

**HEPATITIS *DELTA* VIRUS REPLICATION AFFECTS THE EXPRESSION OF HOST
GENES INVOLVED IN CELL CYCLE**

Gabrielle Goodrum

THESIS SUBMITTED TO THE FACULTY OF GRADUATE AND POSTDOCTORAL
STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN BIOCHEMISTRY

Thesis Supervisor: Dr. Martin Pelchat

Department of Biochemistry, Microbiology, and Immunology
Faculty of Medicine
University of Ottawa

© Gabrielle Goodrum, Ottawa, Canada, 2019

ABSTRACT

The hepatitis *delta* virus (HDV) is the smallest human pathogenic RNA virus and relies heavily on host proteins for its replication. The objective of my research was to observe the effect of HDV replication on host gene expression, using a HEK-293-based cell system engineered to mimic HDV replication. A high-throughput sequencing was performed and allowed to establish a total of 3,561 genes differentially expressed by HDV RNA. Among those genes, 3,278 were upregulated by HDV RNA and 283 downregulated. A Gene Ontology (GO) enrichment analysis was performed on those dysregulated genes and revealed that upregulated genes were predominantly part of these four pathways: RNA processing, G-protein coupled receptor signaling pathway, protein transport, and organelle organization. On the other hand, downregulated genes were part of the nucleosome assembly pathway. The expression of several genes was confirmed by RT-qPCR. Moreover, protein complexes whose expression at the gene level was affected were identified. A total of 30 complexes were found to be significantly affected by HDV replication. Among them, we found many chromatin and histone related complexes. Lastly, a flow cytometry analysis revealed an increase in cell cycle arrest in G0/G1 and a reduction in the percentage of cell in S phase. Moreover, there was a difference in cell size for arrested cells in G0/G1 in HDV replicating cells. Overall, my results support the hypothesis that HDV replication induces cell cycle dysregulation.

ACKNOWLEDGMENTS

I would like to thank all the people who have supported and inspired me during my master. First, I would like to thank Dr. Martin Pelchat for this amazing opportunity, for his guidance and support along with this project. I thank him for the help he has given me with my bioinformatics analysis and for the confidence he had in me as and the freedom he gave me in this project. I would also like to say thank you to Lynda Rocheleau for carrying out many data processing steps. Thanks to Marc-André and Vera for their expert advice for the design of my flow cytometry experiments. I would also like to thank my thesis advisory committee Dr. Tommy Alain and Dr. Marc-André Langlois for their advice and support.

I would like to thank Dorota Sikora, James Butcher and Patrick Taylor for being my mentors during my masters, for having the patience and the kindness to answer my questions and to have taught me so much. Also, thank you to all my friends from the Stintzi and the Fullerton laboratory for being there when I needed it.

I would also like to thank Kevin who was my moral support during all my studies, which encouraged me and believed in me. For his patience to listen to my oral presentations, again and again. Thank you for teaching me so many informatics basic concepts. Finally, I would like to thank my parents for their constant support during my studies, to push me to surpass myself and for believing in me.

I am also grateful for the McGill University and Génome Québec Innovation Center, as well as Dr. John Taylor (Fox Chase Cancer Center, USA) for kindly providing the cell lines. This study was supported by a NSERC discovery grant.

TABLE OF CONTENT

Table of Contents

ABSTRACT	II
ACKNOWLEDGMENTS.....	III
LIST OF ACRONYMS.....	VI
LIST OF FIGURES.....	VIII
LIST OF TABLES.....	IX
CHAPTER 1: INTRODUCTION	1
1.1 DISCOVERY AND CLASSIFICATION	2
1.2 EPIDEMIOLOGY	4
1.3 PATHOGENICITY	5
1.4 VIRION AND GENOME STRUCTURE.....	6
1.5 SYMMETRICAL ROLLING CIRCLE REPLICATION	13
1.6 <i>DELTA</i> ANTIGENS	17
1.6.1 <i>Function of HDag-S</i>	20
1.6.2 <i>Function of HDag-L</i>	21
1.7 POST-TRANSLATIONAL MODIFICATIONS OF HDAGS.....	23
1.8 INTERACTION OF HDV WITH HOST CELLULAR PROTEINS	25
1.8.1 <i>HDag interactions with host proteins</i>	25
1.8.2 <i>HDV RNA interaction with host proteins</i>	27
1.9 MODELS DEVELOPED TO STUDY HDV REPLICATION	32
1.10 HOW HDV AFFECTS ITS HOST CELL.....	34
1.10.1 <i>Proteomic analyses</i>	34
1.10.2 <i>Interferon response</i>	37
1.10.2 <i>Stress and foci formation</i>	40
1.11 RATIONALE, HYPOTHESIS, AND OBJECTIVES.....	42
CHAPTER 2: MATERIALS AND METHODS	43
CELL CULTURE AND INDUCTION.....	44
RNA ISOLATION, LIBRARY CONSTRUCTION AND ILLUMINA DEEP-SEQUENCING.....	44
DIFFERENTIAL EXPRESSION ANALYSIS OF TRANSCRIPT AND DATA PROCESSING.....	45
GENE ONTOLOGY ENRICHMENT	46
IDENTIFICATION OF AFFECTED COMPLEXES.....	46
REVERSE TRANSCRIPTION AND QUANTITATIVE POLYMERASE CHAIN REACTION (qPCR).....	47
WESTERN BLOT ASSAY	48
FLOW CYTOMETRY ANALYSIS	49
<i>Sample preparation</i>	49
<i>Data processing: gating and rescaling</i>	50
<i>Data analysis</i>	51
CHAPTER 3: RESULTS	53
3.1 TRANSCRIPTOME ANALYSIS OF GENES AFFECTED BY HDV	54
3.1.1 <i>Sample preparation and data processing</i>	54
3.1.2 <i>Gene expression analysis</i>	63
3.2 ENRICHMENT OF GENE ONTOLOGY TERM AND BIOLOGICAL PROCESSES AFFECTED BY HDV REPLICATION	71
3.3 WEIGHTED CORRELATION NETWORK ANALYSIS (WGCNA).....	76
3.4 HDV REPLICATION DISRUPTS MAJOR PROTEIN COMPLEXES	82

3.5 FLOW CYTOMETRY REVEALS AN ARREST OF THE CELL CYCLE AND MORPHOLOGICAL DIFFERENCES FOR HDV INDUCED CELLS	89
CHAPTER 4: DISCUSSION	102
4.1 SUMMARY OF FINDINGS	103
HDV REPLICATION AFFECT THE HOST TRANSCRIPTOMIC LANDSCAPE.....	106
HDV REPLICATION PERTURB MAJOR PROTEIN COMPLEXES.....	112
FLOW CYTOMETRY REVEALS AN ARREST OF THE CELL CYCLE AND MORPHOLOGICAL DIFFERENCES FOR HDV INDUCED CELLS.....	118
FUTURE DIRECTIONS	121
CHAPTER 5: CONCLUSION	123
MODEL OF HDV REPLICATION EFFECT IN HOST CELL	125
REFERENCES	127
CONTRIBUTION OF COLLABORATORS	144
SUPPLEMENTARY	145

LIST OF ACRONYMS

Acronym	Definition
293-Ag	293 cells expressing HDAg-S
293-HDV	293 cells allowing the replication of HDV
ADAR	Adenosine deaminase acting on ribonucleic acid
CKII	casein kinase II
CORUM	Comprehensive resource of mammalian protein complexes
CTD	C-terminal domain
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
ER	Endoplasmic reticulum
ERK	Extracellular signal-related kinases
FMO	Fluorescence mines one
FSC	Forward scatter
Fwd	Forward
GAPDH	Glyceraldehyde-3-phosphatedehydrogenase
GFOLD	Generalized fold change
GO	Gene ontology
GPCR	G-protein coupled receptor
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B Virus
HCC	Hepatocellular carcinoma
HCV	Hepatitis C virus
HDAg	Hepatitis <i>delta</i> antigen
HDAg-L	Hepatitis <i>delta</i> Antigen-Large
HDAg-S	Hepatitis <i>delta</i> Antigen-Small
HDRNP	HDV Ribonucleoproteins
HDV	Hepatitis <i>delta</i> Virus
HRP	Horseradish Peroxidase
HUH-7	Human hepatocellular carcinoma cell
INF	Interferon

NES	Nuclear export signal
NESI	NES-interacting protein
NIS	Nuclear import signal
NTCP	Sodium-taurocholate co-transporter polypeptide
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PI	Propidium iodide
PKR	Protein kinase R
PSF	Polypyrimidine tract-binding protein-associated splicing factor
PVDF	Polyvinylidene difluoride
qPCR	Quantitative Polymerase Chain Reaction
Rev	Reverse
RIP	RNA Immunoprecipitation
RNA	Ribonucleic Acid
RNAP	Host DNA-dependant RNA polymerase
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein
RPKM	Reads Per Kilobase Million
RT	Reverse Transcription
RT-qPCR	Real-time quantitative PCR
SSC	Side scatter
SVP	Subviral particule
TBS	Tris-buffered saline
TBST	Tris-buffered saline with tween 20
TET	Tetracycline
Trp	Tryptophan
WGCNA	Weighted Correlation Network Analysis
WHO	World Health Organisation

LIST OF FIGURES

FIGURE 1.1 HEPATITIS <i>DELTA</i> VIRUS GENOMIC RNA, ANTIGENOMIC RNA AND MRNA.....	12
FIGURE 1.2. SYMMETRICAL ROLLING CIRCLE MODEL OF HDV REPLICATION	16
FIGURE 1.3. EDITING OF ANTIGENOMIC HDV RNA BY ADAR1.....	19
FIGURE 3.1. OVERVIEW OF THE SIX DIFFERENT CONDITIONS USED FOR SUBSEQUENT EXPERIMENTS AND SEQUENCING.....	58
FIGURE 3.2. VALIDATION OF HDV INDUCTION AND RNA INTEGRITY.	60
FIGURE 3.3. RNA QUALITY, CONCENTRATION, AND INTEGRITY DATA AS WELL AS THE NUMBER OF BASES SEQUENCES AND THE NUMBER OF SEQUENCES READS OBTAINED BY ILLUMINA SEQUENCING.....	62
FIGURES 3.4. CLASSIFICATION OF GENES ACCORDING TO THEIR EXPRESSION CHANGE BETWEEN DIFFERENT HEK-293, 293-AG AND 293-HDV CONDITIONS	67
FIGURE 3.5. VALIDATION OF RNASEQ IDENTIFIED GENE EXPRESSION CHANGES	69
FIGURE 3.6 BIOLOGICAL PROCESS INTERACTION NETWORK OF ALL AFFECTED PATHWAYS DYSREGULATED BY HDV ACCUMULATION	75
FIGURE 3.7. SIGNED WEIGHTED GENE CORRELATION NETWORK ANALYSIS (WGCNA) OF GENES AFFECTED BY HDV INDUCTION.....	79
FIGURE 3.8. UNSIGNED WEIGHTED GENE CORRELATION NETWORK ANALYSIS (WGCNA) OF GENES AFFECTED WITH HDV INDUCTION.	81
FIGURE 3.9. LIST OF COMPLEXES AFFECTED BY HDV REPLICATION	86
FIGURE 3.10. LIST OF COMPLEXES AFFECTED BY HDV REPLICATION USING A SIGNED ANALYSIS.....	88
FIGURE 3.11. FLOW CYTOMETRY ANALYSIS OF CELL CYCLE PHASE. A-B) FLOW CYTOMETRY USING	93
FIGURE 3.12. CHANGES IN THE CELL DISTRIBUTION OF DIFFERENT PHASES OF THE CELL CYCLE BETWEEN HDV AND HDV TET AFTER 12, 24 AND 36 HOURS.....	95
FIGURE 3.13. FLOW CYTOMETRY ANALYSIS OF 293-HDV CELLS INDUCED AND 293-HDV CELLS NON INDUCED AFTER 24H	97
FIGURE 3.14. DENSITY PLOT OF CELLULAR AGGREGATE IN HEK-293, 293-AG AND HDV CELLS WITHOUT AND WITH TETRACYCLINE INDUCTION AT DIFFERENT TIME POINT.....	99
FIGURE 3.15 PHASE CONTRAST COMPARISON OF NON-INDUCED AND INDUCED HEK-293, 293-AG AND 293-HDV CELLS	101
FIGURE 5.1. WORKING MODEL OF THE EFFECT OF HDV REPLICATION IN HOST CELL. ...	126
FIGURE S1. VALIDATION OF DESIGNED PRIMERS USING PCR.	149
FIGURE S2. SCRIPT FOR GENE CLASSIFICATION ACCORDING TO EXPRESSION CHANGE	150
FIGURE S3. SCRIPT FOR PROTEIN COMPLEX ANALYSIS USING A SIGNED AND UNSIGNED APPROACH	154
FIGURE S4. DETAILED LIST OF AFFECTED OF UNSIGNED COMPLEXES WITH EACH COMPONENT RNA EXPRESSION.....	159
FIGURE S5. SCRIPT FOR FLOW CYTOMETRY DATA PROCESSING INCLUDING GATING AND RESCALING AND DATA ANALYSIS.	163

LIST OF TABLES

TABLE 1.1. HOST PROTEIN INTERACTING WITH HDV RNA AND/OR HDAGS.....	29
TABLE 1.2. LIST OF HOST PROTEIN DIFFERENTIALLY EXPRESSED IN THE PRESENCE OF HDV RNA AND/OR HDAGS.....	37
TABLE 2.1. PRIMERS SEQUENCES USED FOR RT-QPCR ANALYSIS.	51
TABLE 3.1. NUMBER OF READ ABLE TO MAP TO HDV FOR EACH CELL TYPE IN THE RNA- SEQ DATA.....	63
TABLE 3.2. RNA-SEQ AND QPCR VALUES OF GENES SELECTED FOR GENE EXPRESSION VALIDATION IN HDV REPLICATING CELLS.....	70
TABLE 3.3 PERCENTAGE OF CELLULAR AGGREGATES IN HEK-293, 293-AG AND HDV CELLS WITHOUT AND WITH TETRACYCLINE INDUCTION AT DIFFERENT TIME POINT.....	91

Chapter 1: INTRODUCTION

1.1 Discovery and classification

Hepatitis *delta* virus (HDV) was discovered in 1977 by *Rizzetto et al* who identified a novel antigen present among patients positive for the hepatitis *beta* virus (HBV) (1). This new antigen was found exclusively in the nuclei of hepatocytes derived from patient with chronic hepatitis B surface antigen (HBsAg). Originally considered a variant of the HBV nucleocapsid, a subsequent study showed that it was distinct from HBV and it was identified as a new hepatitis agent. This antigen was termed the *delta* antigen (HDAg or δ) and was later associated with HDV. This unique pathogen was classified in its own genus, the *Deltavirus* (2). To this day, a total of 8 major clades have been classified among the *Deltavirus* genus, with a sequence divergence associated with geographic origin, which can vary by as much as 40% in sequence identity (3, 4).

HDV is a defective and satellite virus of HBV, requiring HBV envelope proteins (HBsAg) for its propagation. Despite being a satellite virus of HBV, HDV shares no sequence similarity with its helper virus and replicates independently from it. Hepatitis *delta* virus is unique among animal viruses. Its single-stranded circular RNA genome adopts a rod-like secondary structure, which is similar to viroids, which are plant pathogens. Additionally, HDV and viroids share a common “rolling circle” mechanism for their replication. HDV replication occurs in the nucleus similar to viroids of the family *Pospiviroidae*. Considering their small size and their limited protein-coding capacity, HDV and viroids rely heavily on host proteins and redirect host components to achieve their life cycle. For instance, HDV and viroids both redirect host DNA-dependant RNA polymerase II (RNAP II) from their normal host DNA templates for their replication. HDV and viroids from the *Avsunviroidae* family also share the

use of ribozyme domains to perform auto-cleavage reactions. In terms of size, the HDV genome is much smaller than other mammalian viruses. With a genome of 1.7kb, HDV is considerably smaller than other human RNA pathogens (influenza A virus is about 14kb) but yet bigger than viroids (~250-400nt) (5). Although they share many common features, viroids do not encode for proteins and do not require a helper virus, where HDV encodes for one protein existing as two isoforms (6), the *delta* antigens, and needs HBV for propagation (6). With all these similarities between HDV and viroids, they have been proposed to be evolutionarily related (7).

A recent phylogenetic analysis of different full-length HDV sequences confirmed a total of eight HDV genotypes (8). HDV genotype distributions were mapped and revealed a geographical distribution profile where genotype 1 was worldwide distributed, genotype 2 and 4 primarily in Asia, while genotype 3 and 5-8 were geographically localized in South America and in Africa respectively. Each genotype could subsequently be classified into subtypes. The sequence similarity and genetic distances between the different classes exhibit a further distance with genotype 3 while genotype 1 showed more divergence among its own subtype. Among all the HDV genotypes found globally, the genotype 1 is the most frequent and the most widespread around the world (9) and has been associated with more severe outcome than genotype 2 (10). The sequences of the different genotypes can largely vary, and this influences some aspects of the virus virulence such as its packaging efficacy and can influence the disease outcome (10, 11). Genotype 3 has been associated with a more severe form of liver disease, rapidly progressing to fulminant hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC) (12, 13).

1.2 Epidemiology

HDV infection has been reported worldwide but is predominantly found in underdeveloped countries that do not have easy access to HBV vaccination and lack awareness on the risks of infectious disease transmission. The most affected areas are Central and West Africa, Asia, Pacific islands, the Middle East, Eastern Europe South America, and Greenland (14). An estimate of 15 to 20 million people are co-infected or superinfected with HDV, which represents roughly 5% of people affected with HBV worldwide. Since the implementation of HBV vaccination in 1980, the percentage of HDV prevalence went from more than 20% to 5-10% and varies considerably depending on the country (9). The amount of people affected by HDV does not necessarily correlate with the prevalence of HBV (15). The World Health Organisation (WHO) declared the creation of World Hepatitis Day, on July 28th, in 28 countries and worldwide efforts have made drastic steps towards elimination of hepatitis by increasing access to therapy, increasing vaccination and emphasizing the importance of safe injection. Several countries do not report or do not systematically test for HDV specific antibodies in HBV infected patients, causing an incomplete global estimation of the number of people affected (16).

Like HBV, HDV transmission occurs via contact with infectious body fluid. Transmission routes include sexual transmission and the use of contaminated syringes. HDV can also be transmitted from mother to child (17). Although HBV perinatal transmission is well documented, some studies suggest that perinatal transmission of HDV is uncommon (18). HBV transmission can occur through perinatal transmission (through the placenta), vertical transmission (mother to child) or horizontal transmission (infected household to child) (19). Interestingly, there was a case of a mother carrying both viruses and although the detection of anti-HDV antibody was detected at birth, it disappeared after 3 - 4 months suggesting false

positive results originating from the mother (20). As has been noted, there is some evidence showing that a coinfecting mother with HBV-HDV could pass on HBV without transmitting HDV suggesting that mono-transmission of HBV from a co-infected person is possible (21).

HDV infection in combination with HBV or hepatitis C virus (HCV) results in a decrease in HBV/HCV replication although there were no differences in HBsAg levels between HBV mono-infected and HBV-HDV co-infected patients (22). It is estimated that 1/3 of HDV patients in Central Europe are co-infected with HCV with no sign of additional perturbation compared to HBV-HDV co-infected patients. However, co-infection with HBV, HCV, and HDV has been linked a predisposition to advanced liver disease and higher rates of cirrhosis (22, 23).

1.3 Pathogenicity

HDV pathogenicity is still poorly understood and even though some treatments are available on the market, long-term antiviral responses are rare, and relapse is often observed during long-term follow-up. Among these treatments, interferon-alpha ($\text{INF}\alpha$), pegylated interferon alpha (PEG- $\text{INF}\alpha$) and a combination of PEG- $\text{INF}\alpha$ with adefovir are the most used therapies for HDV infected patients despite their low success rates (16, 24, 25). Interferon production rises upon HDV replication, however, a recent study reported that this rise in INF does not lead to HDV eradication (26). Although active HDV replication induces interferon beta ($\text{INF}\beta$) and interferon lambda ($\text{INF}\lambda 1/2/3$) production, the innate immune response is significantly reduced 7 days post infection. The interferon response displayed only limited inhibition of less than 50% of HDV replication and only at an early infection stage. These results showed that IFN therapy is not significantly effective to eradicate HDV infection. HDV

was suggested to become resistant to IFN treatment after the infection is settled (5-11 days). Thus, the only curative treatment option for HDV infected patients remains liver transplantation, although this technique remains susceptible to the risk of viral recurrence (9, 27, 28).

Among the 20 million peoples co-infected or super-infected with HDV, 80% leads to cirrhosis and/or hepatocellular carcinoma (HCC) (29). The HDV genotype was shown to have an important relationship with pathogenesis as well as the treatment response and patient outcome (29).

1.4 Virion and genome structure

The HDV virion is composed of two main components: the envelope and the ribonucleoprotein (RNP). The envelope is made of HBV envelope proteins and has an average diameter of 36nm. It is composed of three viral proteins: the small (S-HBsAg), the medium (M-HBsAg) and the large (L-HBsAg) surface antigens, which are integrated in a lipid bilayer originating from the host cell (30). Those three proteins are produced from a single long open reading frame (ORF) using three different start codons and sharing one stop codon. This ORF contains three regions: Pre-S1, pre-S2 and S. The small antigen is encoded by the S region, which is also found at the C-terminal of the medium and large form of the protein. The medium antigen has, in addition, the pre-S2 region at its N-terminal while the large antigen possesses all three regions: preS1, pre-S2, and S (31, 32). These envelope proteins are subject to post-translational modifications, including N-glycosylation, as well as other modifications occurring on different domains that confer essential functions (33).

The S-HBsAg is required to assemble the envelope protein while L-HBsAg is needed for infectivity. Both S-HBsAg and M-HBsAg were shown to be essential for virion morphogenesis (34, 35). The HBV virion envelope is predominantly composed of the small protein, but envelopes composed exclusively of S-HBsAg were shown to lack the capacity to bind to primary hepatocytes. Indeed, L-HBsAg is required in small quantities, where it plays a crucial role for virion entry into hepatocytes, by mediating receptor binding and allowing virion release and infectivity (36, 37). However, overexpression of L-HBsAg was shown to inhibit subviral particles release by retaining them bound to the ER (36). The pre-S1 region of L-HBsAg is crucial to the interaction with the sodium-taurocholate co-transporter polypeptide (NTCP), which was found to be the entry receptor used by both HBV and HDV (34, 35). This receptor is predominantly expressed in the liver and maintains bile salt homeostasis (34). Because HDV hijacks HBV envelope proteins, they both use the same mechanism of infectivity and therefore they share a common tropism to human hepatocytes.

HBV envelope proteins can self assemble and can independently be secreted as empty subviral particles (SVPs) deprived of viral components that are consequently non-infectious. These SVPs can be produced in extensive quantities, about 10^3 to 10^6 fold more than mature virion particles (38). HDV, which is produced in higher concentrations than HBV (10^{10} to 10^{11} virions per mL for HDV and 10^8 to 10^9 for HBV), takes advantage of this overproduction for its own encapsidation (39).

The HDV RNP consists of a mixture of HDAg (HDAg-S and HDAg-L) in association with HDV genomic RNA in an approximate ratio of 200:1 (40, 41). This ribonucleoprotein complex is packaged in HBV SVPs to form a fully mature virion. Only the genomic RNA is

packaged into the HDV virion. An infected liver accumulates around 300,000 copies of genomic strand HDV RNA per average cell. The genomic strand RNA is 5 to 22 times more abundant than the antigenomic strand (50 000/cell) and 500 times more abundant than HDV mRNA transcripts (1 000/cell) (**Figure 1.1**) (40, 42, 43).

HDV genomic RNA is exported to the cytoplasm shortly after its synthesis although a considerable proportion remains in the nucleus (44). HDV ribonucleoproteins were also shown to shuttle between the nucleus and cytoplasm compartment (45). The RNP export is mediated through HDAg-L nuclear export signal (NES) located at its C-terminal. Although HDAg-L is essential for RNP export, it was shown not to be essential for the genomic RNA cytoplasm export. Antigenomic HDV RNA almost exclusively remains in the nucleus although some traces were found in the cytoplasmic compartment early in the infection (44).

HDV is the smallest known human RNA virus. Its genome consists of a small (~1,680 nucleotides) single-stranded, circular RNA molecule and folds into an unbranched, rod-like structure due to 74% self-complementarity (**Figure 1.1**) (46–48). This structure confers the genome some stability from degradation by exonucleases and the DICER endonuclease by preventing their easy access to cleavage sites (6). The HDV genome contains a single open reading frame (ORF), located on the antigenomic RNA strand, coding for two viral proteins: the small and the large antigen (HDAg-S and HDAg-L). These two proteins are mostly identical in sequence except that HDAg-L (214 amino acids) contains an additional 19-20 amino acids at its C-terminus resulting from RNA editing of antigenomic HDV RNA at a location corresponding to the termination codon of the HDAg-S (195 amino acids) gene by the host adenosine deaminase that acts on RNA (ADAR-1) (49, 50). Despite being mostly

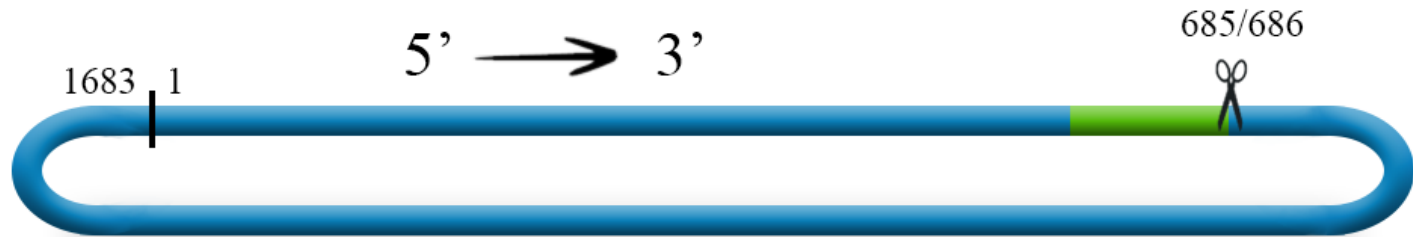
identical in sequence, these two proteins have distinct functions: HDAg-S is essential for HDV accumulation, while HDAg-L is necessary for virion assembly (51).

Both genome and antigenome RNAs possess a self-cleaving ribozyme domain essential for virus replication. Present at position 685 for the genome molecule and at the position 901 for the antigenomic strand, those motifs allow the nascent multimeric linear transcript created by the replication steps to be cleaved into unit-length species (52). This post-transcriptional step is necessary for HDV replication. Those ribozymes are composed of short domains of approximately 85 nucleotides and catalyze a magnesium-dependant trans-esterification reaction 33 nucleotides past the poly(A) site (53).

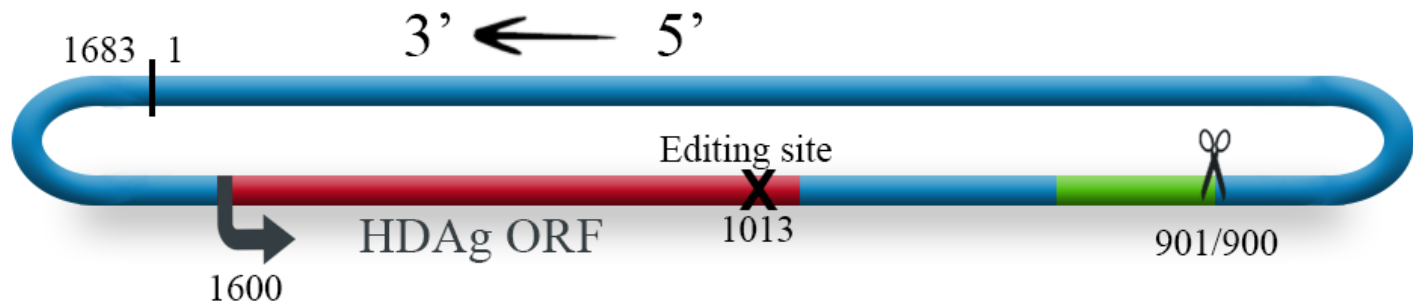
The exact replication site of HDV RNA and the cellular localization of its antigens remains controversial, and it is influenced by the system used to replicate/study HDV replication, by the active or inactive state of replication and by the seeding of the cells (45, 54–57). The genomic and antigenomic strands of HDV were found to localize in different nuclear compartments (58). The antigenomic RNA is located in the nucleolus while the genomic RNA is distributed more diffusely in the nucleoplasm and localizes with PML (promyelocytic leukemia protein) bodies during genome transcription. HDV genomic RNA is exported soon after its synthesis into the cytoplasm and it is found in approximately equal amounts in both the cytoplasm and nucleus, while the antigenomic strand remains exclusively in the nucleus (44). It is now established that HDV genomic RNA is continuously exported into the cytoplasm (44). The HDV small and large antigens location are also subject to many conflicting studies, some suggesting localization in the nucleolus, others showed a nucleoplasm location while others demonstrating both. Several studies have reported that the HDAGs were generally found in the nucleolus. HDAGs also colocalize with RNAP II, PML

and SL1(58). However, in the absence of the HDV genome, the HDAGs are restrained to the nucleus. In contrast, HDV RNA in the absence of HDAGs remains exclusively cytoplasmic, which most likely results from the absence of the NLS found on the HDAG-S which normally carries out nuclear transport (45).

Genome
(300 000/cell)



Antigenome
(50 000/cell)



mRNA
(1 000/cell)



Figure 1.1 Hepatitis *delta* virus genomic RNA, antigenomic RNA and mRNA. The *delta* ribozymes are represented by the green boxes on both strands and the cleavage sites are displayed by the scissor symbols. The HDAG ORF is represented on the antigenome strand by the red box, the arrow at position 1630 marks the transcription start site and the X present in the ORF box represents the ADAR1 editing site located at the stop codon for HDAG-S.

1.5 Symmetrical rolling circle replication

HDV does not encode its own replicase and there are no known HDV DNA intermediates (59, 60). Accordingly, at least one host RNA polymerase (RNAP), traditionally thought to accept only DNA as templates, is involved in the replication and transcription of HDV. HDV genome replication and transcription takes place in the nucleus. HDAg mRNA is post-transcriptionally processed with a 5'-cap and a 3'-poly(A) tail, which are typical features of transcripts generated by RNAP II (61, 62). HDV RNA accumulation in cultured cells is also sensitive to low doses of α -amanitin, an RNAP II inhibitor (63). RNAP II associates with HDV RNA, both in infected cells and *in vitro*, and forms an active pre-initiation complex on HDV RNA, similar to what is observed on DNA promoters (64, 65). Additionally, because antigenomic RNA synthesis is more resistant to α -amanitin, the involvement of at least one other host RNAP in HDV replication has been proposed (58, 66, 67). Consistent with this hypothesis, RNAP I and RNAP III can associate with both polarities of HDV RNA and RNAP I colocalizes with HDV antigenomic RNA (58, 68). Moreover, the RNAP I-specific transcription factor SL1 was coimmunoprecipitated with HDAg and antibody against SL1 resulted in a 80% reduction of antigenomic RNA synthesis (58). Although the role of RNAP II in both HDV replication and transcription is well established, there is still some controversy for the involvement of RNAP I and III.

Even though it requires the HBV envelope proteins for virion production, HDV replicates independently of HBV and relies entirely on its host for replication, as HDV RNA genome replication and RNP assembly were shown to be carried in absence of HBV (69). HDV replicates via a symmetrical, rolling circle mechanism (**Figure 1.2**). Replication of the

infectious circular monomer of genomic RNA produces linear, multimeric strands, which are cleaved to monomers by endogenous ribozymes and ligated by a yet unidentified host enzyme, yielding antigenomic circular monomers. These antigenomic molecules then serve as templates for genomic RNA using the same steps, or serve as templates for the synthesis of the HDAG mRNA (70, 71).



Figure 1.2. Symmetrical rolling circle model of HDV replication.

HDV genomic strand (blue) is used as a template to generate the antigenomic RNA using host polymerase to generate a multimeric linear strand. This strand is self-cleaved by endogenous *delta* ribozymes present on both genomic and antigenomic strands (not shown) and ligated by host enzyme to form the antigenomic monomeric strand (pink). The same steps are repeated on the antigenomic strands to produce a newly synthesized genomic strand. The genomic template is also used for the antigen mRNA synthesis. The cleavage sites are displayed by the scissor symbols.

1.6 Delta antigens

An immunoblotting assay only detected two proteins of 24 and 27 kDa in the serum of patients with chronic HDV, these were later termed the small and large antigens (HDAg-S and HDAg-L) (72). *Weiner et al.* (1988) determined that these two proteins were encoded within a single ORF (73). This sole functional ORF is located on the antigenomic-sense RNA and is located at position 1600 – 1013. Despite the fact that only one HDV ORF has been demonstrated to be expressed, there are several other hypothetical ORF sequences greater than 300 nt in HDV and sequence analysis of HDV RNA genome and antigenome revealed numerous potential ORFs containing start codons (ATG) that could encode functional polypeptides (47). Among the possible ORF sequences present on HDV genome, five were subsequently tested but only one polypeptide reacted against HDV infected antisera (73).

The two forms of HDV antigens are mostly identical in sequence except that HDAg-L contains 19 additional amino acids at its C-terminus resulting from RNA editing of antigenomic HDV RNA at a location corresponding to the termination codon of HDAg-S by a host adenosine deaminase that acts on RNA (ADAR-1) (**Figure 1.3**) (49, 50). Despite being mostly identical in sequence, these two proteins have distinct functions: HDAg-S is essential for HDV accumulation, while HDAg-L is necessary for virion assembly (51). The ratio of antigen present in a virus particle compared to the number of HDV RNA molecules is approximately 200:1. This high number of HDAGs has been proposed to protect HDV RNA against nuclease activity (40).

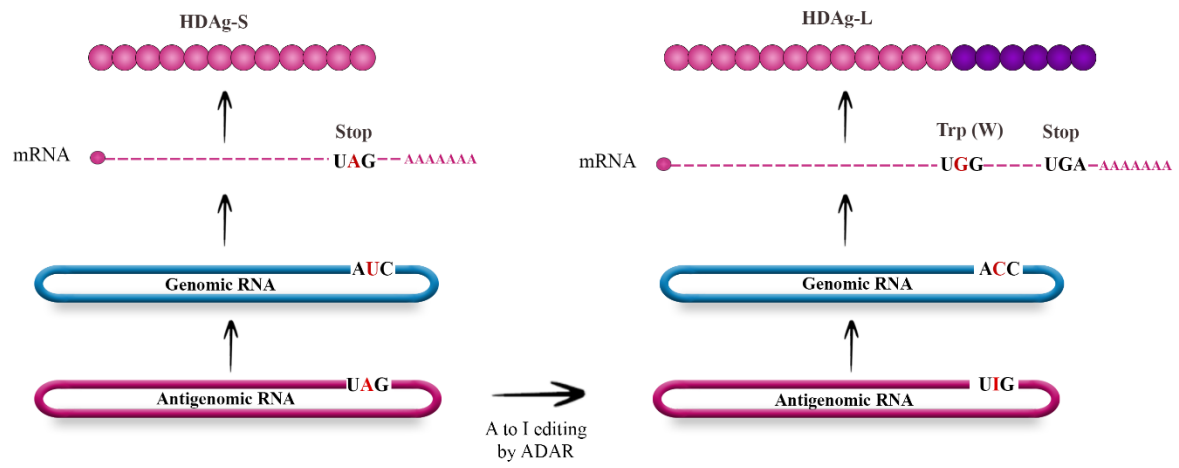


Figure 1.3. Editing of antigenomic HDV RNA by ADAR1.

The genomic strand is used to transcribe HDAg mRNA which is then translated into the small antigen (HDAG-S). Later during the replication cycle, ADAR1 edits the HDV antigenomic strand and deaminates the stop codon (UAG, position 1014) which leads to a change of adenosine (A) to inosine (I). During the next round of replication, this modification becomes an ACC in the genomic strand and results in a tryptophan (Trp /W) codon when it is transcribed into an mRNA, thus causing the production of a longer protein called the large antigen (HDAG-L).

1.6.1 Function of HDAg-S

The HDAg-S is a 24 kDa protein of 195 amino acids essential for HDV genome transport to the nucleus, where its transcription occurs (74) and for viral accumulation (64, 75). This protein contains many features including an RNA-binding domain, a nuclear localization signal (NLS) and a coiled-coil domain (CCD) (76). The RNA-binding domain found on HDAg-S as well as on the HDAg-L is important to accomplish various functions including nuclear and cytoplasmic shuttling, viral replication and virion packaging. This binding domain was first predicted to consist of two arginine-rich motifs termed the ARM I and ARM II (77). Other studies found another RNA binding domain in HDAGs N-terminal region (position 2 to 27) (78). A recent study has shown that the ARM regions are not required for HDV RNA specific binding (79) and that it is actually the HDAg N-terminal domain that directly interacts with HDV RNA in a discrete complex. This interaction occurs at various positions in this terminal region involving a pattern recognition. The structural conformation adopted by HDV RNA into a rod-like structure also plays an important role in its binding to the HDAGs (80). As for the other HDAg-S domains, the NLS is located within the N-terminus as well and is essential for HDV RNA transport into the nucleus (81).

The CCD is located from position 12 to 60 in the HDAGs amino acid sequence and is necessary for its dimerization. The HDAGs are known to assemble into higher order structures and this multimer organization has an important role in the binding of HDV RNA and the formation of RNP (82). The HDAGs form dimers through their CCD domain and were demonstrated to associate in octamers (82, 83). Full-length HDV is bound by roughly 5 oligomers, each containing 8 HDAg monomers, for a total of approximately 32 to 40 HDAg monomers per full-length HDV RNA (82). In addition, the HDAg-S CCD is important for its *trans* activator function in HDV replication (84). HDAGs also possess chaperone activities that

modulate the ribozyme cis-cleavage activity found on HDV RNA and this feature was hypothesized to be important for HDV replication (85, 86).

In addition to mediating HDV RNA transport into the nucleus, HDAg-S plays a role in efficient transcription initiation and/or elongation *in vivo* (65). HDAg-S has been proposed to stimulate transcription elongation by RNAP II by displacing the negative elongation factor (NELF) (87). Despite those studies, it has been demonstrated that HDV initiation can occur in the absence of HDAg-S *in vitro* (65).

1.6.2 Function of HDAg-L

HDAg-L possesses 19 additional amino acids at its C-terminus, for a total length of 214 amino acids, resulting from RNA editing of the termination codon of the small antigen catalyzed by ADAR-1. Aside from the extended C-terminal sequence found in HDAg-L, it shares the same domains as HDAg-S. This extended sequence contains a proline-rich motif that act as a nuclear export signal (NES) located from the residues 198 to 210, which differentiates it from HDAg-S (88). Despite being mostly identical in sequence to HDAg-S, its function is distinct. HDAg-L is necessary for HDV virion assembly, where it's additional C-terminal domain allows a protein-protein interaction with HBsAg (70, 89–91). For a long time, HDAg-L was thought to inhibit HDV replication. HDAg-L was shown to inhibit HDV genomic RNA synthesis but not the antigenomic, although these results were achieved using plasmids cDNAs overexpressing HDAg-L (90).

HDAg-L is a nucleoplasmic shuttling protein that can travel between the nucleus and cytoplasm (88). This shuttling is crucial for HDV virion assembly, particularly because HDAg-L creates physical contacts between the RNP and HBsAg (43). HBV envelope protein

synthesis takes place at the endoplasmic reticulum (ER) membrane and the empty subviral particles (SVPs) get assembled at the pre-Golgi membrane (88). The RNPs were proposed to form in the nucleus where it would be exported into the cytoplasm by HDAG-L. It is important to note that cytoplasmic HDAG-L specifically interacts with clathrin heavy chain (CHC) which has been shown to be important for HDV virus particle production. This interaction with cytoplasmic HDAG-L also competes with other protein binding to the CHC site, resulting in a disruption of normal clathrin-mediated endocytosis and protein transport (92). According to this model, this interaction between HDAG-L and CHC occurs in the trans-Golgi network and this proximity with HBsAg would initiate RNP packaging.

HDV RNP export from the nucleus to the cytoplasm has been subject to numerous speculations. Among those speculations, HDAG-L nuclear export was proposed to be mediated by the chromosome region maintenance 1 (CRM1)-independent pathway. The evidence supporting this pathway were the proline-rich residues present in HDAG-L NES sequence and its insensitivity to leptomycin B (93–95). CRM1 is the most common and best characterized nuclear export receptor. Many host-cellular proteins were proposed to be involved in HDV nuclear export such as the nuclear export signal-interacting protein (NESI), lamin A/C, nucleoporins and more recently the TAP-Aly complex (96–98). The interaction of HDAG-L with NESI was the first proposed interaction to be the key for HDV RNP trafficking out from the nucleus since NESI inhibition significantly reduced HDV genomic RNA cytoplasmic accumulation, although trace amount was still present (96). The packaging of the genomic RNA into viral particles was also completely abolished while HDAG-L packaging was reduced by 40%. Thereafter, the interaction of lamin A/C and nucleoporins with HDAG-L was also demonstrated with coimmunoprecipitation in addition to their colocalization with NESI (97). NESI was proposed to interact simultaneously with lamin A/C and nucleoporins at the nuclear

membrane through interaction domains within its sequence, in addition to binding HDAg-L (97). Knockdown of lamin A/C efficiently reduced the nuclear export of the large antigen and the formation of viral particles. On the other hand, HDAg-L was more recently found to colocalize with the cellular export receptor TAP and Aly in the nucleus (98). The C-terminal of HDAg-L directly interacts with TAP N-terminus. Other viruses like the herpes simplex virus 1 or the Epstein-Barr virus require one or both of these proteins to mediate nuclear export of their own components (98). It should be noted that HDV genotypes 1 and 2 have important differences in their HDAg-L C-terminal and this could influence the selected nuclear export method. The expression of all these host interacting proteins seems to have a direct influence on HDAg-L nuclear export capacity but whether all those proteins are required for CRM1-independent nuclear export is still unclear. Also, as discussed previously, HDV RNA was shown to be exported into the cytoplasm without the help of HDAg-L, which suggests that HDV RNA mediated transport can be achieved by an HDAg-L independent mechanism (44).

1.7 Post-translational modifications of HDAGs

A large number of host proteins have been identified to have a role in the HDV life cycle. Among them, many proteins involved in post-translational modifications have been found to interact with HDV components. The *delta* antigens are subject to several post-translational modifications which regulate their functions and modulates their activity. The cellular enzymes involved in post-transcriptional modifications play a crucial role in the progression of the HDV life cycle, including during HDV transcription, replication and antigen localization (99). The antigens are subject to phosphorylation, acetylation, methylation, and sumoylation and HDAg-L is also subject to isoprenylation (77, 100–106).

Because HDV lacks those post-translational enzymes, it depends on host components to accomplish these modifications and to pursue viral progression.

Among the proteins involved in post-translational modification of the antigens, the extracellular signal-related kinases 1 and 2 (ERK 1/2) was found to phosphorylate HDAg-S at serine-177, a required modification for the interaction of HDAg-S with RNAP II during antigenomic replication (107). Another kinase, casein kinase II (CKII) was shown to phosphorylate HDAg-S at the conserved Ser-2 position, which is required for its activity in HDV replication (101). The double-stranded RNA-activated protein kinase R (PKR) was also shown to phosphorylate serine residues at positions 177 and 180, and threonine at position 182 (102). In addition, HDAg-S can be sumoylated on multiple lysine residues which was shown to enhance the synthesis of genomic RNA and synthesis of mRNAs, but not the antigenomic RNA (105).

The extended 19 amino acid sequence on HDAg-L is isoprenylated at cysteine 211 which causes a conformational change in the protein. This modification has been suggested to be performed by the farnesyltransferase (FTase) enzyme and the resulting change in HDAg-L conformation would cause the main functional difference between the two antigens (108). Isoprenylation was proposed to play an important role in the inhibitory function that HDAg-L has on replication and this modification is associated with virion packaging. Mutation of cysteine 211 and serine 123 had a direct negative effect on cytoplasmic transport of HDAg-L which also affected HDAg-L interaction with HBsAg at the RE and virion secretion (99). Several other proteins are also reported to post-translationally modify HDAGs, including protein farnesyltransferase (FTase), protein arginine methyltransferase 1 (PRMT1), p300

cellular acetyltransferase, small ubiquitin-related modifier isoform 1 (SUMO1), and Ubc9 (109).

1.8 Interaction of HDV with Host Cellular Proteins

Due to its small size and limited protein-coding capacity, HDV relies heavily on host proteins to facilitate its replication and transcription (6, 74, 110). In order to complete its replication and infection cycle, HDV must take advantage of various types of host proteins involved notably in sub-cellular localization, in post-translational modifications, in RNA synthesis, in gene regulation as well as several others whose roles in HDV life cycle are not yet elucidated (reviewed by (109)). Several proteins have been identified to interact with HDV RNA and/or HDAGs and thus these interactions can be classified into two distinct categories: those interacting with the RNA directly and those interacting with the antigens.

1.8.1 HDAG interactions with host proteins

More than 100 proteins were identified to interact with HDAG-S in a co-immunoprecipitation assay coupled with mass spectrometry performed by *Cao et al.* (75). Their results revealed the interaction of HDAG-S with 9 out of the 12 RNA polymerase II subunits, as well as an interaction with the DRB-sensitivity-inducing factor (DSIF), PKR kinase, heterogeneous ribonucleoproteins (hnRNPs) and ZNF326, which were already known to be involved in HDV life cycle (75, 87, 102) (**Table 1.1**). These experiments have also established an association of HDAG-S with different helicases, with transcription-related proteins, with RNA binding and processing proteins, as well as proteins involved in cell cycle

division and apoptosis (CCAR1, CDC5L). Along with those, HDAg-S associated with chromatin and histones (Chromodomain helicase-DNA-binding protein 4, Centrosome-associated protein 350 (CEP350), centrosomal protein 170kDa isoform alpha, H2A, and H4 histones) and with a G-protein coupled receptor (75). Interestingly, more than 25% of these proteins have been shown to be necessary for HDV RNA accumulation (75).

Additional HDAGs interacting proteins include several implicated in sub-cellular localization such as nucleolin (C23), karyopherin 2 α (111), NESI (96), clathrin heavy chain (92, 112) and nucleophosmin (B23) (**Table 1.1**) (113). As mentioned earlier, HDV RNP export from the nucleus to the cytoplasm is mediated through HDAG-L. To achieve this, HDAG-L interacts with particular host proteins including NESI, lamin A/C, nucleoporins and the TAP-Aly complex (96–98). Although it is still unclear if they are all required to achieve cytoplasmic export, they have all shown to interact with HDAG-L. The HDAG-L is now known to directly interact with the N-terminal domain of CHC through its C-terminal domain (92), which possesses a putative clathrin box motif, a site known to be bound by other clathrin-binding proteins (114). Only the cytoplasmic-localized HDAG-L was found to bind CHC, and this interaction occurred at the trans-Golgi network where it could promote virion assembly (92).

HDAG possesses a sequence similarity with the A polypeptide chain from the NELF (negative elongation factor) a transcription elongation factor that acts with DSIF (DRB sensitivity-inducing factor) to cause transcriptional pausing of RNAP II. DSIF binds directly to RNAP II and this association allows the additional binding of NELF which induces transcriptional pausing. One of the four NELF subunits, NELF-A, is critical for RNAP II binding. *Handa et al.* have shown, through a sequence alignment between NELF-A and HDAG, that the 60-aa region of NELF-A previously determined to be critical for RNAP II

binding had weak sequence homology to the C-terminal 66-aa region of HDAG (115), although NELF-A and HDAG do not directly interact (115). HDAG competitively binds on the usual NELF-A site present on RNAP II thus displacing DSIF/NELF complex and enhancing RNAP II elongation (87, 116).

Many other HDAGs interacting proteins are implicated in RNA synthesis including RNAP II, transcription factors such as Ying Yang 1 (YY1), histone H1e and many others (**Table 1.1**) (109). The helicase MOV10, part of the RISC complex, was also found to be interacting with HDAG (75, 117). Although many proteins have been shown to interact with the *delta* antigens, a large majority of these proteins have not been associated with a specific role in the pathogenesis of HDV, thus their functions are still uncertain, and need further validating.

1.8.2 HDV RNA interaction with host proteins

Many host proteins were found to associate with HDV RNA genome. Among these, ADAR1 was shown to interact with HDV antigenome to catalyze editing of the amber codon from UAG to UIG which allows the production of the HDAG-L, a step essential for viral progression (118). In addition to interacting with HDAGs, PKR also interacts with HDV RNA (119). The specific binding of PKR with HDV RNA activates PKR kinase activity, which phosphorylates HDAG-S at Ser-177, Ser-180, and Thr-182, and these phosphorylations are proposed to protect viral RNA from nuclease digestion and to have a role in its stability (102, 119). Phosphorylations of HDAG-S has been shown to be critical for HDV genomic RNA accumulation (120).

It is now well established that RNAP II has a crucial role in HDV genome replication and transcription. Along with RNAP II association with HDAGs, it also interacts with HDV RNA genome. Our laboratory has previously shown that RNAP II interacts with both polarities of the HDV RNA genome using RNA immunoprecipitation (RIP) (64). These results are supported by the finding that HDV RNA products are highly sensitive to low levels of α -amanitin, a selective inhibitor of RNAP II and RNAP III (121), and the fact that HDAG mRNA exhibits typical features of transcripts generated by RNAP II such as a 5'-cap and a 3'- poly (A) tail (61, 62). Additionally, the non-small nuclear ribonucleoprotein (non-snRNP) spliceosomal factor SC35, a component of nucleoplasmic speckles known to be active sites of transcription by RNAP II as well as RNA processing, was found to interact with HDV RNA (65). SC35 associates with a 199nt segment of HDV RNA genome shown to contain an RNAP II promoter (65).

Another host protein interacting with HDV RNA is the glyceraldehyde 3-phosphate dehydrogenase (GAPDH). This protein is known to be part of glycolysis but is also implicated in many other processes such as DNA repair (122), transcriptional activation (123), cytoskeletal dynamics and apoptosis (124, 125). GAPDH was shown to bind HDV antigenomic RNA at a UC-rich domain between positions 379 and 414 (126). This binding improved *cis*-cleavage by increasing HDV ribozyme activity of the antigenomic strand from an efficacy of 34% to 62%.

Recently, the main three paraspeckle proteins (PSF, p54nrb, and PSP1), were demonstrated to interact with HDV RNA (118, 127, 128). The polypyrimidine tract-binding protein-associated splicing factor (PSF) has also been shown to be an HDV RNA interacting protein. In fact, PSF binds to the terminal stem-loop domains of both polarities of the HDV

genome (129). PSF was first found to be required for spliceosome formation as well as multiple other steps in the splicing process (130, 131). In addition, p54nrb-PSF heterotetramer were also implicated in HDV replication (118, 132). They are both involved in transcription, in RNA processing (pre-mRNA splicing), and may modulate transcriptional activity.

In addition to these host proteins interacting with HDV RNA, eukaryotic translation elongation factor 1A1 (eF1A1), heterogeneous nuclear ribonucleoprotein L (hnRNP-L) and the serine/arginine (SR)-rich proteins (ASF/SF2) were also found (118).

Table 1.1. Host protein interacting with HDV RNA and/or HDAGs.
(Modified from *Goodrum and Pelchat, 2018* (133))

Host Protein	Function	Interaction	System used	Reference
Double-stranded RNA-activated protein kinase R (PKR)	Phosphorylation (S117, S180, T182)	RNA HDAG-S	HepG2, HeLa	(102)
Casein Kinase II (CKII)	Phosphorylation (S2, S213)	HDAG-S	HuH-7	(101)
Protein Kinase C (PKC)	Phosphorylation (S210)	HDAG-L	HuH-7	(101)
Extracellular signal-related kinases 1 and 2 (ERK1/2)	Phosphorylation (S177)	HDAG-S	HEK-293T	(107)
Protein farnesyltransferase (FTase)	Isoprenylation with farnesyl (C211)	HDAG-L	Cos-7, d H189, HuH-7, NIH3T3	(100, 134–136)
Protein arginine methyltransferase 1 (PRMT1)	Methylation (R13)	HDAGs	HuH-7	(103)
P300 cellular acetyltransferase	Acetylation (K72)	HDAGs	HeLa, HuH-7, and HepG2	(104, 137)
Small ubiquitin-related modifier isoform 1 (SUMO1)	Sumoylation of lysine residues	HDAG-S	HuH-7	(105)
Ubc9	Sumoylation of lysine residues	HDAG-S	HuH-7	(105)
karyopherin (importin) 2 α	Nuclear import	HDAG-S	BRL	(111)

Host Protein	Function	Interaction	System used	Reference
Nuclear export signal-interacting protein (NESI)	Nuclear import	HDAg-L	HuH-7, HepG2, COS7	(97, 138)
Lamin A/C	Nuclear stability, chromatin structure and gene expression	HDAg-L	HuH-7	(97)
Nucleolin (C23)	Nucleolar localization, shuttling, RNA synthesis/accumulation	HDAGs	COS7, HuH-7, BHK-21	(139)
Clathrin heavy chain	Exocytosis	HDAg-L	HepG2, COS7, HuH-7	(92, 112)
TAP (or NXF1)	Cellular export receptor	HDAg-L	HuH-7	(98)
Aly (or REF)	Export adapter in nuclear export of mRNA	HDAg-L	HuH-7	(98)
Nucleophosmin (B23)	Nucleolar localization, shuttling, RNA synthesis/accumulation	HDAGs	HuH-7, HepG2	(113)
DRB sensitivity-inducing factor (DSIF)	Relieves transcriptional repression; stimulates elongation by RNAP II	HDAGs	HeLa	(87)
<i>delta</i> interacting protein A	Transcriptional regulation	HDAGs	HEK-293	(140)
Yin Yang 1 (YY1)	RNA synthesis/accumulation	HDAGs	HeLa, HuH-7, HepG2	(137)
Histone H1e	RNA synthesis/accumulation	HDAg-S	COS7, HuH-7	(141)
MOV10	RNA remodeling	HDAGs	HuH-7, HEK-293	(117)
Smad3	Host gene expression	HDAGs	HuH-7, Cos7	(142)
c-Jun	Host gene expression	HDAGs	HuH-7, Cos7	(142)
TRAF2	Host gene expression	HDAGs	HEK-293, HuH-7	(143)
SL1	Pol I-associated transcription factor	HDAGs	HuH-7	(58)
ZNF326	Transcription elongation	HADg-S	HEK-293	(75)
CCAR1	Helicase	HDAg-S	HEK-293	(75)
CDC5L	Helicase	HDAg-S	HEK-293	(75)

Host Protein	Function	Interaction	System used	Reference
Chromodomain helicase-DNA-binding protein 4 (CHD4)	Remodeling of chromatin	HDAg-S	HEK-293	(75)
Centrosome-associated protein 350 (CEP350)	Microtubule-organization at the centrosome	HDAg-S	HEK-293	(75)
Centrosomal protein 170kDa isoform alpha	Microtubule organization	HDAg-S	HEK-293	(75)
H2A and H4 Histones	Histone components	HDAg-S	HEK-293	(75)
Probable G-protein coupled receptor 179 precursor	Signal transduction	HDAg-S	HEK-293	(75)
SC35	Splicing factor	HDAg-S, gRNA	HuH-7	(57)
Adenosine deaminase acting on RNA (ADAR 1)	Post-transcriptional modification of HDV antigenome	agRNA	HuH-7, HEK-293	(49)
Glyceraldehydes 3-phosphate dehydrogenase (GAPDH)	Enhances <i>delta</i> ribozyme activity	agRNA	HeLa	(118, 126)
RNAP I	Antigenome synthesis	RNA	HeLa	(68)
RNAP II	Genome synthesis, mRNA synthesis, Antigenome synthesis	HDAg-S, RNA	HeLa	(64, 65)
RNAP III	Unknown	RNA	HeLa	(68)
Polypyrimidine tract-binding protein associated splicing factor (PSF)	Nuclear processes	RNA	HEK-293, HuH-7	(127, 129)
54 kDa nuclear RNA-binding protein (p54nrb)	Nuclear processes	RNA	HEK-293, HeLa	(118, 127)
Paraspeckle protein 1 (PSP1)	Unknown	RNA	HEK-293	(127)
Heterogeneous nuclear ribonucleoprotein L (hnRNPL)	mRNA processing	RNA	HeLa	(118)
Arginine/serine-rich splicing factor (ASF)	Splicing	RNA	HeLa, HEK-293	(118, 128)
Eukaryotic elongation factor 1A1 (eEF1A1)	Ribosomal aa-tRNA transport, gene expression	RNA	HeLa	(118)

1.9 Models developed to study HDV replication

Although animal models such as woodchucks have been used in the past, the lack of good cellular systems to study HDV replication is a limitation in the field. Indeed, co-infection or super-infection with HBV does not allow to study HDV replication independently of its helper virus. This need of HBV envelope for infection has been a limiting step for *in vitro* studies. However, since HDV replication does not depend on its helper virus, HDV replication can easily be achieved in different cell types using transfection or expression methods. Initiation of HDV replication requires HDAg-S, which is either delivered by the virion in case of an infection or is provided *in trans* by a plasmid or by other sources. Many experimental models have been developed through the years to study HDV including animal models for infection or transfection in different cell culture.

Infection in animals models, including woodchucks, and chimpanzees, had been achieved in the past but requires the presence of HBV, otherwise, HDV does not propagate resulting in a single replication cycle (144–147). Mouse model have been also used to study HDV infection (148–150). However, recent studies found a bird and snakes HDV-like agent without any detection of HBV-like sequences suggesting another mechanism of infection that is non-HBV related (151). Moreover, samples from patient liver cells infected with both viruses were also used in the past to study HDV (1). Aside from infections methods, DNA and RNA transfection have also been performed to study HDV genome replication. Those experimental systems have exploited different cell types including chimpanzee, woodchuck and mouse liver cells as well as many human cell lines such as HuH-7, HEK-293 and HepG2 (26, 75, 152, 153).

Until recently, HDV had only been found in the presence of its helper virus, HBV, and thus was only able to infect hepatic cells. However, HDV RNA was recently found in the salivary gland tissue of Sjögren's syndrome patients who tested negative for HBV, which raises the question of whether HDV could spread using HBV-independent mechanisms, such as by taking advantage of other viruses (154). A recent published paper showed that HDV RNPs could be enveloped by six other virus surface glycoproteins unrelated to HBV and HDV (155). HDV virus particles were propagated by HCV and other virus including vesiculovirus and flavivirus (155). Alternatively, Vakraku et al. propose that the RNA sequences found in Sjögren's patients are likely instead from an HDV-like sequence rather than the actual HDV virus itself (unpublished) (156). Despite this recent discovery, it was previously thought that HDV replication was limited to the primary hepatocyte since it shares HBV envelope proteins and as a result, share the same mechanism of entry into hepatocytes and thus a high selectivity for hepatocyte NTCP receptors.

The John M. Taylor laboratory has developed a system to study HDV genome replication using a HEK-293 based cellular system (153). This cellular system was developed in human embryonic kidney cells (HEK-293) where they used a Flp-In T-Rex cell line to generate a stable cell line expressing the HDAg-S open reading frame under a tetracycline-inducible promoter. This cell line was designated as 293-Ag and allows for the conditional expression of HDAg-S. They subsequently transfected 293-Ag cells with HDV RNA sequences containing a 2-nucleotide deletion in the HDAg-S ORF disrupting the translation of a functional HDAg-S. Those cells were called 293-HDV and allow for the initiation of HDV genome replication. After tetracycline induction of HDAg-S expression, these cells were

shown to fully support HDV replication with an average of 4.3×10^4 copies of HDV RNA after 24h.

The cellular system chosen to study HDV, whether is it for replication, or localization, or interaction with other host components, influences the results obtained. For example, nucleolar localization of HDAg-S was observed when it was provided by a plasmid transfection, but this localization has never been observed in infected liver tissues (57). The expression of HDAg-S alone versus with HDV genome replicating also influences the localization of the small antigen (57).

The examples provided above highlight how the system chosen to study HDV can influence the results obtained. For example, the Cunha laboratory performed three proteomic studies assessing how HDV affect host proteomes (152, 157, 158). Two of them used HuH-7 cells (HuH-7 and HuH-D12) with different mechanisms for expressing HDV. HuH-7 cells are transfected with plasmids, while HuH-D12 is a stably transfected cell line with a plasmid containing a trimer of HDV genomic cDNA. The third study used the HEK-293 system developed by the Taylor laboratory described above. These three studies each presented different overall results on HDV impact on the host proteome. We can also detect a difference in the amount of HDV RNA copies in the cell depending on the cellular system used (54). Thus, HDV replication efficiency is influenced by the system and the mode used to drive the replication.

1.10 How HDV affects its host cell

1.10.1 Proteomic analyses

Because HDV interacts with numerous host proteins, several studies were initiated to understand the consequences of these interactions on gene and protein expression. The effect

of the accumulation of HDAg and/or HDV RNA has been investigated using proteomic analyses in several experimental systems (**Table 1.2**). Protein expression analysis were performed by the Cunha laboratory to identify changes in host protein expression in the presence of HDV components (152, 157, 158). Their first proteomic experiment used a hepatocyte-derived cellular carcinoma cell line (HuH-7) transfected with plasmids coding for HDV genomic RNA, antigenomic RNA or each HDAg individually followed by a 2-D gel electrophoresis and mass spectrometry analysis (152). A total of 32 proteins were reported as being differentially expressed. Among those proteins, a total of 10 proteins were affected for cells transfected with HDAg-S or HDAg-L, while cells expressing genomic or antigenomic HDV RNA had 4 and 8 proteins differentially expressed respectively. These proteins were found to be associated with pathways involved in the regulation of nucleic acid and protein metabolism, transport, signal transduction, apoptosis, and cellular growth regulation. Among these identified proteins we can find vigilin, tubulin alpha 6, Eukaryotic translation initiation factor 2 subunit 1 and several others (**Table 1.2**). The dysregulation of triosephosphate isomerase (TPI), heat shock protein 105 (HSP105) and heterogeneous nuclear ribonucleoprotein D (hnRNP D) RNA expression was validated using real-time PCR (although this does not validate their protein expression) (152).

One year later, the same laboratory published another paper about HDV proteomics using a Huh7-D12 cell line as a model system. This system has been stably transfected with a plasmid containing the full-length HDV genomic cDNA derived from an infected woodchuck (157, 159, 160) and has been shown to express the viral antigens as well as HDV genomic and antigenomic RNA (157, 161). Using a similar experimental approach, they identified 23 differentially expressed proteins. Most of these proteins could be grouped in the following cellular pathways: regulation of nucleic acid metabolism, regulation of cell growth and/or

maintenance, and energy pathways. Like the previous study, TPI was dysregulated along with histone H1-binding protein (NASP), La protein, lamin C and others (**Table 1.2**). The differential expression of these proteins was validated with western blots.

Another proteomic analysis reported a total of 89 proteins affected out of 3000 analyzed using an MS-based quantitative proteomics approach (158). A HEK-293 cell system was used to mimic HDV replication (153). In these cells, HDAg-S and HDV RNA altered the expression of 49 and 40 proteins, respectively. A higher number of proteins were differentially expressed by the small antigen on its own than as compared to conditions with the addition of the HDV RNA. Differential expression of p53, ELAV, Transportin1 (TNPO1), Cofilin-1, Eukaryotic translation initiation factor 3 subunit D (EIF3D), and 10 kDa heat-shock protein (HSPE) were further validated using western blots. Gene ontology (GO) enrichment analysis revealed that the proteins up-regulated in the presence of HDAg-S were part of those two categories: translational regulator activity and ribonucleotide binding. In contrast, down-regulated proteins were part of the GO term cellular components category. Moreover, the EIF2 signaling pathway was found to be one of the most affected pathways by HDV RNA accumulation with a total of 17 proteins differentially expressed. In this cell line, HDAg-S was reported to upregulate translational regulators and ribonucleotide binding activities and disrupts pathways related to glycolysis/gluconeogenesis, pyruvate metabolism and cell cycle, more precisely the G2/M DNA damage checkpoint. These results are consistent with cell cycle disruption by HDV, but not with previous results showing a reduction of relative cell number in S and G2/M phase, with an increase in G1/G0 phase upon induction of HDV replication in this cellular system (153).

1.10.2 Interferon response

It has been reported that active HDV replication induces the production of interferon beta (IFN- β) and interferon lambda (INF- λ 1/2/3) (26). The pattern recognition receptor MDA5 has been identified as the key component in HDV recognition. Found in the cytosolic compartment, MDA5 was proposed to detect HDV during HDV RNP export to form new virions. This receptor detects long dsRNAs and higher-ordered RNA structures, and therefore the HDV rod-shaped secondary structure could be important for MDA5 recognition. Although IFN production rises upon HDV replication, it does not lead to HDV eradication. The innate immune response is widely reduced 7 days after infection but only a moderate effect is observed on HDV replication. Thus, even after the HDV-mediated innate immune response through MDA5 sensing, the interferon response displayed only limited inhibition of less than 50% of HDV replication, and this reduction was only seen at an early infection stage, showing that IFN therapy may not be significantly effective to eradicate HDV infection.

Table 1.2. List of host protein differentially expressed in the presence of HDV RNA and/or HDAGs.

Host protein	Biological function	System used	Reference
P53	tumor suppressor and the regulation of cell cycle	HEK-293	(158)
Heat shock 10 kDa protein (HSPE)	Chaperone, efficient protein folding	HEK-293	(158)
ELAV-like protein 1	c-myc stabilization	HEK-293	(158)
Transportin 1	Receptor for nuclear localization signals	HEK-293	(158)
Eukaryotic Translation Initiation Factor 3 Subunit D (EIF3D)	Translation initiation factor activity	HEK-293	(158)
Cofilin 1	ILK signaling pathway	HEK-293	(158)
14-3-3 σ	Signal transduction	HEK-293	(158)
FAM136A	Nuclear-encoded mitochondrial gene	HEK-293	(158)
BRI3BP	Tumorigenesis, p53/TP53 stabilization	HEK-293	(158)

Host protein	Biological function	System used	Reference
Histone H1 binding protein (NASP)	Signal transduction; Cell communication	HEK-293	(75)
Triosephosphate isomerase (TPI)	Metabolism; Energy pathways	HEK-293	(75)
Polyadenylate binding protein (PABP)	RNA metabolism	HEK-293	(75)
Rho GDP dissociation inhibitor (GDI)	GTPase activator	HEK-293	(75)
Guanine nucleotide-binding protein	Signal transduction pathway	HEK-293	(75)
Brebrin 1	Cell growth and/or maintenance	HEK-293	(75)
Keratine 8	Cell growth and/or maintenance	HEK-293	(75)
Vinculin	Cell growth and/or maintenance	HEK-293	(75)
Lamin C	Cell growth and/or maintenance	HEK-293	(75)
Acetyl-CoA acetyltransferase	Metabolism; Energy pathways	HEK-293	(75)
Zinc finger protein 326	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	HEK-293	(75)
High mobility group box 1	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	HEK-293	(75)
Guanine nucleotide binding protein	Signal transduction; Cell communication	HEK-293	(75)
Serum albumin	Transport	HEK-293	(75)
Heterogeneous nuclear ribonucleoprotein D (hnRNP D)	mRNA metabolism and transport	HuH-7	(152)
Heat shock protein 105 (HSP105)	Prevents the aggregation of misfolded proteins	HuH-7	(152)
Annexin IV	Regulation of early stages of apoptosis	HuH-7	(152)
Proteasome activator	Metabolism; energy pathways	HuH-7	(152)
NADH2 dehydrogenase (ubiquinone) flavoprotein 1 precursor	Metabolism; energy pathways	HuH-7	(152)
Adenylate kinase 2B	Metabolism; energy pathways	HuH-7	(152)
Eukaryotic translation initiation factor 2 subunit 1	Protein metabolism	HuH-7	(152)
Serine (or cysteine) proteinase inhibitor	Protein metabolism	HuH-7	(152)
Heat shock 60 kDa protein	Protein metabolism	HuH-7	(152)
CKAP4 protein	Cell growth and/ or maintenance	HuH-7	(152)

Host protein	Biological function	System used	Reference
Tubulin alpha 6	Cell growth and/ or maintenance	HuH-7	(152)
Keratin 8 & Keratin, type I cytoskeletal 19	Cell growth and/ or maintenance	HuH-7	(152)
Dihydropyrimidinase related protein 2	Neurogenesis	HuH-7	(152)
TRIM 28 protein	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	HuH-7	(152)
DNA structure-specific endonuclease FEN1	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	HuH-7	(152)
Ribonuclearprotein La	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	HuH-7	(152)
High density lipoprotein binding protein (vigilin)	Transport	HEK-293, HuH-7	(75, 152)
N-ethylmaleimide-sensitive factor attachment protein	Transport	HEK-293, HuH-7	(75, 152)
Sorting nexin 5	Transport	HEK-293, HuH-7	(75, 152)
Dopamine receptor interacting protein 4	Apoptosis	HuH-7	(152)
Interferon β/λ	Signaling proteins	HepG2, HuH-7, HepaRG	(26)
Interleukin 8 (IL8)	Antiviral protein	HEK-293	(127)
Nuclear Enriched Associated Transcript 1 (Neat1)	Scaffold protein	HEK-293	(127)
Clusterin	Role in tumorigenesis	HuH-7	(162)

1.10.2 Stress and foci formation

A recent paper published by our laboratory showed the involvement of paraspeckles components in HDV replication (127). Two major components of the paraspeckles were previously found to interact with HDV RNA: PSF and p54nrb (118, 129). Paraspeckles are nuclear bodies found in proximity of nuclear speckles. Their roles are still not fully understood, but can be induced by stress conditions and can sequester different RNA-processing/regulatory proteins (163, 164). These foci have been shown to be stress-inducible and they act through sequestration of different RNA-processing/regulatory proteins (163, 164). We identified the paraspeckle protein 1 (PSP1) as a novel RNA interacting with HDV genome. We also demonstrated PSP1 importance in HDV replication, in addition to PSF and p54nrb, using siRNA mediated knockdown. A considerable reduction of more than 90% of HDV RNA abundance was observed when those 3 paraspeckles components were knocked down, without affecting the level of HDAg-S. Interestingly, upon HDV replication, PSP1 was also shown to be delocalized outside of the nucleus. Although PSP1 was noticeably relocating into the cytoplasm upon HDV replication, a reduction of PSP1 in both nuclear and cytoplasmic fractions was observed, despite the fact that total cellular levels of PSP1 remained unchanged. Thus, we have suggested that PSP1 is localizing in insoluble aggregates.

The study further found that PSF also relocates to the cytoplasm upon HDV replication. This change in PSF localization was previously observed with BReast tumor Kinase (BRK) phosphorylation of PSF upon EGF stimulation, which leads to the same kind of cytoplasmic shift in addition to leading to a cell cycle arrest (164). John M. Taylor and his team have shown a similar cell cycle arrest happening in G0/G1 and this together with Mendes et al.'s proteomic analysis proposed that proteins implicated in cell cycle regulation were

affected upon HDV replication (153, 158). Along with PSP1 relocalization, we have revealed a co-localization of PSP1 with PABP, a stress granule marker, thus providing additional evidence that HDV causes cellular stress. Lastly, the levels of long non coding RNA Nuclear Enriched Associated Transcript 1 (Neat1) RNA was increased ~2-fold in HDV replicating cells (127). Neat1 is a fundamental part of paraspeckles formation, acting as a structural RNA implicated in the formation and maintenance of those nuclear structures (165). These results suggest that HDV induces a cellular stress response, and are consistent with the observed arrest in cellular cycle at the G1/G0 phase upon induction of HDV replication (153).

It was also demonstrated that HDAG-L could activate the nuclear factor kappa B (NF- κ B) as well as the signal transducer and activator of transcription-3 (STAT-3) which is consistent with the cellular stress response induced by HDV (166, 167). Reactive oxygen species (ROS) production was increased and suspected to be induced through HDAG-L activation of oxidative stress pathways via NADPH oxidase-4 (167). Excessive ROS production has been previously linked with HBV replication and protein synthesis (168). Activation of these pathways could also be involved in the production of stress foci observed with HDV replication and could be linked to HDV liver pathogenesis (127). In addition to these findings, morphological changes were observed upon HDV replication. Cells became rounder and smaller, a phenomenon that can be observed at early apoptotic stages (127, 153). However, many details are still missing on what precisely triggers stress foci formation in HDV replicating cells.

1.11 Rationale, hypothesis, and objectives

Both HDV RNA and its antigens interact with numerous host proteins involved in regulating gene expressions. Moreover, many previous results pointed towards a cell cycle dysregulation, but several questions remained unsolved. **Therefore, my hypothesis was that the accumulation of HDV RNA affects host gene expression involved in cell cycle.**

Proteins in a cell usually work in coordination with other proteins to achieve a certain function. If only one protein in the complex is affected, this might not affect the function of the overall complex. Because there is no existing analysis to my knowledge looking at gene expression from a complex point of view, I have decided to undertake this avenue. Therefore, I wanted to observe HDV replication effect on the expression of genes for proteins known to be part of protein complexes. Lastly, there had been multiple reports linking cells with active replicating HDV being subjected to a cell cycle arrest. However, there had been little rigorous validation of these observations and thus the link between HDV replication and cell cycle arrest remained poorly explored. Therefore, I wanted to observe cell cycle progression using a flow cytometry approach.

The objectives of my thesis were the following:

Objective 1: Observe the effect of HDV replication on host genes expression by high throughput RNA sequencing.

Objective 2: Identify the protein complexes affected by HDV replication.

- I. Identify protein complexes affected by HDV accumulation (bioinformatic).
- II. Select protein present in complexes affected and validate their expression using RT-qPCR.

Objective 3: Observe cell cycle progression using a flow cytometry approach

Chapter 2: MATERIALS AND METHODS

Cell culture and induction

Cells were kindly provided by Dr. John M. Taylor (Fox Chase Cancer Center, Philadelphia, USA). This cellular system was designed using a line of human embryonic kidney cells (Flp-In T-REx-293 cells; Invitrogen), which was used as control (HEK-293). The same cell type was stably transfected with HDAg-S cDNA under the control of a tetracycline-inducible promoter (TET), using Lipofectamine 2000 to generate an HDAg-S expression cell line (293-Ag). Then they transfected 293-Ag cells with HDV RNA sequences containing a 2-nucleotide deletion in the HDAg-S ORF disrupting the translation of a functional HDAg-S to generate a cell line capable of replicating HDV (293-HDV). HEK-293, 293-Ag, and 293-HDV cells were cultured at 37 °C with 5% CO₂ in Dulbecco's Modified Eagle's Medium (DMEM) with 10% calf serum. For HEK-293 cells, zeocin and blasticidin were used for HEK-293 cells at concentrations of 100 µg/mL and 5 µg/mL, respectively as selection markers. For 293-Ag and 293-HDV, hygromycin B and blasticidin were used at concentrations of 200 µg/mL and 5 µg/mL, respectively to maintain both the δ Ag and TET repressor genes. Cells were passaged every week with a dilution of 1/10 to obtain about 10% confluency ($\sim 2.2 \times 10^5$ cells). Induction of HDAg-S expression was carried out with the addition of 1 µg/mL of tetracycline (TET) performed on cells having reached $\sim 60\%$ confluence.

RNA isolation, library construction and Illumina Deep-Sequencing

HEK-293, 293-Ag, and 293-HDV cells were cultured in duplicates, one of which was induced with TET and the other was not, constituting the six different conditions used for subsequent experiments (**Figure 3.1**). Total RNA extractions were performed 48 hours after induction with TRIzol reagent (Invitrogen) following the manufacturer's recommendations.

The total RNA concentration in the extractions was quantified by spectrophotometry (Nanodrop) at 260nm and migrated on a 1.5% agarose gel to assess ribosomal RNA integrity (~10 µg of RNA in 5 µL, 4 µL of loading dye (Thermo Fisher) and 1µL of SYBR Green II RNA gel stain (Thermo Fisher)). To validate HDV induction, RNA concentrations were normalized to 500 ng across all the samples and reverse-transcribed using the iScript cDNA synthesis kit (Biorad) with random primers according to manufacturers instructions. Following the reverse transcription, a PCR reaction was performed using primers to amplify the regions containing the HDV ribozymes. Once the RNA quality was accessed and the induction validated, a total of 1200 ng of total RNA from each of the 6 samples was sent to the McGill University and Génome Québec Innovation Center where Illumina TruSeq rRNA-depleted Stranded libraries were prepared, multiplexed and sequenced using the Illumina HiSeq 2500 System (McGill University and Genome Quebec Innovation Centre, Montreal, Canada). Raw sequencing data will be deposited on the Sequence Read Archive of NCBI.

Differential expression analysis of transcript and data processing

The raw fastq file containing the reads of the 6 samples were de-multiplexed with cutadapt v1.8.1. Transcriptome reads were mapped using TopHat (v2.0.10) on the human reference genome (*Homo_sapiens.GRCh38.83*) and mapped reads were analyzed with the GFOLD (generalized fold change) algorithm (169). All further data analyses were performed using in-house R scripts (Figure S2). First, a threshold of 1 was used for Reads Per Kilobase Million (RPKM >1) to remove unexpressed or unmapped genes from the analysis. Genes were classified according to their expression change based on the GFOLD value. First, 293-Ag and 293-HDV cells gene expression were compared to 293 control cells to determine the difference of expression of these cell lines without TET induction. These expression results

made it possible to determine the basal expression of the promoter without the expression of tetracycline. The same steps were repeated for the tetracycline-induced cells, therefore, comparing 293-Ag-induced TET and 293-HDV TET with HEK-293 TET. Subsequently, each gene in the induced conditions were compared with those uninduced in the same corresponding cell type (293-Ag TET vs 293-Ag and 293-HDV TET vs 293-HDV) to remove the tetracycline effect. Thereafter, each gene from HEK-293 condition was compared to 293-Ag gene expression, followed by a comparison of 293-Ag gene expression to 293-HDV, according to whether the gene was overexpressed, under-expressed or if there was no change. A cutoff of 2-fold change in gene expression was used to differentiate between affected and non-affected genes. In total, nine categories were created to illustrate the change in expression affected by antigens, HDV RNA or both.

Gene ontology enrichment

A Gene Ontology enrichment was performed on each of the categories mentioned above except for categories with no changes. Genes names were extracted from each category and a GO term enrichment analysis was performed using PANTHER Overrepresentation Test (release 20170413) and the GO Ontology database (released 2017-05-25) with the default Bonferroni FDR correction (170). REVIGO was used to summarize the GO terms enriched using SimRel semantic similarities and medium similarity (0.7), and the clusters were visualized using Cytoscape (v.3.5.1) (171).

Identification of affected complexes

For the protein complex analyses, datasets were obtained from CORUM (03.05.2017 release) and in-house R scripts were used to perform the analyses (Figure S3). A first cutoff of a 1.5-fold change was used to distinguish significant and non-significant expression

changes. Thereafter, we selected complexes that had 75% or more of their proteins affected and performed a second analysis using a lower cutoff of 1.2-fold change to include proteins that change slightly with virus replication thus between 1.5 and 1.2-fold change. Both signed and unsigned analyses were performed. Fisher's exact test was performed and only the complexes composed of at least 3 proteins significantly enriched (p-values ≤ 0.05) were retained. A detailed list for each affected unsigned complex is present in Figure S4.

Reverse transcription and quantitative Polymerase Chain Reaction (qPCR)

For this experiment, 293-HDV and 293-HDV +TET were plated and grown until reaching 70% confluence. One plate was induced with the addition of 1mg/mL of TET in its respective media while the non-induced control plate media was replaced with the same amount of fresh media. Cells were incubated for another 48h before extraction in a cell incubator at 37°C with 5% CO₂. Total RNA was extracted using TRIzol and the RNA pellet was resuspended in distilled and deionized water (ddH₂O). The RNA integrity was accessed by running the RNA sample on a 1.5% agarose gel and the concentration was quantified using a spectrophotometer. The RNA concentration was normalized to 500ng across all the samples prior to reverse transcription. RNA was reverse-transcribed using the iScript cDNA synthesis kit (Biorad) with random primers according to manufacturers instructions. GAPDH, 18S, and B-globulin were tested as reference genes and their expression across the different conditions did not significantly change. GAPDH was selected as the reference gene for the subsequent qPCR experiments to normalize the gene expression across the different experiments. qPCR was performed using IQ Syber Green Supermix kit (Biorad), following the manufacturer's specifications and using a Chromo4 Real-Time Detector (Bio-Rad). The primer used are listed in **Table 2.1**. Primers were designed using Primer-BLAST from NCBI using Bio-rad's primer

design recommendations from their instruction manual as well as from recommendations in “A practical approach to Rt-qPCR” from *Nguyen et al.* (172, 173). Some primer sequences were taken from previously published manuscripts (see Table 2.1). Primers were purchased from ThermoFisher Scientific and were resuspended with ddH₂O. A regular PCR was used to observe the specificity and the length of the amplicon (Figure S1). Moreover, primer efficiency was tested for each designed primer prior to performing our qPCR validation experiments. For each gene, qPCR was performed using three technical replicates as well as three biological replicates. Relative quantification was performed with the $2^{-\Delta\Delta C_t}$ method by *Likav et al* (48) using GAPDH as reference gene. Pearson correlation coefficient (r) between RNA-seq and qPCR data were calculated in R.

Western blot assay

Cells were washed with cold-ice Phosphate Buffered Saline (PBS) prior proceeding with the cell lysis. Whole cell lysate was performed using a NP-40 lysis buffer (150mM sodium chloride, 1% Np-40, 50 mM Tris pH 8.0) with protease and phosphatase inhibitors. Once the cells were harvested in the lysis buffer, they were transferred into a pre-cooled microcentrifuge tube and they were kept in the buffer with constant agitation for 30 minutes at 4°C. Cells were centrifuged for 20 min at 12,000 rpm, the supernatant was harvested for the subsequent experiments and the pellet was discarded (protocol from Abcam: sample preparation for western blot). The protein samples were mixed with Laemmli 4X buffer, heated at 95°C for 5 minutes and migrated using polyacrylamide gel electrophoresis SDS-PAGE (10%) in running buffer (25 mM Tris-HCl, 200 mM Glycine, 0.1% SDS) at 100 volts. The proteins were transferred at room temperature for 1h to a PVDF membrane in transfer buffer (48mM Tris, 39mM glycine, 20% methanol, 0.037% SDS). The PVDF membrane was

blocked in 5% Bovine Serum Albumin (BSA) for 1 hour at room temperature in 1X Tris-buffered saline Tween (TBST; 200 mM Tris, 5 M NaCl, pH 7.5, 0.1% (v/v) Tween) and washed three times with TBST. The membrane was incubated with primary antibodies diluted in TBST solution with 3% BSA. The primary antibodies used were the following: aurora kinase B (rabbit polyclonal, # ab2254, Abcam) diluted 1/1000 and α -tubulin (mouse monoclonal #ab7291) diluted 1/500. The α -tubulin was used as a loading control. After 1h incubation with primary antibody, three washes with TBST were performed and the membranes were incubated with the proper secondary antibody in TBST +3% BSA (w/v). The secondary antibodies used were goat anti-rabbit IgG HRP (polyclonal, #ab6721, Abcam) diluted 1/3000 and rabbit anti-mouse IgG HRP (polyclonal, # ab6728, Abcam) diluted 1/10,000. The membranes were again washed 3 times in TBST for 20 minutes at room temperature with gentle agitation followed by 1 wash in Tris-buffered saline (200 mM Tris, 5 M NaCl, pH 7.5) for 5 minutes. The membranes were visualized using ECL reagent according to the manufacturer's recommendations (Thermo Scientific #32106). The membranes were then exposed to a photosensitive film and scanned. Densitometric measurement of the bands was performed for the digitized blots using Photoshop and normalized with the α -tubulin control.

Flow cytometry analysis

Sample preparation

HEK-293, 293-Ag, and 293-HDV cells were cultured in 6 well plates with and without tetracycline addition and were collected at various time points (12h, 24h, 36h). Controls included HEK-293 and 293-Ag induced and non induced as well as unstained induced and non-induced 293-HDV cells. Flow cytometry was performed using propidium iodide (PI) (BD

Biosciences, #550825) staining. Flow cytometry staining was performed using BD Bioscience's recommended assay procedure. Briefly, cells were harvested and washed in PBS. Cells were pelleted and washed by adding 10mL of PBS. Cells were centrifuged at 1,000 rpm for 10 minutes and the supernatant aspirated. Cells were fixed by adding 5mL of ice-cold 70% ethanol drop-wise while vortexing the cell suspension and were then sorted for 4 hours at -20°C in 70% ethanol. About a million fixed cells were aliquoted per sample and washed once with PBS followed by a wash in stain buffer (BD Bioscience #554656). Cells were centrifuged at 1,000 rpm, the supernatant was removed and cells were stained with PI/RNase staining by resuspending cells in 0.5mL PI/RNase buffer (BD Bioscience #550825) followed by a 15 min incubation at room temperature. Tubes were stored at 4°C and protected from light prior to analysis. Cells clumps were mechanically broken by filtration using a 40µm filter before analysis. Data were acquired on an LSR Fortessa cell analyzer (BD Biosciences) and analyzed using a custom R script (Figure S5).

Data processing: gating and rescaling

The R packages used for the flow cytometry analysis and for the creation of figures are the following: flowCore, MESS, ggplot2, reshape, and gridExtra (174–178). First, gating strategies were employed to exclude debris and cell doublets from the analysis. Subcellular debris and doublets can be distinguished from single cells using the cell size, thus a threshold was set to eliminate values that were below or equal this threshold using forward scatter values. Secondly, the peak from the G0/G1 were aligned between all the different conditions. This correction of the cell phase allowed to have a more uniform distribution and corrects misalignment of the instrument. Once the different populations of cells were rescaled to superimposed in all samples, it was impossible to compare the samples to each other since

every sample had different number of cells count. Therefore, the number of cells was adjusted among all sample based on the sample with the lowers cells and this rescale allowed to report the data on the same basis to make a comparison between the different data sets.

Data analysis

Using those adjusted data, density plots were generated for each of the conditions mentioned above thus allowing to compare the percentage of cells found in each phase of the cell cycle. Using manual cutoffs, the percentage of cells in each phase of the cell cycle has been determined.

Moreover, another analysis performed on the flowcytometry data was to compare the difference in cell size and granulocyt between 293-HDV cells with and without induction. Using the forward and side scatter (FSC and SSC) parameters given by the flow-cytometry, we were able to compare the cell size and granulocyt. Based on the same parameters determined to separate the phases of the cell cycle, we were able to extract more details about cell size and granulocyt according to which phase of the cell cycle cells were in.

Table 2.1. Primers sequences used for RT-qPCR analysis.

Primer name	Primer sequence
GAPDH Forward ¹	5'- CTGTTTCGACAGTCAGCCGCATC -3'
GAPDH Reverse ¹	5'- GCGCCCAATACGACCAAATCCG -3'
B2-microglobulin Forward ²	5'- GGCTATCCAGCGTACTCCAA -3'
B2-microglobulin Reverse ²	5'- TCACACGGCAGGCATACTC -3'
18S RNA Forward ³	5'- CGGACAGGATTGACAGATTGATAGC -3'
18S RNA Reverse ³	5'- TGCCAGAGTCTCGTTTCGTTATCG -3'
HDV-ribozyme Forward	5'-CCCTCGGTAATGGCGAATG -3'
HDV-ribozyme Reverse	5'-CCCAGTGAATAAAGCGGGTT -3'

Primer name	Primer sequence
INO80 Forward	5'- TACACTCAGGATGCCCCCTT -3'
INO80 reverse	5'- TGGAACTACTCTTGAGCGCC -3'
AURKB Forward	5'- CAGGGCTGCCATATAACCTGA -3'
AURKB Reverse	5'- CTAGCACAGGCTGACGGGGC -3'
CDT1 Forward ⁴	5'- ATCCGCACCGACACCTAC -3'
CDT1 Reverse ⁴	5'- TCTGAAGCCCACGTCTGT -3'
MCM2 Forward	5'- AGCACTTGATGAACTCGGGG -3'
MCM2 Reverse	5'- GTGGAGACAAAGCACTCGGA -3'
PDRG1 Forward	5'- GTGGATGCGCCTCTTACCAT -3'
PDRG1 Reverse	5'- CACGGCAGTTCATACTGGGA -3'
SH2D3C Forward ⁵	5'- GCCACATGAAGAGGCGAAGCAT -3'
SH2D3C Reverse ⁵	5'- GGCACAGTTTCGGATGGAGTCT -3'
SYT4 Forward	5'- GAGTGGACCTCCCTCTTTGC -3'
SYT4 Reverse	5'- GCTGAAAACAACAGCGCAGA -3'
HMGCS1 Forward	5' - CTCTAGGTGTGCTCCTGAATCAG - 3'
HMGCS1 Reverse	5' - CCTACTTCAGACCTTGAAGTGGA - 3'
LDLR Forward	5' - AGCAATGGCGGCTGCCAGTATCT - 3'
LDLR Reverse	5' - CTGTGGTCTTCTGATAGACGGGG - 3'
APOD Forward	5' - GCCACCCCAGTTAACCTCACAGA - 3'
APOD Reverse	5' - TTGGGGCAGTTCACCTGGTCTGT - 3'

1: Kmied et al. 2014 (179)

2 : Beeharry et al. 2018 (127)

3 : Nde et al. 2010 (180)

4 : Jian-na et al. 2008 (181)

5: Zeng L et al. 2016 (182)

CHAPTER 3: RESULTS

Author contribution

All experiments were performed by Gabrielle Goodrum except for the following:

Data processing from the raw sequencing files was performed by Lynda Rocheleau which included: cutadapt, tophat, and Gfold. Moreover, the WGCNA analysis was also performed by L. Rocheleau (**Figure 3.7A & 3.8A**).

The flow cytometry experiments in section 3.5 were performed by Gabrielle Goodrum and the bioinformatics scripts, data analysis and figures (3.11, 3.13 & 3.14) were performed in collaboration with Martin Pelchat.

3.1 Transcriptome analysis of genes affected by HDV

3.1.1 Sample preparation and data processing

To observe the impact of HDV genome replication and/or antigens production on host genes expression, a high-throughput RNA sequencing (RNA-seq) method was used to analyze the transcriptomic landscape of host cell replicating HDV. In this study, we used a HEK-293 based cellular system, developed by John M. Taylor to mimic HDV replication. This cellular system includes HEK-293 control cells, 293-Ag cells that are stably transfected with HDAg-S cDNA conditionally expressing HDAg-S in the presence of tetracycline (TET), and finally, 293-HDV cells that are 293-Ag cells containing HDV RNA with a 2 nucleotide deletion in the HDAg-S ORF, disrupting the expression of its own HDAg-S. HEK-293, 293-Ag, and 293-HDV cells were cultured in duplicates, one of which was induced with TET and the other was not, constituting the six different conditions used for subsequent experiments (**Figure 3.1**). Total RNA was extracted with TRIzol two days after induction, which corresponds to the maximal level of HDAg-S and HDV genome expression(153).

Once the RNA was extracted from the six samples, the RNA concentration was measured by spectrophotometry and migrated on an agarose gel to assess the ribosomal RNA quality (**Figure 3.2A**). Ribosomal RNA (rRNA) is part of about 80% total RNA found in the cell, while mRNA represents only about 1-3% of total RNA. As a result, mRNA integrity is usually estimated by looking at rRNA quality, assuming those rRNA populations reflects the mRNA population. Two sharp bands are visible at 1400bp and 2800bp as well as a smaller one around 100bp, which correspond to the 18S ribosomal RNA, 28S ribosomal RNA and transfer RNAs with other small RNAs, respectively. The 28S rRNA is approximately twice the intensity of 18S which is also a good indication of intact RNA. Observation of the small

RNA product around 100bp suggests that the RNA are largely intact and not degraded by RNase (183). HDV induction was validated by performing a PCR (**Figure 3.2B**). Results from the PCR show a band on the agarose gel corresponding to the expected HDV product at around 212bp. Observation of a faint band in the PCR control with a molecular weight smaller than 100nt corresponds to HDV primer dimerization. An amount of 1200ng of the total RNA for each sample was sent to the McGill University and Génome Québec Innovation Center for the creation of multiplexed cDNA libraries and to perform the Illumina HiSeq 2500 high-throughput sequencing. Prior to the creation of the library, they quantify and analyzed our RNA using a Bioanalyser (**Figure 3.3A**). Likewise, the two peaks positioned around 2000 and 400nt correspond to 18S and 28S rRNA, the peak around 100nt correspond to tRNA and the one at 25nt is a marker used for calibration. They also performed an RNA quality test using $OD_{260/280}$ and $OD_{260/230}$. The $OD_{260/280}$ ratio of all the sample was in an acceptable range indicating there was no significant protein contamination in our sample, but $OD_{260/230}$ the ratio was slightly lower indicating the presence of contaminants absorbing at 230nm like residual phenol originating from TRIzol reagent used for the extraction. However, all the ratios were within the acceptable ranges (**Figure 3.3B**). Moreover, the RNA Integrity Number for all the samples was 10, which is the most intact RNA according to their system going from 1 to 10, where 1 has the most degraded profile and 10 is the most intact.

Once the quantification and quality of RNA access, a multiplex library was created with our 6 samples. A ribosomal RNA depletion was performed prior to constructing the sequencing library. Multiplex libraries are created by adding individual “barcode” sequences to each DNA fragment from the same sample. This allows for the pooling of multiple libraries and for everything to be sequenced simultaneously (i.e our 6 different conditions) in a single sequencing run. Those specific sequences or barcodes are necessary to trace the origin of the

sample for the analysis. Once created, the library was sequenced using Illumina HiSeq 2500 high throughput sequencing. The number of reads per sample is shown in **Figure 3.3C**, with an average number of 9,780 million bases sequenced per sample and a total number of 235 million reads.

Sequencing reads were received as raw files (fastq) and were de-multiplexed with cutadapt (184). Subsequently, the sequencing reads were mapped to the human reference genome (*Homo_sapiens.GRCh38.83*) and to HDV RNA sequences using TopHat (185). The GFOLD (generalized fold change) algorithm was used to access expression changes between our 6 different conditions (**Figure 3.1**). GFOLD assigns reliable statistics for expression changes based on the posterior distribution of log fold change (169). R in-house scripts were used for data processing and statistical analysis.

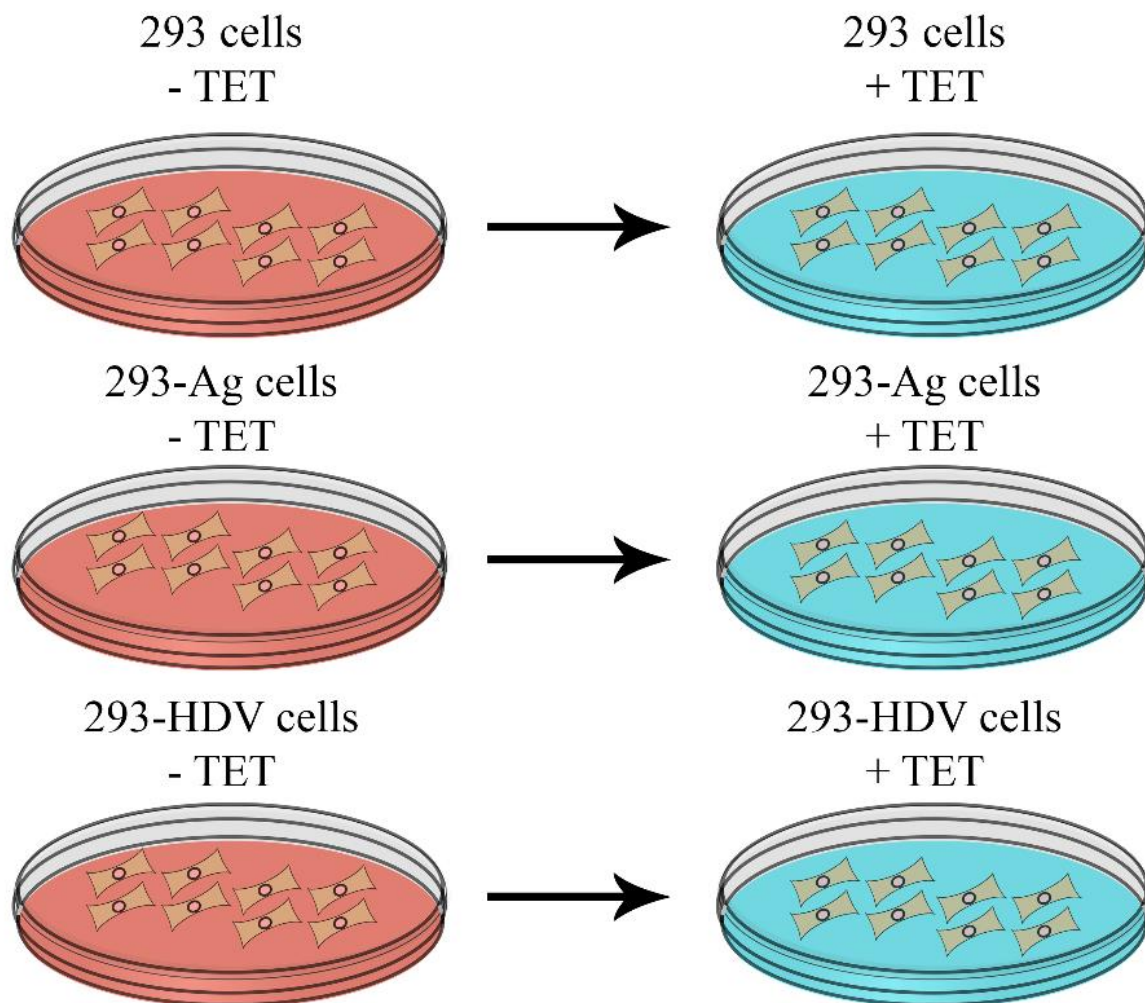


Figure 3.1. Overview of the six different conditions used for subsequent experiments and sequencing. The three-cell type used in this study are: HEK-293 (293), 293-Ag and 293-HDV. The three cell types were cultured in duplicates, one of which was induced with TET (+TET, colored in blue) and the other was not (-TET, colored in pink). Those 6 conditions were used for total RNA extraction and were sent to McGill University and Génome Québec Innovation Center for sequencing.

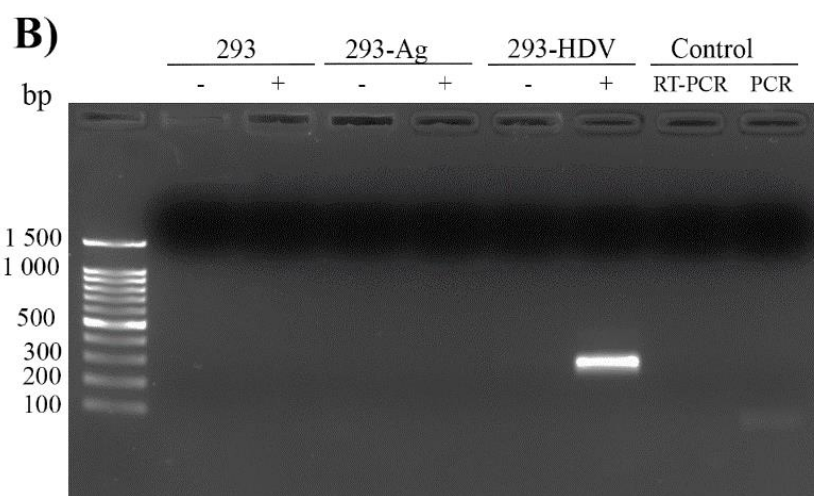
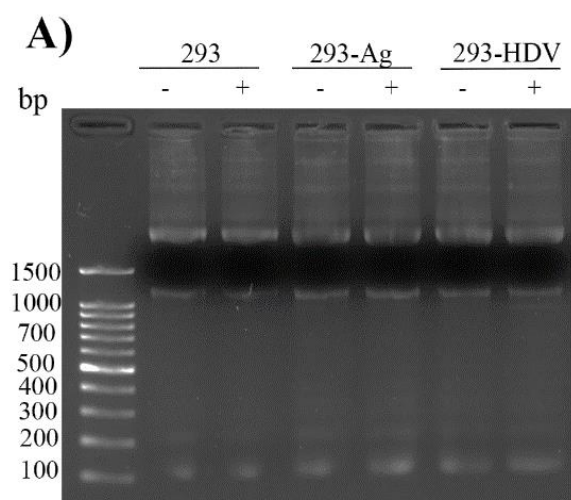
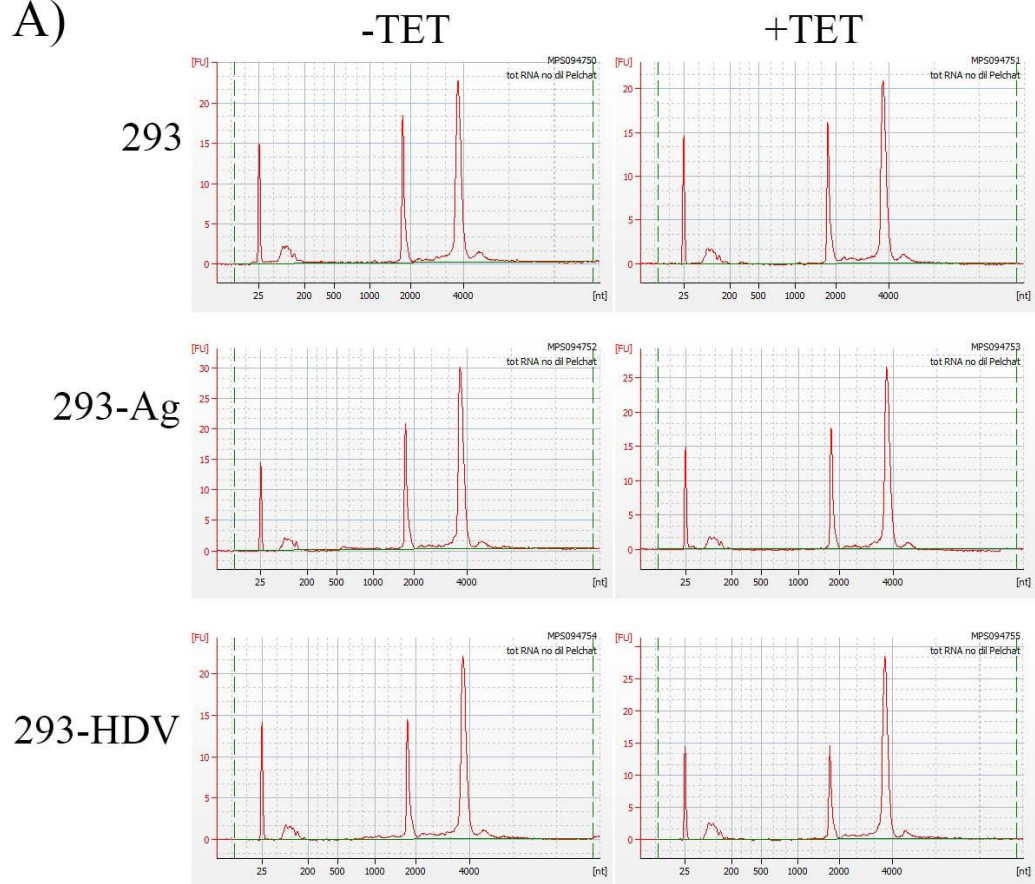


Figure 3.2. Validation of HDV induction and RNA integrity. **A)** Observation of RNA integrity after TRIzol extraction of the six-cell sample (Figure 3.1): HEK-293 (293), 293-Ag and 293-HDV, non induced (-) and induced (+). Total RNA was extracted using TRIzol and quantified using a spectrophotometer. A total of 500ng was mixed with 4 μ L of loading dye and 1 μ L of SYBR Green II for RNA gel staining and loaded on a 1.5% agarose gel. The DNA ladder use was GeneDireX 100bp (10 μ L). **B)** RNA was reverse-transcribed using random primers and a PCR was performed using HDV-ribozyme primers. The RT-PCR control corresponds to the reverse transcribed control performed using the reverse transcription master mix without RNA template (replaced by water), while the PCR control represents the PCR master mix without DNA (substitute by water).

A)



B)

Sample	260/230	260/280	RIN
293	1.57	1.87	10
293+	1.25	1.85	10
Ag	1.79	1.85	10
Ag+	2.00	1.83	10
HDV	1.68	1.84	10
HDV+	2.02	1.88	10

C)

Sample	Number of bases sequenced	Number of sequence reads
293	10900M	43.8M
293+	8900M	35.6M
Ag	9820M	39.3M
Ag+	9090M	36.3M
HDV	8990M	36.0M
HDV+	10900M	43.8M
	Average	Total
	9780M	235M

Figure 3.3. RNA quality, concentration, and integrity data as well as the number of bases sequences and the number of sequences reads obtained by Illumina sequencing. A) Evaluation of RNA quality of the six RNA samples by McGill University and Génome Québec using a Bioanalyser. HEK-293 (293), 293-Ag and 293-HDV, non induced (-TET) and induced (+TET). **B)** RNA quality assessment using 260/230 and 260/280 ratio as well as the RNA integrity number (RIN). **C)** Illumina deep sequencing results of the number of bases sequenced and the number of sequenced reads for each sample from the rRNA-depleted stranded library.

3.1.2. Gene expression analysis

To observe HDV replication effect on host gene expression, first I confirmed HDV induction using the RNA-seq data obtained from the analysis. To assess how many HDV sequences were produced following the induction, sequencing reads were mapped to HDV genome (**Table 3.1**). Some sequencing reads originating from HEK-293 cells with and without tetracycline were mapped to HDV genome, which are probably originating from false HDV-like sequences. Tetracycline induction in HEK-293 cells did not increase the number of HDV reads compared to the non-induced one. A higher number of HDV-like sequence were mapped in 293-Ag cells for a total of 1,832 and 5,386 sequences for the non-induced and the induced cells respectively. Those values were subtracted from the subsequent 293-HDV reads. The non-induced 293-HDV cell had a total of 76,622 reads aligning to the HDV genome, originating from promoter leakage, while the induced 293-HDV had 1,761,864 reads. This difference corresponds to an increase of ~23 time the amount of HDV RNA produced in induced cells compared to non induced. Low levels of transcription often occur in promoter inactive state, which is often referred to as promoter leakage. This basal transcription rate can be influenced by the architecture of the promoter (186).

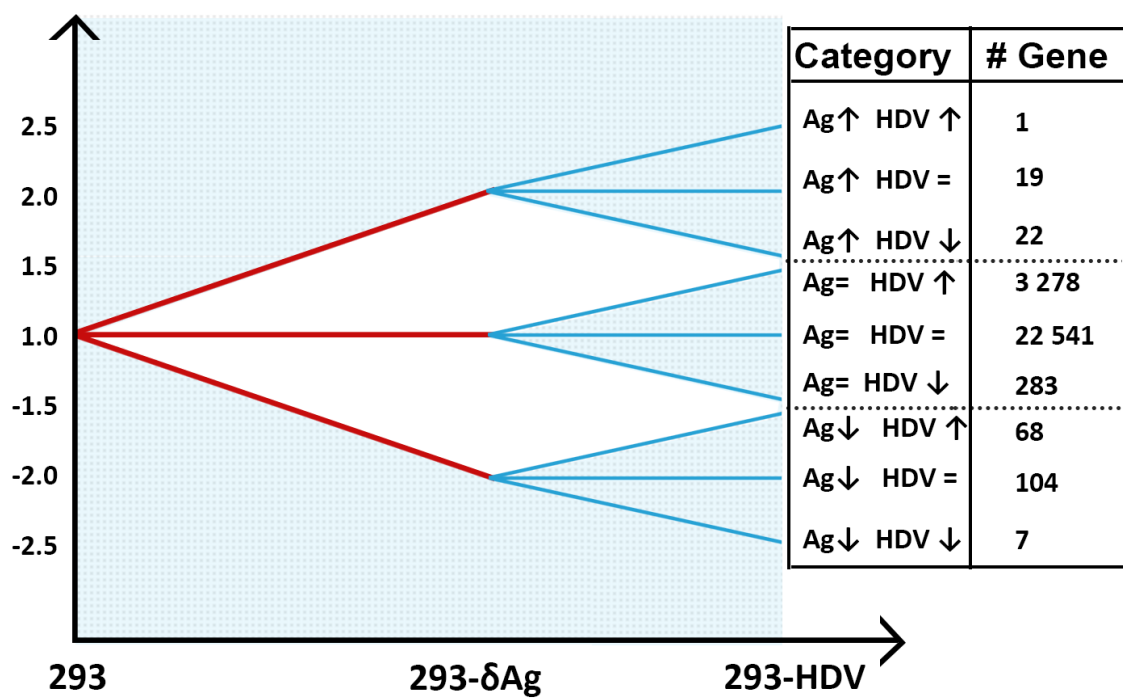
Table 3.1. Number of read able to map to HDV for each cell type in the RNA-seq data.

Cell type	Number of HDV reads	
	-TET	+TET
HEK-293	50	45
293-Ag	1,832	5,386
293-HDV	76,622	1,761,864

Next, genes were classified according to their expression, using in-house R scripts, to establish which genes were differentially expressed under the 6 different conditions (Figure S2). Genes having RPKM (read per kilobase per millions reads) equal to zero were removed from the analysis since those genes were either not expressed or no reads were mapped to these genes. Genes were classified according to their expression change based on the GFOLD value. First, non-induced 293-Ag and 293-HDV genes expression were compared to HEK-293 control cells. Although the expression of the small antigen and the replication of HDV should not be induced without tetracycline, the promoters usually have some basal leakage and thus this comparison of 293-Ag and 293-HDV to HEK-293 allowed us to subsequently remove those false positive in our analysis. The same steps were repeated for the tetracycline-induced cells. Subsequently, each gene in the induced (TET) cells were compared with those uninduced in the same corresponding cell type (293-Ag TET vs 293-Ag and 293-HDVTET vs 293-HDV). These steps allowed the removal of tetracycline effect on gene expression. Additionally, we kept only the genes present among all the different conditions, since certain genes were not present in all samples, as it was impossible to compare the expression of this gene if it was absent in any of the conditions. Subsequently, normalized genes were separated into categories according to whether the gene was overexpressed, under-expressed or if there was no change (cutoff of 2) (**Figure 3.4**). Each gene from the HEK-293 condition was compared to 293-Ag gene expression, followed by a comparison of 293-Ag to 293-HDV, according to the three criteria mentioned above. This creates a total of 9 categories for gene expression as detailed in **Figure 3.4**. A total of 3,782 genes were differentially expressed according to their fold change expression and among those, 3,561 genes were affected by HDV RNA alone, 123 were affected by the antigens only and 98 genes were affected by both (**Figure 3.4**). Among the genes affected by HDV RNA, 3,278 were up-regulated and 283

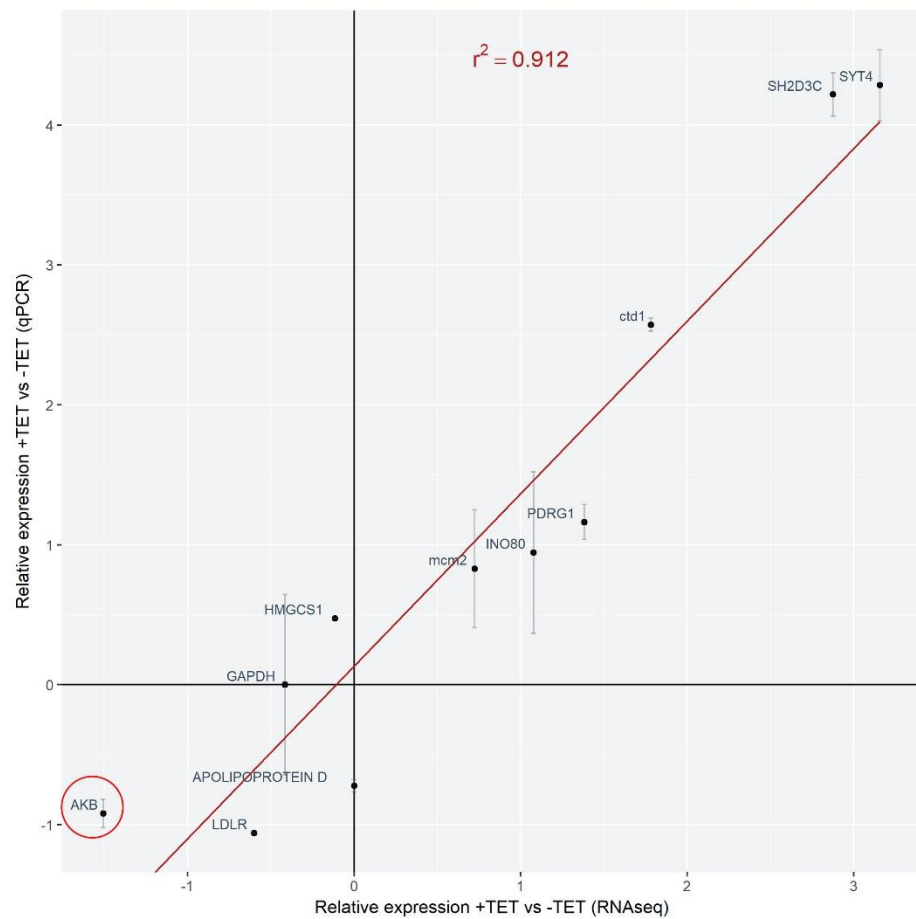
down-regulated, whereas the presence of the antigens induced the up-regulation of 42 genes and the down-regulation of 179 genes in total. Although several genes are influenced by the presence of the antigen alone, they represent only a small portion (3.3%) of the differentially expressed genes, which is a considerably small portion of the genes affected. Among all the genes, 22,541 genes remained unchanged by HDV and/or the antigen.

To validate the accuracy of the RNA-seq data, RT-qPCRs were performed on selected genes according to their expression value. This validation of RNA-seq data with qPCR was required since the sequencing was performed once (n=1). Genes selected for RT-qPCR validation are found in **Table 3.2**. The validation was done using the 293-HDV induced and non-induced conditions after 48h of TET induction. Total RNA was extracted, quantified with a spectrophotometer and the RNA integrity verified on an agarose gel. RNA concentration was normalized and reverse-transcribed. PCR was used to observe the specificity and the length of the amplicon (Figure S1) and primer efficiency was tested for each designed primer. qPCR was performed on selected genes selected and relative quantification was performed using GAPDH as reference gene. The expression of genes confirmed by RT-qPCR can be accessible in **Figure 3.5A**. A coefficient of determination of 0.912 ($R^2 = 0.912$) was determined between RNA-seq and qPCR results. This coefficient indicates that the RNA-seq data correlates well with validated the RT-qPCR data, indicating that we can rely on these values for future analysis. In addition, a western blot was performed on one of the genes validated by RT-qPCR (aurora kinase B/ AURKB) to show the downregulation in the protein level as well (**Figure 3.5B**). A reduction of 2.84x in gene expression was observed with the RNA-seq data and a 1.9x decrease with RT-qPCR (**Figure 3.5A**) while a reduction of 87% in AURKB protein expression was observed in HDV replicating cells when normalized to α -tubulin (**Figure 3.5C**).

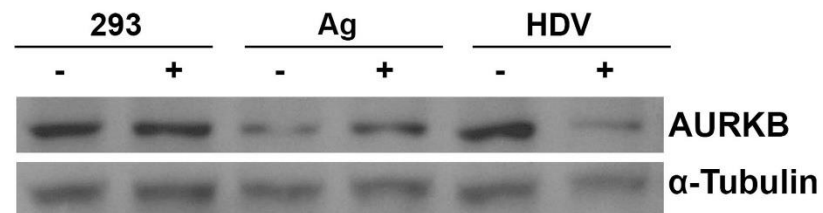


Figures 3.4. Classification of genes according to their expression change between different HEK-293, 293-Ag and 293-HDV conditions. Classification of genes according to their expression using a cutoff of 2-fold change. A total of nine categories were created to illustrate the change in expression affected by antigen, HDV RNA or both. The red line represents the expression change between HEK-293 and 293-Ag and the blue lines represent the difference between 293-Ag and 293-HDV. Upregulated genes are represented by ↑, downregulated by ↓ and no change by =. The number of genes found in each category are indicated in the column “#Gene”. The blue dotted background represents the genes that had a change in expression and have been used for the subsequent study.

A)



B)



C)

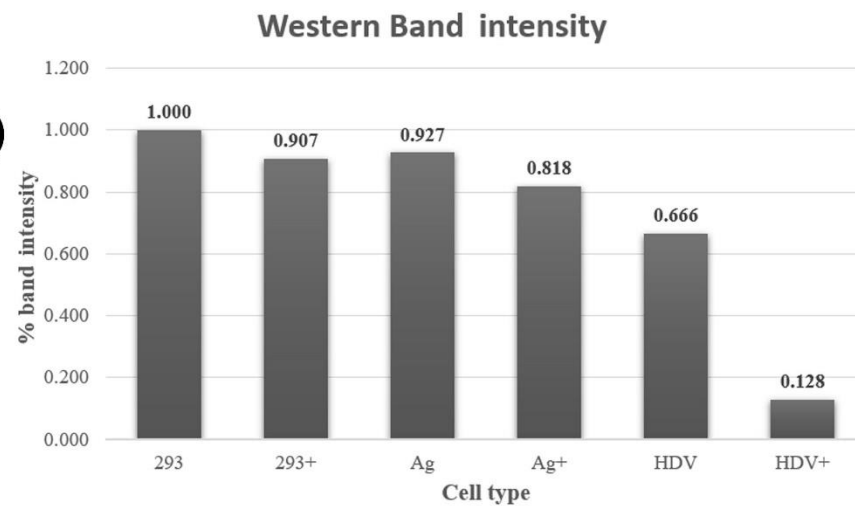


Figure 3.5. Validation of RNAseq identified gene expression changes

A) Effect of HDV accumulation on different host genes expression. 293-HDV cells were cultures in duplicate, one induced and one non-induced with TET. After 48h total RNA was extracted, quantified and the quality access. The RNA concentration was normalized among all samples and reverse transcribed. HDV induction was validated. Quantification by RT-qPCR of the expression of 11 host genes levels (listed in **Table 3.2**) normalized to the amount of GAPDH and to un-induced 293-HDV cells. RT-qPCR values represent the mean of three biological replicates (n=3) and the standard deviation is represented by the error bars for each gene. The expressions values are represented in log2. The Pearson correlation between the RNAseq and qPCR data was 0.912 ($R^2 = 0.912$). The gene AURKB circled in red represent the gene selected for western blot analysis in panels B and C. **B)** Western Blotting showing the relative amount of AURKB compared to α -Tubulin in a whole cell fractionation of HEK-293 (293), 293-Ag (Ag) and 293-HDV (HDV) induced (+) and non-induced (-) cells. **C)** Densitometric measurement of the western blot bands in B) using Photoshop after normalization with α -Tubulin and to the 293 control.

Table 3.2. RNA-seq and qPCR values of genes selected for gene expression validation in HDV replicating cells. The “Ensembl #” corresponds to the unique gene annotation associated with each transcript. The “protein acronym” corresponds to the acronym of the gene name or transcript name. The “RNA-seq” corresponds to the fold changes values obtained with RNA-seq comparing 293-HDV uninduced expression with 293-HDV induced expression after normalization, while the “qPCR” corresponds to the value obtained using qPCR comparing 293-HDV un-induced expression with 293-HDV induced normalized to GAPDH. Data are represented as linear values as well as in log2.

		Fold change		Log2 (fold change)	
Ensembl #	Protein acronym	RNA-seq	qPCR	RNA-seq	qPCR
ENSG00000178999	AURKB	0.35	0.53	-1.51	-0.92
ENSG00000088356	PDRG1	2.61	2.24	1.38	1.16
ENSG00000128908	INO80	2.11	1.92	1.08	0.94
ENSG00000073111	MCM2	1.65	1.78	0.72	0.83
ENSG00000167513	CTD1	3.44	5.95	1.78	2.57
ENSG00000132872	SYT4	8.91	19.49	3.16	4.28
ENSG00000095370	SH2D3C	7.34	18.61	2.88	4.22
ENSG00000111640	GAPDH	0.75	1.00	-0.42	0.00
ENSG00000189058	APOD	1.00	0.61	0.00	-0.72
ENSG00000112972	HMGCS1	0.92	1.39	-0.12	0.48
ENSG00000130164	LDLR	0.66	0.48	-0.60	-1.06

3.2 Enrichment of gene ontology term and biological processes affected by HDV replication

High-throughput RNA sequencing delivers a remarkable amount of details about the transcriptional landscape by measuring the expression levels but also give insights about alternative splicing, novel splice variants and RNA editing (187–189). However, working with large amounts of data can be difficult and presents major challenges, but Gene ontology allows one to easily interpret datasets and makes the analysis more intuitive for biologists. Gene Ontology delivers a controlled and unified vocabulary for the annotation of gene and their products. It describes and classifies based on three aspects: molecular function, cellular component, and biological process. These annotations allow to find relationships between gene function, gene products, biological process, molecular function and cellular component (190). As *Ashburner et al.* explained, this relationship reflects the unavoidable fact that “particular protein may function in several processes, contain domains that carry out diverse molecular functions and participate in multiple alternative interactions with other proteins, organelles or locations in the cell” (190).

Subsequent analysis focused on uncovering which biological processes were affected by HDV RNA and/or the antigens. A Gene Ontology (GO) enrichment analysis was performed on the genes present in each differentially expressed gene category mentioned above by extracting the names of the genes in each category (**Figure 3.4**). The enrichment analysis yielded no results for those small categories that had <100 genes present. Only the two largest categories (Ag=HDV↓ and Ag=HDV↑) had significantly enriched GO terms. Overall, 129 terms were significantly enriched in the up-regulated genes affected by HDV RNA. A total of 19 GO terms were found affected for downregulated genes. Moreover, 2 GO terms were found in the category where the gene expression decreased in presence of antigen as well as HDV

RNA, but no cluster was enriched for this category. The number of GO terms enriched for each analysis is accessible in the supplementary section **Table S1 and S2**. The GO terms enriched were then summarized with REVIGO, using a clustering algorithm relying on semantic similarity measures which allow to remove redundant GO terms (191). Among the processes enriched for upregulated genes, 4 principal clusters were enriched: RNA processing, the G-protein coupled receptor signaling pathway, protein transport and organelle organization (p-values < 0.05) (**Figure 3.6A**). Whereas, a single cluster was enriched for the downregulated genes: nucleosome assembly (p-values < 0.05) (**Figure 3.6B**).

As Beeharry *et al.* had recently reported the presence of cellular stress in HDV, I was intrigued to see how HDV affects stress-related genes in the cell. The first step in this analysis was to identify all genes that are associated with cellular stress. The AmiGO database provides for searches to be performed and for all genes associated with specific GO term to be identified. The genes associated with the following GO terms were downloaded for further analysis: “GO 0034063 :stress granule assembly”, “GO 0033554: cellular response to stress”, “GO 0010494: cytoplasmic stress granule”, "GO 0001725 stress fiber" and “GO 0090400: stress-induced premature senescence”. These genes were then compared to those dysregulated by HDV replication (cutoff of 2), thus is allowed to sorted dysregulated genes related to stress. Out of the 5,236 stress related genes identified in the GO term search, 498 genes were significantly affected by HDV induction. Key genes identified were the cyclic AMP-dependent transcription factor (ATF3) which was upregulated 9.7-fold with HDV, Growth arrest and DNA damage-inducible protein GADD45 alpha (GADD45A) upregulated by 3.9 fold and Fibroblast growth factor 21 (FGF21) by 3.8 fold. The ATF3 transcription factor is induced by many stress factors and regulates genes involved in the cell cycle, apoptosis, cell adhesion and several other pathways (192). GADD45A is a stress inducible gene and interacts

with many cellular proteins involved in cell cycle checkpoint, DNA repair and signal transduction (193).

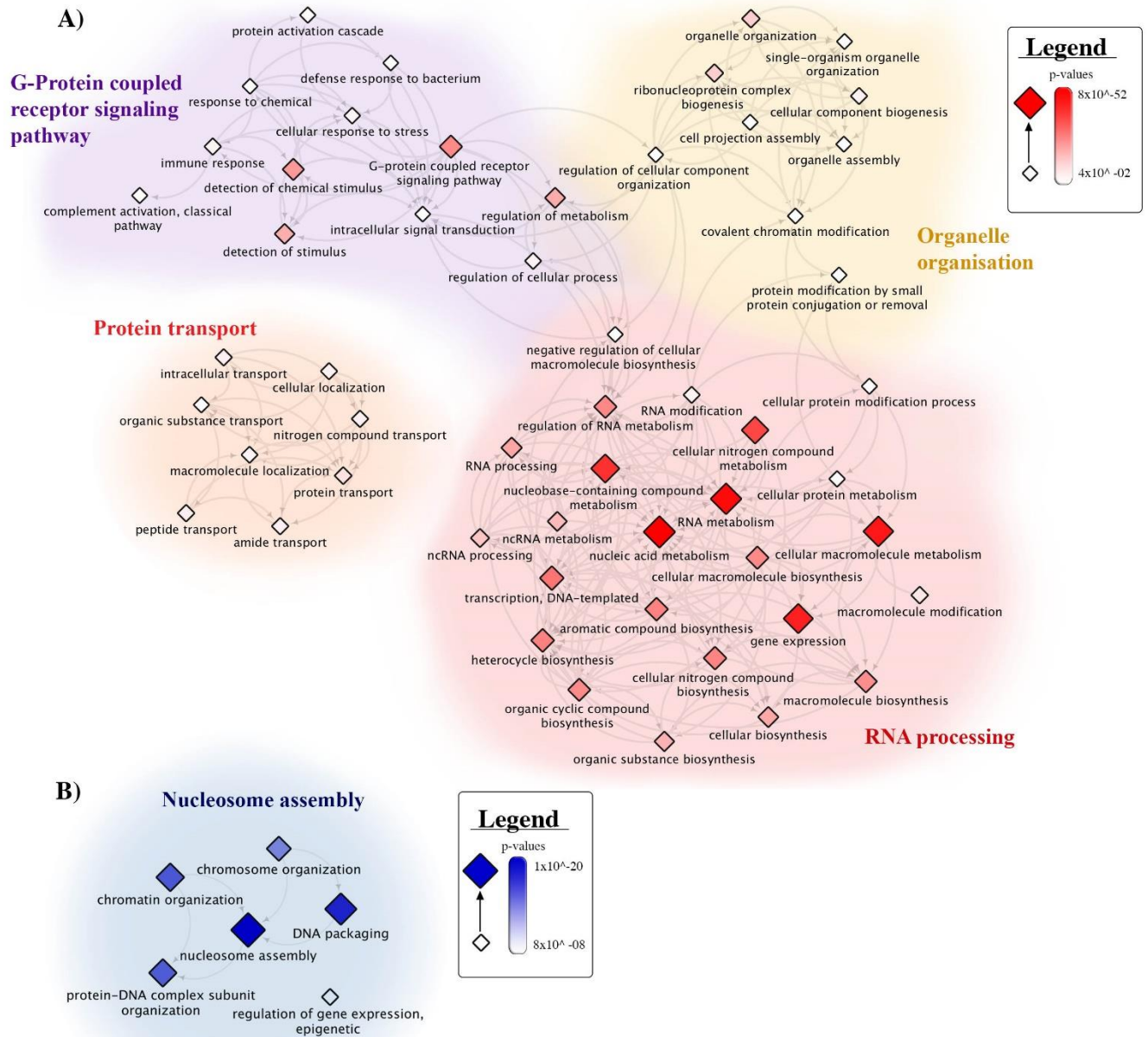


Figure 3.6 Biological process interaction network of all affected pathways dysregulated by HDV accumulation. **A)** GO enrichment analysis of all the genes that had no expression difference with the antigen but were upregulated in HDV replicating cells. GO analysis was performed on upregulated genes (Ag=,HDV↑), REVIGO was used to summarize GO terms and cluster were visualized using Cytoscape. Each node corresponds to a gene, where clusters represent genes that are part of the same biological process, and the edges between nodes represent a relationship between those paired. Four major clusters were enriched in this analysis: RNA processing, the G-protein coupled receptor signaling pathway, protein transport and organelle organisation (p-values < 0.05). **B)** The same analysis was performed for unchanged gene expression in the presence of the antigen alone but were downregulated in HDV replicating cells (Ag=,HDV↓). The only cluster enriched for downregulated genes was: nucleosome assembly. Node color intensities correlate with p-values, where the lowest p-values correspond to a more pronounced color and the highest p-values closer to a white color. Red nodes represent upregulated genes and blue represent downregulated genes. Node sizes are proportional to p-value, where the lower p-values correspond to larger sizes.

3.3 Weighted Correlation Network Analysis (WGCNA)

In order to confirm the previous results, a second analysis was performed using a method which has already proved its worth in the bioinformatics field. Weighted gene correlation network analysis (WGCNA) is a statistic approach designed for big data analysis and allows to performed network modeling base on gene expression correlation (194). It clusters genes by expression pattern, identifies highly-connected hubs and creates modules which represent groups of co-expressed genes with similar expression profiles. WGCNA make correlations between individual pairs of genes, but also to genes that share the same neighbors. It then creates hierarchical clustering, creating a dendrogram that clusters similarly expressed genes into discrete branches and where those branches can be cluster in different modules (194).

The method previously used for classifying gene is considered as a signed analysis (section 3.2). A signed analysis takes into consideration negative and positive values and processes the data separately while an unsigned analysis takes the absolute fold change values and thus identify genes having similar profile of expression independently of a negative or positive correlation values. Here, we designed both a signed and unsigned network analysis using the same data generated from the RNA-seq using the six conditions mentioned previously. The WGCNA was performed on gene affected by HDV replication (cutoff 2). A total of 6 modules were generated with the WGCNA for the signed analysis (**Figure 3.7A**). A GO enrichment analysis was performed on genes clustered in the different modules generated by the WGCNA analysis (**Figure 3.7B-D**). The largest module, which includes the vast majority of genes with similar expression pattern, is represented by a turquoise color (**Figure 3.7A**). The enrichment performed on this particular module generated four major biological

processes significantly enriched: RNA processing, cellular metabolism, detection of chemical stimulus in sensory perception and protein transport (**Figure 3.7B**). The other modules in this signed analysis contained fewer genes with similar pattern of expression, and thus the GO analysis performed on those genes either gave no enrichment or very little. The enrichment performed on the brown module resulted in three major biological processes enriched: Chromatin organization, gene silencing and immune response (**Figure 3.7C**). For the blue category, only one cluster was enriched: nucleosome assembly (**Figure 3.7D**). Those results are similar to those obtained previously with the in-house script analysis (**Figure 3.6**) since they both share RNA processing and protein transport, which seems to be sharing a similar correlation in expression. Although this analysis did not find GPCR as being part of the processes affected. It is interesting to observe the difference between these two distinct way gene classification. For instance, in the previous analysis nucleosome assembly and chromatin organization were both classified as being downregulated, however with WGCNA it is possible to differentiate between two different subclasses of downregulated genes which possess different pattern of expression.

The WGCNA analysis was repeated on the same subset of genes affected by HDV although using an unsigned analysis, thus using the absolute fold change value. A total of four-modules were clustered (**Figure 3.8C**) and among them, only one had GO terms enriched (turquoise module). The principal biological processes enriched were: RNA processing, G-protein coupled receptor signaling pathway, cellular metabolism, organelle organisation, and protein transport and some smaller cluster (**Figure 3.8D**). The results obtained in those analysis were very similar, almost identical to those found previously (**Figure 3.6**). Those independent analysis, using WGCNA, validated the results previously obtained using an in-house script, although the scripts I generated allowed to have better control on the analyzes.

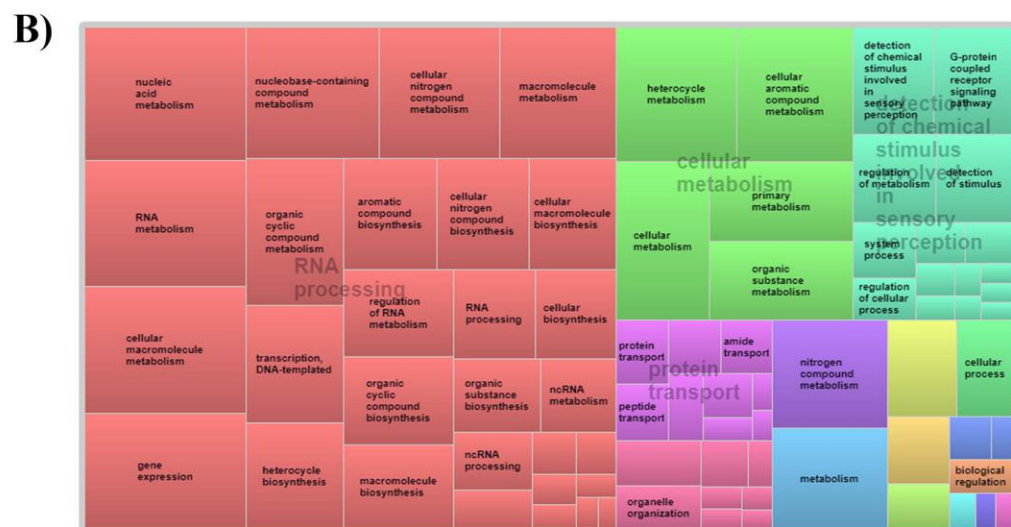
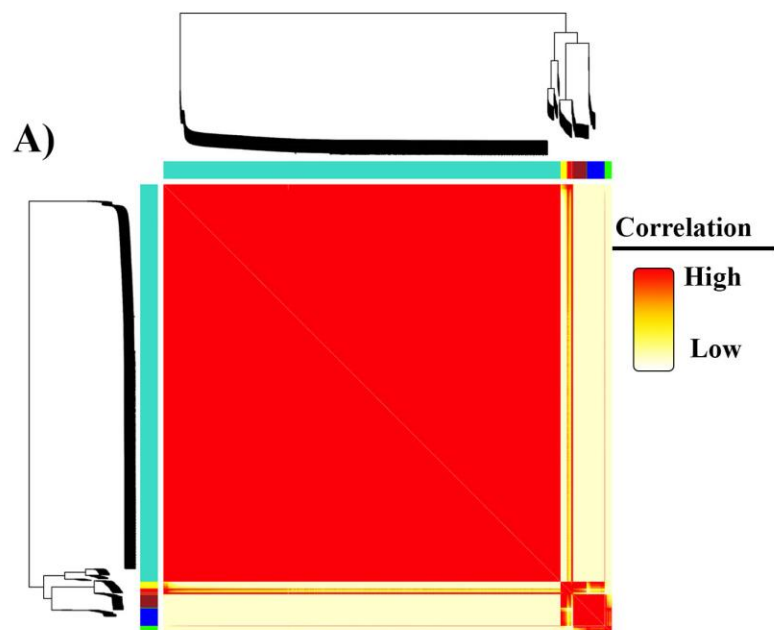


Figure 3.7. Signed weighted gene correlation network analysis (WGCNA) of genes affected by HDV induction.

A) Genes affected by HDV replication (cutoff 2) analyzed using WGCNA signed analysis and genes were clustered by expression pattern, which are represented with a dendrogram and a correlation network heatmap. The red color in the heatmap represents a high correlation between gene expression where yellow represent low correlation. Genes with a similar expression pattern were clustered are represented as color modules: turquoise, brown, blue, yellow, red and green. **B)** Go enrichment analysis performed on genes enriched in the turquoise module. Four major biological processes were enriched: RNA processing, cellular metabolism, detection of chemical stimulus in sensory perception and protein transport. **C)** Go enrichment analysis performed on genes enriched in brown module. Three major biological processes were enriched: Chromatin organization, gene silencing, and immune response. **D)** Go enrichment analysis performed on genes enriched in blue module. Only one biological process was enriched: nucleosome assembly

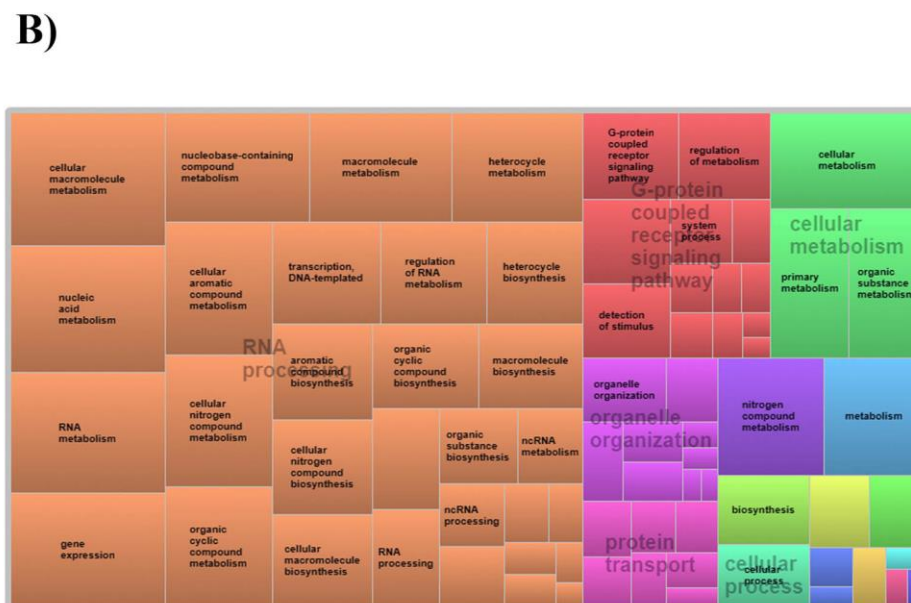
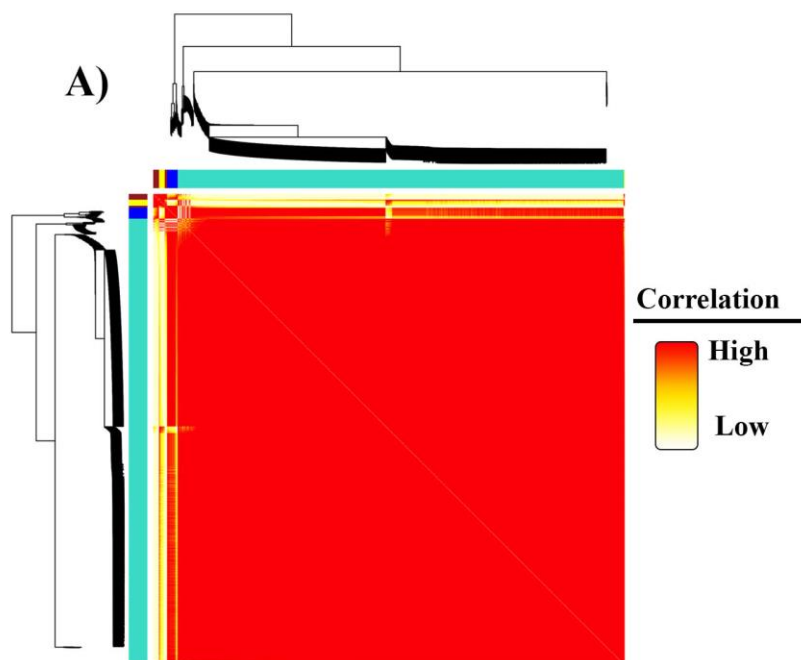


Figure 3.8. Unsigned weighted gene correlation network analysis (WGCNA) of genes affected with HDV induction.

A) Genes were clustered by expression pattern and are represented by a dendrogram and a correlation network heatmap. The red color in the heatmap represent a high correlation between gene expression where yellow represent low correlation. Genes with a similar expression pattern were clustered are represented as color module (turquoise, blue, red, yellow, and brown. **B)** Go enrichment analysis performed on genes enriched in the turquoise module. Five major biological processes were enriched: RNA processing, G-protein coupled receptor signaling pathway, cellular metabolism, organelle organisation, and protein transport.

3.4 HDV replication disrupts major protein complexes

Transcriptomics analysis can measure the abundance of RNA transcripts present in a cell, which can also give a good overview of a perturbation in the corresponding protein level. Even if the correlation between the amount of RNA transcribed and the resultant protein produce are not always representative, it can give an overview of what is affected in the cell (195, 196). Using our RNA expression data, I wanted to look at RNA transcripts expression of genes that are involved in protein complexes. Proteins in a cell usually work in coordination with other proteins to achieve a certain function. If only one protein in the complex is affected, this might not affect the function of the overall complex function, but if many proteins in a complex are dysregulated, there is a greater chance that this complex and its function is affected in the cell. To the best of my knowledge, this “complex-centric” approach has not been applied to previous transcriptomic results in the HDV field. Therefore, a personalized approach to look at affected complexes was developed. To identify a pattern in genes encoding for protein complexes that may be affected by HDV accumulation, the CORUM protein complex database was used to obtain protein complexes. The core complexes were used for the analysis, which consists of a reduced dataset of protein complexes free of redundant entries. Subsequent analyses were done in R using customized scripts (Figure S3). The core complex dataset provided by CORUM included 2,693 different complexes originating from different organisms. I sorted those complexes by selecting only complexes of human origin which limited the analysis to 1,915 entries. The UniProt entries provided by CORUM was converted to use the Ensembl annotation used in the previous analysis. In order to identify affected complexes, expression values for each of the genes were added in the corresponding gene in the dataset. A two-step analysis was used to look at complex expressions. First, a cutoff of a 1.5-fold change was used to distinguish affected and non-affected gene expression.

This cutoff was decided after many trial and error to test which cutoff made it possible to discern genes that have a change of expression without being too stringent in the analysis. To ensure that we were not being too rigid with our cut off values, I carried out a second analysis. Complexes which had 75% or more of their proteins affected at the RNA level were selected for a second analysis. Using a lower cutoff of 1.2 on the remaining 25% of “non affected” proteins of these complexes, it allowed to include those whose RNA fold change expression slightly change between 1.5-1.2 fold change with the virus replication. The number of proteins present in each complex as well as the percentage of these protein affected, using the cut off mentioned above, were calculated for each complex. P-values of each complex were calculated using the Fisher's exact test. Only the complexes with at least 3 proteins with p-value ≤ 0.05 were retained. The analysis was performed using both an unsigned and a signed approach as described in section 3.3. The unsigned analysis allows looking at all the complexes that are affected independently of whether genes were up/downregulated within the same complex. A total of 30 complexes were identified as significantly affected following HDV replication for the unsigned approach (**Figure 3.9A & Figure S4**). With an average of 12 proteins per complex, the smallest complex contained only 3 proteins while the largest had 78 proteins. A gene ontology enrichment analysis was performed on the GO terms associated with the affected complexes (**Figure 3.9B**). A total of 5 clusters were enriched including DNA-template transcription and elongation, chromosome segregation, chromatin remodeling, cellular response to DNA damage stimulus and synaptic vesicle transport. As for the signed analysis, it processes positive and negative values separately and thus can identify complexes where most of the members are impacted in the same manner. A total of 17 complexes were identified in the signed analysis (**Figure 3.10A**). The GO enrichment for those 17 complexes delivered similar results with 4 clusters enriched: chromosome organization, chromosome

segregation, cellular response to DNA damage stimulus and protein phosphorylation (**Figure 3.10B**). In both experiments, the ribosome complex was the biggest complex affected with 41% out of 78 protein.

A)

SNARE complex II	5	100 %	1.6e-03
H2A-H2B-TDIF2-PCNA	4	100 %	5.8e-03
LINC core	4	100 %	5.8e-03
CDC2-PCNA-CCNB1-GADD45G	4	100 %	5.8e-03
CDC2-PCNA-CCNB1-GADD45A	4	100 %	5.8e-03
Ubiquitin E3 ligase II	4	100 %	5.8e-03
SNARE complex	4	100 %	5.8e-03
Mitotic checkpoint	4	100 %	5.8e-03
Ubiquitin E3 ligase	4	100 %	5.8e-03
Chromosomal passenger complex CPC	3	100 %	2.1e-02
IRF2BP2-IRF2BP1-IRF2BPL	3	100 %	2.1e-02
TAK1	3	100 %	2.1e-02
CTGF/Hcs24-actin	3	100 %	2.1e-02
NDC80 kinetochore	3	100 %	2.1e-02
Codanin-1-Asf1	22	95 %	2.5e-11
core-histone-TDIF2-TDT	22	95 %	2.5e-11
Histone H3.1	25	92 %	1.8e-11
Histone H3.3	20	90 %	8e-09
H2AX complex II	20	80 %	1.5e-06
ASF1-interacting protein complex	11	64 %	1.3e-02
Emerin regulatory	13	62 %	1.1e-02
ASF1-histone containing complex	13	62 %	1.1e-02
TLE-Histone H3	8	62 %	4.1e-02
B-WICH	10	60 %	3.2e-02
Intraflagellar transport complex B	10	60 %	3.2e-02
H2AX complex	29	59 %	4.2e-04
TNF-alpha/NF-kappa B signaling I	12	58 %	2.5e-02
TNF-alpha/NF-kappa B signaling complex II	12	58 %	2.5e-02
RC complex	13	54 %	4.1e-02
Ribosome	78	41 %	6.6e-03

NbProt

Percentage Values

p_values

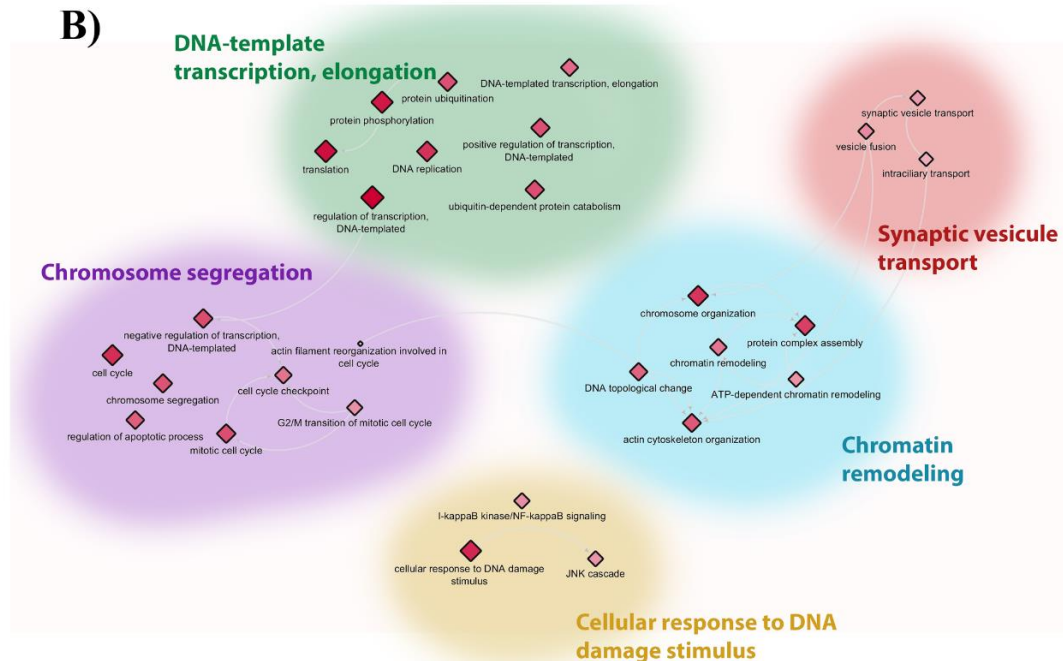


Figure 3.9. List of complexes affected by HDV replication using an unsigned analysis. A)

The complete list of complexes that were significantly affected by HDV, considering the absolute values of expression. The “NbProt” column represents the number of proteins present in the complex (≥ 3), the column “%” represents the percentage of protein affected in the complex, and the p-value column represents the P-value associated with the complex (≤ 0.05).

B) GO enrichment analysis of the affected complexes. A total of 91 GO terms were associated with the affected complex and gave 5 major enriched clusters. Node color intensities correlate with p-values, where the lowest p-values correspond to darker color and the highest p-values to a lighter color. Node sizes are also proportional to p-value, where the lower p-values correspond to larger sizes.

A)

Complex name	Mitotic checkpoint	4	100 %	5.8e-03
	NDC80 kinetochore	3	100 %	2.1e-02
	IRF2BP2-IRF2BP1-IRF2BPL	3	100 %	2.1e-02
	TAK1	3	100 %	2.1e-02
	Chromosomal passenger complex CPC	3	100 %	2.1e-02
	core-histone-TDIF2-TDT	22	82 %	1.7e-07
	Codanin-1-Asf1	22	73 %	1.3e-05
	Histone H3.1	25	72 %	4.6e-06
	H2AX complex II	20	70 %	9.3e-05
	Histone H3.3	20	70 %	9.3e-05
	TLE-Histone H3	8	62 %	4.1e-02
	B-WICH	10	60 %	3.2e-02
	TNF-alpha/NF-kappa B signaling	12	58 %	2.5e-02
	H2AX complex I	29	55 %	1.5e-03
	Emerin regulatory	13	54 %	4.1e-02
	ASF1-histone	13	54 %	4.1e-02
	Ribosome	78	41 %	6.6e-03
		NbProt	Percentage Values	p_values

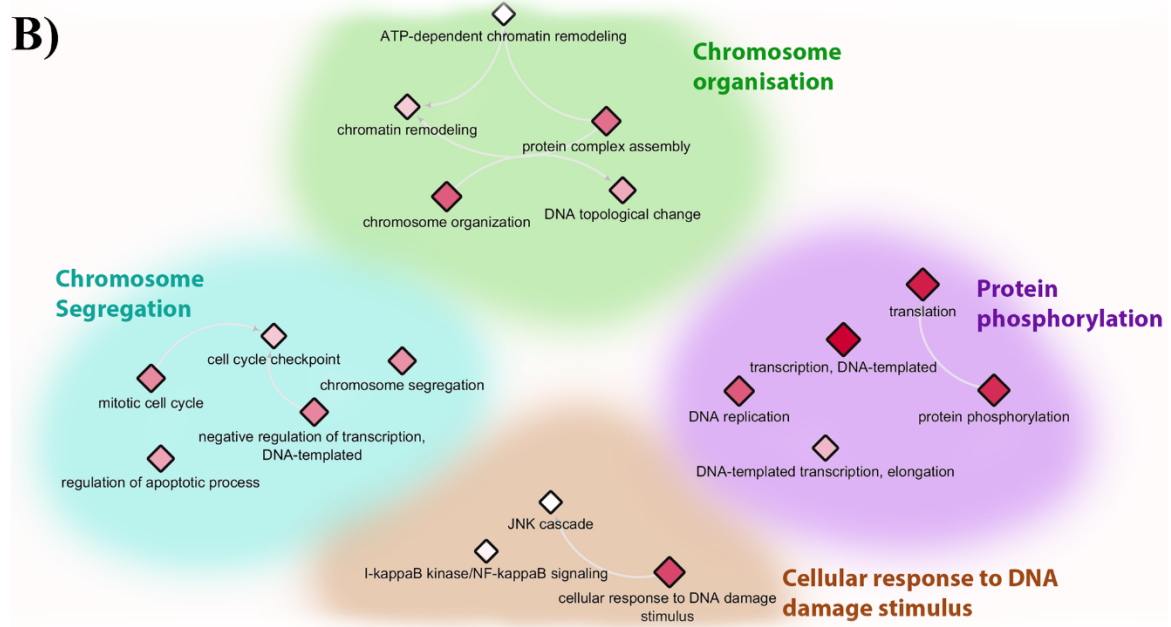


Figure 3.10. List of complexes affected by HDV replication using a signed analysis.

A) The complete list of complexes that were significantly affected by HDV where all the gene members were affected by HDV in the same manner (signed analysis). The “NbProt” column represents the number of proteins present in the complex (≥ 3), the column “%” represents the percentage of protein affected in the complex, and the p-value column represents the P-value associated with the complex (≤ 0.05). **B)** GO enrichment analysis of the affected complexes. A total of 52 GO terms were associated with the affected complex and gave 4 major enriched clusters. Node color intensities correlate with p-values, where the lowest p-values correspond to darker color and the highest p-values to a lighter color. Node sizes are also proportional to p-value, where the lower p-values correspond to larger sizes.

3.5 Flow cytometry reveals an arrest of the cell cycle and morphological differences for HDV induced cells

Previously a proteomic analysis performed by *Mendes et al.*, revealed a possible cell cycle deregulation at the G2/M DNA damage checkpoint (158). Likewise, these results supports findings of cell cycle disruption by HDV, but does not agree with previous results obtained by Taylor *et al.* (153). Their results displayed a 5-fold increase of relative cell number in G1/G0 phase and a reduction in S and G2/M using flow cytometry with propidium iodide (PI) staining. They also observed a 60% decrease in total number of cell growth compared to uninduced cells, as well as a significant loss of adhesion. The authors suggest that this loss of cell adhesion can lead to apoptotic cell death although staining with annexin V as an apoptotic detection method did not show any difference for adherent vs non-adherent cells. Based on these two independent observations of a disruption of the cell cycle, those results suggest a cell cycle perturbation but there is still many questions left to clarify.

In order to clarify those previous findings, a flow cytometry analysis was performed to look at the cellular cycle. Although several methods are available to observe the cell cycle, Propidium Iodide (PI) was selected to observe a possible dysregulation of the cell cycle. The PI method is based on the analysis of cellular DNA content. PI dye binds to DNA and the intensity of fluorescence is proportional to the amount of DNA in the cell. Thus, this method allows distinguishing cells between G0/G1 vs S vs G2/M, where cells in G2/M have double the DNA content of cells in G0/G1 and cells in S phase have an intermediate amount of DNA. The first experiment was to validate the cell cycle arrest observed by Taylor *et al* (153). HEK-293, 293-Ag, and 293-HDV cells were cultured in duplicate, one of which was induced with TET addition and the other non induced. Approximately 1 million cells were collected 24h post-induction for flow cytometry analysis. Cells were fixed and stained using PI dye and

analyzed using LSR Fortessa cell analyzer and data were further analyzed using custom R script. First, debris and cell doublets were excluded from the analysis using gating strategy. The peak of the spectrum in G0/G1 was aligned between all the different conditions followed by a rescaling of the number of cell analyses in all sample to reported data on the same basis. A comparison between the non-induced and induced 293-HDV cells 24h after induction showed a 15.5% increase of the number of cells in G0/G1 as well as a 14% reduction in the number of cells in S phase (**Figure 3.11A&B**). The experiment was repeated with different induction timepoints (12h, 24h, and 36h) (**Figure 3.11C&D** and **Figure 3.12**). At 12 hours after induction, there was no change in the number of cells present in G0/G1 (**Figure 3.11C&D**). However, an 16.6% increase in the number of cells present in G0/G1 phase was observed 24h post induction (**Figure 3.12**). This accumulation seems to reach a plateau after 36h since there is very little change between 24h (16.6%) and 36h (24%). A 15.2% reduction in cells present in S phase was observed after 24h and a 20% reduction after 36h, which seems to agree with the trends observed in the G0/G1 phase but inversely proportional. The percentage of cells in G2/M did not significantly changed over 36h experimental period compared to what is observed in the other phases.

Flow cytometry is a single cell analysis approach that uses different sets of lasers and that allows to characterize a cells physical property as well as its interior complexity. The forward scatter (FSC) measure the amount of light passing around a cell and this light intensity is proportional to the cell diameter and thus allows for cell sizes to be measured and allows to sort and remove cells clumps from the analysis. Another scatter, the side scatter, measures the quantity of light that is refracted or reflected due to intracellular structure like the nucleus or other granules and can therefore provide information about cell internal complexity such as granulocyt. The comparison of HDV non-induced population with the induced one showed a

reduction in cell size and a decrease in granulocytosis 24h after induction (**Figure 3.13A**). Cells were also analyzed separately according to their stage of the cell cycle and only the cells in G0/G1 possessed those atypical characteristics (**Figure 3.13B,C,D**).

Prior to removing cells doublets or clumps from the analysis, we realized there was an increase number of doublets with HDV replication (**Figure 3.14A**). Therefore, after the removal of cellular debris from the analysis, we used the forward scatter height and area measurement with a manual threshold to separate these doublets from the rest of the cell populations and calculated the relative percentage of aggregate present for each sample. Each induced condition was compared to its non-induced corresponding condition, which was reported to 100 (**Figure 3.14B**). The number of cell aggregate increases more than 1.61x after only 12h and 1.72x and 1.76x after 24 and 36h respectively (**Table 3.3**). Moreover, the morphological changes are also observable using a phase contrast microscopy (**Figure 3.15**) where you can observe smaller and rounder cells which forms more aggregates. This phenomenon could be explainable by a cellular stress induced by the replication of HDV, as it has been demonstrated in Beeharry *et al.* 2018 (127).

Table 3.3 Percentage of cellular aggregates in HEK-293, 293-Ag and HDV cells without and with tetracycline induction at different time point

	% Aggregates		Fold change
	- TET	+ TET	
293 (24h)	5.17	4.98	0.96
Ag (24h)	2.98	2.58	0.87
HDV (12h)	2.57	4.15	1.61
HDV (24h)	4.1	7.07	1.72
HDV (36h)	4.46	7.88	1.77

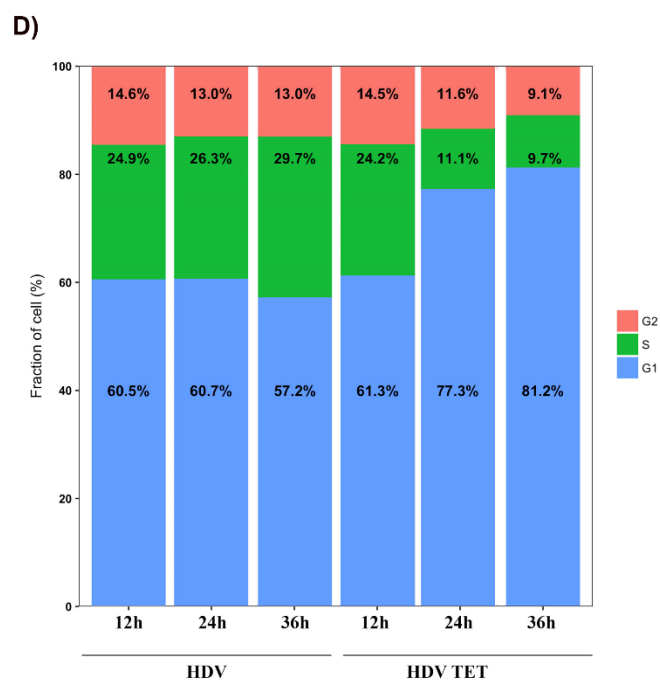
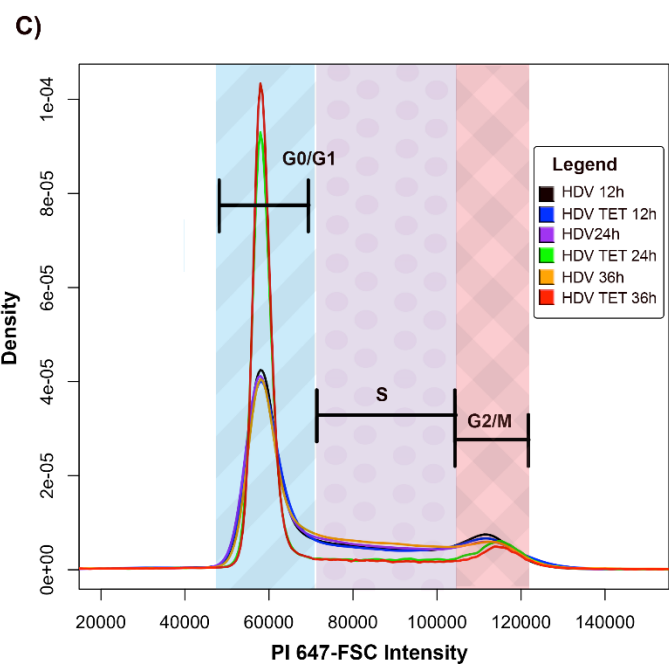
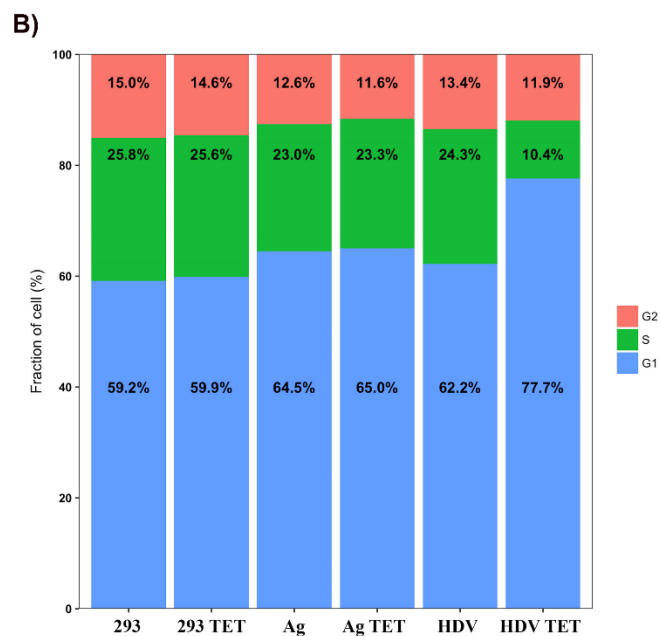
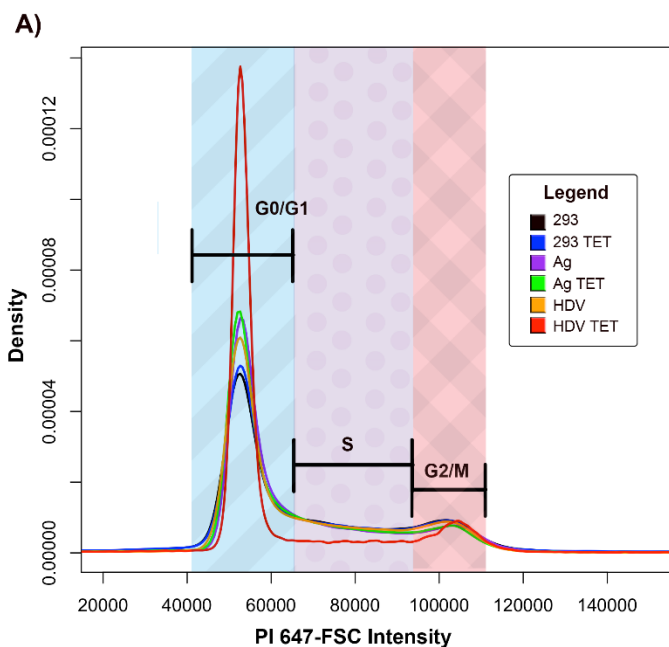


Figure 3.11. Flow cytometry analysis of cell cycle phase. A-B) Flow cytometry using HEK-293, 293-Ag and 293-HDV cells induced and non induced with 1mg/mL of tetracycline and harvest after 24h. **A)** Represent the distribution of cells present in G0/G1, S, and G2-M phase for HEK-293, 293-Ag and 293-HDV cells. Cells were fixed and stained with propidium iodide (PI) were the DNA content was analysed by flow cytometry and plot on a histogram. **B)** The percent of cells distribution in each phase, which was analyzed using R software and in-house scripts. **C-D)** Flow cytometry using 293-HDV cells induced and non induced and harvest at different time points: 12h, 24h, 36h. **C)** Represent the distribution of cells present in the different cycle phase : G0/G1, S, and G2-M. **D)** The percent of cells distribution in each phase according to different time points.

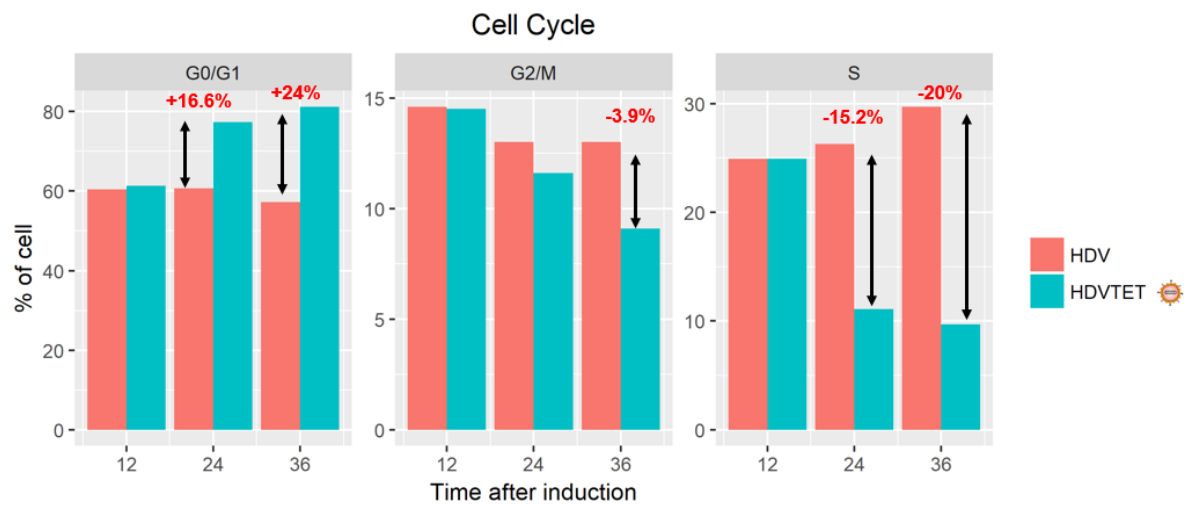


Figure 3.12. Changes in the cell distribution of different phases of the cell cycle between HDV and HDV TET after 12, 24 and 36 hours.

The percentage of cell present in each phase of the cell cycle was access between 293-HDV cell with and without induction at three different time points after tetracycline induction: 12h, 24h and 36h.

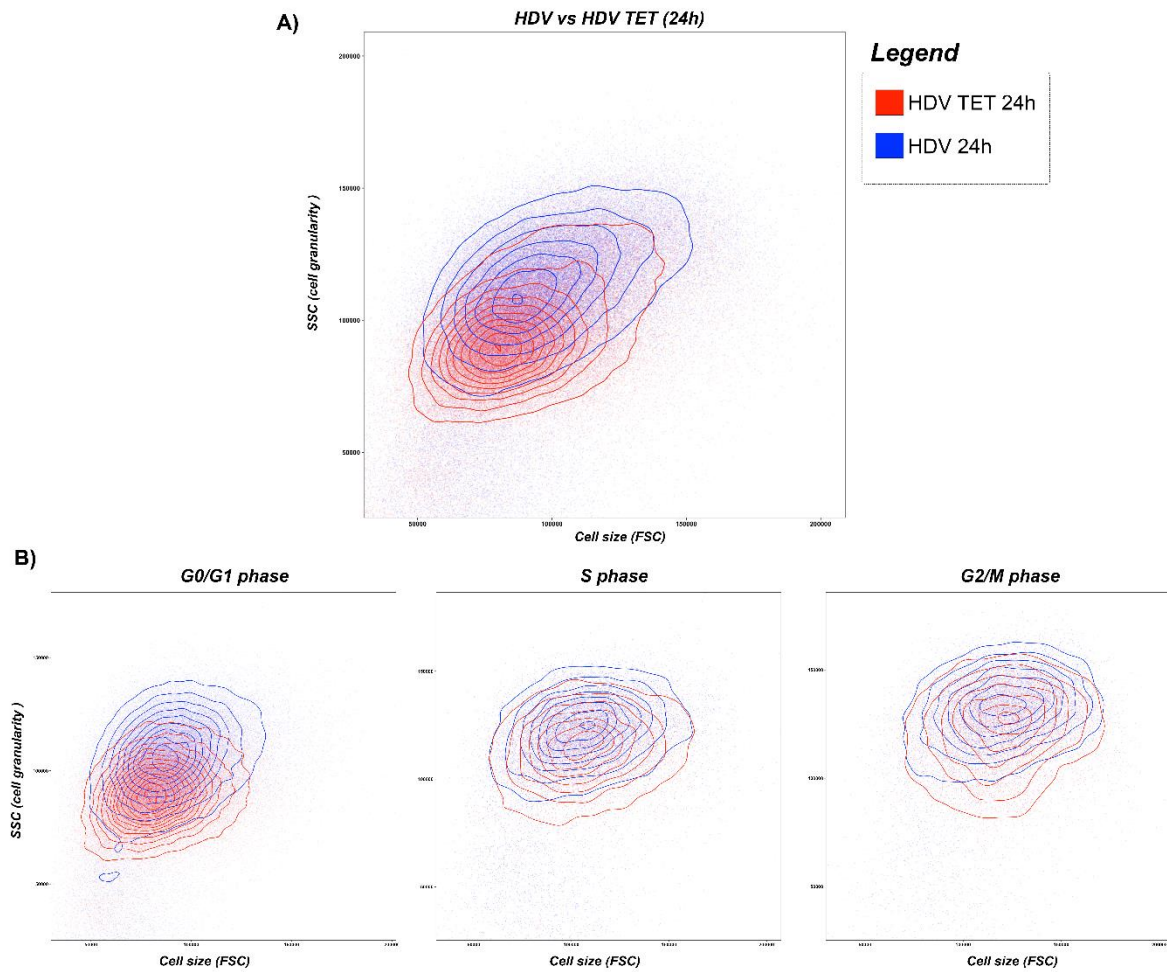


Figure 3.13. Flow cytometry analysis of 293-HDV cells induced and 293-HDV cells non induced after 24h. A) View of the cell size (FSC) and granulocyte (SSC-H) of non-induced and TET induced 293-HDV cells harvested after 24h. All phases of the cell cycle are included. **B)** Cells from A) separated according to their cell cycle phase: G0/G1, S, and G2-M.

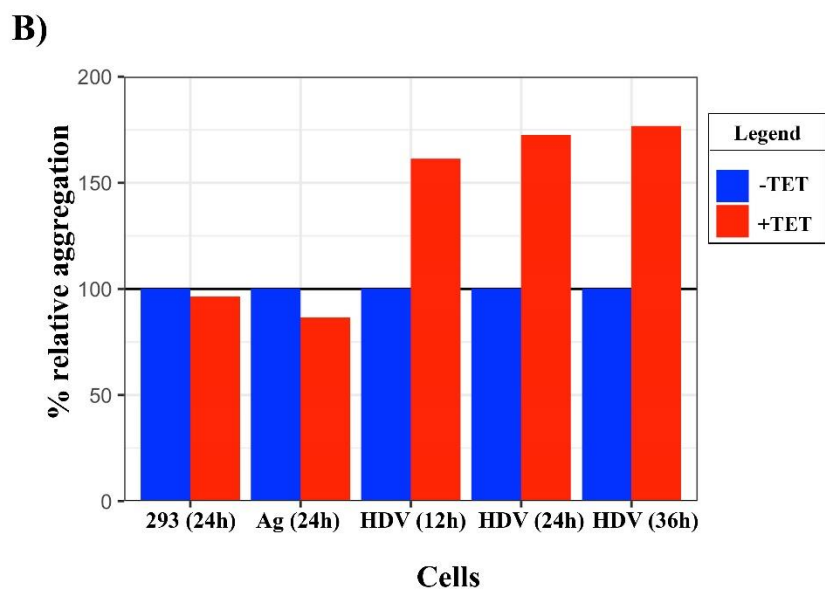
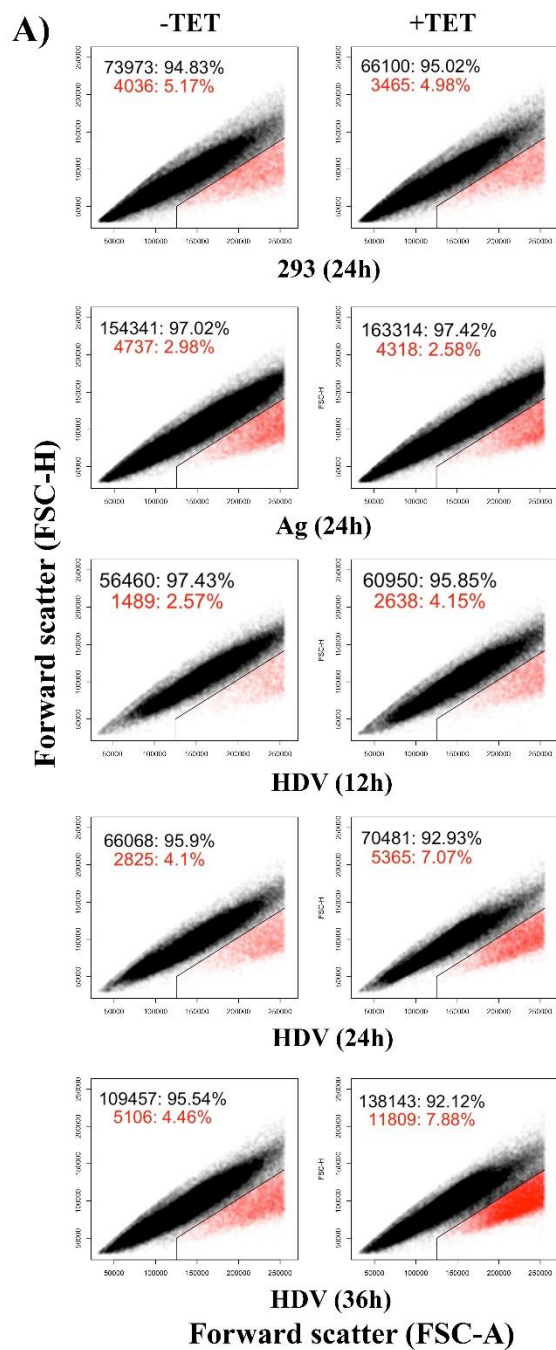


Figure 3.14. Density plot of cellular aggregate in HEK-293, 293-Ag and HDV cells without and with tetracycline induction at different time point. A) Two-parameter density plot using forward scatter height (FSC-H) versus forward scatter area (FSC-A) plot for doublet identification. HEK-293 (293), 293-Ag (Ag), and 293-HDV (HDV) were cultures with (+TET) and without tetracycline (-TET) and harvested at different time points. HEK-293 and 293-Ag cells were harvested 24h after TET induction where 293-HDV samples were revolted 12h, 24h, and 36h post-induction. The same threshold used to eliminate subcellular doublet was used to identify and calculate the percent of doublets present in the different analyzes. Cells doublets are represented in red color where the cells in black represent other cells in the analysis excluding debris previously removed from the analysis. **B)** Bar chart of the relative percentage of aggregate present in every condition in A). Blue bars represent the -TET cell conditions where red bars represent +TET conditions. The graphs were calibrated to have conditions control (-TET) reported to 100 to be able to compare their corresponding induced condition and therefore to have a proportional comparison.

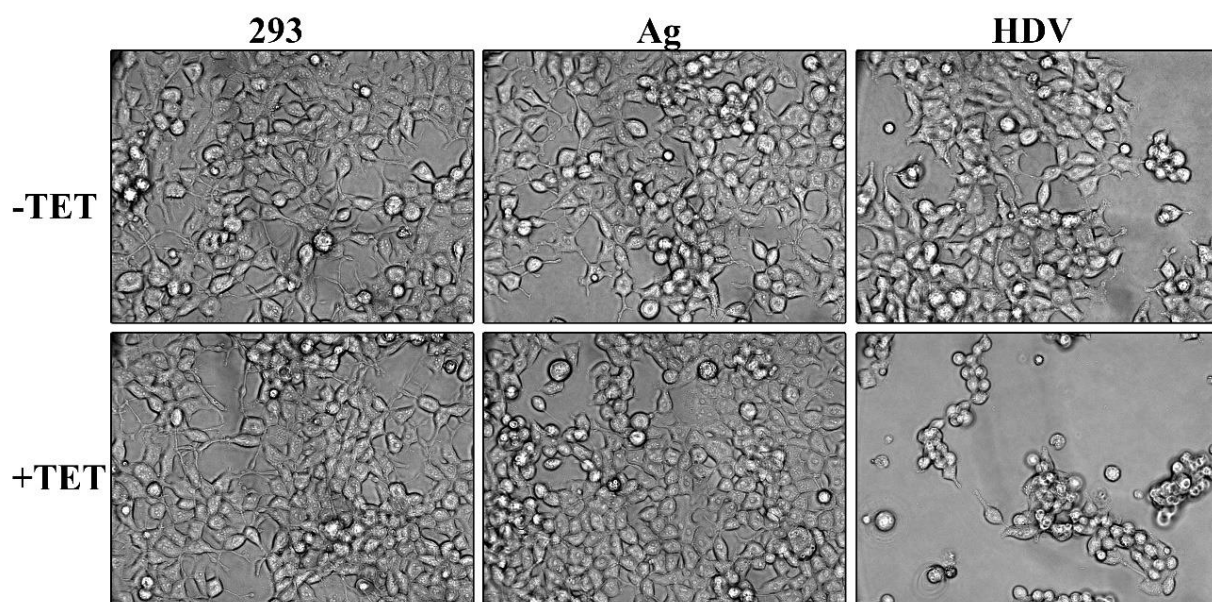


Figure 3.15 Phase contrast comparison of non-induced and induced HEK-293, 293-Ag and 293-HDV cells. The three cell lines HEK-293(293), 293-Ag (Ag), and 293-HDV (HDV) were plated in duplicates at a concentration of 1×10^4 cells and incubated for 24h. Each cell type was then either cultured in absence (-TET) or presence (+TET) of $1 \mu\text{g/ml}$ of tetracycline for 48h. The images were taken using regular light phase of contrast microscopy with a 40x objective.

CHAPTER 4: DISCUSSION

4.1 Summary of findings

In order to observe the effect of HDV replication on host gene expression, we used a HEK-293-based cell system engineered to mimic HDV replication. A high throughput sequencing was performed on HEK-293, 293-Ag, and 293-HDV cells, which were cultured in duplicates, one of which was induced with TET and the other was not. The sequencing results gave an average number of 9,780 million bases sequenced per sample and a total number of 235 million reads. Raw sequencing reads were processed to obtain the expression changes between our six different conditions. By comparing the number of sequencing reads mapped to HDV genome, we determined that HDV induction allowed an increase of ~23 time the amount of HDV RNA produced in induced cells compared to non induced. Genes were further classified into nine different categories according to their expression change. To validate those expressions, RT-qPCR experiments were performed on selected genes. A coefficient of determination of 0.9123 ($R^2 = 0.9123$) was determined between RNA-seq and qPCR results indicated that the RNA-seq data correlates well with validated the RT-qPCR data.

A total of 3,782 genes were differentially expressed and among those, 3,561 genes were affected by HDV RNA alone, 123 were affected by the antigens only and 98 genes were affected by both. Among the genes affected by HDV RNA, 3,278 were up-regulated and 283 down-regulated. Upregulated genes were predominantly part of the following pathways: RNA processing, G-protein coupled receptor signaling pathway, protein transport, and organelle organization. Downregulated genes were shown to be part of the nucleosome assembly pathway. Moreover, 498 genes related to cellular stresses were affected by HDV induction.

In order to confirm the previous results, an independent method was used to confirm the results obtained by our in-house scripts. This second analysis used the WGCNA method was performed using a signed and unsigned analysis. The signed analysis performed on genes affected by HDV replication using a 2-fold change cutoff gave rise to the creation of six modules of genes with similar expression profiles. The largest module, which includes the vast majority of genes with similar expression pattern, was mainly composed of genes part of the following biological processes: RNA processing, cellular metabolism, detection of chemical stimulus in sensory perception and protein transport. Another module resulted in three major biological processes enriched: Chromatin organization, gene silencing and immune response and for the blue module only one cluster was enriched: nucleosome assembly. The results obtained by the signed analysis were similar to those obtained previously with the in-house scripts since they both share RNA processing and protein transport. On the other hand, the unsigned WGCNA analysis generated four modules and only one had enriched GO terms, which were involved in RNA processing, G-protein coupled receptor signaling pathway, cellular metabolism, organelle organisation, and protein transport and some smaller cluster. Those results were very similar, almost identical to those found previously.

A novel approach to identify a pattern in genes encoding for protein complexes was developed. Using an unsigned analysis, I found a total of 30 complexes which were affected by HDV replication. Those complexes were predominantly part of these pathways: DNA-template transcription and elongation, chromosome segregation, chromatin remodeling. With an average of 12 proteins per complex, the smallest complex contained only three proteins while the largest had 78 proteins. Whereas the same analysis performed using a signed method allowed to identify a total of 17 affected complexes mainly part of chromosome organization,

chromosome segregation, cellular response to DNA damage stimulus and protein phosphorylation.

Lastly, using a flow cytometry approach, I found that cells replicating HDV were accumulating in G0/G1. An 15.5% increase of the number of cells in G0/G1 was observed in HDV replicating cells 24h after induction and a 14% reduction in the number of cells in S phase. Using different time points, we observed a 0.8% increase of the number of cells in G0/G1 after 12h, 16.6% increase after 24h and 24% after 36h. While the trend for phase S appeared to be inversely proportional to phase G1/G0, where a 15.2% reduction was observed after 24h and a 20% reduction after 36h. Flow cytometry experiment also showed a reduction in cell size and a decrease in granulocytometry after 24h of induction. This difference in morphology was only observable for cells in G0/G1. Moreover, an increase in the number of cell aggregate was observed in HDV induced cells and cells were smaller and rounder. Those morphologic characteristics were also observable using a phase contrast microscopy.

HDV replication affect the host transcriptomic landscape

HDV highly depends on host components to achieve its replication and transcription and was shown to interact with many host proteins (109). Previous studies have shown that HDV disrupts host cellular proteome (152, 157, 158). Thus, I wanted to investigate how the virus replication impacted host gene expression using a transcriptomic approach. Although HDV usually infects hepatocyte cells and requires HBV for infection, I used a model where HBV was not involved to observe solely HDV effects. In this model, the small antigen is produced in *trans* by a tetracycline-inducible promoter while HDV, which was introduced by transfection, is dependant on the antigen induction to initiate its replication. HDV replication was induced and the RNA was harvested after 48h and sent for high-throughput RNA sequencing. Following the sequencing, gene expression accessed and those results were validated using RT-qPCR on selected genes (**Table 3.2**). Among the genes chosen, the overexpression of CDT1 and MCM2 was validated. They are both involved in the formation of the pre-replication complex necessary for DNA replication. CDT1 and MCM2 expression was upregulated in the presence of HDV with an upregulation of 3.44x for CDT1 and 1.65x for MCM2 in our RNA-seq data and our RT-qPCR data showed a 5.95x and 1.78x increase respectively. The downregulation of Aurora kinase B (AURKB), a mitotic checkpoint kinase that plays a crucial role in the cell cycle was also validated by RT-qPCR showing a 2.84x decrease in our RNA-seq data and a 1.9x decrease in RT-qPCR. Several other genes involved in different pathways have been confirmed (**Figure 3.5A & Table 3.2**). The correlation between the RNA-seq and RT-qPCR data was 0.912 which was excellent and indicated that I could rely on these values with confidence (**Figure 3.5**). Therefore, I classified the genes into

categories according to their expression change and identified the most affected pathways for each different category (**Figure 3.6**). A Gene Ontology enrichment analysis revealed four principal clusters affected positively by HDV RNA: RNA processing, G-protein coupled receptor signaling pathway, protein transport, and organelle organization (**Figure 3.6A**). Only one single cluster was for downregulated genes: nucleosome assembly (**Figure 3.6B**).

The largest cluster affected by HDV upregulated gene was the RNA processing, which englobes many processes (**Figure 3.6**). Among these processes, RNA metabolism, RNA processing, and RNA modification were upregulated. These processes are involved in the conversion of primary RNA transcripts into mature RNA molecules. As we know, HDAG-S interacts with at least 9 of the 12 subunits of the polymerase II complex (75) which is used for HDV replication and HDV mRNA also possess typical features of transcripts generated by the RNAP II including the a 5'-cap and a 3'-poly(A) tail in addition to associate with HDV RNA (64). Both polarities of HDV RNA also interact with RNAP I and RNAP III (58, 68). Since HDV highjacks these host polymerases complexes to perform its own replication, either the host cell increases the production of polymerase to keep producing of its own RNA or the highjacking will negatively impact host transcripts production. HDV also interacts with PSF and p54rnb which is a protein involved in processes like transcription and pre-mRNA splicing (132). Moreover, HDV interacts with ASF/SF2, also implicated in splicing (128, 197). Those HDV interacting protein might influence the upregulation observed in the genes involved in RNA metabolism and processing. Other viruses, such as the influenza A virus, have been shown to highjack the host splicing machinery to produce different mRNA transcripts essential for viral replication (198). Despite those upregulated processes being advantageous for HDV replication, those processes probably negatively affect the host by disrupting its own RNA production and RNA processing.

Overall, there seems to be a tendency for an increase in production of macromolecule building blocks. Organic substance biosynthesis and the nucleobase-containing compound metabolic processes were increased with HDV, which include any cellular metabolic process involving nucleobases, nucleosides, nucleotides and nucleic acids. This boost in basic molecular compound and macromolecule buildings blocks strategy has been exploited by other viruses, using various mechanisms such as by inducing quiescence, to generate more material required for their replication since the amount of available nucleotides is a limiting aspect for replication (199). This increase in resources as well as the increase in the cellular RNA metabolism, processing and modifications probably results to favor and to support exhaustive HDV replication.

The G-protein coupled receptor signaling pathway was the second largest cluster enriched in the GO enrichment analysis performed on upregulated genes. GPCRs are the largest family of cell-surface protein involved in signal transduction (200). They have a wide variety of ligands, and this external signal is converted into intracellular response through biochemical reactions and molecular interactions. GPCRs have many downstream targets and are known to play a role in cell growth, survival, differentiation, migration. GPCRs are also implicated in exocytosis, angiogenesis and neurotransmission. However aberrant activation of GPCRs and their downstream targets can result in tumor growth and development, it can lead an increase to angiogenesis and metastases of cancerous cells and help cancerous cells to evade the immune system. Other virus exploit GPCRs to their advantage like the human immunodeficiency virus (HIV) who uses GPCRs as co-receptors for viral entry, or like Kaposi's sarcoma-associated herpesvirus (KSHV) and Epstein–Barr virus (EBV) who encode their own GPCR, or virus like the human cytomegalovirus who highjack the host receptors (201, 202). In my analysis, stimulus detection was upregulated as well as intracellular signal

transduction and most importantly cellular response to stress (**Figure 3.6**). In a recent paper (Beecharry *et al.* 2017), we demonstrated that HDV replication causes a cellular stress (127). GPCRs can activate many different downstream signaling pathways and the signal transduction initiated by those receptors is not limited to a single signaling pathway. A phenomenon called crosstalk is often observed between signaling pathways such that the signal can activate different intracellular second messenger pathways (i.e those which share a common component that can activate either pathway) and those events play crucial roles in signaling cascades (203). For example, two MAP kinase signaling modules are activated by different type of cellular stresses: JNK and p38. The JNK pathway can be activated by a variety of stresses like oxidative stress and DNA damage stress. Those pathways have been shown to be activated by GPCR signaling (204). I hypothesize that there is a possibility that HDV components stimulated GPCRs using an unknown mechanism that induces the stress observed in 293-HDV cells. The reason why GPCRs related gene expressions are upregulated is unclear. It may be a direct consequence of HDV protein interactions, or an undirected effect observed because of the stress-induced by HDV replication.

Moreover, the transcriptomic analysis of cells with replicating HDV also detected several downregulated genes (**Figure 3.6B**). Among the 283 downregulated genes with HDV exclusively, only one cluster was enriched with the GO analysis: nucleosome assembly. In eukaryotes cells, genomic DNA is highly condensed into chromosome. Chromosomes are composed of DNA which is wound around histones proteins to form the nucleosome and this complex is called chromatin. Chromatin is composed of four histones core particles: H2A, H2B, H3 and H4 (205–207). These core particles assemble to form a histone octamer and the DNA is tightly wrapped along these structures which form the nucleosome. These nucleosomes are connected together with histones H1 which is important to stabilize

chromatin higher-order structure or heterochromatin (208, 209). Histone H1 is involved in the formation of higher order chromatin structures especially in the mitosis phase where it is condensed to the most extreme degree (210–213). Histone H1 stabilizes nucleosome structure and nucleosome spacing therefore plays a role in chromatin compaction. Histone proteins also contain an N-terminal amino acid tail which is subject to post-translational modifications and those changes modify their functions and structure (214). Disassembly and reassembly of chromatin are critical events occurring during events like DNA replication, DNA repair, cell cycle progression and chromosome segregation. During those processes, histones must be removed and re-deposited onto DNA in order to restore chromatin structure (215). This chromatin remodeling is essential for cell survival. Moreover, chromosomes need to be able to condense and decondense for the cell machinery to access DNA for gene expression but also to replicate or to repair itself. Problems in chromosome organization can lead to inappropriate gene expression (216) and this phenomenon could contribute to the important dysregulation of gene expression observed in my transcriptomic analysis since a lot of genes related to nucleosome assembly and DNA packaging are dysregulated. Moreover, problems in chromatin remodeling after DNA repair have been shown to lead to apoptosis through the activation of the DNA damage checkpoint (217). It is important to remember that HDAG-S interacts with the linker histone H1e and this interaction is necessary for HDV replication (141). Histone H1e was also 3.6x less abundant in HDV replicating cells. This perturbation in HDV cells could affect many cellular processes including gene expressions, DNA repair, replication and also affect cell survival.

Two other smaller clusters were enriched for upregulated genes: organelle organization and protein transport. Organelle organization is important to fully support all the cellular processes in a cell but is also a dynamic process. For example, the cell cytoskeleton is extremely

dynamic to provide cell movement but also to performed important cellular functions like mitosis (218). The golgi complex and nucleus also possess dynamic properties (219). As discussed above chromatin organisation and DNA packaging is affected in HDV induced cells, which implies that the internal structure of the nucleus may also be affected. Moreover, we have previously reported the formation of stress granules and increase in NEAT1 foci (the scaffold RNA of paraspeckles) in HDV replicating cells. Lastly, the protein transport cluster was also enriched and this is probably observed as a result of HDV movement in and out of the nucleus and the increase metabolism to support HDV replication (**Figure 3.6A**).

To summarize, HDV replication notably affects the host transcriptomic landscape. With a total of 3 782 gene differentially expressed in this analysis using 293-HDV system, we have established that the virus replication upregulated genes principally involved in RNA processing and G-coupled protein receptors, while HDV replication downregulated genes involved in nucleosome assembly. Those changes in the host cells might be to accommodate HDV replication, but might also be involved in HDV pathogenesis.

If we take a look at Cunha's laboratory proteomic studies (152, 157, 158), two of them were performed in the same cell type (HuH-7) and the same type of analysis but using different strategies to initiate HDV replication. The results obtained in those studies were fairly different and demonstrate how experimental design can influence the final host proteomic results. The system used in our study is provided by a non-hepatocytes cell line where HDAg-L is not being produced, with an unchangeable source of HDAg-S, and where HBV is absent. For this reason, there is the potential that some of the observed results might not be illustrative of host gene expression in a real infection with the presence of HBV.

HDV replication perturb major protein complexes

In the cell, the majority of proteins are associated with one or many partners. About 2/3 of proteins are known to associate into complexes that can have 2 to a hundred different proteins present (220). The majority of complexes are formed from 2 to 4 proteins and a single protein can be part of a multitude of complexes. Core proteins are proteins that are always part of the complex, while there may be other proteins that are temporarily associated in the complex. Identifying complexes that are dysregulated at the expression level can give us good insights to predict which function in the cell will be affected. In this study, I identified 30 complexes that were dysregulated by HDV replication.

Using the CORUM complex dataset, I was able to sort by organism and keep only human complexes. Because many systems of gene annotation are used across the scientific field, I had to convert the protein names provided by the CORUM dataset to the same annotation used in the previous analysis. This conversion of UniProt annotation to Ensembl has a lot of disadvantages, although it is necessary to compare two different dataset using different annotation. Indeed, this conversion causes data loss as certain genes are not present or have not annotated in both UniProt and Ensembl.

GO analysis on the 30 complexes affected by HDV replication (**Figure 3.9**) revealed that many complexes were part of the chromosome segregation, chromatin remodeling and the DNA-templated transcription and elongation pathway. Complexes involved in chromatin remodeling were a major part of the complexes affected. These results are in accordance with the downregulation observed in the nucleosome assembly process previously discussed. Among the complexes in this pathway are: codanin 1-ASF, TLE-Histone H3, B-WICH as well as other histones related complexes. The codanin 1-ASF complex is essential for nucleosome

assembly during replication where they regulate S-phase histone supply (221). Among the 22 proteins in the codanin 1-ASF complex, 95% were affected among which 17 were downregulated and four upregulated. The TLE-Histone H3 complex possesses a total of eight proteins, of which 62% were affected. This complex act as a gene-specific co-repressor who has an important role in gene expression (222, 223). It should be noted that the transducin-like Enhancer of Split (TLE) protein present in the TLE-Histon H3 complex interacts with all four histones but shows highest affinity for the H3 N-terminal tail. TLE also interacts with several histone deacetylases which deacetylate histones and therefore condense chromatin and reduce the accessibility of the transcriptional machinery (224). This complex is also implicated in different signal transduction pathways and was shown to mediate cross talk between pathways like MAPK, Ras and Wnt. A dysregulation of this complex could severely impact the chromatin structure, affect gene expression and was shown to be linked with cancer formation (222–224). While the TLE-Histone H3 complex was downregulated, the B-WICH complex was upregulated. This complex is composed of 10 proteins and 60% of these were positively affected with HDV replication. The B-WICH complex possesses the opposite role of TLE-Histone, thus while the TLE-Histone interacts with histone deacetylase enzymes to maintain chromatin in a transcriptionally inactive state, the B-WICH complex recruits histone acetyltransferases to remodel chromatin to transcriptionally active states (225, 226). There seems to be a tendency in the HDV affected complexes to promote a less condensed form of chromatin. This is due to the fact that downregulation of TLE-Histone complex would negatively impact the chromatin condensation and the upregulation of the B-WICH complex which would favor an open chromatin structure (225). These changes in HDV replicating cells seems to prevent compaction of the chromatin to facilitate the transcription of genes that could be beneficial for HDV.

Moreover, many complexes related to the cell cycle were affected. The GO analysis (**Figure 3.9B**) revealed that many complexes were part of the chromosome segregation pathway, more precisely they were involved in the cell cycle, in mitosis, G2/M transition and also involved in cell cycle checkpoints. These include the chromosomal passenger complex (CPC), the mitotic checkpoint, the NDC80 kinetochore and the RC complex. While the three complexes discussed above are implicated in chromatin related processes, CPC, mitotic checkpoint, NDC80 and the RC complex are more involved in cell cycle related processes. Indeed, the CPC is composed of three proteins (INCENP, CDCA8 and AURKB), which were all downregulated (**Figure 3.9**). This complex is a key regulator for mitosis in particular in chromosome segregation and cytokinesis process. Mitosis is an extremely elaborated process involving major cellular changes which need to be properly performed and synchronized to limit potential errors that may occur during cell division (227). That being said, there are multiple mechanisms developed by the cell to ensure a successful mitosis. The CPC plays many roles in the mitosis progression. First, it is located on chromosome arms where AURKB phosphorylates histone H3 on multiple serines (Ser10 and Ser28) which correlates with chromosome condensation in prophase (228, 229). Secondly, CPC was shown to have a role in spindle assembly and stability in prometaphase where it regulates kinetochore-microtubule attachment (230). In metaphase, CPC plays a role in the kinetochore-microtubule attachment process, in chromosome alignment and plays a role in the central spindle formation/disassembly as well as in the cytokinesis (reviewed by Ruchaud *et al.* 2007 (231)). The CPC complex also regulates the mitotic check point. The CPC protein AURKB cooperate with BUB1 kinases and other components important for this checkpoint including the APC protein (232). Disruption of any member of this complex was shown to disrupt the mitosis progression. The reduced gene expression observed in this analysis for CPC complex most

probably results in a disruption of the complex's activity. In addition to gene expression, I validated the AURKB protein expression which was significantly impaired with a 87% reduction in HDV replicating cells (**Figure 3.5B&C**).

The mitotic checkpoint was also affected in the transcriptomic analysis. The mitotic checkpoint complex, consisting of four proteins which were all downregulated: MAD2L1, CDC20, BUB3 and BUB1. MAD2L1 is a component of the mitotic spindle checkpoint that ensure that all chromosomes are properly aligned for metaphase or otherwise sequesters the Cell Division Cycle 20 protein (CDC20), a protein involved in chromosome segregation (233). The spindle checkpoint can be activated if microtubules are not attached to the kinetochore or if the tension is not well distributed between sister chromatids. The checkpoint activation causes MAD2, BUB1 and BUB3 to interact with Cdc20 and inhibits the anaphase-promoting complex (APC) to degrade the securing protein to initiate chromatid separation. Because this complex is downregulated in my analysis, I believe this might highlight a disturbance at the checkpoint control and thus inhibition of the spindle checkpoint could prevent a cell cycle arrest in G2/M. A perturbation of the mitotic checkpoint could be a mechanism exploited by the virus to get around this checkpoint and progress through mitosis despite possible anomalies in the condensation of sister chromatids.

The last complex involved in cell cycle addressed in this section is the NDC80 kinetochore complex, which is an essential microtubule-binding component of the kinetochore. This small complex is composed of three proteins which were also under-expressed in 293-HDV induced cells (**Figure 3.9**). The NDC80 complex is required for microtubule attachment to kinetochores but also act in close collaboration with the Dam1 complex, which enhances NDC80 ability to attach microtubules and to establish chromosome

biorientation (234, 235). Moreover, NDC80 and its interacting partner Dam1 are substrates of AURKB's phosphorylation (236). When AURKB senses an error in kinetochore-microtubule attachment, AURKB destabilizes the inappropriate interaction between the kinetochore and microtubules. To do this, AURKB phosphorylates NDC80 and its other interacting partners (including Dam1) to disrupt their microtubule binding. By disrupting the microtubule binding activity of those kinetochore proteins, it allows for re-establishment of the kinetochore alignment as well as to release the cell arrest in anaphase caused by the spindle checkpoint. Thus, I believe this could be linked with the downregulation of AURKB, which was shown to be an important kinase in many of the complexes.

In this analysis, many complexes related to the cell cycle were affected. The GO analysis (**Figure 3.9B**) revealed that many complexes were part of the chromosome segregation, more precisely they were involved in the cell cycle, in mitosis, G2/M transition and also involved in cell cycle checkpoint. Those findings support the findings of Mendes *et al.* with the deregulation of G2/M checkpoint (158), although it does not agree with the results obtained by Taylor's team demonstrating a cell cycle arrest in G0/G1 (153). Other RNA and DNA viruses disrupt cell cycle checkpoints to exploit the host cell and to have a favorable environment for their replication and infection. Those viruses employ different strategies like the production of proteins targeting important cell cycle regulators. For instance, the human immunodeficient virus (HIV) induces a cell cycle arrest in G2/M to restrain new cell reproduction which allows it to better evade immune system (237). Herpes simplex virus produce a protein that interferes with kinetochores protein CENP-C and induce its degradation which thus disrupts the kinetochore structure and impairs mitosis (238, 239). Here we uncovered many different cellular complexes that were affected at their RNA expression level. It appears to be evidence that HDV affects chromatin structure integrity and also disrupts the

cell cycle in the mitosis phase. Based on other studies where similar processes were affected or targeted by viruses, I can only assume that these cellular changes are there to support the exhaustive replication of HDV and help it accomplish its viral cycle. However, the mechanism behind how HDV replication affect those complexes is still unclear and needs further research.

It should be noted that the amount of RNA transcript present in a cell is not always representative of the amount of protein produce (195, 196, 240). Some less abundant transcripts in the cell can be more actively translated into proteins while other more abundant RNA can be less actively translated (241, 242). Moreover, by using the terms “upregulation” and “downregulation”, I specifically imply that an upregulation corresponds to an increase in accumulation of that transcript. I acknowledge that some genes might have been less abundant in my analysis due to a higher degradation rate or on the contrary, they could be present in higher quantity as a result of a reduced degradation rate. RNA turnover is an important process that is necessary to maintain a cellular homeostasis (243). RNA degradation by ribonucleases can modulate gene expression and remove non necessary or non-functional RNAs. Modifying enzymes can influence RNA stability to promote RNA degradation and on the opposite can also promote stability thus preventing degradation, as well as influence the RNA processing and translation (244, 245). In the WGCNA signed analysis (**Figure 3.7**) among the RNA processing principal cluster, the RNA metabolism and regulation of RNA processing were among the affected biological processes. This emphasizes the fact that RNA turnover rate could have been affected, although these factors have not yet been assessed in HDV expression systems. As such this analysis cannot deduce the levels of transcripts which are being actively transcribed, and my results solely quantify the transcript levels present in my cellular system at the moment of the analysis.

Flow cytometry reveals an arrest of the cell cycle and morphological differences for HDV induced cells

The cell cycle is a controlled and coordinated series of events and is the essential mechanism by which all eukaryotic cells reproduce. Composed of four major phases; G1 (growth phase), S (chromosome duplication), G2 (growth phase) and M (mitosis), which are regulated by a remarkably robust and reliable control system (246). Cells can also enter a non-replicative or senescent state if the environment is unfavorable. Progression from one phase to the next is controlled by kinases and cyclin-dependant kinases (Cdk) (247). These cell components control the events of the cell cycle and ensure that they are properly timed and coordinated with each other. Several mechanisms have been put in place to regulate the progression of the cell cycle as well as the internal conditions of the cell, which are called cell cycle checkpoints. The three main checkpoints are the ones in G1/S, G2 and G2/M (220, 248). The first one at the beginning of G1 ensure the environment is favorable to initiate a new cell cycle and also controls cell size. The second one is at the beginning of the G2 and validates that the DNA has been properly replicated with no DNA damage and that the environment is favorable to continue progression through the cell cycle. The third one is the spindle assembly checkpoint, which takes place after metaphase but before anaphase. The role of this checkpoint is to assure a proper alignment of chromosome on the mitotic spindle.

The study previously performed by John Taylor's team using the same cell system as the one used in this analysis had demonstrated an accumulation of cells in G0/G1 phase upon induction of HDV replication. Subsequently, the proteomic analysis performed by Cunha's team in 2013, using the same cell system, provided evidence that the G2/M DNA damage

checkpoint was disrupted. These results are consistent with the one obtained in my analysis which strongly suggests a cell cycle perturbation. Indeed, many complexes involved in cell cycle, in cell cycle checkpoints, in chromosome segregation and chromatin remodeling were affected. These results seem to support this cellular disturbance, and prompted my decision to conduct a cell cycle analysis

First a confirmation of the possible cell cycle arrest was required. Therefore, a flow cytometry experiment using PI staining was performed on the three-cell types (HEK-293, 293-Ag and 293-HDV) in induced and non-induced conditions. A 15.5% increase in the relative number of cells in G0/G1 phase was observed compared to the non-induced cells and a 14% reduction in the cells in S phase was also observed. Although this result seemed to suggest a cell cycle arrest in G0/G1, it was interesting to see if this arrest was consistent with induction time. The experiment was therefore repeated using three different time points (12h, 24h and 36h) and the results obtained revealed an increase of 16.6% of relative cell number in G0/G1 phase after 24h and 24.0% 36h post induction although there were no changes after only 12h. This accumulation seems to reach a plateau after 36h since there is very little change between 24h (16.6%) and 36h (24%). It would be necessary; however, to test this hypothesis with more time points. While the number of cells in G0/G1 increased, a reduction of 15% was observed in the S phase after 24h and a reduction 20% after 36h (again with no change after only 12h). These findings are consistent with the results obtained by the Taylor's team. Inducing cell cycle arrest can be a strategy to temporarily escape apoptosis, and to avoid cell division to limit the number of infected cells so that the immune system can be more easily evaded. As discussed above, we observed an increase in the production of essential molecular building blocks. Similarly, a cell cycle arrest in a particular phase like in G0/G1 is a way virus uses to

manipulate the cell in order to take advantage of cellular resources and of favorable conditions for its own replication and propagation.

In addition, we also observed a difference in size for HDV TET cells as compared to non-induced cells (**Figure 3.13A**). Moreover, this size difference in size is present only in cells stopped in G0/G1 (**Figure 3.13B**) and these cells additionally display a decrease in cell granulosity. The size difference is also observable using phase contrast microscopy, where we can physically observe that HDV replicating cells are smaller and rounder and that they have a higher tendency to form aggregates (**Figure 3.15**). A higher percentage of aggregate was also distinguishable by using the flow cytometry forward scatter parameters (**Figure 3.14**). This phenomenon could be explained by a cellular stress induced by the replication of HDV, as it has been demonstrated previously in our laboratory (127).

Since the nucleosome assembly and DNA packaging seems to be severely affected in the addition to a dysregulation of the mitotic checkpoint in the mitosis, it is possible that HDV virus can bypass this control checkpoint and therefore the cell perform mitosis without triggering cell arrest in G2 / M. Despite the evidence of a disruption of chromosome status by our RNA-seq analysis, additional experiments would be required to confirm these findings. Nonetheless, it seems clear that despite these possible chromosome and segregation disturbances in G2/M, the HDV replicating cells manage to complete the mitosis and to go to G0/G1. A cell cycle progression and division with major defect in cell cycle, or a break in chromosome during segregation can lead to catastrophe, trigger apoptosis or induce considerable changes including leading to a cancer or senescence state (220).

Future directions

The GO analyses and protein complex analyses have suggested a perturbation in chromatin remodeling, chromosome segregation and in many complexes related to cell cycle processes. After confirming a disruption of the cell cycle and arrest in G0/G1 with the flow cytometry experiments, it would be essential to better assess what could be triggering this cell arrest. Although several factors could influence a cellular arrest, there was one protein in particular with a crucial role in many of the complexes that were affected in my analysis: AURKB. This kinase has central role in mitosis progression and is a component or a key interacting protein in the chromosome passenger complex, the spindle assembly checkpoint and the NDC80 complex (233–235, 249, 250). Its downregulation at the RNA and protein level were also validated in this work. The decrease in Aurora kinase B expression could have direct consequences on its substrates phosphorylation and thus impact many cell cycle related processes. Therefore, it would be interesting to investigate if the decrease in Aurora kinase B expression had direct consequences on the phosphorylation of its substrate phosphorylation. AURKB phosphorylation is also correlated with mitotic chromosome condensation and for this reason chromosome condensation state could also be explored as well.

Now that I have confirmed a cell cycle arrest in HDV replicating cells and that I have identified how HDV affect the host transcriptomic landscape, we can investigate the mechanisms dictating how HDV causes these changes. Taylor *et al.* have previously shown an increase in the number of cells present in G0/G1 but their staining for annexin V, 7-AAD and trypan blue did not allow to conclude a difference between these cells. Their results suggest that HDV induction does not favor apoptosis or necrosis despite the cell cycle arrest observed and suggest a possible cellular senescence, although this still needs to be confirmed. Preliminary results performed by a colleague, Sergio Armando, agree with the results obtained

by Taylor *et al.* with a very high viability and lower apoptosis rate including a 18% reduction of apoptotic cells present in the G0/G1 phase in HDV replicating cells. Senescence states can be induced by stressor factors, and HDV replication has been previously linked with cellular stress in addition to the 498 genes involved in cellular stress shown to be dysregulated in this present research (127). The next step in this analysis would to explore if the cells arrested in G0/G1 are viable and confirm whether they would go through either apoptosis or necrosis. Moreover, with HDV induction, cells become non-adherent, detach from the plate and becomes rounder (**Figure 3.15**, also observed by Taylor *et al.*). Further research is needed to better characterize the difference between the arrested and the non-arrested cells. Flow cytometry analysis performed on the floating cells collected from the media supernatant compared to the adherent cells could be used to assess if there are differences between factors known to affect cell focal adhesion.

Even if we observe more genes upregulated in this analysis, the analysis of affected protein complexes seems to contain more complexes with under-expressed genes instead of what we may have expected to find with more complexes being upregulated. This interesting finding raises a lot of questions. An analysis of gene promoters would be very interesting and would uncover if the dysregulated genes belong to a common promoter and if they are usually co-regulated. This would allow for an analysis to determine whether there are correlations between dysregulated genes and specific promoters.

CHAPTER 5: CONCLUSION

As the HDV genome is considerably smaller than other mammalian virus and only encodes for one protein containing two isoforms, it relies heavily on the host cell components to perform its replication. Previous studies demonstrated that replication or accumulation of the HDV genome and its antigens affected the host cell at the proteomic level (152, 157, 158). Although some teams had previously looked at the expression of single genes in host cells with HDV and/or HDAGs, to my knowledge, a genome-wide transcriptome analysis study has never been performed. Therefore, this study aimed to investigate how HDV and its components affect the hosts transcriptional landscape. A total of 3,782 genes were shown to be differentially expressed by either HDV and/or HDAGs (**Figure 3.4**). There were 3,278 genes upregulated by HDV RNA and these genes were predominantly from four enriched categories: RNA processing, G-protein coupled receptor signaling pathway, protein transport, and organelle organization (**Figure 5.1**). The 283 downregulated genes affected by HDV were shown to be part of the nucleosome assembly pathway. Moreover, protein complexes with multiple members whose expression at the gene level was affected were identified. A total of 30 complexes were affected and predominantly part of these pathways: DNA-template transcription and elongation, chromosome segregation, chromatin remodeling (**Figure 5.1**). Lastly, using a flow cytometry approach, I found that cells replicating HDV were accumulating in G0/G1 and those cells also had a different morphology. The involvement of HDV dysregulation of those host gene and their role in the accumulation of HDV RNA is still unclear but those findings might enlighten what mechanism or what changes in host cell could be causing this cell cycle arrest observed in HDV replicating cells.

Model of HDV replication effect in host cell

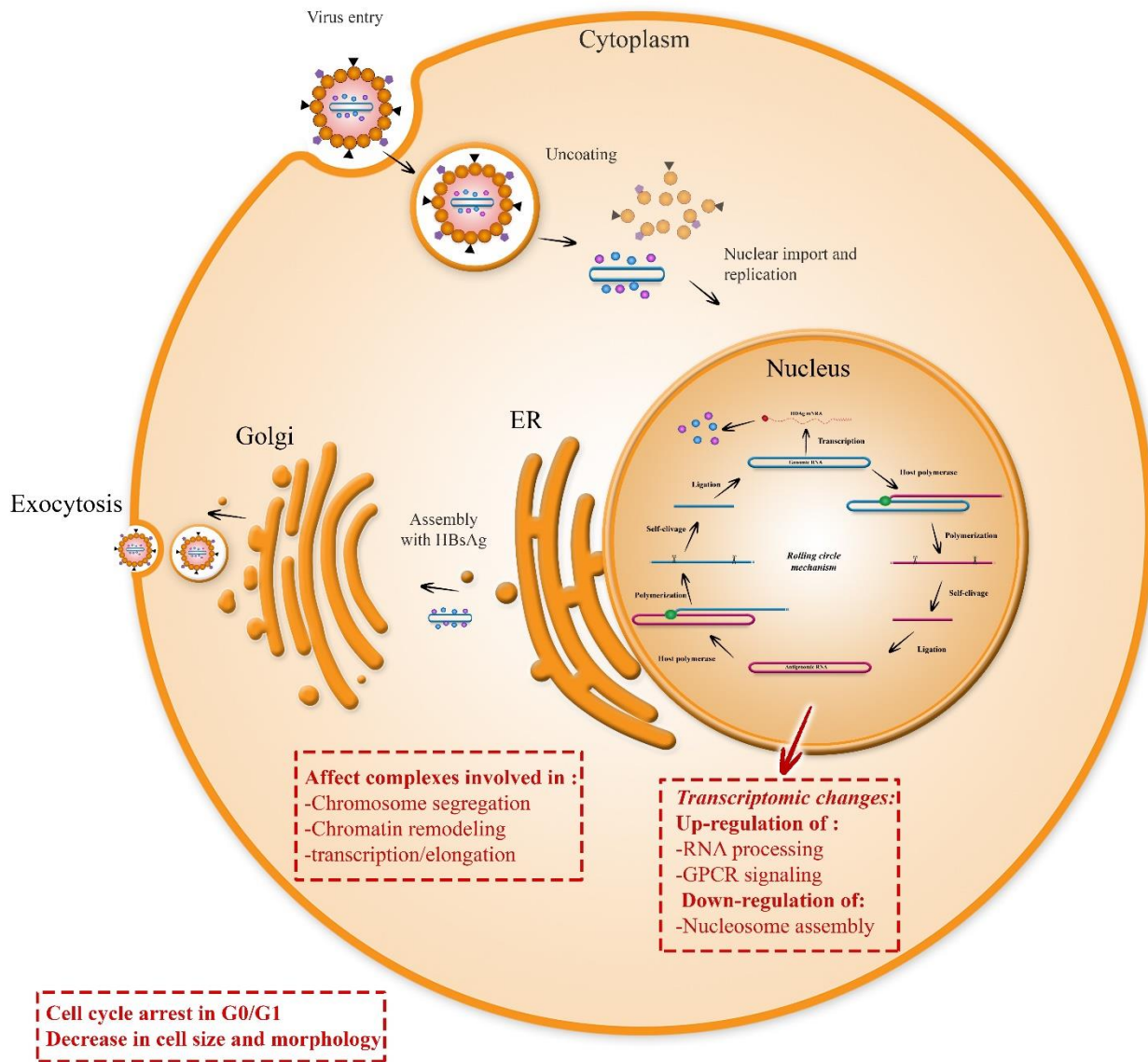


Figure 5.1. Working model of the effect of HDV replication in host cell.

This model represent changes induces by HDV replication in a HEK-293 host cell. HDV induced a multitude of cellular changes in the host when replicating, including changes at the gene expression level, including the expression of genes encoding for protein complex, as well as changes like a cell cycle perturbation. Gene expression analysis implicate in RNA processing and GPCR signaling pathway have shown to be upregulated in the cell while those involved in the nucleosome assembly have been reduced in the host cell as a result of HDV replication. The expression of several genes belonging to the same complex were also been affected by HDV, which provides a clearer picture the functions that are dysregulated during HDV infection. Complexes involved in chromosome segregation, chromatin remodeling and the transcription/elongation were all significantly affected. A cell cycle arrest was also observed in G0/G1 phase as well as a morphological change in cell size and granulocyt and an increase in cellular aggregate.

REFERENCES

1. Rizzetto M, Canese MG, Aricò S, Crivelli O, Trepo C, Bonino F, Verme G. 1977. Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut* 18:997–1003.
2. Wang K-S, Choo Q-L, Weiner AJ, Ou J-H, Najarian RC, Thayer RM, Mullenbach GT, Denniston KJ, Gerin JL, Houghton M. 1986. Structure, sequence and expression of the hepatitis delta (δ) viral genome. *Nature* 323:508–514.
3. Dény P. 2006. Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? *Curr Top Microbiol Immunol* 307:151–71.
4. Le Gal F, Gault E, Ripault M-P, Serpaggi J, Trinchet J-C, Gordien E, Dény P. 2006. Eighth major clade for hepatitis delta virus. *Emerg Infect Dis* 12:1447–50.
5. Lasda E, Parker R, Parker ROY. 2014. Circular RNAs : diversity of form and function. *Circular RNAs : diversity of form and function. Rna* 1829–1842.
6. Flores R, Owens RA, Taylor J. 2016. Pathogenesis by subviral agents: Viroids and hepatitis delta virus. *Curr Opin Virol* 17:87–94.
7. Taylor J, Pelchat M. 2010. Origin of hepatitis δ virus. *Microbiology* 393–402.
8. Miao Z, Zhang S, Ma Z, Hakim MS, Wang W, Peppelenbosch MP, Pan Q. 2018. Recombinant identification, molecular classification and proposed reference genomes for hepatitis delta virus. *J Viral Hepat* 1–8.
9. Wedemeyer H, Manns MP. 2010. Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead. *Nat Rev Gastroenterol Hepatol* 7:31–40.
10. Su C-W, Huang Y-H, Huo T-I, Shih HH, Sheen I-J, Chen S-W, Lee P-C, Lee S-D, Wu J-C. 2006. Genotypes and viremia of hepatitis B and D viruses are associated with outcomes of chronic hepatitis D patients. *Gastroenterology* 130:1625–35.
11. Hsu S, Syu W-J, Sheen I-J, Liu H-T, Jeng K-S, Wu J-C. 2002. Varied assembly and RNA editing efficiencies between genotypes I and II hepatitis D virus and their implications. *Hepatology* 35:665–672.
12. Nakano T, Hadler SC, Orito E, Shapiro CN, Casey JL, Mizokami M, Robertson BH. 2001. Characterization of hepatitis D virus genotype III among Yucpa Indians in Venezuela. *J Gen Virol* 82:2183–2189.
13. Casey JL, Brown TL, Colan EJ, Wignall FS, Gerin JL. 1993. A genotype of hepatitis D virus that occurs in northern South America. *Proc Natl Acad Sci U S A* 90:9016–

- 20.
14. Organization world health. 2018. Hepatitis D.
15. Patrizia Farci. 2003. Delta hepatitis: an update. Elsevier.
16. Lempp FA, Ni Y, Urban S. 2016. Hepatitis delta virus: Insights into a peculiar pathogen and novel treatment options. *Nat Rev Gastroenterol Hepatol* 13:580–589.
17. Navabakhsh B, Mehrabi N, Estakhri A, Mohamadnejad M, Poustchi H. 2011. Hepatitis B Virus Infection during Pregnancy: Transmission and Prevention. *Middle East J Dig Dis* 3:92–102.
18. Yao JL. 1996. Perinatal transmission of hepatitis B virus infection and vaccination in China. *Gut* 38 Suppl 2:S37-8.
19. Umar M, Hamama-tul-Bushra, Umar S, Khan HA. 2013. HBV Perinatal Transmission. *Int J Hepatol* 2013:7.
20. Ramia S, Bahakim H. 1988. Perinatal transmission of hepatitis B virus-associated hepatitis D virus. *Ann l’Institut Pasteur / Virol* 139:285–290.
21. Pinho-nascimento CA, Bratschi MW, Soares CC, Warryn L, Minyem JC, Terezinha M, Moraes B De, Boock AU, Niel C, Pluschke G. 2018. Transmission of Hepatitis B and D Viruses in an African Rural. *mSystems* 3:e00120-18.
22. Heidrich B, Deterding K, Tillmann HL, Raupach R, Manns MP, Wedemeyer H. 2009. Virological and clinical characteristics of delta hepatitis in Central Europe 883–894.
23. Weltman MD, Brotodihardjo A, Crewe EB, Farrell GC, B M, Grierson M. 1995. Coinfection with hepatitis B and C or B , C and 6 viruses results in severe chronic liver disease and responds poorly to in terferon-a treatment.
24. Koytak ES, Yurdaydin C, Glenn JS. 2007. Hepatitis d. *Curr Treat Options Gastroenterol* 10:456–63.
25. Farci P, Chessa L, Balestrieri C, Serra G, Lai ME. 2007. Treatment of chronic hepatitis D. *J Viral Hepat* 14:58–63.
26. Zhang Z, Filzmayer C, Ni Y, Sülmann H, Mutz P, Hiet MS, Vondran FWR, Bartenschlager R, Urban S. 2018. Hepatitis D virus replication is sensed by MDA5 and induces IFN- β / λ responses in hepatocytes. *J Hepatol* 69:25–35.
27. Mederacke I, Filmann N, Yurdaydin C, Bremer B, Puls F, Zacher BJ, Heidrich B, Tillmann HL, Rosenau J, Bock C-T, Savas B, Helfritz F, Lehner F, Strassburg CP, Klempnauer J, Wursthorn K, Lehmann U, Manns MP, Herrmann E, Wedemeyer H. 2012. Rapid early HDV RNA decline in the peripheral blood but prolonged intrahepatic hepatitis delta antigen persistence after liver transplantation. *J Hepatol* 56:115–122.
28. Adil B, Fatih O, Volkan I, Bora B, Veysel E, Koray K, Cemalettin K, Burak I, Sezai Y. 2016. Hepatitis B Virus and Hepatitis D Virus Recurrence in Patients Undergoing Liver Transplantation for Hepatitis B Virus and Hepatitis B Virus Plus Hepatitis D Virus. *Transplant Proc* 48:2119–2123.

29. Rizzetto M. 1983. Chronic Hepatitis in Carriers of Hepatitis B Surface Antigen, with Intrahepatic Expression of the Delta Antigen. *Ann Intern Med* 98:437.
30. Ueda K, Tsurimoto T, Matsubara K. 1991. Three Envelope Proteins of Hepatitis B Virus: Large S, Middle S, and Major S Proteins Needed for the Formation of Dane Particles. *JOURNAL OF VIROLOGY*.
31. Sureau C, Guerra B, Lanford RE. 1993. Role of the large hepatitis B virus envelope protein in infectivity of the hepatitis delta virion. *J Virol* 67:366–72.
32. Lambert C, Prange R. 2007. Posttranslational N-glycosylation of the hepatitis B virus large envelope protein. *Virol J* 4:1–9.
33. Huang WH, Chen CW, Wu HL, Chen PJ. 2006. Post-translational modification of delta antigen of hepatitis D virus. *Curr Top Microbiol Immunol* 307:91–112.
34. Ni Y, Lempp FA, Mehrle S, Nkongolo S, Kaufman C, Fälth M, Stindt J, Königer C, Nassal M, Kubitz R, Sülthmann H, Urban S. 2014. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. *Gastroenterology* 146:1070–1083.
35. Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, Huang Y, Qi Y, Peng B, Wang H, Fu L, Song M, Chen P, Gao W, Ren B, Sun Y, Cai T, Feng X, Sui J, Li W. 2012. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* 2012:1–28.
36. Bruss V, Ganem D. 1991. The role of envelope proteins in hepatitis B virus assembly. *Proc Natl Acad Sci U S A* 88:1059–1063.
37. Hartmann-Stuhler C, Prange R. 2001. Hepatitis B Virus Large Envelope Protein Interacts with 2-Adaptin, a Clathrin Adaptor-Related Protein. *J Virol* 75:5343–5351.
38. Ganemi D, Varmu HE. 1987. The molecular biology of the hepatitis B viruses.
39. Chai N, Chang HE, Nicolas E, Han Z, Jarnik M, Taylor J. 2008. Properties of Subviral Particles of Hepatitis B Virus. *J Virol* 82:7812–7817.
40. Gudima S, Chang J, Moraleda G, Azvolinsky A, Taylor J. 2002. Parameters of Human Hepatitis Delta Virus Genome Replication: the Quantity, Quality, and Intracellular Distribution of Viral Proteins and RNA. *J Virol* 76:3709–3719.
41. Dingle K, Moraleda G, Bichko V, Taylor J. 1998. Electrophoretic analysis of the ribonucleoproteins of hepatitis delta virus. *J Virol Methods* 75:199–204.
42. Chen PJ, Kalpana G, Goldberg J, Mason W, Werner B, Gerin J, Taylor J. 1986. Structure and replication of the genome of the hepatitis delta virus. *Proc Natl Acad Sci U S A* 83:8774–8.
43. Sureau C, Negro F. 2016. The hepatitis delta virus: Replication and pathogenesis. *J Hepatol* 64:S102-16.
44. Macnaughton TB, Lai MMC. 2002. Genomic but not antigenomic hepatitis delta virus RNA is preferentially exported from the nucleus immediately after synthesis and processing. *J Virol* 76:3928–35.

45. Tavanez JP, Cunha C, Silva MCA, David E, Monjardino J, Carmo-Fonseca M. 2002. Hepatitis delta virus ribonucleoproteins shuttle between the nucleus and the cytoplasm. *RNA* 8:S1355838202026432.
46. Kos A, Dijkema R, Arnberg AC, van der Meide PH, Schellekens H. 1986. The hepatitis delta (δ) virus possesses a circular RNA. *Nature* 323:558–560.
47. Wang KS, Choo QL, Weiner a J, Ou JH, Najarian RC, Thayer RM, Mullenbach GT, Denniston KJ, Gerin JL, Houghton M. 1986. Structure, sequence and expression of the hepatitis delta (delta) viral genome. *Nature* 323:508–14.
48. Livak KJ, Schmittgen TD. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 25:402–408.
49. Wong SK, Lazinski DW. 2002. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci U S A* 99:15118–23.
50. Huang C-R, Lo SJ. 2010. Evolution and diversity of the human hepatitis d virus genome. *Adv Bioinformatics* 323654.
51. Taylor J, Pelchat M. 2010. Origin of hepatitis delta virus. *Future Microbiol* 5:393–402.
52. Ferré-D'Amaré AR, Zhou K, Doudna JA. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* 395:567–574.
53. Brown AL, Perrotta AT, Wadkins TS, Been MD. 2008. The poly(A) site sequence in HDV RNA alters both extent and rate of self-cleavage of the antigenomic ribozyme. *Nucleic Acids Res* 36:2990–3000.
54. Han Z, Alves C, Gudima S, Taylor J. 2009. Intracellular Localization of Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation. *J Virol* 83:6457–6463.
55. Bell P, Brazas R, Ganem D, Maul GG. 2000. Hepatitis delta virus replication generates complexes of large hepatitis delta antigen and antigenomic RNA that affiliate with and alter nuclear domain 10. *J Virol* 74:5329–36.
56. Cullen JM, David C, Wang J-G, Becherer P, Lemon SM. 1995. Subcellular distribution of large and small hepatitis delta antigen in hepatocytes of hepatitis delta virus superinfected woodchucks. *Hepatology* 22:1090–1100.
57. Bichko V V, Taylor JM. 1996. Redistribution of the delta antigens in cells replicating the genome of hepatitis delta virus. *J Virol* 70:8064–8070.
58. Li Y-J, Macnaughton T, Gao L, Lai MMC. 2006. RNA-templated replication of hepatitis delta virus: genomic and antigenomic RNAs associate with different nuclear bodies. *J Virol* 80:6478–86.
59. Abbas Z, Afzal R. 2013. Life cycle and pathogenesis of hepatitis D virus: A review. *World J Hepatol* 5:666–675.
60. Moraleda G, Taylor J. 2001. Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. *J Virol* 75:10161–10169.

61. Gudima S, Wu S-Y, Chiang C-M, Moraleda G, Taylor J. 2000. Origin of Hepatitis Delta Virus mRNA. *J Virol* 74:7204–7210.
62. Nie X, Chang J, Taylor JM. 2004. Alternative Processing of Hepatitis Delta Virus Antigenomic RNA Transcripts. *J Virol* 78:4517–4524.
63. Fu TB, Taylor J. 1993. The RNAs of hepatitis delta virus are copied by RNA polymerase II in nuclear homogenates. *J Virol* 67:6965–72.
64. Greco-Stewart VS, Miron P, Abraham A, Pelchat M. 2007. The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology* 357:68–78.
65. Abraham A, Pelchat M. 2008. Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. *Nucleic Acids Res* 36:5201–5211.
66. Modahl LE, Macnaughton TB, Zhu N, Johnson DL, Lai MM. 2000. RNA-Dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol Cell Biol*.
67. Macnaughton TB, Shi ST, Modahl LE, Lai MMC. 2002. Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. *J Virol* 76:3920–7.
68. Greco-Stewart VS, Schissel E, Pelchat M. 2009. The hepatitis delta virus RNA genome interacts with the human RNA polymerases I and III. *Virology* 386:12–15.
69. Kuo MY, Chao M, Taylor J. 1989. Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. *J Virol* 63:1945–50.
70. Hsieh SY, Chao M, Coates L, Taylor J. 1990. Hepatitis delta virus genome replication: a polyadenylated mRNA for delta antigen. *J Virol* 64:3192–8.
71. Lai MMC. 1995. The molecular biology of hepatitis delta virus. 64.
72. Bergmann KF, Gerin JL. 1986. Antigens of hepatitis delta virus in the liver and serum of humans and animals. *J Infect Dis* 154:702–6.
73. Weiner AJ, Choo QL, Wang KS, Govindarajan S, Redeker AG, Gerin JL, Houghton M. 1988. A single antigenomic open reading frame of the hepatitis delta virus encodes the epitope(s) of both hepatitis delta antigen polypeptides p24 delta and p27 delta. *J Virol* 62:594–599.
74. Chou HC, Hsieh TY, Sheu GT, Lai MM. 1998. Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. *J Virol* 72:3684–90.
75. Cao DAN, Haussecker D, Huang Y, Kay MA. 2009. Combined proteomic – RNAi screen for host factors involved in human hepatitis delta virus replication. *Rna* 15:1–9.
76. Lazinski DW, Taylor JM. 1993. Relating structure to function in the hepatitis delta virus antigen. *J Virol* 67:2672–80.
77. Chang MF, Baker SC, Soe LH, Kamahora T, Keck JG, Makino S, Govindarajan S,

- Lai MM. 1988. Human hepatitis delta antigen is a nuclear phosphoprotein with RNA-binding activity. *J Virol* 62:2403–10.
78. Poisson, F, Roingeard P, Dubois F, Calogero RA, Baillou A, Bonelli F, Poisson F, Goudeau A, Roingeard P, Baillou A, Dubois F, Bonelli F, Calogero RA, Goudeau A. 1993. Characterization of RNA-binding domains of hepatitis delta antigen. *J Gen Virol* 74:2473–2478.
 79. Casey JL, Soroush A, Mamedov MR, Daigh LH, Griffin BL. 2013. Arginine-Rich Motifs Are Not Required for Hepatitis Delta Virus RNA Binding Activity of the Hepatitis Delta Antigen. *J Virol* 87:8665–8674.
 80. Chao M, Hsieh SY, Taylor J. 1991. The antigen of hepatitis delta virus: examination of in vitro RNA-binding specificity. *J Virol* 65:4057–4062.
 81. Xia YP, Yeh CT, Ou JH, Lai MM. 1992. Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. *J Virol* 66:914–21.
 82. Lin BC, Defenbaugh DA, Casey JL. 2010. Multimerization of hepatitis delta antigen is a critical determinant of RNA binding specificity. *J Virol* 84:1406–13.
 83. Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM. Structural basis of the oligomerization of hepatitis delta antigen 821–830.
 84. An YXIA, La DMMC. 1992. Oligomerization of Hepatitis Delta Antigen Is Required for both the trans-Activating and trans-Dominant Inhibitory Activities of the Delta Antigen 66:6641–6648.
 85. Huang ZS, Wu HN. 1998. Identification and characterization of the RNA chaperone activity of hepatitis delta antigen peptides. *J Biol Chem* 273:26455–26461.
 86. Wang CC, Chang TC, Lin CW, Tsui HL, Chu PBC, Chen BS, Huang ZS, Wu HN. 2003. Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen. *Nucleic Acids Res* 31:6481–6492.
 87. Yamaguchi Y, Filipovska J, Yano K, Furuya a, Inukai N, Narita T, Wada T, Sugimoto S, Konarska MM, Handa H. 2001. Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science* 293:124–127.
 88. Lee CH, Chang SC, Wu CHHH, Chang MF. 2001. A Novel Chromosome Region Maintenance 1-independent Nuclear Export Signal of the Large Form of Hepatitis Delta Antigen That Is Required for the Viral Assembly. *J Biol Chem* 276:8142–8148.
 89. Lee CZ, Chen PJ, Chen DS. 1995. Large hepatitis delta antigen in packaging and replication inhibition: role of the carboxyl-terminal 19 amino acids and amino-terminal sequences. *J Virol* 69:5332–5336.
 90. Macnaughton TB, Lai MMC. 2011. Large Hepatitis Delta Antigen Is Not a Suppressor of Hepatitis Delta Virus RNA Synthesis once RNA Replication Is Established. *Society* 76:9910–9919.
 91. Hwang SB, Lai MMC. 1993. Isoprenylation Mediates Direct Protein-Protein Interactions between Hepatitis Large Delta Antigen and Hepatitis B Virus Surface

Antigen. *J Virol* 67:7659–7662.

92. Huang C, Chang S-CSC, Yu I-C, Tsay Y-GY-GY-G, Chang M-FM-F. 2007. Large Hepatitis Delta Antigen Is a Novel Clathrin Adaptor-Like Protein. *J Virol* 81:5985–5994.
93. Fornerod M, Ohno M, Yoshida M, Mattaj JW. 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90:1051–60.
94. Nishida E, Fukuda M, Asano S, Nakamura T, Adachi M, Yoshida M, Yanagida M. 1997. CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* 390:308–311.
95. Kudo N, Wolff B, Sekimoto T, Schreiner EP, Yoneda Y, Yanagida M, Horinouchi S, Yoshida M. 1998. Leptomycin B Inhibition of Signal-Mediated Nuclear Export by Direct Binding to CRM1. *Exp Cell Res* 242:540–547.
96. Wang Y, Chang SC, Huang C, Li Y, Lee C, Chang M. 2005. Novel Nuclear Export Signal-Interacting Protein , NESI , Critical for the Assembly of Hepatitis Delta Virus Novel Nuclear Export Signal-Interacting Protein , NESI , Critical for the Assembly of Hepatitis Delta Virus 79:8113–8120.
97. Huang C, Jiang J-Y, Chang SC, Tsay Y-G, Chen M-R, Chang M-F. 2013. Nuclear export signal-interacting protein forms complexes with lamin A/C-Nups to mediate the CRM1-independent nuclear export of large hepatitis delta antigen. *J Virol* 87:1596–604.
98. Huang HC, Lee CP, Liu HK, Chang MF, Lai YH, Lee YC, Huang C. 2016. Cellular nuclear export factors TAP and Aly are required for HDAG-L-mediated assembly of hepatitis delta virus. *J Biol Chem* 291:26226–26238.
99. Tan K-P. 2004. Ser-123 of the large antigen of hepatitis delta virus modulates its cellular localization to the nucleolus, SC-35 speckles or the cytoplasm. *J Gen Virol* 85:1685–1694.
100. Lee C-ZZ, Chen P-JJ, Lai MMC, Chen D-SS. 1994. Isoprenylation of Large Hepatitis Delta Antigen Is Necessary but Not Sufficient for Hepatitis Delta Virus Assembly. *Virology* 199:169–175.
101. Yeh TS, Lo SJ, Chen PJ, Lee YH, Redeker AG, Gerlin JL, Houghton M, Lai MM. 1996. Casein kinase II and protein kinase C modulate hepatitis delta virus RNA replication but not empty viral particle assembly. *J Virol* 70:6190–8.
102. Chen CW, Tsay YG, Wu HL, Lee CH, Chen DS, Chen PJ. 2002. The double-stranded RNA-activated kinase, PKR, can phosphorylate hepatitis D virus small delta antigen at functional serine and threonine residues. *J Biol Chem* 277:33058–33067.
103. Li Y-J, Stallcup MR, Lai MMC. 2004. Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication. *J Virol* 78:13325–34.
104. Mu JJ, Tsay YG, Juan LJ, Fu TF, Huang WH, Chen DS, Chen PJ. 2004. The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine

72 may influence its cellular localization and viral RNA synthesis. *Virology* 319:60–70.

105. Tseng C-H, Cheng T-S, Shu C-Y, Jeng K-S, Lai MMC. 2010. Modification of Small Hepatitis Delta Virus Antigen by SUMO Protein. *J Virol* 84:918–927.
106. Hong S-Y, Chen P-J. 2010. Phosphorylation of serine 177 of the small hepatitis delta antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *J Virol* 84:1430–1438.
107. Chen Y-S, Huang W-H, Hong S-Y, Tsay Y-G, Chen P-J. 2008. ERK1/2-mediated phosphorylation of small hepatitis delta antigen at serine 177 enhances hepatitis delta virus antigenomic RNA replication. *J Virol* 82:9345–58.
108. O'Malley B, Lazinski DW. 2005. Roles of Carboxyl-Terminal and Farnesylated Residues in the Functions of the Large Hepatitis Delta Antigen. *J Virol* 79:1142.
109. Greco-Stewart V, Pelchat M. 2010. Interaction of host cellular proteins with components of the hepatitis Delta virus. *Viruses* 2:189–212.
110. Taylor J. 2006. Structure and replication of hepatitis delta virus RNA. *Hepat delta virus* 20–37.
111. Moroianu J, Hijikata M, Blobel G, Radu A. 1995. Mammalian karyopherin alpha 1 beta and alpha 2 beta heterodimers: alpha 1 or alpha 2 subunit binds nuclear localization signal and beta subunit interacts with peptide repeat-containing nucleoporins. *Proc Natl Acad Sci U S A* 92:6532–6.
112. Wang Y-CC, Huang C-RR, Chao M, Lo SJ. 2009. The c-terminal sequence of the large hepatitis delta antigen is variable but retains the ability to bind clathrin. *Virol J* 6:1–11.
113. Huang WH, Yung BYM, Syu WJ, Lee YHW. 2001. The Nucleolar Phosphoprotein B23 Interacts with Hepatitis Delta Antigens and Modulates the Hepatitis Delta Virus RNA Replication. *J Biol Chem* 276:25166–25175.
114. ter Haar E, Harrison SC, Kirchhausen T. 2000. Peptide-in-groove interactions link target proteins to the beta -propeller of clathrin. *Proc Natl Acad Sci* 97:1096–1100.
115. Narita T, Yamaguchi Y, Yano K, Chanarat S, Wada T, Kim D, Hasegawa J, Omori M, Inukai N, Sugimoto S, Endoh M, Yamada T, Handa H. 2003. Human Transcription Elongation Factor NELF : Identification of Novel Subunits and Reconstitution of the Functionally Active Complex Human Transcription Elongation Factor NELF : Identification of Novel Subunits and Reconstitution of the Functionally Active. *Mol Cell Biol* 23:1863–1873.
116. Yamaguchi Y, Handa H. Hepatitis Delta Antigen and RNA Polymerase II. *Landes Biosci*.
117. Haussecker D, Cao D, Huang Y, Parameswaran P, Fire AZ, Kay MA. 2008. Capped small RNAs and MOV10 in human hepatitis delta virus replication. *Nat Struct Mol Biol* 15:714–721.

118. Sikora D, Greco-Stewart VS, Miron P, Pelchat M. 2009. The hepatitis delta virus RNA genome interacts with eEF1A1, p54nrb, hnRNP-L, GAPDH and ASF/SF2. *Virology* 390:71–78.
119. Circle DA, Neel OD, Robertson HD, Clarke PA, Mathews MB. 1997. Surprising specificity of PKR binding to delta agent genomic RNA. *RNA* 3:438–48.
120. Mu J-J, Chen D-S, Chen P-J. 2001. The Conserved Serine 177 in the Delta Antigen of Hepatitis Delta Virus Is One Putative Phosphorylation Site and Is Required for Efficient Viral RNA Replication. *J Virol* 75:9087.
121. Bensaude O. 2011. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2:103–108.
122. Meyer-Siegler K, Mauro D, Seal G, Wurzer J, DeRiel J, Sirover M. 1991. A human nuclear uracil DNA glycosylase is the 37-kDa subunit of glyceraldehyde-3-phosphate dehydrogenase. *Proc Natl Acad Sci* 88:8460–8464.
123. Zheng L, Roeder RG, Luo Y. 2003. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell* 114:255–266.
124. Hara MR, Agrawal N, Kim SF, Cascio MB, Fujimuro M, Ozeki Y, Takahashi M, Cheah JH, Tankou SK, Hester LD, Ferris CD, Hayward SD, Snyder SH, Sawa A. 2005. S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nat Cell Biol* 7:665–674.
125. Tristan C, Shahani N, Sedlak TW, Sawa A. 2011. The diverse functions of GAPDH: Views from different subcellular compartments. *Cell Signal* 23:317–323.
126. Lin SS, Chang SC, Wang YH, Sun CY, Chang MF. 2000. Specific interaction between the hepatitis delta virus RNA and glyceraldehyde 3-phosphate dehydrogenase: An enhancement on ribozyme catalysis. *Virology* 271:46–57.
127. Beeharay Y, Goodrum G, Imperiale CJ, Pelchat M. 2018. The Hepatitis Delta Virus accumulation requires paraspeckle components and affects NEAT1 level and PSP1 localization. *Sci Rep* 8:1–12.
128. Sikora D, Zhang D, Bojic T, Beeharay Y, Tanara A, Pelchat M. 2013. Identification of a Binding Site for ASF/SF2 on an RNA Fragment Derived from the Hepatitis delta Virus Genome. *PLoS One* 8.
129. Greco-Stewart VS, Thibault CSL, Pelchat M. 2006. Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA. *Virology* 356:35–44.
130. Patton JG, Porro EB, Galceran J, Tempst P, Nadal-Ginard B. 1993. Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev* 7:393–406.
131. Emili A, Shales M, Mccracken S, Xie W, Tucker PW, Kobayashi R, Blencowe BJ, Ingles CJ. 2002. Splicing and transcription-associated proteins PSF and p54 nrb / NonO bind to the RNA polymerase II CTD 1102–1111.

132. Shav-Tal Y, Zipori D. 2002. PSF and p54 nrb /NonO - multi-functional nuclear proteins. *FEBS Lett* 531:109–114.
133. Goodrum G, Pelchat M, Goodrum G, Pelchat M. 2018. Insight into the Contribution and Disruption of Host Processes during HDV Replication. *Viruses* 11:21.
134. Glenn J, Watson J, Havel C, White J. 1992. Identification of a prenylation site in delta virus large antigen. *Science* (80-).
135. Otto JC, Casey PJ. 1996. The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. *J Biol Chem*.
136. Hwang SB, Lai MM. 1994. Isoprenylation masks a conformational epitope and enhances trans-dominant inhibitory function of the large hepatitis delta antigen. *J Virol* 68.
137. Huang W-H, Mai R-T, Lee Y-HW. 2008. Transcription factor YY1 and its associated acetyltransferases CBP and p300 interact with hepatitis delta antigens and modulate hepatitis delta virus RNA replication. *J Virol* 82:7313–7324.
138. Wang Y-H, Chang SC, Huang C, Li Y-P, Lee C-H, Chang M-F. 2005. Novel nuclear export signal-interacting protein, NESI, critical for the assembly of hepatitis delta virus. *J Virol* 79:8113–20.
139. Lee CH, Chang SC, Chen CJ, Chang MF. 1998. The nucleolin binding activity of hepatitis delta antigen is associated with nucleolus targeting. *J Biol Chem* 273:7650–7656.
140. Brazas R, Ganem D. 1996. A cellular homolog of hepatitis delta antigen: Implications for viral replication and evolution. *Science* (80-).
141. Lee C-ZZ, Sheu J-CC. 2008. Histone H1e interacts with small hepatitis delta antigen and affects hepatitis delta virus replication. *Virology* 375:197–204.
142. Choi SH, Jeong SH, Hwang SB. 2007. Large Hepatitis Delta Antigen Modulates Transforming Growth Factor- β Signaling Cascades: Implication of Hepatitis Delta Virus-Induced Liver Fibrosis. *Gastroenterology*.
143. Park C-Y, Oh S-H, Kang SM, Lim Y-S, Hwang SB. 2009. Hepatitis delta virus large antigen sensitizes to TNF-alpha-induced NF-kappaB signaling. *Mol Cells* 28:49–55.
144. Negro F, Korba BE, Forzani B, Baroudy BM, Brown TL, Gerin JL, Ponzetto A. 1989. Hepatitis delta virus (HDV) and woodchuck hepatitis virus (WHV) nucleic acids in tissues of HDV-infected chronic WHV carrier woodchucks. *J Virol* 63:1612–8.
145. Casey JL, Gerin JL. 2006. The Woodchuck Model of HDV Infection, p. 211–225. *In* *Hepatitis Delta Virus*. Springer Berlin Heidelberg.
146. Ponzetto A, Negro F, Popper H, Bonino F, Engle R, Rizzetto M, Purcell RH, Gerin JL. 1988. Serial passage of hepatitis delta virus in chronic hepatitis B virus carrier chimpanzees. *Hepatology* 8:1655–1661.
147. Negro F, Bergmann KF, Baroudy BM, Satterfield WC, Popper H, Purcell RH, Gerin JL. 1988. Chronic Hepatitis D Virus (HDV) Infection in Hepatitis B Virus Carrier

Chimpanzees Experimentally Superinfected with HDV. *J Infect Dis* 158:151–159.

148. Lütgehetmann M, Mancke L V., Volz T, Helbig M, Allweiss L, Bornscheuer T, Pollok JM, Lohse AW, Petersen J, Urban S, Dandri M. 2012. Humanized chimeric uPA mouse model for the study of hepatitis B and D virus interactions and preclinical drug evaluation. *Hepatology* 55:685–694.
149. Polo JM, Lim B, Govindarajan S, Lai MM. 1995. Replication of hepatitis delta virus RNA in mice after intramuscular injection of plasmid DNA. *J Virol* 69:5203–7.
150. Netter HJ, Kajino K, Taylor JM. 1993. Experimental transmission of human hepatitis delta virus to the laboratory mouse. *J Virol* 67:3357–62.
151. Hetzel U, Szirovicza L, Smura T, Prähauser B, Vapalahti O, Kipar A, Hepojoki J. 2018. Identification of a novel deltavirus in *Boa constrictor*. *bioRxiv* 429753.
152. Mota S, Mendes M, Penque D, Coelho A V., Cunha C. 2008. Changes in the proteome of Huh7 cells induced by transient expression of hepatitis D virus RNA and antigens. *J Proteomics* 71:0–8.
153. Chang J, Gudima SO, Tarn C, Nie X, Taylor JM. 2005. Development of a Novel System To Study Hepatitis Delta Virus Genome Replication. *Society* 79:8182–8188.
154. Weller ML, Gardener MR, Bogus ZC, Smith MA, Astorri E, Michael DG, Michael DA, Zheng C, Burbelo PD, Lai Z, Wilson PA, Swaim W, Handelsman B, Afione SA, Bombardieri M, Chiorini JA. 2016. Hepatitis Delta Virus Detected in Salivary Glands of Sjögren’s Syndrome Patients and Recapitulates a Sjögren’s Syndrome-Like Phenotype in Vivo. *Pathog Immun* 1:12–40.
155. Perez-Vargas J, Amirache F, Boson B, Mialon C, Freitas N, Sureau C, Fusil F, Cosset F-L. 2019. Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat Commun* 10:2098.
156. Vakrakou A, Karamichali E, Georgopoulou O, Manoussakis M. 2017. No Detection of an intriguing virus-like sequence in the salivary gland epithelial cells of sjÖgren’s syndrome patients. *BMJ* 76.
157. Mota S, Mendes M, Freitas N, Penque D, Coelho A V., Cunha C. 2009. Proteome analysis of a human liver carcinoma cell line stably expressing hepatitis delta virus ribonucleoproteins. *J Proteomics* 72:616–627.
158. Mendes M, Pérez-Hernandez D, Vázquez J, Coelho A V., Cunha C. 2013. Proteomic changes in HEK-293 cells induced by hepatitis delta virus replication. *J Proteomics* 89:24–38.
159. Kuo MY, Goldberg J, Coates L, Mason W, Gerin J, Taylor J. 1988. Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. *J Virol* 62:1855–61.
160. Cheng D, Yang A, Thomas H, Monjardino J. 1993. Characterization of stable hepatitis delta expressing hepatoma cell lines: effect of HDAG on cell growth. *Prog Clin Biol Res* 382:149–53.

161. Cunha C, Monjardino J, Cheng D, Krause S, Carmo-Fonseca M, Chang D. 1998. Localization of hepatitis delta virus RNA in the nucleus of human cells. *RNA* 4:680–93.
162. Liao F-T, Lee Y-J, Ko J-L, Tsai C-C, Tseng C-J, Sheu G-T. 2009. Hepatitis delta virus epigenetically enhances clusterin expression via histone acetylation in human hepatocellular carcinoma cells. *J Gen Virol* 90:1124–34.
163. Fox AH, Lam YW, Leung AKL, Lyon CE, Andersen J, Mann M, Lamond AI. 2002. Paraspeckles: A novel nuclear domain. *Curr Biol* 12:13–25.
164. Hirose T, Virnicchi G, Tanigawa A, Naganuma T, Li R, Kimura H, Yokoi T, Nakagawa S, Bénard M, Fox AH, Pierron G, Benard M, Fox AH, Pierron G. 2014. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell* 25:169–183.
165. Clemson CM, Hutchinson JN, Sara SA, W A, Fox AH, Chess A, Lawrence JB. 2010. An Architectural Role for a Nuclear Non-coding RNA: NEAT1 RNA is Essential for the Structure of Paraspeckles 33:717–726.
166. Radjef N, Lebon P, Williams V, Goffard A, Hober D, Fagard R, Kremsdorf D, Gordien E, Brichler S, Radjef N, Lebon P, Goffard A, Hober D, Fagard R, Kremsdorf D, Dény P, Gordien E. 2018. Hepatitis delta virus proteins repress hepatitis B virus enhancers and activate the alpha / beta interferon-inducible MxA gene. *J Gen Virol* 90:2759–2767.
167. Williams V, Brichler S, Khan E, Chami M, Dény P, Kremsdorf D, Gordien E. 2012. Large hepatitis delta antigen activates STAT-3 and NF-κB via oxidative stress. *J Viral Hepat* 19:744–753.
168. Waris G, Huh K won, Siddiqui A. 2001. Mitochondrially Associated Hepatitis B Virus X Protein Constitutively Activates Transcription Factors STAT-3 and NF- κB via Oxidative Stress. *Mol Cell Biol* 21:7721–7730.
169. Feng J, Meyer CA, Wang Q, Liu JS, Liu XS, Zhang Y. 2012. GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28:2782–2788.
170. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 13:2129–41.
171. Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q, Bader GD. 2010. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 26:2347–2348.
172. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. 2010. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. *Methods* 50:S1.
173. Protocol. 2009. [Bio-Rad] iQ™ SYBR Green Supermix Instruction Manual. Bio-Rad Technol Inc.
174. Ellis B, Haaland P, Hahne F, And NG, Spidlen J, Jiang M. 2018. flowCore: flowCore: Basic structures for flow cytometry data. R package version 1.46.2.

175. Auguie B, Antonov A. 2017. gridExtra: Miscellaneous Functions for ``Grid`` Graphics. 2.3.
176. Wickham H. 2018. reshape: Flexibly Reshape Data.
177. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 3.1.1.
178. Ekstrøm CT. Miscellaneous Esoteric Statistical Scripts.
179. Wierzbicki PM, Klacz J, Rybarczyk A, Slebiada T, Stanislawowski M, Wronska A, Kowalczyk A, Matuszewski M, Kmiec Z. 2014. Identification of a suitable qPCR reference gene in metastatic clear cell renal cell carcinoma. *Tumor Biol* 35:12473–12487.
180. Nde PN, Johnson CA, Pratap S, Cardenas TC, Kleshchenko YY, Furtak VA, Simmons KJ, Lima MF, Villalta F. 2010. Gene Network Analysis during Early Infection of Human Coronary Artery Smooth Muscle Cells by *Trypanosoma cruzi* and Its gp83 Ligand. *Chem Biodivers* 7:1051–1064.
181. Li J, Feng C, Lu Y, Li H, Tu Z, Liao G, Liang C. 2008. mRNA expression of the DNA replication-initiation proteins in epithelial dysplasia and squamous cell carcinoma of the tongue. *BMC Cancer* 8:395.
182. Zhang W, Kim PJ, Chen Z, Lokman H, Qiu L, Zhang K, Rozen SG, Tan EK, Je HS, Zeng L. 2016. MiRNA-128 regulates the proliferation and neurogenesis of neural precursors by targeting PCM1 in the developing cortex. *Elife* 5.
183. Deutscher MP. 2006. Degradation of RNA in bacteria : comparison of mRNA and stable RNA. *Nucleic Acids Res* 34:659–666.
184. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10.
185. Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111.
186. Huang L, Yuan Z, Liu P, Zhou T. 2015. Effects of promoter leakage on dynamics of gene expression. *BMC Syst Biol* 9:16.
187. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–476.
188. Wahlstedt H, Daniel C, Enstero M, Ohman M. 2009. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome Res* 19:978–986.
189. Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:R14.
190. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, Sherlock

- G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–9.
191. Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6.
 192. Tanaka Y, Nakamura A, Morioka MS, Inoue S, Tamamori-Adachi M, Yamada K, Taketani K, Kawauchi J, Tanaka-Okamoto M, Miyoshi J, Tanaka H, Kitajima S. 2011. Systems Analysis of ATF3 in Stress Response and Cancer Reveals Opposing Effects on Pro-Apoptotic Genes in p53 Pathway. *PLoS One* 6:e26848.
 193. Zhan Q. 2005. Gadd45a , a p53- and BRCA1-regulated stress protein , in cellular response to DNA damage. *Mutat Res Mol Mech Mutagen* 569:133–143.
 194. DiLeo M V., Strahan GD, den Bakker M, Hoekenga OA. 2011. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLoS One* 6.
 195. Griffin TJ. 2002. Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics* 1:323–333.
 196. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M. 2014. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol* 7:548–548.
 197. Xiao SH, Manley JL. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* 11:334–344.
 198. Dubois J, Terrier O, Rosa-Calatrava M. 2014. Influenza viruses and mRNA splicing: doing more with less. *MBio* 5:e00070-14.
 199. Swanton C, Jones N. 2001. Strategies in subversion: de-regulation of the mammalian cell cycle by viral gene products. *Int J Exp Pathol* 82:3–13.
 200. Hill SJ. 2006. G-protein-coupled receptors: past, present and future. *Br J Pharmacol* 147 Suppl:S27-37.
 201. Arvanitakis L, Geras-Raaka E, Varma A, Gershengorn MC, Cesarman E. 1997. Human herpesvirus KSHV encodes a constitutively active G-protein-coupled receptor linked to cell proliferation. *Nature* 385:347–350.
 202. Sodhi A, Montaner S, Gutkind JS. 2004. Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat Rev Mol Cell Biol* 5:998–1012.
 203. Cordeaux Y, Hill SJ. 2002. Mechanisms of Cross-Talk between G-Protein-Coupled Receptors. *Neurosignals* 11:45–57.
 204. Naor Z, Benard O, Seger R. 2000. Activation of MAPK cascades by G-protein-coupled receptors: The case of gonadotropin-releasing hormone receptor. *Trends Endocrinol Metab* 11:91–99.
 205. Olins AL, Olins DE. 1974. Spheroid chromatin units (v bodies). *Science* 183:330–2.
 206. Olins DE, Olins AL. 2003. Chromatin history: our view from the bridge. *Nat Rev Mol*

Cell Biol 4:809–814.

207. Kornberg RD. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science* 184:868–71.
208. Hergeth SP, Schneider R. 2015. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep* 16:1439–1453.
209. Li G, Zhu P. 2015. Structure and organization of chromatin fiber in the nucleus.
210. Baatout S, Derradji H. 2006. About histone H1 phosphorylation during mitosis. *Cell Biochem Funct* 24:93–94.
211. Matsumoto Y, Yasuda H, Mita S, Marunouchi T, Yamada M. 1980. Evidence for the involvement of H1 histone phosphorylation in chromosome condensation. *Nature* 284:181–183.
212. Hergeth SP, Dundr M, Tropberger P, Zee BM, Garcia BA, Daujat S, Schneider R. 2011. Isoform-specific phosphorylation of human linker histone H1.4 in mitosis by the kinase Aurora B. *J Cell Sci* 124:1623–8.
213. Sarg B, Helliger W, Talasz H, Förg B, Lindner HH. 2006. Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. *J Biol Chem* 281:6573–80.
214. Morales V, Richard-Foy H. 2000. Role of Histone N-Terminal Tails and Their Acetylation in Nucleosome Dynamics. *Mol Cell Biol* 20:7230–7237.
215. Linger JG, Tyler JK. 2008. Chromatin Disassembly and Reassembly During DNA Repair. *Biochemistry* 618:52–64.
216. Spector DL. 2003. The Dynamics of Chromosome Organization and Gene Regulation. *Annu Rev Biochem* 72:573–608.
217. Park J-H, Park E-J, Hur S-K, Kim S, Kwon J. 2009. Mammalian SWI/SNF chromatin remodeling complexes are required to prevent apoptosis after DNA damage. *DNA Repair (Amst)* 8:29–39.
218. Alberts B, Johnson A LJ. 2002. The Self-Assembly and Dynamic Structure of Cytoskeletal Filaments. *Molecular Biology of the Cell*, 4th ed.
219. Griffiths G, Fuller SD, Back R, Hollinshead M, Pfeiffer S, Simons K. 1989. The dynamic nature of the Golgi complex. *J Cell Biol* 108:277–97.
220. Alberts B, Johnson A, Lewis J. 2014. *Molecular Biology of the Cell* 6 edition.
221. Ask K, Jasencakova Z, Menard P, Feng Y, Almouzni G, Groth A. 2012. Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *EMBO J* 31:2013–2023.
222. Turki-Judeh W, Courey AJ. 2011. Groucho: A Corepressor with Instructive Roles in Development, p. 1607–1607. *In* *Encyclopedia of Cancer*, 1st ed. Springer Berlin Heidelberg, Berlin, Heidelberg.
223. Agarwal M, Kumar P, Mathew SJ. 2015. The Groucho/Transducin-like enhancer of

split protein family in animal development. *IUBMB Life* 67:472–481.

224. Davie JK, Dent SYR. 2004. Histone Modifications in Corepressor Functions. *Curr Top Dev Biol* 59:145–163.
225. Vintermist A, Böhm S, Sadeghifar F, Louvet E, Mansén A, Percipalle P, Östlund Farrants AK. 2011. The chromatin remodelling complex B-WICH changes the chromatin structure and recruits histone acetyl-transferases to active rRNA genes. *PLoS One* 6.
226. Percipalle P, Farrants AKÖ. 2006. Chromatin remodelling and transcription: be-WICHed by nuclear myosin 1. *Curr Opin Cell Biol* 18:267–274.
227. Nigg EA. 2001. Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol* 2:21–32.
228. Kassardjian A, Rizkallah R, Riman S, Renfro SH, Alexander KE, Hurt MM. 2012. The Transcription Factor YY1 Is a Novel Substrate for Aurora B Kinase at G2/M Transition of the Cell Cycle. *PLoS One* 7:e50645.
229. Crosio C, Fimia GM, Loury R, Kimura M, Okano Y, Zhou H, Sen S, Allis CD, Sassone-Corsi P. 2002. Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases. *Mol Cell Biol* 22:874–85.
230. Carmena M, Wheelock M, Funabiki H, Earnshaw WC. 2012. The chromosomal passenger complex (CPC): From easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol* 13:789–803.
231. Ruchaud S, Carmena M, Earnshaw WC. 2007. Chromosomal passengers: conducting cell division. *Nat Rev Mol Cell Biol* 8:798–812.
232. Morrow CJ. 2005. Bub1 and aurora B cooperate to maintain BubR1-mediated inhibition of APC/CCdc20. *J Cell Sci* 118:3639–3652.
233. Yu H. 2002. Regulation of APC–Cdc20 by the spindle checkpoint. *Curr Opin Cell Biol* 14:706–714.
234. Tien JF, Umbreit NT, Gestaut DR, Franck AD, Cooper J, Wordeman L, Gonen T, Asbury CL, Davis TN. 2010. Cooperation of the Dam1 and Ndc80 kinetochore complexes enhances microtubule coupling and is regulated by aurora B. *J Cell Biol* 189:713–723.
235. Kline-Smith SL, Sandall S, Desai A. 2005. Kinetochore-spindle microtubule interactions during mitosis. *Curr Opin Cell Biol* 17:35–46.
236. Lampson MA, Cheeseman IM. 2011. Sensing centromere tension: Aurora B and the regulation of kinetochore function. *Trends Cell Biol* 21:133–140.
237. Davy C, Doorbar J. 2007. G2/M cell cycle arrest in the life cycle of viruses. *Virology* 368:219–226.
238. Lomonte P, Sullivan KF, Everett RD. 2001. Degradation of Nucleosome-associated Centromeric Histone H3-like Protein CENP-A Induced by Herpes Simplex Virus Type 1 Protein ICP0. *J Biol Chem* 276:5829–5835.

239. Everett RD, Earnshaw WC, Findlay J, Lomonte P. 1999. Specific destruction of kinetochore protein CENP-C and disruption of cell division by herpes simplex virus immediate-early protein Vmw110. *EMBO J* 18:1526–38.
240. Greenbaum D, Colangelo C, Williams K, Gerstein M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol. BioMed Central*.
241. Harford, J. B., and Morris DR. 1997. Cis-Acting mRNA structures in gene-specific translational control, p. 165–180. *In* *Metabolism and Post-transcriptional Gene Regulation*.
242. Zhao BS, Roundtree IA, He C. 2016. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol*.
243. Sokhi UK, Das SK, Dasgupta S, Emdad L, Shiang R, DeSalle R, Sarkar D, Fisher PB. 2013. Human Polynucleotide Phosphorylase (hPNPaseold-35), p. 161–190. *In* *Advances in Cancer Research*. Academic Press.
244. Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik A V., Patil DP, Linder B, Pickering BF, Vasseur J-J, Chen Q, Gross SS, Elemento O, Debart F, Kiledjian M, Jaffrey SR. 2017. Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature* 541:371–375.
245. Roundtree IA, Evans ME, Pan T, He C. 2017. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 169:1187–1200.
246. Harper J V., Brooks G. 2005. The Mammalian Cell Cycle: An Overview, p. 113–154. *In* *Cell Cycle Control*. Humana Press, New Jersey.
247. Satyanarayana A, Kaldis P. 2009. Mammalian cell-cycle regulation: several Cdk, numerous cyclins and diverse compensatory mechanisms. *Oncogene* 28:2925–2939.
248. Barnum KJ, O'Connell MJ. 2014. Cell Cycle Regulation by Checkpoints, p. 29–40. *In* *Methods in molecular biology* (Clifton, N.J.). NIH Public Access.
249. Alushin GM, Ramey VH, Pasqualato S, Ball DA, Grigorieff N, Musacchio A, Nogales E. 2010. The Ndc80 kinetochore complex forms oligomeric arrays along microtubules. *Nature* 467:805–810.
250. Jeyaprakash AA, Klein UR, Lindner D, Ebert J, Nigg EA, Conti E. 2007. Structure of a Survivin-Borealin-INCENP Core Complex Reveals How Chromosomal Passengers Travel Together. *Cell* 131:271–285.

CONTRIBUTION OF COLLABORATORS

Data processing from the raw sequencing files was performed by Lynda Rocheleau which included: cutadapt, tophat, and Gfold. Moreover, the WGCNA analysis was also performed by Rocheleau. The flow cytometry data collection on LSR Fortessa cell analyzer was performed by Vera Tang. Flow cytometry data analysis including bioinformatics scripts, data analysis and figures were performed in collaboration with Martin Pelchat.

SUPPLEMENTARY

Table S1. GO enrichment analysis of the 283 genes down regulated by HDV RNA. A total of 19 biological processes were enriched. GOBPID (Gene Ontology Biological Process ID). P-values were determined by PANTHER Overrepresentation Test using the Bonferroni correction.

Term	GOBPID	P-value
innate immune response in mucosa	GO:0002227	4.40E-03
nucleosome assembly	GO:0006334	1.10E-20
mucosal immune response	GO:0002385	3.14E-02
chromatin silencing	GO:0006342	6.49E-09
chromatin assembly	GO:0031497	1.66E-19
organ or tissue specific immune response	GO:0002251	4.43E-02
nucleosome organization	GO:0034728	1.59E-18
chromatin assembly or disassembly	GO:0006333	3.89E-18
negative regulation of gene expression, epigenetic	GO:0045814	8.68E-08
DNA packaging	GO:0006323	3.09E-18
protein-DNA complex assembly	GO:0065004	8.50E-16
protein-DNA complex subunit organization	GO:0071824	2.14E-14
DNA conformation change	GO:0071103	1.94E-15
gene silencing	GO:0016458	3.93E-05
regulation of gene expression, epigenetic	GO:0040029	1.90E-03
chromatin organization	GO:0006325	1.12E-14
chromosome organization	GO:0051276	3.56E-11
cellular macromolecular complex assembly	GO:0034622	3.66E-04
macromolecular complex assembly	GO:0065003	2.80E-02

Table S2. GO enrichment analysis of the 3 278 genes up regulated by HDV RNA. A total of 130 biological processes were enriched. P-values were determined by PANTHER Overrepresentation Test using the Bonferroni correction.

Term	GOBPID	P-value
tRNA modification	GO:0006400	3.69E-04
RNA methylation	GO:0001510	1.38E-02
RNA modification	GO:0009451	1.22E-05
tRNA processing	GO:0008033	2.16E-03

ncRNA processing	GO:0034470	3.80E-12
ribosome biogenesis	GO:0042254	6.16E-09
tRNA metabolic process	GO:0006399	1.61E-03
rRNA metabolic process	GO:0016072	5.62E-07
rRNA processing	GO:0006364	4.86E-06
ncRNA metabolic process	GO:0034660	1.89E-14
ribonucleoprotein complex biogenesis	GO:0022613	1.88E-10
RNA processing	GO:0006396	1.31E-18
microtubule-based movement	GO:0007018	2.82E-02
cilium assembly	GO:0060271	1.48E-02
cilium organization	GO:0044782	4.31E-02
plasma membrane bounded cell projection assembly	GO:0120031	4.38E-03
cell projection assembly	GO:0030031	7.27E-03
RNA splicing	GO:0008380	1.11E-02
RNA metabolic process	GO:0016070	8.93E-50
mRNA processing	GO:0006397	4.85E-03
nucleic acid metabolic process	GO:0090304	8.37E-52
microtubule-based process	GO:0007017	5.40E-04
transcription, DNA-templated	GO:0006351	1.13E-28
nucleic acid-templated transcription	GO:0097659	1.25E-28
RNA biosynthetic process	GO:0032774	1.00E-28
gene expression	GO:0010467	2.87E-45
covalent chromatin modification	GO:0016569	2.29E-02
intracellular protein transport	GO:0006886	1.09E-05
nucleobase-containing compound biosynthetic process	GO:0034654	5.55E-26
heterocycle biosynthetic process	GO:0018130	5.95E-26
cellular response to DNA damage stimulus	GO:0006974	1.28E-03
aromatic compound biosynthetic process	GO:0019438	1.33E-25
organelle assembly	GO:0070925	8.18E-03
nucleobase-containing compound metabolic process	GO:0006139	6.89E-42
organic cyclic compound biosynthetic process	GO:1901362	6.40E-24
protein transport	GO:0015031	4.66E-08
heterocycle metabolic process	GO:0046483	9.58E-39
cellular nitrogen compound biosynthetic process	GO:0044271	3.23E-25
cellular aromatic compound metabolic process	GO:0006725	2.29E-37
peptide transport	GO:0015833	2.90E-07
establishment of protein localization	GO:0045184	7.67E-08
amide transport	GO:0042886	8.15E-07
cellular macromolecule biosynthetic process	GO:0034645	3.17E-24
intracellular transport	GO:0046907	7.49E-07
cellular nitrogen compound metabolic process	GO:0034641	1.14E-37
regulation of RNA metabolic process	GO:0051252	2.01E-24

cellular macromolecule localization	GO:0070727	9.64E-06
macromolecule biosynthetic process	GO:0009059	2.09E-23
cellular protein localization	GO:0034613	1.15E-05
negative regulation of transcription from RNA polymerase II promoter	GO:0000122	3.83E-02
organic cyclic compound metabolic process	GO:1901360	1.09E-33
regulation of nucleic acid-templated transcription	GO:1903506	4.22E-21
regulation of RNA biosynthetic process	GO:2001141	4.68E-21
regulation of transcription, DNA-templated	GO:0006355	1.27E-20
nitrogen compound transport	GO:0071705	8.23E-07
regulation of cellular macromolecule biosynthetic process	GO:2000112	7.36E-23
protein modification by small protein conjugation or removal	GO:0070647	1.06E-02
regulation of nucleobase-containing compound metabolic process	GO:0019219	4.24E-22
cellular macromolecule metabolic process	GO:0044260	4.09E-47
regulation of macromolecule biosynthetic process	GO:0010556	1.17E-21
protein localization	GO:0008104	2.85E-07
establishment of localization in cell	GO:0051649	1.97E-05
regulation of gene expression	GO:0010468	4.55E-23
cellular response to stress	GO:0033554	1.19E-04
regulation of cellular biosynthetic process	GO:0031326	4.42E-18
negative regulation of cellular macromolecule biosynthetic process	GO:2000113	8.34E-03
single-organism organelle organization	GO:1902589	1.07E-02
cellular localization	GO:0051641	4.29E-06
negative regulation of RNA metabolic process	GO:0051253	3.05E-02
organelle organization	GO:0006996	1.75E-10
regulation of biosynthetic process	GO:0009889	2.67E-17
cellular biosynthetic process	GO:0044249	4.60E-18
negative regulation of nucleobase-containing compound metabolic process	GO:0045934	2.29E-02
macromolecule localization	GO:0033036	1.53E-05
biosynthetic process	GO:0009058	1.41E-16
negative regulation of macromolecule biosynthetic process	GO:0010558	4.20E-02
macromolecule metabolic process	GO:0043170	4.80E-35
organic substance biosynthetic process	GO:1901576	5.08E-16
cellular component biogenesis	GO:0044085	3.96E-06
intracellular signal transduction	GO:0035556	3.16E-02
negative regulation of gene expression	GO:0010629	2.42E-02
regulation of cellular metabolic process	GO:0031323	5.47E-20
regulation of macromolecule metabolic process	GO:0060255	3.08E-19
organic substance transport	GO:0071702	2.97E-03
regulation of nitrogen compound metabolic process	GO:0051171	1.40E-17
macromolecule modification	GO:0043412	1.41E-07
regulation of primary metabolic process	GO:0080090	5.17E-17

cellular metabolic process	GO:0044237	1.20E-32
regulation of metabolic process	GO:0019222	1.97E-17
nitrogen compound metabolic process	GO:0006807	5.04E-28
regulation of cellular component organization	GO:0051128	2.30E-02
primary metabolic process	GO:0044238	4.06E-26
organic substance metabolic process	GO:0071704	9.06E-25
cellular protein modification process	GO:0006464	3.01E-03
protein modification process	GO:0036211	3.01E-03
metabolic process	GO:0008152	1.66E-25
cellular protein metabolic process	GO:0044267	6.46E-04
cellular component organization or biogenesis	GO:0071840	3.41E-07
cellular component organization	GO:0016043	2.11E-03
regulation of cellular process	GO:0050794	2.59E-05
regulation of biological process	GO:0050789	4.70E-03
cellular process	GO:0009987	1.16E-09
biological regulation	GO:0065007	1.96E-03
biological_process	GO:0008150	3.74E-07
response to stimulus	GO:0050896	3.28E-03
multicellular organismal process	GO:0032501	3.73E-04
response to chemical	GO:0042221	4.28E-02
Unclassified	UNCLASSIFIED	0.00E+00
immune system process	GO:0002376	5.29E-04
carbohydrate derivative metabolic process	GO:1901135	3.79E-03
immune response	GO:0006955	1.56E-06
immune effector process	GO:0002252	6.23E-04
system process	GO:0003008	1.69E-10
leukocyte mediated immunity	GO:0002443	1.24E-03
neurological system process	GO:0050877	1.20E-11
humoral immune response	GO:0006959	1.83E-02
defense response to bacterium	GO:0042742	3.10E-02
sensory perception	GO:0007600	4.90E-16
G-protein coupled receptor signaling pathway	GO:0007186	2.53E-24
keratinization	GO:0031424	2.26E-02
detection of stimulus	GO:0051606	9.36E-17
complement activation	GO:0006956	2.79E-02
protein activation cascade	GO:0072376	1.76E-03
humoral immune response mediated by circulating immunoglobulin	GO:0002455	1.31E-02
detection of stimulus involved in sensory perception	GO:0050906	1.42E-17
sensory perception of chemical stimulus	GO:0007606	2.05E-21
complement activation, classical pathway	GO:0006958	2.49E-03
detection of chemical stimulus	GO:0009593	3.72E-21
detection of chemical stimulus involved in sensory perception	GO:0050907	1.58E-20

sensory perception of smell	GO:0007608	3.16E-21
-----------------------------	------------	----------

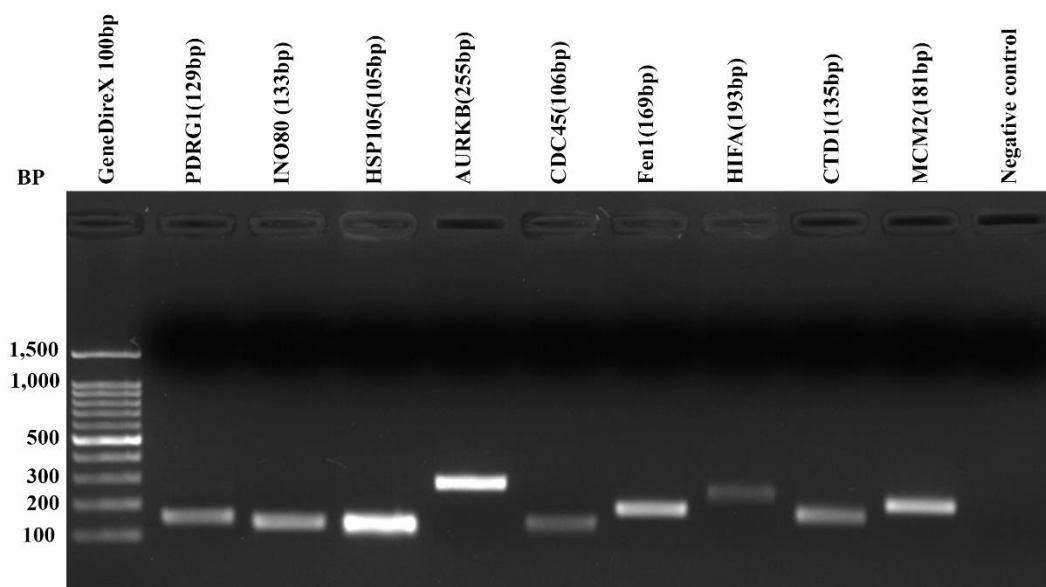


Figure S1. Validation of designed primers using PCR.

Primers were designed using Primer-BLAST from NCBI. Primers were purchased from ThermoFisher Scientific and were resuspended with ddH₂O. HEK-293 cells were plated and grown until reaching confluence followed by a total RNA extraction using TRIzol. The RNA integrity was accessed and was reverse-transcribed using iScript cDNA synthesis kit (Biorad). A PCR was used to observe the specificity and the length of the amplicon for each primer pair and run on a 1.5% agarose gel. The DNA ladder use was GeneDireX 100bp (10 μ L with 1 μ L of SYBR Green I). For the samples, 6 μ L of DNA was used with 4 μ L of loading dye and 1 μ L of SYBR Green I. The negative control include the same master mix used for the sample minus the DNA (using GAPDH primers).

Figure S2. Script for gene classification according to expression change.

```
1. library(ggplot2)
2. library(reshape2)
3.
4. #####
5. #####loading data without tetracycline
6. #####
7. #you can set your working directory here if not using default
8. #setwd()
9.
10. ###Table 293-Ag
11. data.293.Ag <- read.table("293vsAg.diff", sep="\t", header=FALSE)
12. colnames(data.293.Ag )<-c("GeneSymbole", "GeneName", "Ag", "E-
FDR", "log2fdc", "stRPKM", "ndRPKM")
13.
14. ##Table 293-HDV
15. data.293.HDV <- read.table("293vsHDV.diff", sep="\t", header=F)
16. colnames(data.293.HDV )<-c("GeneSymbole", "GeneName", "HDV", "E-
FDR", "log2fdc", "stRPKM", "ndRPKM")
17.
18. #Keep only RPKM values >0
19. data.293.Ag <- data.293.Ag[data.293.Ag$stRPKM >0 & data.293.Ag$ndRPKM>0,]
20. data.293.HDV <- data.293.HDV[data.293.HDV$stRPKM >0 & data.293.HDV$ndRPKM>0,]
21.
22. ##Only keep columns 1 and 3 (GeneSymbole and AG or HDV)
23. data.293.Ag <- data.293.Ag[,c(1,3)]
24. data.293.HDV <- data.293.HDV[,c(1,3)]
25.
26. ##### Merge those table. Default 293 value is=0 since Ag and HDV condition were comp
ares to 293.
27. data.293.Ag.HDV <- merge(data.293.Ag, data.293.HDV, by="GeneSymbole", all=FALSE) #
#all=FALSE puisque on veut juste ceux qui sont communs aux deux et pas seulement to
us les genes.
28. data.293.Ag.HDV$C293 <- 0
29.
30. #Change values to linear values
31. data.293.Ag.HDV <- data.293.Ag.HDV[,c(1,4,2,3)]
32. data.293.Ag.HDVpaslog <- data.293.Ag.HDV
33. data.293.Ag.HDVpaslog[,2:4] <- 2^(data.293.Ag.HDVpaslog[,2:4])
34.
35.
36. #####
37. #####loading data with tetracycline
38. #####
39.
40.
41. ##Table 293-Ag
42. data.293.AgTET <- read.table("293-TETvsAg-TET.diff", sep="\t", header=FALSE)
43. colnames(data.293.AgTET )<-c("GeneSymbole", "GeneName", "Ag", "E-
FDR", "log2fdc", "stRPKM", "ndRPKM")
44.
45. ##Table 293-HDV
46. data.293.HDVTET <- read.table("293-TETvsHDV-TET.diff", sep="\t", header=F)
47. colnames(data.293.HDVTET )<-c("GeneSymbole", "GeneName", "HDV", "E-
FDR", "log2fdc", "stRPKM", "ndRPKM")
48.
49. #Keep only RPKM values >0
50. data.293.AgTET <- data.293.AgTET[data.293.AgTET$stRPKM >0 & data.293.AgTET$ndRPKM>0
,]
```

```

51. data.293.HDVTET <- data.293.HDVTET[data.293.HDVTET$stRPKM >0 & data.293.HDVTET$ndRP
    KM>0,]
52. data.293.AgTET <- data.293.AgTET[,c(1,3)]
53. data.293.HDVTET <- data.293.HDVTET[,c(1,3)]
54.
55. ## Merge those table. Default 293 value is=0 since Ag and HDV condition were compar
    es to 293.
56. data.293.Ag.HDVTET <- merge(data.293.AgTET, data.293.HDVTET, by="GeneSymbole", all=
    FALSE)
57. data.293.Ag.HDVTET$C293 <- 0
58. data.293.Ag.HDVTET <- data.293.Ag.HDVTET[,c(1,4,2,3)]
59. #Change values to linear values
60. data.293.Ag.HDVTETpaslog <- data.293.Ag.HDVTET
61. data.293.Ag.HDVTETpaslog[,2:4] <- 2^(data.293.Ag.HDVTETpaslog[,2:4])
62.
63. #####
64.
65. ##### Value chosen for cutoff
66. a<-2
67.
68.
69. #####
70. ##CREATING A TABLES WITH ALL CATEGORIES OF INDUCED AND NON-INDUCED DATA
71. #####
72.
73. #MERGE BOTH TABLE CREATED PREVIOUSLY
74.
75. Tableau.Final.Comparaison<- merge(data.293.Ag.HDVPaslog, data.293.Ag.HDVTETpaslog,
    by="GeneSymbole", all=TRUE)
76.
77. colnames(Tableau.Final.Comparaison) <-
    c("GeneSymbole", "C293", "Ag", "HDV", "C293 +TET", "Ag +TET", "HDV +TET")
78.
79. #create columns for comparison
80. Tableau.Final.Comparaison$diffAg<- Tableau.Final.Comparaison[,6]/Tableau.Final.Comp
    araison[,3]
81. Tableau.Final.Comparaison$diffHDV<- Tableau.Final.Comparaison[,7]/Tableau.Final.Com
    paraison[,4]
82.
83. ##### Creatting the different categoriesn
84. ## new column :SansTET, AvecTET, Difference
85. ## Different categories: E=Equal, U=Up-regulated, D=Down-
    regulated 0=RPMK 0(1er position = 293 à Ag, 2E position= Ag à HDV)
86.
87. #####
88. # CATEGORIES pour les sans TET
89. #####
90.
91. Tableau.Final.Comparaison$SansTET<- ""
92. Tableau.Final.Comparaison$AvecTET<- ""
93. Tableau.Final.Comparaison$Difference<- ""
94. #
95. # #tableau Ag + (UU, UE, UD)
96. Tableau.Final.Comparaison[ !is.na(Tableau.Final.Comparaison$Ag)& !is.na(Tableau.Fi
    nal.Comparaison$HDV) & Tableau.Final.Comparaison$Ag >= a & (Tableau.Final.Compara
    ison$HDV >= (Tableau.Final.Comparaison$Ag)*a),]$SansTET <- "UU"
97. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Fi
    nal.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag >= a) & (Tableau.Final.Compar
    aison$HDV < ((Tableau.Final.Comparaison$Ag)*a)) & (Tableau.Final.Comparaiso
    n$HDV > ((Tableau.Final.Comparaison$Ag)*(1/a))),]$SansTET <- "UE"

```

```

98. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag >= a) & ((Tableau.Final.Comparaison$HDV <= ((Tableau.Final.Comparaison$Ag)*(1/a) ))) , ]$SansTET <- "UD"
99.
100.
101. # #tableau Ag = (EU, EE, ED)
102.
103. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag < a & Tableau.Final.Comparaison$Ag > (1/a)) & (Tableau.Final.Comparaison$HDV >= (Tableau.Final.Comparaison$Ag)*a ) , ]$SansTET <- "EU"
104. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag < a & Tableau.Final.Comparaison$Ag > (1/a)) & ((Tableau.Final.Comparaison$HDV < (Tableau.Final.Comparaison$Ag)*a ) & (Tableau.Final.Comparaison$HDV > (Tableau.Final.Comparaison$Ag)*(1/a))) , ]$SansTET <- "EE"
105. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag < a & Tableau.Final.Comparaison$Ag > (1/a)) & (Tableau.Final.Comparaison$HDV <= (Tableau.Final.Comparaison$Ag)*(1/a) ) , ]$SansTET <- "ED"
106.
107.
108. # #tableau Ag = (DU, DE, DD)
109. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag <= (1/a)) & (Tableau.Final.Comparaison$HDV >= (Tableau.Final.Comparaison$Ag)*a ) , ]$SansTET <- "DU"
110. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag <= (1/a)) & (Tableau.Final.Comparaison$HDV < ((Tableau.Final.Comparaison$Ag)*a) & (Tableau.Final.Comparaison$HDV > ((Tableau.Final.Comparaison$Ag)*(1/a)))) , ]$SansTET <- "DE"
111. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison$Ag) & !is.na(Tableau.Final.Comparaison$HDV) & (Tableau.Final.Comparaison$Ag <= (1/a)) & (Tableau.Final.Comparaison$HDV <= ((Tableau.Final.Comparaison$Ag)*(1/a))) , ]$SansTET <- "DD"
112.
113. #tableau RPKM=0
114. Tableau.Final.Comparaison[is.na(Tableau.Final.Comparaison$Ag) , ]$SansTET <- "RPKM0"
115.
116.
117.
118.
119.
120. #####
121. # CATEGORIES AVEC TET
122. #####
123.
124. Tableau.Final.Comparaison$AvecTET <- ""
125. Tableau.Final.Comparaison$Difference <- ""
126. #
127. # #tableau Ag + (UU, UE, UD)
128. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Tableau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] >= a & (Tableau.Final.Comparaison[,7] >= (Tableau.Final.Comparaison[,6])*a) , ]$AvecTET <- "UU"
129. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Tableau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] >= a) & (Tableau.Final.Comparaison[,7] < ((Tableau.Final.Comparaison[,6])*a)) & (Tableau.Final.Comparaison[,7] > ((Tableau.Final.Comparaison[,6])*(1/a))) , ]$AvecTET <- "UE"
130. Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Tableau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] >= a) & ((Tableau.F

```

```

    inal.Comparaison[,7] <= ((Tableau.Final.Comparaison[,6])*(1/a) )) , ]$AvecTET
    <- "UD"
131.
132.
133.     # #tableau Ag = (EU, EE, ED)
134.
135.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] < a & Tableau.Final
    .Comparaison[,6] > (1/a)) & (Tableau.Final.Comparaison[,7] >= (Tableau.Final.Co
    mparaison[,6])*a ) , ]$AvecTET <- "EU"
136.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] < a & Tableau.Final
    .Comparaison[,6] > (1/a)) & ((Tableau.Final.Comparaison[,7] < (Tableau.Final.Co
    mparaison[,6])*a ) & (Tableau.Final.Comparaison[,7] > (Tableau.Final.Comparais
    on[,6])*(1/a))) , ]$AvecTET <- "EE"
137.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6] < a & Tableau.Final
    .Comparaison[,6] > (1/a)) & (Tableau.Final.Comparaison[,7] <= (Tableau.Final.Co
    mparaison[,6])*(1/a) ) , ]$AvecTET <- "ED"
138.
139.
140.     # #tableau Ag - (DU, DE, DD)
141.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6]<= (1/a)) & (Table
    au.Final.Comparaison[,7] >= (Tableau.Final.Comparaison[,6])*a ) , ]$AvecTET<- "D
    U"
142.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6]<= (1/a)) & (Table
    au.Final.Comparaison[,7] < ((Tableau.Final.Comparaison[,6])*a ) & (Tableau.Final.
    Comparaison[,7] > ((Tableau.Final.Comparaison[,6])*(1/a)))) , ]$AvecTET<- "DE"
143.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,6]) & !is.na(Ta
    bleau.Final.Comparaison[,7]) & (Tableau.Final.Comparaison[,6]<= (1/a)) & (Table
    au.Final.Comparaison[,7] <= ((Tableau.Final.Comparaison[,6])*(1/a))) , ]$AvecTET<
    - "DD"
144.
145.     #tableau RPKM=0
146.     Tableau.Final.Comparaison[is.na(Tableau.Final.Comparaison[,6]) , ]$AvecTET <-
    "RPKM0"
147.
148.
149.
150.
151.     #####
152.     # CATEGORIES DIFFÉRENCE
153.     #####
154.
155.     Tableau.Final.Comparaison$Difference<- ""
156.     #
157.     # #tableau Ag + (UU, UE, UD)
158.     Tableau.Final.Comparaison[ !is.na(Tableau.Final.Comparaison[,8]) & !is.na(T
    ableau.Final.Comparaison[,9]) & Tableau.Final.Comparaison[,8] >= a & (Tableau.Fin
    al.Comparaison[,9] >= (Tableau.Final.Comparaison[,8])*a), ]$Difference <- "UU"
159.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(
    Tableau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8] >= a) & (Tableau.F
    inal.Comparaison[,9] < ((Tableau.Final.Comparaison[,8])*a)) & (Tableau.Fina
    l.Comparaison[,9] > ((Tableau.Final.Comparaison[,8])*(1/a))), ]$Difference <- "UE"
160.     Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(
    Tableau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8] >= a) & ((Tableau.F
    inal.Comparaison[,9] <= ((Tableau.Final.Comparaison[,8])*(1/a) )) , ]$Differen
    ce <- "UD"
161.
162.

```

```

163.      # #tableau Ag = (EU, EE, ED)
164.
165.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8] < a & Tableau.Final
.Comparaison[,8] > (1/a)) & (Tableau.Final.Comparaison[,9] >= (Tableau.Final.Co
mparaison[,8])*a ) ,]$Difference <- "EU"
166.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8] < a & Tableau.Final
.Comparaison[,8] > (1/a)) & ((Tableau.Final.Comparaison[,9] < (Tableau.Final.Co
mparaison[,8])*a ) & (Tableau.Final.Comparaison[,9] > (Tableau.Final.Comparais
on[,8])*(1/a))) ,]$Difference <- "EE"
167.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8] < a & Tableau.Final
.Comparaison[,8] > (1/a)) & (Tableau.Final.Comparaison[,9] <= (Tableau.Final.Co
mparaison[,8])*(1/a) ) ,]$Difference <- "ED"
168.
169.
170.      # #tableau Ag - (DU, DE, DD)
171.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8]<= (1/a)) & (Table
au.Final.Comparaison[,9] >= (Tableau.Final.Comparaison[,8])*a ) ,]$Difference<-
"DU"
172.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8]<= (1/a)) & (Table
au.Final.Comparaison[,9] < ((Tableau.Final.Comparaison[,8])*a ) & (Tableau.Final.
Comparaison[,9] > ((Tableau.Final.Comparaison[,8])*(1/a)))) , ]$Difference<- "DE
"
173.      Tableau.Final.Comparaison[!is.na(Tableau.Final.Comparaison[,8]) & !is.na(Ta
bleau.Final.Comparaison[,9]) & (Tableau.Final.Comparaison[,8]<= (1/a)) & (Table
au.Final.Comparaison[,9] <= ((Tableau.Final.Comparaison[,8])*(1/a))) , ]$Differen
ce<- "DD"
174.
175.      #tableau RPKM=0
176.      Tableau.Final.Comparaison[is.na(Tableau.Final.Comparaison[,8]) ,]$Difference
<- "RPKM0"
177.
178.
179.      ### Save final table
180.      #write.table(Tableau.Final.Comparaison, file = "TableauRNAseq 12-08-
2016no2.diff", sep="\t", col.names=TRUE, row.names = FALSE, quote = FALSE)

```

Figure S3. Script for protein complex analysis using a signed and unsigned approach

```

1. #####
2. ###Lire tableau final avec TOUS les genes. TableauRNAseq 12-08-2016no2
3. #####
4. #Set your working directory here if not using default
5. #setwd()
6.
7. TableauRNAseqFinal <- read.table("TableauRNAseq 12-08-
2016no2.diff", sep="\t", header=TRUE)
8. TableauRNAseqFinal <- na.omit(TableauRNAseqFinal)
9. TableauRNAseqFinal <- TableauRNAseqFinal[,c("GeneSymbole", "diffHDV")]
10.
11. ###Loading the final table of all gene having a change in expression (cutoff2)

```

```

12. TableauGeneChangement <- read.table("TableauGenesChangements.diff", sep="\t", head=
    r=TRUE)
13. TableauGeneChangement <- TableauGeneChangement[,c("GeneSymbole", "diffHDV")]
14. #####

15. ###Load protein complex dataset
16. #####

17.
18. coreComplexes <- read.csv("coreComplexesCorum.txt", sep="\t", header=TRUE)
19.
20. #####

21. ###Arrangement des colonnes des tableaux
22. #####

23.
24. # install package
25. #source("http://bioconductor.org/biocLite.R")
26. #biocLite("biomaRt")
27.
28. library("BiocInstaller")
29. biocLite("biomaRt")
30.
31. # load biomaRt
32. library(biomaRt)
33. library(org.Hs.eg.db)
34. library(UniProt.ws)
35.
36. mart <- useDataset("hsapiens_gene_ensembl",
37.                    mart = useMart("ENSEMBL_MART_ENSEMBL",
38.                                    host = "useast.ensembl.org"))
39.
40. options(error=traceback)
41. #Choose only human complexes
42. coreComplexes <- coreComplexes[grep("Human", coreComplexes$Organism), ]
43. coreComplexes <- coreComplexes[,c("ComplexID", "ComplexName", "Organism", "subunits.
    UniProt.IDs.", "GO.ID", "GO.description" )]
44.
45. ### cutoff###
46. a<- 1.5
47.
48. testCoreComplex <- coreComplexes
49. testCoreComplex$Ensembl <- ""
50. testCoreComplex$Changements <- ""
51. testCoreComplex$DiffExpression <- ""
52. testCoreComplex$Pourcentage <- 0
53.
54. for(i in 1:nrow(coreComplexes)) {
55.     row <- coreComplexes[i, ]
56.     subunit <- row["subunits.UniProt.IDs."]
57.
58.     chaine <- paste(unlist(subunit), collapse = '') # Convert a list into chain
59.     asStringVector <- strsplit(chaine, " ")[[1]]
60.
61.     # temporary empty variable
62.     temp <- ""
63.     listeDeCode <- c()
64.     for (char in asStringVector) {
65.
66.         if (char == ";") {
67.             listeDeCode <- c(listeDeCode, temp)

```

```

68.     temp <- ""
69.   }
70.   else {
71.       temp <- paste(temp, char, sep = "")
72.   }
73.
74. }
75.
76. if (temp != "") {
77.     listeDeCode <- c(listeDeCode, temp)
78. }
79.
80. print(i)
81. print(listeDeCode)
82.
83. if (!is.null(listeDeCode)) {
84.     mygenes <- getBM( attributes = c("ensembl_gene_id", "entrezgene", "hgnc_symbol"
, "uniprot_gn"), filters = "uniprot_gn", values = listeDeCode, mart = mart)
85.     colnames(mygenes) <- c("GeneSymbole", "entrezgene", "hgnc_symbol", "uniprot_g
n")
86.
87.     mygenes2 <- merge(mygenes, TableauRNAseqFinal, by = "GeneSymbole", all = FALS
E)
88.     mygenes2 <- mygenes2[, c("GeneSymbole", "diffHDV")]
89.
90.
91.     if (nrow(mygenes2) > 0) {
92.         mygenes2$Changement <- NA
93.
94.
95.         ExpressionTemp <- ""
96.         ensemblTemp <- ""
97.         changementsTemp <- ""
98.         pourcentages <- 0
99.         total <- 0
100.
101.         for (j in 1:nrow(mygenes2)) {
102.             row2 <- mygenes2[j, ]
103.             total <- total + 1
104.             valeur <- 0
105.
106.             if (row2["diffHDV"] >= a | row2["diffHDV"] <= (1 / a)) {
107.                 mygenes2[j, "Changement"] <- 1
108.                 pourcentages <- pourcentages + 1
109.                 valeur <- 1
110.             }
111.             else {
112.                 mygenes2[j, "Changement"] <- 0
113.             }
114.
115.             changementsTemp <- paste(changementsTemp, valeur, ";", sep = "")
116.             ensemblTemp <-
117.                 paste(ensemblTemp, mygenes2[j, "GeneSymbole"], ";", sep = "")
118.             ExpressionTemp <-
119.                 paste(ExpressionTemp, mygenes2[j, "diffHDV"], ";", sep = "")
120.         }
121.
122.         changementsTemp <-
123.             substr(changementsTemp, 1, nchar(changementsTemp) - 1)
124.         ensemblTemp <- substr(ensemblTemp, 1, nchar(ensemblTemp) - 1)
125.         ExpressionTemp <-
126.             substr(ExpressionTemp, 1, nchar(ExpressionTemp) - 1)

```

```

127.
128.
129.     #Add to principal table
130.     testCoreComplex[i, "Ensembl"] <- ensemblTemp
131.     testCoreComplex[i, "DiffExpression"] <- ExpressionTemp
132.     testCoreComplex[i, "Changements"] <- changementsTemp
133.     testCoreComplex[i, "Pourcentage"] <- (pourcentages / total) * 100
134.   }
135. }
136.
137. }
138.
139.
140. #####
141. #####
142.     ###Add a column with the protein number
143.     CompleteCoreComplex$NbProt <- " "
144.     library(stringr)
145.     CompleteCoreComplex$NbProt <- str_count(CompleteCoreComplex$Ensembl, ";") +1
146.
147.     TableauRNAseqFinalNomGenes <- read.table("TableauRNAseqFinalNomGenes.diff",
148.     sep="\t", header=TRUE)
149.     colnames(TableauRNAseqFinalNomGenes )<-
150.     c("GeneSymbole", "GeneName", "C293", "Ag", "HDV", "C293TET", "AgTET", "HDVTET",
151.     "diffAg", "diffHDV", "SansTET", "AvecTET", "Difference")
152.     TableauRNAseqFinalNomGenes<- TableauRNAseqFinalNomGenes[,c("GeneSymbole", "di
153.     ffHDV")]
154.     # With a total of 26323 genes with RPKM >0
155.     #####
156.     #####
157.     ###Fisher's exact test table preparation
158.     #####
159.     FisherTable <- CompleteCoreComplex[, c("ComplexID", "ComplexName", "NbProt",
160.     "Pourcentage")]
161.     FisherTable$NbAffecter <- ( CompleteCoreComplex[, "Pourcentage"] * CompleteC
162.     oreComplex[, "NbProt"] )/100
163.     FisherTable <- FisherTable[, c("ComplexID", "ComplexName", "Pourcentage", "Nb
164.     Prot", "NbAffecter")]
165.     #FisherTable$p.value <-"" ##Create empty column for p-value
166.     FisherTable2 <- FisherTable[, c("ComplexID", "NbProt", "NbAffecter")]
167.
168.
169.     ### Found how many Ensembl number used in total
170.     NBTOTAL<- as.character(CompleteCoreComplex[, "Ensembl"])
171.     NBTOTAL<- unlist(strsplit(NBTOTAL, ";"))
172.     FisherTable2$NbTotal <- length(table(NBTOTAL)) ## Compte le nb total de ense
173.     mbl different utilisés.
174.     #FisherTable2$NbTotal <- 2718
175.     NBTOTAL<-
176.     as.data.frame(unique(NBTOTAL)) ##Mettre les ensembl unique dans dataframe
177.     colnames(NBTOTAL )<-c("GeneSymbole")
178.     NBTOTAL2<- merge(NBTOTAL, TableauRNAseqFinalNomGenes, by="GeneSymbole", all
179.     =FALSE) ##Put gene expression
180.     y<- 1.5
181.     AffectedTotal <- NBTOTAL2[NBTOTAL2$diffHDV >= y | NBTOTAL2$diffHDV <= (1 / y
182.     ), ]
183.     #FisherTable2$NbAffectedTotal <- 751

```

```

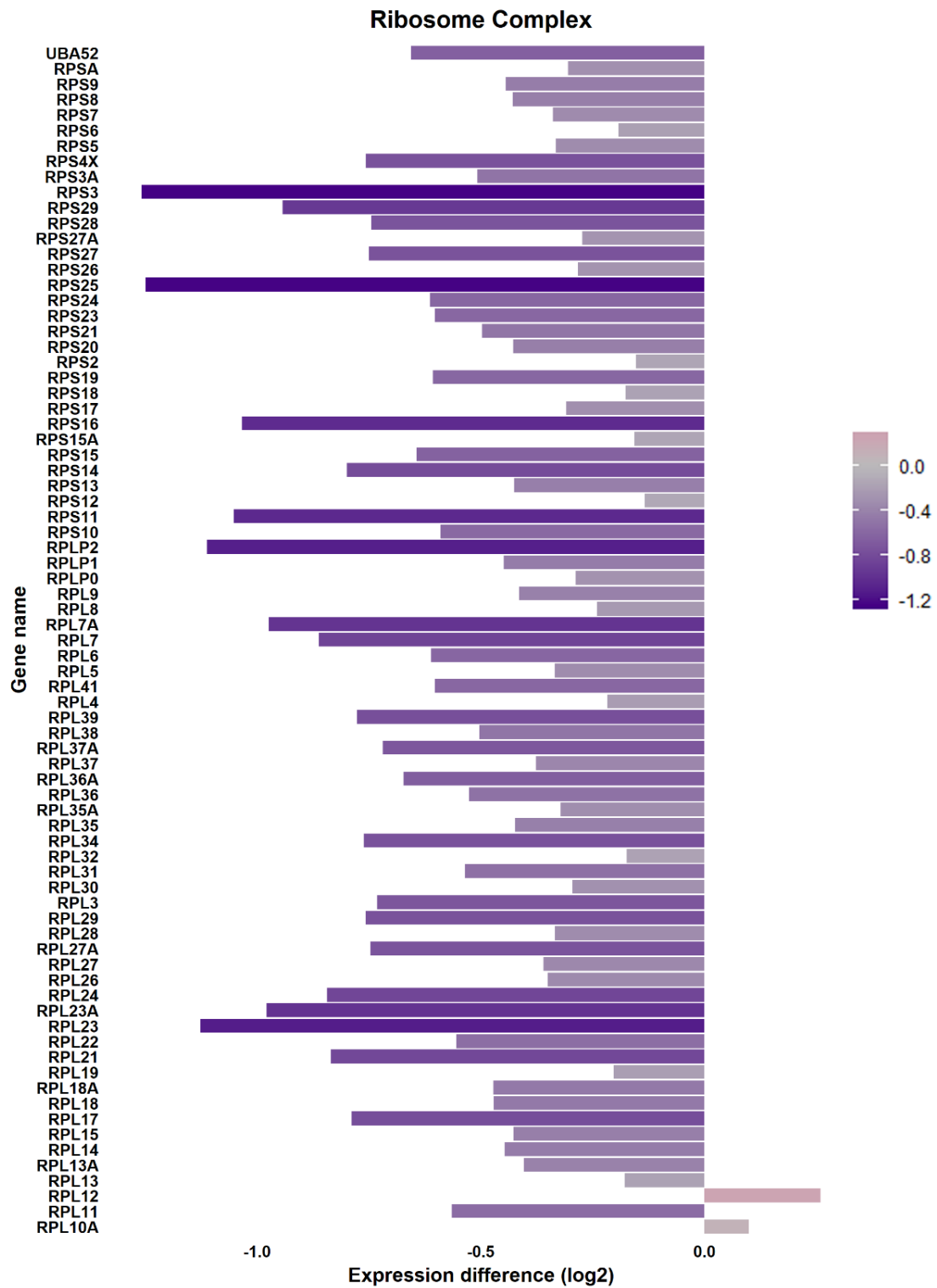
173.     FisherTable2$NbAffectedTotal <- 751 ## Compte le nb total de ensembl differe
      nt utilisés.
174.
175.     #Make Final Fisher table for p-value
176.     FisherTableFinal <-FisherTable2
177.     #colnames(FisherTableFinal )<-c()
178.     FisherTableFinal$R1 <- FisherTableFinal[, "NbAffecter"]
179.     FisherTableFinal$R2 <- FisherTableFinal[, "NbProt"] - FisherTableFinal[, "NbAf
      fecter"]
180.     FisherTableFinal$R3 <- FisherTableFinal[, "NbAffectedTotal"] - FisherTableFin
      al[, "R1"]
181.     FisherTableFinal$R4 <- ( FisherTableFinal[, "NbTotal"] - FisherTableFinal[, "N
      bAffectedTotal" ] ) -FisherTableFinal[, "R2"]
182.     FisherTableFinal$Fin <- FisherTableFinal[, "R1"] + FisherTableFinal[, "R2"] +
      FisherTableFinal[, "R3"] + FisherTableFinal[, "R4"]
183.
184.     FisherTableFinal2<-
      FisherTableFinal[,c("ComplexID", "R1", "R2", "R3", "R4")]
185.
186.
187.     #####
      #####
188.     ###Fisher's exact test
189.     #####
      #####
190.
191.     #alternative='greater'
192.     FisherTableFinal2$p_values2 <- apply(FisherTableFinal2[,c("R1", "R2", "R3",
      "R4")],1, function(x) fisher.test(matrix(x,nrow =2),alternative="greater" )$p.valu
      e)
193.     FisherTableFinal2<-FisherTableFinal2[,c("ComplexID", "p_values2")]
194.
195.     ##Merge tableau Corum complex with p-value
196.     CompleteCoreComplexP_Value<- merge(CompleteCoreComplex, FisherTableFinal2,
      by="ComplexID", all=FALSE)
197.
198.     write.table(CompleteCoreComplexP_Value,file="CompleteCoreComplexP_Value.txt"
      ,sep='\t',row.names = FALSE,quote=FALSE)

```

Figure S4. Detailed list of affected of unsigned complexes with each component RNA expression.

A) The ribosomal complex was 41% affected with a total of 78 proteins. Gene name are on the y axis and gene expression on x axis. **B)** Histones relates complexes affected by HDV including: ASF-1, core-histone-TDIF2-TDT, H2AX complex I and II, histone H3.1, histone H3.3, TLE-histone H3 and two NF-kappa B complex. Gene name are on the y axis and gene expression on x axis. **B)** Other complexes affected by HDV including CPC, NDC80, mitotic checkpoint, SNARE and more. Gene name are on the y axis and gene expression on x axis.

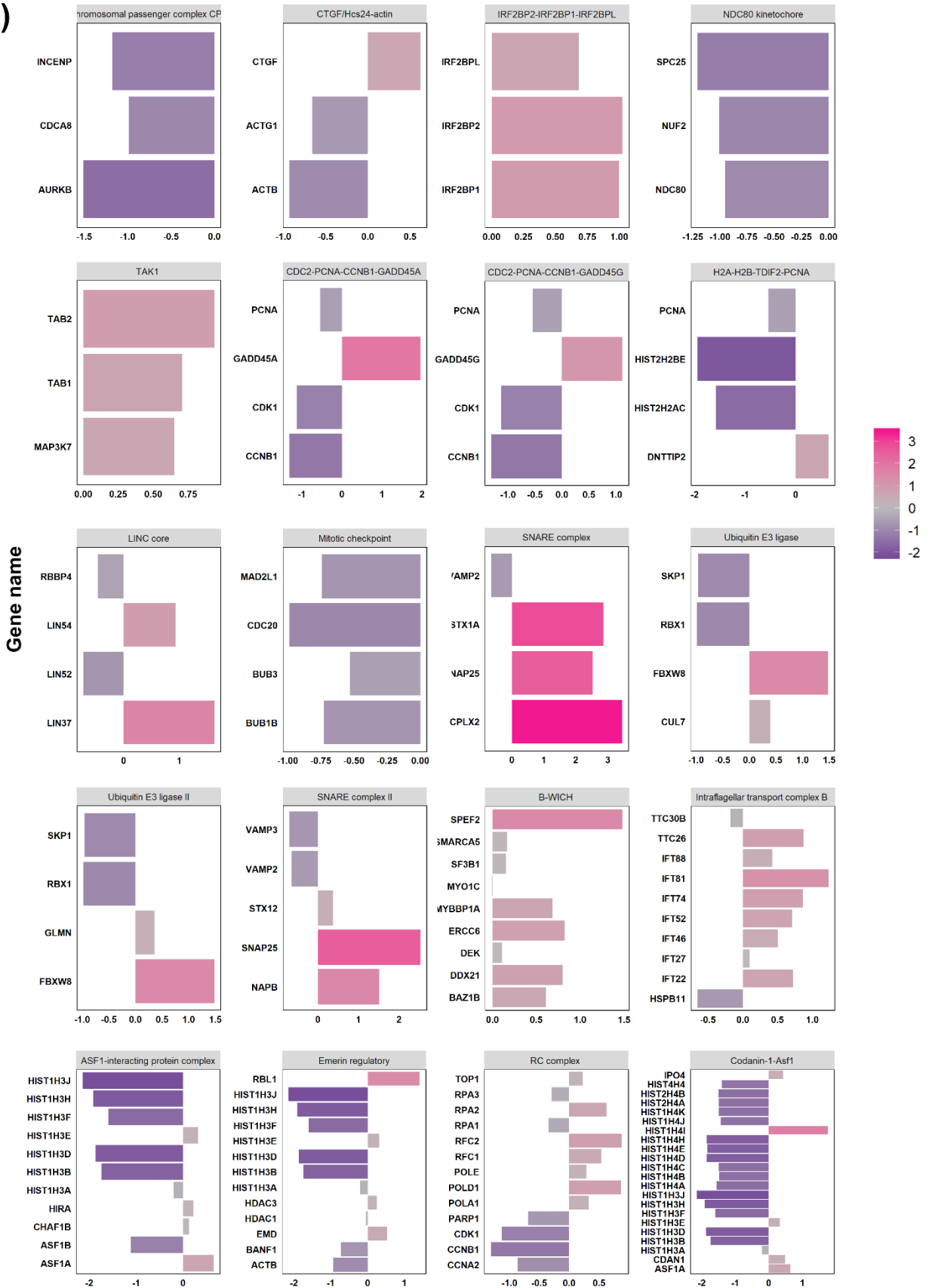
A)



B)



c)



Expression difference

Figure S5. Script for flow cytometry data processing including gating and rescaling and data analysis.

```
1. #Set your working directory here if not using default
2. #setwd()
3.
4.
5. library(flowCore)
6. library(ggplot2)
7. library(reshape)
8. library(gridExtra)
9. library(MESS)
10.
11.
12. #maxfsc <- 255000
13. maxfsc <- 255000000000
14. myinter <- 100000
15.
16. #####
17. #####
18. # IMPORT ALL FLOW DATA
19. #####
20. #####
21. mydata <- read.FCS("../Specimen_001_293 TET 24 h.fcs")
22. mydata <- exprs(mydata)
23. mydata <- as.data.frame(mydata)
24. mydata <- mydata[mydata[,3] < maxfsc,]
25. TET_293_24 <- mydata
26.
27. ## repeat for all the other files
28.
29. #####
30.
31. find_peaks <- function (x, m = 3){
32.   shape <- diff(sign(diff(x, na.pad = FALSE)))
33.   pks <- sapply(which(shape < 0), FUN = function(i){
34.     z <- i - m + 1
35.     z <- ifelse(z > 0, z, 1)
36.     w <- i + m + 1
37.     w <- ifelse(w < length(x), w, length(x))
38.     if(all(x[c(z : i, (i + 2) : w)] <= x[i + 1])) return(i + 1) else return(numbe
ric(0))
39.   })
40.   pks <- unlist(pks)
41.   pks
42. }
43.
44. maxfsc <- 255000
45. #####
46.
47.
48. ##### Remove debris
49. mydata$col <- "#00000005"
50.
51. mydens <- density(mydata[,3])
52. cutoff.debris3 <- mydens$x[find_peaks(-mydens$y)][1]
53. mydata[mydata[,3] < cutoff.debris3,]$col <- "#FF000005"
54.
55.
56. mydens <- density(mydata[,5])
57. cutoff.debris5 <- mydens$x[find_peaks(-mydens$y)][1]
```

```

58. mydata[mydata[,5] < cutoff.debris5,]$col <- "#FF000005"
59.
60. mydata[mydata[,3] > maxfsc,]$col <- "#FF000005"
61.
62. #myx <- 3
63. #myy <- 5
64. #plot(mydata[,myx], mydata[,myy], col=mydata$col, pch=19, xlab=colnames(mydata)[myx], ylab=colnames(mydata)[myy])
65.
66.
67. mydata <- mydata[mydata$col == "#00000005",]
68.
69. #myx <- 3
70. #myy <- 5
71. #plot(mydata[,myx], mydata[,myy], col=mydata$col, pch=19, xlab=colnames(mydata)[myx], ylab=colnames(mydata)[myy])
72.
73.
74. ##### Remove cell clusters
75.
76. mydata$col <- "#00000005"
77.
78. cutoff.clusters <- 2
79. mydata[mydata[,2]/mydata[,3] >= cutoff.clusters,]$col <- "#FF000005"
80. mydata[mydata[,2] > maxfsc,]$col <- "#FF000005"
81.
82. #myx <- 2
83. #myy <- 3
84. #plot(mydata[,myx], mydata[,myy], col=mydata$col, pch=19, xlab=colnames(mydata)[myx], ylab=colnames(mydata)[myy])
85.
86. mydata <- mydata[mydata$col == "#00000005",]
87.
88. TET_293_24 <- mydata
89.
90.
91.
92.
93.
94.
95.
96. #####
97. #####
98. # Align and Rescale PI Density G1/G2 Populations
99. #####
100. #####
101.
102. dens_NoTET_293_24 <- density(NoTET_293_24[,8])
103. dens_TET_293_24 <- density(TET_293_24[,8])
104. dens_NoTET_HDV_24 <- density(NoTET_HDV_24[,8])
105. dens_TET_HDV_24 <- density(TET_HDV_24[,8])
106. dens_NoTET_Ag_24 <- density(NoTET_Ag_24[,8])
107. dens_TET_Ag_24 <- density(TET_Ag_24[,8])
108. dens_NoTET_HDV_12 <- density(NoTET_HDV_12[,8])
109. dens_TET_HDV_12 <- density(TET_HDV_12[,8])
110. dens_NoTET_HDV_36 <- density(NoTET_HDV_36[,8])
111. dens_TET_HDV_36 <- density(TET_HDV_36[,8])
112.
113. G1_dens_NoTET_293_24 <- dens_NoTET_293_24$x[which.max(dens_NoTET_293_24$y)]
114. G1_dens_TET_293_24 <- dens_TET_293_24$x[which.max(dens_TET_293_24$y)]

```

```

115.      G1_dens_NoTET_HDV_24 <- dens_NoTET_HDV_24$x[which.max(dens_NoTET_HDV_24$y)]
116.      G1_dens_TET_HDV_24 <- dens_TET_HDV_24$x[which.max(dens_TET_HDV_24$y)]
117.      G1_dens_NoTET_Ag_24 <- dens_NoTET_Ag_24$x[which.max(dens_NoTET_Ag_24$y)]
118.      G1_dens_TET_Ag_24 <- dens_TET_Ag_24$x[which.max(dens_TET_Ag_24$y)]
119.      G1_dens_NoTET_HDV_12 <- dens_NoTET_HDV_12$x[which.max(dens_NoTET_HDV_12$y)]

120.      G1_dens_TET_HDV_12 <- dens_TET_HDV_12$x[which.max(dens_TET_HDV_12$y)]
121.      G1_dens_NoTET_HDV_36 <- dens_NoTET_HDV_36$x[which.max(dens_NoTET_HDV_36$y)]

122.      G1_dens_TET_HDV_36 <- dens_TET_HDV_36$x[which.max(dens_TET_HDV_36$y)]
123.
124.      NoTET_293_24[,8] <- NoTET_293_24[,8] * (G1_dens_NoTET_293_24/G1_dens_NoTET_2
93_24)
125.      TET_293_24[,8] <- TET_293_24[,8] * (G1_dens_NoTET_293_24/G1_dens_TET_293_24)

126.      NoTET_HDV_24[,8] <- NoTET_HDV_24[,8] * (G1_dens_NoTET_293_24/G1_dens_NoTET_H
DV_24)
127.      TET_HDV_24[,8] <- TET_HDV_24[,8] * (G1_dens_NoTET_293_24/G1_dens_TET_HDV_24)

128.      NoTET_Ag_24[,8] <- NoTET_Ag_24[,8] * (G1_dens_NoTET_293_24/G1_dens_NoTET_Ag_
24)
129.      TET_Ag_24[,8] <- TET_Ag_24[,8] * (G1_dens_NoTET_293_24/G1_dens_TET_Ag_24)
130.      NoTET_HDV_12[,8] <- NoTET_HDV_12[,8] * (G1_dens_NoTET_293_24/G1_dens_NoTET_H
DV_12)
131.      TET_HDV_12[,8] <- TET_HDV_12[,8] * (G1_dens_NoTET_293_24/G1_dens_TET_HDV_12)

132.      NoTET_HDV_36[,8] <- NoTET_HDV_36[,8] * (G1_dens_NoTET_293_24/G1_dens_NoTET_H
DV_36)
133.      TET_HDV_36[,8] <- TET_HDV_36[,8] * (G1_dens_NoTET_293_24/G1_dens_TET_HDV_36)

134.
135.      min.cells.number <- min(c(nrow(NoTET_293_24), nrow(TET_293_24), nrow(NoTET_H
DV_24), nrow(TET_HDV_24), nrow(NoTET_Ag_24), nrow(TET_Ag_24), nrow(NoTET_HDV_12), n
row(TET_HDV_12), nrow(NoTET_HDV_36), nrow(TET_HDV_36)))
136.
137.      dens_NoTET_293_24 <- density(sample(NoTET_293_24[,8], min.cells.number, repl
ace=FALSE))
138.      dens_TET_293_24 <- density(sample(TET_293_24[,8], min.cells.number, replace
=FALSE))
139.      dens_NoTET_HDV_24 <- density(sample(NoTET_HDV_24[,8], min.cells.number, repl
ace=FALSE))
140.      dens_TET_HDV_24 <- density(sample(TET_HDV_24[,8], min.cells.number, replace=
FALSE))
141.      dens_NoTET_Ag_24 <- density(sample(NoTET_Ag_24[,8], min.cells.number, replac
e=FALSE))
142.      dens_TET_Ag_24 <- density(sample(TET_Ag_24[,8], min.cells.number, replace=FA
LSE))
143.      dens_NoTET_HDV_12 <- density(sample(NoTET_HDV_12[,8], min.cells.number, repl
ace=FALSE))
144.      dens_TET_HDV_12 <- density(sample(TET_HDV_12[,8], min.cells.number, replace=
FALSE))
145.      dens_NoTET_HDV_36 <- density(sample(NoTET_HDV_36[,8], min.cells.number, repl
ace=FALSE))
146.      dens_TET_HDV_36 <- density(sample(TET_HDV_36[,8], min.cells.number, replace=
FALSE))
147.
148.
149.
150.
151.      #####
152.      #####

```

```

153.     # Density G1/G2 Populations
154.     #####
155.     #####
156.
157.     png("Densities_24h.png", width=2000, height=2000, res=300)
158.     maxdens <- max(c(dens_NoTET_293_24$y, dens_TET_293_24$y, dens_NoTET_HDV_
159.     24$y, dens_TET_HDV_24$y, dens_NoTET_Ag_24$y, dens_TET_Ag_24$y))
160.     plot(dens_NoTET_293_24, col="black", ylim=c(0, maxdens), xlim=c(20000,15
161.     0000), lwd=2, main="", xlab="")
162.     lines(dens_TET_293_24, col="blue", lwd=2)
163.     lines(dens_NoTET_Ag_24, col="purple", lwd=2)
164.     lines(dens_TET_Ag_24, col="green", lwd=2)
165.     lines(dens_NoTET_HDV_24, col="orange", lwd=2)
166.     lines(dens_TET_HDV_24, col="red", lwd=2)
167.     legend(100000,0.0001, c("NoTET_293_24", "TET_293_24", "NoTET_Ag_24", "TE
168.     T_Ag_24", "NoTET_HDV_24", "TET_HDV_24"), text.col=c("black","blue","purple","green"
169.     ,"orange","red"), bty = "n")
170.     dev.off()
171.
172.     png("Densities_TimeCourse.png", width=2000, height=2000, res=300)
173.     maxdens <- max(c(dens_NoTET_HDV_12$y, dens_TET_HDV_12$y, dens_NoTET_HDV_
174.     24$y, dens_TET_HDV_24$y, dens_NoTET_HDV_36$y, dens_TET_HDV_36$y))
175.     plot(dens_NoTET_HDV_12, col="black", ylim=c(0, maxdens), xlim=c(20000,15
176.     0000), lwd=2, main="", xlab="")
177.     lines(dens_TET_HDV_12, col="blue", lwd=2)
178.     lines(dens_NoTET_HDV_24, col="purple", lwd=2)
179.     lines(dens_TET_HDV_24, col="green", lwd=2)
180.     lines(dens_NoTET_HDV_36, col="orange", lwd=2)
181.     lines(dens_TET_HDV_36, col="red", lwd=2)
182.     legend(100000,0.0001, c("NoTET_HDV_12", "TET_HDV_12", "NoTET_HDV_24", "T
183.     ET_HDV_24", "NoTET_HDV_36", "TET_HDV_36"), text.col=c("black","blue","purple","gree
184.     n","orange","red"), bty = "n")
185.     dev.off()
186.
187.
188.     #####
189.     #####
190.     # Cell cycle Percentages
191.     #####
192.     #####
193.
194.     mycutoffs <- c(42000,65000,95000,115000)
195.
196.     myres <- matrix(nrow=10, ncol=3)
197.     colnames(myres) <- c("G1", "S", "G2")
198.     rownames(myres) <- c("NoTET_293_24", "TET_293_24", "NoTET_Ag_24", "TET_Ag_24
199.     ", "NoTET_HDV_24", "TET_HDV_24", "NoTET_HDV_12", "TET_HDV_12", "NoTET_HDV_36", "TET
200.     _HDV_36")
201.
202.     for(i in 1:nrow(myres)){
203.         mytmp <- get(rownames(myres)[i])
204.         myres[i,1] <- nrow(mytmp[mytmp[,8] > mycutoffs[1] & mytmp[,8] <= mycutoff
205.         fs[2],])
206.         myres[i,2] <- nrow(mytmp[mytmp[,8] > mycutoffs[2] & mytmp[,8] <= mycutoff
207.         fs[3],])
208.         myres[i,3] <- nrow(mytmp[mytmp[,8] > mycutoffs[3] & mytmp[,8] <= mycutoff
209.         fs[4],])
210.     }

```

```

202.     rowsum <- rowSums(myres)
203.     myres[,1] <- myres[,1]/rowsum
204.     myres[,2] <- myres[,2]/rowsum
205.     myres[,3] <- myres[,3]/rowsum
206.
207.
208.     forgraph24 <- myres[grepl("24", rownames(myres)),]
209.     forgraph24 <- melt(forgraph24)
210.     forgraph24$value <- forgraph24$value*100
211.     colnames(forgraph24)[1:2] <- c("cell", "cycle")
212.     forgraph24$cell <- as.character(unlist(strsplit(as.character(forgraph24$cell), "_24"))))
213.     forgraph24$cell <- factor(forgraph24$cell, levels=c("NoTET_293", "TET_293", "NoTET_Ag", "TET_Ag", "NoTET_HDV", "TET_HDV"))
214.     forgraph24$cycle <- factor(forgraph24$cycle, levels=c("G2", "S", "G1"))
215.
216.     graph24 <- ggplot(forgraph24, aes(x=cell, y=value, fill=cycle)) + geom_bar(stat='identity') +
217.       theme_bw() +
218.       scale_y_continuous(expand = c(0, 0), limits=c(0,100.1), breaks=(seq(0,100,by=20))) +
219.       xlab(label="") +
220.       ylab(label="Fraction of cell (%)") +
221.       theme(axis.title.x = element_text(colour = "black"), axis.title.y = element_text(colour = "black")) +
222.       theme(axis.text = element_text(colour = "black", face="bold")) +
223.       theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
224.       labs(fill='')
225.
226.     graph24 <- graph24 + annotate(geom="text", x=forgraph24$cell, y=c(rep(40, 6), rep(82,6), rep(95,6)), label=paste(as.character(format(round(forgraph24$value, 1))), "%", sep=""), fontface = 2)
227.     graph24 <- graph24 + theme(axis.text.x = element_text(angle = 45, hjust = 1))
228.
229.     png("Cell_Cycle_Percents_24h.png", width=2000, height=2000, res=300)
230.
231.     graph24
232.     dev.off()
233.
234.     forgraphHDV <- myres[grepl("HDV", rownames(myres)),]
235.     forgraphHDV <- melt(forgraphHDV)
236.     forgraphHDV$value <- forgraphHDV$value*100
237.     colnames(forgraphHDV)[1:2] <- c("cell", "cycle")
238.     forgraphHDV$cell <- factor(forgraphHDV$cell, levels=c("NoTET_HDV_12", "TET_HDV_12", "NoTET_HDV_24", "TET_HDV_24", "NoTET_HDV_36", "TET_HDV_36"))
239.     forgraphHDV$cycle <- factor(forgraphHDV$cycle, levels=c("G2", "S", "G1"))
240.
241.     graphHDV <- ggplot(forgraphHDV, aes(x=cell, y=value, fill=cycle)) + geom_bar(stat='identity') +
242.       theme_bw() +
243.       scale_y_continuous(expand = c(0, 0), limits=c(0,100.1), breaks=(seq(0,100,by=20))) +
244.       xlab(label="") +
245.       ylab(label="Fraction of cell (%)") +
246.       theme(axis.title.x = element_text(colour = "black"), axis.title.y = element_text(colour = "black")) +
247.       theme(axis.text = element_text(colour = "black", face="bold")) +

```

```

248.         theme(panel.grid.major = element_blank(), panel.grid.minor = ele
249.             ment_blank()) +
250.             labs(fill='')
251.         graphHDV <- graphHDV + annotate(geom="text", x=forgraphHDV$cell, y=c(rep(40,
252.             6),rep(83,6),rep(95,6)), label=paste(as.character(format(round(forgraphHDV$value,
253.             1))), "%", sep=""), fontface =2)
254.         graphHDV <- graphHDV + theme(axis.text.x = element_text(angle = 45, hjust =
255.             1))
256.         png("Cell_Cycle_Percents_TimeCourse.png", width=2000, height=2000, res=300)
257.         graphHDV
258.         dev.off()
259.         forgraphHDV$cell <- factor(forgraphHDV$cell, levels=c("NoTET_HDV_12", "NoTET
260.             _HDV_24", "NoTET_HDV_36", "TET_HDV_12", "TET_HDV_24", "TET_HDV_36"))
261.         graphHDV <- ggplot(forgraphHDV, aes(x=cell, y=value, fill=cycle)) + geom_bar
262.             (stat='identity') +
263.             theme_bw() +
264.             scale_y_continuous(expand = c(0, 0), limits=c(0,100.1), breaks=(
265.                 seq(0,100,by=20))) +
266.             xlab(label="") +
267.             ylab(label="Fraction of cell (%)") +
268.             theme(axis.title.x = element_text(colour = "black"), axis.title.
269.                 y = element_text(colour = "black")) +
270.             theme(axis.text = element_text(colour = "black", face="bold")) +
271.             theme(panel.grid.major = element_blank(), panel.grid.minor = ele
272.                 ment_blank()) +
273.                 labs(fill='')
274.         graphHDV <- graphHDV + annotate(geom="text", x=forgraphHDV$cell, y=c(rep(40,
275.             6),rep(83,6),rep(95,6)), label=paste(as.character(format(round(forgraphHDV$value,
276.             1))), "%", sep=""), fontface =2)
277.         graphHDV <- graphHDV + theme(axis.text.x = element_text(angle = 45, hjust =
278.             1))
279.         png("Cell_Cycle_Percents_TimeCourse_reordered.png", width=2000, height=2000,
280.             res=300)
281.         graphHDV
282.         dev.off()
283.
284.         #####
285.         #####
286.         # Size Scatter Plot
287.         #####
288.         #####
289.
290.         mydata <- NoTET_HDV_24
291.         mydata2 <- TET_HDV_24
292.         mydata$col <- "#0000FF10"
293.         mydata2$col <- "#FF000010"
294.
295.         myx <- 3
296.         myy <- 5
297.
298.         graph <- ggplot(mydata, aes(x=mydata[,myx], y=mydata[,myy])) +
299.             geom_point(cex=0.2, col=mydata$col) +

```

```

295.         geom_point(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]), cex=
0.2, col=mydata2$col) +
296.         geom_density_2d(col=substr(unique(mydata$col ),1,7)) +
297.         geom_density_2d(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]),
col=substr(unique(mydata2$col ),1,7)) +
298.         theme_bw() +
299.         xlab(as.character(colnames(mydata)[myx])) +
300.         ylab(as.character(colnames(mydata)[myy])) +
301.         ggtitle("NoTET_HDV_24 (blue) vs TET_HDV_24 (red)") +
302.         theme(plot.title = element_text(hjust = 0.5)) +
303.         scale_y_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myy], my
data2[,myy])), max(c(mydata[,myy], mydata2[,myy])))) +
304.         scale_x_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myx], my
data2[,myx])), max(c(mydata[,myx], mydata2[,myx])))) +
305.         theme(axis.title.x = element_text(colour = "black"), axis.title.y =
element_text(colour = "black")) +
306.         theme(axis.text = element_text(colour = "black", face="bold")) +
307.         theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank())
308.
309.     png("NoTET_HDV_24vsTET_HDV_24.png", width=3000, height=3000, res=300)

310.     graph
311.     dev.off()
312.
313.
314.     #####
315.     #####
316.     # Size Scatter Plot vs Cell Cycle
317.     #####
318.     #####
319.
320.     mycutoffs <- c(42000,65000,95000,115000)
321.
322.     myx <- 3
323.     myy <- 5
324.
325.     NoTET_HDV_24$col <- "white"
326.     NoTET_HDV_24[NoTET_HDV_24[,8] > mycutoffs[1] & NoTET_HDV_24[,8] <= mycutoffs
[2,],$col <- "#0000FF10" #G1
327.
328.     TET_HDV_24$col <- "white"
329.     TET_HDV_24[TET_HDV_24[,8] > mycutoffs[1] & TET_HDV_24[,8] <= mycutoffs[2,],$
col <- "#FF000010" #G1
330.
331.     mydata <- NoTET_HDV_24[NoTET_HDV_24$col != "white",]
332.     mydata2 <- TET_HDV_24[TET_HDV_24$col != "white",]
333.
334.     graph <- ggplot(mydata, aes(x=mydata[,myx], y=mydata[,myy])) +
335.         geom_point(cex=0.2, col=mydata$col) +
336.         geom_point(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]), cex=
0.2, col=mydata2$col) +
337.         geom_density_2d(col="blue") +
338.         geom_density_2d(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]),
col="red") +
339.         theme_bw() +
340.         xlab(as.character(colnames(mydata)[myx])) +
341.         ylab(as.character(colnames(mydata)[myy])) +
342.         ggtitle("NoTET_HDV_24 (blue) vs TET_HDV_24 (red) G1") +
343.         theme(plot.title = element_text(hjust = 0.5)) +
344.         scale_y_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myy], my
data2[,myy])), max(c(mydata[,myy], mydata2[,myy])))) +

```

```

345.         scale_x_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myx], my
data2[,myx])), max(c(mydata[,myx], mydata2[,myx])))) +
346.         theme(axis.title.x = element_text(colour = "black"), axis.title.y =
element_text(colour = "black")) +
347.         theme(axis.text = element_text(colour = "black", face="bold")) +
348.         theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank())
349.
350.         png("NoTET_HDV_24vsTET_HDV_24_G1.png", width=3000, height=3000, res=300)
351.
352.         graph
353.         dev.off()
354.         NoTET_HDV_24$col <- "white"
355.         NoTET_HDV_24[NoTET_HDV_24[,8] > mycutoffs[2] & NoTET_HDV_24[,8] <= mycutoffs
[3],]$col <- "#0000FF20"      #S
356.
357.         TET_HDV_24$col <- "white"
358.         TET_HDV_24[TET_HDV_24[,8] > mycutoffs[2] & TET_HDV_24[,8] <= mycutoffs[3],]$
col <- "#FF000020"      #S
359.
360.         mydata <- NoTET_HDV_24[NoTET_HDV_24$col != "white",]
361.         mydata2 <- TET_HDV_24[TET_HDV_24$col != "white",]
362.
363.         graph <- ggplot(mydata, aes(x=mydata[,myx], y=mydata[,myy])) +
364.         geom_point(cex=0.2, col=mydata$col) +
365.         geom_point(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]), cex=
0.2, col=mydata2$col) +
366.         geom_density_2d(col="blue") +
367.         geom_density_2d(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]),
col="red") +
368.         theme_bw() +
369.         xlab(as.character(colnames(mydata)[myx])) +
370.         ylab(as.character(colnames(mydata)[myy])) +
371.         ggtitle("NoTET_HDV_24 (blue) vs TET_HDV_24 (red) S") +
372.         theme(plot.title = element_text(hjust = 0.5)) +
373.         scale_y_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myy], my
data2[,myy])), max(c(mydata[,myy], mydata2[,myy])))) +
374.         scale_x_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myx], my
data2[,myx])), max(c(mydata[,myx], mydata2[,myx])))) +
375.         theme(axis.title.x = element_text(colour = "black"), axis.title.y =
element_text(colour = "black")) +
376.         theme(axis.text = element_text(colour = "black", face="bold")) +
377.         theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank())
378.
379.         png("NoTET_HDV_24vsTET_HDV_24_S.png", width=3000, height=3000, res=300)
380.
381.         graph
382.         dev.off()
383.         NoTET_HDV_24$col <- "white"
384.         NoTET_HDV_24[NoTET_HDV_24[,8] > mycutoffs[3] & NoTET_HDV_24[,8] <= mycutoffs
[4],]$col <- "#0000FF20"      #G2
385.
386.         TET_HDV_24$col <- "white"
387.         TET_HDV_24[TET_HDV_24[,8] > mycutoffs[3] & TET_HDV_24[,8] <= mycutoffs[4],]$
col <- "#FF000020"      #G2
388.
389.         mydata <- NoTET_HDV_24[NoTET_HDV_24$col != "white",]
390.         mydata2 <- TET_HDV_24[TET_HDV_24$col != "white",]
391.

```

```

392.     graph <- ggplot(mydata, aes(x=mydata[,myx], y=mydata[,myy])) +
393.       geom_point(cex=0.2, col=mydata$col) +
394.       geom_point(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]), cex=
0.2, col=mydata2$col) +
395.       geom_density_2d(col="blue") +
396.       geom_density_2d(data=mydata2, aes(x=mydata2[,myx], y=mydata2[,myy]),
col="red") +
397.       theme_bw() +
398.       xlab(as.character(colnames(mydata)[myx])) +
399.       ylab(as.character(colnames(mydata)[myy])) +
400.       ggtitle("NoTET_HDV_24 (blue) vs TET_HDV_24 (red) G2") +
401.       theme(plot.title = element_text(hjust = 0.5)) +
402.       scale_y_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myy], my
data2[,myy])), max(c(mydata[,myy], mydata2[,myy])))) +
403.       scale_x_continuous(expand = c(0, 0), limits=c(min(c(mydata[,myx], my
data2[,myx])), max(c(mydata[,myx], mydata2[,myx])))) +
404.       theme(axis.title.x = element_text(colour = "black"), axis.title.y =
element_text(colour = "black")) +
405.       theme(axis.text = element_text(colour = "black", face="bold")) +
406.       theme(panel.grid.major = element_blank(), panel.grid.minor = element
_blank())
407.
408.
409.     png("NoTET_HDV_24vsTET_HDV_24_G2.png", width=3000, height=3000, res=300)

410.     graph
411.     dev.off()
412.
413.
414.
415.     #####
416.     #####
417.     # Make graph and put cutoffs
418.     #####
419.     #####
420.
421.
422.     myx <- 3
423.     myy <- 5
424.     mybins <- 300
425.
426.     graph_NoTET_293_24 <- ggplot(NoTET_293_24) +
427.       geom_hex(aes(NoTET_293_24[,myx], NoTET_293_24[,myy]), bins = mybins) +
428.       scale_fill_gradientn("", colours = rev(rainbow(10, end = 4/6))) +
429.       geom_vline(xintercept = myinter) +
430.       geom_hline(yintercept = myinter) +
431.       xlab(colnames(NoTET_293_24)[myx]) +
432.       ylab(colnames(NoTET_293_24)[myy]) +
433.       ggtitle("NoTET_293_24") +
434.       theme_bw() +
435.       theme(legend.position="none")
436.
437.
438.     mysamples <- c("NoTET_293_24", "TET_293_24", "NoTET_Ag_24", "TET_Ag_24", "No
TET_HDV_24", "TET_HDV_24", "NoTET_HDV_12", "TET_HDV_12", "NoTET_HDV_36", "TET_HDV_3
6")
439.
440.     png("HvsH_plot.png", width=2000, height=4000, res=300)
441.     myx <- 3
442.     myy <- 5
443.     par(mfrow=c(5,2))
444.     for(i in 1:length(mysamples)){

```

```

445.         mydata <- get(mysamples[i])
446.         mydata$col <- densCols(mydata[,myx], mydata[,myy], colramp = colorRampPa
         lette(rev(rainbow(10, end = 4/6))))
447.         plot(mydata[,myx], mydata[,myy], col=mydata$col, pch=19, cex=0.05, xlab=
         colnames(mydata)[myx], ylab=colnames(mydata)[myy], main=mysamples[i])
448.         abline(v=100000, h=100000)
449.     }
450.     dev.off()
451.
452.     png("AvsH_plot.png", width=2000, height=4000, res=300)
453.     myx <- 2
454.     myy <- 3
455.     par(mfrow=c(5,2))
456.     for(i in 1:length(mysamples)){
457.         mydata <- get(mysamples[i])
458.         mydata$col <- densCols(mydata[,myx], mydata[,myy], colramp = colorRampPa
         lette(rev(rainbow(10, end = 4/6))))
459.         plot(mydata[,myx], mydata[,myy], col=mydata$col, pch=19, cex=0.05, xlab=
         colnames(mydata)[myx], ylab=colnames(mydata)[myy], main=mysamples[i])
460.     }
461.     dev.off()
462.
463.
464.     #####
465.     #####
466.     ##### % aggregation
467.     #####
468.     #####
469.
470.     myx <- 2
471.     myy <- 3
472.     mycutoff <- 1.7
473.     maxfsc <- 260000
474.     b <- 63000
475.     m <- (maxfsc/mycutoff-63000)/(maxfsc-125000)
476.     minfsc <- 125000
477.
478.     mysamples <- c("NoTET_293_24", "TET_293_24", "NoTET_Ag_24", "TET_Ag_24", "No
         TET_HDV_24", "TET_HDV_24", "NoTET_HDV_12", "TET_HDV_12", "NoTET_HDV_36", "TET_HDV_3
         6")
479.
480.
481.     par(mfrow=c(5,2))
482.     for(i in mysamples){
483.
484.         mydata <- get(i)
485.         mydata$col <- "#00000005"
486.         mydata[mydata[,myy] <= m*(mydata[,myx]-
         minfsc) + b & mydata[,myx] > minfsc ,]$col <- "#FF000005"
487.         mydata[mydata[,myx] > maxfsc,]$col <- "#0000FF05"
488.         myblack <- paste(nrow(mydata[mydata$col == "#00000005",]), ":", round(1
         00*nrow(mydata[mydata$col == "#00000005",])/nrow(mydata[mydata$col != "#0000FF05",]
         ), 2), "%", sep="")
489.         myred <- paste(nrow(mydata[mydata$col == "#FF000005",]), ":", round(100
         *nrow(mydata[mydata$col == "#FF000005",])/nrow(mydata[mydata$col != "#0000FF05",]),
         2), "%", sep="")
490.
491.         plot(mydata[,myx], mydata[,myy], pch=19, col=mydata$col)
492.         text(x=50000, y=250000, label=myblack, col="black")
493.         text(x=50000, y=240000, label=myred, col="red")
494.         segments(x0=125000, y0=0, x1=125000, y1=63000)
495.         segments(x0=125000, y0= 63000, x1= maxfsc, y1= maxfsc/mycutoff)

```

496.
497.
498. }