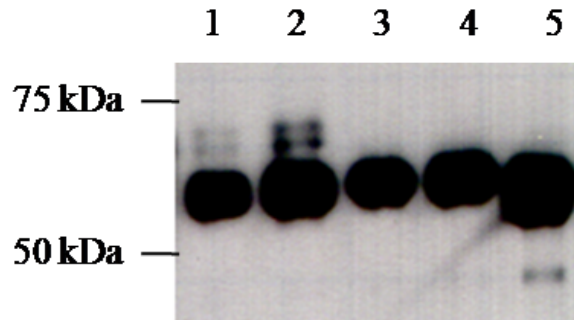
### Glycosylation analysis of CES2<sup>+64</sup>

To determine whether CES2<sup>+64</sup> was glycosylated, preliminary studies were completed to determine which lectin probe to use. Digoxigenin-labeled- Galanthus nivalis agglutinin (GNA), Sambucus nigra agglutinin (SNA), and Datura stramonium agglutinin (DSA) were tested against CES2. Only GNA, which binds to terminal mannose, produced a significant band for CES2 (data not shown). CES2<sup>+64</sup> samples and CES2 were analyzed for glycosylation following both SDS-PAGE and non-denaturing PAGE and subsequent transfer to PVDF membrane. Following SDS-PAGE, all three CES2<sup>+64</sup> samples appeared to be glycosylated (Figure 26A). Because three discrete bands were detected by Coomassie and western for CES2<sup>+64</sup><sub>cells</sub> under non-denaturing conditions, the glycosylation study was also completed for the non-denatured CES2<sup>+64</sup>

samples. As expected CES2 and CES2<sup>+64</sup><sub>media</sub> all had one glycosylated band. The upper two bands of CES2<sup>+64</sup><sub>cells</sub> were glycosylated, but the lower band was not (Figure 26B).

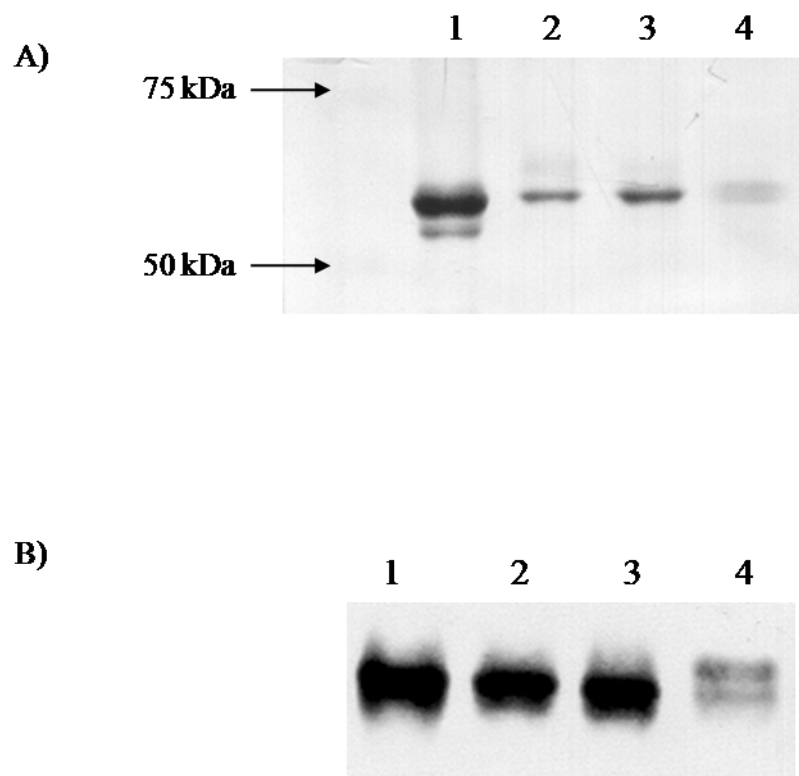


**Figure 24. SDS-PAGE analysis of recombinant CES2<sup>+64</sup> proteins:** Purified recombinant proteins were electrophoresed on an 8% SDS-polyacrylamide gel and stained with Coomassie blue. L) Molecular weight markers; 1) 2 µg CES2; 2) and 3) 5 µg and 10 µg CES2<sup>+64</sup><sub>media 60</sub>, respectively; 4) and 5) 5 µg and 10 µg CES2<sup>+64</sup><sub>media 100</sub>, respectively; 6) and 7) 5 µg and ~9 µg CES2<sup>+64</sup><sub>cells 60</sub>, respectively.



**Figure 25. Western blot analysis of recombinant CES2<sup>+64</sup> proteins:** Purified recombinant proteins were electrophoresed on an 8% SDS-polyacrylamide gel and transferred to a membrane for western blot analysis with anti-CES2 antibody. 1) and 2) 0.1 µg and 0.2 µg CES2<sup>+64</sup><sub>cells 60</sub>, respectively; 3) 0.1 µg CES2<sup>+64</sup><sub>media 60</sub>; 4) 0.1 µg CES2<sup>+64</sup><sub>media 100</sub>; 5) 0.1 µg CES2.

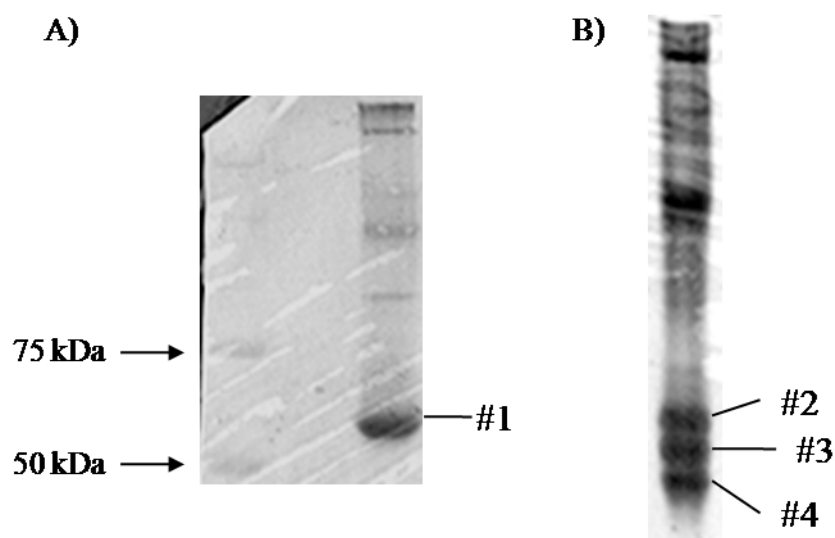




**Figure 26. GNA glycosylation staining of recombinant CES2<sup>+64</sup> proteins:** CES2<sup>+64</sup> samples and CES2 were electrophoresed by (A) SDS-PAGE: 1) CES2 2) CES2<sup>+64</sup><sub>media 60</sub>, 3) CES2<sup>+64</sup><sub>media 100</sub>, 4) CES2<sup>+64</sup><sub>cells 100</sub> and (B) non-denaturing PAGE 1) CES2 2) CES2<sup>+64</sup><sub>media 100</sub>, 3) CES2<sup>+64</sup><sub>media 60</sub>, 4) CES2<sup>+64</sup><sub>cells 100</sub> followed by transfer to membranes. Glycosylation was detected using the DIG Glycosylation kit (Roche). All CES2 and CES2<sup>+64</sup> bands were glycosylated except the third active band of CES2<sup>+64</sup><sub>cells 100</sub>.

### N-terminal sequencing

CES2<sup>+64</sup><sub>media 60</sub> was electrophoresed on by SDS-PAGE and CES2<sup>+64</sup><sub>cells</sub> was electrophoresed by non-denaturing PAGE. The proteins were transferred to PVDF membrane and stained with Coomassie blue. The 60 kDa CES2<sup>+64</sup><sub>media 60</sub> band (#1) and the three CES2<sup>+64</sup><sub>cells</sub> bands (#2, #3, #4) (Figure 27) were excised, treated with PGAP, and sent to Cambridge Peptides for N-terminal sequencing by Edman degradation. Although Coomassie staining indicated sufficient protein, Cambridge Peptide reported that all samples were of “low intensity” indicating that protein recovery from the membrane for sequencing was poor. All sequencing results are summarized in Table 8. The first residue of bands #1, #2, and #3 were not successfully sequenced. The sequence of residues 2-5, Ser-Ala-Ser-Pro, was the same for all four bands. This sequence corresponded to the N-terminal of CES2 indicating that all CES2<sup>+64</sup> bands were CES2 and did not contain the additional 64 N-terminal residues.



**Figure 27. PVDF membranes with CES2<sup>+64</sup> protein bands for N-terminal sequencing:** (A) 11  $\mu$ g CES2<sup>+64</sup> purified from media (B) 57  $\mu$ g of CES2<sup>+64</sup> purified from cells were electrophoresed by SDS-PAGE and non-denaturing PAGE, respectively. Bands #1 through #4 were sent for N-terminal sequence analysis (see Table 8).

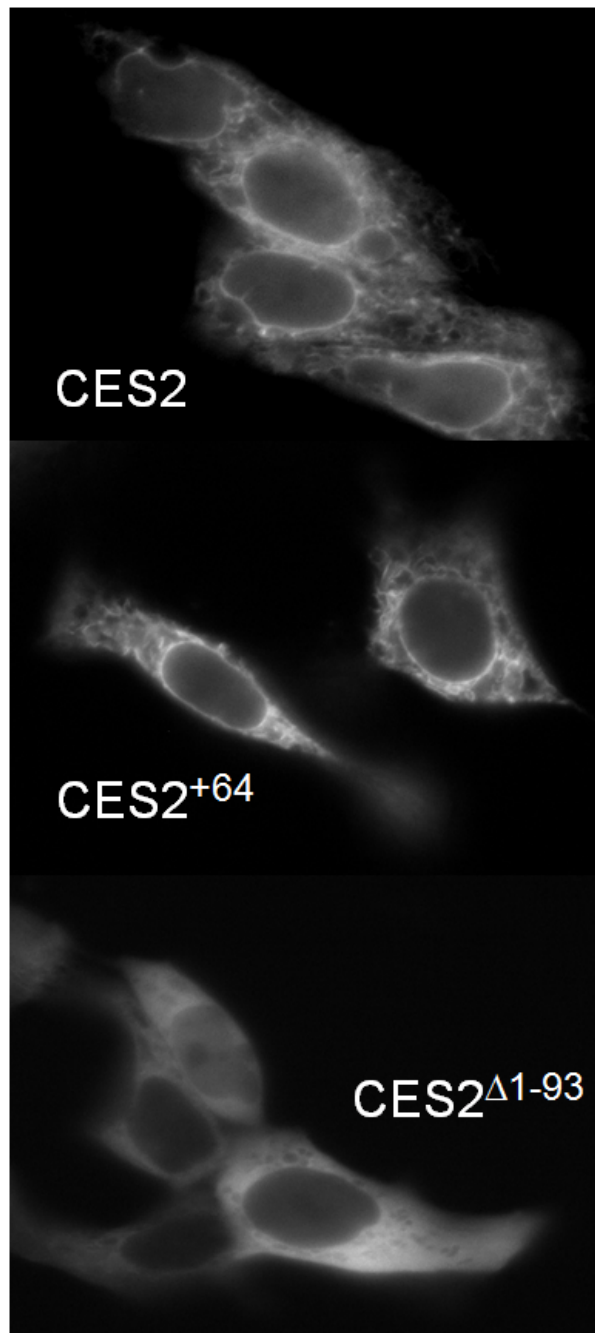
N-terminal Residue	Band #1	Band #2	Band #3	Band #4
1	--	--	--	Q
2	S	S	S	S
3	A	A	A	A
4	S	S	S	S
5	P	P	P	P

**Table 8. N-terminal sequencing results for CES2<sup>+64</sup>:** The sequences were on the edge of detection for all four bands. However, all sequences matched the N-terminal sequence of CES2 indicating that the +64 region had been cleaved. Q=glutamine, S=serine, A=alanine, and P=proline

## V. Sub-cellular localization of CES2 variants

Three CES2-GFP clones were made to study their sub-cellular localization: *pEGFP-CES2*, *pEGFP-CES2<sup>+64</sup>*, and *pEGFP-CES2<sup>Δ1-93</sup>*. The N-terminal coding sequence of *CES2<sup>+64</sup>* and *CES2*, including the signal sequence, and the entire coding region of *CES2<sup>Δ1-93</sup>* were amplified by PCR. Each PCR product was flanked by a Bgl II and Sal I sites at the 5' and 3' ends, respectively. A 5'- Sal I site and a 3'- HTEL coding sequence plus a Not I site were added by PCR to the *EGFP* sequence of pEGFP-N2. The *CES2* variant PCR products and the modified *EGFP* were each cloned into *E. coli* replication vectors and transformed in to *E. coli*. Colonies were selected and grown in culture. Recombinant vectors were isolated and digested with either Bgl II and Sal I or Sal I and Not I. The pEGFP-N2 vector was linearized with Bgl II and Sal I followed by dephosphorylation. Recombinant CES2 variant-GFP constructs were made by ligation of the three pieces, CES2-variant, modified EGFP, and digested pEGFP-N2. The recombinant CES2 variant-GFP constructs were transformed into *E. coli*. Recombinant vectors were isolated by mini-prep from overnight cultures. All constructs were confirmed by DNA sequencing.

The recombinant CES2 variant-GFP constructs and the pEGFP-N2 vector alone were transfected into HCT-15 cells using the FuGene6 reagent (Roche). The HCT-15 cells had been seeded on poly-L-lysine coverslips. An Axiovert 200 microscope (Zeiss) was used to visualize the localization of the chimeric GFP proteins over a 24 to 48-hour period. *CES2* and *CES2<sup>+64</sup>* appeared to localize to the ER while *CES2<sup>Δ1-93</sup>* had a diffuse, cytosolic localization pattern (Figure 28) similar to GFP alone (not shown).



**Figure 28. Localization of CES2 variant-GFP constructs in HCT-15 cells:** CES2 and CES<sup>+64</sup> were localized to the ER while CES2<sup>Δ1-93</sup> remained in the cytoplasm.

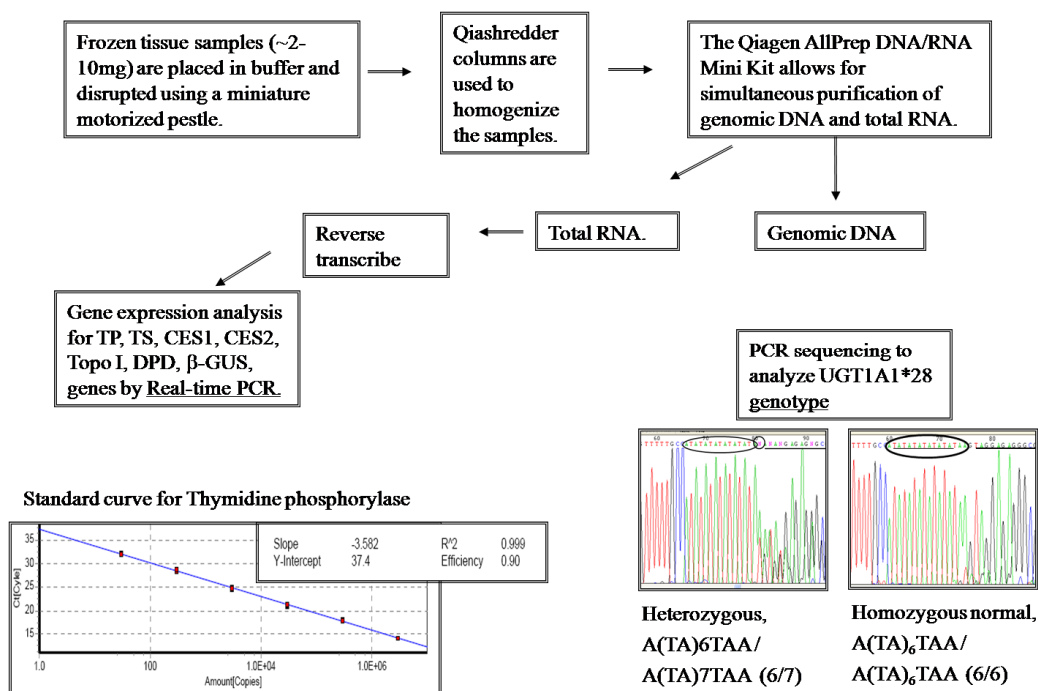
# VIII. The role of CES2, CES1, TOPO I, TP, TS, DPD, $\beta$ -GUS, and UGT1A1 in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine

## Expression of CES2, CES1, TOPO I, TP, TS, DPD, and $\beta$ -GUS in paired tumor and normal rectal biopsy samples

A protocol was successfully developed to obtain high quality mRNA and DNA from small rectal biopsy samples (Figure 29). Genomic DNA and mRNA were isolated from eleven paired tumor and normal biopsy samples using the Allprep DNA/RNA kit (Qiagen). The mRNA and DNA quantity and to a degree quality, by the A260/280 ratio, were analyzed using the ND-1000 Spectrophotometer (Nanodrop). Standard curves were created for each gene from recombinant vectors constructed in our laboratory. Amplification of each standard curve was linear over a 100,000-fold range with respect to copy number ( $r^2 > 0.99$ ). The amplification efficiency was  $\geq 0.85$  for each gene. Melt curve analysis was used to determine the uniformity of the PCR products. All cDNA samples were analyzed in triplicate, and the median expression is reported in Table 9.

## UGT1A1 sequencing

Genomic DNA, isolated from either a normal biopsy or surgery sample was used as a template in PCR to amplify the promoter region of *UGT1A1*. PCR products were purified by agarose gel electrophoresis followed by extraction from the gel. The purified products were sequenced, and the chromatograms were analyzed for TA insertions or deletions (Figure 30). If the chromatogram appeared to contain two sequences, this indicated that the sample was heterozygous for the *UGT1A1* promoter. Four patient samples #4, #5, #6, and #8 were homozygous for (TA)<sub>6</sub>TAA/(TA)<sub>6</sub>TAA (6/6). The other seven patient samples were heterozygous for the *UGT1A1*\*28 allele, (TA)<sub>6</sub>TAA/(TA)<sub>7</sub>TAA (6/7).



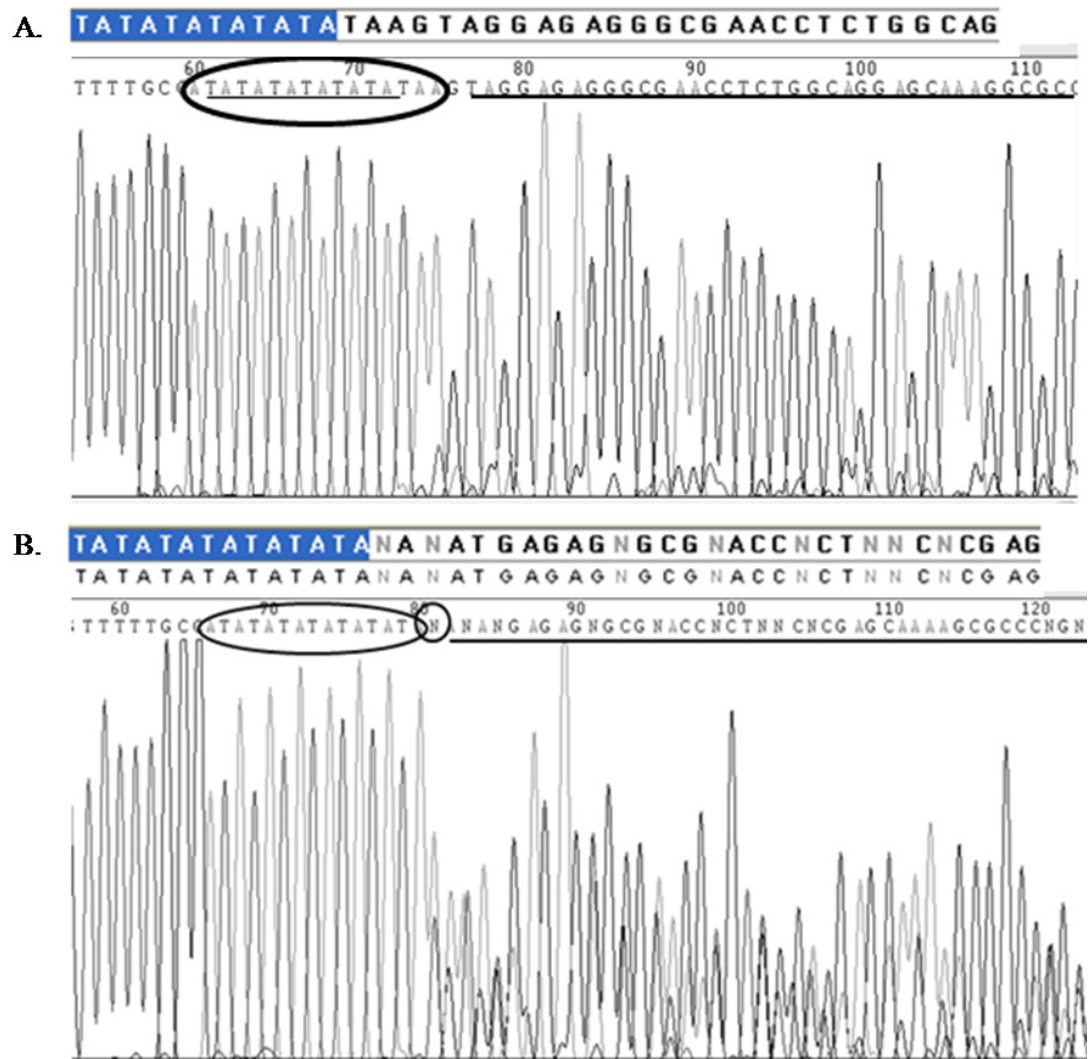
**Figure 29. Summary of the protocol for the HOG GI03-053 rectal tissue samples**

<b>TUMOR</b>	<b>Name</b>	<b>CES2</b>	<b>CES1</b>	<b>TP</b>	<b>TS</b>	<b>B-GUS</b>	<b>TOPOI</b>	<b>DPD</b>
1	<b>Tb 1</b>	3996	29.3	1807	15991	3611	6029	16.3
2	<b>Tb 2</b>	8403	15.3	550	733	525	634	42.9
3	<b>Tb 3</b>	2175	202	1256	3513	506	1013	13.6
4	<b>Tb 4</b>	14872	33	1205	829	930	1381	94.6
5	<b>Tb 5</b>	16531	69.7	2309	860	1615	1463	138
6	<b>Tb 6</b>	1853	142	6659	1676	516	1026	118
7	<b>Tb 7</b>	8151	631	5445	3364	2912	4624	205
8	<b>Tb 8</b>	15804	584	2009	1204	1058	3773	133
9	<b>Tb 9</b>	13778	116	3242	1171	970	4938	133
10	<b>Tb 10</b>	5329	3359	1848	3280	719	1416	42
11	<b>Tb 11</b>	4927	55.1	11838	2775	1498	1696	635

<b>NORMAL</b>	<b>Name</b>	<b>CES2</b>	<b>CES1</b>	<b>TP</b>	<b>TS</b>	<b>B-GUS</b>	<b>TOPOI</b>	<b>DPD</b>
1	<b>Nb 1</b>	15287	144	1319	3207	1763	1983	23.6
2	<b>Nb 2</b>	32059	38.9	1315	2380	1633	2728	4.15
3	<b>Nb 3</b>	27488	226	7606	3255	1500	1848	26.4
4	<b>Nb 4</b>	28636	94.4	2518	795	1848	1453	25.6
5	<b>Nb 5</b>	12599	19.3	382	441	549	950	8.95
6	<b>Nb 6</b>	16871	89.7	386	1238	563	912	5.47
7	<b>Nb 7</b>	21292	56.2	1447	867	1498	1139	11.4
8	<b>Nb 8</b>	7795	11.6	478	689	353	774	2.59
9	<b>Nb 9</b>	10695	61.3	348	400	530	673	6.09
10	<b>Nb 10</b>	11017	43.8	88.6	476	603	583	13.8
11	<b>Nb 11</b>	5729	9.94	86.2	55.3	231	196	3.39

**Table 9. Gene expression data for HOG GI03-053 rectal samples:** Data are reported in copy numbers as determined by real-time PCR.

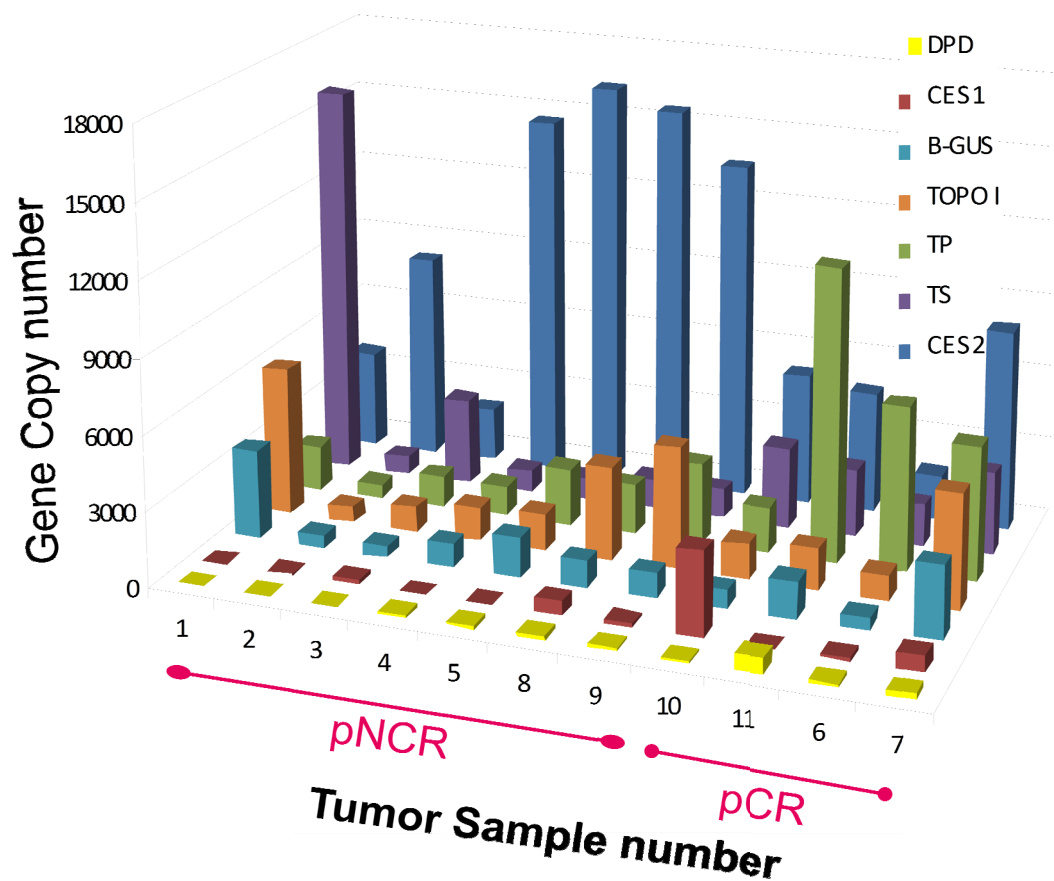




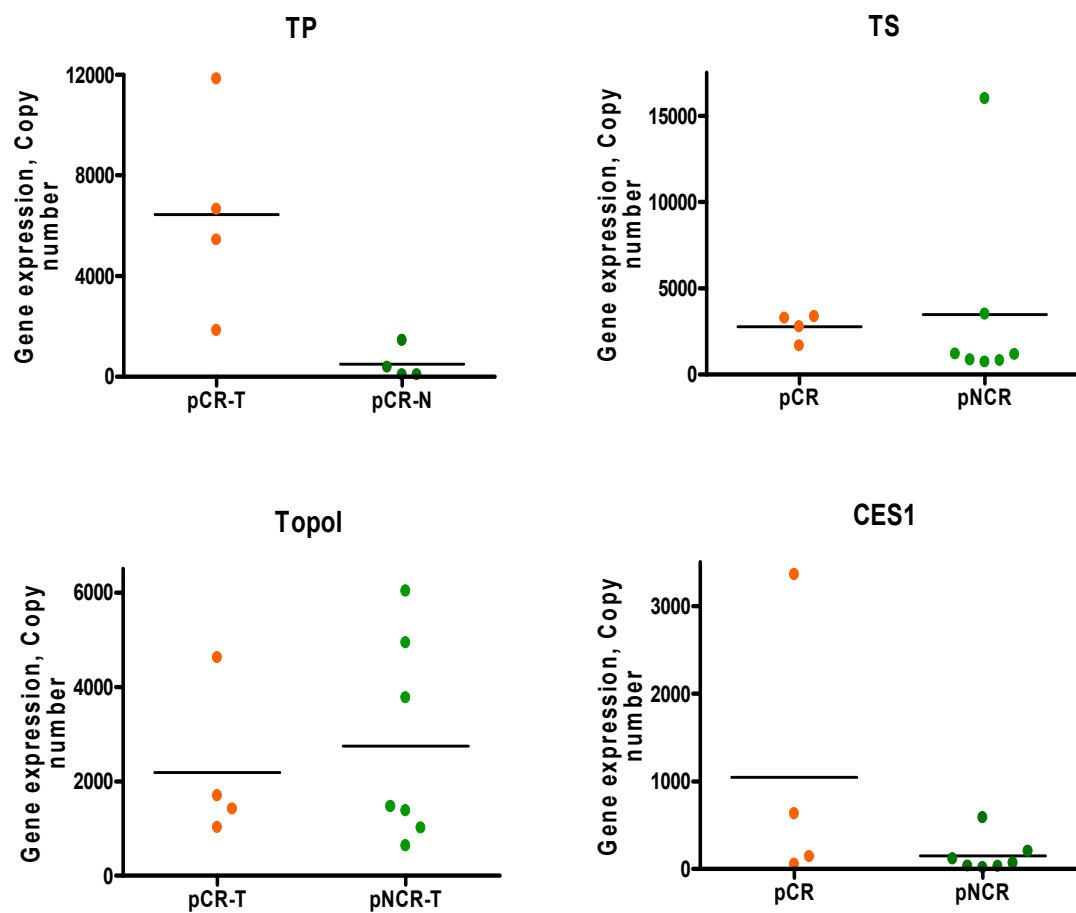
**Figure 30. *UGT1A1* sequencing chromatograms:** (A) Wild type, homozygous (TA)6TAA/ (TA)6TAA (6/6). (B) Heterozygous for the *UGT1A1*\*28 allele, (TA)6TAA/ (TA)7TAA (6/7).

### Clinic data and statistical analysis

Clinical data were collected by members of the HOG GI03-53. Statistical analysis was completed by statisticians in the HOG GI03-53 group. In order to maintain objectivity during analysis of the tissue samples, we were not made aware of the clinical data for individual patients. Patients were classified as exhibiting either a pathologic complete response (pCR) or a pathologic non-complete response (pNCR). Four of the eleven patients had pCR while the remaining seven demonstrated pNCR. Figure 31 shows the rectal tumor gene expression profiles for all eleven patients grouped by pCR and pNCR. The average expressions of individual genes were compared between pCR and pNCR (Figure 32). A significantly higher *CES2* expression was observed in the pNCR group ( $p=0.0291$ ). There was a trend towards higher expression of *CES1*, thymidine phosphorylase (*TP*), and thymidylate synthase (*TS*) in the pCR group. No correlation has been observed between gene expression and toxicity profiles. The collection of clinical data and gene expression data is on-going.



**Figure 31. Expression profiles of HOG GI03-53 complete responders (pCR) and non-complete responders (pNCR):** Real-time PCR was used to determine the expression of *DPD*, *CES1*,  $\beta$ -*GUS*, *TOPO I*, *TP*, *TS*, and *CES2* in rectal tumor samples.



**Figure 32. Comparison between complete responders (pCR) and non- complete responders (pNCR) with respect to the expression of *TP*, *TS*, *TOPO I*, and *CES1***

## DISCUSSION

Carboxylesterases are responsible for the metabolism of numerous endogenous and xenobiotic compounds. Irinotecan and capecitabine are two chemotherapy pro-drugs hydrolyzed by carboxylesterases. Carboxylesterases convert the pro-drug irinotecan to its active form SN-38, which is 1000-fold more cytotoxic. CES2, located predominately in the liver and intestines, is the most important enzyme for this activation (Humerickhouse, et al., 2000; Sanghani, et al., 2004). Irinotecan has shown to be ineffective therapy for lymphomas and gallbladder tumors which lack CES2 expression suggesting that local expression of CES2 in tumors may influence treatment outcome (Xu et al., 2002). High inter-individual variability in outcome has been demonstrated when irinotecan is used in the treatment of colorectal cancer. One of the most debilitating adverse effects, delayed onset diarrhea has limited the use of irinotecan in some patients. Khanna et al. (2000) suggested that gut toxicity may be due to local conversion of irinotecan to SN-38 in the gastrointestinal mucosa. Sanghani et al. (2003) reported significant variation of *CES2* expression in 24 colon tumor samples. We hypothesized that the levels of carboxylesterase in the tumor and surrounding normal tissue may contribute to the inter-individual variation in both response and adverse affects.

The human genome is known to contain fewer than 30,000 genes yet over 90,000 proteins are estimated to comprise the human proteome (Ast, 2004; Xing, 2007). Single nucleotide polymorphisms (SNPs), alternative splicing and copy number variants (CNVs) all contribute to the formation of the complex human proteome. Several *CES2* SNPs have been studied; however, there is no consensus on their affect on CES2 expression and activity. The Database of Genomic Variants (Zhang et al., 2006) does not report any

CNV for *CES2*, unlike *CES1* which has a reported six CNVs. Based on the EST database, we identified alternative splicing in exon 1 and exon 10 of *CES2* (Figure 3). Two potential translation start sites were also noted in exon 1. We proposed that splice variants of *CES2* could contribute to the inter-individual variation in *CES2* expression, activity, and ultimately therapeutic response to irinotecan and capecitabine. We aimed to characterize the expression patterns and activities of several *CES2* splice variants, *CES2*<sup>Δ458-473</sup>, *CES2*<sup>+64</sup>, and *CES2*<sup>Δ1-93</sup>. *CES2* is only one of several enzymes that contribute to the metabolism of irinotecan and capecitabine. We designed a protocol to study the expression of *CES2* and other genes involved with irinotecan and capecitabine metabolism with the ultimate goal of being able to tailor chemotherapy for individual patients.

## **I. Characterization of *CES2* splice variants**

Multi-tissue northern analysis revealed that *CES2* is primarily expressed in the liver and gastrointestinal tract. Expression has also been demonstrated in the heart, kidney, testis, and brain (Satoh et al., 2002). There are three transcripts of approximately 4 kb, 3 kb, and 2 kb in size found in various tissues (Satoh et al., 2002; Wu et al., 2003). There is discrepancy as to whether *CES2* is expressed in skeletal muscle tissue (Satoh et al., 2002; Wu et al., 2003). Wu et al. (2003) reported that different promoters explained the three *CES2* transcripts on northern blot. We proposed that alternative splicing contributes to the multiplicity of the mRNA transcripts. A northern blot was probed for *CES2*<sup>Δ1-93</sup> and *CES2*<sup>+64</sup> to determine if alternative splicing contributed to the multiple transcripts (Figure 9). The sequence for *CES2*<sup>Δ1-93</sup> was found only in the 4.1 kb transcript. Expression of the *CES2*<sup>+64</sup> sequence was found in both the 4.1 kb and 2.8 kb

transcripts as well as a 1.7 kb transcript in the liver and heart. These data suggest that *CES2* is spliced in a tissue specific manner. *Trans*-acting splice factors present in the different tissue are expected to contribute to the tissue-specific variation. The AceView database currently reports eleven splice variants of *CES2* (Thierry-Mieg and Thierry-Mieg, 2006). Variants are designated by a lowercase letter, in decreasing order of protein size, followed by the current release Apr07. The sizes of several reported variants correspond with the different sized *CES2* transcripts revealed by northern analysis. The mRNA for variant *aApr07* is 4177 bp long and is in concordance with the identification of the *CES2*<sup>+64</sup> sequence in the 4.1 kb transcript. Variant *bApr07*, 3901 bp, and variant *cApr07*, 4140 bp, may also contribute to the 4.1 kb transcript visualized by northern analysis. Variant *dApr07* yields a protein with a 93 amino acid deletion at the N-terminus; however, the reported mRNA is 1970 bp. The *CES2*<sup>Δ1-93</sup> sequence was identified in the 4.1 kb transcript, but a 1.9 kb transcript was not visualized by northern. The 1.9 kb mRNA may be expressed in very low abundance interfering with detection by northern. Analysis of Aceview transcripts shows that *only* transcript *aApr07* 4177 bp would be responsible for formation of full-length protein. In our original northern analysis (Satoh et al., 2002), a cDNA probe encompassing exons 7-10 identified only two transcripts of approximately 2 kb and 3 kb in the liver. A 4 kb band was not detected in the liver, a site of high *CES2* expression. Therefore, AceView is an important tool for identifying multiple transcripts, but it is by no means exhaustive at this stage. The complexity surrounding alternate transcripts increases if alternative splicing events are combined. Further complicating the picture are the possible combinations of alternative splice variants and multiple promoters. Additional northern analysis should be performed

with probes specific to the newly reported splice variants to better define CES2 expression. Understanding which CES2 splice variants are expressed in different tissues will allow researchers to focus their studies of CES2 in those tissues. The pharmacokinetics of compounds metabolized by CES2 can be better understood by knowing which splice variants are expressed in a specific tissue.

#### Expression and Activity of CES2 and CES2<sup>Δ458-473</sup>

CES2 is the most widely studied of the CES2 splice variants. The translated form of CES2 has 559 amino acids with a 26 amino acid leader sequence. The active, glycosylated, microsomal form of CES2 has 533 amino acids. Insertions or deletions created from splice variants are described as  $\Delta^{x-y}$  relative to the translated 559 amino acid protein.

Analysis of *CES2* transcripts in the EST database reveals that combinations of alternate ATG start sites and 2 splicing events in exon 1 and exon 10 could lead to 6 potential proteins (Figure 3). A twenty-five nucleotide sequence (shown in bold in Figure 10) unique to the *CES2*<sup>Δ458-473</sup> transcript was identified in 14 ESTs. The transcript is missing the final 48 nucleotides from exon 10, and the resulting protein CES2<sup>Δ458-473</sup> lacks the 16 amino acid residues directly following the active site histidine. The proximity of the deleted amino acids to the active site suggested that CES2<sup>Δ458-473</sup> may differ in activity from CES2.

The expression levels of the *CES2*<sup>Δ458-473</sup> and *CES2* transcripts were analyzed in tumor and normal colon tissue pairs using variant specific primers and real-time PCR. Primers were able to distinguish between alternate splicing events in exon 10 but could not detect any splicing events occurring in exon 1. Thus copy numbers for *CES2*<sup>Δ458-473</sup>



and *CES2* represented the total transcript levels differentiating only between normal and alternate splicing in exon 10, without regard for alternative splicing in exon 1. Both the *CES2*<sup>4458-473</sup> and *CES2* variants were present in all 20 tissue samples, and *CES2*<sup>4458-473</sup> accounted for approximately 6% (SD < 1.3%) of the total *CES2* transcript, suggesting that it was being spliced at a constant rate. There was large inter-individual variation in *CES2* expression among the tumor tissue samples and the normal colon samples, 27-fold and 45-fold respectively. There was no significant difference in *CES2* expression between tumor and normal colon samples which supports the findings of our previous report (Sanghani et al., 2003). In six paired samples, *CES2* expression was greater in the normal tissues, while the other four pairs have higher expression in tumor. This variability suggests that inter-individual variation in expression of *CES2* could account for differences in patient response to irinotecan therapy. *CES2* expression levels, especially in tumor tissue, may dictate whether or not irinotecan therapy is successful while *CES2* expression levels in normal tissue may contribute to the severity of deleterious side-effects.

For all real-time PCR studies copies numbers were determined using standard curves made from plasmids. There is an on-going discussion as to whether the use of housekeeping genes is appropriate for normalization. An ideal housekeeping gene is expressed at constant levels between tissues. It can therefore be used as an internal control to account for errors introduced by the methods used in the quantification process. Studies have shown that even housekeeping genes have a variable range of expression especially between tissues types and between cancerous and normal tissues (Dheda et al., 2004; Rubie et al., 2005). We chose to report absolute copy numbers without

normalization to GAPDH or  $\beta$ -actin. In order to verify the accuracy and precision of the reported copy numbers two reproducibility control experiments were performed. RNA was isolated from five different pieces of the same tumor sample. Less than 1.6-fold variation in the *CES2* copy numbers could be attributed to the entire method from tissue disruption through real-time PCR (Figure 13). A second experiment was completed in which one sample RNA was reverse transcribed in quintuplicate. Less than a 1.3-fold variation in copy number was attributable to reverse transcription alone (Figure 13B). Accordingly we report absolute copy number in this study because the use of a housekeeping gene in these experiments would unduly increase the reported inter-individual variation.

*CES2* expression data were subjected to linear regression analysis with two different measures of carboxylesterase activity, 4-MUA hydrolysis *in vitro* and band density of *CES2* on nondenaturing PAGE with 4-MUA staining. Unlike the previous study (Sanghani et. al. 2003), variant-specific primers were used to quantify the transcript for *CES2* separately from the transcript for *CES2*<sup>4458-473</sup>. *CES2* expression in the colon tumor samples had significant positive correlation with 4-MUA hydrolase activity and band density. These findings are in concordance with the data previously published by Sanghani et al. (2003) although the correlation coefficients are increased. These increased correlations may be attributable to the use of variant specific primers as well as improved methods. *CES2* expression in normal tissue was also found to have significant, positive correlation with 4-MUA hydrolase activity and positive correlation with band density ( $p \leq 0.055$ ). Correlation values between expression and activity for the tumor samples were greater than the corresponding correlation values in normal samples. It is

possible that the tumor samples are more homogenous in cell type than the normal samples, which may include muscular tissue in addition to mucosal tissue.

Microdissection of tissue samples could be employed to obtain a more accurate representation of CES2 expression in each tissue.

If inter-individual variation in response to irinotecan treatment may be explained by variations in CES2 expression, it is important to examine the expression and activities of the CES2<sup>Δ458-473</sup> variant and other splice variants. CES2<sup>Δ458-473</sup> was cloned and expressed to characterize its activity and physical properties. When purifying the CES2<sup>Δ458-473</sup> variant protein, the protein elution profile of CES2 was used as a guide. The CES2<sup>Δ458-473</sup> protein had an elution profile identical to CES2 for purification on concanavalin A and preparative non-denaturing PAGE. Binding of the CES2<sup>Δ458-473</sup> protein to concanavalin A resin indicated that it is glycosylated. CES2 and CES2<sup>Δ458-473</sup> proteins migrated similarly on non-denaturing PAGE (Figure 17D). When incubated with 4-MUA, a substrate for which CES2 has high activity with a reported  $K_{cat}/K_m$  value of 60,000 mM<sup>-1</sup> min<sup>-1</sup> (Pindel et al., 1997), CES2<sup>Δ458-473</sup> demonstrated no activity. The lack of esterase activity demonstrated by the CES2<sup>Δ458-473</sup> protein raised questions regarding its folding. The circular dichroism (CD) spectrum for CES2<sup>Δ458-473</sup> (Figure 17E) shows the presence of a secondary structure similar to that of CES2. The small increase in  $\alpha$ -helical content and small decrease in  $\beta$ -strand content of CES2<sup>Δ458-473</sup> in comparison to CES2 could be expected from the 16 amino acid deletion. Although the CES2<sup>Δ458-473</sup> protein lacked 4-MUA hydrolase activity, it was feasible that it retained activity for other substrates. The CPT-11 hydrolase activity of the CES2<sup>Δ458-473</sup> protein was evaluated. Even at highest concentration of CES2<sup>Δ458-473</sup> protein, no significant

amount of SN-38 (< 5nM) was detected. Based on these results we concluded that the CES2<sup>Δ458-473</sup> protein lacked irinotecan hydrolase activity. To date efforts to identify the CES2<sup>Δ458-473</sup> protein in tissue samples have been unsuccessful because there is currently no antibody specific for the protein sequence flanking the deleted 16 amino acids of CES2<sup>Δ458-473</sup>. CES2<sup>Δ458-473</sup> varies by only 16 amino acids from CES2, so these two variants have not been distinguishable on SDS-PAGE.

#### Characterization of CES2<sup>+64</sup>

There are two in-frame ATGs in exon1 of *CES2*. If translation began at the first ATG, 64 amino acids would be added to the protein at the N-terminus. We hypothesized that these additional amino acids would affect the activity of the protein as well its subcellular localization. Analysis of CES2<sup>+64</sup> by SignalP 3.0 software (<http://www.cbs.dtu.dk/services/SignalP/>) (Bendtsen et al., 2004; Nielsen et al., 1997) predicted no signal sequence in CES2<sup>+64</sup>. Without a signal sequence, CES2<sup>+64</sup> should not be targeted to the ER and should be expected to remain in the cytoplasm as an unglycosylated protein.

The baculovirus system (BD Biosciences Pharmingen) (Sanghani et al., 2004; Sun et al., 2004) was used to express recombinant CES2<sup>+64</sup>. The C-terminal ER retention sequence in CES2<sup>+64</sup> was replaced with a His-tag at its C-terminus to simplify purification. When using the baculovirus system, C-terminally His-tagged recombinant CES2 protein is secreted into the media. Since we predicted CES2<sup>+64</sup> to be a cytosolic protein, recombinant CES2<sup>+64</sup> protein was purified from both the cells and the media. A one-step procedure using nickel affinity chromatography resin was utilized for purification taking advantage of the C-terminal His-tag. From the media, the elution

profile showed two protein peaks at 60mM and 100mM imidazole, CES2<sup>+64</sup><sub>media 60</sub> and CES2<sup>+64</sup><sub>media 100</sub>, respectively. CES2 predominantly elutes at 60 mM imidazole with 60% of the recovered protein in this fraction (Table 7). Both peaks were collected and analyzed separately. Another protein peak was eluted and represented only 11% of the recovered protein (Table 7). Even after the purification process, the CES2<sup>+64</sup><sub>cell</sub> was not a homogenous sample (Figure 21).

Activity was studied using the 4-MUA hydrolysis assay. CES2<sup>+64</sup><sub>media 60</sub> and CES2<sup>+64</sup><sub>media 100</sub> had similar specific activities, 51.5 and 52.6  $\mu\text{mol min}^{-1} \text{mg}^{-1}$ , respectively. The activity of CES2<sup>+64</sup><sub>cell</sub> was lower at 17.6  $\mu\text{mol min}^{-1} \text{mg}^{-1}$ . In contrast native CES2 purified from human liver has a specific activity of 120  $\mu\text{mol min}^{-1} \text{mg}^{-1}$  (Pindel et al., 1997). The impurity of CES2<sup>+64</sup> samples could account for the lower specific activity of media protein versus purified CES2.

The samples were analyzed by non-denaturing PAGE and SDS-PAGE, both followed by western blot analysis. On non-denaturing PAGE, both CES2<sup>+64</sup><sub>media</sub> samples demonstrated one active band similar to CES2. The CES2<sup>+64</sup><sub>cell</sub> sample revealed three active bands. The active CES2<sup>+64</sup><sub>media</sub> bands and all three CES2<sup>+64</sup><sub>cell</sub> bands were recognized on western blot by an anti-CES2 antibody. Western blot analysis also revealed several inactive bands in the CES2<sup>+64</sup><sub>media</sub> samples. While the identity of these inactive bands is unclear, they may represent polymers or different glycosylation states of the protein. It is unlikely that they are due to cross reactivity between the anti-CES2 antibody and an insect cell proteins. Nickel resin was used to purify the recombinant His-tagged CES2<sup>+64</sup>, and all insect cell proteins should have been eliminated because they did not contain the His-tag. Sequence analysis of CES2<sup>+64</sup> predicts a 69 kDa protein.

However, on SDS-PAGE, the CES2<sup>+64</sup><sub>media</sub> samples migrated similarly to CES2, a 60 kDa protein.

CES2 is a glycosylated ER-protein. If the additional 64 amino acids of CES2<sup>+64</sup> prevent recognition of the signal peptide as predicted, the protein should not be targeted to the ER. Without translocation to the ER, CES2<sup>+64</sup> should not be a glycosylated protein. Glycosylation analysis of CES2<sup>+64</sup><sub>cell</sub> and both CES2<sup>+64</sup><sub>media</sub> bands was performed. All active bands were glycosylated except the third CES2<sup>+64</sup><sub>cell</sub> band. The CES2<sup>+64</sup><sub>media</sub> and the three CES2<sup>+64</sup><sub>cell</sub> bands were sent for sequencing by automated Edman degradation. Senter et al. (2001) reported that CES2 is blocked at its N-terminal. When the N-terminus is blocked by pyroglutamic acid, edman degradation is inhibited. Before sequencing, the samples were treated with *Pfu* Pyroglutamate Aminopeptidase to unblock the N-terminus by removing any pyroglutamic acid residues. Although the four samples were of low intensity making sequencing difficult, the sequences for all four samples corresponded to the N-terminal sequence of the mature wild-type CES2 protein. This indicated that all four samples of the recombinant CES2<sup>+64</sup> protein purified from insect cells did not contain the additional 64 N-terminal amino acids.

There are two potential explanations for this observation: 1) the signal sequence was still identified despite the additional 64 amino acids at the N-terminal or 2) translation begins at the second ATG and the first ATG was bypassed. We believe that initiation at the second ATG is unlikely because translation was under the control of the strong polyhedrin viral promoter in this expression vector.

### Sub-cellular localization of CES2 variants

Wild-type CES2 has an endoplasmic reticulum translocation signal peptide at its N-terminus and an ER retention signal sequence at its C-terminus. Therefore, it is targeted to and retained in the ER where it undergoes glycosylation. In CES2<sup>+64</sup> the signal peptide is preceded by a 64 amino acid addition and the signal peptide cleavage site will be between amino acids 90-91. It is widely believed that signal peptides are generally located within the first 70 amino acids of a protein. Based on predictions by SignalP 3.0 software (Bendtsen et al., 2004; Nielsen et al., 1997), we did not expect a signal sequence to be recognized in CES2<sup>+64</sup>. If no signal sequence is recognized in CES2<sup>+64</sup>, then it should not go to the ER or become glycosylated. The CES2<sup>Δ1-93</sup> variant does not possess the signal peptide nor was a different one predicted; and therefore, it also should not be targeted to the ER or undergo glycosylation. We constructed chimeric CES2 variant-GFP constructs and transfected them into HCT-15 cells. As expected, the chimeric CES2-GFP protein exhibited an ER localization pattern and chimeric CES2<sup>Δ1-93</sup>-GFP protein was localized in the cytoplasm. Further research should be done on CES2<sup>Δ1-93</sup>. The deletion of the first 93 amino acids may also significantly affect the structure and activity of the CES2<sup>Δ1-93</sup> protein compared to CES2. Similar to the chimeric CES2-GFP protein, the chimeric CES2<sup>+64</sup>-GFP protein also exhibited an ER localization pattern. This suggests, as with the glycosylation experiment, that 1) the signal peptide is recognized or 2) translation begins at the second ATG. If the signal peptide is recognized, SignalP 3.0 does not accurately predict the signal sequence for ER localization.

## **II. The role of CES2, CES1, TOPO I, TP, TS, DPD, $\beta$ -GUS, and UGT1A1 in the inter-individual variation in response to treatment of rectal cancer with irinotecan and capecitabine**

Capecitabine and irinotecan are pro-drugs used in the treatment of colorectal cancer. Irinotecan is converted to SN-38, a potent topoisomerase I inhibitor, by carboxylesterases (Figure 4). CES2 has a greater affinity for irinotecan and higher catalytic efficiency for irinotecan when compared with CES1. Capecitabine is an oral pro-drug for 5-FU which has been the mainstay of colorectal chemotherapy for over the past 50 years. Studies have shown that capecitabine can successfully replace 5-FU in dosing regimens. The first step in the activation of capecitabine is mediated by carboxylesterases (Figure 5). Both CES1 and CES2 have similar catalytic efficiency for capecitabine. While carboxylesterases play a crucial role in the activation of both irinotecan and capecitabine, they are not the only enzymes that affect the metabolism of these pro-drugs.

The active metabolite of irinotecan SN-38 is inactivated to SN-38G through glucuronidation that is mediated by UDP-glucuronosyltransferases (UGT), especially UGT1A1.  $\beta$ -glucuronidase in both the tissue and the gut flora can convert the inactive SN-38G back to the active form SN-38 (Takasuna et al., 1996). Enterohepatic recirculation is believed to play a role in the complexity of irinotecan metabolism.

After conversion of capecitabine to 5'-DFCR by carboxylesterases, cytidine deaminase, which is highly expressed in the liver and some solid tumors, converts 5'-DFCR to 5'-DFUR (Miwa et al., 1998). Thymidine phosphorylase converts 5'-DFUR to the active 5-FU. Increased expression of thymidine phosphorylase in many tumors guided the development of capecitabine as a pro-drug of 5-FU (Miwa et al., 1998). Elevated thymidine phosphorylase contributes to thymidine salvage as well as to



angiogenesis (Brown and Bicknell, 1998). 5-FU is incorporated into nucleotide metabolites that interfere with DNA and RNA synthesis. FdUMP binds to thymidylate synthase and decreases the production of thymidine. Dihydropyrimidine dehydrogenase inactivates 5-FU.

There is high inter-individual variation in response to the therapy with both irinotecan and capecitabine. Numerous studies have been completed to study how the different genes involved in the metabolism of these drugs contribute to therapeutic efficacy and adverse outcomes. Our laboratory, along with others, have proposed that CES2 may contribute to variations in response to irinotecan (Sanghani et al., 2003; Xie et al., 2002; Xu et al., 2002). Other groups have examined the effects of UGT1A on irinotecan treatment outcomes (Carlini et al., 2005; Gagne et al., 2002; Hanioka et al., 2001; Innocenti et al., 2004; Jinno et al., 2003; Lankisch et al., 2005; Tukey et al., 2002). TA insertions into the TATA box region of *UGT1A1* affect the promoter activity. The *UGT1A1*\*28 allele has an additional TA insertion (TA)<sub>7</sub>TAA and decreases the promoter activity by 70% (Ramchandani et al., 2007). Patients who are homozygous for the *UGT1A1*\*28 allele have demonstrated an increased risk for toxicity, especially neutropenia (McLeod and Watters, 2004; Rouits et al., 2004).  $\beta$ -GUS expression has been correlated to irinotecan induced diarrhea. (Kehrer et al., 2000; Takasuna et al., 1996; Tobin et al., 2006). Expression and activity of topoisomerase I, the target of irinotecan, were found to best predict response to irinotecan (Guichard et al., 1999; Jansen et al., 1997; Pavillard et al., 2004; Sanghani et al., 2003). When treated with 5-FU, colorectal cancer patients whose tumors expressed low levels of *TS*, *TP*, and *DPD* were found to have prolonged survival as compared to patients whose tumors had high expression of

any one of those genes (Salonga et al., 2000). This study supported using more than one predictor of response to increase identification of patients who respond to therapy.

A Hoosier Oncology Group HOG GI03-53 study was designed to evaluate the clinical and pathologic outcomes in rectal cancer patients treated with a combination of irinotecan, capecitabine, and radiation followed by surgery. Biopsy samples were collected from normal and tumor tissue prior to the initiation of therapy. Samples were also collected from tissues excised during surgery. Clinic data including tumor response and adverse outcomes were amassed throughout the study. A protocol was developed to study the expression of eight genes involved with the metabolism of irinotecan and capecitabine (Figure 29). The design of this experiment studies five genes, potential predictors of response, found in metabolic pathway of each drug. *CES1*, *CES2*, *UGT1A1*,  $\beta$ -*GUS*, and *TOPO I* are in the irinotecan pathway while *CES1*, *CES2*, *TP*, *TS*, and *DPD* contribute to the metabolism of capecitabine. The goal of this study is to find correlations between the clinical data and gene expression in order to predict treatment outcomes and tailor chemotherapy for each patient.

Real-time PCR experiments were designed to study the expression of *CES2*, *CES1*, *TOPO I*, *TP*, *TS*, *DPD*, and  $\beta$ -*GUS*. Methods were designed to sequence the promoter region of *UGT1A1*. We report protocol development and preliminary data from the biopsy samples of eleven patients.

RNA and genomic DNA were isolated from the biopsy samples using QIAshredders and the Allprep DNA/RNA kit from Qiagen. RNA was quantified and reverse transcribed using previously described methods (Schiel et al., 2007). For real-time PCR analysis, standard curves with a 100,000-fold linear amplification range with

respect to copy number were designed for *CES2*, *CES1*, *TOPO I*, *TP*, *TS*, *DPD*, and  $\beta$ -*GUS*. Gene expression was reported in absolute copy numbers without normalization to a housekeeping gene (Schiel et al., 2007). Data were collected on normal and tumor rectal tissue sample from eleven patients. We fully expect this protocol to be translatable to studying the biopsy samples from newly accrued patients as well as corresponding surgery samples.

The promoter area of *UGT1A1* was amplified by PCR using primers and reaction conditions described by (Monaghan et al., 1996). Genomic DNA from either normal biopsy or surgery samples was used as a template. No research was found to suggest an increased rate of mutation in the *UGT1A1* promoter in tumor tissue. Therefore sequencing was only completed for normal rectal tissue samples. However, genomic DNA was collected and stored at -70°C for both normal and tumor tissue for any desired future studies. Purified PCR samples were sequenced by the DNA Sequencing Core Facility at Indiana University School of Medicine. Chromatograms of each sample were inspected to determine homozygosity or heterozygosity as well as the number of TA repeats in the TATA box region. The genotype was determined for all eleven patient samples.

Patient accrual and the gathering of clinical data and gene expression data are ongoing for the HOG GI03-53 study. Preliminary results from the first eleven patients indicate that the capecitabine/irinotecan regimen is well tolerated. Correlative studies between individual gene expression and therapy response, classified as pathologic complete response (pCR) or pathologic non-complete response (pNCR), suggest a trend for elevated *TP*, *TS*, and *CES1* expression in pCR patients. This is in contrast to studies

that indicate, tumors responding to 5-FU have low expression of DPD, TS, and TP (Metzger et al., 1998; Salonga et al., 2000); however, capecitabine requires TP for conversion to 5-FU which would support the trend for elevated TP in pCR patients. Recent studies with capecitabine have indicated that TP is predictive of response to capecitabine for breast and gastric cancers (Andreetta et al., 2008; Koizumi et al., 2008). Those studies did not include irinotecan in the therapy regimen. Czejka et al. (2005) reported that capecitabine contributed to decreased formation of SN-38 after irinotecan infusion; however, irinotecan pharmacokinetics was not significantly affected. Further statistical analysis will be performed as more clinical data and gene expression data is accrued. The collection of both biopsy and surgery samples will provide an opportunity to study the effects of irinotecan and capecitabine on gene expression. Kocakova et al. (2007) have found that chemoradiotherapy induces expression of *TP* and *TS* mRNA. Further studies could include the study of protein activity and its correlation with both gene expression and clinical response. Studies of cellular transporters of irinotecan and its metabolites, which are also believed to contribute to the inter-individual response to therapy (de Jong et al., 2007; Mathijssen et al., 2001) could be considered.

### **III. Summary**

CES2 is an important enzyme for the activation of irinotecan and capecitabine, two pro-drugs used for the treatment of colorectal cancer. There is high inter-individual variation in response to treatment with both irinotecan and capecitabine. To characterize CES2 expression and activity as well as its role in the metabolism of irinotecan and capecitabine, basic and translational research was performed. Through analysis of the EST database, two ATG translation start sites and splicing events in exon 1 and exon 10

were identified for *CES2*. Three resulting protein variants  $CES2^{\Delta 458-473}$ ,  $CES2^{+64}$ , and  $CES2^{\Delta 1-93}$  along with *CES2* were studied with regards to expression patterns and enzyme activity. Knowledge of the splice variants yielded insight into the *CES2* expression pattern on northern analysis. Significant variation in *CES2* expression was found among 10 pairs of tumor and normal colon tissue; however, no expression pattern difference was identified between tumor and normal sample pairs.  $CES2^{\Delta 458-473}$ , a variant missing 16 amino acids after the active site histidine, was determined to be inactive although it was expressed at approximately 6% in the 10 paired colon samples. Extensive characterization of  $CES2^{+64}$  indicated that its mature protein product was in fact *CES2* suggesting that the native leader sequence is recognized in spite of the addition of 64 N-terminal amino acids. Sub-cellular localization studies showed that, unlike *CES2* and  $CES2^{+64}$  which localize to the ER,  $CES2^{\Delta 1-93}$  is trafficked to the cytoplasm. The methods employed in studying these variants can be used for characterizing newly discovered *CES2* variants. Real-time PCR methods that were developed for *CES2* have been translated into the study of six additional genes contributing to the metabolism of irinotecan and capecitabine. In collaboration with the Hoosier Oncology Group, we will continue to gather clinical and gene expression data from rectal cancer patients being treated with a combination of capecitabine and irinotecan. The ultimate goal would be to tailor chemotherapy regimens to best fit the drug metabolic profile for each patient.

## REFERENCES

- Aldridge, W.N. (1993). The esterases: perspectives and problems. *Chem. Biol. Interact.* 87, 5-13.
- American Cancer Society. Cancer Facts & Figures 2007. 2007. Atlanta, American Cancer Society.  
Ref Type: Report
- Andreetta, C., Puppini, C., Minisini, A., Valent, F., Pegolo, E., Damante, G., Di, L.C., Pizzolitto, S., Pandolfi, M., Fasola, G., Piga, A., and Puglisi, F. (2008). Thymidine phosphorylase expression and benefit from capecitabine in patients with advanced breast cancer. *Ann. Oncol.*
- Ast, G. (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773-782.
- Bendtsen, J.D., Nielsen, H., von, H.G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783-795.
- Brown, N.S. and Bicknell, R. (1998). Thymidine phosphorylase, 2-deoxy-D-ribose and angiogenesis. *Biochem. J.* 334, 1-8.
- Brzezinski, M.R., Spink, B.J., Dean, R.A., Berkman, C.E., Cashman, J.R., and Bosron, W.F. (1997). Human liver carboxylesterase hCE-1: binding specificity for cocaine, heroin, and their metabolites and analogs. *Drug Metab Dispos.* 25, 1089-1096.
- Budman, D.R., Meropol, N.J., Reigner, B., Creaven, P.J., Lichtman, S.M., Berghorn, E., Behr, J., Gordon, R.J., Osterwalder, B., and Griffin, T. (1998). Preliminary studies of a novel oral fluoropyrimidine carbamate: capecitabine. *J. Clin. Oncol.* 16, 1795-1802.
- Canal, P., Gay, C., Dezeuze, A., Douillard, J.Y., Bugat, R., Brunet, R., Adenis, A., Herait, P., Lokiec, F., and Mathieu-Boue, A. (1996). Pharmacokinetics and pharmacodynamics of irinotecan during a phase II clinical trial in colorectal cancer. Pharmacology and Molecular Mechanisms Group of the European Organization for Research and Treatment of Cancer. *J. Clin. Oncol.* 14, 2688-2695.
- Carlini, L.E., Meropol, N.J., Bever, J., Andria, M.L., Hill, T., Gold, P., Rogatko, A., Wang, H., and Blanchard, R.L. (2005). UGT1A7 and UGT1A9 polymorphisms predict response and toxicity in colorectal cancer patients treated with capecitabine/irinotecan. *Clin. Cancer Res.* 11, 1226-1236.
- Chabot, G.G. (1996). Clinical pharmacology and pharmacodynamics of irinotecan. A review. *Ann. N. Y. Acad. Sci.* 803:164-72., 164-172.
- Charasson, V., Bellotti, R., Meynard, D., Longy, M., Gorry, P., and Robert, J. (2004). Pharmacogenetics of human carboxylesterase 2, an enzyme involved in the activation of irinotecan into SN-38. *Clin. Pharmacol. Ther.* 76, 528-535.

- Couteau,C., Risse,M.L., Ducreux,M., Lefresne-Soulas,F., Riva,A., Lebecq,A., Ruffie,P., Rougier,P., Lokiec,F., Bruno,R., and Armand,J.P. (2000). Phase I and pharmacokinetic study of docetaxel and irinotecan in patients with advanced solid tumors. *J. Clin. Oncol.* 18, 3545-3552.
- Cunningham,D., Pyrhonen,S., James,R.D., Punt,C.J., Hickish,T.F., Heikkila,R., Johannesen,T.B., Starkhammar,H., Topham,C.A., Awad,L., Jacques,C., and Herait,P. (1998). Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. *Lancet.* 352, 1413-1418.
- Cygler,M., Schrag,J.D., Sussman,J.L., Harel,M., Silman,I., Gentry,M.K., and Doctor,B.P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Sci.* 2, 366-382.
- Czejka,M., Schueller,J., Hauer,K., and Ostermann,E. (2005). Pharmacokinetics and metabolism of irinotecan combined with capecitabine in patients with advanced colorectal cancer. *Anticancer Res.* 25, 2985-2990.
- D'Agostino R.B., Belanger A., and D'Agostino,R.B.Jr. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician* 44, 316-321.
- Darlington R.B. (2007). Is Kurtosis Really "Peakedness? *The American Statistician* 24, 19-22.
- de Jong,F.A., de Jonge,M.J., Verweij,J., and Mathijssen,R.H. (2006). Role of pharmacogenetics in irinotecan therapy. *Cancer Lett.* 234, 90-106.
- de Jong,F.A., Scott-Horton,T.J., Kroetz,D.L., McLeod,H.L., Friberg,L.E., Mathijssen,R.H., Verweij,J., Marsh,S., and Sparreboom,A. (2007). Irinotecan-induced diarrhea: functional significance of the polymorphic ABCC2 transporter protein. *Clin. Pharmacol. Ther.* 81, 42-49.
- Dean,R.A., Zhang,J., Brzezinski,M.R., and Bosron,W.F. (1995). Tissue distribution of cocaine methyl esterase and ethyl transferase activities: correlation with carboxylesterase protein. *J. Pharmacol. Exp. Ther.* 275, 965-971.
- Dheda,K., Huggett,J.F., Bustin,S.A., Johnson,M.A., Rook,G., and Zumla,A. (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques.* 37, 112-119.
- Dodds,H.M., Haaz,M.C., Riou,J.F., Robert,J., and Rivory,L.P. (1998). Identification of a new metabolite of CPT-11 (irinotecan): pharmacological properties and activation to SN-38. *J. Pharmacol. Exp. Ther.* 286, 578-583.

- Douillard, J.Y., Cunningham, D., Roth, A.D., Navarro, M., James, R.D., Karasek, P., Jandik, P., Iveson, T., Carmichael, J., Alakl, M., Gruia, G., Awad, L., and Rougier, P. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet*. 355, 1041-1047.
- Eyre, T.A., Ducluzeau, F., Sneddon, T.P., Povey, S., Bruford, E.A., and Lush, M.J. (2006). The HUGO Gene Nomenclature Database, 2006 updates. *Nucleic Acids Res.* 34, D319-D321.
- Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., and Guillemette, C. (2002). Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol. Pharmacol.* 62, 608-617.
- Guichard, S., Terret, C., Hennebelle, I., Lochon, I., Chevreau, P., Fretigny, E., Selves, J., Chatelut, E., Bugat, R., and Canal, P. (1999). CPT-11 converting carboxylesterase and topoisomerase activities in tumour and normal colon and liver tissues. *Br. J. Cancer*. 80, 364-370.
- Gupta, E., Mick, R., Ramirez, J., Wang, X., Lestingi, T.M., Vokes, E.E., and Ratain, M.J. (1997). Pharmacokinetic and pharmacodynamic evaluation of the topoisomerase inhibitor irinotecan in cancer patients. *J. Clin. Oncol.* 15, 1502-1510.
- Haaz, M.C., Riche, C., Rivory, L.P., and Robert, J. (1998). Biosynthesis of an aminopiperidino metabolite of irinotecan [7-ethyl-10-[4-(1-piperidino)-1-piperidino]carbonyloxycamptothecin] by human hepatic microsomes. *Drug Metab Dispos.* 26, 769-774.
- Hanioka, N., Ozawa, S., Jinno, H., Ando, M., Saito, Y., and Sawada, J. (2001). Human liver UDP-glucuronosyltransferase isoforms involved in the glucuronidation of 7-ethyl-10-hydroxycamptothecin. *Xenobiotica*. 31, 687-699.
- Hanioka, N., Ozawa, S., Jinno, H., Tanaka-Kagawa, T., Nishimura, T., Ando, M., and Sawada, J.J. (2002). Interaction of irinotecan (CPT-11) and its active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38) with human cytochrome P450 enzymes. *Drug Metab Dispos.* 30, 391-396.
- Hoff, P.M., Ansari, R., Batist, G., Cox, J., Kocha, W., Kuperminc, M., Maroun, J., Walde, D., Weaver, C., Harrison, E., Burger, H.U., Osterwalder, B., Wong, A.O., and Wong, R. (2001). Comparison of oral capecitabine versus intravenous fluorouracil plus leucovorin as first-line treatment in 605 patients with metastatic colorectal cancer: results of a randomized phase III study. *J. Clin. Oncol.* 19, 2282-2292.
- Hosokawa, M. (2008). Structure and catalytic properties of carboxylesterase isozymes involved in metabolic activation of prodrugs. *Molecules*. 13, 412-431.



- Hotelier,T., Renault,L., Cousin,X., Negre,V., Marchot,P., and Chatonnet,A. (2004). ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. *Nucleic Acids Res.* 32, D145-D147.
- Humerickhouse,R., Lohrbach,K., Li,L., Bosron,W.F., and Dolan,M.E. (2000). Characterization of CPT-11 hydrolysis by human liver carboxylesterase isoforms hCE-1 and hCE-2. *Cancer Res.* 60, 1189-1192.
- Ichikawa,W., Uetake,H., Shiota,Y., Yamada,H., Nishi,N., Nihei,Z., Sugihara,K., and Hirayama,R. (2003). Combination of dihydropyrimidine dehydrogenase and thymidylate synthase gene expressions in primary tumors as predictive parameters for the efficacy of fluoropyrimidine-based chemotherapy for metastatic colorectal cancer. *Clin. Cancer Res.* 9, 786-791.
- Innocenti,F., Undevia,S.D., Iyer,L., Chen,P.X., Das,S., Kocherginsky,M., Karrison,T., Janisch,L., Ramirez,J., Rudin,C.M., Vokes,E.E., and Ratain,M.J. (2004). Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. *J. Clin. Oncol.* 22, 1382-1388.
- Jansen,W.J., Zwart,B., Hulscher,S.T., Giaccone,G., Pinedo,H.M., and Boven,E. (1997). CPT-11 in human colon-cancer cell lines and xenografts: characterization of cellular sensitivity determinants. *Int. J. Cancer.* 70, 335-340.
- Jinno,H., Tanaka-Kagawa,T., Hanioka,N., Saeki,M., Ishida,S., Nishimura,T., Ando,M., Saito,Y., Ozawa,S., and Sawada,J. (2003). Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38), an active metabolite of irinotecan (CPT-11), by human UGT1A1 variants, G71R, P229Q, and Y486D. *Drug Metab Dispos.* 31, 108-113.
- Kehrer,D.F., Yamamoto,W., Verweij,J., de Jonge,M.J., de,B.P., and Sparreboom,A. (2000). Factors involved in prolongation of the terminal disposition phase of SN-38: clinical and experimental studies. *Clin. Cancer Res.* 6, 3451-3458.
- Khanna,R., Morton,C.L., Danks,M.K., and Potter,P.M. (2000). Proficient metabolism of irinotecan by a human intestinal carboxylesterase. *Cancer Res.* 60, 4725-4728.
- Kim,S.R., Nakamura,T., Saito,Y., Sai,K., Nakajima,T., Saito,H., Shirao,K., Minami,H., Ohtsu,A., Yoshida,T., Saijo,N., Ozawa,S., and Sawada,J. (2003). Twelve novel single nucleotide polymorphisms in the CES2 gene encoding human carboxylesterase 2 (hCE-2). *Drug Metab Pharmacokinet.* 18, 327-332.
- Kocakova,I., Svoboda,M., Kubsova,K., Chrenko,V., Roubalova,E., Krejci,E., Sefr,R., Slampa,P., Frgala,T., and Zaloudik,J. (2007). Preoperative radiotherapy and concomitant capecitabine treatment induce thymidylate synthase and thymidine phosphorylase mRNAs in rectal carcinoma. *Neoplasma.* 54, 447-453.
- Koizumi,W., Okayasu,I., Hyodo,I., Sakamoto,J., and Kojima,H. (2008). Prediction of the effect of capecitabine in gastric cancer by immunohistochemical staining of thymidine phosphorylase and dihydropyrimidine dehydrogenase. *Anticancer Drugs.* 19, 819-824.

- Kroetz,D.L., McBride,O.W., and Gonzalez,F.J. (1993). Glycosylation-dependent activity of baculovirus-expressed human liver carboxylesterases: cDNA cloning and characterization of two highly similar enzyme forms. *Biochemistry* 32, 11606-11617.
- Kubo,T., Kim,S.R., Sai,K., Saito,Y., Nakajima,T., Matsumoto,K., Saito,H., Shirao,K., Yamamoto,N., Minami,H., Ohtsu,A., Yoshida,T., Saijo,N., Ohno,Y., Ozawa,S., and Sawada,J. (2005). Functional characterization of three naturally occurring single nucleotide polymorphisms in the CES2 gene encoding carboxylesterase 2 (HCE-2). *Drug Metab Dispos.* 33, 1482-1487.
- Lankisch,T.O., Vogel,A., Eilermann,S., Fiebler,A., Krone,B., Barut,A., Manns,M.P., and Strassburg,C.P. (2005). Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. *Mol. Pharmacol.* 67, 1732-1739.
- Liu,L.F., Desai,S.D., Li,T.K., Mao,Y., Sun,M., and Sim,S.P. (2000). Mechanism of action of camptothecin. *Ann. N. Y. Acad. Sci.* 922:1-10., 1-10.
- Long,R.M., Satoh,H., Martin,B.M., Kimura,S., Gonzalez,F.J., and Pohl,L.R. (1988). Rat liver carboxylesterase: cDNA cloning, sequencing, and evidence for a multigene family. *Biochem. Biophys. Res. Commun.* 156, 866-873.
- Marsh,S., Xiao,M., Yu,J., Ahluwalia,R., Minton,M., Freimuth,R.R., Kwok,P.Y., and McLeod,H.L. (2004). Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics.* 84, 661-668.
- Mathijssen,R.H., de Jong,F.A., van Schaik,R.H., Lepper,E.R., Friberg,L.E., Rietveld,T., de,B.P., Graveland,W.J., Figg,W.D., Verweij,J., and Sparreboom,A. (2004). Prediction of irinotecan pharmacokinetics by use of cytochrome P450 3A4 phenotyping probes. *J. Natl. Cancer Inst.* 96, 1585-1592.
- Mathijssen,R.H., van Alphen,R.J., Verweij,J., Loos,W.J., Nooter,K., Stoter,G., and Sparreboom,A. (2001). Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). *Clin. Cancer Res.* 7, 2182-2194.
- Matlin,A.J., Clark,F., and Smith,C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386-398.
- Mauritz,R., van Groeningen,C.J., Smid,K., Jansen,G., Pinedo,H.M., and Peters,G.J. (2007). Thymidylate synthase and dihydropyrimidine dehydrogenase mRNA expression after administration of 5-fluorouracil to patients with colorectal cancer. *Int. J. Cancer.* 120, 2609-2612.
- McLeod,H.L. and Watters,J.W. (2004). Irinotecan pharmacogenetics: is it time to intervene? *J. Clin. Oncol.* 22, 1356-1359.

Metzger,R., Danenberg,K., Leichman,C.G., Salonga,D., Schwartz,E.L., Wadler,S., Lenz,H.J., Groshen,S., Leichman,L., and Danenberg,P.V. (1998). High basal level gene expression of thymidine phosphorylase (platelet-derived endothelial cell growth factor) in colorectal tumors is associated with nonresponse to 5-fluorouracil. *Clin. Cancer Res.* *4*, 2371-2376.

Miwa,M., Ura,M., Nishida,M., Sawada,N., Ishikawa,T., Mori,K., Shimma,N., Umeda,I., and Ishitsuka,H. (1998). Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *Eur. J. Cancer.* *34*, 1274-1281.

Monaghan,G., Ryan,M., Seddon,R., Hume,R., and Burchell,B. (1996). Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. *Lancet.* *347*, 578-581.

Nielsen,H., Engelbrecht,J., Brunak,S., and von,H.G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* *10*, 1-6.

Nilsen,T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays.* *25*, 1147-1149.

Pavillard,V., Charasson,V., Laroche-Clary,A., Soubeyran,I., and Robert,J. (2004). Cellular parameters predictive of the clinical response of colorectal cancers to irinotecan. A preliminary study. *Anticancer Res.* *24*, 579-585.

Pham,V.C., Henzel,W.J., and Lill,J.R. (2005). Rapid on-membrane proteolytic cleavage for Edman sequencing and mass spectrometric identification of proteins. *Electrophoresis.* *26*, 4243-4251.

Pindel,E.V., Kedishvili,N.Y., Abraham,T.L., Brzezinski,M.R., Zhang,J., Dean,R.A., and Bosron,W.F. (1997). Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. *J. Biol. Chem.* *272*, 14769-14775.

Quinney, S. K. Activation of Anticancer Prodrugs by Human Carboxylesterases. 2004. Purdue University.  
Ref Type: Thesis/Dissertation

Quinney,S.K., Sanghani,S.P., Davis,W.I., Hurley,T.D., Sun,Z., Murry,D.J., and Bosron,W.F. (2005). Hydrolysis of capecitabine to 5'-deoxy-5-fluorocytidine by human carboxylesterases and inhibition by loperamide. *J. Pharmacol. Exp. Ther.* *313*, 1011-1016.

Ramchandani,R.P., Wang,Y., Booth,B.P., Ibrahim,A., Johnson,J.R., Rahman,A., Mehta,M., Innocenti,F., Ratain,M.J., and Gobburu,J.V. (2007). The Role of SN-38 Exposure, UGT1A1\*28 Polymorphism, and Baseline Bilirubin Level in Predicting Severe Irinotecan Toxicity. *J. Clin. Pharmacol.* *47*, 78-86.

- Ratain,M.J. (2000). Insights into the pharmacokinetics and pharmacodynamics of irinotecan. *Clin. Cancer Res.* 6, 3393-3394.
- Robbi,M. and Beaufay,H. (1991). The COOH terminus of several liver carboxylesterases targets these enzymes to the lumen of the endoplasmic reticulum. *J. Biol. Chem.* 266, 20498-20503.
- Rougier,P., Van,C.E., Bajetta,E., Niederle,N., Possinger,K., Labianca,R., Navarro,M., Morant,R., Bleiberg,H., Wils,J., Awad,L., Herait,P., and Jacques,C. (1998). Randomised trial of irinotecan versus fluorouracil by continuous infusion after fluorouracil failure in patients with metastatic colorectal cancer. *Lancet.* 352, 1407-1412.
- Rouits,E., Boisdron-Celle,M., Dumont,A., Guerin,O., Morel,A., and Gamelin,E. (2004). Relevance of different UGT1A1 polymorphisms in irinotecan-induced toxicity: a molecular and clinical study of 75 patients. *Clin. Cancer Res.* 10, 5151-5159.
- Rubie,C., Kempf,K., Hans,J., Su,T., Tilton,B., Georg,T., Brittner,B., Ludwig,B., and Schilling,M. (2005). Housekeeping gene variability in normal and cancerous colorectal, pancreatic, esophageal, gastric and hepatic tissues. *Mol. Cell Probes.* 19, 101-109.
- Sai,K., Kaniwa,N., Ozawa,S., and Sawada,J.I. (2001). A new metabolite of irinotecan in which formation is mediated by human hepatic cytochrome P-450 3A4. *Drug Metab Dispos.* 29, 1505-1513.
- Salonga,D., Danenberg,K.D., Johnson,M., Metzger,R., Groshen,S., Tsao-Wei,D.D., Lenz,H.J., Leichman,C.G., Leichman,L., Diasio,R.B., and Danenberg,P.V. (2000). Colorectal tumors responding to 5-fluorouracil have low gene expression levels of dihydropyrimidine dehydrogenase, thymidylate synthase, and thymidine phosphorylase. *Clin. Cancer Res.* 6, 1322-1327.
- Saltz,L.B., Cox,J.V., Blanke,C., Rosen,L.S., Fehrenbacher,L., Moore,M.J., Maroun,J.A., Ackland,S.P., Locker,P.K., Pirotta,N., Elfring,G.L., and Miller,L.L. (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N. Engl. J. Med.* 343, 905-914.
- Sanghani,S.P., Quinney,S.K., Fredenburg,T.B., Davis,W.I., Murry,D.J., and Bosron,W.F. (2004). Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. *Drug Metab Dispos.* 32, 505-511.
- Sanghani,S.P., Quinney,S.K., Fredenburg,T.B., Sun,Z., Davis,W.I., Murry,D.J., Cummings,O.W., Seitz,D.E., and Bosron,W.F. (2003). Carboxylesterases expressed in human colon tumor tissue and their role in CPT-11 hydrolysis. *Clin. Cancer Res.* 9, 4983-4991.

- Sanghani, S. P., Sanghani, P. C., Schiel, M. A., and Bosron, W. F. Human Carboxylesterases; an update on CES1, CES2 and CES3. *Protein & Peptide Letters* . Accepted for publication 2008.
- Santos,A., Zanetta,S., Cresteil,T., Deroussent,A., Pein,F., Raymond,E., Vernillet,L., Risse,M.L., Boige,V., Gouyette,A., and Vassal,G. (2000). Metabolism of irinotecan (CPT-11) by CYP3A4 and CYP3A5 in humans. *Clin. Cancer Res.* 6, 2012-2020.
- Satoh,T. and Hosokawa,M. (1998). The mammalian carboxylesterases: from molecules to functions. *Annu. Rev. Pharmacol. Toxicol.* 38:257-88., 257-288.
- Satoh,T. and Hosokawa,M. (2006). Structure, function and regulation of carboxylesterases. *Chem. Biol. Interact.* 162, 195-211.
- Satoh,T., Taylor,P., Bosron,W.F., Sanghani,S.P., Hosokawa,M., and La Du,B.N. (2002). Current progress on esterases: from molecular structure to function. *Drug Metab Dispos.* 30, 488-493.
- Schiel,M.A., Green,S.L., Davis,W.I., Sanghani,P.C., Bosron,W.F., and Sanghani,S.P. (2007). Expression and characterization of a human carboxylesterase 2 splice variant. *J. Pharmacol. Exp. Ther.* 323, 94-101.
- Schuller,J., Cassidy,J., Dumont,E., Roos,B., Durston,S., Banken,L., Utoh,M., Mori,K., Weidekamm,E., and Reigner,B. (2000). Preferential activation of capecitabine in tumor following oral administration to colorectal cancer patients. *Cancer Chemother. Pharmacol.* 45, 291-297.
- Schwer,H., Langmann,T., Daig,R., Becker,A., Aslanidis,C., and Schmitz,G. (1997). Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. *Biochem. Biophys. Res. Commun.* 233, 117-120.
- Senter,P.D., Beam,K.S., Mixan,B., and Wahl,A.F. (2001). Identification and activities of human carboxylesterases for the activation of CPT-11, a clinically approved anticancer drug. *Bioconj. Chem.* 12, 1074-1080.
- Singer,M. and Berg,P. (2004). George Beadle: from genes to proteins. *Nat. Rev. Genet.* 5, 949-954.
- Sreerama,N. and Woody,R.W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* 209, 32-44.
- Sreerama,N. and Woody,R.W. (2004). On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* 13, 100-112.
- Stamm,S., Ben-Ari,S., Rafalska,I., Tang,Y., Zhang,Z., Toiber,D., Thanaraj,T.A., and Soreq,H. (2005). Function of alternative splicing. *Gene.* 344:1-20. *Epub; %2004 Dec 10.*, 1-20.

- Sun,Z., Murry,D.J., Sanghani,S.P., Davis,W.I., Kedishvili,N.Y., Zou,Q., Hurley,T.D., and Bosron,W.F. (2004). Methylphenidate is stereoselectively hydrolyzed by human carboxylesterase CES1A1. *J. Pharmacol. Exp. Ther.* *310*, 469-476.
- Takagi,Y., Morohashi,K., Kawabata,S., Go,M., and Omura,T. (1988). Molecular cloning and nucleotide sequence of cDNA of microsomal carboxylesterase E1 of rat liver. *J. Biochem.* *104*, 801-806.
- Takahashi,H., Maeda,Y., Watanabe,K., Taguchi,K., Sasaki,F., and Todo,S. (2000). Correlation between elevated intratumoral thymidine phosphorylase and prognosis of node-positive breast carcinoma undergoing adjuvant doxifluridine treatment. *Int. J. Oncol.* *17*, 1205-1211.
- Takasuna,K., Hagiwara,T., Hirohashi,M., Kato,M., Nomura,M., Nagai,E., Yokoi,T., and Kamataki,T. (1996). Involvement of beta-glucuronidase in intestinal microflora in the intestinal toxicity of the antitumor camptothecin derivative irinotecan hydrochloride (CPT-11) in rats. *Cancer Res.* *56*, 3752-3757.
- Thierry-Mieg,D. and Thierry-Mieg,J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* *7 Suppl 1:S12.1-14*. *Epub;2006 Aug 7.*, S12-S14.
- Tobin,P., Clarke,S., Seale,J.P., Lee,S., Solomon,M., Aulds,S., Crawford,M., Gallagher,J., Evers,T., and Rivory,L. (2006). The in vitro metabolism of irinotecan (CPT-11) by carboxylesterase and beta-glucuronidase in human colorectal tumours. *Br. J. Clin. Pharmacol.* *62*, 122-129.
- Tukey,R.H., Strassburg,C.P., and Mackenzie,P.I. (2002). Pharmacogenomics of human UDP-glucuronosyltransferases and irinotecan toxicity. *Mol. Pharmacol.* *62*, 446-450.
- Twelves,C., Boyer,M., Findlay,M., Cassidy,J., Weitzel,C., Barker,C., Osterwalder,B., Jamieson,C., and Hieke,K. (2001). Capecitabine (Xeloda) improves medical resource use compared with 5-fluorouracil plus leucovorin in a phase III trial conducted in patients with advanced colorectal carcinoma. *Eur. J. Cancer* *37*, 597-604.
- Venables,J.P. (2004). Aberrant and alternative splicing in cancer. *Cancer Res.* *64*, 7647-7654.
- Venables,J.P. (2006). Unbalanced alternative splicing and its significance in cancer. *Bioessays.* *28*, 378-386.
- von,H.G. (1983). Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* *133*, 17-21.
- Walko,C.M. and Lindley,C. (2005). Capecitabine: a review. *Clin. Ther.* *27*, 23-44.

- Wu,M.H., Chen,P., Remo,B.F., Cook,E.H., Jr., Das,S., and Dolan,M.E. (2003). Characterization of multiple promoters in the human carboxylesterase 2 gene. *Pharmacogenetics*. *13*, 425-435.
- Wu,M.H., Chen,P., Wu,X., Liu,W., Strom,S., Das,S., Cook,E.H., Jr., Rosner,G.L., and Dolan,M.E. (2004). Determination and analysis of single nucleotide polymorphisms and haplotype structure of the human carboxylesterase 2 gene. *Pharmacogenetics*. *14*, 595-605.
- Wu,M.H., Yan,B., Humerickhouse,R., and Dolan,M.E. (2002). Irinotecan activation by human carboxylesterases in colorectal adenocarcinoma cells. *Clin. Cancer Res.* *8*, 2696-2700.
- Xie,M., Yang,D., Liu,L., Xue,B., and Yan,B. (2002). Human and rodent carboxylesterases: immunorelatedness, overlapping substrate specificity, differential sensitivity to serine enzyme inhibitors, and tumor-related expression. *Drug Metab Dispos.* *30*, 541-547.
- Xing,Y. (2007). Genomic analysis of RNA alternative splicing in cancers. *Front Biosci.* *12*:4034-41., 4034-4041.
- Xu,G., Zhang,W., Ma,M.K., and McLeod,H.L. (2002). Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with activation of irinotecan. *Clin. Cancer Res.* *8*, 2605-2611.
- Zhang,J., Feuk,L., Duggan,G.E., Khaja,R., and Scherer,S.W. (2006). Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet. Genome Res.* *115*, 205-214.

## CURRICULUM VITAE

Marissa Ann Schiel

### Education

Indiana University School of Medicine, Indianapolis, IN

- Doctor of Medicine, May 2009

Indiana University, Indianapolis, IN

- Doctor of Philosophy in Biochemistry and Molecular Biology, February 2009

University of Illinois, Urbana-Champaign, IL

- Bachelor of Science in Biochemistry, 2002
- Bachelor of Science in Life Sciences, Major Honors Biology, 2002
- Graduated *magna cum laude* and *Phi Beta Kappa* with University Bronze Tablet Distinction

### Research

Biochemistry Ph.D. Student, Laboratory of Dr. William F. Bosron, Ph.D.

Indiana University, Indianapolis, IN 2004-2007

- Thesis: *Human Carboxylesterase 2 Splice Variants: Expression, Activity, and Role in the Metabolism of Irinotecan and Capecitabine*

Undergraduate Researcher, Laboratory of Dr. George Ordal, Ph.D.

University of Illinois, Urbana-Champaign, IL 1999-2002

- Researched the role of the McpB chemoreceptor in *Bacillus subtilis* chemotaxis

Research Intern, Hospital Products Division

Abbott Laboratories, Abbott Park, IL Summer 2001

- Performed hematology related research

### Leadership

Crispus Attucks Student Interest Group

Secretary and Chair of the Constitution Committee 2007-2009

- The mission of the Crispus Attucks Student Interest Group is to expose, educate, and encourage high school students to pursue careers in medicine and healthcare related fields.

Combined Degree Student Committee

Representative 2005 – 2009, Treasurer 2005-2006

- The committee serves as an informal student council for the MSTP program.
- Responsibilities include inviting guest lecturers, collecting resources for students, hosting a website, and organizing volunteer and social activities.



## **Publications**

**Schiel MA**, Green SL, Davis WI, Sanghani PC, Bosron WF, Sanghani SP. Expression and Characterization of a Human Carboxylesterase 2 Splice Variant. *Journal of Pharmacology and Experimental Therapeutics*. 2007 Oct; 323(1): 94-101.

Zimmer MA, Szurmant H, Saulmon MM, **Collins MA**, Bant JS, Ordal GW. The role of heterologous receptors in McpB-mediated signalling in *Bacillus subtilis* chemotaxis. *Molecular Microbiology*. 2002 Jul; 45(2): 555-568.

Zimmer MA, Tiu J, **Collins MA**, Ordal GW. Selective methylation changes on the *Bacillus subtilis* chemotaxis receptor McpB promote adaptation. *Journal of Biological Chemistry*. 2000 Aug; 275(32): 24264-24272.

## **Posters**

Chiorean, E.G., Sanghani, S., Cardenes, H., **Schiel, M.**, Yu, M., Shen, J., LeBlanc, J., Robb, B., Loehrer, P.J., Harvey, J. (2008, January). *Phase II trial of neoadjuvant capecitabine plus irinotecan followed by combined modality capecitabine and radiation therapy for patients with locally advanced rectal cancer: A Hoosier Oncology Group study*. Poster displayed at: American Society of Clinical Oncology 2008 Gastrointestinal Cancers Symposium; Orlando, FL.

**Schiel, M.A.**, Davis, W.I., Fears, S.L., Green, S., Zhai, LM, Sanghani, P.C., Bosron, W.F., and Sanghani, S.P. (2008, May). *Splice Variants of Human Carboxylesterase 2*. Poster displayed at: Great Lakes Drug Metabolism Discussion Group Third Annual Meeting; Indianapolis, IN

**Schiel, M.A.**, Sanghani, S.P., Davis, W.I., and Bosron, W.F. (2006, January). *Characterization of Carboxylesterase 2 splice variants and the treatment of Colorectal Cancer*. Poster presented at: AACR Special Conference New Developments in the Epidemiology of Cancer Prognosis: Traditional and Molecular Predictors of Treatment Response and Survival; Charleston, SC.

## **Memberships**

- American College of Physicians
- Phi Beta Kappa Honor Society
- American Association for Cancer Research, 2006
- Phi Eta Sigma Honor Society
- Alpha Lambda Delta National Academic Honor Society

## **Honors and Awards**

- Scholar-in-Training Award for AACR Special Conference, January 2006
- Dean's List of the University of Illinois
- Edmund J. James Scholar, University of Illinois, 1998-2002
- Robert C. Byrd Honors Scholarship, 1998-2002
- Academic Achievement Scholarship, 1998-2002

**Funding**

- Indiana Genomics Initiative 2002-2004, 2007-2009
- Translational Research Fellowship, 2006-2007
- DeVault Endowment, 2005-2006