Gene synthesis and expression of human torque teno virus VP3: Exploring the cancer-killing potential of an apoptin homolog

A Major Qualifying Project

Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE

In Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science

By

_____

Eric Evan-Browning

_____

McGhee Orme-Johnson

Date: April 30, 2009

Approved:

_____

Professor Destin Heilman, Primary Advisor

Table of Contents

Abstract

Torque teno virus (TTV) is a small, non-enveloped animal virus that contains a single-stranded, circular DNA genome of 3.8 kb. Related to chicken anemia virus (CAV), TTV's ORF 3 may express a 105 residue protein which is homologous to CAV VP3 (apoptin)--a protein shown to multimerize and trigger apoptosis in cancer cells via a mechanism that involves nuclear translocation. Using gene synthesis, we generated the TTV ORF3 and expressed the gene fused to GFP. Here we demonstrate that TTV VP3 localizes to the cytoplasm of H1299 lung cancer cells and shows evidence of cytoplasmic aggregation. This aggregation is strikingly similar to CAV apoptin and may indicate a similar multimerization capacity for TTV VP3. However, an altered localization pattern may suggest a different mechanism of action.

Introduction

Viruses are discovered with some regularity, but often it requires years of research to evince the true behavior, significance, and impact of such agents. In the unique case of torque teno virus (TTV), more than a decade of research has unearthed little about its role in the human body or its viral mechanisms.

Discovered in the serum of a Japanese hepatitis patient in 1997, TTV was initially introduced as a novel hepatitis agent. [19] TTV's heightened prevalence in hepatitis patients has driven considerable investigation as to the virus's role in hepatitis infection as well as other human diseases. However, more than a decade of research has presented no evidence connecting TTV with any known pathology. Through epidemiological studies, it has been shown that TTV is present in over 90% of the adult population. [1] In addition, alternate TTV genotypes have been found in numerous other mammalian species, including chimpanzees, chickens, pigs, cows, and dogs. Despite its near-ubiquity among humans, no evidence has been presented to correlate TTV infection with any kind of human disease; in addition, much about the biological nature of TTV is still unknown. [5,8]

TTV is a small, nonenveloped virus with a single-stranded, antisense DNA genome of length 3.8 kb. Three polycistronic transcripts are produced from the genome, which encode only three putative proteins. Despite this simplicity, TTV is a virus of immense genomic diversity, with over 5 taxonomic groups and 40 genotypes spanning numerous species. TTV is currently classified into an unassigned genus Anellovirus, and although currently not part of taxonomic family, TTV was initially designated to the Circoviridae family, due to a number of distinctive structural similarities. The proteins expressed by TTV have been poorly characterized

[8], with little known about their expression or function. This difficulty is due to a number of significant obstacles regarding its study: with all its genotypes combined, TTV is nearly ubiquitous in all infected species; its lack of known pathogenicity complicates the understanding of its biological function; finally, no culture system has been developed for growing TTV in the laboratory. All of these factors combined make investigations into the nature of TTV exceedingly difficult. [8] One member of Circoviridae, chicken anemia virus (CAV), shows an ORF orientation which closely resembles that of TTV. Significantly, CAV VP3 codes for the much-studied protein apoptin, which is capable of selectively inducing apoptosis in transformed or tumorogenic cells. Due to their similarities in both structure and ORF orientation, TTV may express a protein akin to CAV apoptin, which may have similar cancer-killing activity. [8]

Since 1999, a number of similarities have been drawn between TTV and CAV. Despite some structural differences, these viruses share a number of genomic similarities which have significantly advanced our understanding of TTV, allowing for parallels to be drawn between the proteins expressed by CAV and those of TTV. Central to this project, a team from the Netherlands in 2004 claimed TTV VP3 could induce apoptosis in cancer cells, much like CAV apoptin. [14]

First, we will discuss the current status of TTV, its characteristics, and its history and then we will then detail our study of TTV VP3. Our goal was to synthesize TTV ORF3 and express its associated protein to study its localization and interactions in cancerous cells, and by this, lay the foundation for future investigations into TTV VP3 behavior and potential cancer-killing ability.

Potential pathogenicity of TTV

In the year of TTV's discovery, a percentage of liver disease had no association with any viral agent, yet researchers speculated that other agents must exist. [26] Sparked by the discovery of the hepatitis C virus in 1989, scientists began investigating the presence of other types of viruses in hepatitis cases. They found that a substantial fraction of hepatitis C patients did not display any known virus indicators. In 1997, this search yielded an unknown virus from a group of Japanese post-transfusion hepatitis patients. [19] The researchers isolated a DNA clone of the virus from a serum sample which they initially dubbed the virus "TT", after the initials of the patient in which the virus was first found. This first patient had no indicators of known hepatitis viruses, and so TTV was initially thought to be a new agent for hepatitis, providing the initial drive for TTV investigations.

In 1999, the virus had been renamed torque teno virus, from the Latin meaning "thin necklace" (in reference to the single-stranded circle DNA of the virus). [26] At this point in time, the virus had been identified in two distinct genetic groups. This preliminary study identified the virus in nearly half of all unknown etiology hepatitis patients, as well as a similar number of patients with chronic liver disease with unknown etiology. This result led researchers to speculate to the potential role of TTV in various liver-related diseases. [8]

Until 1998, the disease had only been found in Japanese patients. In that year, serum samples from 126 adult patients in the UK were tested by Naoumov et al., and among these patients 24 were shown to be infected by TTV. In this group, 30 were healthy individuals acting as a control and 72 had chronic liver diseases, 33 of whom had been diagnosed with hepatitis C.

[14] Overall, 25% of all patients with chronic liver disease showed TTV infection, and 10% of the control group had TTV infection. This showed not only that TTV was present in the UK as well as in Japan, but also that the presence of TTV was both heightened among chronic liver patients, but that TTV was undoubtedly prevalent in cases other than those exhibiting liver problems. There was some connection, however, linking the 24 patients with TTV in this study. Of these 24 patients, 15 had a history of either drug use or intravenous transfusions. [13]

Further, two studies provided sound evidence for the widespread presence of TTV among all populations. The first was a Japanese study in 1998, which found evidence for high prevalence of TTV in not just individuals with chronic liver disease, but within the entire Japanese population. [16] The second study, conducted in 1999 by Neil et al. concluded that TTV was present throughout the Brazilian population. Their research focused upon the sera of Brazilian blood donors, most of whom were negative for hepatitis and other chronic hepatic disease. [16]

In 1999, Leary et al. demonstrated that TTV is also present in a number of animal species. [15] In this study, several species of animal were assayed via PCR in order to find the presence of TTV. They conducted this study because a large body of evidence at this point indicated that TTV was widespread, and that perhaps animals too exhibited this virus. In addition, the researchers were interested in determining whether TTV had any specific geographic correlation. In order to test the geographic variable, this study assayed the sera from pigs, sheep, chickens, and cows from a number of geographic locations. TTV was found to be present in all of the animals tested, but not in the same percent populations exhibited by humans. Also, the study examined the sera of six types of primates (the study indicated that the

samples were highly limited), with TTV found in three of the six species examined. [13] Later studies with animals showed little effect of exposure to TTV and that animals with TTV exhibited a negligible amount of liver problems. Questions arose as to whether TTVs in animals were, in fact, specific to non-human species, or whether these animals harbored only human TTV. In 2001, Okamoto et al., found that TTV in lesser primates was distinct from that of TTV in humans. [14] In a follow-up study, Okamoto et al., confirmed that TTV was present in cats, dogs, and that TTV was highly prevalent among pigs. [22]

In 2002, TT virus was shown to have a heightened prevalence in cancer patients. A study of 101 Turkish patients, 61 of whom had various types of hepatic cancer and 45 healthy control subjects, were assayed for TTV infections. Of those, 23 were found to exhibit TTV, and of those 18 were from the maligned liver group. This result presented the possibility that TTV infection, while present broadly among human populations, is present in higher proportions among cancer patients. Additionally, the presence of TTV was observed to be significantly higher in patients with a history of transfusions. This suggested to researchers that transfusions are somehow linked to TTV infection. Additionally, greater risk and prevalence of TTV among cancer patients could be the result of cancer-caused immunodeficiency. [8] In 2003, it was elucidated that TTV is present in many types of tissue, and not just a limited selection of tissues. [24] In 2002, a study by Tokita et al. claimed to find a correlation between TTV load and cancer. Their research showed that a high level of TTV in hepatitis C patients posed a significant risk for those patients to develop liver cancer. [23]

At this point in time, it has been determined that TTV presence within any particular patient can be greatly heightened due to the immunological state of the patient. In a 2001

study, Zhong et al. showed that using a high-TTV concentration cutoff, over 90% of cancer patients showed the presence of TTV, compared to 4% in the control group. [30] This study showed nearly a 100-fold increase in TTV concentration in cancer patients. This study by Zhong et al. suggested that this high level of TTV concentration in cancer patients may be directly related to the immunological state of a patient, and that TTV may not have any other significant linkage to cancer. [30]

Advancements of TTV screening techniques became a crucial part in discovering more about the virus. It is believed that of the major reasons behind the unknown pathogenic effects of TTV is the highly varied nature of methodologies used to study the virus. Several efforts have been made to standardize the study of TTV. Numerous types of PCR have been utilized to screen for TTV, but only a few have been used repeatedly in TTV studies. Most research has shown that the most common genotype of TTV is genotype 1, which is often screened by UTR PCR—which detects the presence of the relatively conserved untranslated region (UTR) of TTV. [23]

Characteristics of TTV

Study of TTV as a pathological agent had revealed that TTV is nearly ubiquitous worldwide among humans, as well as in many mammalian and avian species. [1] Despite nearly eleven years of research into TTV‘s role as a pathological agent, no solidified connection has ever been made between TTV and any type of contagion. So as it stands, TTV appears to be completely innocuous in humans and animal species. TTV is transmitted both fecally and orally, and a recent study by Griffen et al. theorizes that TTV, due to its ease of detection and

prevalence, may be an improved indicator of viral pathogens in drinking water, as opposed to using bacterial indicators. TTV is a small semi-spherical viral particle, with a diameter of approximately 30 nm. TTV particles, when found in circulation within serum, are complexed with immunoglobulin G (IgG), and when found in human feces, TTV appears almost entirely as a free viral particle. Because IgG binds TTV, TTV can be isolated and visualized by electron microscopy. [8]

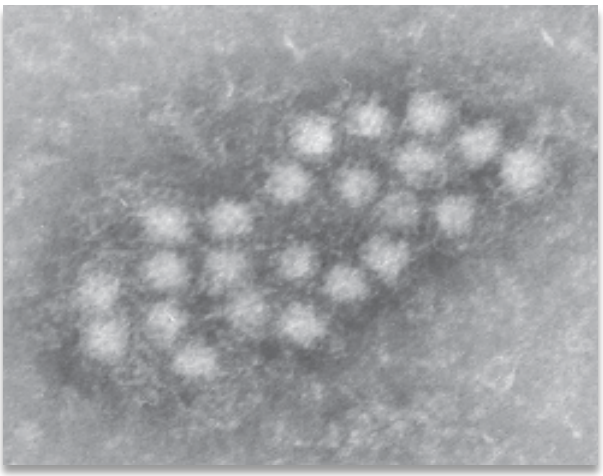

Figure 1: Electron micrograph of aggregated TTV particles (30-30 nm) found in a blood serum sample. [14]

Initially, it was elucidated that TTV was approximately 3.8 kilobases in length with negative polarity. [8] Within a few years, it was shown that TTV was a nonenveloped virus with a single-stranded, circular DNA genome. Several other families of viruses share these characteristics: the bacterial viruses Inoviridae and Microviridae, the plant viruses Geminiviridae and Nanoviridae, and the animal viruses of Circovirdae. Unlike the aforementioned taxonomic groups, TTV is the first human virus with these characteristics. [9] Owing to these similarities, TTV was arbitrarily assigned to the Circoviridae family, alongside viruses such as chicken anemia virus (CAV) and pocine circovirus (PCV). However, TTV has antisense open reading frames (ORF), while those of Circoviridae are ambisense (transcribed on

both strands), with the exception of CAV which is antigenomic. In addition, TTV's genome is significantly larger than the viruses of Circoviridae. [13] Due to these differences between TTV and the viruses belonging to the Circovirdae family, TTV was assigned to the "floating" genus, Anelloviridae (where floating is used to mean a currently unassigned or orphaned taxonomic designation). [10] Recently, it has been proposed that due to the enormous diversity of TTV, these viruses should be assigned as a new taxonomic family, Anelloviridae. Interestingly, Hino et al. have supported the redesignation of CAV from Circoviridae into the same taxonomic family as TTV (Anelloviridae, or whatever the taxonomic designation becomes), or even a family on its own. [9]

In 1999, Miyata et al. published a study on the TTV strain TA278, showing a number of similarities between TTV and CAV. Because an understanding of the similarities between TTV and CAV provides evidence for our later discussion—and because many of these characteristics were initially studied in comparison to those of CAV—we will discuss and contrast them here in some depth.

TTV exhibits broad infection diversity—TTV is present in over 20 mammalian and avian species. In addition, genetic analysis has revealed that TTV has an immense genomic diversity, or heterogeneity, with only 60% homology across all of the known TTV genotypes. [5] The genotypes of TTV have been represented in a number of methods across several studies, and the last few years have seen several reorganizations of the structure of TTV genotypic structuring. In general, TTV is generally organized into five or six groupings of genotypes; each genotype group is based upon several factors. Primarily, the groups are organized by similarity of order of discovery, genetic structure and relative homology. In addition, a number of

groupings exist exclusive of the human genotypes, often depicted in their own cladogram; these being genotypes only found in animal species, with an entire group for primates, as well as one for bovine and porcine genotypes. [8] The genotypes of TTV show so much diversity that it is suspected that TTV may have number of distinct pathologies spread across its myriad genotypes. [7] Part of TTV's diversity, according to Manni et al. is due to a high level of genomic recombination, which they analyzed by studying various phylogenic groups of TTV. In contrast to TTV, this genomic diversity is one area where CAV notably differs from TTV. Unlike TTV, CAV has a fairly conserved genotype. [9]

A TTV variant called the torque teno mini virus (TTMV) was discovered in 2000. [2] It is a 2.9 kilobase virus, but its relative genomic framework is nearly identical to the genomic structure of TTV. TTMV is also exceedingly diverse, and has similarly numerous genotypes to those of TTV. Researchers found that most humans are infected with both TTV and TTMV simultaneously. Scientists believe that a further comparison of TTMV may aid in the understanding of TTV's physiological role, and potential pathological function. [2, 29]

Despite its high level of genotypic variance, TTV has a highly conserved genetic framework. All of its genotypes share a similar length, with most approximately 3.8 kb in length. The variance of TTV, when seen across a large selection of its genotypes is between 3.6 and 3.8 kilobases. [3, 12]

It has been observed that over one-third of the TTV genome is comprised of UTR. [9] This UTR is one unifying characteristic of the various TTV genotypes, with a high degree of sequence conservation among the genotypes, including those genotypes that are considered to have highly divergent sequences. One common feature of the various TTV UTRs is the presence

of an inclusive GC-rich 113 nt region. [27] Within this region is a 36 nt stretch which forms a stem-loop structure and serves as the origin of DNA replication for TTV. Like TTV, CAV has a 36 nt stem-loop forming region at the origin of replication and CAV shares over 80% homology with TTV for this particular region; however, this is the only region of significant sequence homology between CAV and TTV, for the rest of their DNA there is little similarity in their exact nucleotide sequence. [27]

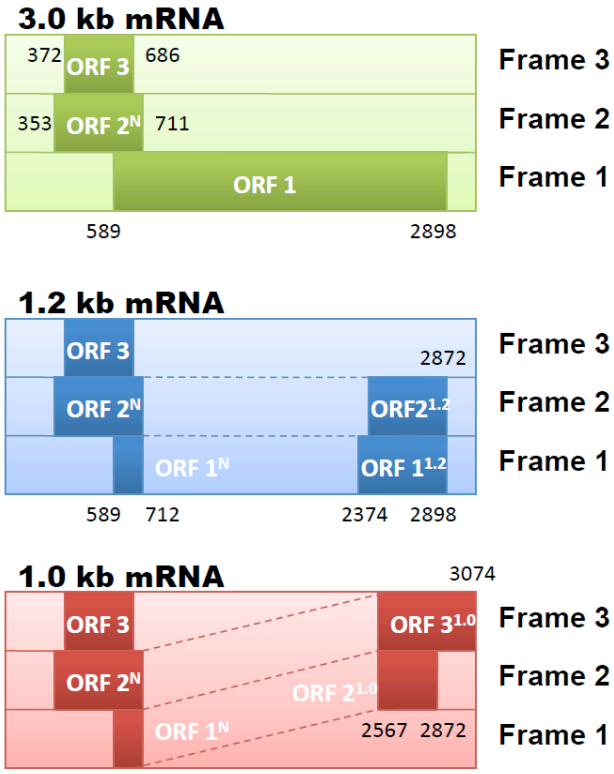

Figure 2: TTV transcribes three messenger RNAs of lengths 3.0 kb, 1.2 kb, and 1.0 kb. For each transcript, ORF 3 codes for TTV-VP3, our protein of interest.

In addition, all genotypes thus far studied by researchers have a very similar set of ORFs and spliced mRNAs. TTV possesses three conserved spliced mRNAs, with lengths of 3.0, 1.2, and 1.0 kb. In TTV, three ORFs are known to exist, here denoted as ORF1, ORF2, and ORF3. [3,7] ORF1 is present in every genotype of TTV, but is only partially conserved across many of the

more distally related genotypes of TTV. ORF2 consistently appears with similar variability. ORF3 is present in all mRNA transcripts, and ORF3 overlaps with ORF2, but in a different reading frame. [12] In TA278, all transcribed mRNAs use the same TATA box and poly-A signal, found on the antisense strand at bases 86-90 and 3073-3079, respectively. Also in TA278, transcriptional regulation is controlled by critical regulatory elements at -76 and -154 nt which govern the TTV promoter. [9] In addition, each mRNA transcribed by TTV is polycistronic, with one ORF in each of its reading frames. [8] CAV also transcribes multiple mRNAs, although most CAV studies have focused upon the transcription of its large, unspliced mRNA of 2.0 kb, and have given little focus to its spliced mRNAs and their putative proteins. Further, the genomic distribution of the putative ORFs of both TTV and CAV are quite similar, providing evidence that the biological functions of their respective viral proteins may share a common function at some level. [8,9, 27]

Due to a number of limitations, studies into the proteins of TTV have been mostly unsuccessful. There is currently no culture system for TTV, making a study of the expression and function of its proteins difficult. In addition, many of the TTV genotypes are present simultaneously, making their individual separation and study complicated. [8]

The similarities in both ORF distribution and genomic structure of TTV and CAV have led researchers to hypothesize, that like CAV, TTV expresses three main viral proteins. [13] However, TTV has been reported in some cases to produce greater than three proteins. One study, using TTV genotype 6, purports that TTV expresses up to seven distinct proteins; however, the remaining four proteins are all coded by the alternately spliced mRNAs of lengths 1.0 and 1.2 kb, and because they are expressed in considerably lower populations than those

proteins coded by ORF1, ORF2 and ORF3, their role in TTV has yet to be investigated. [25] The functions of CAV's three proteins, VP1, VP2, and VP3, have been established to some degree. VP2 and VP3 are believed to be early proteins, as they are often expressed approximately twelve hours after infection. VP1 is likely a late protein, as it is often seen expressed at twenty-four hours or later. VP1 is a protein which has been shown to exist exclusively in the capsid of the virus, and it is believed that this protein forms the capsid. [14, 20] Further, in 2001, Yamaguchi et al. showed that the mutation of a particular glutamine to a histidine results in a loss of pathogenicity for CAV. VP2 is less well characterized, but it is has been speculated, by Noteborn et al. to serve a role as a scaffold protein in the assembly of the virus. Evidence towards this proposed role for VP2 as a scaffold protein was provided by Noteborn and colleagues, who showed that expressing only VP1 or VP2 does not allow for the generation of CAV-specific antibodies (which bind the capsid). [20] This suggests that both VP1 and VP2 must be expressed to generate the fully functional capsid, and so VP2 must play some role in the assembly of VP1 into a capsid structure. In addition, VP2 in both CAV and the ORF2 of TTV have been shown to have phosphatase activity. VP3 of CAV, known as apoptin, is a non-structural protein of 121 residues has been shown to induce apoptosis, within various types of cells after infection. Further, apoptin has been shown to cause apoptosis specifically in transformed or tumorigenic cell lines, but not in normal cells. Additionally, apoptin is able to induce apoptosis independent of p53, a key protein in tumor suppression and cellular growth, which is often overridden by transformed growth. [9] Hino et al., in research submitted for publication, have reported that apoptin is necessary element in both the processes of CAV virion formation and DNA replication. For both CAV and TTV, their spliced mRNAs (for TTV, the 1.0 and 1.2 kb

mRNAs) can also generate proteins, however, if these other proteins play a role in the biological function of either virus it has yet to be demonstrated. [9, 27]

For TTV, ORF1 codes for a protein of 770 amino acids. ORF1 contains several conserved replicase protein motifs, as well as a highly basic region near its N-terminus. These two properties have lead researchers to believe that this ORF codes the major structural protein of TTV, as the two aforementioned properties are both qualities present in circoviral capsid proteins. Less is known about the protein coded by ORF2, although it occupies a similar orientation to that of CAV ORF2, and so it may too act in some sort of structural assembly function. [8] The protein for ORF3, TTV-VP3, is coded for by all three mRNAs. Like apoptin, TTV-VP3 has been reported to be a non-structural protein of 105 residues. To a greater degree than the other two main proteins coded by TTV, TTV-VP3 shows a number of structural qualities to apoptin. Their primary sequences bear little resemblance, but the makeup and distribution of amino acids in TTV-VP3 is perhaps the strongest indicator of their similar biological function. First, TTV-VP3 and apoptin have a high level of proline residues—10.5% for TTV-VP3 and 11.5% for apoptin. Further, these are in a highly dispersed distribution, suggesting that these residues serve a structural function in the respective viral proteins. Another shared quality is their high concentration of threonines—TTV-VP3 has 12.4%, while apoptin has 11.5%—suggesting that threonines play an important structural role for both. Finally, both TTV-VP3 and apoptin contain a section of hydrophobic amino acids close to their N-terminus which forms the site of apoptin multimerization and export. These three similarities are significant enough support the idea that TTV-VP3 may harbor some similar activity to apoptin. [9, 27]

Similarities between TTV and CAV
Each has a single-stranded DNA genome with ORFs on the antisense strand.
36-nt section proximal to the origin of replication for TA278 has more that 80% similarity to that of CAV.
Both TTV and CAV transcribes multiple mRNAs.
Approximate orientation of the three putative ORFs in both viruses is quite similar.
The capsids of both TTV and CAV are each composed solely of their VP1.
CAV ORF 3 expresses apoptin, TTV-VP3 is apparently a similar protein.
Both TTV-VP3 and apoptin have a high level of prolines (10.5% and 11.5%, respectively) and threonines (12.4% for TTV-VP3 and 11.5% for apoptin), as well as section of hydrophobic residues in their N-terminal regions.
Notable differences between TTV and CAV
Different genomic sizes for TTV and CAV—3.8 kb and 2.3 kb, respectively.
TTV displays enormous genomic diversity, with over 40 genotypes, unlike CAV which is highly conserved.
No actually homology of TTV-VP3 and apoptin primary sequence.

Table 1: This table provides a simplified overview of some of the major similarities and differences between the viruses TTV and CAV.

Recently, a study published by Noteborn et al. reported that TTV ORF3 has a similar function to that of the VP3 of CAV, apoptin. [14] In an undisclosed method, these researchers expressed the protein of ORF3 from TA278, genotype 1 of TTV, and transfected the protein in hepatocellular carcinoma cells (HCC). They also claimed that TTV VP3 was able to induce apoptosis in a similar manner to apoptin in CAV, although on a more limited scale. In addition, they made a thorough comparison of the overall genomic framework of TTV and CAV which they used as their foundation for their work with the TTV protein. They named the protein "TTV apoptosis-inducing protein", TAIP, and in the conclusion of the report, Noteborn et al. proposes TTV VP3 as a potential cancer therapeutic. [14] In subsequent years, no further reports have made further headway into the study of TTV VP3, and in the past several years a number of

reports have noted that the research into TTV VP3 by Noteborn et al. lacked clarity. Through correspondence with Mathieu Noteborn and his laboratory in Leiden, Netherlands, it has been established that all known samples of the ORF gene and TTV VP3 in their possession have been lost due to a laboratory move. Despite these setbacks, we believe TTV VP3 remains one of the most intriguing aspects of TTV and merits an exhaustive study.

In the following sections we will discuss our work to synthesize TTV ORF3 and to study the expression and behavior of TTV VP3. By synthesizing the gene for ORF3, we will be able to study the expression of TTVVP3 extensively. Through this work, it is our hope to build a foundation for—and perhaps reignite interest in—this poorly characterized protein, which may indeed share some activity with apoptin.

Materials and Methods

Phylogenic analysis

Based on the GenBank accession number (AB008394) and ORF3 sequence published by Noteborn et al., we positively identified the TTV strain as TA278—the prototypic TTV strain. Sequence analysis and homology comparisons were performed between TA278 and a number of other TTV strains based upon the phylogenic groupings from Jelcic et al., and Hino and Miyata. [8, 6] Special focus was given to determining the homology of TA278's ORF3 with those genotypes of the same phylogenic group. In addition, cross comparisons were made between TA278's ORF3 sequence and those of numerous genotypes, both human and animal, more distantly related (as evidenced by their phylogenic groupings and overall sequence heterology).

TTV ORF3 construct design

TTV VP3 Sequence, 40bp oglionucleotides	**Contains**
1. gc**GAATTC**a***atg***atcaacactaccttaactggcaatggt (*39 nt*)	1.EcoRI restriction site, 2.start codon
2. actcaagtatacttagctcccacgctgctatgtgcgggtg	
3. tcccgacgctgtcgctcattttaatcatcttgcttctgtg	
4. cttcgtgccccgcaaaacccacccccctcccggtccccagc	
5. gaaacctgcccctccgacggctgccggctctcccggctgc	
6. gccagaggcgcccggagatagagcaccatggcctatggct	
7. ggtggcgccgaaggagaagacggtggcgcaggtggagacg	
8. cagaccatggaggcgccgctggaggacc (*28 nt*)	

TTV VP3 Reverse Compliment, 40bp oglionucleotides	**Contains**
1. gc**GGATCC**gtctagcaggtctgcgtcttcgggtcctccag	BamHI restriction site
2. cggcgcctccatggtctgcgtctccacctgcgccaccgtc	
3. ttctccttcggcgccaccagccataggccatggtgctcta	
4. tctccgggcgcctctggcgcagccgggagagccggcagcc	
5. gtcggaggggcaggtttcgctggggaccgggaggggtgg	
6. gttttgcggggcacgaagcacagaagcaagatgattaaaa	
7. tgagcgacagcgtcgggacacccgcacatagcagcgtggg	
8. agctaagtatacttgagtaccattgcca (*28 nt*)	

Table 2: Oligonucleotides used for PCR-directed gene synthesis, forward and reverse complement.

Initial PCR assembly was performed under the influence of several variables, allowing for a broad range of observable product efficiency. The 16 construct oligonucleotides were initially mixed into a single-tube mixture, each segment added at a volume of 1ul. This mixture

(containing the full-length gene in 16 oligonucleotide segments) was split by volume, with half of the sample being set aside at room temperature. The remaining mixture was heated to 95 C and then allowed to cool slowly for an extended period of time at room temperature. This was done in order to facilitate the thermodynamic formation of pre-annealed complexes of oligonucleotides. The pre-annealed mixture along with the remaining mixture were each distributed among numerous reactions, each reaction at a total volume of 20ul. These reactions were conducted in a PCR mix comprised of 1 µl oligonucleotide mixture, 1 µl of *Taq* polymerase (Promega), 0.02 of *Pfu* polymerase (Stratagene), 5 µl of *Taq* buffer, 1 µl of 0.2 mM dNTPs, with remainder volume of distilled water to a volume of 20 µl. *Pfu* polymerase was utilized in conjunction with *Taq* polymerase for its high degree of proofreading ability and processivity.

The base PCR program was varied in regards to stringency and number of total program cycles. Stringency variation consisted of altering the annealing temperature for the second stage of PCR. Half of all reactions were run at 60 C, while the other half were run at low stringency 52 C; for our purposes, these will be considered high and low stringency, respectively. Varying the annealing base temperature, we postulated, may allow for preferential annealing of longer assembled DNA products, perhaps increasing the overall efficiency of generation for full-length genes. Based on the techniques provided by Stemmer et al., we considered 55 cycles to be the maxima for cycle number. In addition, we ran PCR reactions at two lower cycle numbers, 40 and 25 cycles. All of the reactions, in conjunction with the previously mentioned variable divisions, were again subdivided into three batches (one third of all samples in each division), based on cycle number. In this manner, the three variables (pre-annealing, stringency, and cycle number) resulted in 12 unique reactions. Two samples of

each reaction type, for a total of 24 reactions, were run through the PCR program with the appropriate cycle number. The PCR program consisted of repeating cycles of 94 C for 30 seconds, 52 or 60 C for 30 seconds, and 72 C for 30 seconds. All 24 samples were visualized by electrophoresis on a 0.9% agarose gel.

All of the pre-annealed 40 and 55 cycle samples displayed optimal bands, in regards to both size and concentration, and so these 8 PCR products were carried forward and amplified with primers consisting of the N-terminus forward primer and the C-terminus reverse primer and using Pfu polymerase for proofreading. This amplification was done in order to ensure a contiguous full-length product. For this round of amplification, the PCR program consisted of 40 cycles of 94 C for 30 seconds, 52 C for 30 seconds, and 72 C for 30 seconds. With its completion, the various PCR samples were visualized by electrophoresis on a 0.9% agarose gel. Those samples which displayed the optimized band length (approximately 330-340 bp), band range and product concentration were selected. These products were excised from the agarose gel and the DNA product purified via a GeneClean Kit (MP Biomedicals). Duplicate samples of the same variety were combined to increase overall product volume.

Due to the polyadenylation generated by *Taq* polymerase, a final round of PCR reamplification was conducted in the absence of *Pfu* polymerase. Before reamplification, the PCR products were visualized on a 0.9% agarose gel at three dilutions with distilled water to final concentrations of 1/100, 1/1000 and 1/10000 of the original sample concentration. The most refined bands were those of the 1/1000 dilution. These agarose bands were excised and purified as mentioned above. The purified bands were reamplified with *Taq* polymerase using recipe described above, with the PCR program of 40 cycles of 94 C for 30 seconds, 52 C for 30

seconds, and 72 C for 30 seconds. Four samples were selected for continuation: 40-cycle high stringency, 40-cycle low-stringency, 55-cycle high-stringency, and 55-cycle low-stringency. In subsequent discussion, these samples will be designated with the monikers 40H, 40L, 55H and 55L.

Cloning, transformation and sequencing

The synthesized genes, 40H, 40L, 55H and 55L, were cloned into a plasmid expression vector pGEM-T under the control of the bacterial T7 promoter (SP6 promoter in the antisense direction). The vector contains a *lacZ* start codon and *lac* operatorto allow for blue-white screening, and a constitutively-expressed *amp* gene for conferring ampicillan resistance. The DNA samples were ligated into this plasmid vector, creating the constructs pTTV-VP3 40H, 40L, 55H and 55L. The constructs were transformed into competent DH5α *E. coli* cloning strain cells. For instance, DH5α *E. coli* utilizes a mutated RecA allele to prevent integration of plasmids by homologous recombination. The pTTV-VP3 constructs 40H, 40L and 55H were positively selected via blue-white screening from ampicillin plates coated with 0.1 mM IPTG and 40 µg/ml X-gal. These positive colonies were inoculated and grown overnight in 2 ml LB media with 2 µl 100 mg/ml ampicillin and the plasmid constructs were harvested from 1.5 ml of culture using small-scale DNA purification technique [details]. Construct DNA was isolated and resuspended in 1 M TE buffer (10 mM Tris/1mM EDTA) to a final volume of 50 µl. The construct DNA was digested by enzyme restriction of EcoRI and BamHI to verify proper transformation and ligation. The restrictions were conducted at 37 C for 1 hour in the presence of 1 µl RNAse A and then visualized by electrophoresis on a 0.9% agarose gel. With confirmation of the correct insert size,

the remainder of the DH5α *E. coli* cultures (0.5 ml) were used to inoculate three large-scale cultures of 100 ml LB media with 100 µl 100 mg/ml ampicillan which were incubated overnight at 37 C. The cultured cells were harvested, and the cells were lysed and the plasmid vector was purified with a MidiPrep kit (Promega). The construct DNA was resuspended in 300 µl distilled water. Again, the purified DNA was digested by enzyme restriction of EcoRI and BamHI to verify the presence of proper-length insert. Using UV spectroscopy, the concentrations of the three construct samples were determined and used to create final dilutions for sequencing (Macrogen USA Inc.).

Cloning into GFP and 3xFlag vectors

The lone successful sequence, pTTV-VP3 40L, was restricted with EcoRI and BamHI and the insert was purified using a GeneClean Kit (MP Biomedicals). The 40L gene was then cloned into both a pEGFP-c1 vector and a 3xFlag vector (Sigma-Aldrich). The pEGFP-c1 allows for the expression of target proteins with a N-terminal GFP tag. This would create a functional TTV VP3 fusion with a GFP vector, allowing for visualization of protein localization *in vivo* by fluorescence microscopy. This vector is under the control of the cytomegalovirus (CMV) promoter and contains the gene for kanamycin resistance. The 3xFlag vector is under the control of the cytomegalovirus (CMV) promoter region which drives constitutive control and it contains the gene for ampicillan resistance controlled by a seprate promoter [6]. The 3xFlag vector was used to express TTV ORF3 in conjunction with three small protein tags, allowing the protein to be cloned into human cells for killing assays with an unobtrusive tagging system. In addition, 3xFlag is an excellent epitope, being easily bound by a corresponding 3xFlag antibody. For the

pEGFP-cl vector, the insert was ligated at a ratio of 3-to-1 (6 µl:2 µl), insert-to-vector ratio. This construct, pVP3-GFP, was transformed in DH5α *E. coli* and plated on kanamycin plates. For 3xFlag, ligation was conducted at a ratio of 4-to-1, insert-to-vector, and the resultant construct pVP3-Flag vector was transformed into DH5α *E. coli* and plated on ampicillin plates. Positive colonies were selected and grown overnight in 2 ml LB media with 2 µl 100 mg/ml kanamycin or ampicillin and the DNA was harvested from 1.5 ml of culture using small-scale DNA purification technique as detailed above. Both vectors were isolated and resuspended in 50 µl 1 M TE buffer. The construct DNA was digested by EcoRI and BamHI restriction to verify transformation and ligation. The restrictions were conducted at 37 C for 2 hours in the presence of 1 µl RNAse A and then visualized by electrophoresis on a 0.9% agarose gel. With a positive restriction, the remainder of the vector-transformed DH5α *E. coli* cultures were used to inoculate large-scale cultures with different pVP3-GFP and pVP3-Flag samples in 100 ml LB media with 100 ul kanamycin or ampicillin, respectively. After incubating for twenty-four hours, cultured cells were harvested and lysed. The plasmid vector was purified by a MidiPrep kit (Promega). The isolated construct DNA preparation was resuspended in 300 µl distilled water. To confirm the construct's successful isolation, the DNA was digested by restriction with EcoRI and BamHI. Using UV spectroscopy, the concentrations of the construct samples were determined and used to create final dilutions of for sequencing (Macrogen USA Inc.).

Tissue culture and transfection

Tissue culture experimentation was carried out in a tumorogenic human cell line H1299, or NCI-H1299. H1299 is an adherent, cancerous lung cell line derived from the lymph node. [18] After removal from cryostasis, the H1299 cells were maintained in DMEM and 10% PBS.

Transfections were performed in six-well tissue culture plates as described by the Effectene protocol. For each well, a circular coverslip was first placed in the basin and then to each well 2 ml of cell-containing prediluted media was added. This prediluted media was seeded with the passaged cells to reach 70% confluence for the next day. After growing for twenty-four hours, or achieving the proper confluence, the six-well plate was transfected. In this way, the coverslip-bound cells can be transfected and mounted for visualization. Multiple rounds of transfections were done in order study localizations a multiple time points according to the Quiagen Effectene protocol. To each plate, two identical samples and one control were added across the six wells. The control was transfected with a GFP-expressing construct to achieve a uniform GFP localization pattern.

For the pVP3-GFP construct, the transfections were carried out in H1299 cells in two rounds. These were fixed with 4% paraformaldehyde in 1M non-sterile PBS approximately one day following transfection to generate a twenty-four hour sample and approximately two days following transfection to generate the forty-eight hour sample. Once fixed, the coverslips were mounted on slides with 35 µl mounting media consisting of 50% glycerol, 100 mM Tris (pH 7.5), 2% DABCO, and DAPI (nuclear stain).

Apoptosis detection via crystal violet staining

Using the pVP3-Flag construct described previously, H1299 cells were passaged and transfected using the standard Effectene protocol in six-well tissue culture dishes. Two samples transfected with pVP3-Flag construct occupied two wells, while a third, functioning as control, consisted of non-transfected H1299 cells. Cells were observed each day following transfection by simple microscopy, noting changes in confluency between sample and control wells. When obvious confluency differences were noted, the plates were fixed (as previously described). Immediately after fixing, the wells were immersed in PBS and a single drop of crystal violet (methyl violet 10B) dye was applied to each of the three sample wells. Tissue culture tray was lightly swirled and allowed to incubate at room temperature for three minutes. The PBS/crystal violet suspension was aspirated from the cells and cells were promptly washed with PBS (to remove residual crystal violet), agitating for one minute, before this too was aspirated off. Cells were allowed to dry at room temperature.

Results

Phylogenic Analysis of ORF 3

Genotypic analysis of TTV provided us with several interesting observations. At this time, multiple phylogenic trees exist for TTV, but many of these fall into several distinct arrangements. Some are based on ORF 1 homology, drawing the TTV genotypes into clustered groupings; others provide a breakdown based on full genome analysis. In some cases, the human genotypes and lesser mammalian genotypes are separated and viewed independently. Due to the immense diversity of TTV—40% sequence heterology, with over 20 confirmed

human genotypes, as well as many others from other species, including numerous swine, primate and bovine strains—these various published tree groupings and cladograms show an erratic level of relative diversity in their structure and presentation, making a conclusive study of TTV's genotypic relationships a near impossibility. Hino and Miyata present what is perhaps the most lucid phylogenic representation to date. [6] Their analysis provides a bifurcated cladogram, with one part providing full-sequence-based analysis, and the other providing a phylogenic representation based on the sequence similarity of the non-translated regions of TTV. This second cladogram, in addition to the most studied genomic groups, also incorporates many new TTV strains, including a battery of novel bovine TTV strains. We selected this representation for its up-to-date and clear presentation of TTV genotypes, as well as its authors' encompassing understanding of TTV and its current related research. Based on Hino and Miyata's phylogenic representation, the prototypic TTV genotype TA278 resides in group 1, also comprised of JA1, T3PB, TKM1, THEM1, TFC3155. These six members of group 1 are all human strains, and are the most prominent of human TTV varieties. In addition, these genotypes are given numbered prefixes in order to indicate their order of discovery—1-TA278 being the first TTV genotype to be discovered. A genome-wide comparison of sequence homology between these six genotypes from group 1 showed considerable similarity. Further, an analysis of group 1's ORF 3 showed a high degree of sequence homology, with only a few residues changing across the 6 genotypes. Two genotypes, TRM1 and JA9, showed almost complete ORF 3 homology with TA278, differing by only one and two nucleotides, respectively. This analysis provided sufficient evidence that the ORF 3 sequence of TA278 is quite conserved across the entirety of group 1, as defined by Hino and Miyata. Our further analysis, expanding

the comparison of TA278's ORF 3 across the other main genotypic groups showed some retention of homology, with most genotypes retaining more of the TA278 sequence, on average, in comparison to their overall sequence. Although there was still a great deal of variance across genotypes, especially in full-sequence comparisons, a logical consequence of the highly diverse nature of TTV. Importantly, this analysis of was conducted to verify the presence of ORF 3 and to objectively examine ORF 3 sequence diversity.

PCR-directed gene synthesis

Using PCR-directed gene synthesis, we were able to effectively use a number of variable conditions, such as cycle number and degree of stringency, to synthesize the full length TTV ORF 3. When we observed the PCR products by gel electrophoresis, we noticed the bands to be approximately 330 nt in length, which was appropriate for the expected product. Next, we were selected those products which we observed to have the greatest potential fitness—products that displayed the most concentrated bands that at the same time were most constrained to the acceptable range for product length. The bands which we selected were 40L, 40H, 55L and 55H, and these we carried forward to further amplification. The degree of stringency seemed to have some level of effectiveness refining the PCR products, as the high-stringency bands displayed a more well-defined band range, indicating that the products were all at nearly identical lengths; however, this was not so significantly pronounced so as to not pursue those products generated by low-stringency synthesis. Cycle number seemed to be a more critical factor—heavily influencing both product band range and concentration. Some bands not selected for further amplification showed more dispersed products, including several whose

product bands were visualized as a long swath of product, likely indicating partially-formed or aberrant products. Several others PCR products, were not selected because they displayed faint or overly bright bands, indicating overly-low or too-great product concentrations, respectively. In addition, little difference was displayed between pre-annealed bands and those that were not; however, all of the pre-annealed bands were selected over the others as we perceived their bands to be slightly more refined in contrast to those reactions which were not pre-annealed. Despite this, it is difficult to say the overall difference of effectiveness engendered by the use of pre-annealing the PCR reactions.

The gel electrophoresis showed all of the product bands to be of the appropriate length and of a well-defined size, indicating their successful reamplification. In order to reduce the chance of mutagenesis with this lengthy PCR, it was important in this process to conduct the initial full-length assembly in the presence of a DNA polymerase with a high degree of proofreading and processivity—in this case *Pfu* polymerase in conjunction with *Taq* polymerase served this purpose.

Engineering TTV VP3 expression vectors

The samples were successfully transformed into competent E. coli cloning strain DH5α and a large-scale quantity of the plasmid construct pTTV-VP3 was harvested. Analysis by double restriction enzyme digests, EcoRI and BamHI, confirmed the presence of the appropriate length fragment at approximately 310-340 nt in length in the various samples. Despite multiple potentially viable samples, several 40L transformants were the only constructs successfully carried through to the final stages for sequencing. Of the three constructs analyzed for

sequencing, the construct pTTV-VP3 40L3 was the lone successful sequence. For this construct, all bases from the original PCR-directed gene synthesis were successfully retained, with no presence of mutagenesis, damage or frameshift present in the final sample.

The pTTV-VP3 construct was digested and the gene of interest was successfully purified. The synthesized VP3 gene was cloned into both GFP and 3xFlag vectors. After transforming the constructs in DH5α cell, positive colonies for pVP3-GFP and pVP3-Flag were selected from Kanamycin and Ampicillan-enriched plates, respectively. These two types of constructs were grown in large-scale culture and the plasmids were harvested and purified for sequencing. The sequencing analysis confirmed that both constructs were positive for the TTV VP3 gene and that the vectors contained the intact sequence. The two vector constructs were assessed for insert and linearized vector length by restriction enzyme digestion. Visualization by electrophoresis confirmed the presence of the correct length VP3 DNA insert in both plasmid systems. The linearized plasmids appeared at their correct lengths, approximately 4 kb for the linearized GFP vector and 6 kb for the linearized 3xFlag vector.

TTV VP3 expression analysis

Effectene-induced transfection was successfully conducted on H1299 adherent lung carcinoma cells using the pVP3-GFP construct. H1299 cells are immortalized, and as such, can grow—if unhindered—indefinitely. Importantly, H1299 cells have a homozygous partial deletion of the p53 gene and as a result, are unable to express the p53 protein, a tumor suppressor. [18] The H1299 cells were transfected with the pVP3-GFP construct in two separate rounds—one fixed at a twenty-four hour time point, and another fixed at forty-eight hours. This

was done in order to observe the localization tendencies of TTV VP3 over time, perhaps allowing for the visualization of a changing localization pattern, or even the occurrence of VP3-induced apoptosis.

The GFP-tagged VP3 distributions at twenty-four hours (see A-C, figure 5), although not particularly skewed towards any localization trend, appeared in clear contrast when viewed alongside those H1299 cells observed from the forty-eight hour transfections. At 24 hours, there were at least three comingled states of relatively equal distribution present among the transfected cells. Of all the observed patterns, perhaps the most interesting localization pattern consisted of strong, punctate GFP signals situated in the cytoplasm of the cell, as displayed in A and B in figure 5. These appeared as both singleton entities, as well as clustered groupings. In most of these cases, the GFP signal appeared as a uniformly bright, clean-edged foci, with little occurrence of any distribution gradient along the edges. This may suggest a trend towards cytoplasmic, proteinaceous self-aggregation at this temporal stage. Among the transfected cells, GFP-tagged VP3 localization distributions could be classified into four distinct populations: cells with relatively uniform GFP-tagged protein distribution, both cytoplasmic and potentially nuclear, with fusion protein loci; Cells with fairly even cytoplasmic distributions, with punctate GFP signals; cells with purely cytoplasmic distributions but only in explicit nodes and clusters without any diffuse distribution; cells with predominantly nuclear distributions with the presence of bright GFP-tagged VP3 loci. The most common characteristic was the presence of the aforementioned GFP-tagged protein nodes. In addition, when cytoplasmic, the punctate signals were sometimes oriented in close relation to the nuclear membrane. In most of these instances, the tagged proteins were clustered in tight groups near the nuclear

membrane (see A, figure 5). Some nuclei, especially those with lobed or curved oblong shapes, displayed these tagged proteins in the deepest point in their curve or their most prominent cleft. Those cells with both nuclear and cytoplasmic distributions showed distinctly weaker GFP signals in their nuclei, perhaps indicating that the nuclear concentration of tagged protein was either lessened, or that the strength of the cytoplasmic signal was such that it was able to visibly permeate through the nucleus (in such a case, the nucleus may indeed be relatively free of the tagged protein). One other characteristic of the twenty-four hour pVP3-GFP transfected cells was a thread-like diffusion pattern which the tagged protein displayed at the cellular periphery (see A, figure 5).

At forty-eight hours, pVP3-GFP transfected cells showed a markedly different distribution—nearly all of the cells displayed a primarily cytoplasmic distribution (D and E, figure 5). These transfections all had clear nuclei (free of GFP-tagged protein), with all nuclear GFP signaling the obvious consequence of strong GFP-tagged protein concentration around the nuclear membrane exterior. The general cytoplasmic pattern was that the tagged protein was distributed throughout the cytoplasm, and with less evident protein foci, although these were still present. At this later stage, the punctate, GFP-tagged proteins occupied relatively diminutive loci within the cell. These aggregations were often widespread and diffuse in the cytoplasm, appearing as a multitude of fine, bright points in a wash of green coloration. Similar to the twenty-four hour transfections, a number of cells in this second round showed punctate protein aggregations proximal to the exterior of the nuclear envelope. These foci were often larger than the median-sized protein aggregations—either the result of the amplified brightness

of densification of small protein aggregations, or perhaps these punctate GFP signals indicate the presence of larger tagged-protein aggregates.

Apoptosis detection via crystal violet staining

After transfection, H1299 cells were observed once per day over a five-day period. On the fourth and fifth days, sizable differences in confluency were observed. Cells were fixed with 4% paraformaldehyde on the fifth day post-transfection. Staining with crystal violet dye provides a strongly-colored visual indicator as to the degree of cell confluency, allowing for differences to be measured by simple unaided observation. For our samples, a clear difference was observed between the control and those cells transfected with pVP3-Flag, Figure 6. For those cells transfected with the pVP3-Flag construct, there is a sharp contrast in cell density in comparison to the control cells, indicating significant cell death among the transfected cell populations.

Discussion

Benefits and applications of PCR-directed gene synthesis

This project was extensively benefited by the use of PCR-directed gene synthesis. This fairly novel technique proved to have a number of intrinsic benefits on the overall flexibility and efficiency of creating, verifying, and utilizing a synthesized gene. By working with a number of overlapping sense and antisense DNA segments, it was possible to assemble the full-length TA278 VP3 in a secure and cost-efficient method, in a variation of gene synthesis which Stremmer et al. refer to as "gene shuffling". [28] Each segment was purchased as a short,

single-stranded 40 nt oligonucleotide—significantly more cost-effective than purchasing a full-length gene. Each fragment was also more stable than a larger DNA molecule, and allowed for accurate temperature stability information about each particular segment. Further, any small mutation or error could be pin-pointed and dealt with on the scale of a small, facile oligonucleotide and not on the scale of a full length gene. Also, using numerous segments allowed for versatile amplification, where each oligonucleotide can act as an individual primer; future experimentation can also be benefited by the replacement of these short segments with mutated fragments to readily conduct mutagenesis. Based on our success, PCR-directed gene synthesis, if implemented properly, could be used on significantly larger stretches of DNA. By ensuring an adequate number of cycles it may be possible to construct genes of thousands of bases in length with a high level of efficiency.

Finally, this type of segmented assembly can also be applied to the creation of future alternate constructs of TA278's ORF3 in order to study the individual functions of the sequence on a section-by-section basis. By simply designing and purchasing more oligonucleotides, engineered with appropriate restriction sites their ends, it becomes possible to assemble new constructs. For instance, one may wish to determine the location of the nuclear export sequence (NES) of TTV VP3, and by designing various sectional constructs it will be possible to knock-out nuclear export of the protein, allowing for a greater knowledge of TTV VP3's sectional properties.

In this way, PCR-directed gene synthesis provides a great advantage this type of experimentation, allowing for assembly of genes in modular, easy to manipulate sections. For

this particular project, the groundwork laid out by the use of this "gene shuffling" technique will likely prove highly beneficial to future iterations of TTV VP3 studies.

The enigma of TTV VP3

Despite the abundance of information provided by the florescence microscopy images of H1299 transfected with pVP3-GFP, the overall behavior of the protein and its potential *in vivo* role remain largely a mystery. The fusion TTV VP3-GFP fusion protein, as evidenced by its twenty-four and forty-eight hour distributions, has a great deal of time-dependent localization. In the case of the twenty-four hour samples, the fusion protein displayed a wide range of distributions, both cytoplasmic and nuclear. For those cells fixed at forty-eight hours, the fusion protein had clearly shifted towards to a primarily cytoplasm-exclusive distribution. Understanding this general localization shift may aid in developing an understanding of the dynamic temporal behavior of TTV VP3.

For those cells fixed at twenty-four hours, there seems to be little correlation between the fusion protein's presence and any particular region within the cytosolic environ. Two cytoplasmic regions showed noteworthy distributions: the external nuclear periphery and the cytoplasmic periphery. At the nuclear periphery, the there was a common tendency for the clustering of punctate foci. This could be cause by many dynamic behaviors, both cellular and protein-based. It is possible that the export of the fusion protein allows for the concentrate the protein in localized region of the nuclear exterior. If all TTV VP3 were exported by a particular family of nuclear pore complex, then it may be possible for this trend of clumping and clustering to be accentuated in a particular region around the nuclear envelope. TTV VP3's

behavior may be controlled by some degree of concentration dependence, such that at only at certain concentrations does it aggregate or localize to a particular cellular region

The aggregated fusion proteins also seemed to appear in the nucleus of the one-day transfected cells. However, many of these nuclear signals are the result of punctuate foci, and although they appear in the same region as the nuclear stain, they may in fact be on either exterior face of the nuclear envelope. Other cases are clearer, especially when the fusion protein is observed to be dispersed fairly uniformly. In these cases, the GFP signal appears to be just as strong throughout the cell, indicating that perhaps the fusion protein exists in both the cytoplasm and nucleus. Nonetheless, it is possible that the GFP signals are predominantly cytoplasmic and only appear to reside in the nucleus.

The punctate foci, which were so predominantly displayed by many of the one-day cells, seem to play a key behavioral role for VP3 and perhaps provide a clue to the potential biological function of the protein. Due to these fusion protein foci, VP3 seems to have a propensity for large-scale aggregation. This would indicate that the functional protein design either allows for or causes this self-aggregation. It would seem likely that the proteins conformation allows for this aggregation with minimal processing, as the aggregates distinctly appear in both the cytoplasm and nucleus in independent cases.

The aggregated fusion protein complexes are clearly of highly variable nature, as displayed by the many sizes and shapes of the punctate foci. These sizes can vary from highly diminutive to obtrusive, almost organelle-sized aggregates, and their shapes both circular and asymmetric. This may indicate that the aggregation is less structured and the potential aggregated structure more the result of random interactions and less based on uniformly-

repeated linking. This aggregation could be the result of a weak protein-protein interaction, which when compounded by the presence of the protein's high concentration and close proximity, may result in a glomming together of the proteins. In this way, a number of weak protein-protein interactions may have the capacity to build synergistically, ultimately aggregating into a stable multi-protein complex. The proteins may also be interacting covalently, and so in their interaction the proteins would be reversibly bonded to one another. The protein-protein linking involved in this sort of aggregation could be the result of a particular primary or secondary structure interaction, but this may not be the case due to the almost sporadic nature of the fusion protein complexes. This variety of binding could either be based on random interactions or perhaps binding is made more favorable via conformational changes within a particular bound protein.

The presence of the GFP tag on the fusion protein may also act to disturb its *in vitro* behavior, as the protein itself is only 105 residues in length, and so even the small GFP fused protein (238 amino acids) dwarfs VP3, and could have any number of spurious effects on replicating *in vivo* protein multimerization.

In concurrence with Noteborn's observations of the behavior of TTV-VP3, at two days after transfection the H1299 cells displayed a completely cytoplasmic distribution. [16] Obviously, this is a significant transition in overall localization when compared to those distributions observed for the one-day transfections. Unlike the one-day transfections localization—a variable distribution in both the nucleus and cytoplasm, as well as pronounced punctate aggregations—the two-day transfections showed little variance in localization. The two-day transfected cells displayed multimerization, although in markedly different from the

distribution visualized at the earlier time point. In this case, the aggregates were uniformly small and highly dispersed through much of the cytoplasm, giving the fusion protein signals a spattered appearance. Near the peripheral interior of the cell membrane the fusion protein signals appear lessened, and conversely, the signals appear to be denser proximal to the nuclear envelope. The trend towards spattered-looking distributions of TTV-VP3 may be due to the oversaturation of the cell with these proteins from unconstrained expression of VP3. Perhaps this greater level of concentration disturbs the protein's ability to form larger aggregates.

Simply from its fully cytoplasmic localization two days after transfection, we learned a significant fact about TTV VP3. Due to this obvious localization shift from a seemingly unlocalized distribution to a purely cytoplasmic one, we have established that TTV VP3 must have a functional nuclear export sequence (NES). This is significant as it indicates a notable characteristic about TTV VP3's behavior, which in turn may enable a greater understanding of the protein's intended biological function. Further, it provides us with an inherent future goal: to locate the NES and study its effect on TTV VP3 when disrupted. If it is truly similar to apoptin, TTV VP3 may also have a nuclear localization sequence and a multimerization sequence—the existence and location of which may be revealed by future research.

The cytoplasmic distribution may be due to a number of factors. If TTV VP3 also has a functional nuclear localization sequence (NLS), perhaps the nuclear export of TTV VP3 is occurring at a greater rate than its import to the nucleus, and so the state that we observe is composed of a skewed equilibrium of export and import. Another possibility is that the fusion protein is fully sequestered to the cytoplasm at this point of observation. Perhaps this could be

caused by a blockage or clog due to the volume of aggregated proteins. Also, the aggregated proteins may have become too generally large relative to the passable pores in their relevant NPC. The first explanation seems the most likely from our observational data, since there is an obvious signal gradient present near the nuclear periphery.

In addition, there were slightly greater cell concentrations for the control at the forty-eight hour time point, when compared to the pVP3-GFP transfected cells at this same time point. Indeed, this may indicate some degree of apoptotic potential, as the lysed cells, loosened and degraded, would have been washed away in the fixing stage of slide preparation. This observation is speculative and circumstantial, and further, is not fully substantiated by the transfection results; however, it merits further exploration into this potential phenomenon.

Due to similar observed confluence levels (by general observational approximation of living cell density) at both one-day and two-day transfection time periods, and because at these time points we were unable to observe any apoptosis evidence outright, it seems that if apoptotic activity exists for TTV VP3 it must be induced at some later time.

Noteborn et al. observed the VP3 to first take on a finely granular distribution, a description which matches our own observations. He observed that this granular stage transitioned into large-scale aggregation which led to the transfected cells becoming apoptotic and fragmenting at three to four days. Overall, our two-day observation is consistent with that of Noteborn and colleagues. Further, Noteborn et al. indicated that the localization of TTV VP3 was cytoplasmic prior to inducing apoptosis; however, he did not further examine or discuss the possible mechanism for this apoptosis. To learn whether TTV VP3 truly exerts apoptotic

activity upon tumorogenic cells, we must conduct a longer period of post-transfection observation before fixing.

Another powerful clue in revealing potential apoptotic activity of TTV VP3 was presented by the simple procedure of crystal violet staining. The noticeable differences in live cells between transfected and non-transfected cell samples is a strong indicator of TTV VP3's activity. One significant component of this result is the consistency of confluency observed between both sample wells, which are sparsely populated in contrast to the matted cell growth observed in the control well. However, as the differences in confluence levels were first observed at four days, further work must be conducted to study TTV VP3's localization at this time point.

One major difference between our observations for TTV VP3 and those localizations observed for apoptin is that our two-day transfections displayed cytoplasmic localization while apoptin displays aggregation in the nucleus prior to inducing apoptosis. Although the observed localization for TTV VP3 is different that observed for apoptin, the observed localization of the protein coded by PCV ORF3 also induces apoptosis like apoptin, but in the cytoplasm. [4] Perhaps the apoptotic activity of TTV VP3 is more similar to that PCV VP3; however, the discovery of any further similarity will require a more in depth investigation into potential mechanistic similarities.

Future investigations

With the foundation established by this initial research into TTV VP3, any number of routes of study lay open to future investigative endeavors. An important follow-up investigation may be first to conduct a more extensive search into—and to conclusively demonstrate—the apoptotic activity of TTV VP3. This may be accomplished through a more extensive study into the protein's localization—perhaps using our designed pVP3-GFP construct, but for more extended temporal periods post-transfection. Further studies using the pVP3-3xFlag construct we designed during our research may reveal present more evidence about TTV VP3's apoptotic ability. 3xFlag's unobtrusive size may allow for a greater understanding or approximation of TTV VP3's natural behavior when unencumbered by fusion with GFP. By western blotting 3xFlag-tagged VP3 using a native gel, it may be also possible to gain some approximation as to the true size and number of units composing each aggregated multimer..

In addition to learning about TTV VP3's potential apoptotic activity, there is much that can be learned about the role that ORF3's primary sequence plays in determining the dynamic behavior and localization of TTV VP3. By creating sectional constructs, as discussed in the previous section, it may be possible to analyze the role of specific sections in controlling TTV VP3. If the loss of a particular section eliminates or alters some aspect of its localization or behavior, then this may indicate the section's role in the protein's function. For instance, if a construct is made, minus one or two 20 nt oligonucleotides from the N-terminal region (replaced with a new oligonucleotide with appropriate start codons and restriction sites), and the protein is observed to no longer aggregate, then perhaps later studies could be focused upon the removed sequence's role in multimerization. Or similarly, if the protein is seen to exist

in both the nucleus and cytoplasm two days after transfection then perhaps the truncated section could play some role in signaling nuclear export.

Future work with TTV VP3 will likely involve altering the function of the protein through site-directed mutagenesis. With greater knowledge regarding the particular functions of TTV ORF 3's primary sequence, future investigators will be able to design primers and change the residues produced by specific codons through PCR. In this way, mutated TTV ORF 3 could be transfected and observed to study the general effects of the mutation, perhaps disturbing localization, activity, or aggregation.

Investigating the expression and behavior of TTV ORF3 proteins from other genotypes may also provide insight into the biological significance of these proteins. Perhaps if similar investigations to this one are conducted upon the ORF 3 of other TTV genotypes, researchers may gain a greater understanding of the degree of conserved activity for this protein across the myriad genotypes of TTV.

Conclusions

At this stage, much remains to be understood about TTV and the potential of its VP3. Years of study have done little to bring us closer to an understanding of TTV's pathology. In-depth studies of TTV's proteins and its biological behavior have been greatly impeded. Although researchers have been working to uncover the properties and behavior of TTV's proteins, we are still left in the dark regarding the role these proteins play in TTV infection and *in vivo* behavior.

Through our investigations, we have demonstrated that TTV VP3 shows dynamic similarities to apoptin from CAV, by forming aggregations in the cytoplasm. Although this cytoplasmic localization differs from CAV's apoptin, it shows similarities to the VP3 of porcine circoviruses which form cytoplasmic aggregates and have been shown to selectively trigger apoptosis akin to apoptin. [4] Finally, our work demonstrates that TTV VP3 is fully excluded from the nucleus at two days after transfection, indicating the presence of a NES in the structure of TTV VP3.

By expressing the protein from TTV ORF 3 in a careful and controlled manner we have provided a framework of understanding which future researchers can build upon and grow the understanding of TTV VP3. In the process of accomplishing this, we have demonstrated the overall effectiveness of PCR-directed gene synthesis in generating full-length gene constructs from a number of overlapping oligonucleotides. In addition, we have established a genetic toolbox of ready-to-study genes precloned into a number of expression vectors. Above all, we hope that through our work we were able to provide solid grounding for investigations to come and that our research and findings were engaging enough to perpetuate and propel future studies into TTV VP3.

Figures

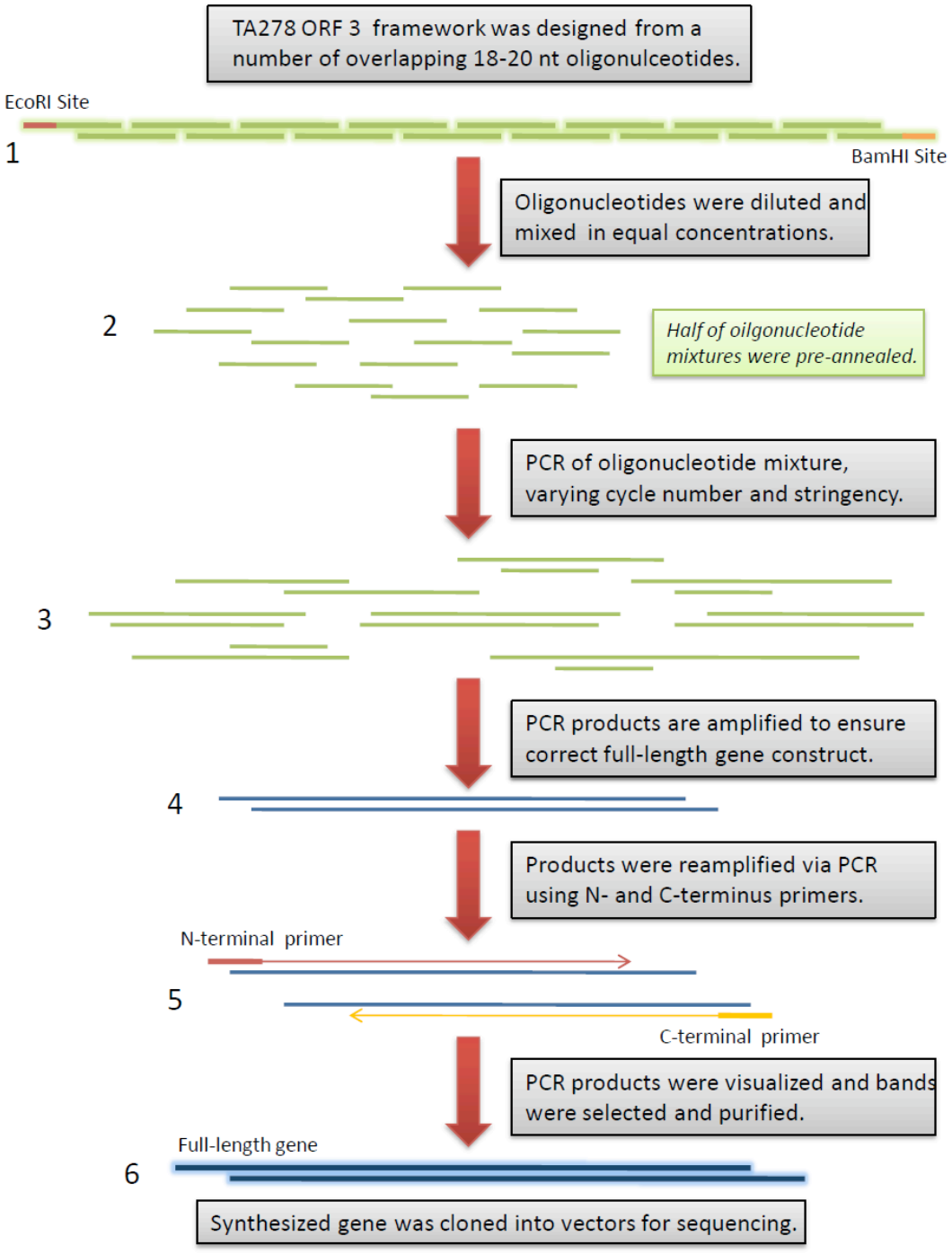

Figure 3: General visual overview of PCR-directed gene synthesis.

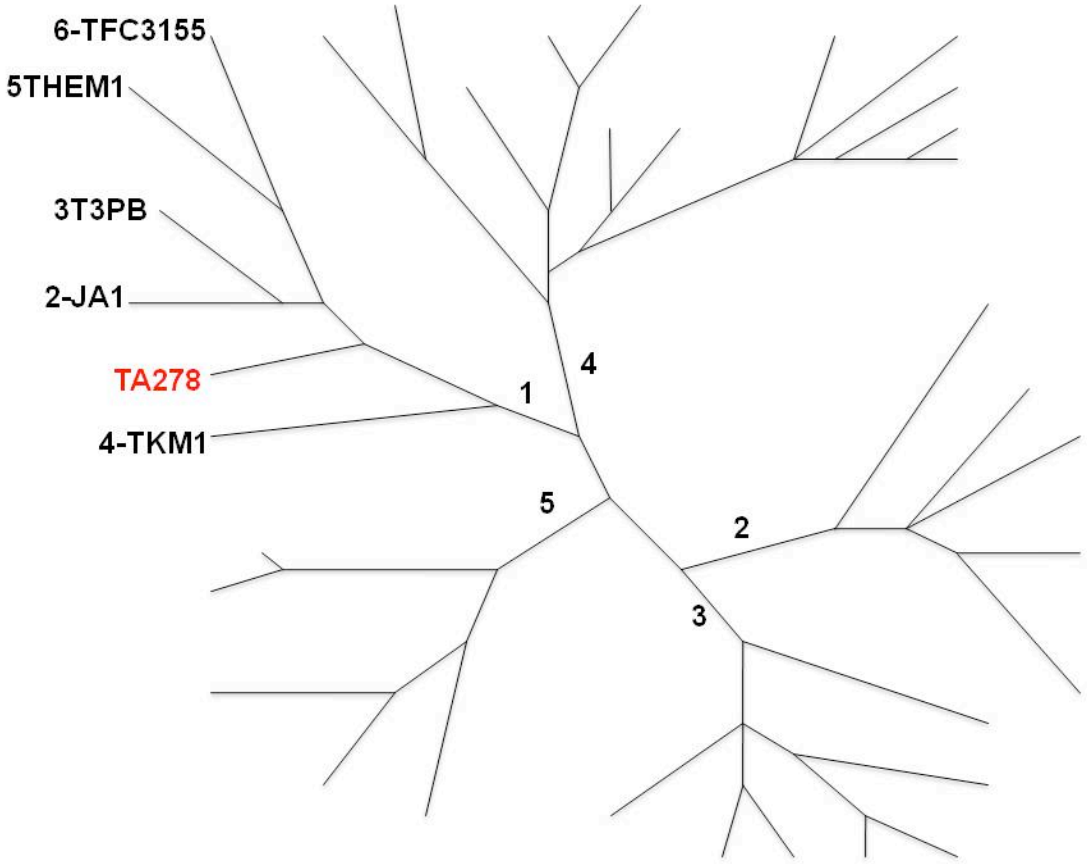

Figure 4: Cladogram representation of TTV group one in association with abbreviated phylogenic distribution, as adapted from Hino and Miyata. [9]

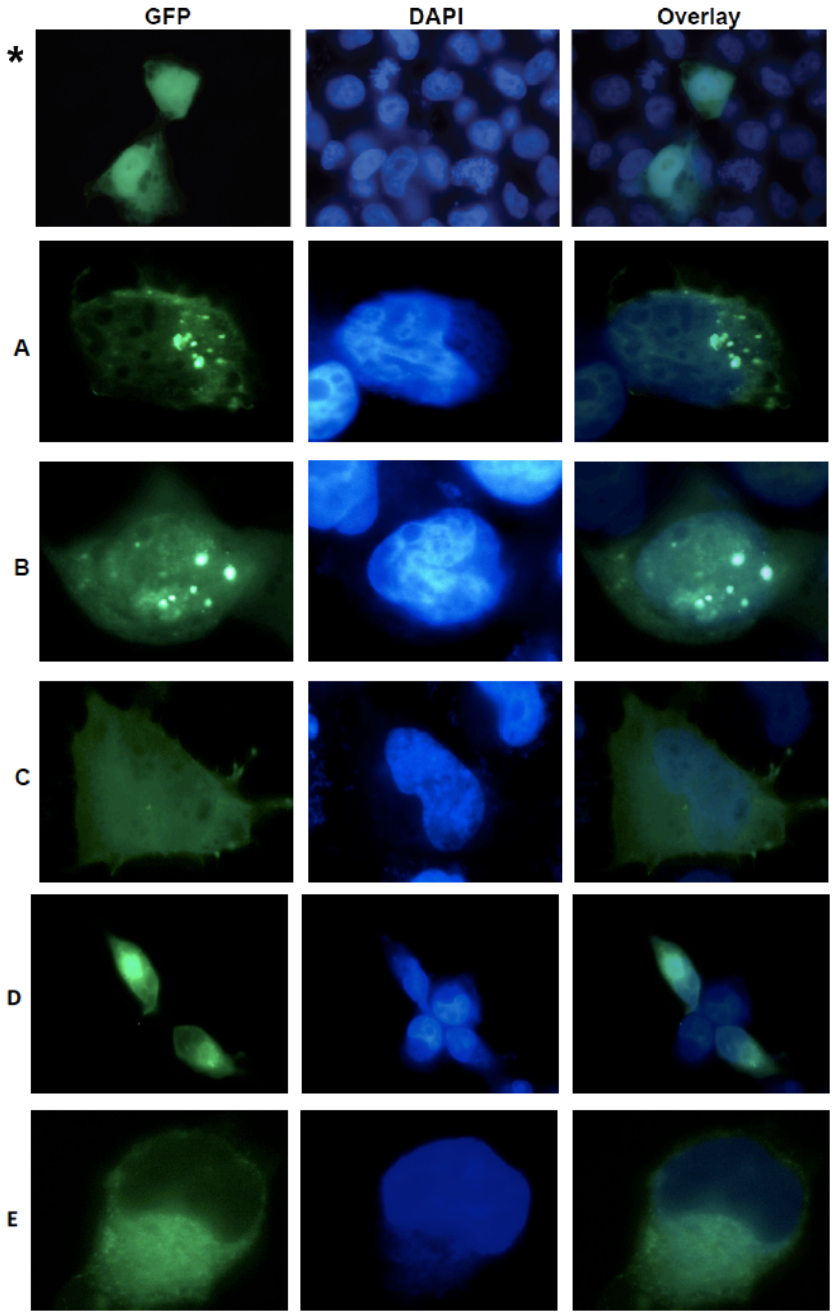

Figure 5: Fluorescence microscopy images of H1299 small-cell lung carcinoma cells transfected with construct pVP3-GFP. Cells transfected for one day are shown in panels A-C, while cells transfected for two days are shown in panels D,E. The first column displays fusion protein localization signals and the second shows the same image with filtered for the nuclear stain DAPI. The Third column shows an overlay of the two images with one image transposed upon the other to show both GFP signals and DAPI simultaneously. Control GFP cells shown in * panel.

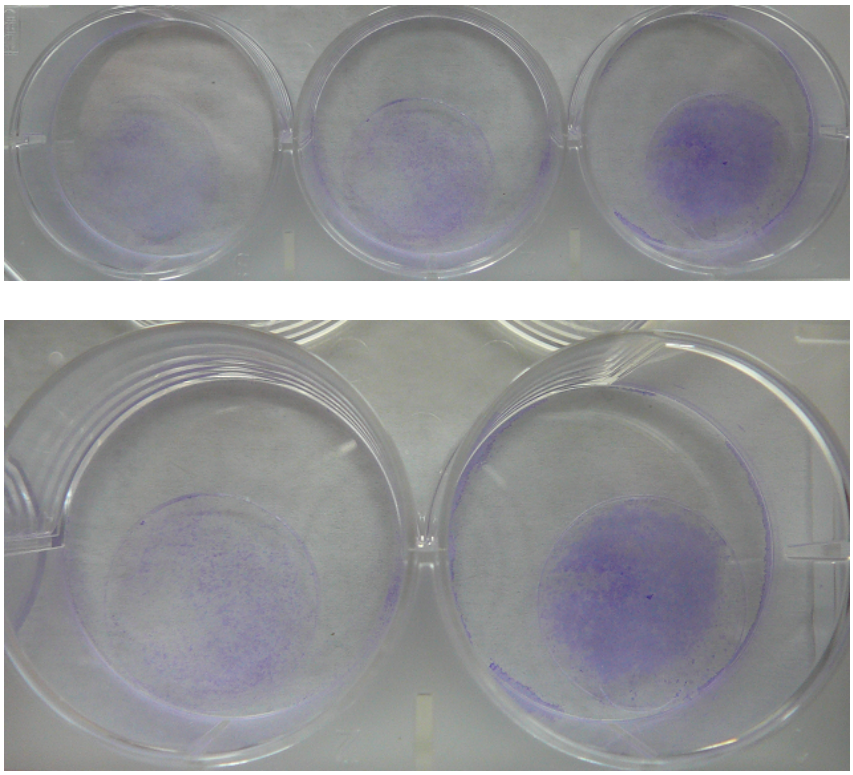

Figure 6: Crystal violet-stained H1299 cells transfected with pVP3-Flag construct, fixed at five days after transfection. Three wells were used for transfection and crystal violet staining (top), with the left and middle wells containing transfected H1299 cells, while the rightmost well contained non-transfected cells (control). Detail of the middle and rightmost wells are shown in the lower panel of the figure.

Works Cited

1. Abe, K., Inami, T., Asano, K., Miyoshi, C., Masaki, N., Hayashi, S., Ishikawa, I., Takebe, Y., Win, K. M., El-Zayadi A. R., Han K. H., Zhang D. Y., (1999). TT virus infection is widespread in the general populations from different geographic regions. J. Clin. Microbiol. 1999, 37:2703-2705.

2. Biagini, P., de Micco, P., de Lamballerie, X., (2006). Identification of a third member of the Anellovirus genus ('small anellovirus') in French blood donors. Arch. Virol. 151, 405–408.

3. Erker, J. C., Leary, T. P., Desai, S. M., Chalmers, M. L. & Mushahwar, I. K. (1999). Analyses of TT virus full-length genomic sequences. J. Gen. Virol. 80, 1743–1750.

4. Finsterbusch, T., Mankertz, A., Porcine circoviruses—Small but powerful. Virus Res. (2009).

5. Hallett, R. L., Clewley, J. P., Bobet, F., McKiernan, P. J. & Teo, C. G. (2000). Characterization of a highly divergent TT virus genome. J. Gen. Virol. 81, 2273–2279.

6. Herman, R., Hueurmann, K., Brizzard, B., (2009). 3xFlag: Ultra-sensitive detection of recombinant proteins. Sigma-Aldrich.

7. Hino S., (2006). TTV, a new human virus with single-stranded circular DNA genome. Rev Med Virol 12: 151–158.

8. Hino, S., Miyata, H., (2007). Torque teno virus (TTV): current status. Reviews in Medical Virology 17, 45–57.

9. Hino, S., Prasetyo, A. A. Relationship of Torque Teno Virus to Chicken Anemia Virus. (Book) Current Topics in Microbiology and Immunology, *TT Viruses: The Still Elusive Human Pathogens,* (eds.) de Villiers, E.-M., zur Hausen, H., 2009.

10. ICTVdB. The universal virus database of the International Committee on Taxonomy of Viruses. http://www.ictvdb.rothamsted.ac.uk/index.htm.

11. Jelcic, I., Hotz-Wagenblatt, A., Hunziker, A., zur Hausen, H., and de Villiers, E.-M., Isolation of Multiple TT Virus Genotypes from Spleen Biopsy Tissue from a Hodgkin's Disease Patient: Genome Reorganization and Diversity in the Hypervariable Region. J. Virol. 73, 1052-1064.

12. Kamahora, T., Hino, S., Miyata, H., (2000). Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. J. Virol. 74, 9980–9986.

13. Khudyakov, Y. E., Cong, M. E., Nichols, B., Reed, D., Dou, X. G., Viazov, S. O., Chang, J., Fried, M. W., Williams, I., (2000). Sequence heterogeneity of TT virus and closely related viruses. J. Virol. 74, 2990–3000.

14. Kooistra, K., Zhang, Y. H., Henriquez, N. V., Weiss, B., Mumberg, D., Noteborn, M. H., (2004). TT virus-derived apoptosis-inducing protein induces apoptosis preferentially in hepatocellular carcinoma-derived cells. J. Gen. Virol., 85: 1445–1450.

15. Leary TP, Erker JC, Chalmers ML, Desai SM, Mushahwar IK. Improved detection systems for TT virus reveal high prevalence in humans, non-human primates and farm animals. J Gen Virol 1999; 80: 2115–2120.

16. Miyata, H., Tsunoda, H., Kazi, A., Yamada, A., Khan, M. A., Murakami, J., Kamahora, T., Shiraki, K. & Hino, S. (1999). Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. J Virol 73, 3582–3586.

17. Naoumov, N. V., E. P. Petrova, M. G. Thomas, and R. Williams. 1998. Presence of a newly described human DNA virus (TTV) in patients with liver disease. Lancet 352:195–197.

18. NCI-Navy Medical Oncology Branch cell line supplement. J. Cell Biochem. Suppl. 1996; 24:1-291.

19. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem. Biophys. Res. Commun. 1997; 241: 92–97.

20. Noteborn MH. Chicken anemia virus induced apoptosis: underlying molecular mechanisms. Vet. Microbiol. 2004; 98: 89–94.

21. Okamoto H, Fukuda M, Tawara A, et al. Species specific TT viruses and cross-species infection in nonhuman primates. J. Virol. 2000; 74: 1132–1139.

22. Okamoto, H., Takahashi, M., Nishizawa, T., Tawara, A., Fukai, K., Muramatsu, U., Naito, Y. & Yoshikawa, A. (2002). Genomic characterization of TT viruses (TTVs) in pigs, cats and dogs and their relatedness with species-specific TTVs in primates and tupaias. J. Gen. Virol. 83, 1291–1297.

23. Okamoto, H., Nishizawa, T. & Takahashi, M. (2004). Torque teno virus (TTV): molecular virology and clinical implications. In Viral Hepatitis: Molecular Biology, Diagnosis, Epidemiology and Control, pp. 241–254. Edited by I. K. Mushahwar. California: Elsevier.

24. Pollicino T, Raffa G, Squadrito G, Costantino L, Cacciola I, Brancatelli S, Alafaci C, Florio MG, Raimondo G: TT virus has ubiquitous diffusion in human body tissues: analyses of paired serum and tissue samples. J Viral Hepat 2003, 10:95-102.

25. Qiu, J., Kakkola, L., Cheng, F., Ye, C., Soderlund-Venermo, M., Hedman, K., Pintel, D.J., 2005. Human circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. Journal of Virology 79, 6505–6510.

26. Simmonds P, Prescott LE, Logue C, Davidson F, Thomas AE, Ludlam CA: TT virus—part of the normal human flora? J. Infect. Dis. 1999, 180:1748-1750.

27. de Smit M. H., Noteborn, M. H. M., Apoptosis-Inducing Proteins in Chicken Anemia Virus and TT Virus. (Book) Current Topics in Microbiology and Immunology, TT Viruses: The Still Elusive Human Pathogens, (eds.) de Villiers, E.-M., zur Hausen, H., 2009.

28. Stremmer, Willem P. C., Crameri, Andreas, Kim D. Ha, Thomas M. Brennan and Herbert L. Heyneker. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene, 1995; 164:49-53.

29. Takahashi, K., Iwasa, Y., Hijikata, M., Mishiro, S., (2000). Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. Arch. Virol. 145: 979–993.

30. Zhong S, Yeo W, Tang MW, Lin XR, Mo F, Ho WM, Hui P, Johnson PJ. 2001. Gross elevation of TT virus genome load in the peripheral blood mononuclear cells of cancer patients. Ann. N Y Acad. Sci. 945-56.