Characterization of a novel truncated isoform of murine AGPAT4 arising from predicted splice variants

by

Chia Chun Joey Hung

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Kinesiology

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The acylglycerophosphate acyltransferase (AGPAT)/lysophosphatidic acid acyltransferase (LPAAT) family of enzymes is responsible for the conversion of lysophosphatidic acid (LPA) to phosphatidic acid (PA), a precursor of *de novo* triacylglycerol and glycerophospholipid synthesis. Specifically, a fatty acyl group is attached to the *sn-2* position of LPA to form PA. There are eleven AGPAT/LPAAT enzymes identified in humans and mice. Murine AGPAT4/LPAATδ is an outer mitochondrial membrane (OMM) protein, which is highly expressed in whole brain and brain subregions and is linked to phospholipid regulation, learning and memory function, as well as muscle force contractility and adipose tissue structure. The murine AGPAT4 reference protein variant NP_080920.2 has been experimentally characterized, while a truncated form of the AGPAT4 enzyme (generated by the coding regions of at least two different predicted splice variant transcripts, XM_006523348.4 and XM_006523347.3) was predicted to exist through the gene prediction tool Gnomon but had not yet been studied. This thesis reports the first identification and quantification of the predicted splice variants of murine *Agpat4* in various tissues, and reports a heterogenous expression of the variants among the tissues studied, although in all tissues, the reference transcript was much more abundant. A novel cloning vector expressing the truncated AGPAT4 protein variant with a hemagglutinin tag was generated using molecular cloning for use in protein-protein interaction studies. Protein pull-down assays were performed to investigate if the truncated AGPAT4 protein variant directly interacts with the reference AGPAT4 protein, but did not yield evidence of an interaction. Finally, a series of scaled co-transfection experiments for reference AGPAT4 and truncated AGPAT4 were performed to determine whether the truncated AGPAT4 variant plays a role in modulating the levels of the reference AGPAT4 variant. Analyses of immunoblots revealed that the reference AGPAT4 protein modulated levels of the truncated

AGPAT4 variant, but expression of the truncated protein variant did not significantly modulate the reference AGPAT4 protein *in vitro.* In short, this thesis has determined *Agpat4* splice variants X1, X2, and X3 are endogenously synthesized in various murine tissues, and that truncated AGPAT4 protein levels may be modulated by the co-expression of reference AGPAT4 protein, *in vitro.*

Acknowledgements

I would first like to thank my supervisor, Dr. Robin Duncan, from the bottom of my heart for her undying support and guidance. I have learned much from you, not only in our field of research but as an individual as well. I definitely would not be where I am and who I am today without your efforts.

Secondly, I would like to thank my committee members Dr. Ken Stark and Dr. Joe Quadrilatero for their time and insightful comments provided during my time working on this thesis. I never expected myself to be presenting and discussing two years' worth of work to you when I sat in your Undergraduate classes. I would also like to thank my fellow lab members Dr. Fernanda Fernandes, Ashkan Hashemi, John Chan, Michelle Tomczewski, and Kalsha De Silva for their support in my research.

Finally, I want to thank my mother, her partner, and my family for their constant care and (attempted) understanding. Even though you all never really understood what running a gel meant, you were always there for me. Also, thank you to an endless list of people that have been with me since day one. I will not be able to name any specific names and the antics we did, or allegedly did; the result of that might end up longer than this thesis.

I have learned a lot from my time as a master's student, research techniques aside. You will have a bad day every now and then but try to go for a walk and think about the nice things, no matter how small they might be. Sometimes it do be a 10 on the pain scale, but don't forget that there's still cool things to do and people to see!

Dedication

I would like to dedicate this thesis to my grandmother and my good friend Eric Yeung. Thank you both for your lessons, rest in peace.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The acylglycerophosphate acyltransferases (AGPAT) are a family of 11 related enzymes that play critical roles in the endogenous synthesis of glycerolipids and glycerophospholipids [1, 2]. Their physiological roles differ between tissues such as the brain, liver, adipose and skeletal muscle [3]. The importance of AGPAT activity in varying tissue types and the diversity of physiological roles that they play is further emphasized in *Agpat* gene ablation studies, where consequences range from impaired learning and spatial awareness, to seizures, hepatosteatosis, and impaired muscle force contractility [3-5]. In recent literature, including studies from our laboratory, AGPATs 1-5 have been found to display true lysophosphatidic acid acyltransferase (LPAAT) activity, preferring the use of lysophosphatidic acid (LPA) as the acyl-acceptor substrate, and are thus named LPAATs α-ε, respectively [3]. AGPATs 6-11 show differing substrate specificities, preferring other lysophospholipids and glycerol-3-phosphate (G3P) instead of LPA as the primary acyl-acceptor. Therefore, AGPATs 6-11 have largely been renamed to specific lysophospholipid acyltransferases (LPLATs) and glycerol-3-phosphate acyltransferases (GPATs) [3].

The reference AGPAT4/LPAATδ protein (ref. AGPAT4) in murine models has been characterized by previous members of the Duncan Lab as well as other external labs. It was experimentally determined that ref. AGPAT4 is present endogenously in a multitude of tissues and highly expressed in the brain, heart, and skeletal muscle of murine models [2, 6]. As such, ref. AGPAT4 displays physiological roles within those tissues, which will be discussed in the next chapter. After successfully determining physiological roles and characteristics of ref. AGPAT4, past Duncan Lab members subsequently began to investigate predicted splice variants. Prior lab

members generated a novel amplicon by reverse transcriptase-polymerase chain reaction (RT-PCR) by designing a primer within the protein coding region that is shared by two predicted *Agpat4* mRNA sequences, predicted splice variants X2 and X3. The genetic sequences used to design the primers were derived from accession numbers XM_006523348.4 (Variant X2) and XM_006523347.3 (Variant X3) on the Nucleotide database of the National Centre for Biotechnology Information. Genetic splice variants are of interest in the scientific world due to the diverse and often physiologically significant roles they may have in different tissue subsets [7]. Finding an endogenously synthesized truncated variant of AGPAT4 was both curious and plausibly significant, which leads into the overarching aim of this thesis.

Due to the important role that AGPAT4 plays in the *de novo* synthesis of glycerophospholipids [1], this thesis aims to perform an initial characterization of the protein translated from the coding region of the *Agpat4* X2 and X3 genetic variants. This work begins by determining the existence and relative tissue expression levels of the reference *Agpat4* transcript and the predicted *Agpat4* splice variants X1, X2, and X3, by reverse transcriptase real-time quantitative polymerase chain reaction (RT-qPCR) using unique N-terminal primer sequences designed to distinguish the variants. Attempts to elucidate the function of truncated AGPAT4 involved protein pull-down assays via immobilized metal affinity chromatography (IMAC) to determine if truncated AGPAT4 formed a protein complex with the reference AGPAT4 protein. A unique expression vector coding for a chimeric truncated AGPAT4 tagged with hemagglutinin was generated through molecular cloning to allow for co-immunoprecipitation assays as a second method to test for interactions. Finally, graded cell transfections using expression vectors for both reference AGPAT4 and truncated AGPAT4, each with unique protein affinity tags was conducted,

and lysates were immunoblotted to determine if the truncated AGPAT4 protein variant was involved in modulating reference AGPAT4 protein levels *in vitro* and *vice versa.*

Chapter 2

Biochemical Foundations

2.1 Glycerolipid and Glycerophospholipid Synthesis

Glycerolipids and glycerophospholipids are synthesized *de novo* through the actions of GPATs, LPAATs, and LPLATs [8, 9]. Their synthesis begins with the glycerol phosphate pathway, utilizing the molecule glycerol-3-phosphate (G3P), which is acylated typically at the *sn-1* position by a GPAT homologue using a fatty acyl-CoA, creating LPA [8-10]. These enzymes show little specificity for different fatty acyl species and tend to utilize fatty acyl-CoAs based on abundance, which results in the predominant incorporation of saturated fatty acids (SFA) and monounsaturated fatty acids (MUFA) at the *sn-1* position. LPA is further converted into phosphatidic acid (PA) by an LPAAT, which typically acylates the *sn-2* position, using specific fatty acyl CoA groups listed in Table 1 [8, 9]. To clarify, the donor species listed in Table 1 are known acyl donors identified through experimental studies, but the full spectrum of possible fatty acyl CoA donor species has not yet been fully defined. PA is an important molecule, since it is the common precursor to the phospholipids phosphatidylglycerol (PG), phosphatidylinositol (PI), and cardiolipin (CL) [8, 10]. Furthermore, PA can be converted into diacylglycerols (DAGs) through a dephosphorylation reaction catalyzed by phosphatidic acid phosphatase (PAP) enzymes, also known as Lipins [9, 11]. DAGs can either be converted to triacylglycerols (TAGs) by a member of the diacylglycerol acyltransferase (DGAT) family [9, 12], or can be converted to phosphatidylcholine (PC), phosphatidylethanolamine (PE), or phosphatidylserine (PS) via the CDP-choline or CDP-ethanolamine branches of the Kennedy Pathway [9, 13].

2.2 Land's Phospholipid Remodeling Cycle

Following the *de novo* synthesis of glycerophospholipids via the Kennedy pathway, phospholipid remodeling occurs through the Land's cycle [14-16]. This alteration arises through the actions of phospholipase and acyltransferase enzymes, with most remodeling occurring at the *sn-2* position as a result of the action of phospholipase A_2 enzymes, which cleave a fatty acyl chain at this site, and members of the LPLAT family of enzymes that catalyze reacylation [16]. Due to this remodeling process, tissues in eukaryotic organisms are able to achieve unique and characteristic phospholipid profiles. For example, more mature phospholipids are found to carry arachidonic acid or stearic acid in their *sn-2* positions, while newer molecules of phosphatidic acid typically do not display the same profiles [17]. It is also important in shaping the general characteristic of phospholipids in the body. While the *sn-1* positions of phospholipids are typically found esterified with saturated or monounsaturated fatty acyl groups arising from *de novo* synthesis by GPATs, the *sn-2* position of cellular phospholipids is often more highly enriched in polyunsaturated fatty acyl chains, due to the process of remodeling that occurs to a greater extent at the *sn-2* position [16, 18].

2.3 Glycerol-3-Phosphate Acyltransferases

In the Kennedy Pathway, GPATs are the first group of enzymes that catalyze the reaction to convert G3P to LPA by esterifying G3P at the *sn-1* position [8, 9]. GPATs are recognized as the rate limiting step in the *de novo* synthesis of glycerolipids and glycerophospholipids, and they display the lowest specific activity of all the acyltransferases in the Kennedy Pathway [9, 19, 20]. To date, there are four identified GPATs, numbered from 1-4. GPATs 1 & 2 are mitochondrial enzymes while GPATs 3 & 4 are microsomal [1, 9]. GPAT1 displays preference towards utilizing saturated fatty acyl-CoA groups as donors to acylate G3P, but GPATs 2-4 do not show any

particular preference to saturated acyl donor species [8]. GPATs 3 & 4 were originally classified as AGPAT8 (LPAATθ) and AGPAT6 (LPAATζ) respectively. Functional studies showed that neither of these enzymes exhibited AGPAT activity, but rather GPAT activity. The four GPATs share the distinct HXXXXD acyltransferase activity sequence found in Motif I with LPAATs, LPCATs, LPEATs, and ALCAT1 [1, 8].

2.4 Acylglycerophosphate Acyltransferases

The AGPAT family of enzymes are responsible for catalyzing the second step of the Kennedy Pathway, by acylating the *sn-2* position of lysophosphatidic acid to form phosphatidic acid [1, 3, 8, 21]. Originally, 11 AGPATs were found to exist in both humans and mice but have now been split into AGPATs 1-5 and AGPATs 6-11 [3, 22]. AGPATs 1 through 5 display true LPAAT activity, preferring LPA as substrate, resulting in specific renaming to LPAATs [2, 3].

2.5 AGPAT/LPAAT Fatty-Acyl CoA Donors, Tissue Expression, and Subcellular Localization

Out of the five "true" LPAAT enzymes, AGPAT1/LPAATα has been experimentally shown to utilize saturated fatty acyl-CoA donors, including myristoyl-CoA (14:0), pentadecanoyl-CoA (15:0), palmitoyl-CoA (16:0), and stearoyl-CoA (18:0), to form PA, in addition to using oleoyl-CoA (18:1n-9) and arachidonoyl-CoA (20:4n-6) [3, 22, 23]. AGPAT2 and AGPAT3 both utilize oleoyl-CoA and arachidonoyl-CoA as fatty acyl donors to form PA [23]. Lastly, AGPAT4 and AGPAT5 utilize oleoyl-CoA as fatty acyl donors to esterify LPA, although fatty acyl preference has not yet been fully tested for these homologues [3, 6, 24].

The tissue expression and subcellular localization of AGPATs 1-5 have been investigated in humans and mice. There are similarities between species in expression and localization, likely

reflecting the unique roles that the individual LPAATs play in specific tissues and within cells. Table 1 summarizes the tissue expression of the AGPATs and known respective fatty acyl-CoA donor species usage.

AGPAT1/LPAATα has been shown to be ubiquitously expressed in tissues throughout the body [22, 23]. The expression level of AGPAT1 varies between humans and mice. Murine models show universally high levels of AGPAT1 across diverse tissues including the lungs, heart, brain, liver, spleen, thymus, kidney, stomach, skeletal muscle, and both brown and white adipose tissues [1, 22]. Human expression of AGPAT1 is similar to that of mice in terms of ubiquity, with the exception that AGPAT1 is virtually undetectable in skeletal muscle of humans and has more varying levels of expression between tissues [23]. Humans have the highest level of expression of AGPAT1 in the testis, followed by pancreas and adipose tissue [23]. AGPAT1 has been shown to localize to the endoplasmic reticulum in cells [22, 23].

AGPAT2/LPAATβ shares approximately 34% sequence identity with AGPAT1, but does not display the broad tissue distribution of AGPAT1 [25]. In humans, AGPAT2 has been shown to be more highly restricted to white adipose tissue, followed by pancreas and liver [23]. In mice, AGPAT2 expression is seen largely in adipose tissue, followed by liver and then the kidneys, stomach, and skeletal muscle, with less in the heart and lungs, and a complete absence from the brain [22]. The significance of AGPAT2 in adipose tissue has been demonstrated, since loss of function of AGPAT2 has been linked to a form of congenital generalized lipodystrophy, which causes near complete losses of adipose tissue at birth [25]. Similar to AGPAT1, subcellular localization studies have indicated that AGPAT2 localizes to the endoplasmic reticulum [3].

AGPAT3/LPAATγ was among the first AGPAT enzymes to be characterized [1]. Like AGPAT1, it is found ubiquitously in all tissues, albeit at varying levels [3, 22]. In humans,

AGPAT3 is most highly expressed in the testes, pancreas, and kidneys, followed by the spleen, lungs, liver, and white adipose tissue [24]. Murine tissue expression of AGPAT3 is similar to humans, with the testes also showing the highest expression level [1, 3, 24, 26]. This has been attributed to the involvement of AGPAT3 in the maturation of the testes [26]. Like AGPAT1 and AGPAT2, AGPAT3 localizes to the endoplasmic reticulum [1].

A more detailed discussion of AGPAT4/LPAATδ is provided later in this chapter in a dedicated section. However, briefly, AGPAT4/LPAATδ has more recently been characterized by members of our own lab, as well as researchers in Italy [21] and Japan [27]. AGPAT4 is currently poorly characterized in human models. In murine models, AGPAT4 is most highly expressed in the brain, followed by the heart and skeletal muscle [2, 6]. AGPAT4 has been shown to localize to multiple subcellular regions by different laboratories, with those being the outer mitochondrial membrane [6], *trans*-golgi apparatus [21], and endoplasmic reticulum [27].

AGPAT5/LPAATε expression in humans is highest in the testes, followed by the brain and heart [24]. In murine models, AGPAT5 is also expressed predominantly in the brain, heart, and skeletal muscle tissue [22]. AGPAT5 subcellular localization studies reveal that AGPAT5 is found mainly in the mitochondria and the endoplasmic reticulum [24].

Table 1: Summary of Tissue Expression and known Fatty Acyl-CoA donor species utilized by AGPATs 1-5/LPAATs α-ε

Enzyme	**Tissue Expression**	**Fatty Acyl-CoA Donor Species**	**References**
AGPAT1/LPAATα	Murine: Universal, mostly in testes, liver, and pancreas.	Myristoyl-CoA (14:0) Pentadecanoyl-CoA (15:0) Palmitoyl-CoA (16:0) Stearoyl-CoA (18:0) Oleoyl-CoA (18:1n-9) Arachidonoyl-CoA (20:4n-6)	[1, 3, 22, 23]
	Human: Universal apart from skeletal muscle, mostly in testes, pancreas, and white adipose tissue.		
AGPAT2/LPAATβ	Murine: Largely in White Adipose Tissue, followed by Liver, and Kidney.	Oleoyl-CoA (18:1n-9) Arachidonoyl-CoA (20:4n-6)	[1, 3, 22, 23, 25]
	Human: Largely in White Adipose Tissue, followed by Pancreas, and Liver.		
AGPAT3/LPAATγ	Murine: Universal, mostly in testes, pancreas, and kidney.	Oleoyl-CoA (18:1n-9) Arachidonoyl-CoA (20:4n-6)	[1, 3, 22, 24, 26]
	Human: Universal, mostly in testes, pancreas, and kidney.		
AGPAT4/LPAATδ	Murine: Brain, Heart, and Skeletal Muscle.	Oleoyl-CoA (18:1n-9)	[1-3, 6, 21, 22]
	Human: N/A		
AGPAT5/LPAATε	Murine: Brain, Heart, and Skeletal Muscle.	Oleoyl-CoA (18:1n-9)	[3, 22, 24]
	Human: Testes, Brain, and Heart.		

2.6 AGPAT4/LPAATδ

Several functions and roles of AGPAT4/LPAATδ have been elucidated by members of the Duncan Lab in previous years, both *in vitro* and *in vivo*. AGPAT4 has been characterized to regulate brain PC, PE, and PI levels, have an impact on learning and memory, and affect adipose tissue biology [28] and skeletal muscle fiber type and force contractility [4-6].

Results from both *in vitro* overexpression studies and *in vivo* knockout studies provide evidence of a role for AGPAT4 in regulating phospholipid content in cells and the brain. In vitro assays confirmed that AGPAT4 only has activity as an LPAAT. However, the Duncan Lab has shown that the cellular content of phosphatidylinositol significantly increased by 72% in Sf9 insect cells relative to control cells following AGPAT4 overexpression, while levels of other phospholipids, including PA, did not change significantly [6]. Furthermore, a follow-up murine *Agpat4* knockout study of whole brain phospholipid content showed a complimentary trend: a significant 52% decrease in cellular PI content relative to wild type mice, as well as a 39% decrease in phosphatidylcholine and a 32% decrease in phosphatidylethanolamine [6]. When taken together, these results indicate that although AGPAT4 is a true LPAAT, it functions *in vivo* to support the biosynthesis of various downstream phospholipids.

To further expand on the physiological role of AGAPT4 *in vivo*, the Duncan lab pursued studies focusing on learning and memory, due to the apparent role of AGPAT4 regulation of brain phospholipid content. Using the Morris Water Maze, *Agpat4* knockout mice were evaluated alongside their wild type littermates to determine if differences existed in spatial learning and memory. Results of the study indicated that, indeed, *Agpat4* gene ablation correlated with significantly poorer outcomes in knockout mice performing the Morris Water Maze test relative to wild type littermates [5]. This reduction in cognitive function of *Agpat4* knockout mice was

hypothetically attributed to the decrease of brain PC, PE, and PI content, which was further linked to a decrease in NMDA and AMPA receptor subunits in the brain [5].

In other work, *Agpat4* knockout mice were found to display alterations in muscle properties, namely fiber type alterations and decreases in force contractility. Previous members of our lab and collaborators from our university investigated the role of AGPAT4 in soleus and extensor digitorum longus (EDL), which represent oxidative and glycolytic fiber types, respectively. Using immunohistochemistry, it was determined that *Agpat4* ablated mice had a reduction in type I and type IIA muscle fibers in the glycolytic EDL muscle [4]. Electrical stimulation tests of muscle force contractility revealed significant decreases in soleus contractile force in mice lacking AGPAT4 [4]. These changes in skeletal muscle have not yet been explained fully but have been conjectured to be related to decreasing pyruvate dehydrogenase activity and alterations in skeletal muscle phosphatidic acid and phosphatidylethanolamine content in *Agpat4* knockout models [4].

Agpat4 deficiencies in murine models were also examined for any changes in adipose tissue physiology. In wild type mice, *Agpat4* was found to have relatively consistent expression across various white adipose tissue depots [28]. Ablation of *Agpat4* revealed a heterogeneity between certain white adipose tissue depots, namely the epididymal fat depot in male *Agpat4* knockouts, which displayed a 40% increase in adipose tissue weight relative to wild-type mice [28]. This difference was found to not be caused by alteration in changes of metabolic processes, food intake, or varying activity levels [28]. Rather, it was elucidated that a hypertrophic change in the adipocytes found in the epididymal fat of male *Agpat4* knockouts was the likely source of differences seen [28]. To support this finding, epididymal white adipose tissue depots of *Agpat4* ablated male mice were sectioned and compared to the same depot in wild type mice. The mean

cross-sectional area of *Agpat4*$^{-/-}$ male epididymal white adipose tissue was, indeed, found to be significantly larger in size relative to their control littermates [28]. To explain this phenomenon of increased TAG content, changes in lipolysis were investigated and found to be significantly downregulated in the epididymal fat depots of male *Agpat4* knockout mice. Specifically, the lipolytic enzyme adipose triglyceride lipase (ATGL) was found to be expressed at 52% lower level in *Agpat4* ablated male mice relative to their wild-type littermates [28].

2.7 Acyltransferase Structure

The complete characterization of an enzyme not only requires a physiological approach but should include a molecular component as well. By creating an image of the three-dimensional structure of an enzyme, it becomes possible to link molecular findings with specific mechanistic properties of the enzyme in question [29]. The gold-standard approach for imaging three-dimensional structures of proteins is known as X-ray crystallography, and it is heavily favoured for determining the structures of proteins and other macromolecules [30]. However, the nature of proteins which contain several transmembrane domains makes it difficult to effectively utilize this technique [31]. Members of the acyltransferase family are proposed to contain multiple transmembrane domains [8], and indeed, this property is a significant limitation of crystallization of proteins [31]. Membrane proteins must be extracted from their bilayer environment and be preserved in an external, artificially created lipid environment before the crystallization technique may be applied [31]. Being part of the extensive Membrane-Bound O-Acyltransferase (MBOAT) family of enzymes, GPATs, AGPATs, and LPLATs are therefore subject to this limitation [21]. Despite this difficulty, the first successful three-dimensional structure of a GPAT enzyme was synthesized in 2001 from squash chloroplasts [32]. The model revealed two major structural domains in the tertiary structure of the GPAT, and within Domain

II, a distinct region of hydrophobic amino acid residues ending with positively charged amino acid residues that are flanked by the highly conserved HXXXXD catalytic motif [32]. More recently, a bacterial LPAAT enzyme termed PlsC had its crystal structure successfully synthesized in 2017 [21, 33]. Although PlsC is bacterial, it is proposed by other research groups that due to certain residue similarities, information gathered from the PlsC crystal structure may be extrapolated to mammalian AGPAT4/LPAATδ as preliminary speculation [21, 33]. As a mammalian model of AGPAT4 has yet to be crystallized, a two-dimensional membrane topology has been proposed [21]. It is suggested that AGPAT4 likely contains three distinct transmembrane domains, with the four acyltransferase motifs located in the cytoplasm of the cell [21]. The N-terminus of AGPAT4 is projected to be contained within the mitochondrial lumen while the C-terminus is located in the cytoplasm [21]. The proposed membrane topology of human AGPAT4 is depicted in the following Figure 1.

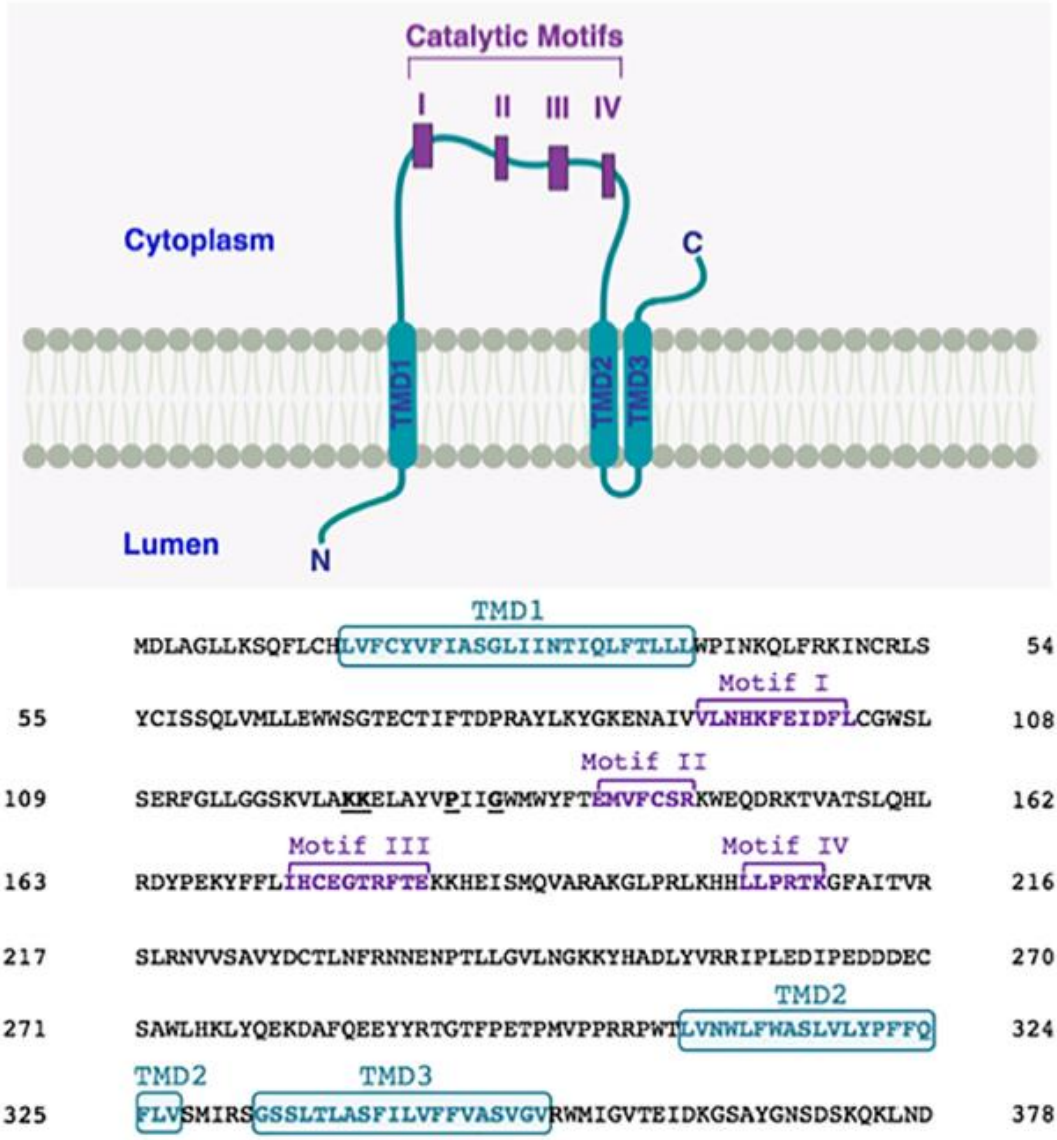


Figure 1: The proposed membrane topology of human AGPAT4 adapted from Zhukovsky et al., (2019) comprising of specific amino acid sequences corresponding to transmembrane domains (TMD) and catalytic motifs, and their projected locations within the cell. Transmembrane domains are anchored within the mitochondrial lumen while the catalytic motifs are suspended in the cytoplasm of the cell.

2.8 Protein Interactions

The concept of proteins forming oligomeric or multimeric complexes is well established as a factor in proper biological function [34]. Following folding into tertiary structures, individual proteins may further combine to form quaternary structures, which are classified as either homocomplexes or heterocomplexes [34]. Homocomplexes are defined as protein-protein interactions that occur between identical (Prefix: homo-) oligomers while heterocomplexes are interactions occurring between non-identical (Prefix: hetero-) oligomers [34]. These complexes may be further defined by either obligate or non-obligate folding structures [34]. Oligomers involved in obligate protein-protein interactions are not found *in vivo* as individual structures on their own due to structural instability, and are thus functionally obligate as well [34]. Oligomers which form non-obligate protein-protein interactions may be found *in vivo* as stable, independent proteins [34]. With regards to the role of accessory proteins stabilizing an overarching protein complex, understanding the stability and folding processes of proteins is crucial to understanding their basic function [35]. Fundamentally, across multiple fields of science, equilibrium and stability are critical and protein structures and their complexes are no different [35]. Optimally, folded proteins and their higher order complexes exist in the lowest free energy state possible, otherwise risking denaturation [35]. So, upon forming either homocomplexes or heterocomplexes, major proteins may go through important conformational changes which allow them to partake in what is termed as “disorder-to-order transitions” [36]. Other than stability, proteins forming complexes serves another purpose: the regulation of function [37]. Of particular interest, with regards to this thesis, is allosteric regulation, defined as the regulation of enzyme function based on conformational change through an interaction outside of the active site [37]. This external interaction may be caused by small molecules such as amino acids, pathway by-

products such as cyclic AMP, and finally by protein-protein interaction [37]. With this knowledge, speculation regarding the function of truncated AGPAT4 regulating reference AGPAT4 via interaction was established and is of interest for potentially characterizing its function.

2.9 Molecular Cloning

Molecular cloning is a vital tool for expressing proteins of interest for study or use in molecular biology [38, 39]. By taking advantage of recombinant DNA molecules, molecular cloning provides researchers with the ability to manipulate gene and protein expression [39]. At its core, molecular cloning involves a cloning vector and a DNA insert [38]. The cloning vector importantly consists of a promotor region to drive gene expression, a Multiple Cloning Site (MCS) flanked by several sequences recognized by restriction enzymes, and often a protein affinity tag for protein purification purposes or a fluorescent protein tag for imaging purposes [38]. DNA inserts are comprised of the researcher's gene of interest and are commonly manipulated by PCR to add desired restriction enzyme target sequences for ligation into the cloning vector [38]. Restriction enzymes can be utilized to selectively cut and expose nucleotide overhangs known as "sticky ends" in the MCS to allow for ligation of the desired DNA insert [38]. Promotors can be selectively chosen for protein expression in bacteria or mammalian cells and are essential for initiating a high rate of transcription of the gene of interest [38]. Additionally, antibiotic resistance genes are typically included in cloning vectors to assist researchers in easily selecting growing DH5-α bacterial colonies that have taken in the appropriate plasmid [39]. For the purposes of this thesis, pCMV-3Tag-3A cloning vectors are used primarily. This vector features a strong cytomegalovirus promotor (pCMV) that drives high transcription rates in mammalian cells, a series of three FLAG protein affinity tags that can be

used for labeling, if inserts are cloned in-frame, or silenced, if inserts contain a stop codon at the 3' end, and a kanamycin resistance gene sequence for use in growing and isolating bacterial colonies.

2.10 The Prediction of Novel Splice Variants

It is estimated that over 95% of documented genes have variants with alternative splicing [40, 41]. The National Centre for Biotechnology Information (NCBI) contains an expansive collection of gene transcripts that have been experimentally produced and validated through characterization studies. Said transcripts are given the prefix "NM_", which represents an mRNA molecule that has been curated and published. These authenticated mRNA sequences are subsequently presented with their protein product counterparts that are given the prefix "NP_". Both mRNA and protein sequences are then assigned a numerical value for identification (ie. NM_026644.2 and NP_080920.2).

However, NCBI also possesses a gene prediction software named Gnomon, which can identify possible additional transcript variants using an algorithm [42]. These predicted transcript variants are given similar prefixes to published counterparts, those prefixes being "XM_" and "XP_" which represent a predicted mRNA sequence and predicted protein product, respectively. To provide predicted transcripts and proteins, Gnomon gene prediction functions based on a two-part system of homology searching with *ab initio* prediction [42]. A basic overview of Gnomon is depicted in figure 2. Homology searching involves a search whereby a query gene is automatically matched to existing sequences in the available data libraries, yielding matching pairs of nucleotides with accuracy represented in percentages [43]. *Ab initio* prediction refers to the prediction of transcript variants using information found only in the known genomic sequence of a specific organism [44]. This combined system of homology searching and *ab initio* prediction increases

the accuracy of prediction for the projected transcript variants and proteins [42]. However, no prediction can be certain without experimental investigation. Thus, the XM_ and XP_ sequences require experimental studies both to determine if they exist *in vivo*, and to understand the physiological role that they play in cells. Characterizing alternative splicing is therefore critical for the understanding of physiological function and complexity of the proteins encoded by the genome [40].

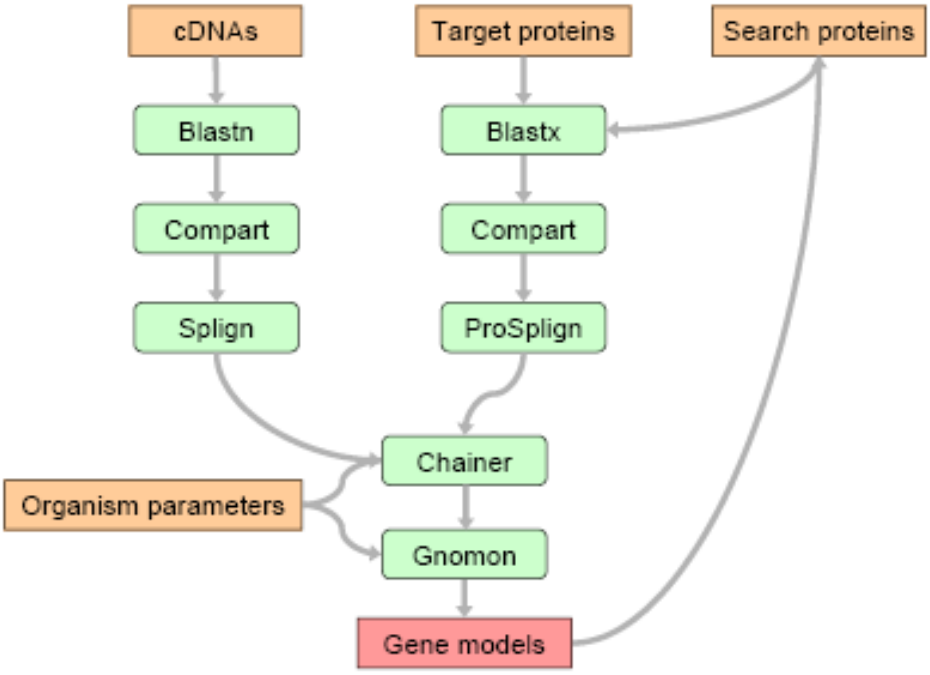


Figure 2: Overview of the Gnomon algorithm process adapted from Souvorov et al., (2010). Initial input parameters are indicated in orange boxes. Green boxes denote different programs and algorithms used to generate possible splice variants and sequences. The red box denotes final gene model output, which can then be re-added into the database of search proteins.

2.11 A Novel Truncated AGPAT4 Protein Arising from Predicted Splice Variants

In addition to the reference sequence for murine *Agpat4* (NM_026644.2), NCBI lists three alternative predicted mRNA transcripts for murine *Agpat4* – transcript variant X1 (XM_006523345.3), transcript variant X2 (XM_006523348.4), and transcript variant X3 (XM_006523347.3), derived from alternate splicing. Transcript variant X1 is predicted to encode a protein (XP_006523408.3) that is 80 amino acids longer than the reference protein (NP_080920.2), with the additional amino acids found upstream of the reference protein's start site. However, transcript variant X1 still maintains the conserved acyltransferase motifs I through IV shared with the reference protein (Fig. 3). Transcript variant X2 and transcript variant X3 both encode for the same predicted protein sequence (listed as XP_006523410.1 and XP_006523411.1, respectively), but differ in their 5' untranslated sequences. That predicted protein sequence is shorter than the characterized reference sequence (Fig. 3). Specifically, transcript variants X2 and X3 are predicted to encode for a protein that is missing the first and highly conserved "Motif I" among the acyltransferase enzyme families, HXXXXD, which is responsible for the catalytic activity of acyltransferase enzymes [1, 8]. Both variants are, however, predicted to code for an alternate, 10 amino acid N-terminal sequence that is unique from the characterized, reference form (Fig. 3).

Although the predicted alternative splice variants have not yet been verified in experimental reports in the literature, previous members of the Duncan Lab have successfully used PCR to amplify a fragment corresponding to the protein coding region of predicted transcript variants X2 and X3, which includes the novel N-terminal sequence, using murine cDNA isolated from whole brain (see Fig. 4). This demonstrates the presence of this alternative sequence in cells, *in vivo.* Because this transcript encodes for a smaller protein, we called the amplicon "Truncated

Agpat4" or "Trunc. *Agpat4*". The existence of an endogenously synthesized enzyme, predicted to be catalytically inactive due to the absence of the catalytic HXXXXD motif, is curious and, given the various physiological roles that AGPAT4 plays throughout a multitude of tissues, warrants investigation.

Figure 3: Multiple protein sequence alignment of murine reference AGPAT4 (NP_080920.2) with predicted variants X1 (XP_006523408.1), X2 (XP_006523410.1), and X3 (XP_006523411.1). Motif I (amino acid sequence HKFEID) is preserved in variant X1 but is not present in variants X2 and X3. Motif II (amino acid sequence FCTR), Motif III (amino acid sequence EGTR), and Motif IV (amino acid P) are preserved throughout all murine AGPAT4 transcript variants. Transcript variants X2 and X3 have unique N-terminal start sites further downstream than the AGPAT4 reference protein, causing Motif I to be absent.

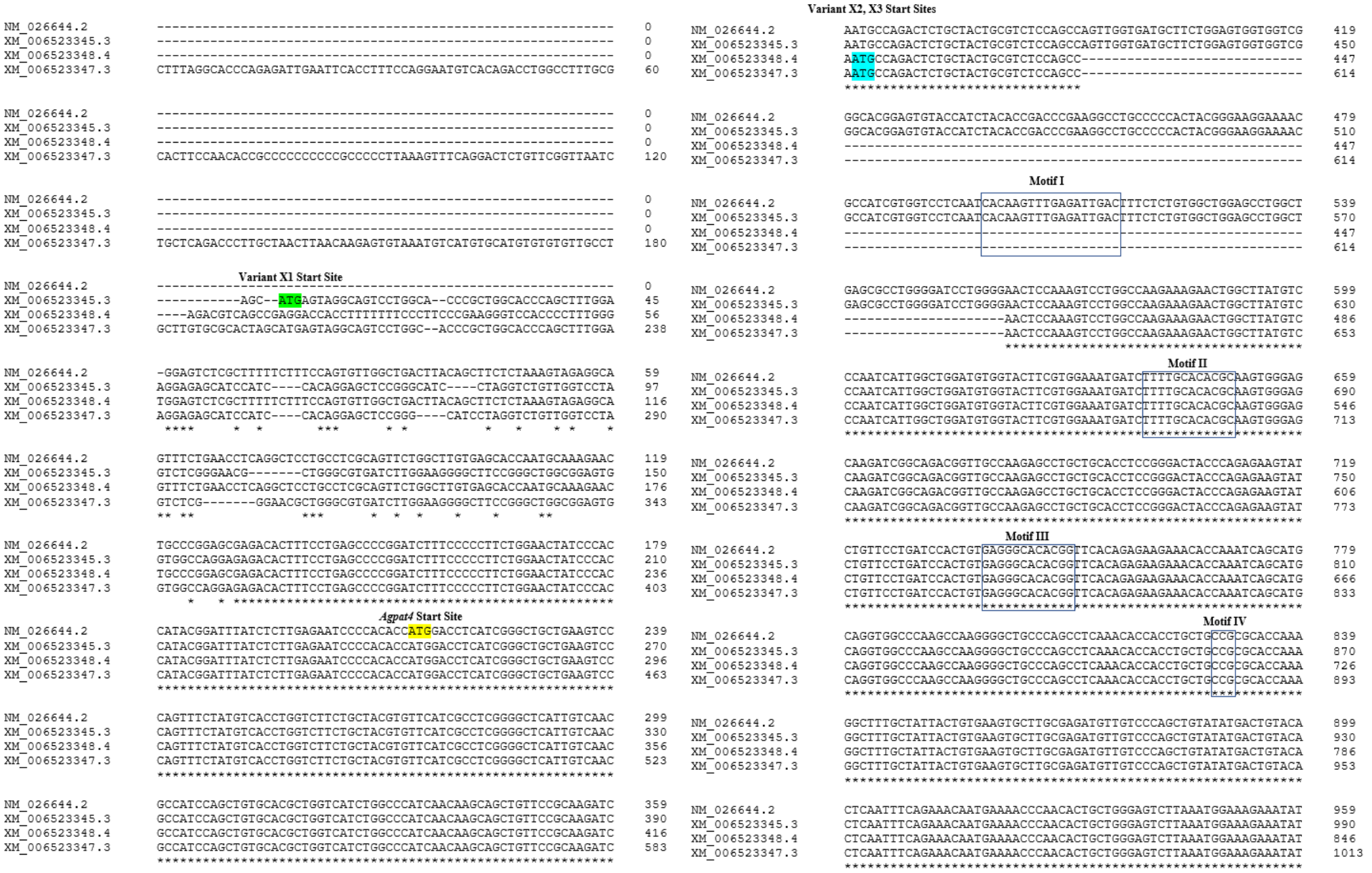

Variant X2, X3 Start Sites
Variant X1 Start Site
Motif I
Motif II
Motif III
Motif IV
Agpat4 Start Site

Figure 4: Multiple mRNA sequence alignment of reference murine *Agpat4* (NM_026644.2), variant X1 (XM_006523345.3), variant X2 (XM_006523348.4), and variant X3 (XM_006523347.3). The varying translational start sites are highlighted to showcase the unique N-terminal regions. The Motif I mRNA coding region is absent from transcript variant X2 and X3; mRNA coding regions for Motifs II, III, and IV are conserved among all variants. All variants of *Agpat4* share a distinct stop codon as depicted.

Prior research has been conducted on enzymes that also have truncated splice variants, with results often showing that truncated enzymes can play a regulatory role. For example, in one instance where the UGT2B7 enzyme was investigated, the splice variant also produced a catalytically non-functional truncated protein, and it worked to inactivate the primary reference enzyme via protein-protein interaction [45]. However, it is critical that direct experimental investigation is performed to understand the function and/or regulatory role of novel splice variants, since this cannot be inferred by structure alone. It is possible that a catalytically non-functional splice variant could simply act as a decoy, causing competitive inhibition of the reference variant by binding substrate, but not using it, and allowing the cell to rapidly titrate out activity of the reference variant. Alternate splice variants could also affect activity of the reference isoform by direct interaction. For example, alternate transcripts could bind to the reference form, altering the subcellular localization of the complex. This could result in targeting of the truncated-reference dimer to a site that is deficient in substrate, reducing activity, or targeting to a site where substrate is abundant, increasing activity. It could also result in allosteric regulation of the reference enzyme by inducing a conformational change, preventing, or increasing, access to substrate. Alternatively, dimerization of a truncated transcript with a reference transcript could target the complex for degradation or increase stability of both proteins. Direct experimental investigation is required for each of these possibilities.

This study will aim to determine whether interactions between truncated AGPAT4 and reference AGPAT4 result in increased or decreased levels of each protein. To better understand the possible physiological role of this variant and other predicted transcripts, differences in tissue expression of the truncated splice variant compared to the reference transcript will also be determined.

Chapter 3

Tissue expression profiles of reference *Agpat4* and predicted splice variants

3.1 Rationale and Objective

Expression of the reference *Agpat4* transcript has been characterized in several tissues using RT-PCR and RT-qPCR analysis. However, to the best of my knowledge, this analysis has not been performed for any of the predicted splice variants of *Agpat4*, specifically (*i.e.* transcript variants X1, X2 or X3), which would provide evidence that they exist. Additionally, this analysis also would distinguish the total transcript expression from expression of individual variants, which likely has been assessed as a total transcript pool in at least some prior publications, given the high degree of homology between the reference and predicted transcripts. Thus, assessment of the relative expression of the reference transcript, as well as the predicted transcripts, using primers that can distinguish these isoforms, would provide novel data on this gene. As a catalytically non-functional protein variant, it may be reasoned that the expression of predicted transcripts that encode for the truncated isoform of AGPAT4 (i.e. X2 and X3) in tissues is minimal, while expression of the predicted X1 transcript, which expresses a protein that is likely to be catalytically functional, is likely higher. Given that experimental evidence has not yet been found substantiating the presence of these alternate predicted isoforms, they are all likely to be expressed at lower levels than the reference variant. However, direct experimental evidence using an *in vivo* model of murine tissues is needed to verify both the existence of these transcripts, and their expression relative to the reference transcript. Transcript variant X1 was studied only in this chapter, and not in subsequent chapters that addressed the function of the truncated protein, since it does not translate into a truncated AGPAT4 protein.

Primary Objective: To determine the tissue expression profile of the predicted *Agpat4* transcript variants X1, X2 and X3, relative to the reference *Agpat4* variant.

3.2 Hypothesis

The predicted *Agpat4* variants X1, X2 and X3, will be detected in tissues. The variants that are predicted to produce a truncated AGPAT4 protein lacking Motif I will be minimally expressed in various tissues derived from C57Bl6/J mice relative to the reference *Agpat4* gene. The X1 variant, which is predicted to produce a longer AGPAT4 protein, will be expressed at higher levels than the X2 and X3 transcripts, but at lower levels than the reference transcript.

3.3 Methods, Materials, and Study Design

3.3.1 Tissue Collection

Tissue expression of truncated *Agpat4* and reference *Agpat4* was performed using brain, heart, kidney, white adipose tissue, and skeletal muscle (extensor digitorum longus and soleus), testes, lungs, liver, and brown adipose tissue from C57BL/6J mice. These tissues were chosen since prior reports in the literature have all exhibited some level of expression of the reference *Agpat4* gene. Organs were harvested from male mice 3-6 months of age that were fed a standard chow diet.

3.3.2 RNA Extraction and cDNA Generation

The following TRIzol® protocol is adapted from Invitrogen's TRIzol® reagent data sheet [46]. All surfaces and equipment used during the RNA extraction protocol were wiped down using RNaseZAP^{TM} cleaning agent to eliminate any RNases present in the environment. A 100 mg section of organ tissue was cut and placed into a 5 mL Falcon tube. One millilitre of ice-cold TRIzol® reagent was added to the tube and the sample was homogenized using a Polytron

Homogenizer at the highest setting. The subsequent mixture was incubated on ice for 5 minutes. Two hundred microlitres of chloroform were added to the sample, mixed by gentle inversion, and incubated on ice for 2-3 minutes. Using a refrigerated centrifuge, the sample was spun down at 4^{o}C for 15 minutes at 12,000 × g. Following centrifugation, the upper aqueous layer containing RNA is carefully transferred to a clean 1.5 mL microcentrifuge tube. Five hundred microlitres of isopropanol is added to the RNA, gently mixed by inversion, and centrifuged at 4^{o}C for 10 minutes at 12,000 × g. The supernatant was carefully extracted and discarded, leaving a pellet at the bottom of the microcentrifuge tube, which was resuspended using 1 mL of 75% ethanol and vortexed. The sample is then centrifuged again at 4^{o}C for 5 minutes at 7500 × g. The resulting supernatant is extracted and disposed. The sample pellet is air-dried for 5-10 minutes and resuspended in 50 µL of ddH_2O. The sample is subjected to 60^{o}C for 15 minutes using a heat block to enable resuspension. RNA sample purity was determined using a NanoDrop^{TM} 2000 spectrophotometer, with acceptable purity (260/280) falling between 1.8 to 2.0. RNA samples are diluted with ddH_2O to reach a final concentration of 2 µg and aliquoted to 10 µL in clean 1.5 mL microcentrifuge tubes.

To generate cDNA from RNA samples, 2 µL of 10x RT Buffer, 0.8 µL of 25x dNTPs, 2 µL of 10x RT Random Primers, 1 µL of reverse transcriptase, and 4.2 µL of ddH_2O is added to each of the previous 10 µL aliquoted RNA samples. The samples are placed in a Bio-Rad T100 Thermal Cycler and run on the following settings: 25^{o}C for 10 minutes, 37^{o}C for 2 hours, 85^{o}C for 5 minutes, and finally samples are held at 4^{o}C until they are collected. Resulting cDNA samples are diluted 1:5 using ddH_2O and stored at -20^{o}C.

3.3.3 Real-time PCR (qPCR)

qPCR settings are adapted from prior work performed in the Duncan Lab [2]. A Master Mix containing 5 µL of SYBR^{TM} Green (Bio-Rad), 0.5 µL each of 25 µM Forward and Reverse primers, and 3 µL ddH_2O per sample was added to the necessary wells in a 96-well plate. One µL of cDNA was added to each well. qPCR thermocycler settings are as follows: 95^{o}C for 2 minutes and 40 cycles of (95^{o}C for 10 seconds and 60^{o}C for 20 seconds). A melt curve analysis was run following the qPCR protocol to ensure primer specificity, and an amplification curve was generated to ensure that all amplification reactions are equally efficient. When plotted, melt curve analyses should display a well-defined peak and a consistent melting temperature of independent PCR products when graphed, indicating appropriate primer specificity for creating the desired amplicon [47]. Amplification curve plots should indicate similar cycle counts on the x axis when fluorescence thresholds are passed for each independent tissue and gene analyzed [47]. Relative expression of truncated *Agpat4* to reference *Agpat4* was determined using the ΔΔ Ct method. A list of primers used in qPCR experiments are listed in Table 2 on the following page; primer locations on each *Agpat4* transcript variant are depicted on Figure 5 on page 30. The specifically designed sequences of forward primers designed to recognize unique nucleotide sequences of *Agpat4* transcripts, allowing for the differential amplification of specific transcripts, in conjunction with melt curve analyses, will ensure gene expression is specific to each desired transcript [47].

Table 2: List of primers used in qPCR protocols to determine reference *Agpat4* and predicted transcript variants X1, X2, and X3 expression in tissues.

Primer Pair	Gene	Primer Sequence*	Amplicon Size
1	Reference *Agpat4* (NM_026644.2)	F: 5' - AGTGTTGGCTGACTTACAGC - 3' R: 5' - CCGTATGGTGGGATAGTTCCAG - 3'	165 bp
2	*Agpat4* Transcript Variant X1 (XM_006523345.3)	F: 5' - ATGAGTAGGCAGTCCTGGCA - 3' R: 5' - CCGTATGGTGGGATAGTTCCAG - 3'	214 bp
3	*Agpat4* Transcript Variant X2 (XM_006523348.4)	F: 5' - GTCAGCCGAGGACCACCTTT - 3' R: 5' - CCGTATGGTGGGATAGTTCCAG - 3'	239 bp
4	*Agpat4* Transcript Variant X3 (XM_006523347.3)	F: 5' - CGCACTAGCATGAGTAGGCA - 3' R: 5' - CCGTATGGTGGGATAGTTCCAG - 3'	223 bp

*Forward and Reverse primers are denoted by "F" and "R", respectively.

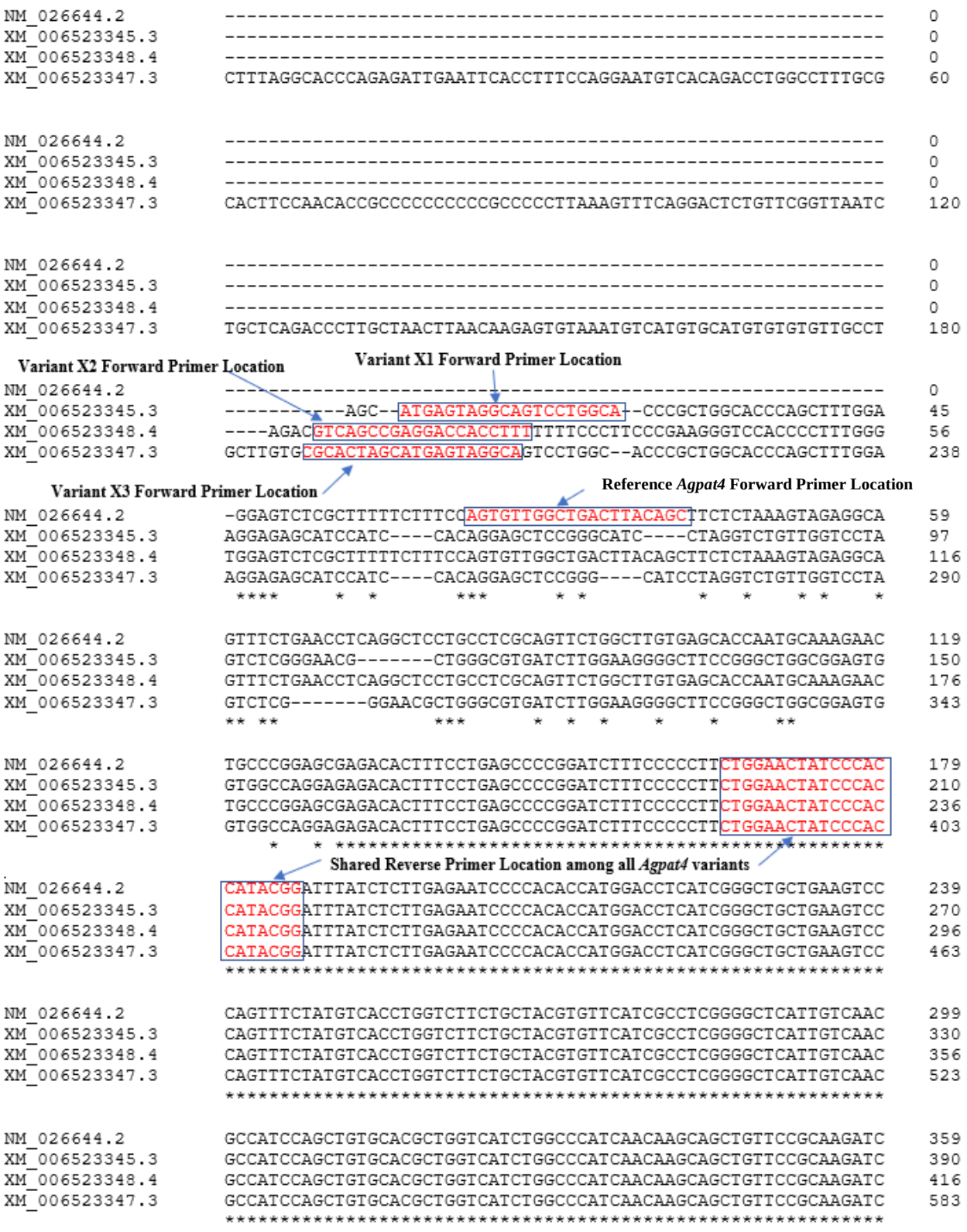


Figure 5: Multiple mRNA sequence alignment showing reference *Agpat4* and *Agpat4* transcript variant primer locations. Reference *Agpat4* and all transcript variants may share the same reverse primer due to having forward primers designed in unique 5' regions.

3.4 Statistical Analysis

Expression of the transcript variants was compared by 1-way ANOVA with multiple comparisons, and differences were analyzed by Tukey's post-hoc test.

3.5 Results

3.5.1 *Agpat4* variants X1, X2 and X3 are expressed in mice, but at lower levels relative to the reference *Agpat4* gene, and display heterogenous expression

The tissue expression of the known and predicted splice variants of *Agpat4* were analyzed by RT-qPCR. In all tissues examined, *Agpat4* variants X1, X2, and X3 were determined to be present, albeit at low abundance relative to the reference *Agpat4* transcript. This thesis is the first to report that these *Agpat4* gene variants do exist *in vivo* and display heterogeneous expression levels in multiple tissues.

The reference *Agpat4* gene expression was found to be the most abundant in all ten selected tissues relative to *Agpat4* transcript variants X1, X2, and X3 as depicted in Figures 6 and 7 on the following pages, and differences were highly statistically significant ($P<0.0001$). This result was expected, since the reference *Agpat4* gene codes for the AGPAT4 protein that has been studied to the largest extent and is known to be involved in physiologically important roles in multiple tissues [3-6]. The largest relative difference in expression levels were observed in the brown adipose tissue depots (BAT) and liver, shown in Figure 7B and 7D, respectively. In liver, the reference *Agpat4* transcript was found to be ~23 times more abundant than the X1 transcript, more than 2000 times more abundant than the X3 transcript, and more than 200,000 times more abundant than the X2 transcript. In BAT, the reference *Agpat4* transcript was found to be over 70 times more abundant than the X1 transcript, more than 10,000 times more abundant

than the X3 transcript, and almost 400,000 times more abundant than the X2 transcript. On the other hand, in the brain, kidneys, and testes, shown in Figures 6A, 6C, and 7E, respectively, the reference *Agpat4* gene is expressed at levels at least 10-fold higher than the other transcripts, which are all present at similarly low levels. In heart, EDL, and lung, shown in Figures 6B, 6D and 6E, respectively, the X1, X2 and X3 transcripts are expressed at similar levels, with the reference transcript expressed at much higher levels, ranging from ~75- to ~100-fold higher. In WAT (Figure 7A), the X1 transcript is expressed at significantly higher levels than the X2 or X3 transcripts, but all are significantly less abundant than the reference transcript, which was expressed at ~80-fold, ~800-fold, and ~8000-fold higher levels than each of the alternate transcripts, respectively. In soleus muscle (Figure 7C), a similar pattern is evident, with the X1 transcript expressed at >30-fold higher levels than the X2 or X3 transcripts, and the reference transcript expressed at ~25-fold higher levels than the X1 transcript, ~800-fold higher levels than the X2 transcript, and ~3000-fold higher levels than the X3 transcript.

The *Agpat4* transcript variant X1 is predicted to encode an AGPAT4 protein that contains an extra 80 amino acids upstream from the reference AGPAT4 protein's starting methionine residue. While this longer predicted protein variant was not the focus of this thesis, we included it in this gene expression study to encompass expression of the known and predicted *Agpat4* transcript variants as a whole. The observed results still proved to be of interest, as the *Agpat4* variant X1 was significantly more abundant relative to the truncated variants X2 and X3 in the epidydimal white adipose tissue (Figure 7A), brown adipose tissue (Figure 7B), soleus (Figure 7C), and liver (Figure 7D). When analyzed, *Agpat4* variant X1 displays the largest significant difference in expression relative to both the truncated variants X2 and X3 in the soleus, liver, and

BAT ($P<0.0001$). In the epididymal WAT depot, variant X1 only displays a small significant increase in expression relative to truncated variant X3 ($P<0.05$) but not truncated variant X2.

Results from the RT-qPCR experiments agreed with the hypothesis that the *Agpat4* gene variants X2 and X3 overall display minimal expression in the tissues studied relative to the reference *Agpat4* transcript. Between the X2 and X3 variants, gene variant X2 appeared to be the more highly expressed variant in most tissues, with exceptions seen in the liver and brown adipose tissue depot. Comparative analyses between the two truncated variants revealed a significant difference in the liver and BAT depot where the X3 variant was expressed at significantly higher levels relative to the X2 variant ($P<0.0001$). Further statistical analyses in the remaining tissues did not reveal any significant differences in gene expression between the two truncated variants. Of interest, *Agpat4* variant X2 was found to be expressed more highly in the brain than any other tissue relative to the reference *Agpat4* variant. Given the fact that the reference gene is most highly expressed in the brain as well [6], results showing a similar pattern with a novel variant indicates a potentially important role for the protein that it produces.

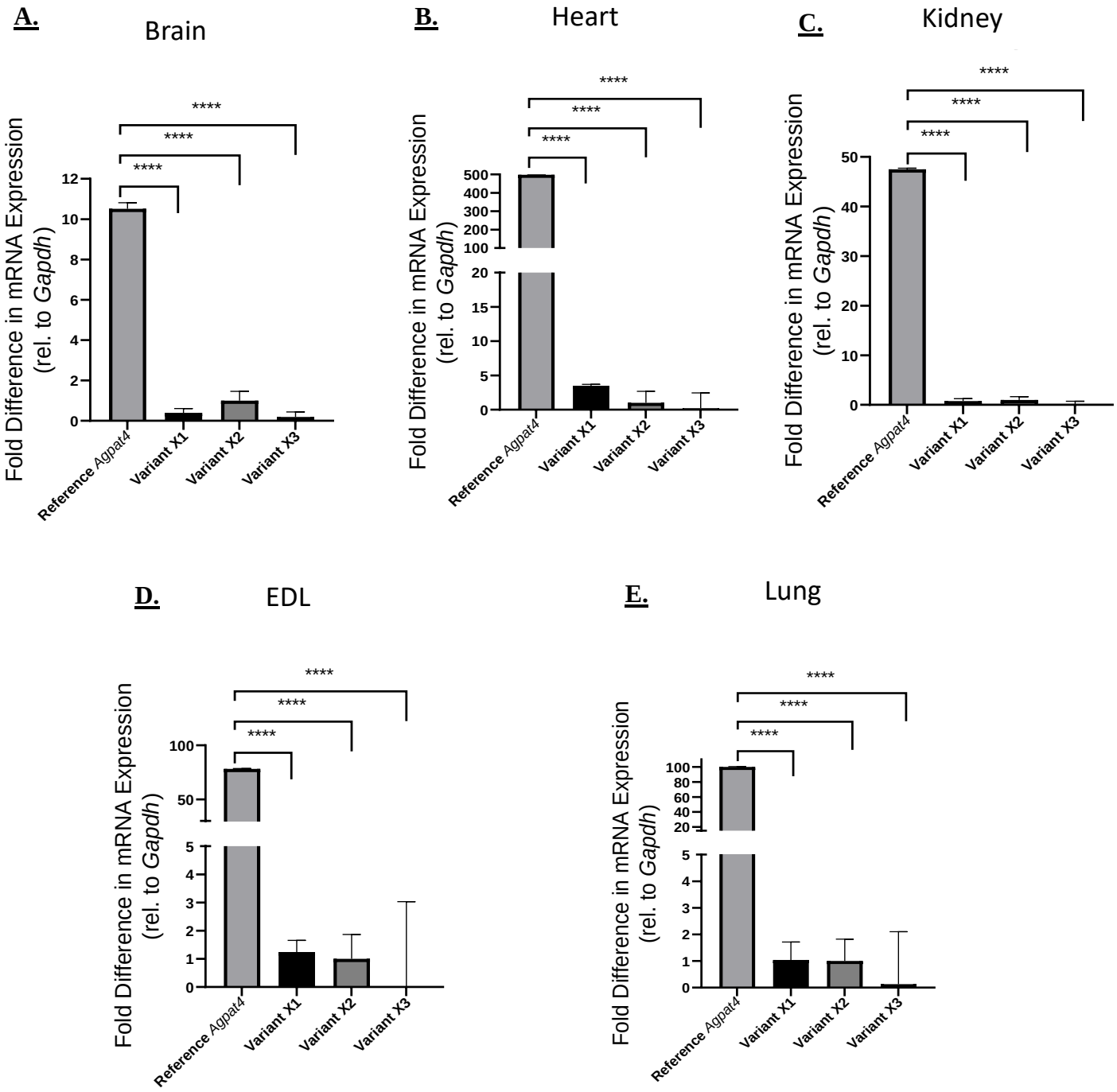


Figure 6: Expression of the reference and predicted *Agpat4* mRNA transcripts in brain, heart, kidney, extensor digitorum longus (EDL) muscle, and lung. Reference *Agpat4* transcript was observed to have significantly higher expression than *Agpat4* predicted splice variants X1, X2, and X3 in the mentioned tissues. No significant differences were detected between predicted splice variants X1, X2, and X3. RT-qPCR was performed using specific primers designed to recognize and distinguish between the unique splice variants. Values were normalized to *Gapdh* as a loading control and are represented in fold difference relative to the X2 variant. Data are presented as means ± S.E.M. Differences are as marked, **** P<0.0001.

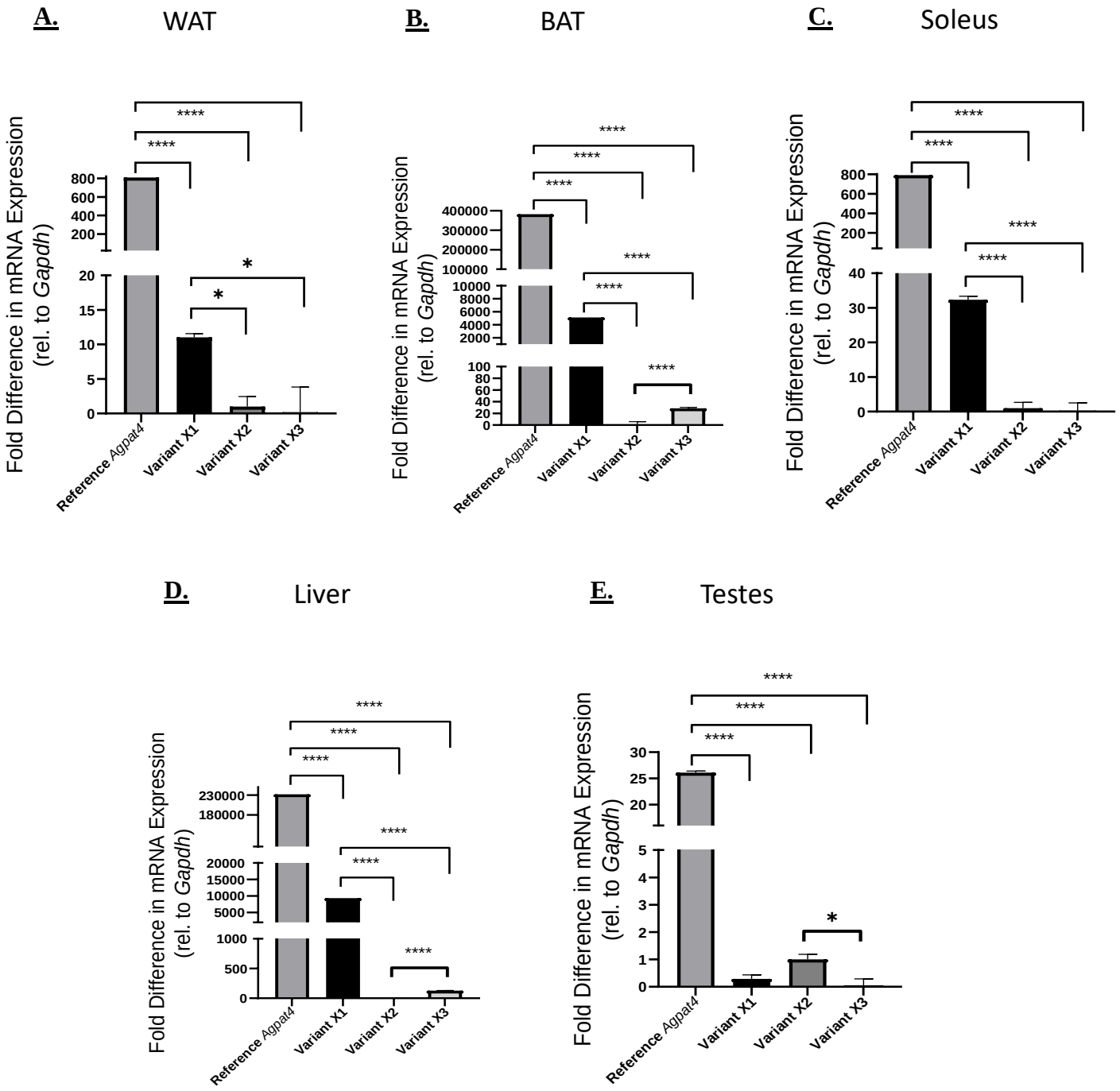


Figure 7: Expression of the reference and predicted *Agpat4* mRNA transcripts in epididymal white adipose tissue (WAT), brown adipose tissue (BAT), soleus muscle, liver, and testes. Reference *Agpat4* had significantly higher expression than *Agpat4* predicted splice variants X1, X2, and X3. *Agpat4* predicted variant X1 was detected at significantly higher levels in WAT, BAT, soleus, and liver relative to predicted variants X2 and X3. *Agpat4* predicted variant X2 was detected at significantly higher levels in the testes. *Agpat4* predicted variant X3 was detected at significantly higher levels than predicted variant X2 in the BAT and liver. RT-qPCR was performed using specific primers designed to recognize and distinguish between the unique splice variants. Values were normalized to *Gapdh* as a loading control and are represented in fold difference relative to the X2 variant. Data are presented as means ± S.E.M. Differences are as marked, * $P<0.05$, **** $P<0.0001$.

3.5.2 *Agpat4* Transcript Predicted Splice Variants X1, X2, and X3 display highest expression levels in murine brain.

Utilizing the same raw data from the previous section, expression of *Agpat4* predicted splice variants X1, X2, and X3 in each tissue were normalized to the lowest expressing tissue per variant to determine the spectrum of gene expression between tissues. Statistical analysis showed that for all the predicted transcript variants, whole brain displayed significantly higher levels of transcripts X1, X2, and X3 compared to the other tissues studied. The reference *Agpat4* transcript did not display the highest expression levels in the brain, contrary to previously published findings from our lab. Rather, the testes showed the highest tissue expression of reference *Agpat4* followed by the lungs, and whole brain after. Figure 8 on the following page depicts the heterogenous expression between tissues for each *Agpat4* transcript variant.

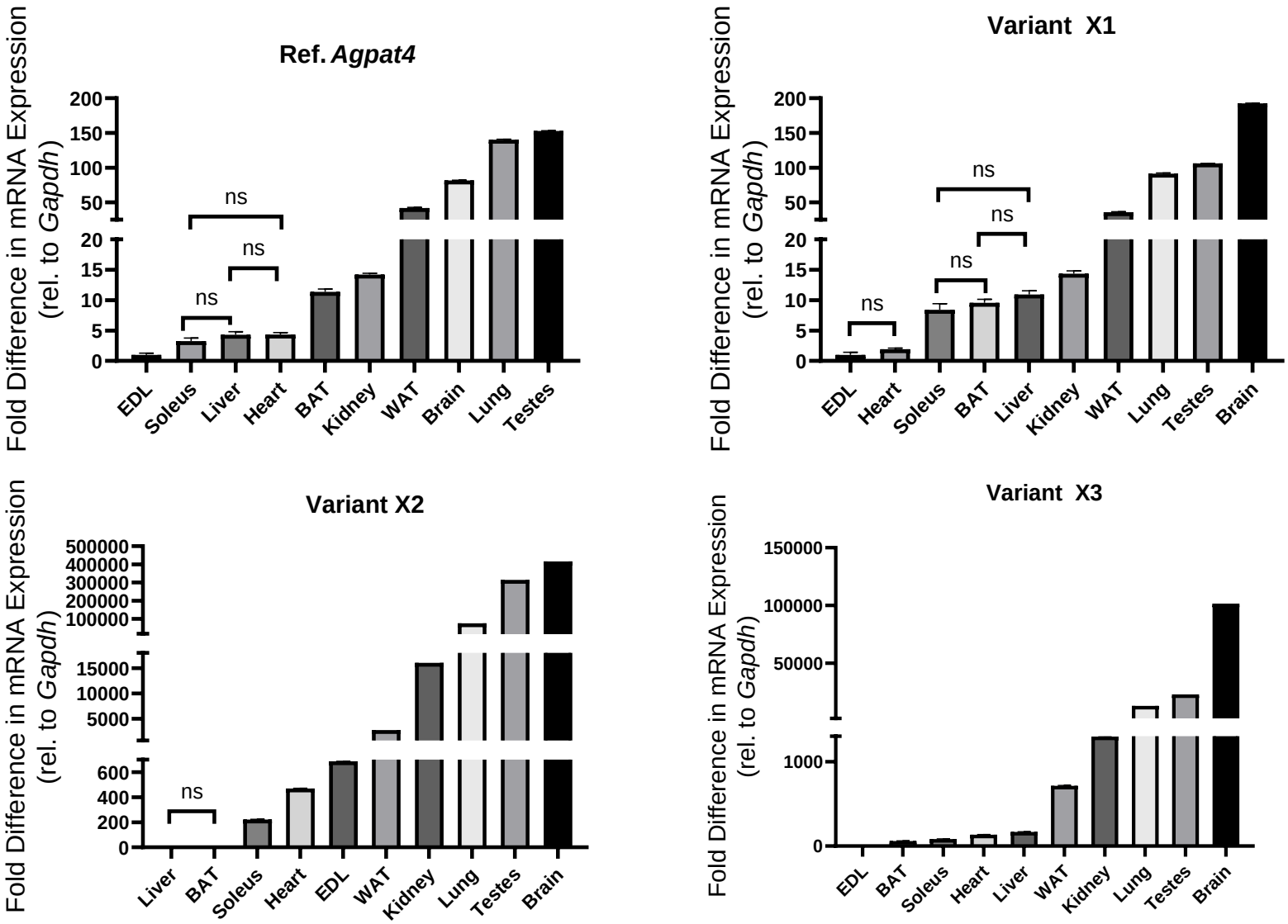


Figure 8: Comparison of expression of reference *Agpat4* and predicted transcript variants X1, X2, and X3 between tissues studied. The brain, testes, and lungs displayed the highest tissue expression of all transcript variants, but the testes displayed significantly higher expression of reference *Agpat4*, while the brain displayed the highest tissue expression of all three *Agpat4* predicted splice variants. RT-qPCR was performed using specific primers designed to recognize and distinguish between the unique splice variants. Values were normalized to *Gapdh* as a loading control and are represented in fold difference relative to the lowest expressing tissue, per transcript variant. Data are presented as means ± S.E.M. All values are significantly different, unless marked "ns", denoting no significance.

3.6 Discussion

We first examined tissues that were known to express the reference *Agpat4* variant more abundantly, namely the brain, heart, kidneys, white adipose tissue, extensor digitorum longus, and soleus [3, 6]. Of these six tissues, the X2 variant was found to be the second most abundant transcript, after the reference transcript, only in brain. In the other tissues, the X1 transcript tended to be more abundant, or the X1 and X2, or X1, X2 and X3 transcripts tended to be similarly abundant.

To follow up on this work, I then investigated tissues where the reference *Agpat4* gene is reportedly less abundant - specifically the liver, lungs, brown adipose tissue, and testes [3, 6, 27]. The purpose of researching these tissues that did not endogenously express *Agpat4* at higher levels was to attempt to elucidate a possible function of the *Agpat4* splice variants X2 and X3, namely looking into a possible modulation or reduction of endogenous levels of reference *Agpat4*. Results did not indicate higher expression levels of any of the *Agpat4 splice* variants relative to the reference gene. Of the four organs studied in this second round of RT-qPCR, the testes displayed higher levels of the X2 *Agpat4* variant relative to variant X3, and the liver and BAT expressed higher levels of variant X3 relative to X2. Since the resultant proteins translated by either the *Agpat4* X2 or X3 variants are exactly the same in coding and amino acid sequence, the varying expression levels observed may be attributed to differences in the 5' untranslated regions in their mRNA sequences [48], which may regulate the stability of these X2 and X3 transcripts [49]. Additionally, since these splice variants are also formed from different transcriptional initiation sites, it is possible that differences in genomic regulators, such as transcription factors or repressors, may also play a role [50]. Given the very wide array of possible factors that may play a role in differentially regulating the expression of the different

transcripts, determining the reason for the varied expression of the two truncated variants will take considerable additional study.

Agpat4 variant X1, coding for a larger AGPAT4 protein, was found to be expressed at significantly higher levels relative to the truncated X2 and X3 variants in the soleus, liver, brown adipose tissue, and at significantly higher levels relative only to the truncated X3 variant in the epidydimal white adipose tissue depot. Since the additional 80 amino acid moieties are found upstream of the reference AGPAT4 protein start site, my current speculation to explain these findings is that the additional amino acids may be representative of a unique subcellular localization signal, or as some modulator of tertiary or quaternary folding structure. As a preliminary investigation step, I referred to a subcellular localization prediction tool (PSORT) as described in [51]. The X1 variant is predicted to localize to the plasma membrane of cells rather than the mitochondria, endoplasmic reticulum, or *trans*-Golgi apparatus. A figure depicting the prediction results from PSORT may be found in the Appendix. An interesting similarity between the five tissues that expressed significantly higher levels of the X1 variant (liver, soleus, WAT, and BAT) is that they are either oxidative tissues or more involved in lipid metabolism and lipid storage [52-54].This may point to a potential role of the X1 variant in tissues that display more oxidative properties or have more functional roles in processing lipids. The predicted amino acid sequence of variant X1 still contains all four of the acyltransferase family conserved motifs, so it may be deduced that the overall enzymatic function remains the same in terms of converting lysophosphatidic acid to phosphatidic acid. However, *Agpat4* transcript variant X1 or the resultant protein product predicted to be translated from this transcript has not been studied beyond this, and thus the true nature of its existence is still unknown.

Expression of reference *Agpat4* and predicted splice variants revealed that between tissues, murine whole brain expressed significantly higher levels of all three predicted splice variants of *Agpat4*. This finding is consistent with previous findings in the literature, of which reference *Agpat4* is highly expressed in whole brain of mice [6]. Interestingly, the qPCR results from this study pertaining to the reference transcript showed inconsistencies with published findings. This chapter shows reference *Agpat4* expressed at significantly higher levels in testes and lungs, followed by whole brain. I speculate the differences in findings may be attributed to the forward primer used in this study to be specific to the 5' untranslated region and the reverse primer of the reference *Agpat4* transcript. Different primers used in qPCR studies typically have innate differences in amplification efficiencies, which may have led to the discrepancy between results [55]. Notably, amplification efficiencies of the primers have not yet been calculated, and will be included in future work. These differences may also be attributable to differences in the age and metabolic state of the mice, since the *Agpat4* gene is regulated with changing metabolic conditions [2].

In summary and to the best of my knowledge, this study is the first to show evidence that the predicted *Agpat4* splice variants X1, X2 and X3 are, indeed, expressed in murine tissues. As hypothesized, however, the reference variant was the most highly expressed transcript relative to the predicted splice variants. Murine whole brain displayed significantly higher expression of predicted *Agpat4* splice variants X1, X2, and X3. However, it is important to consider that this study only covers this tissue expression profile in wild-type, male mouse models that are 3-6 months of age (equivalent to a young adult in humans). It is reasonable to suggest that the expression of the three *Agpat4* transcript variants could be influenced by numerous factors such as feeding, aging, disease, or sex differences, which have been associated with alterations of

gene expression [56-58]. These factors may upregulate or downregulate expression of the reference *Agpat4* transcript and/or the other splice variants.

Chapter 4

Synthesis of a new plasmid expressing truncated Agpat4 tagged with hemagglutinin.

4.1 Rationale and Objective

At the start of this thesis, the Duncan Lab had only one plasmid that could express truncated AGPAT4 in mammalian cells - specifically pEGFP-Trunc.AGPAT4. When transfected into mammalian cells, pEGFP vectors enable host cells to express a 27 kDa Green Fluorescent Protein that, when exposed to a blue laser, will cause the GFP to glow bright green, wherever it is present in cells. Truncated AGPAT4 is estimated to be ~30 kDa (based on an average amino acid mass of 110 Da), and when translated in-frame with a GFP molecule, is approximately 57 kDa in total. Under conventional use in confocal microscopy, GFP should not disrupt subcellular localization. Indeed, it is used extensively in cell biology because it does not disrupt targeting. However, when investigating a potential protein-protein interaction between FL. AGPAT4 and Trunc. AGPAT4, the 27 kDa size of GFP could potentially cause interference in dimerization. It has been noted in the literature that larger protein tags may result in problems of functional activity in the protein structure of which it is bound [59]. Thus, a critical goal before proceeding with interaction experiments was to clone a new plasmid containing truncated AGPAT4 with a small affinity tag, but without GFP. A table listing the relevant plasmids pertaining to this thesis are shown later in this chapter in table 4a and 4b.

Primary Objective: To create a new plasmid to express the truncated splice variant of *Agpat4*, based on the coding region of the X2 and X3 transcripts, with a hemagglutinin protein tag.

4.2 Materials, Methods, and Cloning Strategy

4.2.1 Generation of a new Amplicon via PCR

Using PCR, a new amplicon was generated that contained the coding sequence for Trunc. AGPAT4, with a hemagglutinin (HA) tag at the C-terminus, and restriction sites at the N-terminus and C-terminus ends. Primers used to create the new amplicon, using pEGFP-Trunc.AGPAT4 plasmid as a template, are as shown in the following table:

Table 3: List of primers used in PCR protocols to modify pEGFP-Tr. AGPAT4 plasmid.

Primers	**Direction**	**Primer Sequence**	**Amplicon Size**
Bgl II-Trunc. AGPAT4	Forward	5'–CTCAGATCTACCATGCCAGACTCT–3'	870 bp
Trunc. AGPAT4-HA-Stop-XhoI	Reverse	5'-ACTCGAGCTATGCATAGTCCGGGACGTCAT AGGGATAGTCCGTTTGTTTCCGTTTGTTGTC – 3'	

Primers were re-hydrated and diluted to 100 µM using ddH_2O. A polymerase chain reaction (PCR) master mix was made using 2 µL of 10x PCR Buffer + $MgCl_2$, 0.5 µL of 10 mM dNTPs, 0.1 µL each of 100 µM Forward and Reverse Primers, and 0.1 µL of Taq Polymerase [6]. Master Mix was added to 1 µL of pEGFP-Trunc.AGPAT4 plasmid DNA. Total volume was filled to 20 µL with ddH_2O. Samples were run on the following setting using a Bio-Rad T100 Thermal Cycler: Initialization: 95°C for 3 minutes, 40 cycles of (Denaturation: 95°C for 30 seconds, Annealing: 55°C for 30 seconds, Elongation: 72°C for 1 minute), and Final Elongation: 72°C for 10 minutes. The PCR product was electrophoresed through a 1% agarose-TAE gel for isolation and purification using a gel purification kit from Qiagen.

4.2.2 Cloning of the Amplicon to Form pCMV-Trunc. AGPAT4-HA tag

The PCR process described above resulted in an amplicon containing a sequence that codes for the truncated AGPAT4 protein with a Bgl II restriction site at the 5' end, as well as an HA tag, stop codon, and Xho I restriction site at the 3' end. The fragment was purified from the PCR reaction mixture by agarose gel electrophoresis, and the gel-cleaned fragment was then ligated into pGEM-T-Easy vectors using T4 DNA ligase to generate pGEM-Trunc.AGPAT4-HA tag. Use and optimization of the pGEM-T-Easy vector system was derived from the Promega technical manual [60]. After an overnight ligation reaction at room temperature, pGEM-Trunc. AGPAT4-HA tag was transformed into competent DH5-α cells by heat shock. Transformed bacteria were plated on agar plates containing ampicillin to produce individual bacterial colonies.

Colonies were selected, grown, and plasmid produced was purified and checked by restriction digest for the presence of an insert of the appropriate size. Clones containing an 870 bp insert were sent for sequencing by The Centre for Applied Genomics (TCAG) sequencing facility. Sequence results were verified by comparing results to the National Center for Biotechnology Information's (NCBI) RefSeq database using accession number XM_006523348.4. A verified clone was grown in a 200 mL stock to produce sufficient plasmid for use in subcloning, and pGEM-Trunc.AGPAT4-HA tag was then digested with Bgl II and Xho I, and the resulting fragments were electrophoresed on a 1% agarose-TAE gel. The band corresponding to the amplicon was excised and purified, and ligated into a pCMV-3Tag-3A vector that had been linearized by restriction digest with BamHI and XhoI, which produces compatible cohesive ends, and purified by electrophoresis in a 1% TAE agarose gel. The ligation mixture was transformed into DH5-α cells. Transformed bacteria were plated on agar containing

kanamycin, and selected colonies were grown and purified. Positive clones were identified by restriction digest using Not I and Xho I and verified by sequencing by the TCAG facility.

Table 4a: List of Duncan Lab plasmids utilized throughout this project and their associated Protein Tags, General Descriptions, and Uses

Plasmid	**Protein Tags**	**Description**	**Uses**
pEGFP-Trunc.AGPAT4 (In Duncan Lab inventory)	Green Fluorescent Protein (GFP)	Mammalian expression vector containing ***Truncated*** *Agpat4* in the multiple cloning site (MCS) followed by an EGFP coding region to create a chimeric Trunc. AGPAT4 protein with GFP attached on the C-terminus.	Co-Immunoprecipitation + Immunoblotting for GFP to determine if interaction between Ref. AGPAT4 and Trunc. AGPAT4 exists.
pCMV-3TAG-3A (In Duncan Lab inventory)	FLAG Tag (3x FLAG)	Mammalian expression vector that **does not** contain any specific gene in the MCS, followed by 3 FLAG Tags on the C-terminus.	Molecular cloning as the final mammalian expression vector to express gene of interest. Secondary Use: Control vector in experiments to ensure non-specific reactions from the pCMV host vector are non-impactful to results.
pCMV-Full.AGPAT4-6HIS-FLAG (In Duncan Lab inventory)	Polyhistidine Tag (6x HIS) FLAG Tag (3x FLAG)	Mammalian expression vector containing ***Reference*** *Agpat4* in the MCS followed by a polyhistidine tag specifically comprised of 6 HIS residues, then followed by 3 FLAG Tags on the C-terminus.	Co-Immunoprecipitation and IMAC experiments testing for interactions between Ref. AGPAT4 and Trunc. AGPAT4.
pCMV-Full.AGPAT4-*FLAG* (In Duncan Lab inventory)	FLAG Tag (3x FLAG)	Mammalian expression vector containing ***Reference*** *Agpat4* in the MCS followed by 3 FLAG Tags on the C-terminus.	Co-Immunoprecipitation experiments testing for interactions between Ref. AGPAT4 and Trunc. AGPAT4.

Table 4b: List of plasmids utilized throughout this project and their associated Protein Tags, General Descriptions, and Uses

Plasmid	**Protein Tags**	**Description**	**Uses**
pCMV-Trunc.AGPAT4-HA (Synthesized as part of Aim #2)	Hemagglutinin Tag (HA)	Mammalian expression vector containing ***Truncated*** *Agpat4* in the MCS followed by a Hemagglutinin Tag on the C-terminus. A stop codon was inserted after the Hemagglutinin Tag to exclude the 3x FLAG tag coding region present in the pCMV host vector.	Co-Immunoprecipitation experiments testing for interactions between Ref. AGPAT4 and Trunc. AGPAT4.
pGEM-T-Easy (In Duncan Lab inventory)	N/A	Bacterial expression vector used as an amplification tool for creating high copy numbers of a selected gene of interest without the need of PCR.	Utilized in molecular cloning to create multiple copies of Trunc. AGPAT4 insert to be subcloned into pCMV-3TAG-3A vectors.

4.2.3 Generation and purification of a Bgl II-Trunc.AGPAT4-HA Tag-Xho I amplicon

Utilizing PCR, pEGFP-Trunc.AGPAT4 plasmids were successfully amplified forming new Bgl II-Trunc.AGPAT4-HA-Xho I amplicons without GFP. Following one PCR run using 20 samples, only 2 sample tubes contained the correct HA tagged amplicon. Since the purpose of this PCR step was to introduce two new restriction enzyme sites, as well as a sequence coding for a hemagglutinin (HA) tag, the specificity of the primers I used for the coding sequence of truncated AGPAT4 was low. This likely contributed to the low amplification efficiency that was noted, due to the larger overhangs of non-complimentary primer sequences in both the forward and reverse primers. Bgl II-Trunc.AGPAT4-HA tag-Xho I amplicon size was predicted to be 870 base pairs in size. Confirmation of correct amplicon size was determined by electrophoresing the linearized fragments in a 1x TAE gel and imaging the bands under UV light as depicted in the image on the following page (Figure 9) . The fragments were excised from the gel and purified to isolate the DNA insert for further use in cloning.

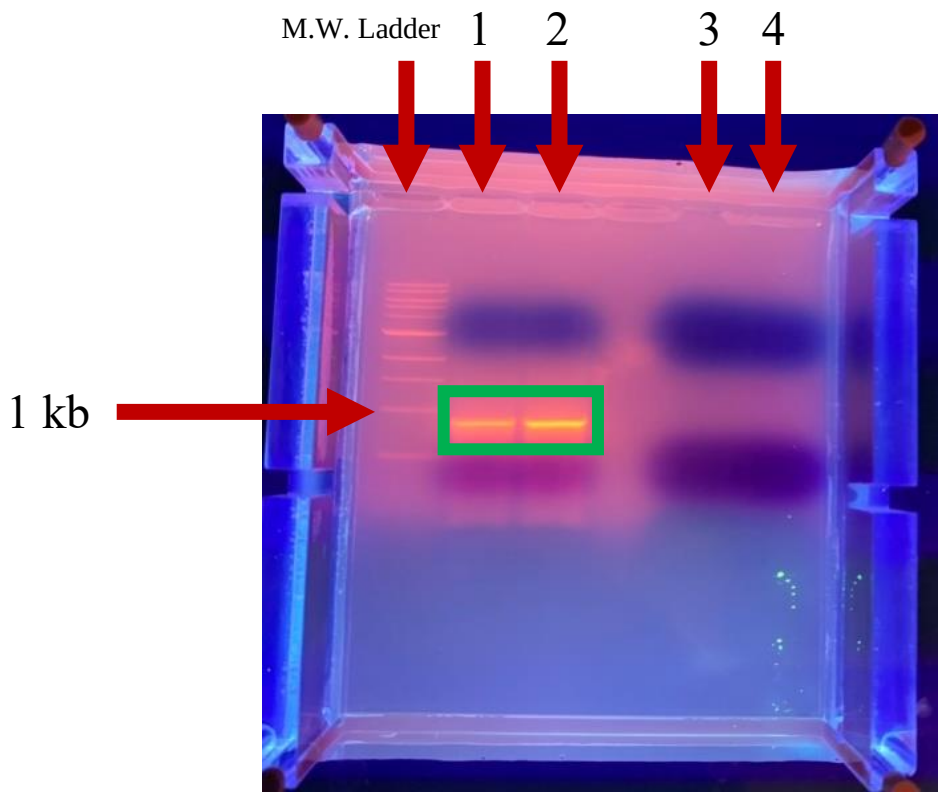


Figure 9: UV gel image showing the correct 870 bp band containing the Bgl II-Trunc.AGPAT4-HA Tag- Xho I amplicon in Lanes 1 and 2, denoted by the green box. Lanes 3 and 4 contained unsuccessful PCR samples which failed to amplify the linearized amplicon, and do not display the fluorescent bands when exposed to UV light like Lanes 1 and 2 do. There are two visible dye fronts in this gel image, coloured blue for the upper front and purple for the lower front. They are utilized as a visual check to determine where the DNA amplicon is contained, as DNA is not visible during the electrophoresis process. This is to prevent accidentally running the DNA off the gel and losing the samples entirely. There is a presence of primer dimers in Lanes 1 and 2 which may be observed as faint orange bands below the lower purple dye front. They are not of any significance and occur randomly when primers exhibit some self-complementarity for themselves in the PCR sample tube solution during PCR.

4.2.4 Subcloning of the Bgl II-Trunc.AGPAT4-HA Tag-Xho I amplicon into pGEM-T-Easy

Gel-purified Bgl II-Trunc.AGPAT4-HA Tag-Xho I fragments were ligated into pGEM-T-Easy host vectors and transformed into competent DH5-α cells via heat shock. DH5-α colonies were plated on agar, individually selected and grown, purified, then checked for insertion creating a new plasmid, pGEM-BglII-Trunc.AGPAT4-HA Tag-Xho I plasmid, by restriction enzyme digestion with Not I (since Not I sites flank the insertion site of this vector) followed by electrophoresis through 1x TAE gels. This process took numerous attempts over the course of 3 months of attempted ligations of the host vector and plasmid insert DNA. On average, 30 to 40 colonies of DH5-α cells were checked per week during this subcloning process, which means that ligation was successful in less than 1% of colonies screened. The final ligation, which contained a single successfully transformed colony of DH5-α cells is depicted in Figure 10. The single colony of DH5-α cells containing the pGEM-BglII-Trunc.AGPAT4-HA Tag-Xho I plasmid was grown into a larger stock, and DNA was purified from it, and a sample was sent for sequencing at the TCAG facility.

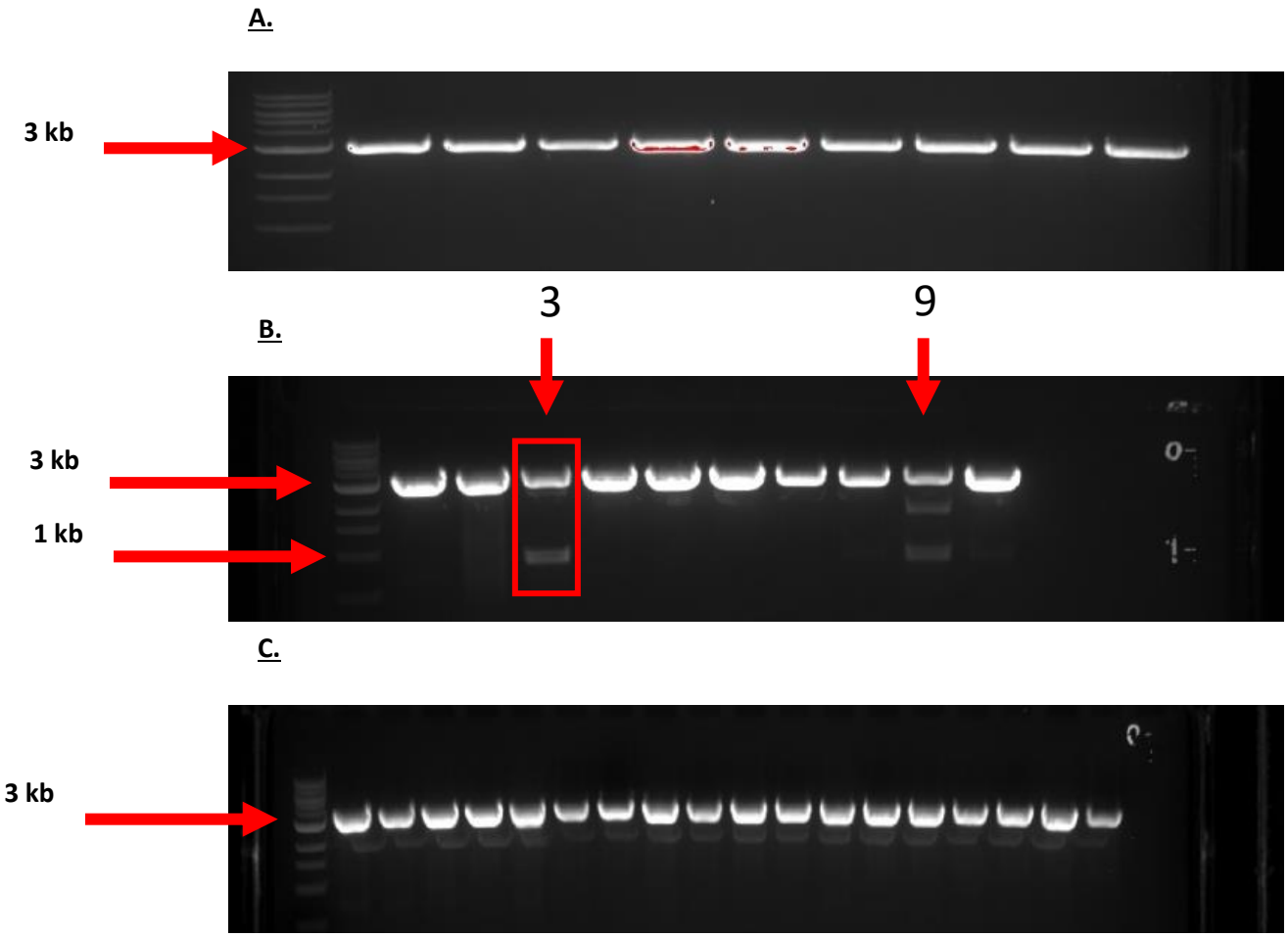


Figure 10: UV gel images depicting restriction digests attempting to identify DH5-α colonies containing the desired pGEM-BglII-Trunc.AGPAT4-HA Tag-Xho I plasmid. **A.** No colonies contained the plasmid of interest. **B.** A single colony, seen in Lane #3 and highlighted in a red box, contained the desired plasmid, indicated by the presence of two bands. The upper band was approximately 3 kb in size, indicative of the linearized pGEM host vector, while the lower band was located just under the 1 kb ladder marker, indicative of the 870 bp Bgl II-Trunc.AGPAT4-HA Tag-Xho I insert. **C.** No colonies contained the plasmid of interest.

Note: Lane #9 in image **B.** shows presence of 3 unique bands which is indicative either of random star activity from the restriction enzymes, or insertion of an improper insert. Although the lower band is approximately just below the 1 kb ladder marker, the fact that random cuts were made by the restriction enzyme deemed this sample to likely be incorrect, and therefore it was not selected as a successful colony.

4.2.5 Subcloning of the Bgl II-Trunc.AGPAT4-HA Tag-Xho I amplicon from pGEM-T-Easy into pCMV-3TAG-3A

Following confirmation that the subcloned pGEM-Trunc. AGPAT4-HA Tag plasmid had the correct DNA sequence, the DH5-α colony containing the plasmid was grown out to large quantities, and the plasmid DNA was purified, restriction digested using BglII/XhoI, electrophoresed through 1x TAE gels, and gel purified (to separate the insert from the pGEM-T-Easy vector) to mass-produce the Bgl II-Trunc.AGPAT4-HA Tag-Xho I insert. To note, a new stop codon was included in the PCR primer after the Xho I restriction site. The stop codon was added in order to express the necessary Trunc. AGPAT4 with the HA tag but without the 3x FLAG tag that is normally included in the pCMV-3TAG-3A cloning vector. The pCMV-3TAG-3A vector was digested using BamH1/XhoI , then gel purified to remove the fragment generated from the multiple cloning site, in order to linearize it and prepare it with compatible cohesive ends for ligation with the insert. Of importance, BglII and BamH1 are different restriction sites but generate complementary cohesive overhangs following restriction digest. However, after ligation, the resulting site would no longer be recognized by either BglII or BamH1 and would need to be cut by an entirely different enzyme, if one exists. The alternative to this would be moving upstream to the next available restriction site and working with that instead in the case that a restriction digest is needed. The gel purified insert was ligated into pCMV-3TAG-3A, which was transformed again into DH5-α cells via heat shock. The DH5-α cells were subjected to the same procedure as written in 4.3.2. This time, approximately one month of ligation reactions were attempted before successfully generating a pCMV-Trunc.AGPAT4-HA Tag plasmid. Since the BglII and BamH1 overhangs had formed a new indigestible site, I moved upstream to the next available restriction site recognized by the enzyme Not I to perform the

following digest. The final gel electrophoresis run showed a band corresponding to the size of linearized pCMV-3TAG-3A (~4.2 kb) and a second band corresponding to the predicted size of the insert (870 bp), indicating a high likelihood that this bacterial clone contained the successfully subcloned plasmid. The representative UV image is depicted in Figure 11.

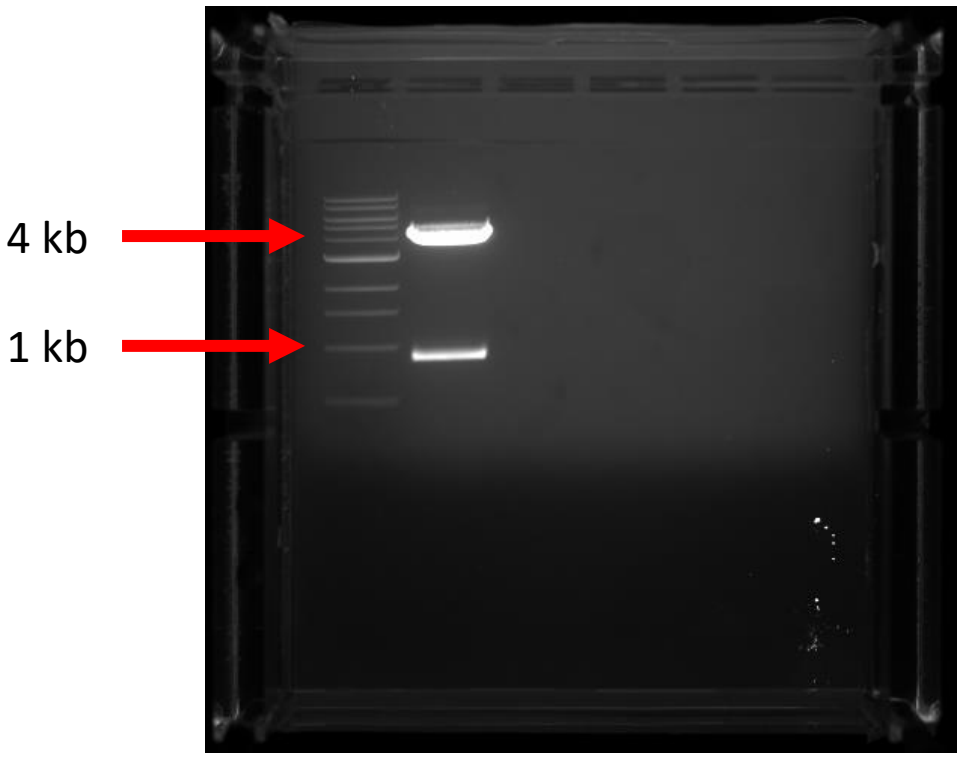


Figure 11: UV gel image depicting restriction digest attempting to identify DH5-α colonies which contained the desired pCMV-Trunc. AGPAT4-HA Tag plasmid. The upper band was approximately 4.2 kb in size which is indicative of the pCMV-3TAG-3A host vector. The lower band was once again located underneath the 1 kb ladder marker, indicative of the 870 bp Bgl II-Trunc. AGPAT4-HA Tag-Xho I insert. The digest was performed using Not I and Xho I restriction enzymes.

4.3 Results

After confirmation via gel electrophoresis that the pCMV vector had been ligated successfully with the Trunc. AGPAT4-HA Tag insert, the plasmid was sent out for sequencing at the TCAG sequencing facility. Sequencing results are depicted in Figure 17 in the Appendix. The sequencing was reviewed to ensure the plasmid was accurately encoding the desired Bgl II and Xho I restriction sites, truncated AGPAT4 protein, hemagglutinin tag, and newly inserted stop codon.

4.4 Discussion

After roughly one term of molecular cloning, a novel pCMV-Trunc.AGPAT4-HA Tag plasmid was successfully generated, and a large stock was made for the Duncan Lab inventory. By replacing the original GFP tag from the starting pEGFP-Trunc. AGPAT4 plasmid, the hope was to streamline future work involving truncated AGPAT4 by eliminating the possible negative effects of the large GFP tag outside of confocal microscopy. It was necessary to perform the cloning in two separate phases, where the first phase was to integrate the Bgl II-Trunc. AGPAT4-Ha Tag-Xho I insert into the pGEM-T-Easy cloning vector, and the second phase was to integrate the insert into a pCMV-3TAG-3A vector. The pGEM-T-Easy vector is a bacterial expression vector, meaning it is specifically for use in transformed bacteria [60]. If one were to transfect HEK-293 cells with pGEM vectors containing a gene of interest, the cells would not express anything, since the vector lacks a mammalian promoter. The resulting protein encoded by pCMV-3TAG-3A would be the truncated AGPAT4 variant, with a hemagglutinin (HA) tag. It should be noted that pCMV-3TAG-3A contains a 3x FLAG tag. However, the insert that was used was designed with a stop codon after the HA tag, and therefore the 3x FLAG tag would not

be expressed or attached to the encoded protein. Despite the need for this additional cloning consideration, this vector was used since it offers a high level of expression in mammalian cells.

Since two subcloning steps were utilized, some discussion is merited. It is possible to proceed directly from synthesis of the PCR amplicon directly into subcloning into the pCMV-3TAG-3A vector. While this may seem like the most direct method, the efficiency of uptake of synthetic DNA from PCR is lower than the uptake by bacteria of DNA from another plasmid vector that was amplified in bacteria. As a result, cloning into a simple vector, such as pGEM-T-Easy, is frequently a first step that is followed by subsequent subcloning. A feature of taq-mediated PCR is the fact that taq polymerase will attach an additional adenine group to the 3' ends of the PCR product [61]. The T in pGEM-T-Easy stands for a thymine group which is specifically included at the 3' ends of the pGEM cloning vector sequence to make subcloning PCR fragments, with their additional complementary adenine overhangs, much simpler, and typically with a higher rate of success [60]. The major challenge evident during subcloning attempts was the random chance of successful ligation between DNA insert and host vector, and the low concentration of insert available from PCR was limiting in this process. Once a ligation reaction is set up, the researcher is left at the mercy of hoping the T4 DNA Ligase enzyme can manage to adhere the plasmid together in solution. Despite the difficulties, the desired plasmid was accurately synthesized in this chapter and future work involving expressing truncated AGPAT4-HA could be done.

Chapter 5

Investigation of a protein-protein interaction between reference AGPAT4 and Trunc. AGPAT4.

5.1 Rationale and Objective

Investigating a direct interaction between the truncated and reference variants of AGPAT4 is important for understanding the potential physiological role that Trunc. AGPAT4 plays. Other enzymes with endogenously synthesized, truncated variants have been shown in the literature to directly interact with their reference counterparts to inhibit enzyme activity [45]. Determining whether dimerization occurs between the two proteins will help to identify part of the mechanism by which truncated AGPAT4 acts. These functions may include enzyme activation or inhibition via dimerization, or in the case that interaction does not occur, then direct competition between the two enzymes may be considered. Interaction studies were conducted *in vitro* by overexpression of reference AGPAT4 and truncated AGPAT4 using plasmids in the HEK-293 cell line.

Principal Objective: To determine whether the truncated AGPAT4 protein interacts with the reference AGPAT4.

5.2 Hypothesis

Truncated AGPAT4 interacts with reference AGPAT4, and the proteins will pull-down together in co-immunoprecipitation or IMAC affinity purification assays.

5.3 Materials, Methods, and Study Approach

5.3.1 Overview of Approach

In order to determine whether an interaction exists between the truncated and reference variants of AGPAT4, two plasmids expressing either truncated AGPAT4 or reference AGPAT4, each with a unique protein tag, were transfected into Human Embryonic Kidney (HEK-293) cells. Specific plasmids and quantities used are listed in Tables 3a, 3b, and 3c. Transfected cells were cultured, harvested, and lysed, with the subsequent cell lysates being used in an immunoprecipitation protocol to pull down for their respective protein tags. If an interaction exists, immunoprecipitation for an HA tag (attached to Trunc. AGPAT4) and immunoblotting for FLAG tag (attached to Full. AGPAT4) would reveal a protein band on the final blot, and vice versa.

5.3.2 Cell Culture and Transfection

HEK-293 cells were plated on 10 cm treated tissue culture dishes containing growth media consisting of high glucose Dulbecco's Modified Eagle's Medium (DMEM), 10% fetal bovine serum (FBS), and 1% penicillin streptomycin (PS) [62]. Cultures were maintained at 37^{o}C with 5% CO_2 in a Thermo Fisher Scientific Heratherm^{TM} Microbiological Incubator. HEK-293 cultures were regularly passaged after reaching ~80% confluency using Trypsin-EDTA. Transfection of HEK-293 cells used for co-immunoprecipitation, immobilized metal affinity chromatography (IMAC), and immunoblotting were performed using jetPRIME®, according to the manufacturer's instructions.

5.3.3 Immobilized Metal Affinity Chromatography (IMAC) and Co-Immunoprecipitation (Co-IP)

The following protocol is adapted from the Immunoprecipitation (IP) Protocol by Abcam [63]. Two HEK-293 cultures at ~80% confluency were transfected, one with a combination of pCMV- Trunc.AGPAT4-HA tag and pCMV-Full.AGPAT4-6His tag (experimental group for IMAC) or pCMV-Full.AGPAT4-FLAG tag (experimental group for Co-IP) and the second with only pCMV- Trunc.AGPAT4-HA tag and pCMV-3-TAG-3A as an empty vector control to bring DNA concentrations used in transfections to an equivalent level, using jetPRIME® transfection reagent. Following a 24-hour incubation, transfected cells were harvested using trypsin-EDTA, and the trypsinization reaction was quenched using DMEM containing 10% FBS. Cells were transferred from their respective culture dishes to 15 mL Falcon tubes and centrifuged at 1000 RPM for 5 minutes to pellet cells. Excess DMEM was suctioned off without disturbing the cell pellets. Using ice cold 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, ddH_2O to 1 L, pH 7.4) solution, cell pellets were washed and re-pelleted by centrifugation at 1000 rpm for 3 minutes, and washes were repeated until the PBS was clear and colourless, indicating that the trace phenol-red indicator dye from cell media has been washed away and cell samples were ready for either IMAC or Co-IP procedures.

For the IMAC procedure, excess PBS was carefully pipetted off the cell pellets and pellets were resuspended in equilibration buffer (0.5 M 1x PBS, 10 mM Imidazole), then transferred to 1.5 mL microcentrifuge tubes. Cells were lysed on a slurry of ice water using a sonicator at medium-high setting by 6 short pulses and 3 seconds of rest between each pulse. Sonicated cells were centrifuged at 3000 × g for 10 minutes at 4^{o}C to pellet cell debris. As samples were centrifuging, 50 µL per sample of Ni-NTA beads were washed with ddH_2O and

briefly centrifuged to pellet the beads. The ddH_2O was pipetted off and beads were washed and briefly centrifuged again with equilibration buffer to pellet the beads. The wash equilibration buffer was pipetted off, then fresh equilibration buffer was added to the beads at 100 µL of equilibration buffer per sample. Once the 10-minute centrifugation of the sonicated cells was completed, supernatant was transferred to 1.5 mL microcentrifuge tubes and 100 µL of beads suspended in equilibration buffer was added to each sample. Samples were gently rocked at 4^{o}C for 1 hour. After rocking, beads were centrifuged at 1500 × g for 5 minutes and supernatant saved for immunoblotting. Ni-NTA beads were then ready for immunoblotting procedure.

For the Co-IP procedure, excess PBS was carefully pipetted off the cell pellets and pellets were resuspended in IP Buffer, pH 7.4 (150 mM NaCl, 5 mM EDTA pH 8.0, 50 mM Tris pH 7.4, 1% v/v Triton-X 100) containing 1% v/v protease inhibitor cocktail (PIC) then transferred to 1.5 mL microcentrifuge tubes. Cell samples were slowly lysed over a duration of 15 minutes on ice, with vigorous vortexing every 5 minutes. Lysed cells were centrifuged at 12,000 × g for 20 minutes at 4^{o}C to separate cell debris from the samples. The resulting supernatant was carefully pipetted out into new 1.5 mL microcentrifuge tubes. Twenty microlitres of supernatant were separated into different 1.5 mL tubes and stored at -80^{o}C for use in immunoblotting as input lanes. Five microlitres of Protein G beads were added to each of the remaining samples, and samples were gently rocked at 4^{o}C for 30 minutes to preclear any non-specific proteins binding to the Protein G beads. Samples were then centrifuged at 12,000 × g for 30 seconds at 4^{o}C to pull down the Protein G beads, and the precleared supernatant was carefully pipetted out into new 1.5 mL microcentrifuge tubes. Five microlitres of the appropriate antibodies were added to the precleared supernatant and rocked gently overnight at 4^{o}C. The following day, 20 µL of Protein G beads were added to the samples and rocked gently at 4^{o}C for 4 hours. Samples were then

centrifuged at 12,000 × g for 30 seconds at 4°C to pull down the Protein G bead-antibody-protein complex and supernatant was removed carefully without disturbing the beads. The bead-antibody-protein complex was washed with ice-cold IP buffer without PIC by gentle inversion, then centrifuged at 12,000 × g for 30 seconds at 4°C to pellet the beads. The wash IP buffer was carefully removed. The wash steps were repeated 4 additional times to remove as much unreacted protein from the samples as possible. Samples were then ready for immunoblotting.

5.3.4 Immunoblotting

Samples were mixed at a 1:1 ratio with 2x Laemmli Sample Buffer (40 µL of 10% SDS, 120 µL of 1 M Tris at pH 6.8, 20 µL of 1% bromophenol blue, 260 µL of ddH_2O, and 20 µL of fresh 1 M dithiothreitol) and boiled for 5 minutes at 95°C. Samples will then be subjected to electrophoresis through SDS-PAGE gels. Following appropriate resolution indicated by a 10-250 kD protein standard, samples were transferred to a nitrocellulose membrane using a semi-dry transfer system. Membranes were blocked using either 5% milk-1x TBST (20 mM Tris Base, 150 mM NaCl, 0.1% v/v Tween 20, ddH_2O to 1 L, pH 7.4) or 5% bovine serum albumin (BSA)-1x TBST for 1-2 hours at room temperature. The choice of blocking solution used was dependent on the antibody manufacturer's recommendation. The following day, the blot was washed using 1x TBST solution 3 times, 5 minutes per wash at room temperature. Primary antibody was diluted 1:1000 in either 5% milk-1x TBST or 5% BSA-1x TBST depending on the blocking solution used. The blot was incubated in primary antibody solution overnight at 4°C. Following primary antibody incubation, the blot was washed 3 times, 5 minutes per wash at room temperature with 1x TBST solution. Secondary horseradish peroxidase conjugated antibody was diluted 1:5000 in 5% milk-1x TBST or 5% BSA-1x TBST as per the blocking solution. The blot was incubated for 2-3 hours at room temperature in secondary antibody solution, then washed

again in 1x TBST 3 times, 5 minutes per wash at room temperature. Finally, the blot was incubated for 1-2 minutes at room temperature in Luminata® chemiluminescence solution [64] then imaged using a Bio-Rad ChemiDoc Imaging System to view the resultant protein bands.

5.4 Results

5.4.1 Truncated AGPAT4-HA does not pull-down with Reference AGPAT4-HIS

The immunoblots depicted below were selected from numerous attempts at protein pull-downs. Figure 12 on the following page depicts an immunoblot using Immobilized Metal Affinity Chromatography as the protein pull-down method, where the nickel beads attract the 6x HIS tag attached to the C-terminal end of the reference AGPAT4-6HIS. The final immunoblot was incubated with an anti-HA Tag antibody to detect the interaction protein, Truncated AGPAT4-HA Tag. The input lanes (1 and 2) show a signal for the presence of Trunc. AGPAT4-HA Tag, but both the lane containing the negative control and the interaction lanes (3 and 4, respectively) contain no signal. Interestingly, we noticed that input lane 1, where both reference and truncated AGPAT4 are co-expressed, showed a reduction in chemiluminescent signal relative to input lane 2 where truncated AGPAT4 is expressed with a control plasmid. Table 5 shown below the immunoblot details the contents of each numbered lane in the immunoblot experiment.

Blot #1: IMAC HIS Tag, Blot for HA

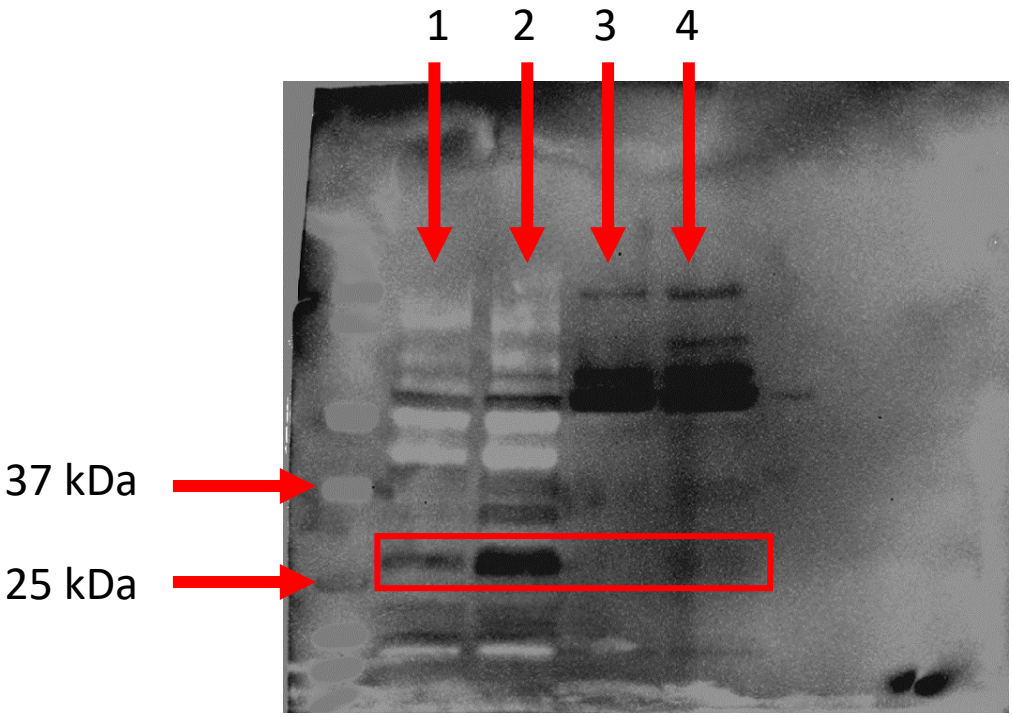


Figure 12: Immunoblot depicting IMAC pulldown of the 6xHIS tag in the full-length AGPAT4-6HIS protein, from cell lysates transfected both with pCMV-Full-*Agpat4-HIS* and pCMV-*Trunc. Agpat4-HA tag*, and blotted for HA tag. The predicted size of the truncated AGPAT4-HA construct is 31 kDa and can be visualized clearly in lanes 1 and 2. However, although equal amounts of the plasmid encoding for truncated AGPAT4-HA were transfected into cells in lanes 1 and 2, when a plasmid encoding full-length AGPAT4-6HIS is co-transfected (lane 1), levels of the truncated variant are much lower In lane 4, a specific increase in truncated AGPAT4-HA would demonstrate an interaction with AGPAT4-6HIS, and this is not seen. Lane 3 shows a negative control for the IMAC experiment, where a sample containing only the truncated AGPAT4-HA protein was subjected to IMAC to ensure it was not non-specifically binding to the nickel beads without the presence of histidine.

Table 5: Cell transfection combinations and Immobilized Metal Affinity Chromatography conditions for the contents of Lanes 1 through 4 in the above immunoblot.

	Transfection Contents and IMAC Conditions
Lane #1	Crude Lysates pCMV-Full-*Agpat4-HIS* **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #2	Crude Lysates pCMV-3TAG-3A **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #3	IMAC pCMV-3TAG-3A **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #4	IMAC pCMV-Full-*Agpat4-HIS* **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**

5.4.2 Reference AGPAT4-FLAG does not co-immunoprecipitate with Truncated AGPAT4-HA

Figure 13 on the following page shows an immunoblot where co-immunoprecipitation with an anti-FLAG Tag antibody was used as the protein pull-down technique. The final immunoblot was incubated with an anti-HA Tag antibody for detection. The input lanes 1 and 2 show a similar pattern as the blot in section 5.4.1, where the input lane #2 containing the two co-expressed AGPAT4 variants displays a reduction in chemiluminescent signal relative to lane #1 containing only the truncated AGPAT4-HA variant. Negative control lane 3 and protein interaction lane 4 both do not display signs of chemiluminescent signal, suggesting no interaction is occurring. Table 6 shown below the immunoblot details the contents of each numbered lane in the immunoblot experiment.

Blot #2: Co-IP FLAG Tag, Blot for HA Tag

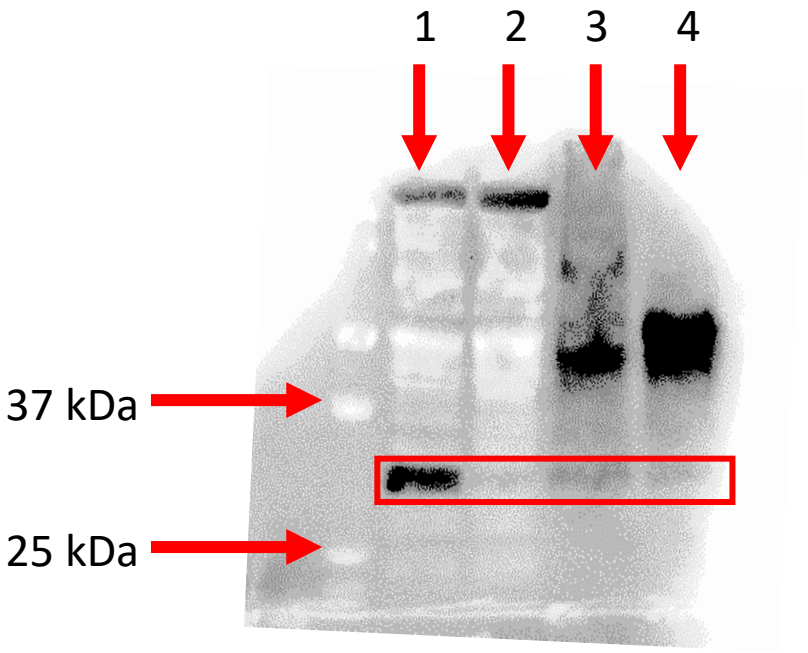


Figure 13: Immunoblot depicting a representative co-immunoprecipitation experiment. In this experiment the reference AGPAT4-FLAG homologue was expressed in cells together with the truncated AGPAT4 variant (tagged with HA), and cells were immunoprecipitated for FLAG, and then blotted for HA tag. The approximate 31 kDa band of truncated AGPAT4-HA tag is clearly visible in lane 1 where it was transfected alone, but the truncated AGPAT4-HA tag band in lane 2 is extremely faint when co-transfected with full-length AGPAT4-FLAG tag, indicating lower levels of the truncated variant. In lane 4, a specific increase in truncated AGPAT4-HA would demonstrate an interaction with FLAG-tagged reference AGPAT4, and this is not seen. Lane 3 shows a negative control for the co-immunoprecipitation experiment, where lysates do not contain any HA-tagged protein.

Table 6: Cell transfection combinations and Co-Immunoprecipitation conditions for the contents of Lanes 1 through 4 in the above immunoblot.

	Transfection Contents and Co-IP Conditions
Lane #1	Crude Lysates pCMV-3TAG-3A **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #2	Crude Lysates pCMV-Full-*Agpat4-FLAG* **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #3	Co-IP with non-specific antibody pCMV- Full-*Agpat4-FLAG* **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**
Lane #4	Co-IP with Anti-FLAG Antibody pCMV-Full-*Agpat4-FLAG* **(5 µg)** pCMV-*Trunc. Agpat4-HA* **(5 µg)**

5.4.3 Truncated AGPAT4-GFP does not co-immunoprecipitate with Reference AGPAT4-FLAG

Figure 14 on the following page shows the final variation of co-immunoprecipitation that was attempted, where an anti-GFP antibody was used instead of the typical anti-FLAG tag antibody. The resultant immunoblot was incubated with an anti-FLAG antibody. Similar to the previous blots, the input lane #2 containing the co-expressed proteins displays a weaker band relative to the input lane #1 containing AGPAT4-FLAG and a control plasmid. The negative control lane #3 and the protein interaction lane #4 both do not display any evidence of chemiluminescent signal, indicative of no interaction occurring. Table 7 shown below the immunoblot details the contents of each numbered lane in the immunoblot experiment.

Blot #3: Co-IP GFP Tag, Blot FLAG Tag

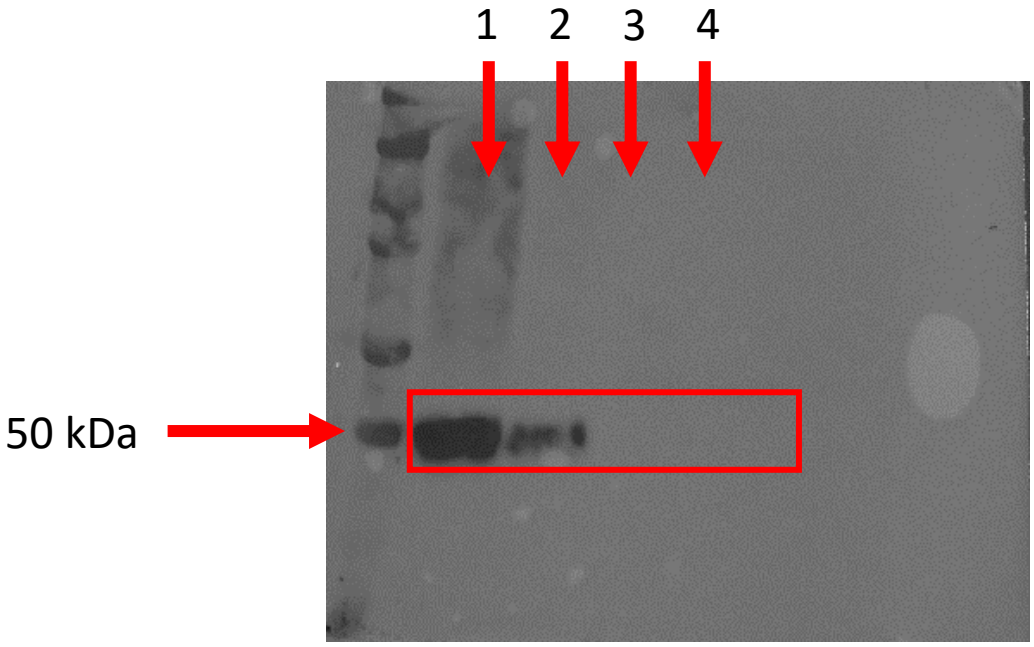


Figure 14: Immunoblot depicting a representative co-immunoprecipitation experiment. In this experiment the reference AGPAT4-FLAG homologue was expressed in cells together with the truncated AGPAT4 variant (tagged with GFP), and cells were immunoprecipitated for GFP, and then blotted for FLAG tag. The approximate 47 kDa band of AGPAT4-FLAG tag is clearly visible in lane 1 where it was transfected alone, but the AGPAT4-FLAG tag band in lane 2 displays a fainter signal when co-transfected with truncated AGPAT4-GFP tag, indicating lower levels of the AGPAT4-FLAG homologue. In lane 4, a specific increase in AGPAT4-FLAG would demonstrate an interaction with GFP-tagged truncated AGPAT4, and this is not seen. Lane 3 shows a negative control for the co-immunoprecipitation experiment, where lysates do not contain any FLAG-tagged protein.

Table 7: Cell transfection combinations and Co-Immunoprecipitation conditions for the contents of Lanes 1 through 4 in the above immunoblot.

	Transfection Contents and Co-IP Conditions
Lane #1	Crude Lysates pCMV-3TAG-3A **(5 µg)** pCMV-Full-*Agpat4-FLAG* **(5 µg)**
Lane #2	Crude Lysates pCMV-Full-*Agpat4-FLAG* **(5 µg)** pEGFP-*Trunc. Agpat4* **(5 µg)**
Lane #3	Co-IP with non-specific antibody pCMV- Full-*Agpat4-FLAG* **(5 µg)** pEGFP-*Trunc. Agpat4* **(5 µg)**
Lane #4	Co-IP with Anti-GFP Antibody pCMV-Full-*Agpat4-FLAG* **(5 µg)** pEGFP-*Trunc. Agpat4* **(5 µg)**

5.5 Discussion

To investigate a potential protein-protein interaction between truncated and reference AGPAT4, HEK-293 cells were transfected with a combination of plasmids listed in Tables 5, 6, and 7, and then purified with either Immobilized Metal Affinity Chromatography or co-immunoprecipitation techniques and, finally, immunoblotted for the respective tags on either truncated AGPAT4-HA, reference AGPAT4-FLAG, or reference AGPAT4-6His. The purpose of using two separate techniques lies in the key difference between the two: IMAC uses nickel beads which strongly bind to the polyhistidine tags present on the desired chimeric proteins, while Co-IP relies on antibodies and their unique affinities to their target proteins [65, 66]. During the early stages of this chapter's work, only co-immunoprecipitation was being used to attempt a protein interaction pull-down. However, concerns regarding non-specific antibody binding during the co-immunoprecipitation experiments led to attempting the Immobilized Metal Affinity Chromatography (IMAC) pull-down instead. IMAC procedures are marginally more reliable due to the natural affinity that polyhistidine tags have for nickel; histidine strongly chelates nickel, and this pull-down technique attempts to exploit that fact [65]. Since crude cell lysates were being used for these protocols, the likelihood of the antibodies used in co-immunoprecipitation randomly binding to proteins and showing false bands on the final immunoblot would be at a higher risk. Additionally, IMAC eliminates the presence of heavy and light immunoglobulin bands associated with IP on immunoblots, which can disrupt imaging.

Analysis of immunoblot results did not yield any sign of direct protein-protein interaction between reference and truncated AGPAT4, since there were no clearly discernable bands in the protein purified interaction lanes in any final immunoblotting products, which are represented in Lane #4 throughout all experiments. In some of the non-depicted immunoblots there was a band

in the final interaction lane. However, in all instances this was accompanied by a protein band also in the negative control Lane #3, indicating that this was a non-specific interaction. The fact that both of these protein isolation techniques were used, and neither protocol clearly showed a difference, strongly suggests that the reference form of AGPAT4 and the truncated AGPAT4 proteins do not interact. However, my protocol was limited to detection of strong protein-protein interactions, as methods of chemical cross-linking between proteins for detection of weak or transient interactions were not used [67]. It is possible that the IP buffer used to lyse my cells contained one or more reagent(s) that disrupted weak bonding between proteins.

Although it was disappointing to find a lack of interaction, another interesting pattern emerged across separate immunoblot experiments. Examination of Lanes 1 and 2 showed evidence of a reduction in the chemiluminescent signal in the labelled "Input" lanes when both plasmids were co-expressed. When blotting for reference AGPAT4-FLAG, reference AGPAT4-6HIS, or truncated AGPAT4-HA, the input lane containing a combination of the truncated and reference proteins (i.e. Lane #2) displayed a weaker signal than the input lane with only one of either of the reference or truncated proteins and a control plasmid (i.e. Lane #1). Notably, this occurred despite the use of equal plasmid amounts during cell transfections (i.e. 5 μg of each plasmid was used per plate, for a total of 10 μg of transfected plasmids in total). However, it should also be noted that in these lanes, protein concentrations were not checked or normalized prior to loading. When plates were harvested, crude lysates were prepared for pull-down assays, and both plates were treated in an identical manner. Typically, prior to pull-down, a portion of the lysate (typically 50 μl) of the 1 mL of clarified lysate, would be set aside for loading in the 'input' control lanes (i.e. lanes 1 and 2). Thus, initially, we believed that the immunoblotting pattern, showing lower expression of either the reference or truncated protein, indicated a

possible degradation effect that may be linked to the simultaneous presence of both the truncated and the reference AGPAT4 proteins. However, it was also possible that the reduction in signal intensity in immunoblotting could have resulted from a reduction in cell numbers, since equal protein concentrations were not verified for loading. In Chapter 6, experiments investigate the former idea (that co-expression affected levels each of the other protein). However, based on results from Chapters 5 and 6 combined, the latter notion, that co-expression of the two variant proteins affects the number of cells, appears plausible. Although not directly investigated in this thesis, further work exploring this idea is proposed in the Future Directions section.

To conclude, this study did not find any evidence of interaction between the truncated and reference AGPAT4 proteins. It is, however, important to note here that both the co-immunoprecipitation and IMAC protocols did not account for weak or transient interaction between the two proteins, but rather only investigated for strong interactions. However, a possible modulation effect was observed in immunoblot results, and was followed-up in the next study.

Chapter 6

Investigation of effects of AGPAT4 protein variants on protein expression of the alternate isoform

6.1 Rationale and Objectives

Based on preliminary immunoblot data produced in Chapter 6, it appeared that co-expression of truncated AGPAT4 decreased protein expression of reference AGPAT4, and *vice versa*. This was rationalized as potentially indicating that Trunc.AGPAT4 may be induced to function in the regulation of cellular AGPAT4 activity by targeting reference AGPAT4 for degradation. The work in this chapter therefore investigates whether Trunc.AGPAT4 does, indeed, reduce levels of reference AGPAT4 in cells, and also whether reference AGPAT4 affects levels of Trunc.AGPAT4. Experiments investigating the potential modulation effects of reference and truncated AGPAT4 will be conducted *in vitro* using plasmids expressing reference and truncated AGPAT4 proteins in HEK-293 cells.

Primary Objective: To determine whether truncated AGPAT4 protein regulates levels of reference AGPAT4 protein in cells, and vice versa.

6.2 Hypothesis

Increasing expression of truncated AGPAT4 will reduce levels of reference AGPAT4 *in vitro,* while increasing expression of reference AGPAT4 will reduce levels of truncated AGPAT4.

6.3 Materials, Methods, and Study Design

6.3.1 Study Overview

Studies to determine if truncated AGPAT4 can reduce reference AGPAT4 levels *in vitro* were performed through a series of scaled cell transfections, where either the reference or truncated plasmid was transfected at a single, constant level of 5 ug per plate, and the other variant was transfected at increasing levels (with a control plasmid used to ensure that total DNA transfected remained constant at 10 ug per plate). In the first experimental paradigm, five plates of HEK-293 cells were transfected with specific combinations of pCMV-3TAG-3A (used as a control plasmid to ensure transfection of equal amounts of DNA) and increasing amounts of pCMV-Trunc. AGPAT4-HA tag, together with 5 μg pCMV-Full AGPAT4-FLAG tag per plate, as described in Table 8b shown on the following page. Following a 24-hour incubation, transfected HEK-293 cells were harvested and lysed, then the crude cell lysates were immunoblotted for FLAG tags to visualize Full. AGPAT4-FLAG protein levels. The inverse experimental paradigm was also performed, where pCMV-3-TAG-3A was co-transfected with increasing amounts of pCMV-Full AGPAT4-FLAG, and 5 μg pCMV-Trunc.AGPAT4-HA tag and, as described in Table 8a shown on the following page. Lysates were then immunoblotted for HA tags to observe levels of the truncated variant. Cell culturing and transfection methodology and immunoblotting methodology was essentially as described in Chapters 5.3.2 and 5.3.4, respectively.

Table 8a: List of plasmid combinations/concentrations for transfections to be immunoblotted for **HA Tag.**

Plasmid	**Plate #1**	**Plate #2**	**Plate #3**	**Plate #4**	**Plate #5**
pCMV-3TAG-3A	5 µg	4 µg	3 µg	2 µg	0 µg
pCMV-*Trunc. Agpat4-HA tag*	5 µg	5 µg	5 µg	5 µg	5 µg
pCMV-*Full Agpat4-FLAG tag*	0 µg	1 µg	2 µg	3 µg	5 µg

Table 8b: List of plasmid combinations/concentrations for transfections to be immunoblotted for **FLAG Tag.**

Plasmid	**Plate #1**	**Plate #2**	**Plate #3**	**Plate #4**	**Plate #5**
pCMV-3TAG-3A	5 µg	4 µg	3 µg	2 µg	0 µg
pCMV-*Trunc. Agpat4-HA tag*	0 µg	1 µg	2 µg	3 µg	5 µg
pCMV-*Full Agpat4-FLAG tag*	5 µg	5 µg	5 µg	5 µg	5 µg

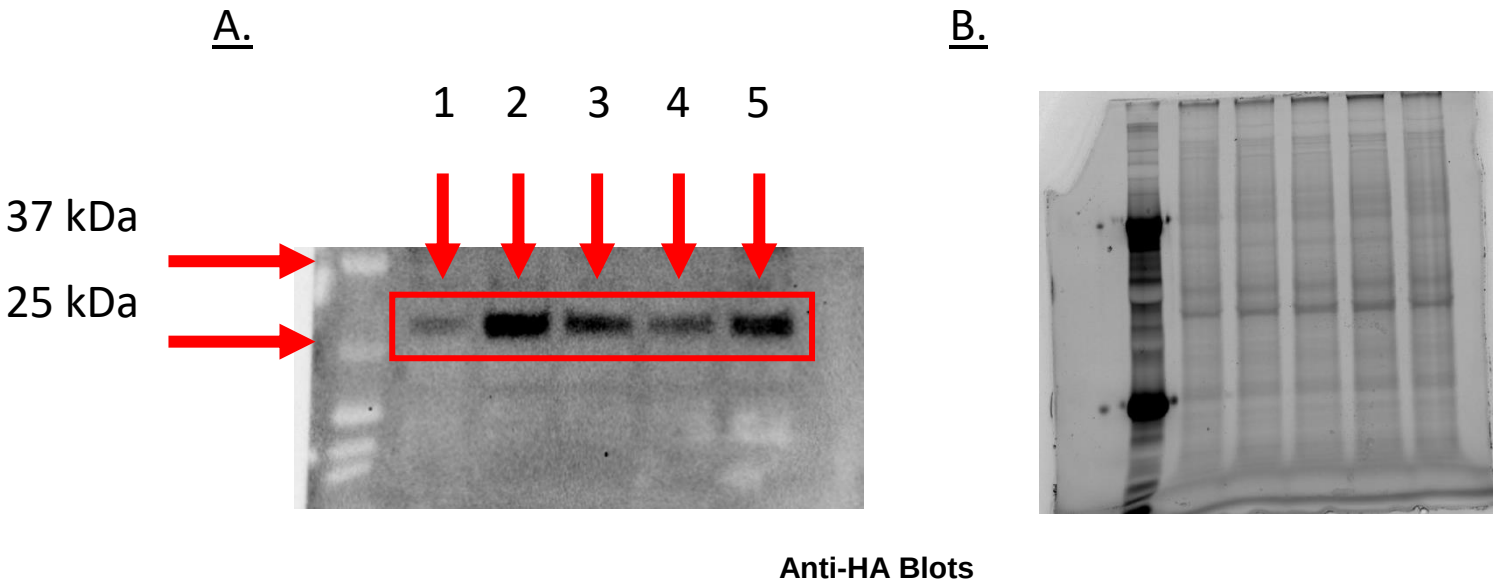
6.4 Statistical Analysis

Chemiluminescent signal was normalized to stain-free gels, then compared by 1-way ANOVA with multiple comparisons, and differences were analyzed by Tukey's post-hoc test.

6.5 Results

6.5.1 Expression of Reference AGPAT4 modulates immunodetectable levels of Truncated AGPAT4

When transfecting HEK-293 cells with a constant amount of plasmid expressing truncated AGPAT4 with a gradually increasing amount of plasmid expressing reference AGPAT4, a pattern of modulation of the truncated AGPAT4-HA protein was observed in immunoblots. The addition of 1 μL of reference pCMV-Full. AGPAT4-6HIS plasmid significantly increased the immunodetectable level of the truncated AGPAT4-HA, as indicated by an increase in band density. Furthermore, there appears to be a dose-dependent nature to this effect, where further increases in co-transfection of cells with pCMV-Full. AGPAT4-6HIS, while still increasing immunodetectable truncated-APGAT4-HA levels over levels seen when no reference AGPAT 4 was added, did so to a significantly lesser extent. An interesting finding consistent throughout three independent immunoblot experiments under the same parameters was observed in the final lane (Figure 15A), which is representative of an equal transfection amount of *reference Agpat4* plasmid and *truncated Agpat4* plasmid (5 μL each). This band was observed to increase truncated-AGPAT4 in calculated density compared to cells expressing only the truncated-AGPAT4 form alone, which was different from results in Chapter 5. Relevant images and data are presented in Figure 15 on the following page.



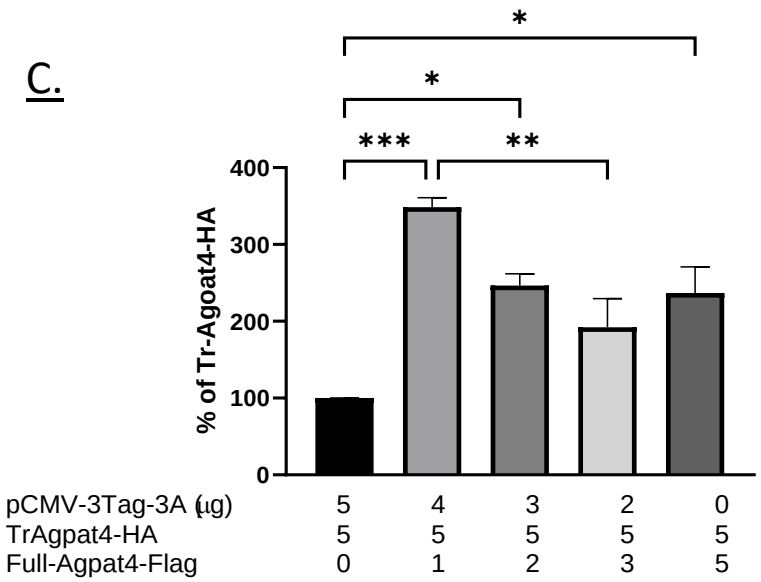


Figure 15: **A.** Immunoblot image depicting crude cell lysates blotted with an anti-HA Tag antibody. 31 kDa bands representative of truncated-AGPAT4-HA Tag are denoted by the red box. **B.** Corresponding stain-free gel used to normalize band densities to loaded protein concentrations. **C.** Band density analysis of the anti-HA Tag immunoblots (N=4) represented in % of Truncated AGPAT4-HA Tag seen on the blots. Data are means ± S.E.M. *P<0.05, **P<0.01, ***P<0.001.

Note: Columns in the graph depicted in Figure 15C. from left to right are respective of Lanes 1 through 5 in Figure 15A.

6.5.2 Truncated AGPAT4 does not significantly modulate immunodetectable levels of Reference AGPAT4

For the second phase of this experiment, HEK-293 cells were once again transfected, however, this time with a constant amount of pCMV-full-AGPAT4-FLAG Tag plasmid and a stepwise increase of pCMV-truncated AGPAT4-HA Tag plasmid. Under these parameters, immunoblotting for the reference AGPAT4-FLAG chimera yielded no significant differences. Relevant images and data are presented in figure 16 on the following page.

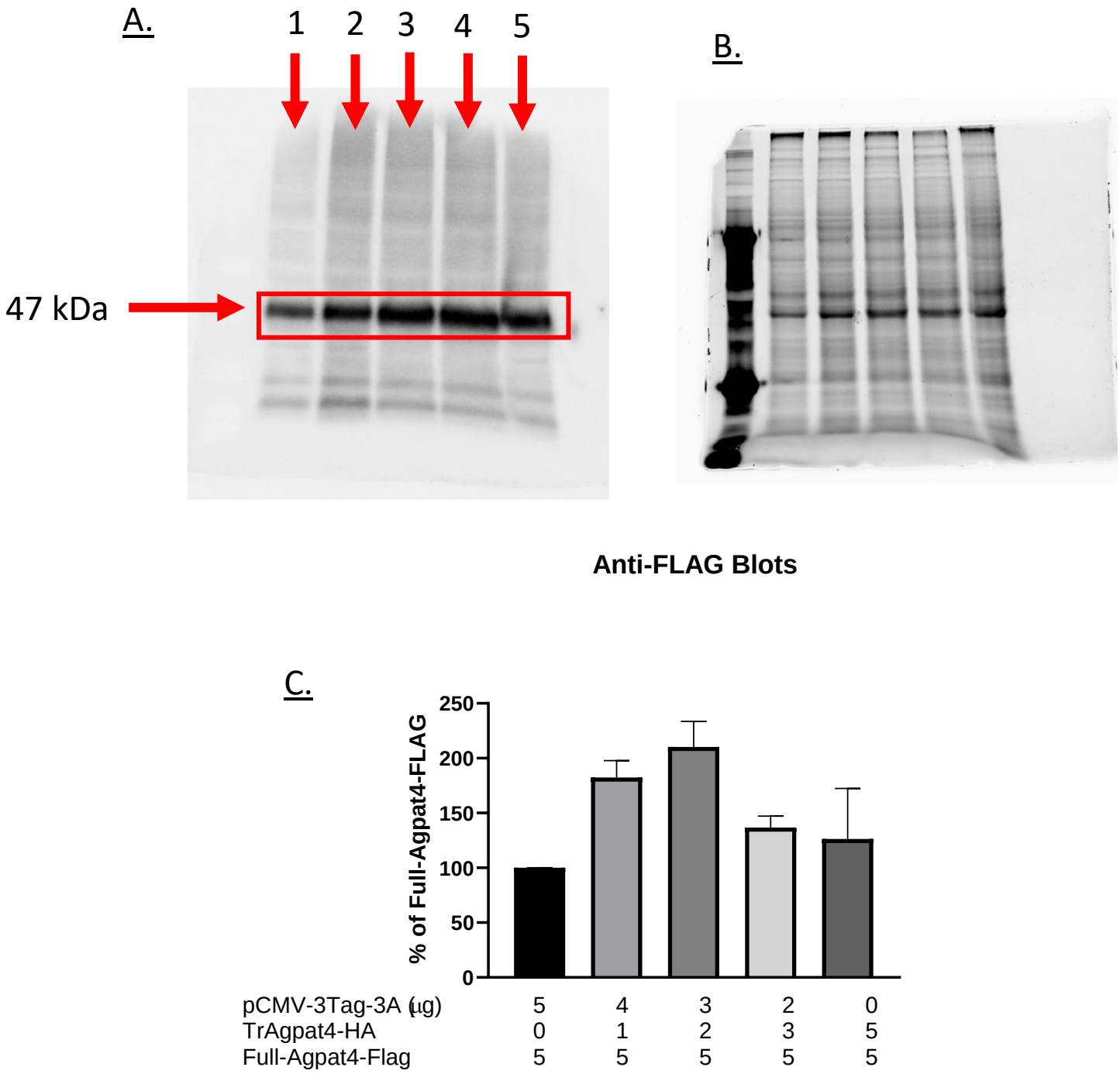


Figure 16: **A.** Immunoblot image depicting crude cell lysates blotted with an anti-FLAG Tag antibody. 47 kDa bands representative of AGPAT4-FLAG Tag are denoted by the red box. **B.** Corresponding stain-free gel used to normalize band densities to loaded protein concentrations. **C.** Band density analysis of the anti-FLAG Tag immunoblots (N=3) represented in % of Full-AGPAT4-FLAG Tag seen on the blots, and normalized to densities of the corresponding lane on the stain-free gel. Data are means ± S.E.M. There are no significant differences observed between any groups.

Note: Columns in the graph depicted in Figure 16C. from left to right are respective of Lanes 1 through 5 in Figure 16A.

6.6 Discussion

To study a potential role for truncated AGPAT4 in cell physiology, I designed experiments to further investigate the findings from prior co-immunoprecipitation and IMAC studies, which had suggested that the truncated AGPAT4 variant may act to regulate levels of the reference AGPAT4 homologue, and *vice versa*. Plasmid vectors expressing either truncated AGPAT4-HA or reference AGPAT4-6HIS were co-transfected in varying amounts (Tables 8a and 8b) into HEK-293 cells for 24 hours. The cells were then lysed, and the resulting lysates were used in immunoblotting experiments.

Based on the results, the reference AGPAT4 protein appears to play a role in positively modulating levels of the truncated AGPAT4 protein. An interesting finding from this section can be seen by comparing Lanes 1 and 2 of the anti-HA tag immunoblot analysis shown in section 6.5.1. There is significantly ($p=0.0002$) more truncated AGPAT4-HA detected when 1 μg of reference AGPAT4 plasmid was co-transfected with 5 μg of truncated AGPAT4 compared to when no reference AGPAT4 plasmid was co-transfected. This suggests that reference AGPAT4-6HIS helps to stabilize or prevent degradation of the truncated AGPAT4-HA. However, this effect was largest when the plasmids encoding the truncated and reference forms were co-transfected at a 5:1 ratio. When the amount of vector expressing reference AGPAT4-6HIS was increased to a higher ratio, this effect was less pronounced, although still statistically significant in some groups.

The reason for the increase in observed quantity of the truncated protein in the presence of the reference protein is not currently known. One possible reason for this observation is that the truncated AGPAT4 variant may be unstable by itself and breaks down readily in the cells and/or in solution. There are documented examples of proteins that are structurally and

functionally obligate, and do not have long half-lives when they are isolated outside of their quaternary structure [68]. Addition of the reference variant may cause some binding heterocomplex to occur before degradation is triggered. However, it is notable that co-IP experiments failed to detect an interaction, indicating that if this does, indeed, occur, it is likely transient and/or weak. A second interesting observation in the immunoblots characterizing the stability of the truncated AGPAT4-HA variant comes from Lane #5, where 5 μg of truncated AGPAT4 is co-transfected with 5 μg of reference AGPAT4, and this results in significantly more truncated AGPAT4-HA than in cells not co-transfected with reference AGPAT4. This was different than the effect observed in the earlier studies on AGPAT4 co-IP and IMAC, where co-transfection at equal levels resulted in *apparent decreases* in both reference and truncated proteins. The reason for this discrepancy is not certain. Speculative reasoning for such a phenomenon may be that rather than dimerization between reference and truncated AGPAT4 leading to decreasing levels, there may actually be formation of a trimer, either with reference AGPAT4 or a third unknown protein. Ternary degradation structures have been observed, and the presence of too many of the enzymes involved in binding and breakdown may lead to what is known as a "hook effect" [69]. This creates a scenario where the degradation effect of truncated AGPAT4 is attenuated due to competitive inhibition between multiple proteins involved in the breakdown of truncated AGPAT4. Another possibility to explain the aforementioned observation is that too much reference AGPAT4 may cause unregulated binding to occur at both the N and C terminal ends of truncated AGPAT4, leading to the formation of an unrecognizable or incompatible structure to form, ultimately preventing proper breakdown. However, the most likely explanation is that it may reflect unequal loading in earlier experiments, possibly resulting

from lower cell numbers in plates transfected with both the reference and truncated AGPAT4 isoforms.

With regards to levels of the reference AGPAT4-6HIS form, although a similar general trend appeared, where adding small amounts of the truncated AGPAT4 variant to 5 μg of the reference protein appears to increase levels of the protein somewhat, the effect was smaller, and did not reach significance. It is plausible that adding additional immunoblots may produce some small significance, but after inclusion of an extra biological replicate to make an n=4, still no significance was detected. The speculative reason as to why such a large difference is not seen is simply that the reference AGPAT4 protein is possibly more stable than the truncated variant, and less affected by the truncated AGPAT4 protein. Since reference AGPAT4 has been characterized as a fully functional protein in multiple tissues on its own, it is likely a non-obligate protein and its detectable levels are not dependent on it forming a complex with other proteins [68].

In brief, this study concludes that *in vitro* overexpression of reference AGPAT4 appears to significantly modulate the detectable levels of truncated AGPAT4 in immunoblot experiments. The inverse does not display the same results, as *in vitro* overexpression of truncated AGPAT4 does not appear to have a significant effect on immunodetectable levels of reference AGPAT4. The differences seen are attributed to the possible innate characteristic of reference AGPAT4 as a non-obligate protein structure, while truncated AGPAT4 likely relies on binding or interaction with external proteins to prevent degradation *in vitro.*

Chapter 7

General Discussion, Limitations, and Future Directions of Study

7.1 General Discussion

Over 95% of documented genes are predicted to have unique splice variants [40, 41]. Of these, the vast majority of their functions are unknown [40]. In many cases, only the sequences are catalogued in gene databases, without much or any context at all. AGPAT4 was found to have three predicted splice variants. In prior work by the Duncan Lab, the coding region of the X2 and X3 variants was amplified from mouse brain-derived cDNA, indicating the presence of part of those transcripts endogenously. However, little else was known about the variants, or the truncated protein coded for by the mouse brain-derived transcript.

Thus, in initial work to characterize tissue expression of the reference *Agpat4* mRNA relative to the predicted variants, I designed primers for all of the predicted AGPAT4 splice variants, and the reference variant. Because of significant sequence overlap, instead of designing primers around the coding region, I selected forward primers that fell within the unique 5' untranslated regions of the four variants of AGPAT4, meaning the resultant amplicons generated by PCR would be unique to each predicted transcript variant. Results did end up showing that the X1, X2, and X3 transcripts, as well as the reference transcript, were expressed in multiple tissues in wild-type C57BL6/J mice, although at relatively low levels compared to the reference *Agpat4* transcript.

To establish preliminary data regarding the possible physiological function of truncated AGPAT4, this thesis probed for an interaction between truncated AGPAT4 and the reference AGPAT4 variants. There are many instances of specialized protein complexes, where the

truncated protein variant acts as a regulator of the original protein through interaction [45]. If an interaction were found to exist, that would have provided avenues for further experimentation. Regrettably, the results of this thesis did not find any evidence of protein-protein interaction between truncated AGPAT4 and reference AGPAT4 after several attempts at co-immunoprecipitation and another protein complex isolation technique, IMAC.

The failure to find a specific interaction between the truncated AGPAT4 and reference AGPAT4 variants did, however, suggest an alternate possible functional interaction between these homologues. A specific pattern of decreased protein levels appeared to occur when the truncated and reference AGPAT4 forms were co-expressed at equal levels in cells. Notably, however, loading was performed using a measured aliquot of the lysates, rather than an adjusted amount based on relative protein concentration in each sample. When examining protein bands created by blotting cell lysates, either the reference AGPAT4 variant or the truncated AGPAT4 variant appeared to be more highly expressed when co-transfected into HEK-293 cells with a control plasmid vector, than when co-transfected with one another. Cases of truncated protein variants acting as a mechanism of degradation to their reference counterparts have been reported in the literature, and so a potential role for truncated AGPAT4 in regulating levels of reference AGPAT4 was initially hypothesized and considered [45].

In order to test the hypothesis that truncated AGPAT4 was linked to the regulation of reference AGPAT4 levels, and *vice versa*, the immunoblotting experiments were replicated with additional groups. Notably, in these experiments, protein concentrations in lysates were determined, and protein loading was controlled according to total protein concentration assessments, which was lacking in the study reported in Chapter 5. By transfecting increasing amounts of plasmid expressing either the reference or truncated protein to the other in cell

transfections, we expected to see a gradual dose-dependent decrease in the resultant protein bands of the respective immunoblots. However, results showed that expression of reference AGPAT4 tended to increase levels of the co-expressed truncated AGPAT4, although a maximal effect was seen for the lowest ratio chosen, while the truncated variant had less of an effect on levels of the reference protein, but still tended to be positive in nature. It would be beneficial to repeat this experiment with an increased sample size to determine if a significant increase in either truncated AGPAT4 or reference AGPAT4 can be consistently observed between groups.

The findings of this thesis are of value, especially with regards to understanding tissues such as the brain where AGPAT4 is particularly abundant and serves an important physiological role in [3, 6, 28]. Comparisons between Chapter 5 and 6 also suggest that co-expression of the truncated and reference AGPAT4 proteins may affect cell numbers, which suggests effects on proliferation or apoptosis, or both. Based on this work and knowledge from the literature, future studies extending from this work should examine whether changes in stability of proteins are mechanistically involved in the increased levels of truncated AGPAT4 that were detected in cells co-expressing reference AGPAT4. Future studies should also examine the effects of these protein isoforms alone and together on cell health and cellular processes.

7.2 Limitations

In this thesis on the characterization of truncated AGPAT4, there were some limitations. Two of the studies utilized expression vectors to overexpress reference AGPAT4 and truncated AGPAT4 *in vitro.* The major issue which presents itself with these methods is that overexpression is not representative of the true endogenous expression of the two proteins. While the results produced with plasmids are of interest, the use of expression vectors in this way makes extrapolating *in vitro* findings to *in vivo* processes difficult. On the other hand, expression

vectors do provide the benefit of controlling the amount of plasmid, and therefore protein being expressed, to some degree. Utilizing, for example, an *in vivo* cell line that endogenously expresses reference and/or truncated AGPAT4 takes away the ability to control and modulate the expression levels of the two proteins. The studies performed in this thesis, especially those in Chapter 6, would not have been possible using a cell line that expresses both reference and truncated AGPAT4 innately. Another limitation of note comes from Chapter 5, where the co-immunoprecipitation and Immobilized Metal Affinity Chromatography methodology did not account for a possible weak and/or transient interaction between reference and truncated AGPAT4. Instead, the protocol was performed with the assumption that reference AGPAT4 strongly interacted with truncated AGPAT4. There is potential for the cell lysis buffer I used to disrupt weak bonding between individual amino acids of the two proteins. Making use of a protein crosslinking method such as formaldehyde linkage or using another chemical cross-linker may have been warranted, but ultimately was not done in this study. Additionally, in Chapter 5, immunoblot experiments had their proteins loaded by volume rather than a standard protein concentration that would be generated from a procedure such as a Bradford assay or a BCA assay. There were observed differences between Chapter 5 and 6, where the co-expression lanes containing 5 μg of each plasmid expressing reference and truncated AGPAT4 exhibited different protein band densities. The differences in results may be attributed to decreasing cell number post-transfection, which would not have been accounted for in Chapter 5 where there was no standardized concentration in loading of samples. Finally, this thesis largely reports methodology and findings from *in vitro* models. To transition to *in vivo* work, it would be necessary to develop unique murine models that have specific transcript variants of *Agpat4* knocked out to fully characterize each individual variant. However, before this can be done, there is much important

consideration to work through. Factors such as exon overlap between variants, promotor regions overlapping between introns and exons, and even how other genes in the same chromosomal region may be affected, must be considered before knockouts can be made and interpreted. It may be speculated that these unique knockouts of transcript variants may be generated in murine models, but likely with a high degree of difficulty to ensure there are no unwanted effects that come with these genetic modifications. Rather, a better genetic approach could involve the use of siRNA-mediated knockdown of gene expression, using siRNA targeting regions of mRNA specific to each transcript (e.g., the 5'UTR).

7.3 Future Directions

Within the previous decade, the first five members of the AGPAT family of enzymes has been characterized as critical lipid enzymes involved in the production of cellular phosphatidic acid. More recently, AGPAT4 was found to be an important regulator of brain phospholipid content [6], as well as a regulator of skeletal muscle phosphatidylcholine and phosphatidylethanolamine content [70]. With knowledge from the literature, it is known that truncated splice variants of enzymes may regulate their reference counterparts [45]. This thesis study reveals truncated AGPAT4 likely plays a role in the modulation of reference AGPAT4, and vice-versa, but more downstream work is still required to draw accurate conclusions as to what truncated AGPAT4 may impact physiologically. It is imperative for future research to quantify enzymatic activity of the reference AGPAT4 variant while co-expressed with the truncated enzyme variant. It could also be beneficial to include *in vitro* studies which manipulate pH of the cellular environment to determine any changes from the norm when truncated AGPAT4 is co-expressed with AGPAT4, considering the optimal pH of AGPAT4 being 7.4 [6]. Following enzymatic activity studies, quantification of enzyme products of AGPAT4 should be

evaluated with lipidomic work to discern whether the presence of truncated AGPAT4 affects phosphatidic acid production, as well as other downstream glycerolipids and glycerophospholipids. Of particular interest would be phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol since it has been discerned that AGPAT4 plays a significant upstream role in regulating these phospholipid species in murine brain [6]. As results from this thesis study show, the X1, X2, and X2 variants of *Agpat4* expression across various murine tissues is low relative to the reference *Agpat4* transcript. Future *in vivo* research would be wise to include factors such as feeding alterations in mice, aging studies, male and female studies, and possibly disease states to discern if these changes affect tissue expression of the X1, X2, and X3 *Agpat4* variants. RT-PCR should also be performed to determine if the X1, X2, and X3 variants, including their 5' and 3' coding regions, can be amplified from a natural tissue, like brain, and these verified sequences should be deposited in GenBank. Finally, inclusion of qualitative data from confocal microscopy should be investigated as well. It is known that reference AGPAT4 localizes to the outer mitochondrial membrane [6], *trans*-golgi apparatus [21], and endoplasmic reticulum [27]. However, it is not known if the truncated AGPAT4 variant follows these same characteristics. It is plausible that, due to the alteration of the N-terminus of truncated AGPAT4, a different subcellular localization signal is procured.

References

1. Yamashita, A., et al., *Glycerophosphate/Acylglycerophosphate acyltransferases.* Biology (Basel), 2014. **3**(4): p. 801-30.
2. Bradley, R.M., et al., *Acute Fasting Induces Expression of Acylglycerophosphate Acyltransferase (AGPAT) Enzymes in Murine Liver, Heart, and Brain.* Lipids, 2017. **52**(5): p. 457-461.
3. Bradley, R.M. and R.E. Duncan, *The lysophosphatidic acid acyltransferases (acylglycerophosphate acyltransferases) family: one reaction, five enzymes, many roles.* Curr Opin Lipidol, 2018. **29**(2): p. 110-115.
4. Bradley, R.M., et al., *Lpaatdelta/Agpat4 deficiency impairs maximal force contractility in soleus and alters fibre type in extensor digitorum longus muscle.* Biochim Biophys Acta Mol Cell Biol Lipids, 2018. **1863**(7): p. 700-711.
5. Bradley, R.M., et al., *Mice Deficient in lysophosphatidic acid acyltransferase delta (Lpaatdelta)/acylglycerophosphate acyltransferase 4 (Agpat4) Have Impaired Learning and Memory.* Mol Cell Biol, 2017. **37**(22).
6. Bradley, R.M., et al., *Acylglycerophosphate acyltransferase 4 (AGPAT4) is a mitochondrial lysophosphatidic acid acyltransferase that regulates brain phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol levels.* Biochim Biophys Acta, 2015. **1851**(12): p. 1566-76.
7. Ule, J. and B.J. Blencowe, *Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution.* Mol Cell, 2019. **76**(2): p. 329-345.
8. Takeuchi, K. and K. Reue, *Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis.* Am J Physiol Endocrinol Metab, 2009. **296**(6): p. E1195-209.
9. Lee, J. and N.D. Ridgway, *Substrate channeling in the glycerol-3-phosphate pathway regulates the synthesis, storage and secretion of glycerolipids.* Biochim Biophys Acta Mol Cell Biol Lipids, 2020. **1865**(1).
10. Kitson, A.P., K.D. Stark, and R.E. Duncan, *Enzymes in brain phospholipid docosahexaenoic acid accretion: a PL-ethora of potential PL-ayers.* Prostaglandins Leukot Essent Fatty Acids, 2012. **87**(1): p. 1-10.
11. Carman, G.M. and G.S. Han, *Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis.* J Biol Chem, 2009. **284**(5): p. 2593-7.
12. Bell, R.M., Coleman, R. A., *Enzymes of Glycerolipid Synthesis in Eukaryotes.* Annual Review of Biochemistry, 1980. **49**: p. 459-487.
13. Kennedy, E.P., Weiss, S. B., *The function of cytidine coenzymes in the biosynthesis of phospholipides.* The Journal of Biological Chemistry, 1956. **222**(1): p. 193-214.
14. Lands, W.E.M., *Metabolism of Glycerolipides: A Comparison of Lecithin and Triglyceride Synthesis.* The Journal of Biological Chemistry, 1958. **231**(2): p. 883-888.
15. Shindou, H., et al., *Recent progress on acyl CoA: lysophospholipid acyltransferase research.* J Lipid Res, 2009. **50 Suppl**: p. S46-51.
16. Shindou, H. and T. Shimizu, *Acyl-CoA:lysophospholipid acyltransferases.* J Biol Chem, 2009. **284**(1): p. 1-5.
17. Yamashita, A., et al., *Acyltransferases and transacylases that determine the fatty acid composition of glycerolipids and the metabolism of bioactive lipid mediators in mammalian cells and model organisms.* Prog Lipid Res, 2014. **53**: p. 18-81.
18. Shindou, H., et al., *Identification of membrane O-acyltransferase family motifs.* Biochem Biophys Res Commun, 2009. **383**(3): p. 320-5.

19. Coleman, R.A., Reed, B. C., Mackall, J. C., Student, A. K., Lane, M. D., Bell, R. M., *Selective Changes in Microsomal Enzymes of Triacylglycerol Phosphatydylchoine, and PHosphatidylethanolamine Biosynthesis during Differentiation of 3T3-L1 Preadipocytes.* The Journal of Biological Chemistry, 1978. **253**(20): p. 7256-7621.
20. Coleman, R.A., Lee, D. P., *Enzymes of triacylglycerol synthesis and their regulation.* Progress in Lipid Research, 2004. **43**(2): p. 134-176.
21. Zhukovsky, M.A., et al., *The Structure and Function of Acylglycerophosphate Acyltransferase 4/ Lysophosphatidic Acid Acyltransferase Delta (AGPAT4/LPAATdelta).* Front Cell Dev Biol, 2019. **7**: p. 147.
22. Lu, B., Jiang, Y. J., Zhou, Y., Xu, F. Y., Hatch, G. M., Choy, P. C., *Cloning and characterization of murine 1-acyl-sn-glycerol 3-phosphate acyltransferases and their regulation by PPARα in murine heart.* Biochemical Journal, 2005. **385**: p. 469-477.
23. Agarwal, A.K., et al., *Human 1-acylglycerol-3-phosphate O-acyltransferase isoforms 1 and 2: biochemical characterization and inability to rescue hepatic steatosis in Agpat2(-/-) gene lipodystrophic mice.* J Biol Chem, 2011. **286**(43): p. 37676-91.
24. Prasad, S.S., A. Garg, and A.K. Agarwal, *Enzymatic activities of the human AGPAT isoform 3 and isoform 5: localization of AGPAT5 to mitochondria.* J Lipid Res, 2011. **52**(3): p. 451-62.
25. Hollenback, D., et al., *Substrate specificity of lysophosphatidic acid acyltransferase beta -- evidence from membrane and whole cell assays.* J Lipid Res, 2006. **47**(3): p. 593-604.
26. Yuki, K., et al., *Characterization of mouse lysophosphatidic acid acyltransferase 3: an enzyme with dual functions in the testis.* J Lipid Res, 2009. **50**(5): p. 860-9.
27. Eto, M., H. Shindou, and T. Shimizu, *A novel lysophosphatidic acid acyltransferase enzyme (LPAAT4) with a possible role for incorporating docosahexaenoic acid into brain glycerophospholipids.* Biochem Biophys Res Commun, 2014. **443**(2): p. 718-24.
28. Mardian, E.B., et al., *Agpat4/Lpaatdelta deficiency highlights the molecular heterogeneity of epididymal and perirenal white adipose depots.* J Lipid Res, 2017. **58**(10): p. 2037-2050.
29. K. Ravi Acharya, M.D.L., *The advantages and limitations of protein crystal structures.* Trends in Pharmacological Sciences, 2005. **Volume 26**(Issue 1): p. 10-14.
30. Smyth, M.S., Martin, J. H. J., *XRay Crystallography.* Journal of Clinical Pathology: Molecular Pathology, 2000. **53**: p. 8-14.
31. Harkey, T., et al., *The Role of a Crystallographically Unresolved Cytoplasmic Loop in Stabilizing the Bacterial Membrane Insertase YidC2.* Sci Rep, 2019. **9**(1): p. 14451.
32. Turnbull, A.P., Rafferty, J.B., Sedelnikova, S.E., Slabas, A.R., Schierer, T.P., Kroon, T.M., Simon, J.W., Fawcett,. T., Nishida, I., Murata, N., Rice,. D.W., *Analysis of the Structure, Substrate Specificity, and Mechanim of Squash Glycerol-3-Phosphate (1)-Acyltransferase.* Structure, 2001. **9**: p. 347-353.
33. Robertson, R.M., et al., *A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer.* Nat Struct Mol Biol, 2017. **24**(8): p. 666-671.
34. Jones, S., Thornton, J.M., *Principles of protein-protein interactions.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**: p. 13-20.
35. Rumfeldt, J.A., et al., *Conformational stability and folding mechanisms of dimeric proteins.* Prog Biophys Mol Biol, 2008. **98**(1): p. 61-84.
36. Perkins, J.R., et al., *Transient protein-protein interactions: structural, functional, and network properties.* Structure, 2010. **18**(10): p. 1233-43.
37. Cooper, G.M., *The Cell: A Molecular Approach, 2nd edition*, in *Regulation of Protein Function*. 2000, Sinauer Associates: Sunderland (MA).
38. Lessard, J.C., *Molecular cloning.* Methods Enzymol, 2013. **529**: p. 85-98.

39. Tirabassi, R. *Foundations of Molecular Cloning - Past, Present and Future*. N/A.
40. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. **40**(12): p. 1413-5.
41. Matlin, A.J., F. Clark, and C.W. Smith, *Understanding alternative splicing: towards a cellular code.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 386-98.
42. Souvorov, A., Kaspustin, Y., Kiryutin, B., Chetvernin, V., Tatusova, T., Lipman, D., *Gnomon - NCBI eukaryotic gene prediction tool*. 2010, National Center for Biotechnology Information: Bethesda, MD.
43. Cui, X., et al., *Homology search for genes.* Bioinformatics, 2007. **23**(13): p. i97-103.
44. Wang, Z., Y. Chen, and Y. Li, *A Brief Review of Computational Gene Prediction Methods.* Genomics, Proteomics & Bioinformatics, 2004. **2**(4): p. 216-221.
45. Menard, V., et al., *Modulation of the UGT2B7 enzyme activity by C-terminally truncated proteins derived from alternative splicing.* Drug Metab Dispos, 2013. **41**(12): p. 2197-205.
46. Invitrogen. *TRIzol Reagent*. Available from: http://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf.
47. Bustin, S. and J. Huggett, *qPCR primer design revisited.* Biomol Detect Quantif, 2017. **14**: p. 19-28.
48. Samal, S.K., *Leader Sequence*, in *Brenner's Encyclopedia of Genetics*. 2013. p. 203-205.
49. van der Velden, A.W., Thomas, A. A. M., *The role of the 5' untranslated region of an mRNA in translation regulation during development.* The International Journal of Biochemistry & Cell Biology, 1999. **31**: p. 87-106.
50. Lambert, S.A., et al., *The Human Transcription Factors.* Cell, 2018. **172**(4): p. 650-665.
51. Horton, P., et al., *WoLF PSORT: protein localization predictor.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W585-7.
52. Alves-Bezerra, M. and D.E. Cohen, *Triglyceride Metabolism in the Liver.* Compr Physiol, 2017. **8**(1): p. 1-8.
53. Komiya, Y., et al., *Mouse soleus (slow) muscle shows greater intramyocellular lipid droplet accumulation than EDL (fast) muscle: fiber type-specific analysis.* J Muscle Res Cell Motil, 2017. **38**(2): p. 163-173.
54. Lehnig, A.C. and K.I. Stanford, *Exercise-induced adaptations to white and brown adipose tissue.* J Exp Biol, 2018. **221**(Pt Suppl 1).
55. Li, K. and A. Brownley, *Primer Design for RT-PCR*, in *RT-PCR Protocols: Second Edition*, N. King, Editor. 2010, Humana Press: Totowa, NJ. p. 271-299.
56. de Magalhaes, J.P., J. Curado, and G.M. Church, *Meta-analysis of age-related gene expression profiles identifies common signatures of aging.* Bioinformatics, 2009. **25**(7): p. 875-81.
57. Keleher, M.R., et al., *Maternal high-fat diet associated with altered gene expression, DNA methylation, and obesity risk in mouse offspring.* PLoS One, 2018. **13**(2): p. e0192606.
58. Penner-Goeke, S. and E.B. Binder, *Epigenetics and depression.* Dialogues Clin Neurosci, 2019. **21**(4): p. 397-405.
59. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification.* Curr Protoc Protein Sci, 2013. **73**: p. 9 9 1-9 9 23.
60. Promega. *pGEM-T and pGEM-T Easy Vector Systems*. Available from: https://www.promega.com/-/media/files/resources/protocols/technical-manuals/0/pgem-t-and-pgem-t-easy-vector-systems-protocol.pdf.
61. Clark, D.P., Pazdernik, N.J., *Molecular Biology, 2nd edition*. 2013, Academic Cell.
62. Scientific, T. *Useful Numbers for Cell Culture*. Available from: https://www.thermofisher.com/ca/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html.

63. Abcam. *Immunoprecipitation Protocol: General immunoprecipitation procedure and required reagents*. Available from: http://docs.abcam.com/pdf/protocols/Immunoprecipitation_protocol.pdf?mi_u=1157442.
64. MilliporeSigma. *Luminata Western Chemiluminescent HRP Substrates*. Available from: https://www.emdmillipore.com/CA/en/product/Immobilon-Crescendo-Western-HRP-substrate-100-mL,MM_NF-WBLUR0100#anchor_DS.
65. Block, H., et al., *Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC)*, in *Guide to Protein Purification, 2nd Edition*. 2009. p. 439-473.
66. Lin, J.S. and E.M. Lai, *Protein-Protein Interactions: Co-Immunoprecipitation.* Methods Mol Biol, 2017. **1615**: p. 211-219.
67. Tang, X. and J.E. Bruce, *Chemical Cross-Linking for Protein–Protein Interaction Studies*, in *Mass Spectrometry of Proteins and Peptides: Methods and Protocols*, M.S. Lipton and L. Paša-Tolic, Editors. 2009, Humana Press: Totowa, NJ. p. 283-293.
68. Acuner Ozbabacan, S.E., et al., *Transient protein-protein interactions.* Protein Eng Des Sel, 2011. **24**(9): p. 635-48.
69. Pettersson, M. and C.M. Crews, *PROteolysis TArgeting Chimeras (PROTACs) - Past, present and future.* Drug Discov Today Technol, 2019. **31**: p. 15-27.
70. Lee, S., et al., *Skeletal muscle phosphatidylcholine and phosphatidylethanolamine respond to exercise and influence insulin sensitivity in men.* Sci Rep, 2018. **8**(1): p. 6531.

Appendix

pCMV-*Trunc. Agpat4-HA Tag*- Forward Sequencing

tmccgtggtggcggccgctctagcccgggcggatctaccatgccagactctgctactgcg
tctccagccaactccaaagtcctggccaagaaagaactggcttatgtcccaatcattggc
tggatgtggtacttcgtggaaatgatcttttgcacacgcaagtgggagcaagatcggcag
acggttgccaagagcctgctgcacctccgggactacccagagaagtatctgttcctgatc
cactgtgagggcacacggttcacagagaagaaacaccaaatcagcatgcaggtggcccaa
gccaaggggctgcccagcctcaaacaccacctgctgccgcgcaccaaaggctttgctatt
actgtgaagtgcttgcgagatgttgtcccagctgtatatgactgtacactcaatttcaga
aacaatgaaaacccaacactgctgggagtcttaaatggaaagaaatatcacgctgactgc
tacgttcggaggatccccatggaggacattccggaggatgaggacaagtgctctgcctgg
ttacacaagctctaccaggagaaggatgcctttcaggaggaatactacaggacaggggtc
ttcccagagactccctgggttcccccacggcggccctggtctctggtcaactggttgttc
tgggcatcgctgctgctctacccttttcttccagttcctagttagcatggtcagcagcggt
tcctcggtgacgctggccagcttggtcctcatcttctgtatggcctccatgggagttcga
tggatgattggcgtgacagaaatcgacaagggctctgcctacggcaacatcgacaacaaa
cggaaacaaacggactatccctatgacgtcccggactatgcatagctcgaggattacaag
gatgacgacgataaggactataaggacgatgatgacaaggactacaagatgatgacgata
atagggcccggtaccttaattaattaagtaccaggtaagtgtacccaattcgccctatag
tgagtcgtatacattcactcgatcggctcgctgatcagcctcgactgtgcttctagtgca
gccatctgtgttgccctcccgtgctcttgacctgcagtgccactcccactgtcttctaat
aaatgagaaatgcatcgcattgtcttgataggktcattctattctctcgcg

pCMV-*Trunc. Agpat4-HA Tag*- Reverse Sequencing

gckgctactggtacttattattaaggtaccgggccctatttatcgtcatcatctttgtag
tccttgtcatcatcgtccttatagtccttatcgtcgtcatccttgtaatcctcgagctat
gcatagtccgggacgtcatagggatagtccgtttgtttccgtttgttgtcgatgttgccg
taggcagagcccttgtcgatttctgtcacgccaatcatccatcgaactcccatggaggcc
atacagaagatgaggaccaagctggccagcgtcaccgaggaaccgctgctgaccatgcta
actaggaactggaagaaagggtagagcagcagcgatgcccagaacaaccagttgaccaga
gaccagggccgccgtgggggaacccagggagtctctgggaagacccctgtcctgtagtat
tcctcctgaaaggcatccttctcctggtagagcttgtgtaaccaggcagagcacttgtcc
tcatcctccggaatgtcctccatggggatcctccgaacgtagcagtcagcgtgatatttc
tttccatttaagactcccagcagtgttgggttttcattgtttctgaaattgagtgtacag
tcatatacagctgggacaacatctcgcaagcacttcacagtaatagcaaagcctttggtg
cgcggcagcaggtggtgtttgaggctgggcagccccttggcttgggccacctgcatgctg
atttggtgtttcttctctgtgaaccgtgtgccctcacagtggatcaggaacagatacttc
tctgggtagtcccggaggtgcagcagctcttggcaaccgtctgccgatcttgctcccact
tgcgtgtgcaaaagatcatttccacgaagtaccacatccagccaatgattgggacataag
ccagttctttcttggccaggactttggagttggctggagacgcagtagcagagtctggca
tggtagatccgcccgggctagagcggccgccaccgcgtgagctccagcttttgttcccttt
tagtgaggttaattcgagcttggcgtaatcgctagcggatctgacggttcactaaccagc
tctgctatatagacctcccaccgtacmcgcctacgccatttgcgtcaatgggggcggatt
gttacgacatttggaagtccgtgactgggtcaacaatccatgacgtcaatgggggkggag
aacttgggaaaatcacccgcccacgt

Figure 17: TCAG sequencing results for newly synthesized plasmid pCMV-*Trunc. Agpat4-HA* tag from Chapter 5. Start codons are indicated by highlighted red text. The newly inserted hemagglutinin (HA) tag is indicated by purple text. The newly inserted stop codons are indicated by red text. The newly inserted Xho I restriction site is indicated by green text.

```
TCAG        atgccagactctgctactgcgtctccagccaactccaaagtcctggccaagaaagaactg      60
Variant     atgccagactctgctactgcgtctccagccaactccaaagtcctggccaagaaagaactg      60
            ************************************************************

TCAG        gcttatgtcccaatcattggctggatgtggtacttcgtggaaatgatcttttgcacacgc      120
Variant     gcttatgtcccaatcattggctggatgtggtacttcgtggaaatgatcttttgcacacgc      120
            ************************************************************

TCAG        aagtgggagcaagatcggcagacggttgccaagagcctgctgcacctccgggactaccca      180
Variant     aagtgggagcaagatcggcagacggttgccaagagcctgctgcacctccgggactaccca      180
            ************************************************************

TCAG        gagaagtatctgttcctgatccactgtgagggcacacggttcacagagaagaaacaccaa      240
Variant     gagaagtatctgttcctgatccactgtgagggcacacggttcacagagaagaaacaccaa      240
            ************************************************************

TCAG        atcagcatgcaggtggcccaagccaaggggctgcccagcctcaaacaccacctgctgccg      300
Variant     atcagcatgcaggtggcccaagccaaggggctgcccagcctcaaacaccacctgctgccg      300
            ************************************************************

TCAG        cgcaccaaaggctttgctattactgtgaagtgcttgcgagatgttgtcccagctgtatat      360
Variant     cgcaccaaaggctttgctattactgtgaagtgcttgcgagatgttgtcccagctgtatat      360
            ************************************************************

TCAG        gactgtacactcaatttcagaaacaatgaaaacccaacactgctgggagtcttaaatgga      420
Variant     gactgtacactcaatttcagaaacaatgaaaacccaacactgctgggagtcttaaatgga      420
            ************************************************************

TCAG        aagaaatatcacgctgactgctacgttcggaggatccccatggaggacattccggaggat      480
Variant     aagaaatatcacgctgactgctacgttcggaggatccccatggaggacattccggaggat      480
            ************************************************************

TCAG        gaggacaagtgctctgcctggttacacaagctctaccaggagaaggatgcctttcaggag      540
Variant     gaggacaagtgctctgcctggttacacaagctctaccaggagaaggatgcctttcaggag      540
            ************************************************************

TCAG        gaatactacaggacaggggtcttcccagagactccctgggttcccccacggcggccctgg      600
Variant     gaatactacaggacaggggtcttcccagagactccctgggttcccccacggcggccctgg      600
            ************************************************************

TCAG        tctctggtcaactggttgttctgggcatcgctgctgctctacccttttcttccagttccta      660
Variant     tctctggtcaactggttgttctgggcatcgctgctgctctacccttttcttccagttccta      660
            ************************************************************

TCAG        gttagcatggtcagcagcggttcctcggtgacgctggccagcttggtcctcatcttctgt      720
Variant     gttagcatggtcagcagcggttcctcggtgacgctggccagcttggtcctcatcttctgt      720
            ************************************************************

TCAG        atggcctccatgggagttcgatggatgattggcgtgacagaaatcgacaagggctctgcc      780
Variant     atggcctccatgggagttcgatggatgattggcgtgacagaaatcgacaagggctctgcc      780
            ************************************************************

TCAG        tacggcaacatcgacaacaaacggaaacaaacggac---    816
Variant     tacggcaacatcgacaacaaacggaaacaaacggactga    819
            ***********************************
```

Figure 18: TCAG sequencing results aligned with mRNA coding region of *Agpat4* Variant X2/X3. Sequencing results indicate a successfully cloned plasmid insert relative to the truncated variant coding region. It should be noted the stop codon TGA present in the variant coding reference sequence is missing in the insert sequence due to intentional removal during cloning.

Input Sequence

QUERY (458 aa)

MSRQSWHPLA PSFGRRASIH RSSGHPRSVG PSLGNAGRDL GRGFRAGGVV
ARRDTFLSPG SFPLLELSHH TDLSLENPHT MDLIGLLKSQ FLCHLVFCYV
FIASGLIVNA IQLCTLVIWP INKQLFRKIN ARLCYCVSSQ LVMLLEWWSG
TECTIYTDPK ACPHYGKENA IVVLNHKFEI DFLCGWSLAE RLGILGNSKV
LAKKELAYVP IIGWMWYFVE MIFCTRKWEQ DRQTVAKSLL HLRDYPEKYL
FLIHCEGTRF TEKKHQISMQ VAQAKGLPSL KHHLLPRTKG FAITVKCLRD
VVPAVYDCTL NFRNNENPTL LGVLNGKKYH ADCYVRRIPM EDIPEDEDKC
SAWLHKLYQE KDAFQEEYYR TGVFPETPWV PPRRPWSLVN WLFWASLLLY
PFFQFLVSMV SSGSSVTLAS LVLIFCMASM GVRWMIGVTE IDKGSAYGNI
DNKRKQTD

queryProtein WoLFPSORT prediction plas: 12, mito: 9.5, cyto_mito: 8, cyto: 5.5, nucl: 3, extr: 1, pero: 1

PSORT features and traditional PSORTII prediction

32 Nearest Neighbors

id	site	distance	identity	comments
GLP2_RAT	plas	1672.84	15.704%	[Uniprot] SWISS-PROT45:Integral membrane protein.
SCG2_XENLA	plas	2215.8	14.6305%	[Uniprot] SWISS-PROT45:Integral membrane protein.
SCAG_XENLA	plas	2266.39	13.4848%	[Uniprot] SWISS-PROT45:Integral membrane protein.
GLSL_HUMAN	mito	2471.68	12.7907%	[Uniprot] SWISS-PROT45:Mitochondrial. GO:0005739; C:mitochondrion; Evidence:TAS.
HEM0_BRARE	mito	2554.94	15.5556%	[Uniprot] SWISS-PROT45:Mitochondrial matrix.
GLP2_HUMAN	plas	2576	16.6966%	[Uniprot] SWISS-PROT45:Integral membrane protein. GO:0005886; C:plasma membrane; Evidence:TAS.
ADHL_GADMO	cyto	2681.73	14.3791%	[Uniprot] SWISS-PROT45:Cytoplasmic.
FANA_HELAS	plas	2710.21	14.4%	[Uniprot] SWISS-PROT45:Integral membrane protein.
FPGT_HUMAN	cyto	2882.68	12.4579%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO:0005737; C:cytoplasm; Evidence:TAS.
EGL3_RAT	cyto_mito	2981.54	15.2174%	[Uniprot] SWISS-PROT45:Mitochondrial and cytoplasmic.
CYA5_RAT	plas	2992.55	11.7274%	[Uniprot] SWISS-PROT45:Integral membrane protein.
ORC4_MOUSE	nucl	3157.2	11.9826%	[Uniprot] SWISS-PROT45:Nuclear.
GP17_HUMAN	plas	3200.1	18.7773%	[Uniprot] SWISS-PROT45:Integral membrane protein. GO:0005887; C:integral to plasma membrane; Evidence:TAS.
SES1_XENLA	nucl	3215.48	12.8364%	[Uniprot] SWISS-PROT45:Nuclear.
CYA5_RABIT	plas	3292.32	11.2342%	[Uniprot] SWISS-PROT45:Integral membrane protein.
SCAG_RAT	plas	3311.25	14%	[Uniprot] SWISS-PROT45:Integral membrane protein.
GCSP_HUMAN	mito	3319.02	12.8431%	[Uniprot] SWISS-PROT45:Mitochondrial.
ODPT_HUMAN	mito	3334.42	15.4684%	[Uniprot] SWISS-PROT45:Mitochondrial matrix.
NUAM_BOVIN	mito	3345.59	13.7552%	[Uniprot] SWISS-PROT45:Matrix and cytoplasmic side of the mitochondrial inner membrane.
HEM1_RAT	mito	3350.06	13.0841%	[Uniprot] SWISS-PROT45:Mitochondrial matrix.
P2X2_HUMAN	plas	3366	14.7368%	[Uniprot] SWISS-PROT45:Integral membrane protein.
NUAM_HUMAN	mito	3420.48	14.9931%	[Uniprot] SWISS-PROT45:Matrix and cytoplasmic side of the mitochondrial inner membrane.
PAFA_CHICK	extr	3455.39	14.5652%	[Uniprot] SWISS-PROT45:Extracellular.
TNI3_MOUSE	cyto	3488.29	15.4839%	[Uniprot] SWISS-PROT45:Cytoplasmic and nuclear (By similarity). GO:0005737; C:cytoplasm; Evidence:IDA.
PSS1_CRILO	plas	3519.49	11.3445%	[Uniprot] SWISS-PROT45:Integral membrane protein.
SPYA_RABIT	pero	3533.59	12.854%	[Uniprot] SWISS-PROT45:Peroxisomal.
EKI1_HUMAN	cyto	3548.35	11.0151%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO:0005737; C:cytoplasm; Evidence:NAS.
AG22_MOUSE	plas	3551.81	14.8148%	[Uniprot] SWISS-PROT45:Integral membrane protein.
ACDB_RAT	mito	3553.58	12.987%	[Uniprot] SWISS-PROT45:Mitochondrial matrix.
RAG2_RABIT	nucl	3560.52	13.9623%	[Uniprot] SWISS-PROT45:Nuclear.
HMGL_CHICK	mito	3599.53	11.5721%	[Uniprot] SWISS-PROT45:Mitochondrial matrix.
GLNE_RAT	cyto	3616.17	14.1274%	[Uniprot] SWISS-PROT45:Cytoplasmic.

Figure 19: Subcellular localization prediction for *Agpat4* splice variant X1. The PSORT program output predicts the highest likelihood of localization of predicted variant X1 to the plasma membrane of cells.