

**ALV INDUCED TUMORIGENESIS: Identifying Clonal Progression of Tumors
and Novel Cancer Associated Genes.**

by

Sanandan Malhotra

A dissertation submitted to Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

December 8th, 2017

Abstract

Avian leukosis virus (ALV) is a simple retrovirus that infects and causes cancer in chickens. The virus integrates into the chicken genome and it can alter the expression of surrounding genes via its strong promoter and enhancer elements. This perturbation of gene expression near ALV integration sites induces a wide range of tumors. Chickens infected with ALV subgroup A mostly develop B-cell lymphomas, while those infected with ALV subgroup J develop myeloid tumors as well as hemangiomas. Since ALV acts as a mutagen by integrating into the chicken genome, it serves as a good insertional mutagenesis tool.

We analyzed the preferences of ALV integrations by infecting different cells (CEF, DT40 and HeLa) in tissue culture. We also analyzed integrations in early and late ALV-induced neoplasms. We infected 5- and 10-day old chicken embryos with ALV subgroup A and J to induce tumorigenesis, and then identified viral integration sites using an Illumina high-throughput sequencing platform. This allowed us to determine how the ALV integration pattern is altered due to oncogenic selection during tumorigenesis. We were also able to empirically define the clonal progression of tumors and identify the underlying biological pathways and gene players implicated in oncogenesis.

We identified clonally expanded integrations in the MET gene in ALV subgroup A and J induced hemangiomas. We also showed that ALV induces strong overexpression of MET in the tumors that harbor these integrations. MET is a known proto-oncogene, involved in angiogenesis but has not been

previously implicated in the development of hemangiomas.

We also identified the TERT promoter region as a common ALV integration hotspot in many B-cell lymphomas. This led us to observe the activation of a novel lncRNA in the TERT promoter region, named TERT antisense promoter-associated (TAPAS) RNA. We showed that ALV drives the overexpression of TAPAS and contributes to the development of lymphomas.

Lastly, we reported a novel transcript in the human TERT promoter region, named human TAPAS (hTAPAS) lncRNA. hTAPAS expression inversely correlates with TERT expression in tumors, and its expression is not affected by the TERT promoter mutation. Knock down of hTAPAS results in increased TERT expression, and its ectopic overexpression down regulates TERT expression. With a role in regulating TERT expression, hTAPAS is implicated in regulating TERT homeostasis.

Acknowledgements

I would like to thank Dr. Karen Beemon for her guidance throughout my PhD. I appreciate her letting me use my own ideas and instilling creativity in my work. She has been supportive of my work and has been very generous with her time to discuss ideas to take the work forward. I respect Karen for her intellectual curiosity, which has greatly benefited my research. I am very glad I joined her lab as a graduate student and I feel very fortunate to have worked with her.

I would also like to thank my thesis committee members, Dr. Joe Gall, Dr. James Taylor, and Dr. Andy Hoyt for all their time and support. They have been very supportive of my work and have made significant contributions to my scientific ideas.

I would like to thank our collaborators, late Dr. Paul Neiman (Fred Hutchinson Cancer Center) and Dr. Robin Morgan (University of Delaware) for their work with chickens and the generation of tumors. Their work served as a solid foundation for a lot of my research and data analysis. I would also like to thank Dr. Mallory Freeberg and Dr. James Taylor for their collaboration to analyze genomics data from cancer patients.

I am also very grateful to all the people I have worked with in the Beemon Lab. I would especially like to thank Dr. James Justice, who was an exceptional mentor and a great friend. His patience and composed attitude helped me a lot to gain success in my work. A lot of my findings were built upon his work, which served as a great platform for me, and I am very grateful to him for that. I would also like to thank Shelby Winans, a fellow graduate student who joined the lab at

the same time as me. Shelby has been an excellent scientific companion and has always pushed me to do my best. I am also grateful to Yingying Li, our lab manager, for being a great source of assistance with all my lab work.

I would like to thank Abigail Fuchsman and Marla Tharp for their friendships, which have helped me during my struggling phases of graduate school. I deeply value these friendships and will cherish them for life.

I would like to thank my parents, Sanjiv and Sampada Malhotra for their unconditional love and support, and always having my back. Last, but not the least, I would like to thank my best friend and fiancé, Dhriti Dhawan, for being a constant pillar of support and believing in me when I did not believe in myself. Dhriti has been a constant source of support for me in every sense: socially, morally, emotionally and intellectually, and I feel very fortunate to have her as a part of my life. I could not have done any of the things in graduate school or in life if it weren't for the great support network of my friends, family and fiancé.

To my late grandmother, Anjana Malhotra

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	vii
List of Tables	xi
List of Figures	xii
List of Abbreviations	xvi
Chapter 1. Introduction to Avian Leukosis Virus	1
Introduction to avian retroviruses	2
The life cycle of ALV.....	3
ALV as an insertional mutagenesis tool	8
ALV integrations help define the clonal progression of tumors.....	10
ALV integrations allow identification of novel cancer associated genes	13
Chapter 2. Selection for Avian Leukosis Virus integration sites determines the clonal progression of B-cell lymphomas	18
Summary.....	19
Introduction	21
Results	24
ALV integration is enriched within genes in cultured cells and in tumors	24
ALV integrations are enriched at the 5' end of gene bodies in cultured cells and in tumors.....	29
ALV integration near transcription start sites and CpG islands varies in cultured cells versus tumors	31
ALV integrations are enriched in expressed transcriptional units in cultured cells and in tumors	35
ALV integration is targeted to highly spliced transcriptional units	38
Distribution of ALV integrations in B-cell lymphomas	42
ALV integrations are most clonal in metastases	46
ALV clonality increases with progressing stages of lymphomagenesis	49

Proviral load (PVL) increases with the progressing stages of tumorigenesis ..	53
Ontologies of genes near ALV integrations highlight biological processes associated with tumor progression	56
Discussion	63
Materials and Methods	69
Tissue culture infections	69
Tumor induction and tissue isolation	70
Ethics statement.....	71
Integration site mapping and quantification	71
Sequencing Analysis	72
Analysis of ALV integrations with respect to different genome annotations and features	74
Analysis of ALV integrations with respect to gene expression and levels of splicing	75
Proviral Load quantification	75
Gene Ontology Analysis	77

Chapter 3. The MET gene is a common integration target in ALV subgroup J induced chicken hemangiomas	78
Summary.....	79
Introduction	81
Results	85
Characterization of an American ALV-J isolate PDRC-59831	85
Sequencing of ALV-J/cellular integration junctions.....	89
Some ALV-J integrations are present in expanded clones.....	94
<i>MET</i> intron 1 is an expanded common integration site.....	96
<i>MET</i> mRNA expression is elevated in tumors containing <i>MET</i> proviral integrations.....	100
Discussion.....	102
Materials and Methods	105
Tumor induction.....	105

Phylogenetic analysis of ALV-J isolate PDRC-59831 <i>env</i> gene	106
DNA extraction and deep sequencing	106
Sequencing Analysis	108
<i>MET</i> gene expression.....	109
 Chapter 4. Avian leukosis virus activation of an antisense RNA upstream of TERT in B-cell lymphomas	110
Summary.....	111
Introduction	112
Results	115
TERT promoter is a common ALV integration site in B-cell lymphomas ...	115
Novel antisense (TAPAS) RNA is transcribed from TERT bidirectional promoter	118
TAPAS RNA expression is elevated in tumors with integrations in the TERT promoter	121
Viral transcripts splice into exon 4 of TAPAS RNA.....	123
TAPAS RNA is expressed in normal chicken tissues and during development.....	127
TAPAS RNA is conserved	129
Discussion.....	132
Materials and Methods	136
Tumor induction.....	136
DNA and RNA isolation	136
Reverse transcription, PCR amplification and Sanger sequencing	137
Detection of splice variants and ALV-TAPAS RNA fusion transcripts	137
High throughput sequencing.....	138
Quantitative PCR.....	138
Evolutionary analysis.....	139
 Chapter 5. hTAPAS, a novel long non-coding RNA in the TERT promoter region, regulates TERT expression	140
Summary.....	141

Introduction	143
Results	146
An antisense transcript is expressed upstream of hTERT	146
hTAPAS exhibits features of a long non-coding RNA.....	153
Sub-cellular localization of hTAPAS and hTERT	156
hTAPAS expression negatively correlates with hTERT expression in cancer patients.....	160
Knockdown of hTAPAS upregulates hTERT expression in HEK-293 cells... ..	164
hTAPAS expression does not correlate with the TERT promoter mutation... ..	166
Discussion.....	170
Materials and Methods	178
Analysis of RNA-seq and CAGE data	178
Evolutionary conservation and coding potential analysis	179
Fractionation experiments in HEK-293 cells	179
RNA isolation and qRT-PCR	179
hTAPAS knockdown experiments using antisense oligonucleotides	180
Generating hTAPAS constructs and performing luciferase assays.....	181
Chapter 6. Future directions	182
MET can be targeted for therapeutic intervention in development of hemangiomas.....	183
Identification of cooperating gene players in ALV induced tumorigenesis.....	186
Functions of human TAPAS RNA	187
Role of HBV integrations in altering hTAPAS	188
References	190
Curriculum Vitae	227

List of Tables

Table 2.1: Percent of integrations observed for different genome features via HOMER bioinformatics analysis	26
Table 2.2: List of different normal and infected tissues analyzed for ALV integrations.....	28
Table 2.3: Putative cooperating gene players that might cooperate with TERT and MYB in tumorigenesis.....	61
Table 2.4: List of gene players targeted by the different families of transcription factors that are enriched in tumors	62
Table 3.1: Tissues collected and neoplasms observed	91
Table 3.2: Top 20 common integration sites in ALV-J induced hemangiomas ...	93
Table 5.1: Number of RNA-seq samples from TCGA for each tissue type for each cancer type investigated for <i>hTERT</i> and <i>hTAPAS</i> expression.....	150

List of Figures

Figure 2.1: The percentage of ALV integration events in genes	27
Figure 2.2: The distribution of ALV integration events in genes	30
Figure 2.3: Distribution of ALV integrations with respect to transcription start sites.	33
Figure 2.4: Distribution of ALV integrations in proximity to CpG islands	34
Figure 2.5: Influence of the expression of transcriptional units on ALV integrations	37
Figure 2.6: Distribution of ALV integrations with level of splicing	40
Figure 2.7: Correlation between ALV integrations and the number of alternative transcripts in human HeLa cells	41
Figure 2.8: Distribution of integration sites in ALV infected tissue culture	44
Figure 2.9: Distribution of integration sites in ALV-induced B-cell lymphomas, and clonal expansion in metastases	45
Figure 2.10: A sample set of individual primary (bursa) and secondary (liver, spleen and kidney) tumors from the same bird	48
Figure 2.11: The OCI values of the different tissue samples	50
Figure 2.12: Top 10 most clonally expanded integrations for metastasized liver tumor D2L	51
Figure 2.13: Clonally expanded integrations from different slices of a tumor, suggest heterogeneity of bursa	52

Figure 2.14: Proviral Load (PVL) ranges between different neoplasms.....	54
Figure 2.15: Proviral load values are calculated for different tissues, using the env, gag or LTR regions of ALV	55
Figure 2.16: Functional classification of gene ontologies overrepresented among the clonally expanded integrations	60
Figure 3.1: Sequence comparison and phylogenetic analysis of env genes of ALV-J isolates	88
Figure 3.2: Library preparation	92
Figure 3.3: Distribution of integration sites in ALV-J-induced hemangiomas.....	95
Figure 3.4: The top 10 most abundant integrations identified in the ALV-J induced head, leg and ALV-A induced kidney hemangiomas	98
Figure 3.5: ALV-J integration in the MET gene in chicken hemangiomas and non-tumor tissues	99
Figure 3.6: <i>MET</i> mRNA expression measured by qRT-PCR.....	101
Figure 4.1: The TERT promoter region is a common site of ALV proviral integration in lymphomas.....	117
Figure 4.2: Schematic of the TAPAS gene.....	120
Figure 4.3: Expression of TAPAS RNA and <i>TERT</i> in ALV-induced B-cell lymphomas.....	122
Figure 4.4: Viral RNAs splice into exon 4 of TAPAS RNA.....	125

Figure 4.5: <i>TERT</i> and TAPAS RNA are expressed at comparable levels in adult tissues and during chick development.....	128
Figure 4.6: The TAPAS gene is conserved in avian species.....	131
Figure 5.1: An antisense RNA, named <i>hTAPAS</i> , is expressed in the <i>hTERT</i> promoter region.....	148
Figure 5.2: Normalized and stranded RNA-seq (Bedgraph) transcription coverage for <i>hTERT</i> (green) and <i>hTAPAS</i> (blue) expression are depicted for different ENCODE cell lines.....	149
Figure 5.3: Pipeline of TCGA RNA-seq data analysis.....	151
Figure 5.4: Gene models and expression of hTERT and hTAPAS in tumors...	152
Figure 5.5: hTAPAS does not exhibit conservation.....	154
Figure 5.6: hTAPAS does not exhibit protein-coding potential.....	155
Figure 5.7: Sub-cellular localization of hTERT and hTAPAS.....	158
Figure 5.8: Proportion of primary tumor samples with detectable expression of hTERT or hTAPAS or both.....	162
Figure 5.9: hTERT and hTAPAS expression are inversely correlated in primary tumors.....	163
Figure 5.10: hTAPAS knockdown in HEK-293 cells results in increased hTERT expression.	165
Figure 5.11: hTAPAS and hTERT expression (FPKM) in 8 different human liver cancer cell lines.....	168

Figure 5.12: Presence of the hTERT promoter mutation does not alter hTAPAS expression.....169

Figure 5.13: TAPAS is a viral integration hotspot in chicken and human tumors..
.....177

List of Abbreviations

ALV	avian leukosis virus
ERV	endogenous retrovirus
HIV	human immunodeficiency virus
LTR	long terminal repeat
MLV	murine leukemia virus
FV	foamy virus
HTLV	human T-cell leukemia virus
PIC	preintegration complex
RSV	Rous sarcoma virus
TSS	transcription start site
UIS	unique Integration site
TCGA	the cancer genome atlas
TERT	telomerase reverse transcriptase
TAPAS	TERT antisense promoter associated

Chapter 1. Introduction to Avian Leukosis Virus

Introduction

Over a century ago, two Danish researchers, Vilhelm Ellerman and Oluf Bang, began the work of studying avian retroviruses. They showed that avian leukosis (a leukemia of chickens), could be transmitted by cell-free filtrates (Ellermann and Bang, 1908). Three years later, an American scientist, Peyton Rous (a Johns Hopkins university alumnus), discovered a virus named the Rous sarcoma virus (RSV). Rous showed that the virus could be transmitted through cell free extracts of solid chicken sarcoma tumors (Rous, 1911). This discovery of the first solid tumor oncogenic virus led to Rous being awarded the Nobel Prize in 1966. We now know that RSV is derived from the avian leukosis virus (ALV) and contains the *src* oncogene. RSV and ALV have since been studied extensively and have made significant contributions to our present day understanding of cancer as well as of retroviral entry, integration and replication (Weiss and Vogt, 2011).

1.1 Introduction to avian retroviruses

The avian leukosis virus (ALV) is a simple exogenous retrovirus that belongs to the Alpharetrovirus genus of the family Retroviridae. The ALV genome is a single stranded RNA, composed of three genes – *gag*, *pol* and *env*, flanked by long terminal repeat sequences (LTRs). The *gag* gene encodes the structural proteins of capsid (CA), matrix (MA) and nucleocapsid (NC) and the protease enzyme (PR). The *pol* gene encodes the viral enzymatic components that include the reverse transcriptase (RT) and integrase (IN) enzymes. The *env* gene encodes the surface glycoprotein

(SU) and transmembrane (TM) proteins. The flanking LTRs contain strong promoter and enhancer elements that drive the expression of viral genes. The chicken genome contains many endogenous retroviruses (ERVs), some of which are related to exogenous ALVs (Bolisetty et al., 2012).

There are 11 ALV subgroups named A, B, C, D, E, F, G, H I, J and K (Fadly and Nair, 2008). These are classified based on differences in their env coding sequences, and use of different host cell receptors for gaining entry into the host cell. This in turn also determines their viral host range and pathogenicity. The host cell surface receptors tva, tvc, and chNHE1 (Chai and Bates, 2006) allow viral entry of ALV subgroups A, C, and J respectively. The host protein tvb on the other hand allows entry of ALV subgroups B, D, and E (Adkins et al., 2000; Barnard and Young, 2003; Brojatsch et al., 1996).

1.2 The life cycle of ALV

Nuclear Entry and Proviral Integration

Initially an ALV virion interacts with the host cell surface receptor via its env glycoproteins. This interaction induces an endosomal uptake of the virion into the host cell. As endocytosis progresses, the pH in the endosome becomes more acidic which triggers fusion of virion envelope with the host cell plasma membrane. This fusion mechanism mediates the entry of the virion capsid into the host cell cytoplasm (Barnard and Young, 2003). Once the capsid is released into the host cell cytoplasm it loosens its structure to allow entry of host cell dNTPs into the capsid particle. This step allows the

initiation of reverse transcription of the viral genome in the cellular cytoplasm.

Reverse transcription is a complex process divided into multiple steps. Initially, the packaged RT enzyme makes use of the host cellular dNTPs and packaged tryptophan tRNAs to convert the viral RNA genome into a double-stranded DNA. This process of reverse transcription is completed in the host cell nucleus (Werner et al., 2002), after which the proviral DNA genome integrates into the host genome. The viral integrase (IN) protein forms a preintegration complex (PIC) along with the host cell proteins. This complex facilitates proviral integration into the host genome via a two-step mechanism. Initially, the last 2 bases (TT) from the 3' ends of the proviral genome are cleaved by IN. PIC then binds the host cell DNA and allows IN to catalyze a strand transfer reaction via attacking the phosphodiester bonds on opposite strands of the host DNA (Bushman et al., 1990; Engelman et al., 1991). Once integration is completed, the host cell repair machinery fills the stretches of gaps on either ends of the integration site. As a result of this repair mechanism, ALV integration sites are flanked by a 6 nucleotide repeat sequence (Hughes et al., 1981). The FACT (facilitates chromatin transcription) complex, a chromatin remodeler, was recently reported to promote ALV integration *in vivo* (Winans et al., 2017a).

Transcription and Nuclear Export

The viral LTRs direct the process of proviral transcription. The host transcription factors bind in the U3 region of the LTR and in turn facilitate RNA Polymerase II transcription. The proviral transcripts then undergo transcriptional processing, making use of the host cellular machinery. This includes 5' capping, splicing, 3' end cleavage and polyadenylation of proviral transcripts, followed by their nuclear export.

A replication-competent provirus generates a single primary RNA transcript. This transcript serves as the template for the translation of *gag* and *pol*, and also serves as the genome material to be packaged into new virions. A portion of these full-length primary transcripts also undergo splicing to generate the *env* mRNA.

Furthermore, the ALV genome is composed of different RNA elements that allow efficient transcription, export, and translation of the long unspliced proviral transcripts. For example the suboptimal 3' splice sites (McNally and Beemon, 1992), and cis-acting RNA elements serve to protect a portion of the proviral transcripts from splicing. The negative regulator of splicing (NRS) is one such cis-acting RNA element, which acts as a faux 5' splice site and recruits the spliceosome. As a result, this sequesters the 3' splice site away from the 5' splice site and reduces the splicing efficiency (Giles and Beemon, 2005).

In order to mediate nuclear export of its transcripts, the ALV genome utilizes a 100 nt direct repeat (DR) sequence, present in the 3' untranslated

region (UTR) of the proviral transcript. The DR element forms a stable stem loop structure, which allows nuclear export by interaction with the cellular nuclear export factor Tap (LeBlanc et al., 2007).

Since the proviral transcript contains a stop codon at the end of *gag*, the downstream region appears as a long 3' UTR, which can be targeted for degradation by the cellular nonsense mediated decay (NMD) machinery. In order to avoid this NMD degradation, the ALV genome contains a 400 nt sequence immediately downstream of the gag termination codon, named the RNA stability element (RSE) (Weil et al., 2009).

Translation

The full-length proviral transcript serves as a template for the translation of the Gag-Pro and Gag-Pro-Pol polyproteins. Even though *gag* and *pol* are encoded on the same transcript, ALV synthesizes more structural Gag proteins than the enzymatic Pol proteins, as is required for new virions. This is facilitated via a short A-U rich “slippery sequence” upstream of the gag termination codon, followed by a pseudoknot sequence (Jacks et al., 1988). The pseudoknot pauses the ribosome over the slippery sequence and occasionally slips it backward (~5% of the time) by a single nucleotide. This slippage induces a -1 frameshift, and renders the gag termination codon out of frame. This in turn allows the ribosome to read through until the next termination codon of pol, thereby generating the Gag-Pro-Pol polyprotein.

The spliced proviral transcript serves as template for the Env polyproteins. The *env* splice donor sequence is placed within *gag*, downstream of the *gag* start codon. As a result the first six amino acids of Gag protein are a part of the beginning of the Env polyprotein (Ficht et al., 1984; Swanstrom and Wills, 1997). Post translation, the Env polyprotein undergoes proteolytic cleavage into three fragments. These fragments are glycosylated in the endoplasmic reticulum (ER), and undergo folding to form a trimer (Einfeld and Hunter, 1988). This processed functional Env trimer is exported to the cell surface to bind the host receptor that is also processed by the secretory pathway. This mechanism of the Env binding to the cellular receptor inhibits any further infection of the cell via other virions that utilize the same host cell surface receptor. This phenomenon is called super infection resistance.

Virion Assembly and Budding

The Gag proteins are the primary factors that mediate the assembly and budding processes of new virions. The Gag protein is transiently imported to the nucleus after its synthesis, where the NC domain of Gag interacts with the proviral RNA packaging sequence (Ψ) (Gudleski et al., 2010; Scheifele et al., 2002). A nuclear export signal within the p10 domain of Gag then allows export of the ribonucleoprotein (RNP) complex via the CRM1 Pathway (Gudleski et al., 2010; Scheifele et al., 2005).

The RNP complex undergoes phosphoinositide-dependent (Nadaraia-Hoke et al., 2013) trafficking from the nucleus. The membrane-binding domain (MBD) at the N-terminus of Gag mediates stable association with the plasma membrane (Verderame et al., 1996). At the plasma membrane the processed Env proteins, viral polyproteins, two linked genomic RNAs and tryptophan tRNAs assemble into the virion particle that buds from the cell surface. The Gag protein and different cellular proteins are involved in facilitating this process (Pincetic and Leis, 2009; Swanstrom and Wills, 1997). Shortly after budding, the viral protease cleaves the polyproteins, which allows the virion to undergo maturation and makes it capable of infection.

1.3 ALV as an insertional mutagenesis tool

ALV integration events into the host genome can up regulate the expression of nearby host genes via the strong enhancer and promoter elements in the viral LTRs. ALV is a slowly transforming virus since it lacks an oncogene in the viral genome. ALV integrations can induce tumor formation by integration near or within cancer related host genes, and alter their expression and function. The viral genome could also induce the expression of truncated gene products or alter the mechanisms of splicing or polyadenylation of host genes (Jiang et al., 1997). Since enhancer elements can function at far away loci via DNA looping, ALV can also perturb the expression of host genes that are at very distant sites from the viral

integrations (Li et al., 2014). This process, discovered with ALV, is called viral insertional mutagenesis (Hayward et al., 1981; Payne et al., 1982). Therefore, ALV serves as an insertional mutagenesis tool that alters the host gene activity in different ways.

Retroviruses like HIV-1 preferentially integrate into actively transcribed and highly spliced genes (Schröder et al., 2002; Singh et al., 2015). The host targeting factor LEDGF (lens epithelial derived growth factor), a general transcriptional co-activator, was found to play a role in facilitating HIV-1 integration into gene regions (Maertens et al., 2003). Murine leukemia virus (MLV) on the other hand preferentially integrates in proximity to transcription start sites and CpG islands (Mitchell et al., 2004; Wu et al., 2003). MLV's preference for the TSS is mediated by the binding of the host bromodomain and extraterminal domain (BET) proteins to the MLV integrase (Sharma et al. 2013). In contrast to HIV-1 or MLV, ALV has previously been reported to integrate relatively randomly with only slight preferences for transcribed genes (Barr et al., 2005; Mitchell et al., 2004; Narezkina et al., 2004; Withers-Ward et al., 1994). In this work (Chapter 2), we analyze ALV integrations in a high throughput fashion, in chicken and human cells to determine its integration preferences. Our studies are an improvement from previous reports in terms of both the number of proviral integrations analyzed, and the robustness of the reference genome utilized.

If ALV integration occurs near or within an oncogene, the host cell can undergo clonal proliferation due to oncogenic transformation. This ALV

induced clonal expansion can trigger tumor development. Identification of common viral integration sites in tumors can therefore be used to identify novel genes that are implicated in oncogenesis. ALV integration occurs in a quasi-random fashion and exhibits minimal discrimination for integration near genome features relative to other well-studied retroviruses. Therefore ALV, as an insertional mutagen, allows the study of biological processes and underlying gene players involved in oncogenesis with minimal bias. In order to observe selection for oncogenic characteristics, we compare integration sites in cultured cells with those in ALV-induced tumors, as described in Chapter 2.

1.4 ALV integrations help define the clonal progression of tumors

As identified by previous work, ALV tumors are considered to be clonal (Neiman et al., 2003; Yang et al., 2007a). Clonal expansion in cancer is the phenomena of a small number of precursor cells undergoing transformation and proliferating into expanded clones, thus giving rise to the tumor. The empirical degree of oligoclonality and extent of clonal expansion in different stages of tumorigenesis has previously been defined for HTLV-1 and HTLV-2 induced tumors (Berry et al., 2012; Gillet et al., 2011). We analyzed ALV infection in tumors by quantifying the clonal abundance and distribution of integrations during progression of tumors, as described in Chapter 2.

Oncogenes were originally identified through study of avian viruses (Beemon and Rosenberg, 2012; Hayward et al., 1981; Swanstrom et al., 1983). ALV therefore, serves as a useful tool to study the transformation and clonal proliferation of B-cells. Our work here, studying these ALV unique integration sites (UISs), reflects their role in inducing lymphomas in chicken B-cells. B-cell development in chicken and mammals is a very similar process (Kohonen et al., 2007). These similarities are evident at levels of molecular changes and gene regulatory networks (Koskela et al., 2003; Weill et al., 2004; Wu et al., 2004). For example, the results on the c-myc-regulated genes are in agreement between human and chicken studies in that growth and energy control genes are consistently up regulated (Coller et al., 2000).

Even though mice serve as a good mammalian model, in terms of oncogenesis they differ in some fundamental ways from humans. Unlike human telomerase, the mouse telomerase enzyme is active in normal somatic cells (Hackett and Greider, 2002). This discrepancy between humans and mice is important because telomerase activation is a critical step in the human oncogenic process, with telomerase activation seen in approximately 90% of human cancers (Garcia et al., 2007a; Shay and Wright, 2011). Similar to human expression, chicken telomerase expression is down regulated in most somatic tissues (Delany and Daniels, 2004). Furthermore, chicken telomeres shorten with age, and telomerase activity is important for oncogenesis (Delany et al., 2000). Therefore, chicken serves

as an advantageous model over mouse, to study oncogenic events. Since, common and conserved mechanisms are fundamental, study of genetic alterations in chickens as a model organism can help unravel the way human systems are regulated in B-cell lymphomagenesis.

Cancers exist in a number of stages, characterized by a spectrum of divergent cells and genetic changes. Each cancer is unique, and in any given tumor, the clonal structure shifts over time (Greaves and Maley, 2012). This evolution involves a process of clonal selection and expansion. The underlying dynamics are complex with highly variable patterns of genetic diversity and resultant clonal architecture.

In cases of ALV-induced B-cell lymphomas, the bursa serves as the primary organ of malignant transformation and site of tumorigenesis (Baba and Humphries, 1985; Neiman et al., 2003). Infected chickens typically develop multiple primary neoplastic follicles in the bursa, some of which may eventually form primary tumors. The development of these neoplasms is a multi-stage process (Baba and Humphries, 1985; Cooper et al., 1968; Neiman et al., 1980a). In order to examine the clonality of lymphomagenesis, we studied different stages of cancer progression. These include inflammation, neoplastic follicles, primary tumors in the bursa, and metastatic tumors at secondary sites. The stage of neoplastic follicles in the bursa are part of tumor progression towards transformation and malignancy (Baba and Humphries, 1985; Neiman et al., 1985). Metastases of primary tumors are observed from bursa to secondary sites of liver, spleen, and kidneys.

In an emerging picture of B-cell malignancy, understanding tumor progression is an important piece of the puzzle. Cancers acquire, via mutational and epigenetic changes, a variety of traits that trigger clonal expansion, via proliferation, migration and invasion. These properties are alterations to normal developmental and physiological cellular processes (Greaves and Maley, 2012). Genomic analysis of ALV integrations across progressing stages of a tumor helps determine the extent of diversity within the tumor and highlights the different mutational processes that operate at different times in tumor initiation, evolution, and metastasis. In Chapter 2, we show that clonal expansion of ALV-infected B-cells is a key feature of malignant transformation in tumors. Furthermore, we analyze the ontology of genes flanking integrations to identify biological pathways that are deregulated in lymphomas. Our work reveals that five major gene functions contribute to clonal dominance of lymphomagenesis: regulation of proliferation, differentiation, immune response, apoptosis, and phosphorylation.

1.5 ALV integrations allow identification of novel cancer associated genes

While ALV can infect different bird species like ducks and turkeys (Li et al., 2013), chickens are the natural hosts of ALV infections. When ALV integration occurs within or near an oncogene it can potentially up regulate its expression via strong viral enhancers and promoters in the LTRs, and

drive tumorigenesis. These integration events can induce a wide range of tumors, the most common of which is lymphoid leukosis. This is a B-cell lymphoma that originates in the bursa of Fabricius, an avian specific organ that serves as the site of B-cell development (hematopoiesis). The lymphoma can then metastasize to the distant organs of the liver, spleen and kidney (Fadly and Nair, 2008). Other ALV induced tumors include hemangiomas, myeloid leukosis and erythroblastomas (Beemon and Rosenberg, 2012; Justice IV and Beemon, 2013; Justice et al., 2015a). ALV infections predominantly spread horizontally via direct or indirect contact, but can also be transmitted vertically from hen to egg (Justice IV and Beemon, 2013).

ALV induced tumors can vary between 8 and 24 weeks, and are dependent on the type of viral strain and the age of the bird at the time of infection. As an indiscriminant insertional mutagenesis tool, common ALV integrations in tumors have previously allowed the identification of several cancer associated genes like *MYC*, *MYB*, *BIC* (miR 155 precursor) and *TERT* (Baba and Humphries, 1986; Clurman and Hayward, 1989; Hayward et al., 1981; Justice et al., 2015b; Yang et al., 2007a).

Initial studies mapping integrations in long latency B-cell lymphomas identified *MYC* as a viral integration hotspot. Most of the tumors (approximately 80%) had integrations in the *MYC* intron 1 region (Hayward et al., 1981; Neel et al., 1981), and led to the expression of a viral fusion transcript with *MYC* RNA. These tumors, containing *MYC* integrations, were

identified in chickens infected with ALV at 2-7 days post hatching. Such infections resulted in development of late onset B-cell lymphomas, around 4-6 months after infection.

Later work in late onset B-cell lymphomas also identified integrations in the *BIC* gene i.e. the precursor to the noncoding microRNA mir-155. Interestingly, the *BIC* integrations were frequently observed in tumors with the *MYC* integrations, and were implicated in later stages of lymphomagenesis (Clurman and Hayward, 1989; Tam et al., 1997).

Further work showed that infection of 10-day-old chicken embryos with a recombinant ALV strain (EU-8) resulted in rapid onset B-cell lymphomas. The *MYB* locus was identified as the viral integration hotspot in these tumors (Kanter et al., 1988). Most of the integrations occurred in intron 1, upstream of the AUG initiation codon, and induced formation of truncated *MYB* viral fusion transcripts. It was later identified that the EU-8 genome possesses a deletion in the NRS element, and therefore, allows increased efficiency of viral read-through transcription, and viral splicing to downstream genes (Smith et al., 1997)

Previous studies have made use of low-throughput techniques to identify hotspots of proviral integrations to identify genes that drive ALV-induced tumors. In chapters 2, 3 and 4 of this work, we use high throughput techniques to study hotspots of ALV integrations in tumors to identify additional gene players, which are implicated in oncogenesis. We make use of an insertional mutagenesis screen via use of an ALV variant, LR-9, with a

mutation in the NRS (G919A), which induces rapid-onset B-cell lymphomas (Polony et al., 2003).

The ALV subgroup J, ALV-J, was discovered in the late 1980's (Bai et al., 1995), and has since caused huge economic losses to the poultry industry, especially in China. ALV-J is believed to have originated from a recombination event between ALV and an endogenous retroviral element (Fadly and Smith, 1999; Sacco et al., 2000; Venugopal, 1999). The virus induces a different spectrum of tumors compared to other ALV subgroups. These mostly include high frequency of myeloid leukosis and hemangiomas (Cheng et al., 2010; Payne et al., 1992). The molecular basis of oncogenesis in these tumors was until recently not well understood. In our work, as described in Chapter 3, we identify the MET gene as a common viral integration target in ALV- J and ALV-A induced hemangiomas (Justice et al., 2015a).

The telomerase reverse transcriptase (TERT) promoter region was identified in our lab as a common integration hotspot for clonally expanded ALV integrations in rapid-onset B-cell lymphomas (Justice et al., 2015b; Yang et al., 2007a). In Chapter 4 these integration sites are re-analyzed to reveal the activation of a novel antisense lncRNA named the TERT antisense promoter associated (TAPAS) RNA. This is the first report of a lncRNA being activated by retroviral insertional mutagenesis. To extend our work to human cancers, as described in Chapter 5, we also identify a lncRNA in the human TERT promoter region named human TAPAS

(hTAPAS). Since hTAPAS down-regulates TERT expression, we propose its role in maintaining telomere homeostasis.

**Chapter 2. Selection for Avian Leukosis Virus integration sites
determines the clonal progression of B-cell lymphomas**

Adapted from:

Malhotra S., Winans S., Lam G., Justice J., Morgan R., Beemon K. (2017).
Selection for avian leukosis virus integration sites determines the clonal
progression of B-cell lymphomas. PLoS Pathog. 2017 Nov
3;13(11):e1006708.

Summary

Avian leukosis virus (ALV) is a simple retrovirus that causes a wide range of tumors in chickens, the most common of which are B-cell lymphomas. The viral genome integrates into the host genome and uses its strong promoter and enhancer sequences to alter the expression of nearby genes, frequently inducing tumors. In this study, we compare the preferences for ALV integration sites in cultured cells and in tumors, by analysis of over 87,000 unique integration sites. In tissue culture we observed integration was relatively random with slight preferences for genes, transcription start sites and CpG islands. We also observed a preference for integrations in or near expressed and spliced genes. The integration pattern in cultured cells changed over the course of selection for oncogenic characteristics in tumors. In comparison to tissue culture, ALV integrations are more highly selected for proximity to transcription start sites in tumors. There is also a significant selection of ALV integrations away from CpG islands in the highly clonally expanded cells in tumors. Additionally, we utilized a high throughput method to quantify the magnitude of clonality in different stages of tumorigenesis. An ALV-induced tumor carries between 700 and 3000 unique integrations, with an average of 2.3 to 4 copies of proviral DNA per infected cell. We observed increasing tumor clonality during progression of B-cell lymphomas and identified gene players

(especially TERT and MYB) and biological processes involved in tumor progression.

Introduction

Avian leukosis virus (ALV) is a simple retrovirus that causes cancer, primarily B-cell lymphomas in chickens (Beemon and Rosenberg, 2012; Hayward et al., 1981; Justice IV and Beemon, 2013). The ALV genome does not contain a viral oncogene and induces aberrant host gene expression via use of strong viral enhancer and promoter elements. Relative to other well studied retroviruses like HIV-1 and MLV, ALV was shown to integrate relatively randomly into the host genomic DNA, with little bias for genomic features (Barr et al., 2005; Kvaratskhelia et al., 2014; Mitchell et al., 2004; Narezkina et al., 2004). The FACT (facilitates chromatin transcription) complex, a chromatin remodeler, was recently reported to promote ALV integration (Winans et al., 2017a).

ALV-induced lymphomas develop in a multistage process, appearing initially as neoplastic follicles in the bursa, which develop into primary bursal tumors. Primary tumors can then metastasize to distant organs and form secondary tumors (Neiman et al., 2003). Cellular transformation occurs through multiple genetic changes in oncogenes and tumor suppressor genes, as well as noncoding RNAs (Huarte, 2015). These oncogenic changes can occur via different genetic mechanisms; insertional mutagenesis by retroviruses is one such mechanism. Viral integration into the host genome can alter host gene expression and induce cancer development by selection of cells with oncogenic changes (Justice IV and

Beemon, 2013). In turn, common viral integration targets observed in multiple tumors can help identify oncogenes (Baba and Humphries, 1986; Clurman and Hayward, 1989; Hayward et al., 1981; Jiang et al., 1997; Justice et al., 2015a, 2015b; Nehyba et al., 2016; Yang et al., 2007a). Consequently, retroviral-mediated lymphomagenesis in chickens provides an excellent experimental model system for analysis of neoplastic change in tumors of B-cell lineage (Neiman, 1994).

In tumors ALV integrations can induce cancer development by selection of cells with tumorigenic mutations. Analysis of ALV integrations in B-cell lymphomas allows identification of host genes that might contribute to development of tumors and their progression. We have previously identified common proviral integrations in ALV-induced B-cell lymphomas, notably the *TERT* promoter region, and hemangiomas (Justice et al., 2015a, 2015b; Yang et al., 2007a). Since, selection in tumorigenesis alters the pattern of viral integration sites, we analyze integrations in cultured cells, to identify preferences of ALV integrations in an unbiased way without the influence of oncogenic selection. We investigate how ALV integrations in tissue culture correlate with previously unreported genomic features of splicing and distribution within transcription units. We observe an enrichment of ALV integrations in the 5' end of gene bodies, proximal to CpG islands and transcription start sites, as well as a preference for expressed and highly spliced genes. No association was observed with levels of alternative splicing.

In order to determine the effects of selection for oncogenic characteristics, we compare integration sites in cultured cells with those in ALV-induced B-cell lymphomas. ALV tumors are considered to be clonal, as determined by previous work (Neiman et al., 2003; Yang et al., 2007a). However, the clonality of these neoplasms has not been empirically defined. We analyze ALV infection in tumors by quantifying the clonal abundance and distribution of integrations during progression of tumors. Using the statistical Gini index, we calculate the empirical degree of oligoclonality and extent of clonal expansion in different stages of tumorigenesis (Berry et al., 2012; Gillet et al., 2011). Furthermore, the gene ontology analysis of host genes most proximal to proviral sites provides insight into underlying gene players and their contribution to oncogenic transformation.

We observe that ALV integrations are especially strongly selected for proximity to transcription start sites in tumors. Quantifying the clonality index and average number of integrations per cell within individual tumors helps determine the clonal architecture and hierarchies of lymphomagenesis. Thus, our work helps unravel how the integration sites of ALV are selected for in oncogenesis and play a role in the clonal progression of tumors. This is the most in depth analysis for ALV integration sites and is novel in terms of being able to lead from a stepwise early infection (in tissue culture) through to early and late tumor development.

Results

ALV integration is enriched within genes in cultured cells and in tumors

Via deep sequencing, we analyzed approximately 85,000 unique ALV integration sites (UISs) in tissue culture and tumors, as summarized in **Table 2.1**. Randomly generated integration sites were used as a control for all subsequent analysis. ALV integrations were analyzed for different ALV subgroups (A, C and J) in different infected cell types, including the chicken embryo fibroblasts (CEF), DT-40, a chicken B-cell lymphoma cell line, and human HeLa cells. After mapping the UISs to the host genome, we used the HOMER bioinformatics tool (Heinz et al., 2010) to associate integrations with specific annotated genomic sites, such as genes, exons, introns and CpG islands. ALV integrations were observed to be near random for most of the analyzed genomic features (**Table 2.1**).

The analysis of ALV integrations in CEFs, using the ensembl *Gallus gallus* 4 genome, suggested that the majority of the integrations (approximately 60%) occur in intergenic regions, similar to the random events (73%) (**Table 2.1**). However, we observed a significant bias for ALV integration into genes (approximately 40%) relative to random events (27%) (t test, p -value 0.006) (**Figure 2.1**). We also observed a slight enrichment for integrations near LINE sequences, gene promoters, simple repeat and satellite DNA sequences; however, these were not statistically significant (data not shown). Independent analysis of all these features were consistent,

for infections with ALV subgroup C in DT-40 B cells and in HeLa cells, suggesting that these preferences are not cell type specific.

To study selection of ALV integration sites in tumors, we sequenced 72 tissue samples from 41 different birds (**Table 2.2**). We obtained 17.2 million reads, originating from viral integrations in neoplasms and non-tumor tissues, which were mapped to 71,368 UISs. Similar to the integration pattern in cultured cells, integrations in the ALV-induced tumors, at the primary sites of bursa or secondary metastases, showed a significant enrichment for genes, relative to random (t test, p -value 0.022) (**Figure 2.1**).

Genome Feature	Random (Galgal 4) (%)	ALV-A CEF (%)	ALV-C CEF (%)	ALV-J CEF (%)	ALV-C DT-40 (%)	ALV-A Tumors (%)	Random (hg19) (%)	ALV-C HeLa (%)
+/- 5kb of TSS	6	8.6	10.9	8.3	9.8	15.4	7.9	13.7
TTS	0.7	0.8	0.9	0.5	1	1.1	0.9	1.3
LINE	6.8	9.9	9.3	7.8	10.7	7.5	7.2	7.8
SINE	0.2	0.2	0.1	0.1	0	0.1	0.1	0.1
Promoter	0.6	1.1	1.2	0.8	0.6	1.4	0.5	1.2
CpG Island	1.1	3.9	2.7	2.3	2.5	1.4	1.2	3.1
Low complexity	0.6	0.4	0.2	0.3	0.3	0.3	0.5	0.2
Simple repeat	0.6	0.7	2.1	5.2	1.5	1.9	0.8	1.9
Satellite	0.3	1.9	3.4	1.4	4.5	1.2	0.5	0.9
Genes	27	38	38.7	41.1	40	37.6	25.4	36.3
Intergenic	73	62	61.3	58.9	59.6	62.4	74.6	63.7

HOMER bioinformatics analysis was performed for infections with different ALV subgroups (A, C and J) in tissue culture (CEF, DT-40 and HeLa cells) and tumors. Percentage of integrations for each category is depicted. As a control, analysis was repeated for a matched number of random sites in the chicken genome (Galgal 4) or human genome (hg19).

<https://doi.org/10.1371/journal.ppat.1006708.t001>

Table 2.1 Percent of integrations observed for different genome features via HOMER bioinformatics analysis. Analysis was performed for the chicken (Galgal4) or human (hg19) genomes. The percent of ALV integrations analyzed for different ALV subgroups (A, C and J) in tissue culture (CEF, DT40 and HeLa cells) or tumors are indicated. Matched numbers of random sites were used as control for each analysis.

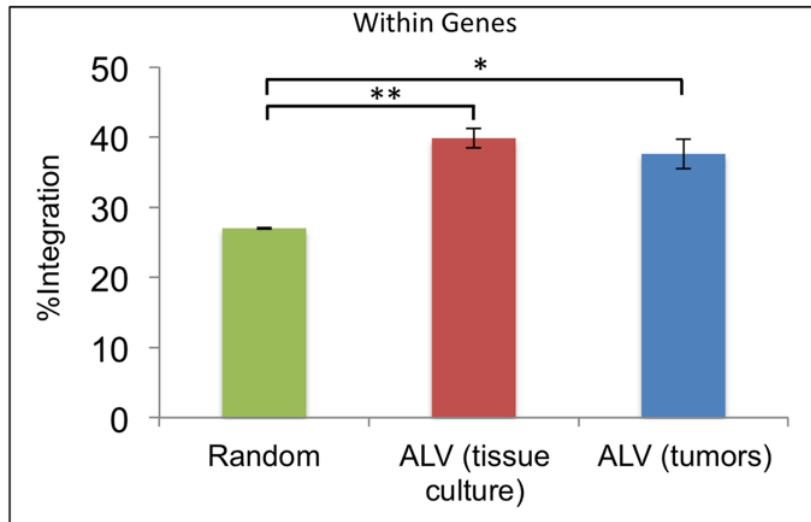


Figure 2.1 The percentage of ALV integration events are determined in tissue culture and in tumors within gene bodies (transcription start site to termination site), relative to a control number of random events, using the ensembl *Gallus gallus* 4 genome.

Type	Tissue	#Samples
Normal	Bursa	2
Normal	Liver	1
Non-tumor	Brain	2
Non-tumor	Kidney	1
Inflammation	Liver	2
Neoplastic/Transformed follicle	Bursa	9
Primary Tumor	Bursa	25
Metastases	Liver	13
Metastases	Kidney	10
Metastases	Spleen	7

Table 2.2 List of different normal and infected tissues analyzed for ALV integrations.

ALV integrations are enriched at the 5' end of gene bodies in cultured cells and in tumors

We analyzed the distribution of ALV integrations within transcriptional units by dividing the gene bodies into 10 equal segments or bins. Then, we calculated the number of integrations within each bin to determine the density of ALV integrations within a given part of a transcriptional unit. Relative to the matched control set (11.99%), there was a significant enrichment of ALV integrations towards the 5' end of the gene body in both tissue culture (16.25%) (*t* test, p-value 0.031), and in tumors (16.22%) (*t* test, p-value 0.004). This bias is most distinct within the first 10% of the gene body, i.e. in proximity to the transcriptional start site (**Figure 2.2**).

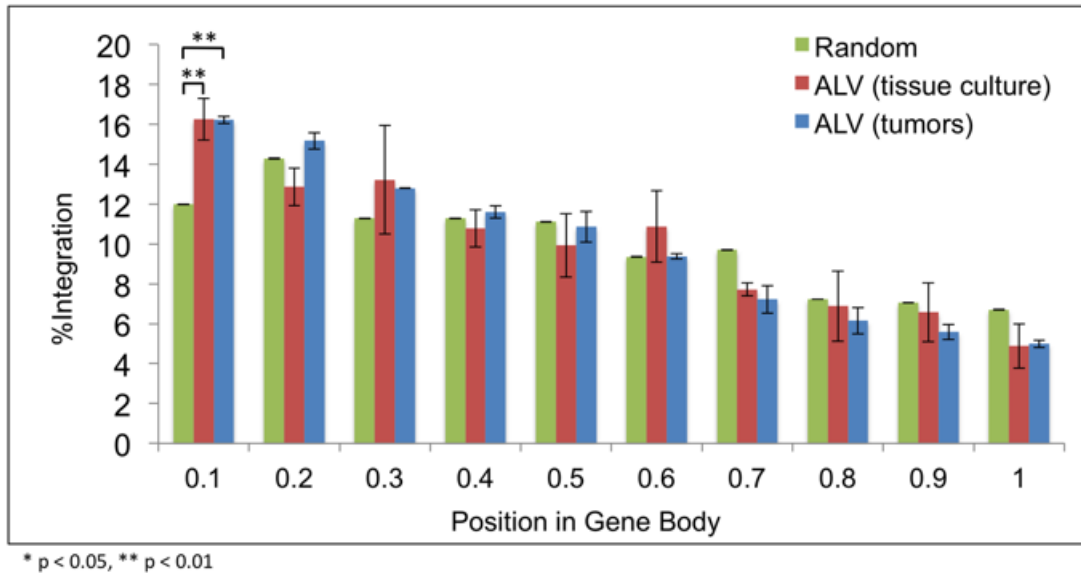


Figure 2.2 Each transcription unit was divided into 10 equal bins beginning from the 5' end of the transcription start site. The percentages of ALV integrations were calculated for each bin. Random integration events were used as a relative control.

ALV integration near transcription start sites and CpG islands varies in cultured cells versus tumors

To determine the pattern of integrations surrounding transcription start sites (TSS), we plotted the observed ALV integrations in cultured cells and in tumors with respect to the nearest TSS, extending over 50 kb on either end (**Figure 2.3A**). Relative to the random integration sites within 5 kb of the TSS, we observed a nearly 2-fold enrichment of ALV integration sites in tissue culture (t test, p -value 0.042) (**Figure 2.3B**). We observed an even greater frequency of integrations near TSS in tumors, with nearly 3-fold enrichment relative to random events within the 5 kb window (t test, p -value 0.006). This suggests that clonal selection of integration sites in tumors has a strong bias for proximity to TSS, probably to promote induction of aberrant gene expression in tumorigenesis. In addition, consistent with previously reported data from our lab (Justice et al., 2015b), relative to the surrounding region, we observed a drop in the integration frequency in the vicinity of the TSS in tumors (**Figure 2.3A**). Interestingly, we also observed this drop in tissue culture, suggesting it is a result of integration preference and not due to selection in tumors (**Figure 2.3A**).

Based on our initial HOMER analysis, we observed a 3-fold preference for integrations within CpG islands (approximately 3%) relative to random sites (1.1%) in cultured cells (t test, p -value 0.041) (**Figure 2.4A**). In contrast to tissue culture, the percentage of integrations within CpG islands was not enriched in tumors (1.4%), and appeared near random. The same

was true for the most clonally expanded integrations (0.96%), with 5 or more breakpoints (see Materials and Methods). To further investigate this, we determined the frequency of ALV integrations in the area immediately surrounding CpG islands (**Figure 2.4B**). Analyzing ALV integrations in tissue culture, we found that in the 1 kb region surrounding CpG islands, there was a 1.5 fold enrichment of integration relative to random (t test, p-value 0.047). If the window is expanded to 5 kb flanking the CpG island, the enrichment becomes more pronounced. Nearly 33% of all integrations are observed within 5kb of CpG islands, which is a 1.7 fold enrichment relative to the matched random control (21%) (t test, p-value 0.043).

In contrast, this enrichment for integration near CpG islands in cultured cells was not observed in tumors (**Figure 2.4B**). Frequency of integration within 1 or 5 kb of CpG islands in tumors was not significant relative to random events. When the same analysis was done for only the most clonally expanded integrations, with 5 or more breakpoints (see Materials and Methods), there is a striking depletion of integrations within 5kb of CpG islands (11.8%) relative to the total integration set in tumors (27.7%) as well as a matched random control (21%) (**Figure 2.4B**). This suggests that in tumors, there is an enrichment of integrations away from CpG islands (t test, p-value 0.045).

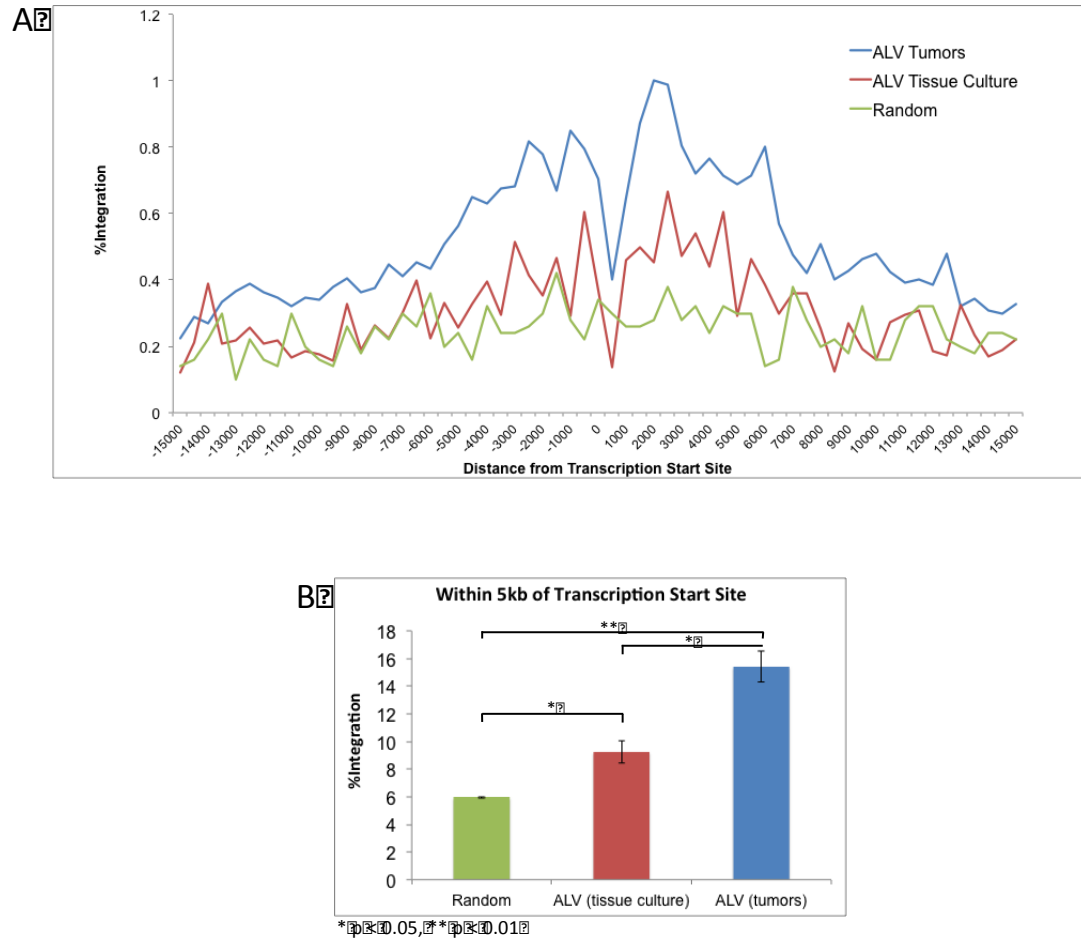


Figure 2.3 Distribution of ALV integrations with respect to transcription start sites. ALV integrations in tissue culture (n=15,416) and in tumors (n=71,368) were compared to a control number of random integration events in the ensembl *Gallus gallus* 4 genome. (A) A plot for the percentage of integration events within a 100 kb window of transcription start sites (TSS) is shown, with division into 500-bp bins. The blue and red lines represent ALV integrations in tumors and tissue culture respectively, and the green line represents the random control events. A preference for integration around TSSs is observed. (B) Percentage of integrations within 5 kb of the TSS are calculated for ALV integrations in culture and tumors versus random data.

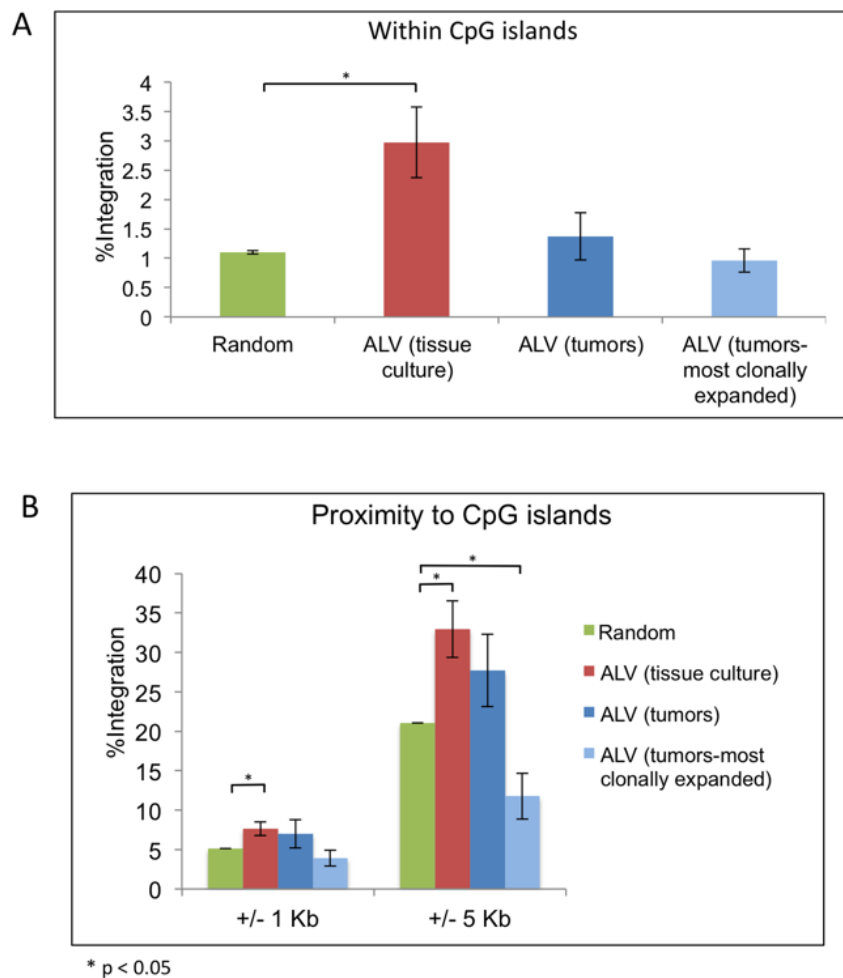


Figure 2.4 Distribution of ALV integrations in proximity to CpG islands. The percentage of ALV integration events in tissue culture and in tumors are determined for (A) within CpG islands or (B) within 1kb or 5 kb of CpG islands, relative to control number of random events. The most clonally expanded integrations in tumors include integrations with 5 or more breakpoints.

ALV integrations are enriched in expressed transcriptional units in cultured cells and in tumors

We next investigated ALV integration as a function of expression levels of the most proximal transcriptional units. In order to determine the background expression levels of host genes, we analyzed RNA-seq data for CEF, DT-40 and HeLa cells. We divided the whole transcriptome into 13 bins and determined the percentage of integrations within each bin, to observe any enrichment above background. While nearly 22.5% of the chicken RefSeq gene bodies (6,060) are not expressed in CEFs, only 8.2% of integrations occur in this expression bin. Thus, we observe a significant depletion in the percentage of integrations within or in proximity to unexpressed genes (t test, p -value 0.006) (**Figure 2.5**). For expressed genes, there was no bias observed for ALV integration with the level of gene expression, relative to random events. Thus, ALV integrates randomly in proximity to genes with low, intermediate or high levels of expression.

Since integrations can occur in intergenic regions at distant loci from transcriptional units, we asked whether ALV integrations within the transcriptional unit might exhibit a bias towards expressed genes. To address this, we repeated our analysis for only those integrations that occur within the gene body (between the transcription start and termination sites). Overall, we observed a similar integration pattern relative to background, with a more pronounced depletion of integrations within unexpressed genes.

In contrast to random events (25.4%), there was a nearly 6-fold decrease in preference for ALV integrations in genes with no expression (3.9%) (t test, p -value 0.003). Therefore, while ALV preferentially integrates near or within expressed genes, there is no preference for the level of gene expression. We also correlated ALV integrations with the gene expression levels via analysis of RNA-seq data for a subset of the ALV-induced lymphomas. Similar to our findings in cultured cells, we observed that in tumors ALV integrations are selected for proximity to or within expressed genes, but there is no distinct bias for varying gene expression levels (data not shown).

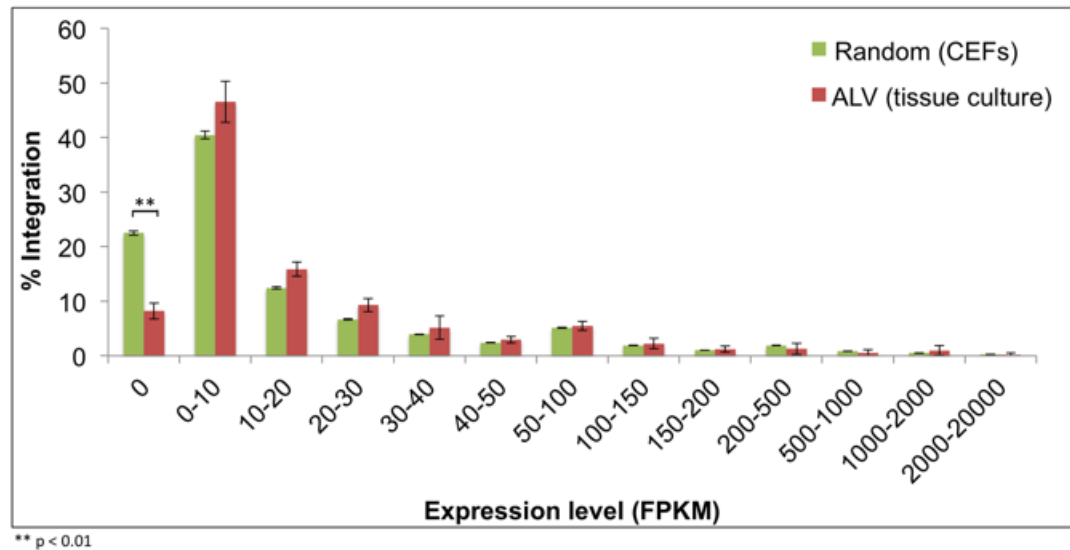


Figure 2.5 Influence of the expression of transcriptional units on ALV integrations. Expression levels of the chicken RefSeq transcriptional units (6,060) in CEFs were analyzed using available RNA-seq data sets, as described in materials and methods. Gene expression levels were divided, based on FPKM (Fragments Per Kilobase of transcript per Million mapped reads) expression values, into 13 bins. Numbers of integrations that occur near or within genes were then plotted into the bins as a percentage of the total and compared to random events.

ALV integration is targeted to highly spliced transcriptional units

HIV has been previously reported to preferentially integrate into genes that are highly spliced, i.e. have a greater number of introns (Singh et al., 2015). In order to determine whether ALV might display any similar preferences, we correlated ALV integration with the levels of mRNA splicing and alternative splicing. We associated the percent of integrations with the number of introns of the most proximal transcriptional unit, using the ensembl *Gallus gallus* 4 genome. While by random chance 12.1% of integrations are predicted to occur in unspliced transcriptional units, only 7.7% (*t* test, p-value 0.004) and 6.4% (*t* test, p-value 0.002) of ALV integrations in cultured cells and tumors, respectively, occurred within this range. Therefore, there is a significant lack of integration into unspliced genes in tumors. By random chance, the vast majority of the integrations occurred in transcriptional units with 1-19 introns (71.1%). For this window of splicing, ALV integrations in tissue culture (68.3%) appear close to random as well. On the other hand, nearly 22.1% of ALV integrations in tissue culture fall into highly spliced genes with 20 or more introns, which is a significant enrichment above random events (16.6%) (*t* test, p-value 0.012). Furthermore, ALV integrations in tumors (47.1%) have a greater enrichment for integration within or proximity to transcriptional units with 10 to 39 introns, relative to random sites (39.3%) (*t* test, p-value 0.026). Thus, we observe that ALV has a bias for integration into spliced genes with enrichment for higher levels of gene splicing (**Figure 2.6**).

We also asked whether the levels of ALV integration are associated with the number of spliced isoforms of the proximal transcription unit. The chicken genome is not well characterized for reporting alternative spliced variants of genes. Since the human genome is better characterized than the chicken genome, we repeated our analysis for ALV integrations in a HeLa cell line (**Figure 2.7**). However, our analysis did not identify any preference for integrations relative to the level of alternative splicing of proximal transcriptional units.

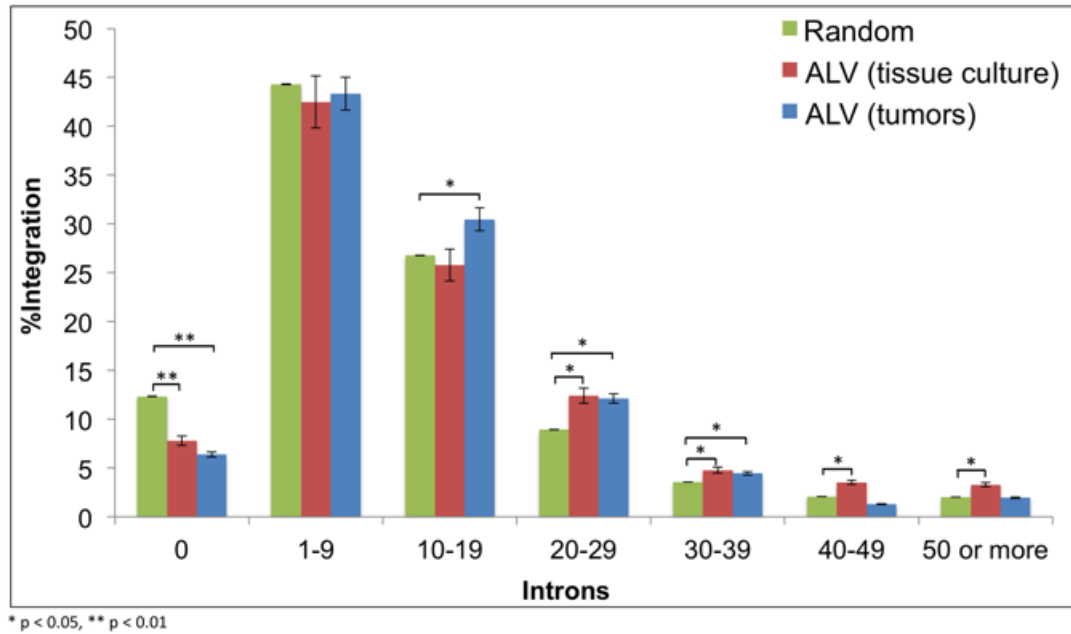


Figure 2.6 Distribution of ALV integrations with level of splicing

Correlation of ALV integrations in tissue culture and in tumors with the number of introns in the transcription units is depicted, using the ensembl *Gallus gallus* 4 genome. All the transcriptional units in the chicken genome are divided into 7 bins based on their number of introns, with the first bin depicting unspliced transcripts (with 0 introns). The X-axis shows the number of introns, and the Y-axis shows the percentage of integrations that occur near or within genes with the corresponding number of introns.

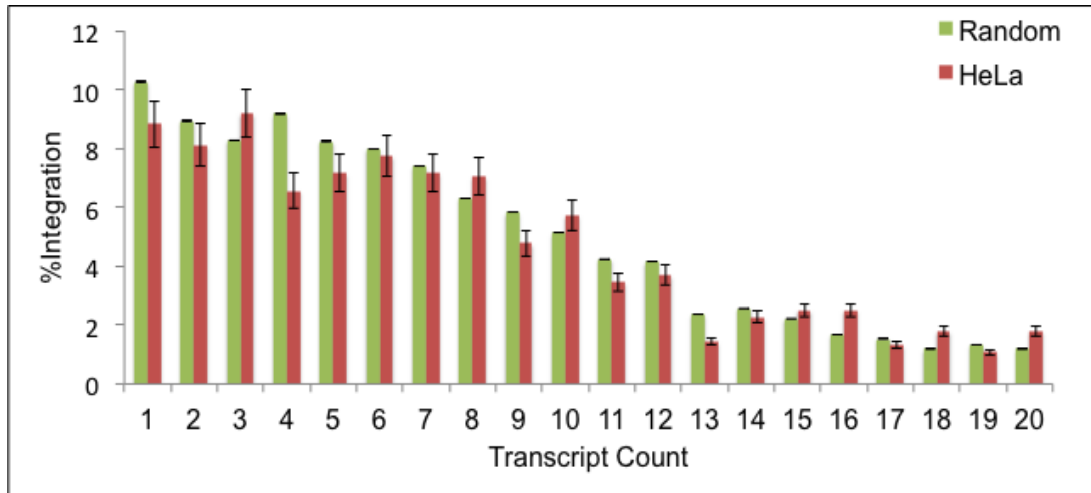


Figure 2.7 Correlation between ALV integrations and the number of alternative transcripts in human HeLa cells. Each transcription unit is assigned to a group based on the number of known transcripts that originate from it. Percentage of random and ALV integration events are plotted within each group for comparison.

Distribution of ALV integrations in B-cell lymphomas

In order to measure the relative abundance, or clonal expansion, of UISs within a tissue, we quantified the number of sonication breakpoints for each site, as described previously (Justice et al., 2015a). The number of breakpoints reveals the extent of clonal expansion of the UIS. In tissue culture we identified 16,978 unique sonication breakpoints, i.e. an average of 1.1 breakpoints per integration site (**Figure 2.8**). The vast majority of these integrations (82.9%) had a single breakpoint, suggesting that these integrations were not clonally expanded. On the other hand, in tumors we identified 92,951 unique sonication breakpoints (**Figure 2.9**). The average number of breakpoints per integration was 1.3, with the vast majority of integrations (67.6%) showing only a single sonication breakpoint. In contrast to 17.1% of integrations in tissue culture, 32.3% of integrations in tumors had two or more breakpoints, revealing that a large fraction of the infected cells are from expanded clones. Moreover, a large number of integrations (approximately 13,000) in tumors had very high number (10 or more) of breakpoints.

The distribution of viral integration sites in the neoplasms is depicted in a composite pie chart (**Figure 2.9**). The most highly expanded clones, each with 70 or more breakpoints, are highlighted in a table. The table depicts 28 UISs, indicating the gene most proximal to the integration, respective tumor, and its corresponding number of breakpoints. The maximum observable number of breakpoints is limited by the length of deep

sequencing reads and, in the case of highly abundant integration sites, the probability of repeated sonication at the same genomic position. Thus, it is important to note that our standard breakpoint analysis is an underestimate of the fraction of the infected cells with expanded clones (Gillet et al., 2011).

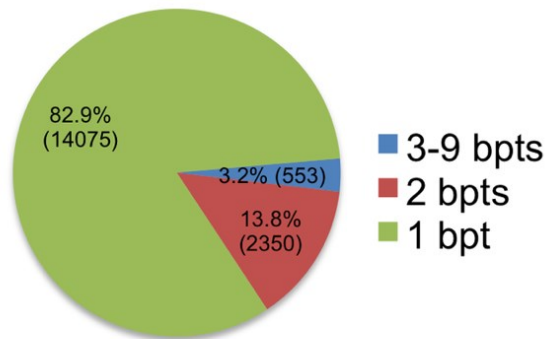


Figure 2.8 Distribution of integration sites in ALV infected tissue culture

(A) The pie chart depicts a total of 16,978 breakpoints that were identified in tissue culture (CEFs, DT-40s and HeLa cells). The single breakpoints (82.9%) and the expanded breakpoints (17.1%) are highlighted as separate slices of the pie.

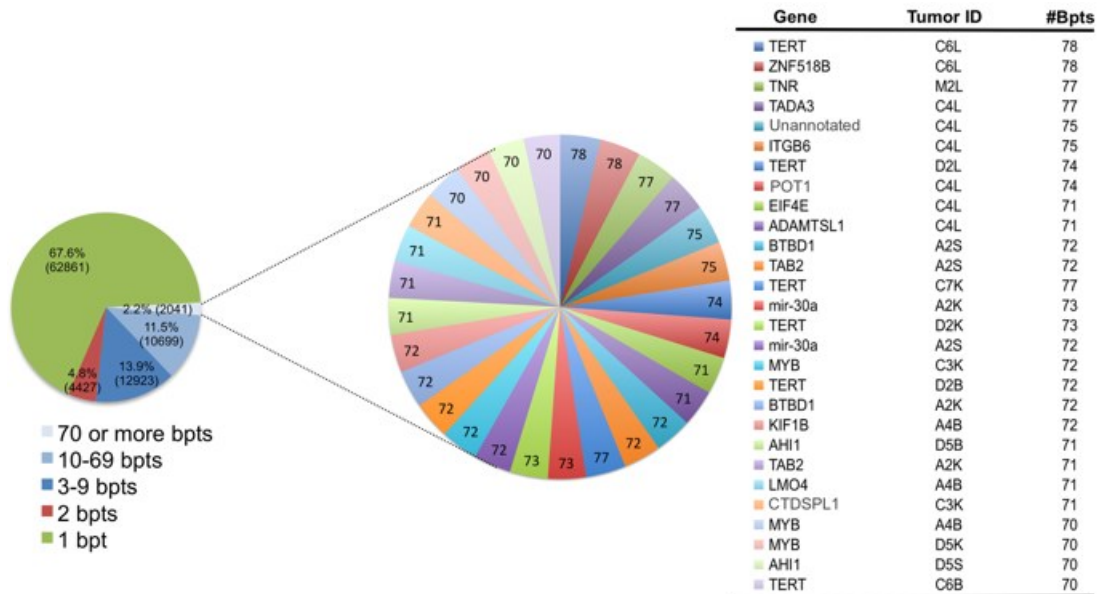


Figure 2.9 Distribution of integration sites in ALV-induced B-cell lymphomas, and clonal expansion in metastases. A total of 92,951 breakpoints were identified in 69 different ALV-induced neoplasms, from 41 different birds. Genes in proximity to integrations with 70 or more breakpoints are highlighted in a separate pie. Each slice represents a unique integration site, and the size of the slice represents the number of breakpoints for that site. The table shows integrations with the highest number of breakpoints along with the tumor in which that integration was identified. The tumor ID defines the bird number (A1, B1, C2, D2 etc.) along with the respective tissue of bursa (B), liver (L), kidney (K) or spleen (S) harboring the integration.

ALV integrations are most clonal in metastases

Cancers exist in a number of stages, characterized by a spectrum of divergent cells and genetic changes. Each cancer is unique, and in any given tumor the clonal structure shifts over time, which involves the clonal selection and expansion of cells. (Greaves and Maley, 2012). In cases of ALV-induced B-cell lymphomas, the bursa serves as the primary organ of malignant transformation and site of tumorigenesis (Baba and Humphries, 1985; Neiman et al., 2003). Infected chickens typically develop multiple primary neoplastic follicles in the bursa, some of which may eventually form primary tumors. The development of these neoplasms is a multi-stage process (Baba and Humphries, 1985; Cooper et al., 1968; Neiman et al., 1980a). In order to examine the clonality of lymphomagenesis, we studied different stages of cancer progression. These include inflammation, neoplastic follicles, primary tumors in the bursa, and metastatic tumors at secondary sites. The stage of neoplastic follicles in the bursa are part of tumor progression towards transformation and malignancy (Baba and Humphries, 1985). Metastases of primary tumors are observed from bursa to secondary sites of liver, spleen, and kidneys.

Analyzing the different neoplasms, we observed an increasing extent of clonality with the advancing stages of tumorigenesis. This clonal expansion can be represented by a pie chart, where each pie represents an individual tumor. A given slice of a pie represents a UIS, and the size of the

slice corresponds to its relative clonal abundance, denoted by the respective number of sonication breakpoints for that UIS.

The most clonally expanded integrations in secondary tumors can be investigated by comparison of metastasized versus non-metastasized neoplasms within an individual bird (**Figure 2.10**). For example, the different tumors in bird A2, all harbor overlapping UISs with an increasing level of clonal expansion in the metastasized neoplasms. For example, the expanded clones of UISs at *TAB2*, *BTBD1* and *mir-30a* integrations appear in a mix of other clonally expanded integrations in the primary bursa tumor, where as the secondary liver, kidney and spleen tumors are more homogenous. This suggests increasing clonal homogeneity of the metastasized tumors relative to the bursa.

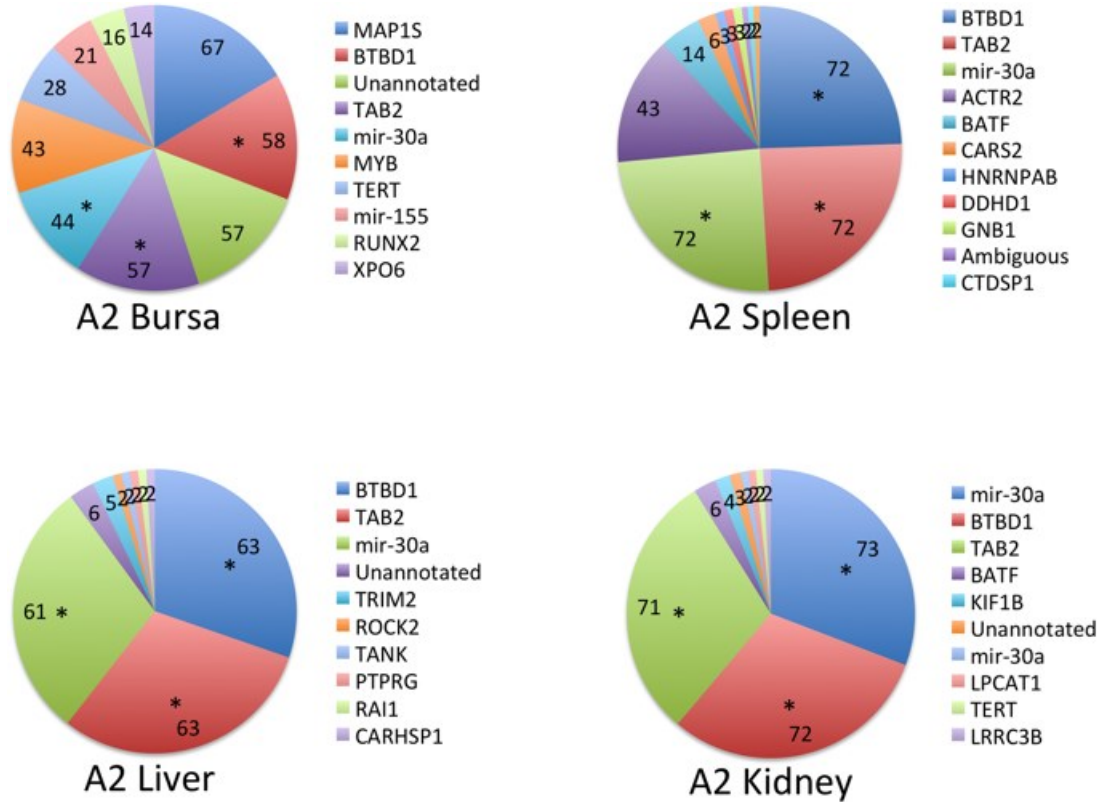


Figure 2.10 A sample set of individual primary (bursa) and secondary (liver, spleen and kidney) tumors from the same bird are illustrated. The top 10 most clonally abundant UIs are depicted within each pie chart, with the corresponding number of sonication breakpoints indicated on each slice. The list of the most proximal host genes is denoted next to the pie chart. UIs observed in multiple tumors of the same bird are denoted with an “*”. While, the primary tumor in the bursa exhibits heterogeneous distribution for clonally expanded UIs, the metastasized neoplasms exhibit a more homogenous distribution of expanded clones.

ALV clonality increases with progressing stages of lymphomagenesis

In order to empirically estimate the clonality of different neoplasms we made use of an objective parameter called the oligoclonality index (OCI) as defined by the Gini co-efficient index (Cook et al., 2014; Gillet et al., 2011). The OCI defines the clonal abundance of a tissue on an objective scale of 0 to 1. In theory, a tissue with perfect monoclonality would have an OCI value of 1. Conversely, an entirely polyclonal tissue would have an OCI value of 0. The OCI values of the neoplastic subtypes, in representative stages of lymphomagenesis, are depicted in **Figure 2.11**. The plot represents the OCI values in ascending order with tumor progression, suggesting an increased magnitude of clonal expansion within these neoplasms. As was previously depicted by the pie charts of bursal tumors, OCI value is lower in these samples compared to metastases. The OCI for the metastasized tumors was significantly greater, in some cases close to 1.

Additionally, in order to further validate the OCI values, we investigated the ALV integrations and their extent of clonal expansion in different slices of representative neoplasms (**Figure 2.12**). A sample with high clonal homogeneity would exhibit a highly uniform pattern of integrations and corresponding clonal abundance in different slices of the tumor, as depicted for tumor D2L. Conversely, a sample with lower OCI value such as a bursa tumor exhibited more clonal heterogeneity in different portions of the neoplasm (**Figure 2.13**).

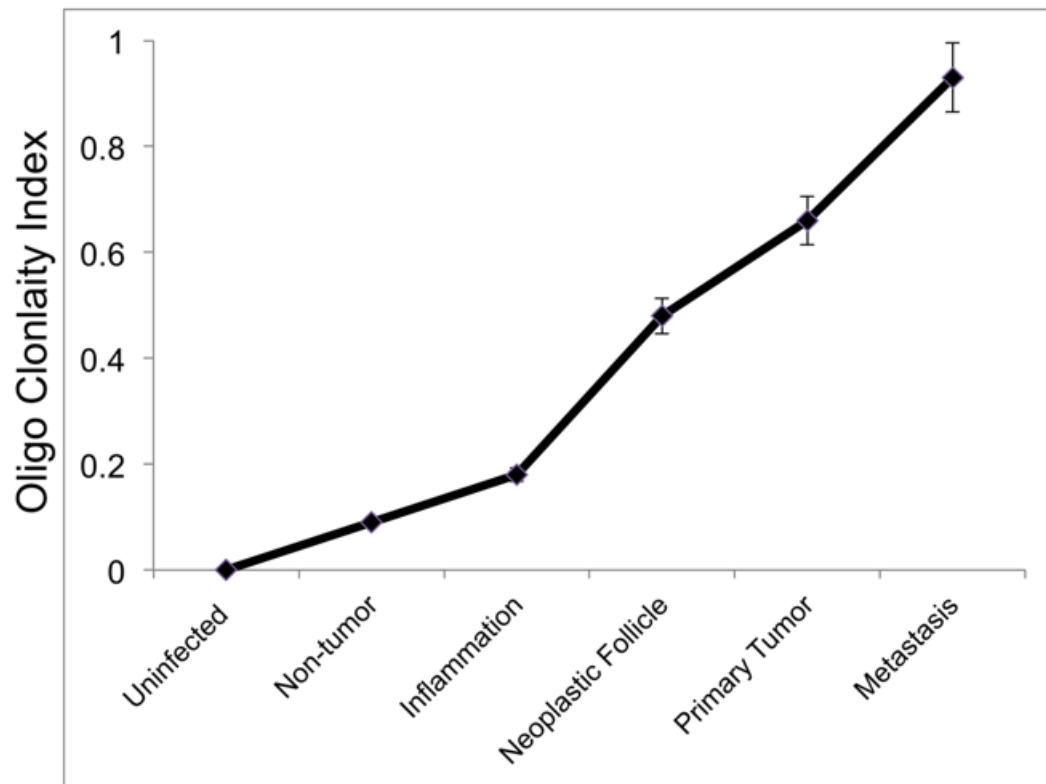


Figure 2.11 The OCI values of the different tissue samples are plotted, which provide an empirical estimate of homogeneity. The OCI values range between 0.03 to 0.96. The increasing OCI values for later stages of tumorigenesis suggest increasing tissue homogeneity.

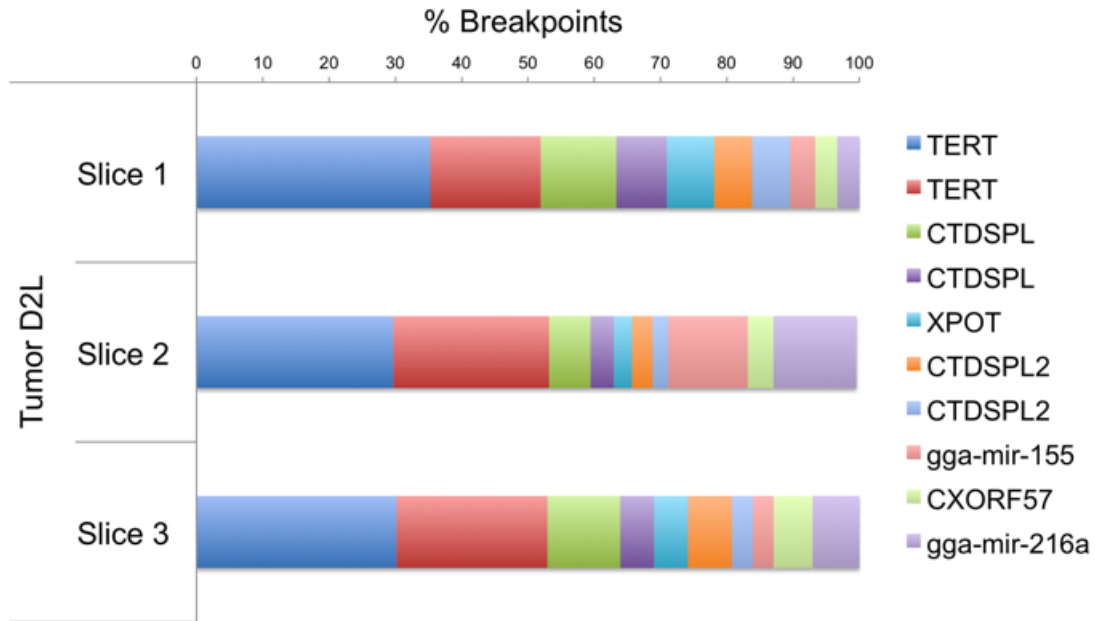


Figure 2.12 Top 10 most clonally expanded integrations for metastasized liver tumor D2L, are illustrated. Individual bars represent UISs and the corresponding extent of clonal expansion (as breakpoints) from 3 different parts of the same neoplasm. Slices were chosen randomly from three distinct portions of tumor mass, including peripheral and interior regions.

C2B

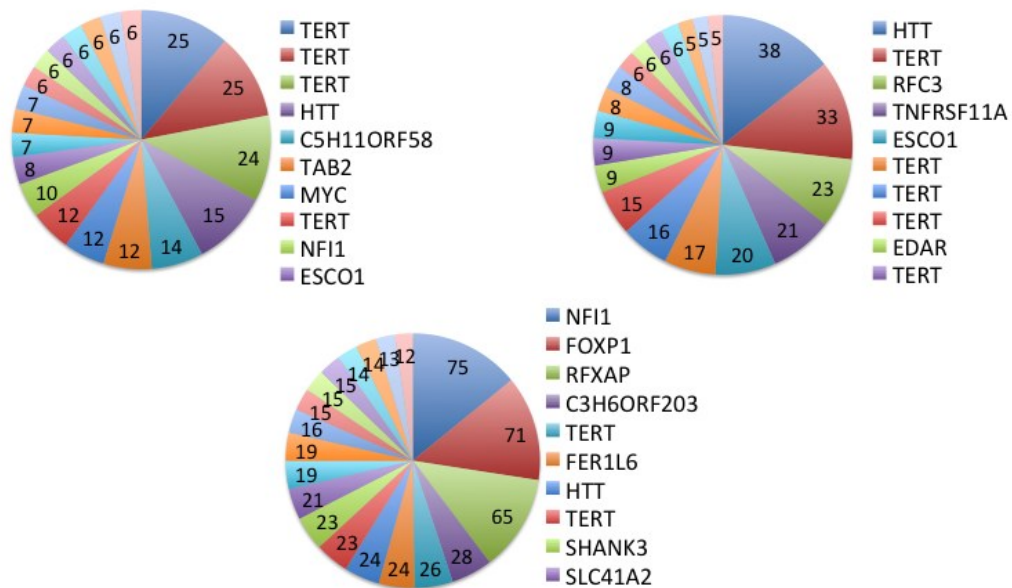


Figure 2.13 Clonally expanded integrations from different slices of a tumor, suggest heterogeneity of bursa. Top 10 most clonally expanded integrations from different slices of primary tumor (C2B) are illustrated. Individual pie charts represent UISs and corresponding extent of clonal expansion (as breakpoints) from a different slice of the tissue. Slices were chosen randomly from three distinct portions of tumor mass, including peripheral and interior regions.

Proviral load (PVL) increases with the progressing stages of tumorigenesis

In addition to the extent of clonal expansion, we also wanted to investigate the average number of proviral integrations within tumors, defined as its proviral load (PVL). Twenty randomly selected tumors across different stages of lymphomagenesis were isolated for this analysis. The PVL varies widely between the tissues, ranging on an average from approximately 2.3 to 4 copies per infected cell of a tumor (**Figure 2.14**). We observed a correlation between the PVL and different stages of tumor progression, suggesting that a higher PVL is associated with later stages of tumorigenesis.

ALV *env* regions are more distinct and offer greater specificity to distinguish ALV proviral genomes from endogenous retrovirus genomes in the host chicken genome. However, proviruses undergo varying amounts of genome rearrangements and deletions during cellular transformation and oncogenesis, especially in the *env* region, probably to evade cellular immune surveillance. Therefore, we utilized LTR specific primers to estimate the PVL. We assume each provirus contains 2-LTRs therefore, the PVL values described here may be underestimates of the actual PVL among infected neoplasms due to the possibility of solo LTRs. Additional analyses of PVL estimates, via use of *gag* or *env* regions of ALV, also exhibit similar association of PVL with the progression of tumors (**Figure 2.15**).

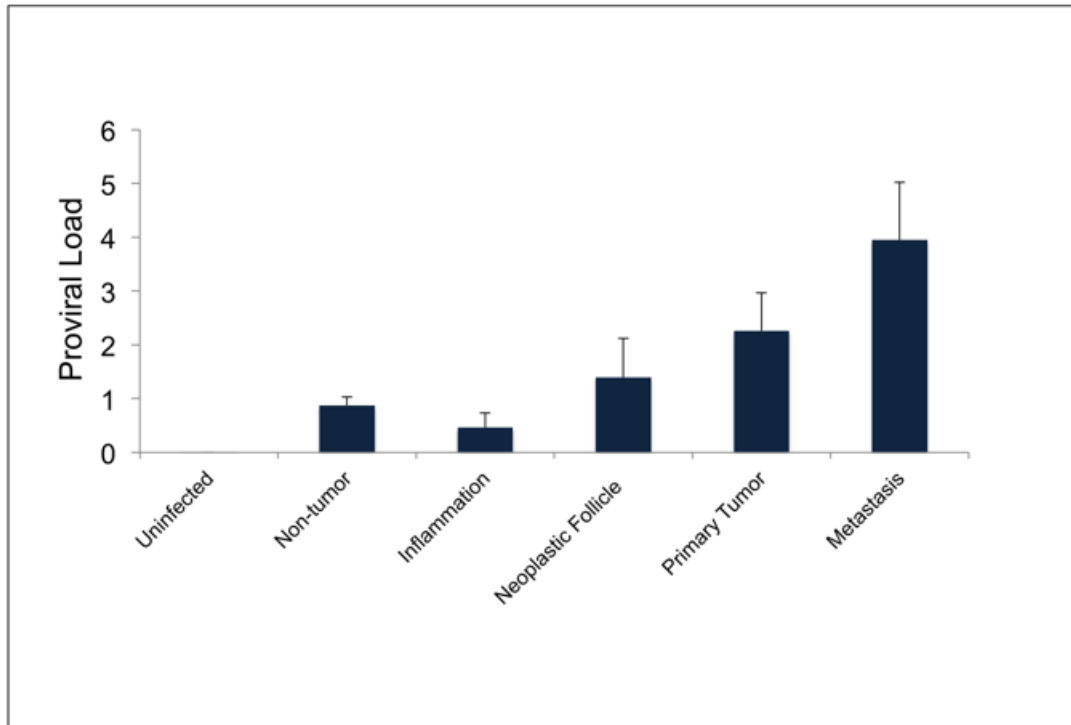


Figure 2.14 Proviral Load (PVL) ranges between different neoplasms.

PVL values are represented for different malignant and non-malignant tissues, including uninfected normal tissue as control. PVL ranges between 0.5 to 4 copies per cell, with an increasing trend for later stages of tumorigenesis.

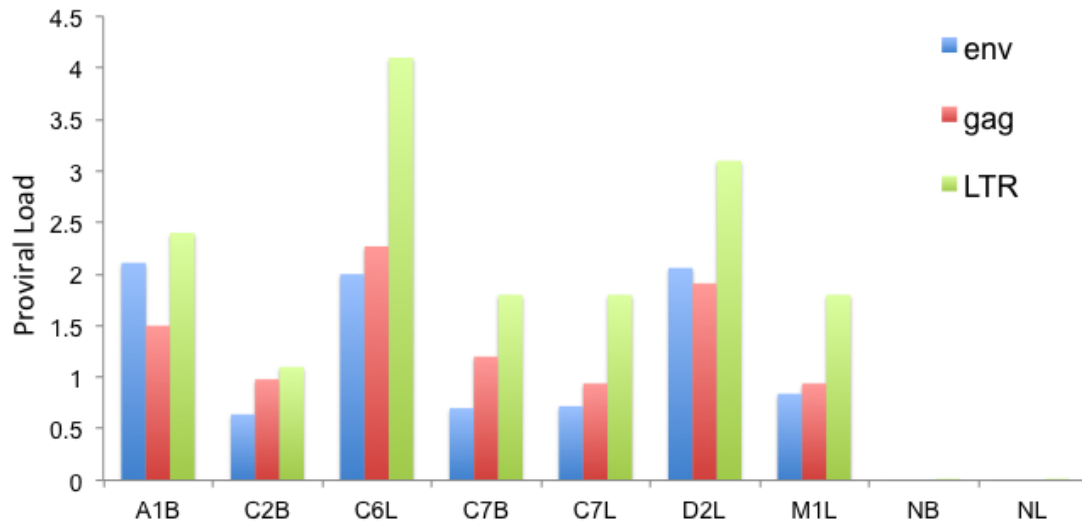


Figure 2.15 Proviral load values are calculated for different tissues, using the env, gag or LTR regions of ALV. Uninfected normal tissues of normal bursa (NB) and normal liver (NL) are depicted as controls.

Ontologies of genes near ALV integrations highlight biological processes associated with tumor progression

Cancers acquire, via mutational and epigenetic changes, a variety of traits that trigger clonal expansion, via proliferation, migration and invasion. These properties are alterations to normal developmental and physiological cellular processes (Greaves and Maley, 2012). We wanted to further investigate how ALV integrations near host genes, in certain functional categories, confer oncogenic advantages on the infected B-cell clone. To test this, we used the G:profiler analysis software to analyze the ontology of the nearest host gene upstream or downstream of each integration site. G:profiler allows functional profiling of genes within a neoplasm, in a quantitative fashion (Reimand et al., 2007). Queries are ordered with more importance given to the more expanded clones within tumors. The results showed a significant overrepresentation of genes in five cellular pathways: cell differentiation, phosphorylation, immune response signaling, proliferation and regulation of apoptosis (**Figure 2.16**). These biological processes show a greater enrichment in the secondary tumors (malignant) than in neoplasms with low or intermediate clonality.

These changes reflect the temporal order of genetic alterations and underlying cellular processes acquired in the progression of B-cell lymphomagenesis. These processes were associated with the host gene nearest to the viral integration sites, regardless of its transcriptional orientation relative to the provirus. Furthermore, when we analyzed the GO

terms associated with the most clonally expanded integrations in individual tumors, the above mentioned biological processes appeared repeatedly in many tumors. This suggests that the gene players involved in these biological processes exhibit a degree of cooperativity to trigger oncogenic transformation. For example, among the most clonally expanded UISs in **Figure 2.10** *TAB2* is known to be involved in immune response signaling, *BTBD1* plays a role in cellular differentiation and *mir-30a* has known roles in regulating cell proliferation, migration and invasion (Chen et al., 2017; Pisani et al., 2004; Zhu et al., 2012).

Furthermore, we identified *TERT* and *MYB* among the most clonally expanded integrations in 15 independent tumors from 9 different birds, and in 15 independent tumors from 11 different birds, respectively. *TERT*, the catalytic component of telomerase, has known roles in immortalization, senescence and apoptotic signaling (Koh et al., 2015). *MYB*, a transcription factor, functions in regulation of cellular differentiation and proliferation (Sandberg et al., 2005). Clonally expanded *MYB* integrations co-occur in half of the birds with *TERT* tumors. Additionally, we also observed up regulated *MYB* expression in *TERT* tumors without any ALV integrations near or within *MYB* ((Yang et al., 2007a) and data not shown). This suggests a possible cooperation between *MYB* and *TERT*. Genes proximal to other clonally expanded integrations, which occur in the *TERT* or *MYB* tumors, might also cooperate with them for inducing oncogenic transformation. Of particular interest, integrations in *CTDSPL* and *CTDSPL2* are frequently clonally

expanded, along with *TERT* and *MYB* (Winans et al., 2017b). *miR-155* integrations were also frequently seen with *MYB* in tumors (**Table 2.3**). These putative cooperating gene players are involved in varying biological processes such as differentiation, proliferation, apoptosis, phosphorylation, immune response signaling, immortalization and DNA damage repair (**Table 2.3**).

In order to determine common pathways activated in multiple individual tumors, we also analyzed genes near the clonally expanded integrations within individual tumors. A number of common transcription factor target gene networks were identified as common integration sites in various tumors. Among the common targets of ALV integration, the most enriched are the genes targeted by the E2F, EGR, WT1 and SP families of transcription factors (**Table 2.4**). E2F is a well-characterized protein family that mediates both cell proliferation and apoptosis (Polager and Ginsberg, 2009). EGR (early growth response) is a family of nuclear proteins that function as transcriptional regulators and target genes required for regulating differentiation and mitogenesis (Sukhatme et al., 1988). The SP (specificity protein) and WT (Wilms tumor) family of transcription factors are involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, and response to DNA damage (Black et al., 2001; Deniaud et al., 2006; Dupuis-Maurin et al., 2015). This suggests that the ALV integrations in these genes are at the intersection of events of tumorigenic

transformation. These gene players might cooperate to trigger the oncogenic characteristics, thus resulting in tumor formation.

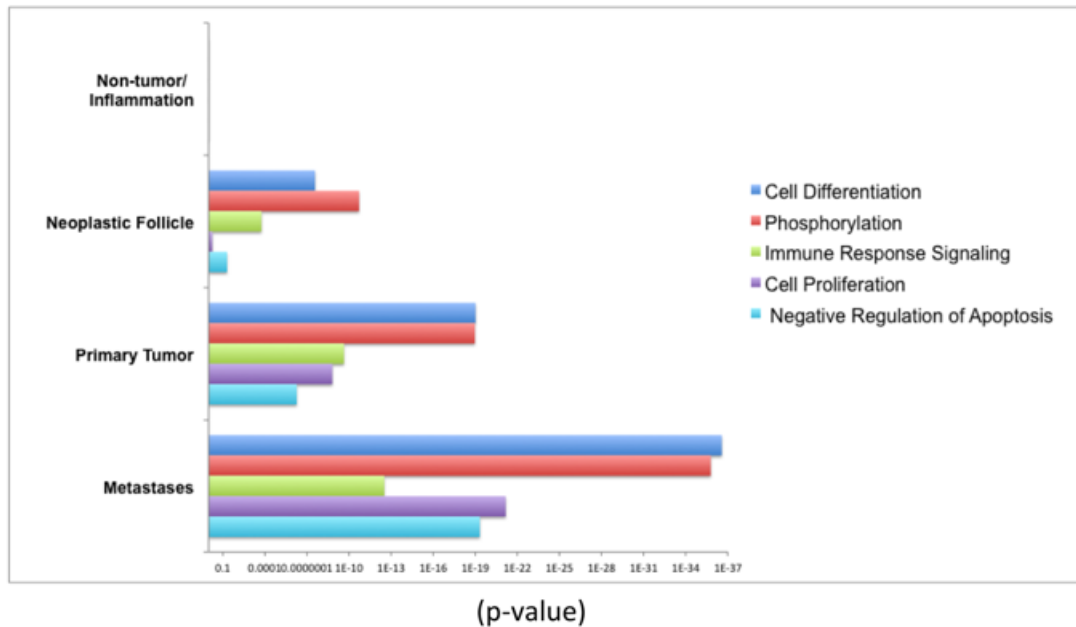


Figure 2.16 Functional classification of gene ontologies overrepresented among the clonally expanded integrations. Five functional categories that are significantly overrepresented among the neoplastic follicles, primary tumors and metastases, are illustrated as a bar graph. The horizontal bars represent statistical significance as defined by $-\log(p\text{-values})$. The integrations are analyzed by g:profiler software using a random list of integrations ($n=5000$) as background. Since there are no overrepresented pathways involving the integration sites in inflammation or non-tumor samples, the bars are not visible. Due to use of varying sample size, all the analysis was adjusted for multiple testing with unequal sample size using the Mann Whitney U test.

Pathway	TERT Cooperating players (14 tumors, 8 birds)	MYB Cooperating players (11 tumors, 7 birds)
Differentiation	MYB (A1,B4,A5,C7), SELIL (B4), HTT (C2), ZNF518B (C6), AHI1 (A5)	LMO4 (A4), AHI1 (A5,D5), SELIL (B4), NR3C1 (D5)
Phosphorylation	SMEK1 (C6), CCNA2 (A1), TAB2 (C2), AKT1 (C2)	CCNA2 (A1), DCAKD (E5)
Proliferation	MYB (A1,B4,A5,C7), EGFR (C6), AKT1 (C2)	NR3C1 (D5), HELLS (E5)
Cell survival/apoptosis	EGFR (C6), CTDSPL2 (D2,E3), AKT1 (C2)	TERT (A1,A5,B4,C7), KIF1B (A4), MYC (215), mir-155 (E1,E2,E4)
DNA Repair	CCNA2 (A1), SMEK1 (C6)	CCNA2 (A1), HELLS (E5)
Immune Response	TPN (A1), TAB2 (C2),	TPN (A1), FAM19A2 (D5)
Immortalization	CTDSPL (D2)	TERT (A1,A5,B4,C7,E2), CTDSPL (C3,D5,E2)

Table 2.3 Putative cooperating gene players that might cooperate with TERT and MYB in tumorigenesis.

Transcription Factor Family	#Tumors with enrichment
E2F	20
EGR	20
ETF	13
SP	12
WT1	12
RNF	10
AP	9
ZF	8
KROX	7
MOVO	2

Table 2.4 List of gene players targeted by the different families of transcription factors that are enriched in tumors.

Discussion

We report the analysis of more than 85,000 UISs leading from a stepwise early infection (in tissue culture) through to early and late tumor development. With only slight preferences for some genome features, we show that ALV exhibits a relatively random integration pattern compared to other well-studied retroviruses. Due to this relatively minimal discrimination, ALV serves as a good insertional mutagenesis tool to study tumorigenesis. We utilize the OCI to empirically define magnitude of clonality in different stages of tumorigenesis. Consistent with the clonal expansion hypothesis, we observe that ALV clonality increase with progressing stages of tumorigenesis (Greaves and Maley, 2012). We also identify putative cooperating gene players (especially *TERT* and *MYB*) and the underlying biological processes of cell differentiation, phosphorylation, immune response signaling, proliferation and regulation of apoptosis involved in tumor progression.

ALV displays a preference for highly spliced and expressed transcriptional units, and for proximity to the 5' end of a gene, most prominently around both sides of TSSs. Similar trends in the integration sites of different ALV subgroups were observed via an independent analysis in CEF, DT-40 and HeLa cells, suggesting that the ALV integration preference is not cell type specific. We observe a further enrichment of integrations near the TSSs in tumors, as a consequence of tumorigenic selection. ALV integrations are also enriched within CpG islands as well as near spliced and

expressed genes. There is a significant selection of ALV integrations away from CpG islands in the highly clonally expanded cells (5 or more breakpoints). Since DNA methylation is often observed in CpG islands, ALV integrations near CpG islands may be more susceptible to repression by methylation (Moore et al., 2013). Thus, cancer cells are likely enriched for integrations away from CpG islands, that are more likely to remain transcriptionally active.

We also report an enrichment of ALV integrations within gene bodies as nearly 40% of integrations are found in genes relative to 27% at random integration sites. Barr et al. (2005) reported a similar bias for ALV integrations into transcriptional units relative to matched random sites (Barr et al., 2005). Barr et. al 2005 also reported that ALV favors integrations in transcriptional units with higher expression levels (Barr et al., 2005). However, our data does not support these observations. These differences could be explained by our use of different methods to measure gene expression. Since we use RNA-seq data, in lieu of microarrays used previously (via 249 probe sets), our expression values might be different for the host genome. Moreover, we analyzed 15,416 proviral integration sites in this study compared to their analysis of 658 integrations, which should offer a more comprehensive analysis.

We observed a semi-random integration pattern of ALV in contrast to the integration patterns of other retroviruses. MLV and FV exhibit a strong preference for integrations within TSS or CpG islands (Sharma et al., 2013;

Trobridge et al., 2006; Wu et al., 2003). HIV-1 on the other hand, has a strong preference for integrating into transcriptional units with higher expression levels (Barr et al., 2005; Mitchell et al., 2004). ALV on the other hand shows a more random distribution of integration sites, similar to HTLV and MMTV (Derse et al., 2007; Faschinger et al., 2008).

In an emerging picture of B-cell malignancy, understanding tumor progression is an important piece of the puzzle. Here, we show that clonal expansion of ALV-infected B-cells is a key feature of malignant transformation in tumors. Approximately 100 to 500 UISs have been observed for HIV in peripheral blood lymphocytes (Laskey et al., 2016; Maldarelli et al., 2014). On the other hand, nearly 500-5000 UISs have been observed for a typical HTLV-1 host (Gillet et al., 2011). We observe approximately 700 to 3000 UISs in individual tumors induced by ALV mutagenesis. The most clonally expanded viral integrations appear to be early events in tumorigenesis and are expanded during progression of tumors. Therefore, this pattern of selection and expansion defines the clonal evolution of this cancer.

The distribution of the clone abundances can be quantified by an OCI value. Late stage B-cell neoplasms are associated with higher OCI values than non-tumors, and the PVL is also observed to correlate with the progressing stages of tumorigenesis. Interestingly, while the *gag* and *env* ratios appear very similar, LTR ratios are elevated for some individual tumor samples. This suggests that over the course of tumorigenesis, there are

likely more deletions and rearrangements acquired in the *gag* and *env* regions of the viral genome. Further work will be necessary to identify the factors that may influence proviral expression and genome modifications, such as DNA methylation, in tumorigenesis.

We observed a correlation between the PVL and different stages of tumor progression. As determined by the PVL analysis, a single cell in a tumor has multiple (2.3 to 4) copies of ALV. Since multiple UISs might contribute to oncogenic transformation, a higher PVL is associated with later stages of tumorigenesis. Therefore, loss of super-infection resistance could be involved in tumorigenesis. Alternately, deletions in the *env* region of proviruses, identified from previous work in our lab, might allow overcoming of super-infection resistance (Nehyba et al., 2016).

Analysis of the ontology of genes flanking integrations demonstrated a functional overrepresentation of certain known and other novel genes that are deregulated in many lymphomas (Justice et al., 2015b). Consistent with present concepts of oncogenesis and lymphomagenesis, GO analysis revealed that five major gene functions contribute to clonal dominance: regulation of proliferation, differentiation, immune response, apoptosis, and phosphorylation. Of these, cell differentiation and phosphorylation appear to be significantly altered in earlier stages of tumor progression. Interestingly, we observed possible cooperativity between *TERT* and *MYB*, which might function together to induce oncogenic transformation. Further analysis via single cell sequencing would be useful to investigate this cooperativity.

Studying ALV integrations, our work depicts a comprehensive investigation into the role of ALV-induced lymphomas in chickens. The value of our work can be extended to mammalian systems. B-cell development in chicken and mammals is a very similar process (Kohonen et al., 2007). These similarities are evident at levels of molecular changes and gene regulatory networks (Koskela et al., 2003; Weill et al., 2004; Wu et al., 2004). Although mice serve as a good mammalian model, in terms of oncogenesis they differ in some fundamental ways from humans. Unlike humans, the mouse telomerase enzyme is active in normal somatic cells (Hackett and Greider, 2002). This discrepancy between humans and mice is important because telomerase activation is a critical step in the human oncogenic process, with telomerase activation seen in approximately 90% of human cancers (Garcia et al., 2007a; Shay and Wright, 2011). Similar to human expression, chicken telomerase expression is down regulated in most somatic tissues (Delany and Daniels, 2004). Furthermore, chicken telomeres shorten with age, and telomerase activity is important for oncogenesis (Delany et al., 2000). Therefore, chicken serves as an advantageous model over mouse, to study oncogenic events. Since, common and conserved mechanisms are fundamental, study of genetic alterations in chickens as a model organism can help unravel the way human systems are regulated in B-cell lymphomagenesis.

Limited information is available about the molecular mechanisms of lymphomagenesis, and the role of selective clonal expansion. Cells containing certain integration sites can undergo selective expansion in

tumors, resulting in abundant clonal populations. We observed that in the course of tumor progression, the more transformed neoplasms contained integrations with a high number of breakpoints (Greaves and Maley, 2012). Via our genomic analysis of ALV integrations across progression of B-cell lymphomas, we are able to provide insights into the biological processes associated with initiation, progression, and metastasis of tumors.

Materials and Methods

Tissue culture infections

Chicken embryo fibroblasts (CEFs) were cultured in medium 199 (Thermo Fisher Scientific) supplemented with 2% tryptose phosphate, 1% fetal calf serum, 1% chicken serum, and 1% antibiotic at 39°C and 5% CO₂. Viruses were generated by transfecting CEFs via electroporation, with vectors RCASBP(A) and RCASBP(C) to generate viral titers of subgroups A and C respectively (Hughes, 2004). ALV-J virus (Malhotra et al., 2015) was generated from homogenates of tumors with ALV-J integrations, by passing it through a 0.22 micrometer pore size filter. The collected supernatant from tumors was in turn used to infect CEFs. CEFs were grown at approximately 40% confluency and were infected with ALV subgroup A, C or J at an MOI of 1-2. The cells were collected at 48 hours and 120 hours post infection for DNA isolation. DT-40 cells were cultured in Dulbecco's modified eagle medium (Thermo Fisher Scientific), 10% fetal calf serum, 5% chicken serum, 5% tryptose phosphate, and 1% antibiotic at 37°C and 5% CO₂. DT-40s were grown at approximately 40% confluency and were infected with ALV subgroup C at an MOI of 1-2. The cells were collected at 48 hours post infection for DNA isolation. HeLa cells (ATCC) were cultured in Dulbecco's modified eagle medium, 10% fetal bovine serum (FBS), and 1% antibiotic at 37°C and 5% CO₂. To generate ALV pseudo-typed with vesicular stomatitis virus glycoprotein (VSV-G), CEFs were co-transfected via electroporation, with pMD.G (VSV-G envelope plasmid) and RCASBP(C) plasmid (Burns et

al., 1993). Viral supernatant was collected after 48 h, filtered through a 0.22-micrometer filter, and concentrated by polyethylene glycol (PEG) precipitation (10% PEG8000) (Cepko, 2001). This concentrate of pseudo-typed ALV was used to infect HeLa cells and cells were collected 48 hours post infection for DNA isolation.

Tumor induction and tissue isolation

5 or 10-day-old chicken embryos were injected with ALV-LR9, ALV- Δ LR9, ALV-G919A, or ALV-U916A as described previously (Justice et al., 2015b; Polony et al., 2003). Chickens were observed daily and were euthanized when apparently ill or at 10-12 weeks. A total of 69 tissues were selected for characterization by high-throughput sequencing (**Supplementary Table 2**). Three uninfected tissues and several non-tumor tissues from infected birds were sequenced to serve as controls. All of the B-cell lymphomas included in the study were rapid-onset lymphomas, arising within 10-12 weeks. LR-9 is an ALV subgroup A recombinant virus consisting of *gag*, *pol*, and *env* genes derived from UR2-associated virus and LTRs derived from ring-necked pheasant virus (Simon et al., 1987). ALV- Δ LR-9 contains a deletion in the *gag* gene, causing increased splicing to downstream genes (Smith et al., 1997). ALV-G919A contains a silent mutation in the NRS (Polony et al., 2003). Tumors were collected from primary bursal (B) tissue or metastasized liver (L), kidney (K) or spleen (S) tissues.

Ethics statement

Five- and ten-day old chicken embryos were injected with virus. Chickens injected include inbred SC White Leghorn line embryos (Hy-Line International, Dallas Center, IA), and SPAFAS embryos (Charles River). Chickens were euthanized at 10-12 weeks post infection. Institutional Animal Care and Use Committee (IACUC) approval was obtained at the University of Delaware and Fred Hutchinson Cancer Research Center.

Integration site mapping and quantification

DNA from ALV infected cultured cells or tumor samples were isolated. The sequencing libraries were prepared as described previously (Justice et al., 2015a). Five micrograms of purified genomic DNA was sonicated with a Bioruptor UCD-200. End repair, A-tailing, and adapter ligation were performed as described previously (Gillet et al., 2011) (adapter short arm, P-GATCGGAAGAGCAAAAAAAAAAAAAAAAAA, and adapter long arm, CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T, where “X’s” denote the barcode sequence, “P” denotes phosphorylation, and “*” denotes a phosphorothioate bond). Nested PCR was performed to enrich the library for proviral junctions. The first PCR was 23 cycles and employed an ALV-specific primer (CGCGAGGAGCGTAAGAAATTTTCAGG) between the 3’ LTR and *env* and a primer (CAAGCAGAAGACGGCATACGAGAT) within the adapter that was attached by ligation in the previous step. In the second round of PCR, a

primer

(AATGATACGGCGACCAACCGAGATCTACACTCGACGACTACGAGCACATGCATGAAG) near the 3' end of the LTR was used. This primer ended 12 nucleotides short of the junction between viral and genomic DNA. This primer was paired with an adapter-specific primer on the opposite side of the fragment, which overlapped the adaptor's bar code sequence (CAAGCAGAAGACGGCATAACGAGATXXXXXX). Libraries were quantified by quantitative PCR (qPCR) and then underwent single-end 100-bp multiplexed sequencing on the Illumina Hi-Seq 2000. A custom sequencing primer (ACGACTACGAGCACATGCATGAAGCAGAAGG) was used which hybridized near the end of the viral 3' LTR, 5 nucleotides short of the proviral/genomic DNA junction. The resulting reads could be validated as genuine integrations by verifying that they began with the last 5 nucleotides of the proviral DNA, CTTCA. The last two nucleotides of the unintegrated proviral DNA, TT, are cleaved by ALV integrase upon integration, so the lack of these 2 nucleotides in the read acted as further validation of a true viral integration.

Sequencing Analysis

Reads were first curated with a custom python script to remove sequences that did not begin with the last five nucleotides of viral DNA, "CTTCA" (Justice et al., 2015a, 2015b). The files were then uploaded to Galaxy (Blankenberg et al., 2010a; Giardine et al., 2005; Goecks et al., 2010), which

was used to perform downstream analyses. In Galaxy, first the quality scores were converted to Sanger format with FastQ Groomer v1.0.4 (Blankenberg et al., 2010b). CTTCA and adapter sequences were then trimmed using the Galaxy Clip tool v1.0.1. This tool also removed reads containing an N and reads less than 20 nucleotides in length after adapter removal. The remaining reads were mapped with bowtie [66] using the *Gallus gallus* 4.0 genome (Nov. 2011). Sequences were aligned using a seed length of 28 nucleotides, with a maximum of 2 mismatches permitted in the seed. All alignments for a read were suppressed if more than one reportable alignment existed. This was done to prevent multiple mapping and ensure that reads correspond to only unique integration sites. 100,000 random mapped reads were selected from each sample to be used for further analysis. If less than 100,000 reads were present for a sample, all available reads were used. A custom Perl pipeline developed in the lab was used to analyze the aligned reads output from bowtie (Justice et al., 2015a, 2015b). This custom pipeline identified unique integration locations, and calculated the number of reads and sonication breakpoints for each integration site. It also identified hotspots of integration and common integration sites among multiple samples. Integrations from two unrelated barcodes on the same sequencing lanes were omitted via our pipeline. The pipeline source code is available upon request. The integration sites identified in our work are deposited at the NCI Retrovirus Integration Database (RID) (<https://rid.ncifcrf.gov/>) (Shao et al., 2016).

Analysis of ALV integrations with respect to different genome annotations and features

Reads for the junctions of proviral integration and genomic DNA were mapped with Bowtie (Langmead et al., 2009a). Only reads that mapped uniquely to the genome were utilized for further analysis. This step filtered out reads that originate from repetitive elements. Mapped reads from all samples were then combined into a single file and analyzed with HOMER (Heinz et al., 2010). HOMER calculates the enriched features at each integration locus as well the proximity to closest transcription start site. A random integration control data set was generated with Bedtools Random (Quinlan and Hall, 2010). The genomic DNA sequences corresponding to the genomic coordinates obtained from Bedtools Random were extracted from the Gallus gallus 4 genome using the Galaxy tool Extract Genomic DNA. Control sequences were mapped with Bowtie and analyzed with HOMER using the same parameters as for ALV integrations. Proximity to CpG islands was determined using the WindowBed tool in Galaxy (Goecks et al., 2010). We note that our calculations are subject to certain biases. This includes, but is not limited to, an underestimate of the chicken or human genome sizes due to unsequenced gaps or overlapping sequences. Furthermore, an aberrant karyotype, which might exist in the transformed HeLa (Macville et al., 1999) or the DT-40 cells (Chang and Delany, 2004), was not taken into account for our analysis. However, as previously determined by Narezkina et al. (2004), despite the aberrant karyotype in HeLa cells, the ratio between

the genome size and the gene number in HeLa cells is equivalent to that of the normal human genome (Narezkina et al., 2004).

Analysis of ALV integrations with respect to gene expression and levels of splicing

The ensembl *Gallus gallus* 4 genome was utilized to obtain reference information for the transcript count and number of introns for all transcriptional units in the chicken genome. If an integration occurs within a gene, then the corresponding gene is used for all subsequent analysis. If an integration occurs in an intergenic region, then the nearest gene is used for all subsequent analysis. RNA-seq data for analysis of CEFs was downloaded from the public Sequence Read Archive (SRA) database (SRA accession no. SRP107761) (Leinonen et al., 2011). A custom Python script was utilized to associate the expression, transcript count and number of introns of a gene with the number of ALV integrations proximal to or within the given gene. A matched random control set, generated as mentioned above, was used as a control. The Python source code is available upon request.

Proviral Load quantification

PVL was measured by quantitative polymerase chain reaction (qPCR) of ALV-LR9 for *env* (primers CCTGAAACCCAGTGCATAAGG and CTAGCTGTGCAGTTCACCGT), *gag* (primers

GTTTAGAGAGGTTGCCCCGAC and GTCAATGATCACCGGAGCCC) and LTR (CGAACCACTGAATTCCGCAT and GAATCAACGGTCCGGCCATC); and *HMG14b* (primers ACTGAAGAGACAAACCAAGAGC and CCAGCTGTTTTAGACCAAAGAATAC) using Q SYBR green Supermix (Bio-Rad) according to the manufacturer's protocol on a Bio-Rad C1000 thermal cycler/CFX96 Real-Time System. We assumed a single copy of *env* and *gag* and 2 copies each of *HMG14b* and the LTR per cell. *HMG14b* is a known single copy gene in the chicken genome and thus, was used as a housekeeping reference gene (Srikantha et al., 1990). Thermal cycling conditions were 95°C for 20 seconds and 40 cycles each of 95°C for 1 second followed by 53°C for 30 seconds. Quantitative PCR (qPCR) was performed in duplicate, with each sample present in technical duplicate during each run. The results were normalized to those for normal bursa using the comparative threshold cycle (*CT*) method.

Statistical analysis for oligoclonality index calculations

Statistical analysis for clonality index was carried out using R version 2.15.2 (<http://www.R-project.org/>). The oligoclonality index (OCI; Gini coefficient) was calculated using the R package *sonicLength* (<http://soniclength.r-forge.r-project.org/>) as described previously (Berry et al., 2012; Gillet et al., 2011).

Gene Ontology Analysis

Functional profiling of genes and ontology analysis for the clonally expanded integrations was conducted with g:profiler, via using an ordered query option (<http://biit.cs.ut.ee/gprofiler/>) (Reimand et al., 2007).

**Chapter 3. The *MET* gene is a common integration target in ALV
subgroup J induced chicken hemangiomas**

Adapted from:

Malhotra S, Justice J 4th, Lee N, Li Y, Zavala G, Ruano M, Morgan R, Beemon K. (2015). Complete genome sequence of an American avian leukosis virus subgroup J isolate that causes hemangiomas and myeloid leukemia. *Genome Announc.* 9;3(2).

and

Justice J 4th, **Malhotra S**, Ruano M, Li Y, Zavala G, Lee N, Morgan R, Beemon K. (2015). The *MET* gene is a common integration target in avian leukosis virus subgroup J-induced chicken hemangiomas. *J Virol.* 89(9).

Summary

Avian leukosis virus, subgroup J (ALV-J), is a simple retrovirus that can cause hemangiomas and myeloid tumors in chickens and is currently a major economic problem in Asia. Here we characterize ALV-J strain PDRC-59831, a newly studied US isolate of ALV-J. Five-day-old chicken embryos were infected with this virus, and the chickens developed myeloid leukosis and hemangiomas within two months after hatching. To investigate the mechanism of pathogenesis, we employed high throughput sequencing to analyze proviral integration sites in these tumors. We found expanded clones with integrations in the *MET* gene in two of the five hemangiomas studied. This integration locus was not seen in earlier work characterizing ALV-J induced myeloid leukosis. *MET* is a known proto-oncogene that acts through a diverse set of signaling pathways and is involved in many neoplasms. We show that tumors harboring *MET* integrations exhibit strong overexpression of the *MET* mRNA. These data suggest that ALV-J induces oncogenesis by insertional mutagenesis, and integrations in the *MET* oncogene can drive overexpression of *MET* and contribute to the development of hemangiomas.

Additionally, we identify *MET* as the most clonally expanded integration event in an ALV-A induced kidney hemangioma (among tumors studied in Chapter 2). Thus, *MET* is observed as a clonally expanded integration site for different ALV subgroups, in different tumors from

independent studies. This therefore, suggests implications of *MET* in hemangiogenesis.

Introduction

Avian leukosis viruses (ALVs) are classified into subgroups based on their envelope gp85 surface glycoprotein (SU), viral cross-neutralization patterns, and host range. Most avian retroviruses are classified as subgroups A, B, C, D, or E. ALV of subgroup J (ALV-J) was first isolated in 1988 in the United Kingdom; the prototype strain HPRS-103 causes primarily myeloid leukosis, but can induce other tumor types at low incidence (Payne et al., 1992). ALV-J is thought to have originated from a recombination event between an exogenous ALV and an ancient endogenous avian (EAV) retroviral element (Bai et al., 1995; Smith et al., 1999). This recombination event incorporated the endogenous retroviral *env* into ALV-J.

Since its discovery in the UK, a variety of ALV-J strains have been characterized in diverse geographical areas, including North America, Europe, East Asia, Australia, and the Middle East (Benson et al., 1998; Fenton et al., 2005; Landman et al., 2002; Malkinson et al.; Sung et al.). It is believed that these isolates derive from a single common ancestor, and are not the result of independent recombination events (Benson et al., 1998). The types of neoplasms caused by ALV-J vary and can be influenced by the specific strain of ALV-J that has infected the bird. Most often, the virus induces tumors of myeloid origin (as with HPRS-103), but some strains induce primarily hemangiomas. These are vascular tumors found in the skin or visceral organs that originate from the endothelial cells that line blood vessels (Fadly and Nair, 2008). Other ALV-J strains are capable of inducing

both myeloid tumors and hemangiomas. We found that the strain used in this study falls into the third category, inducing both hemangiomas and myeloid tumors at high incidence.

Besides myeloid tumors and hemangiomas, ALV-J has been shown to induce other types of tumors at low frequency such as skeletal myelocytomas, renal tumors, histiocytic sarcomas, and others (Payne et al., 1992; Venugopal, 1999). This pathology contrasts sharply with that of the more studied ALV-A, which induces mainly B-cell lymphomas but also erythroblastomas (Kanter et al., 1988; Neiman et al., 1980b; Nilsen et al., 1985).

ALV-J infection can cause significant economic loss due to reduced egg production, stunted growth, and early death. The economic losses have been particularly extensive in China where the virus commonly infects poultry (Fadly and Smith, 1999; Gao et al., 2010; Malkinson et al., 2004; Sung et al., 2002). It has been recently shown that ALV-J infection is not limited to domesticated chickens. In fact, infection with ALV subgroups A, B, and J appears to be widespread in wild fowl throughout China (Jiang et al., 2014; Li et al., 2013).

ALVs do not carry a viral oncogene and instead cause neoplasia through insertional mutagenesis (Beemon and Rosenberg, 2012). In order to complete the viral life cycle, all retroviruses must integrate into the genomic DNA of the infected cell. Thus, the provirus can act as a mutagen, landing within a gene and ablating its function. Alternatively, because the virus has

potent enhancers and promoters in its long terminal repeats (LTRs), ALV can induce the expression of genes located near an insertion site. This process can drive tumor formation if the provirus integrates near and perturbs expression of cancer-related genes.

Virus-induced mutagenesis can be exploited to identify genes that may play a role in driving development of neoplasms. For example, the virus can be used to induce tumor formation, and common integration loci can be identified. These common integration sites (CISs) flag a genomic locus as potentially harboring an oncogene or tumor suppressor. Viral insertional mutagenesis screens have been fruitful in identifying cancer genes in several model systems (Beemon and Rosenberg, 2012). ALV is an especially useful virus for such a screen because it integrates in a largely random fashion, with only slight preference for active transcriptional units (Barr et al., 2005; Mitchell et al., 2004). This ensures that as many genomic loci as possible are probed for oncogenic potential by the virus.

Earlier studies implicated several genes as drivers of tumorigenesis in ALV-induced neoplasms. In ALV-A induced B-cell lymphomas, common integration sites were identified near or within *MYC*, *MYB*, *MIR-155*, and *TERT* genes (Clurman and Hayward, 1989; Hayward et al., 1981; Payne et al., 1982; Yang et al., 2007a). The study of proviral integrations within ALV-J-induced neoplasms has only recently begun. Early work has shown *MYC*, *TERT*, and *ZIC1* to be targets of proviral integration in ALV-J induced myeloid leukosis (ML) (Li et al., 2014).

In this study, we conducted an insertional mutagenesis screen to identify the genes involved in ALV-J induced tumors. To identify viral integration sites in these tumors, we employed high-throughput sequencing on the Illumina platform. We identified intron 1 of the *MET* gene as a common integration site in hemangiomas. *MET* encodes a well-studied receptor tyrosine kinase that binds hepatocyte growth factor/scatter factor and plays important roles in normal development and a wide range of human cancers (Graveel et al., 2013). Because we observed integrations near a known oncogene in multiple tumors, we hypothesize that ALV-J tumors, like those induced with ALV-A, are generated by insertional mutagenesis.

Furthermore, we identify *MET* as a clonally expanded integration event in an ALV-A induced kidney hemangioma. Together, our work shows that *MET* is a common integration hotspot in ALV induced tumors and therefore is implicated in development of hemangiomas.

Results

Characterization of an American ALV-J isolate PDRC-59831

We report the complete genome sequence of an avian leukosis virus subgroup J (ALV-J) isolate PDRC-59831, which causes myeloid leukosis and hemangiomas in chickens. This is an American ALV-J isolate, which was found in a 38-week-old broiler breeder chicken on a farm in Georgia in 2007. The first ALV-J strain, HPRS-103 was isolated in the UK in 1989 from meat-type chickens with myeloid leukosis (ML) (Payne et al., 1991). ALV-J has caused huge economic losses to the poultry industry worldwide (Fenton et al., 2005; Landman et al., 2002; Malkinson et al., 2004; Sung et al., 2002; Xu et al., 2004) resulting from drastically reduced egg production, stunted growth, bleeding tissues, and increased mortality (Cui et al., 2009; Gao et al., 2010). ALV-J typically induces ML tumors, but some strains primarily induce hemangiomas (Pan et al., 2011). The isolate sequenced in this study induces both hemangiomas and ML tumors at high incidence. This new ALV-J isolate, designated PDRC-59831 (*Poultry Diagnostic and Research Center*, Athens, Georgia) was isolated from a 38-week-old broiler breeder chicken.

We inoculated PDRC-59831 into five-day-old SPAFAS embryos (Charles River) via the yolk sac route. All experimental chickens were euthanized when ill or by 12 weeks of age. Seven ALV-J infected chickens survived and were all found to have hemangiomas, ML, or both (Justice et al., 2015a). The whole genome of PDRC-59831 was then amplified via PCR from genomic DNA of 4 different tissues from 3 infected birds. Sequence

assembly and multiple sequence alignment were done using SnapGene™ and ClustalOmega™.

Comparison of the PDRC-59831 sequence to the original English isolate, HPRS-103 (Accession no. Z46390), showed that the *gag*, *pol* and *env* sequences share a nucleotide identity of 97.2%, 97.7% and 95.0% respectively. In contrast, comparison to the other fully sequenced American isolate ADOL-7501 (Accession no. AY027920), which also induces ML and hemangiomas in chicken, showed homologies of 95.1%, 97.2% and 91.2% for these respective regions. We also observed the presence of open reading frames (uORFs) in the RNA leader sequence that are known to modulate viral gene expression (Donzé and Spahr, 1992; Moustakas et al., 1993). We found that in comparison to the three uORFs found in RSV (Accession no. J02342) and HPRS-103, PDRC-59831 has only two uORFs (uORF1 and uORF3).

Several studies reported certain genetic alterations in ALV-J strains that primarily induce hemangiomas in chickens. For example, an 11-nucleotide deletion was observed in the LTR-U3 region of ALV-J strains SCDY1 and NHH (Shi et al., 2011). Additionally, two different 19-nucleotide insertions in the 5' untranslated region (5'UTR) and in the U3 region were identified in hemangioma-inducing strains JL093-1, SD09DP03, and HLJ09MDJ-1 (Pan et al., 2011). None of these sequence alterations were observed in PDRC-59831. Instead, PDRC-59831 has more sequence similarity to the ML-inducing prototype strain HPRS-103 in these regions.

This suggests that these genetic alterations are not necessary to induce avian hemangiomas. Furthermore, a 205-nucleotide deletion in the 3' UTR that leads to higher oncogenicity and increased mortality in infected chickens has also been identified in some ALV-J strains (Wang et al., 2012).

Sequence alignment identified a similar deletion in PDRC-59831. This observed deletion of 216 nucleotides almost encompasses the previously described 205-nucleotide deletion.

Our work provides additional understanding of the variations in ALV-J genomes, which can help determine evolutionary relationships among viral populations. The complete genome sequence of the PDRC-59831 isolate was submitted to GenBank, and the assigned accession number is **KP284572**. The same viral isolate was used for characterizing ALV-J common integration sites in the chicken genome (10).

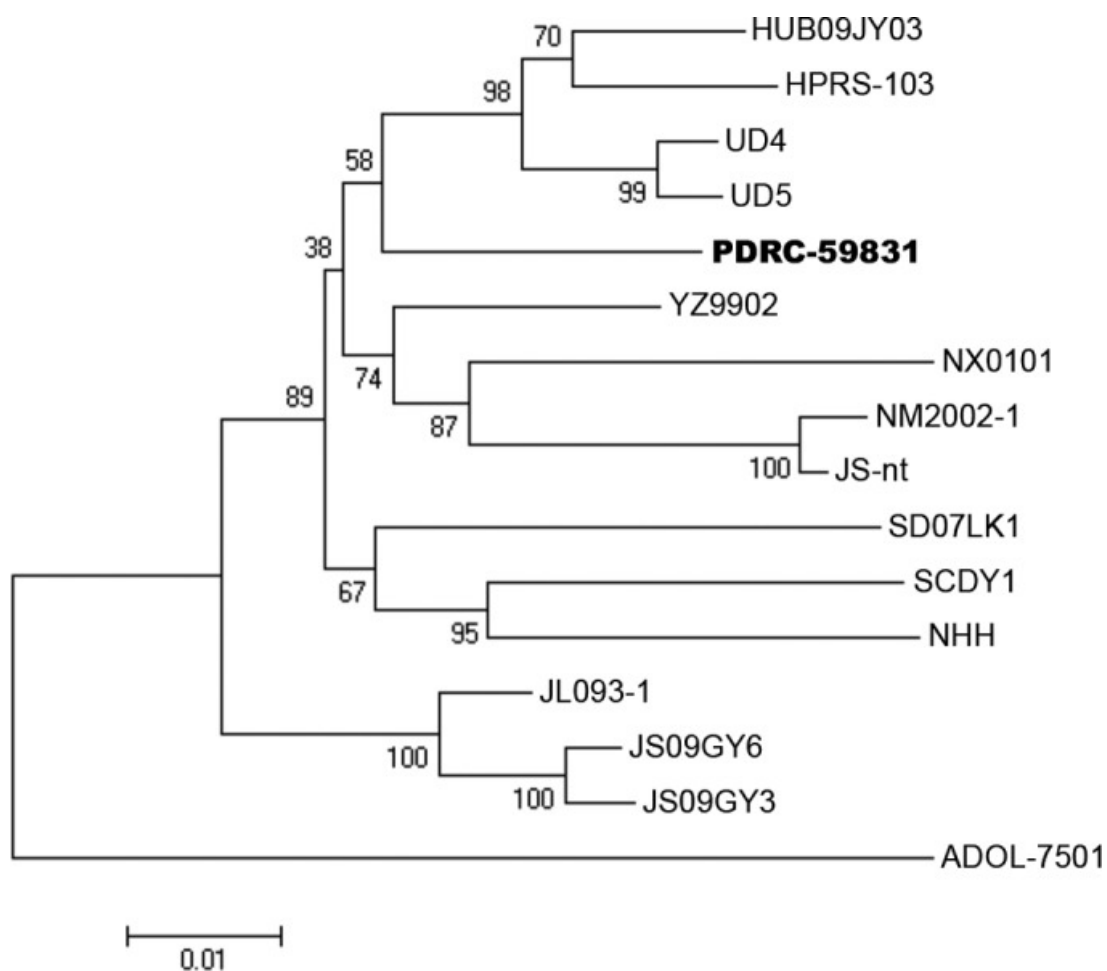


Figure 2.1 Sequence comparison and phylogenetic analysis of env genes of ALV-J isolates. Phylogenetic analysis was conducted by using the neighbor-joining method. The optimal tree, with a sum of branch length of 0.37621, is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Strain PDRC-59831 (in boldface type) is shown to cluster closest to UD4, UD5, and YZ9902 and on the same branch as prototype ALV-J isolate HPRS-103.

Sequencing of ALV-J/cellular integration junctions

All 7 birds infected with PDRC-59831 developed tumors (hemangiomas, myeloid tumors, or both) by twelve weeks of age. A total of six hemangiomas and four myeloid tumors were observed in total (**Table 3.1**). Genomic DNA was isolated from these tumors, randomly sheared by sonication, and proviral junctions were amplified in two successive rounds of nested PCR as outlined in **Figure 3.2**. Proviral integration/genomic DNA junctions were deep sequenced on the illumina Hi-seq 2000 and mapped onto the Gallus gallus genome.

We chose to sonicate the DNA to induce random fragmentation. This introduces different sonication breakpoints to unique DNA molecules across multiple cells. Sometimes multiple cells in the sample carry the same viral integration site due to clonal expansion. When this is the case, shearing the DNA from these cells can produce multiple fragments that share the same integration site but have different sonication breakpoints. By quantifying these breakpoints after deep-sequencing, we were able to determine the relative abundance of an integration with respect to the other integrations in the sample. We refer to integrations that exhibit more than one sonication breakpoint as “expanded clones”.

A total of 25,961 integrations were identified among five ALV-J induced hemangiomas analyzed. Integrations were mapped and quantified with a custom data analysis workflow. 15,738 integration “hotspots” were identified among all myeloid tumor, hemangioma, and non-tumor tissues.

These hotspots were defined as any genomic loci containing at least 2 integrations where each integration is separated by no more than 5kb. The top 20 hotspots are shown in **Table 3.2**. Although these loci are sites of frequent integration, breakpoint analysis showed that these integrations are not found in highly expanded clones.

Interestingly, endogenous retrovirus LOC101750146 is tied for the most frequent target of integration among the hemangiomas we analyzed, with a total of 29 integrations (**Table 3.2**). However, none of these clones were highly expanded and many integrations were also seen in non-tumor tissues, so these integrations are unlikely to be relevant to tumor induction.

Bird	Neoplasm observed in tissue									
	Bursa	Liver	Spleen	Brain	Kidney	Breast	Leg	Intestine	Lung	Head
B1	X	X	X	X	—	M	H	—	—	H
B2	X	X	X	X	—	—	—	—	H	—
B3	X	X	X	X	H	—	—	—	—	—
B4	X	X	X	X	—	—	—	H	—	—
B5	X	M	X	X	—	—	—	—	—	—
B6	X	X	X	X	M	—	—	—	—	—
B7	X	H	X	X	M	—	—	—	—	—
C1	X	X	X	X	—	—	—	—	—	—
C2	X	X	X	X	X	X	X	—	—	—

Table 3.1 Tissues collected and neoplasms observed. X, no neoplasm; H, hemangioma; M, myeloid tumor. —, tissue was not collected, and no neoplasm was observed during dissection. The tissues from bird 7 (B7) were not analyzed in this study.

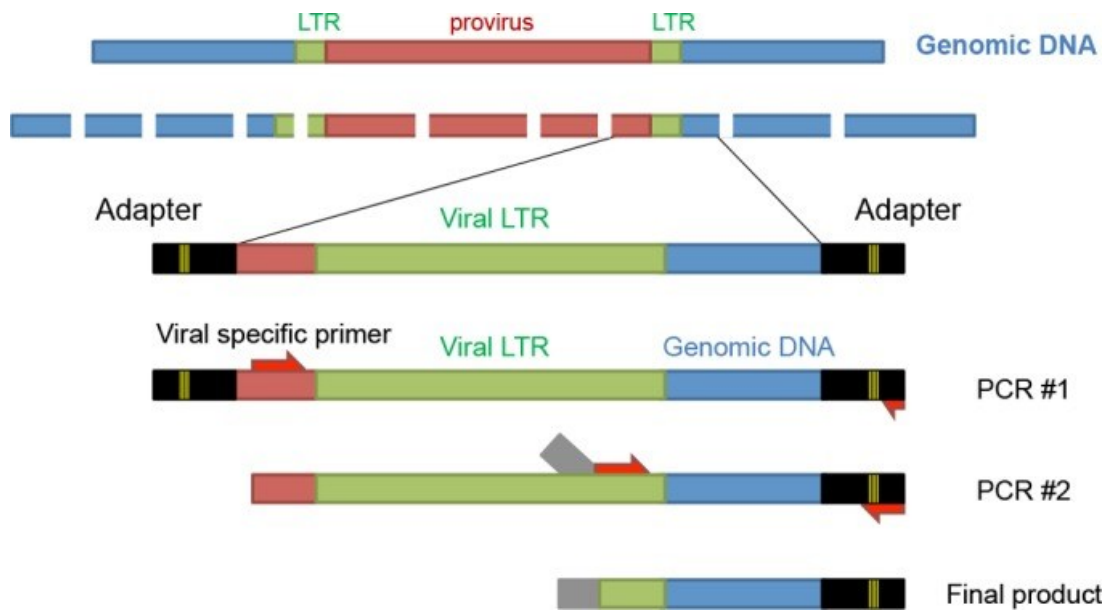


Figure 3.2 Library preparation. Genomic DNA was isolated from tumor and nontumor samples. DNA was sonicated, end repaired, and A tailed. Adapters containing barcodes were ligated onto the DNA fragments, and they underwent two rounds of PCR (nested). The final product underwent multiplexed sequencing on the Illumina Hi-Seq 2000 platform.

Gene	No. of common integrations	Avg no. of breakpoints
LOC101750146	29	1.0
<i>ELF1</i>	29	1.3
<i>LAPTM4A/WDR35</i>	19	1.0
<i>MAML2</i>	17	1.1
<i>SPRY1/ANKRD50</i>	17	1.5
<i>NFKBIA</i>	17	1.1
<i>MAP4K4</i>	16	1.1
<i>ADARB1/GLS/STAT1</i>	16	1.2
<i>SSBP3</i>	15	1.1
<i>GSTA3/ICK</i>	15	1.1
<i>ATAD2/FBXO32</i>	15	1.2
<i>ADD3</i>	14	1.4
<i>RUNX3</i>	14	1.1
<i>SPIDR</i>	14	1.1
<i>RREB1</i>	14	1.1
<i>SRGAP3</i>	14	1.2
<i>INPP5A</i>	13	1.2
<i>ANKRD9</i>	13	1.1
<i>TNFRS21</i>	13	1.0
<i>GRHL1</i>	13	1.0
<i>MET</i>	3	16.0

Table 3.2 Top 20 common integration sites in ALV-J-induced hemangiomas.

Hot spots were defined as any genomic locus containing at least two integrations within 5 kb of each other. By definition, hot spots are flanked on either side by at least 5 kb of sequence lacking an integration. The number of integrations observed in hemangiomas and the average number of breakpoints are listed for each hot spot. MET is shown for comparison.

Some ALV-J integrations are present in expanded clones

In order to measure the relative abundance of each integration, we quantified the number of sonication breakpoints for each site. Previous studies have shown that sonication breakpoints can be used as a measure of clonal expansion; if an integration has more than one breakpoint, the cell carrying that integration has undergone clonal expansion (Berry et al., 2012).

88.5% of the integrations we identified exhibited only a single sonication breakpoint, suggesting that these integrations were not clonally expanded, while 11.5% had two or more breakpoints, evidence of clonal expansion. The 17 most highly expanded clones are shown in **Figure 3.3**. Each of these integrations had 10 or more breakpoints. The tumor which we refer to as a “head hemangioma” is a hemangioma that developed subcutaneously outside the skull.

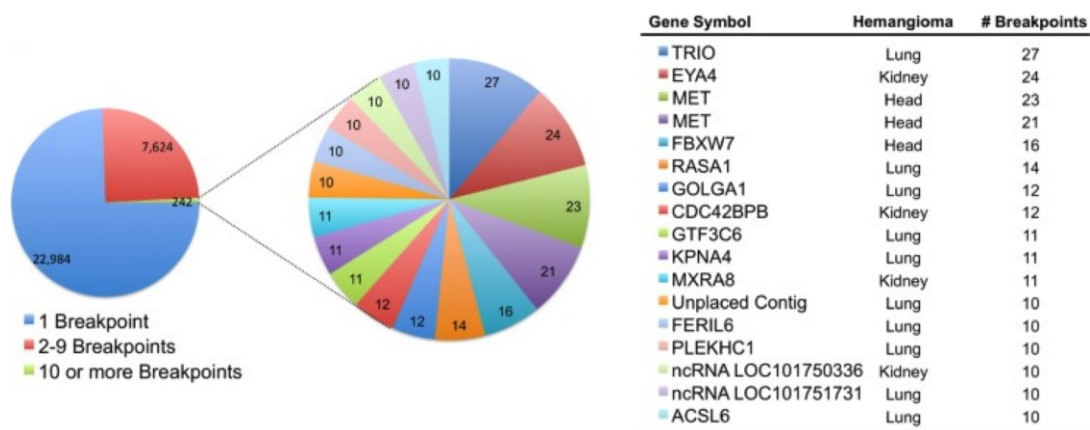


Figure 3.3 Distribution of integration sites in ALV-J-induced hemangiomas. A total of 30,850 breakpoints were identified in ALV-J-induced hemangiomas. Genes that had 10 or more breakpoints are highlighted in a separate pie. Each slice represents a unique integration site, and the size of the slice represents the number of breakpoints for that site. The table shows integrations with the highest number of breakpoints along with the tumor in which that integration was identified. ncRNA, noncoding RNA.

***MET* intron 1 is an expanded common integration site**

Breakpoint analysis revealed four multi-breakpoint integrations within the first intron of the *MET* gene, 3 from ALV-J induced hemangiomas and 1 from an ALV-A induced hemangioma. These integrations occurred in a head (2 integrations), leg and kidney hemangiomas. Strikingly, these *MET* integrations represented the most highly expanded clones in these tumors (**Figure 3.4**). To confirm the deep sequencing results, genomic DNA/proviral junctions from this locus were PCR-amplified and sequenced. This verified that these integrations were present only in the expected tumors, and in the orientation and location as predicted by deep sequencing (data not shown). Interestingly, the two *MET* integrations in the head hemangioma occurred in different orientations and were determined by deep sequencing (of the 3' ends) to be offset by only six nucleotides. We were able to verify the 3' end of each of the proviral junctions by PCR and sequencing, but were unable to amplify the 5' junction in both cases. This suggests that these two proviruses exist adjacent to each other on the same strand and in opposite orientations. The six nucleotide offset observed by deep sequencing could be the result of the characteristic 6 nucleotide duplication that occurs upon ALV integration (Hughes et al., 1981). Several lower breakpoint *MET* integrations were observed in infected non-tumor controls (**Figure 3.5**), but none were seen in myeloid tumors.

Of the 4 multi-breakpoint integrations identified in *MET* hemangiomas, all existed in a tight cluster within intron 1, a region spanning 1599 bp. This

suggests that ALV-J integration within this specific region of *MET* gave these tumor cells a selective advantage and may have driven hemangiomagenesis in these birds. Both of these tumors had at least one highly abundant integration in the same orientation as the *MET* gene. Since the first exon of *MET* is non-coding, this integration pattern suggests that the virus may be inducing over-expression of MET protein.

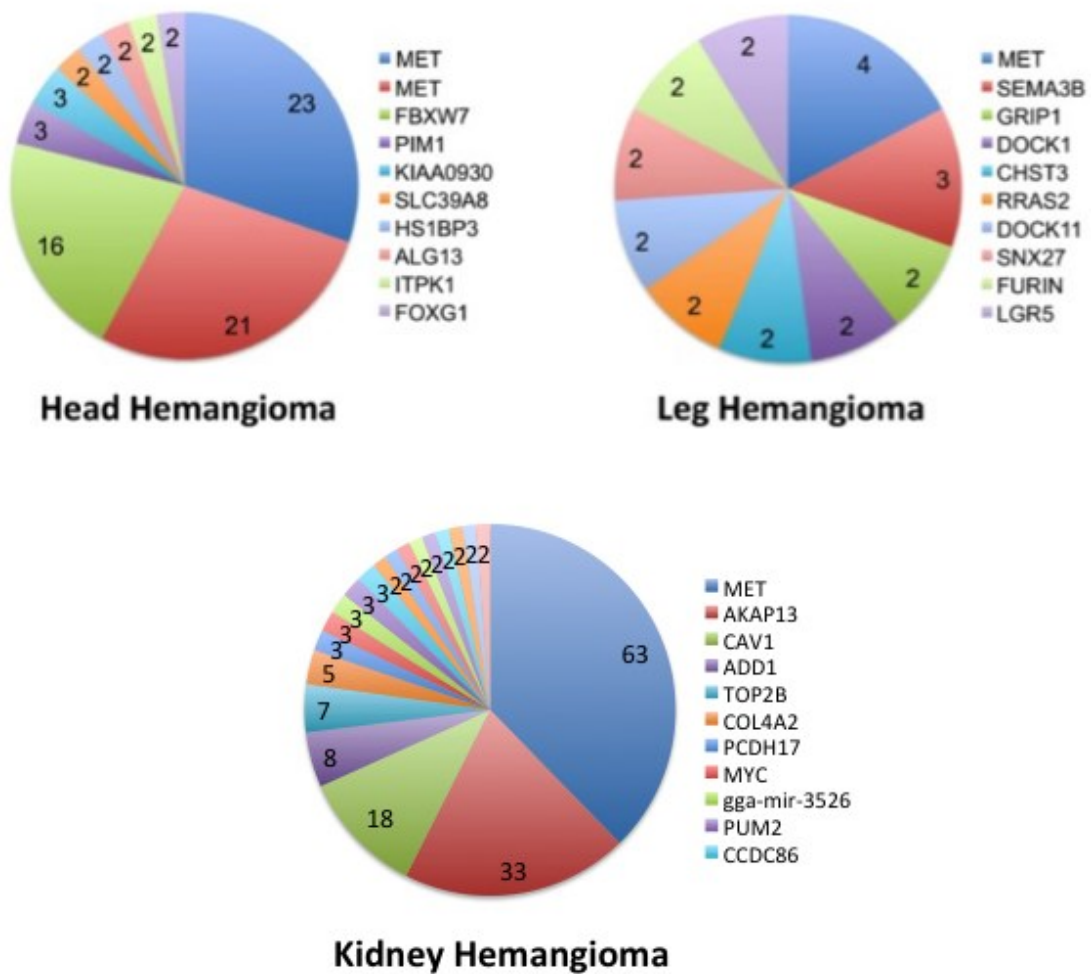


Figure 3.4 The top 10 most abundant integrations identified in the ALV-J induced head, leg and ALV-A induced kidney hemangiomas are shown. Integrations in the MET gene represent the most abundant integration sites in these tumors.

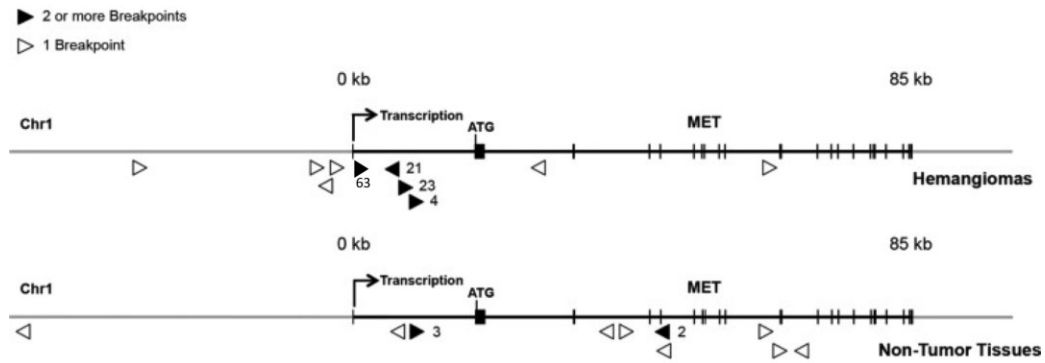


Figure 3.5 ALV-J integration in the MET gene in chicken hemangiomas and non-tumor tissues. There were 10 integration sites at the MET locus in 6 hemangiomas (top) as well as 10 integration sites in 24 non-tumor tissues (bottom) from infected birds. No MET integrations were observed in myeloid tumors. Integrations where multiple breakpoints were observed are represented as solid arrows, and the corresponding number of breakpoints is noted. Arrowheads represent the genomic location and the transcriptional orientation of each provirus. Most of the multibreakpoint integrations were observed in the intron 1 region of the MET gene.

MET mRNA expression is elevated in tumors containing MET proviral integrations

Next we wanted to determine the effects of *MET* intron 1 integrations on gene expression in these tumors. ALV LTRs are known to harbor strong promoter and enhancer sequences that promote gene expression near viral integration sites (Beemon and Rosenberg, 2012). *MET* mRNA expression was analyzed by qRT-PCR (**Figure 3.6**). As expected, we found that *MET* mRNA expression correlated closely with the abundance of the *MET* integration sites in the sample as measured by breakpoint analysis. For example, the head hemangioma which has two integrations (21 and 23 breakpoints), exhibited the highest *MET* expression of any tumor. Expression was a full 131 fold higher than the kidney hemangioma which lacked any high abundance *MET* integrations. The same is true for hemangioma of the leg which also contains a *MET* expanded clone and shows higher levels of *MET* mRNA expression compared to those samples which lack *MET* integrations. This suggests that *MET* expression is induced by viral integration and is playing a causal role in hemangioma development.

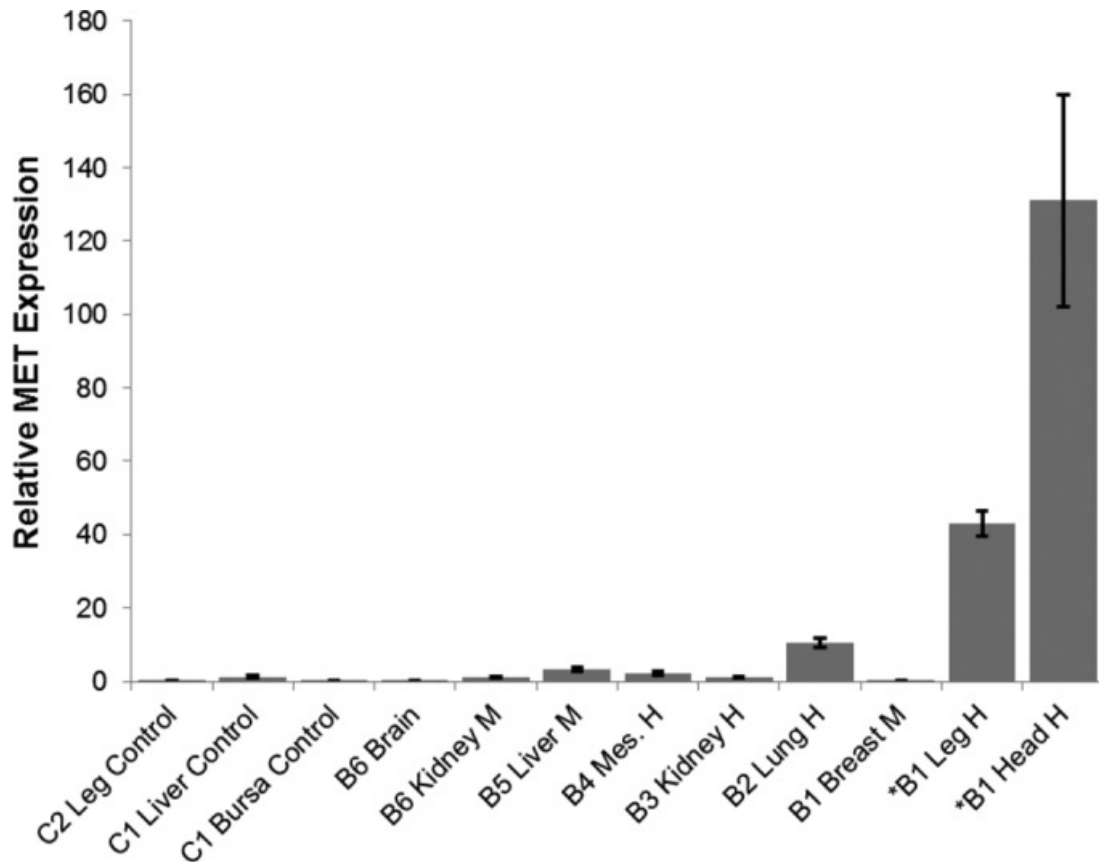


Figure 3.6 *MET* mRNA expression measured by qRT-PCR. Hemangiomas (H) and myeloid tumors (M) as well as nontumor and uninfected controls (labeled control) are shown. The *MET* expression level was normalized to that of *GAPDH* and is relative to that of the hemangioma with the lowest *MET* expression level, B3 kidney hemangioma. Error bars represent standard errors of the means. Tumors containing highly abundant *MET* integrations are noted (asterisks). These tumors exhibited markedly increased levels of *MET* mRNA expression compared to those in other tumors and nontumor controls.

Discussion

In this study we present the first integration profiles of ALV-J induced hemangiomas. Hemangiomas are vascular tumors characterized by uncontrolled angiogenesis (Mabeta and Pepper, 2011). They can occur in many species, including humans. In fact, infantile hemangioma is the most common tumor of early childhood, estimated to develop in 7-10% of infants (Tan et al., 2000). They present as benign tumors that grow post-natally for 8-12 months then typically undergo a slow process of self-involution that can last several years (Blei, 2005; Takahashi et al., 1994).

Previous studies analyzing ALV-induced neoplasms mostly focused on ALV-A induced B-cell lymphoma and identified *MYC*, *MYB*, *MIR-155*, and *TERT* as common integration sites in these tumors (Clurman and Hayward, 1989; Hayward et al., 1981; Payne et al., 1982; Yang et al., 2007a). Only recently have ALV-J induced neoplasms been the subject of similar work. A recent study of ALV-J induced myeloid leukosis showed that *MYC*, *TERT*, and *ZIC1* are targets of proviral integration in myeloid tumors (Li et al., 2014). Interestingly, we did not identify these genes as CISs in this study. This may be due to the fact that only three myeloid tumors were analyzed. It is conceivable that with a larger cohort of tumors these CISs may have been observed as well.

In this study we implicate *MET* overexpression as a causal agent in the development of ALV-J and ALV-A induced hemangiomas in chickens. Interestingly, although viral induced tumors have been the subject of much

study, to the best of our knowledge no previous work has implicated *MET* in tumor induction by insertional mutagenesis. This suggests that *MET* may play a unique role in inducing hemangiomas rather than other types of viral induced tumors.

MET has been studied extensively, and has important roles in both development and cancer. The MET protein is a receptor tyrosine kinase that binds hepatocyte growth factor/scatter factor (HGF/SF), and can activate an array of downstream signaling pathways including PI3K-AKT, RAC1-CDC42, RAP1 and RAS-MAPK (Gherardi et al., 2012). Previous work has shown that *MET* is activated in many types of human cancer via mutation, amplification, and protein overexpression, and *MET* activation correlates with poor prognosis in cancer patients (Abounader and Laterra, 2005; Garcia et al., 2007b; Peghini et al., 2002).

It has also been shown that *MET* plays an important role in angiogenesis, a process that is crucial to the development of hemangioma. For example, activation of the HGF/SF-MET pathway is now understood to be a potent inducer of angiogenesis, specifically in endothelial cells – the same type of cells that give rise to hemangiomas (Abounader and Laterra, 2005; Birchmeier et al., 2003; Bussolino et al., 1992; Grant et al., 1993). In addition, the HGF/SF-MET pathway can suppress TSP1, a negative regulator of angiogenesis, and can induce expression of VEGFA, a pro-angiogenic gene. In this way the HGF/SF-MET pathway controls an “angiogenic switch”, turning on angiogenesis (Zhang et al., 2003).

Additionally, it has been shown that activating mutations in *MET* can induce hemangiosarcomas in mice (Graveel et al., 2004). Viewed in this context it does not seem surprising that ALV-J-induced *MET* overexpression can lead to hemangiomagenesis in chicken. Whether or not *MET* can act as a causal agent in human hemangioma has yet to be determined, but this may be an avenue for further research.

In addition to *MET*, other CISs were identified as hotspots of integration (**Table 3.2**). Notably, none of these integrations were seen in highly expanded clones and they were also observed in non-tumor tissue. This suggests that these CISs are unlikely to be relevant to tumor induction. In contrast, some tumors exhibited non-*MET* integrations in highly expanded clones. For example *TRIO* and *EYA4* both contained integrations with many breakpoints. *TRIO* has guanyl exchange factor (GEF) activity and regulates Rho family GTPases, which coordinate cytoskeletal rearrangement and cell migration (Bellanger et al., 1998; Blangy et al., 2000). *EYA4* is a member of the eyes absent (EYA) family of proteins, it has phosphatase activity and may function in eye development as a transcriptional activator (Borsani et al., 1999). Our data suggest that these genes may also be capable of contributing to hemangioma formation, but because similar integrations were not observed in multiple tumors we cannot establish with certainty that they are drivers of oncogenesis and not merely passenger integrations. Further studies involving a larger cohort of birds may help identify other common integration sites and genes that drive ALV-J induced hemangiomagenesis.

Materials and Methods

Tumor induction

ALV-J strain PDRC-59831 was isolated from a 38-week-old broiler breeder chicken. The case was recorded on 5/30/2007 on a farm near Danielsville, GA. In this study, ALV-J strain PDRC-59831 was inoculated into thirty 5-day-old SPAFAS embryos (Charles River) via the yolk sac route. Four embryos died at day 10 of embryogenesis, leaving twenty six viable embryos. Of these, eleven hatched (11/26; 42%), five pipped but could not complete hatching, and ten embryos did not hatch and died. In comparison, 90% (19/21) of the uninoculated control chickens hatched. Chickens were observed daily and euthanized when apparently ill or at 12 weeks of age. Of the eleven ALV-J-infected chickens, one was euthanized at week 2 and three died at week 11 for reasons unrelated to infection. One died at week 5 and one was euthanized at week 7, both had tumors. Five were euthanized at 12 weeks of age, all of which had tumors. In total, tumor tissue was obtained from six ALV-infected chickens (**Table 3.1**). The early death of bird #7 prevented the collection of useful tissue. Tumors were classified by gross examination, tissue samples were collected and flash frozen, and then stored at -80C.

The ALV-A induced kidney hemangioma is tumor C8K, was induced and isolated as described in the materials and methods section of Chapter 2.

Phylogenetic analysis of ALV-J isolate PDRC-59831 *env* gene

The sequence of the ALV-J *env* gene was determined by Sanger sequencing (NCBI accession number KP284572). Phylogenetic analysis was conducted using the Neighbor-Joining method (Saitou and Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (Fig. 1) (Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. The analysis involved 16 nucleotide sequences. Accession numbers are as follows: HUB09JY03 (HQ634811), HPRS-103 (Z46390), UD4 (AF307951), UD5 (AF307592), PDRC-59831 (KP284572), YZ9902 (HM235670), NX0101 (DQ115805), NM2002-1 (HM235669), JS-nt (HM235667), SD07LK1 (FJ216405), SCDY1 (HQ425636), NHH (HM235668), JL093-1 (JN624878), JS09GY6 (GU982310), JS09GY3 (GU982308), ADOL-7501 (AY027920). Codon partitions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. These analyses were conducted in MEGA6 (Tamura et al., 2013).

DNA extraction and deep sequencing

50 mg of tissue was first homogenized with Kimble-Chase Kontes™ Pellet Pestle™, and then digested with Proteinase K at 50°C for 15 hours. The sample was then phenol extracted, put through a 25 gauge needle ten times, and ethanol precipitated; this procedure was repeated. The sample

was then treated with 2ug RNase A for 1 hour at 37°C. Phenol extraction, shearing, and ethanol precipitation were repeated, and DNA concentration was measured with a Thermo Scientific Nanodrop 2000c. 5ug of purified genomic DNA was sonicated with a Bioruptor™ UCD-200. End repair, A-tailing, and adapter ligation were performed as described by Gillet et al. (Gillet et al., 2011) (adapter short arm: p-gatcgggaagagcaaaaaaaaaaaaaaaaaa, adapter long arm: caagcagaagacggcatacagagatxxxxxgtgactggagttcagacgtgtgctcttccgatac*t where the x's denote the barcode sequence). Multiplexed barcodes each differed by at least two nucleotides. Two rounds of PCR (nested) were conducted to enrich the library for proviral junctions. The first PCR reaction had 23 cycles, and employed one ALV-J specific primer (gggactgtagcatgtataggcgctgag) between 3'LTR and *env*, and a second primer within the adapter ligated on earlier (caagcagaagacggcatacagagat). The second round of PCR employed a primer (aatgatacggcgaccaccgagatctacactcgacgattgcgagcacctgaatgaag) at the 3' end of the LTR, twelve nucleotides short of the junction between virus and genomic DNA, as well as a nested adapter primer overlapping the barcode sequence within the adapter (caagcagaagacggcatacagagatxxxxx) (Fig. 2). Libraries were quantified by qPCR, and then underwent single-end 100 bp multiplexed sequencing on the Illumina Hi-Seq 2000. A custom sequencing primer (acgattgcgagcacctgaatgaagtgaagg) was used that hybridized near the end of the viral 3' LTR, five nucleotides short of the junction between viral

and genomic DNA. This allowed for reads to be validated as genuine integrations, by verifying that each read begins with the last 5 nucleotides of the proviral DNA. In all, 9,759,304 junction reads were obtained.

Sequencing Analysis

Reads were first filtered with a custom python script to remove sequences that did not begin with the last five nucleotides of viral DNA, “CTTCA.” A total of 1,352,053 reads (13.85%) were discarded because they did not begin with this verification sequence. The files were then uploaded to Galaxy (Blankenberg et al., 2010a; Giardine et al., 2005; Goecks et al., 2010), which was used to perform downstream analyses. In Galaxy, first the quality scores were converted to Sanger format with FastQ Groomer v1.0.4 (Blankenberg et al., 2010b). Adapters were then trimmed using the Galaxy Clip tool v1.0.1. This tool also removed reads containing an N (14,871, 0.18% of reads removed at this step) and reads less than 20 nucleotides in length after adapter removal (4,289,006, 51.02%). The remaining reads (4,103,370, 48.81%), were mapped with bowtie (Langmead et al., 2009b) using the Gallus gallus 4.0 genome (Nov. 2011). 100,000 random mapped reads were selected from each sample to be used for further analysis. If less than 100,000 reads were present for a sample, all available reads were used. A custom Perl pipeline was developed to analyze the aligned reads output from bowtie. Briefly, reads containing sequencing errors were filtered and read counts and breakpoints were quantified. Integrations found in multiple samples were assigned to the sample with highest number of

breakpoints. Integration hotspots were identified by walking the genome. If two integrations were within 5kb of each other they were placed into an integration group. Additional integrations were included in the group until an integration-free span of 5kb was reached. Files were annotated with refseq features and the orientation and distance to the nearest gene were calculated for each integration. The source code is available upon request.

***MET* gene expression**

RNA transcripts were isolated from all five hemangiomas, three myeloid tumors, and four non-tumor controls with RNA-Bee (Tel-Test). First-strand synthesis was performed with a poly-dT primer and Superscript III reverse transcriptase (Invitrogen) following the manufacturer's protocol. Quantitative reverse transcription PCR (qRT-PCR) was conducted with iQ SYBR Green Supermix (Biorad) according to manufacturer's protocol. *MET* primers (aggacattttgggtgtgtgt, aactgagccacttcttcag) were designed using primer3plus software (www.primer3plus.com). Thermal cycling was conducted on a Biorad C1000 Thermal Cycler/CFX96 Real-Time System. *MET* expression was normalized to *GAPDH* using primers published previously (Heidari et al., 2008). PCR was repeated four times and each sample was present in duplicate during each run. Results were normalized to B3 Kidney Hemangioma using the comparative C_T method. Melt curves were conducted to verify specificity of the primers.

**Chapter 4. Avian leukosis virus activation of an antisense RNA
upstream of *TERT* in B-cell lymphomas**

Adapted from:

Nehyba J.*, **Malhotra S.***, Winans S.*, O'Hare T, Justice J. 4th, Beemon K. (2016). Avian leukosis virus activation of an antisense RNA upstream of TERT in B-cell lymphomas. J Virol. 90(20). (* indicates equal contribution).

Summary

Avian leukosis virus (ALV) induces tumors by integrating its proviral DNA into the chicken genome and altering expression of nearby genes via strong promoter and enhancer elements. Viral integration sites that contribute to oncogenesis are selected in tumor cells. Deep sequencing analysis of B-cell lymphoma DNA confirmed that the telomerase reverse transcriptase (*TERT*) promoter is a common ALV integration target. Twenty-six unique proviral integration sites were mapped between 46 and 3552 nt upstream of the *TERT* transcription start site, predominantly in the opposite transcriptional orientation of *TERT*. RNA-seq analysis of normal bursa revealed a transcribed region upstream of *TERT* in the opposite orientation, suggesting the *TERT* promoter is bidirectional. This transcript appears to be an uncharacterized antisense RNA. We have previously shown that *TERT* expression is up regulated in tumors with integrations in the *TERT* promoter region. We now report that the viral promoter drives expression of a chimeric transcript, containing viral sequences spliced to exons 4 through 7 of this antisense RNA. Clonal expansion of cells with ALV integrations driving over expression of this *TERT* antisense RNA suggest it may have a role in tumorigenesis. These data suggest that ALV integrations in the *TERT* promoter region drive overexpression of a novel antisense RNA and contribute to the development of lymphomas.

Introduction

Avian leukosis virus (ALV) is a simple retrovirus that does not carry a viral oncogene but induces tumors by insertional mutagenesis (1,2). ALV typically induces B-cell lymphomas, but can also induce erythroblastomas, hemangiomas, and myeloid tumors (Beemon and Rosenberg, 2012; Justice IV and Beemon, 2013; Justice et al., 2015a). Proviral integration can up-regulate expression of nearby genes via strong enhancers and promoter elements in the viral long terminal repeat (LTR) sequences. An advantage of using ALV as an insertional mutagenesis tool is its relatively random integration pattern, with only a slight preference for actively transcribed sites (Barr et al., 2005; Justice et al., 2015b; Narezkina et al., 2004). A reduced integration bias allows us to map proviral integrations in many genomic locations and to observe selection of integration sites with oncogenic potential. Previous studies have shown common integration sites in ALV-induced lymphomas in *MYC*, *MYB*, *BIC* (miR 155 precursor), and telomerase reverse transcriptase (*TERT*) genes (Baba and Humphries, 1986; Clurman and Hayward, 1989; Hayward et al., 1981; Justice et al., 2015b; Yang et al., 2007a). High throughput sequencing revealed multiple integration sites in a series of rapid-onset B-cell lymphomas (6). The *TERT* promoter region was identified as the most clonally expanded of these integrations, suggesting this is an early event in tumorigenesis (Justice et al., 2015b). Twenty-six unique integration sites were identified in this region in multiple independent tumors (Justice et al., 2015b).

Telomerase is a ribonucleoprotein complex that adds repeat sequences to chromosome ends. It contains a catalytic protein component, TERT, as well as a non-coding telomerase RNA template component (TERC). Elevated telomerase activity has been detected in more than 90% of all human cancers (Shay and Wright, 2011). In addition, many human tumors have a point mutation in the *TERT* promoter at -124 or -146 nt upstream of the *TERT* translation start site (Heidenreich et al., 2014). These mutations up-regulate *TERT* expression (Atala, 2015; Horn et al., 2013; Huang et al., 2013; Killela et al., 2013). Elevated telomerase activity maintains telomere lengths and prevents apoptotic signaling, thus allowing continual proliferation and long-term viability of cancer cells (Blasco, 2005). It has also been shown that *TERT* can promote oncogenesis independent of the reverse transcriptase function of telomerase (Koh et al., 2015).

Telomerase activity in most somatic cells is limited by the availability of TERT protein, and expression of *TERT* is tightly regulated at the transcriptional level through epigenetic modifications in the promoter region (Delany and Daniels, 2004; Zhu et al., 2010). In addition, extensive alternative splicing of the *TERT* transcript has been shown to generate inactive variants that decrease telomerase activity (Hrdlickova et al., 2012; Saebøe-Larssen et al., 2006; Withers et al., 2012). Both human and chicken telomerase expression is down regulated in most normal somatic tissues (Liu et al., 2002; Taylor and Delany, 2000). Furthermore, human and chicken telomeres shorten with age, and telomerase activity is important for

oncogenesis (Delany et al., 2000). In contrast, mice express telomerase in normal somatic cells and have longer telomeres than humans or chickens (Hackett and Greider, 2002). Therefore, chicken serves as a good model to study oncogenic events of TERT activation and signaling.

We previously reported that ALV integrations upstream of *TERT* cause a 2-4 fold up-regulation of *TERT* expression in rapid-onset B-cell lymphomas (Yang et al., 2007a). However, these integrations were in the opposite transcriptional orientation to *TERT*, unlike most previously characterized common integration sites in ALV-induced tumors (Clurman and Hayward, 1989; Hayward et al., 1981; Kanter et al., 1988). In this work, we show that these integrations also drive over-expression of a novel antisense transcript, associated with the bidirectional *TERT* promoter, which we call TAPAS (TERT Antisense Promoter ASsociated) RNA. The ALV integrations result in chimeric transcripts with ALV leader sequences spliced into exon 4 of the 7-exon TAPAS RNA.

Results

***TERT* promoter is a common ALV integration site in B-cell lymphomas**

In order to identify genes contributing to the formation of ALV-A induced rapid-onset B-cell lymphomas, high throughput sequencing of proviral – host DNA junctions was previously performed (Justice et al., 2015b). Common integration sites in the host genome that contribute to tumorigenesis are present in multiple tumor cells and thus are overrepresented in the deep sequencing data. The *TERT* promoter region was identified as the most clonally-expanded, common integration site with integrations present in seven different lymphomas from five birds (**Figure 4.1A**). We analyzed 19 of the most clonally expanded, unique integrations from both primary bursal tumors (B) and tumors metastasized to the liver (L). Three of the clonally expanded integrations were present in multiple tissues from the same bird. The integration sites ranged from 46 nt to 3552 nt upstream of the *TERT* transcription start site. The majority of the proviral integrations (16/19) were in the opposite transcriptional orientation to *TERT*. Four out of 7 lymphomas, termed C7B, C6L, C7L, and D2L had integrations only in the opposite orientation. The remaining three tumors – A1B, C2B, and C2L – harbored integrations in the same orientation as *TERT*, but also contained integrations in the opposite orientation that were more clonally expanded. In contrast, no integrations in the *TERT* promoter region were identified in any non-tumor tissues of infected birds (**Figure 4.1B**). The observation of proviral integrations in this region in multiple tumors suggests

that ALV integration in the *TERT* promoter contributes to driving lymphomagenesis in these birds.

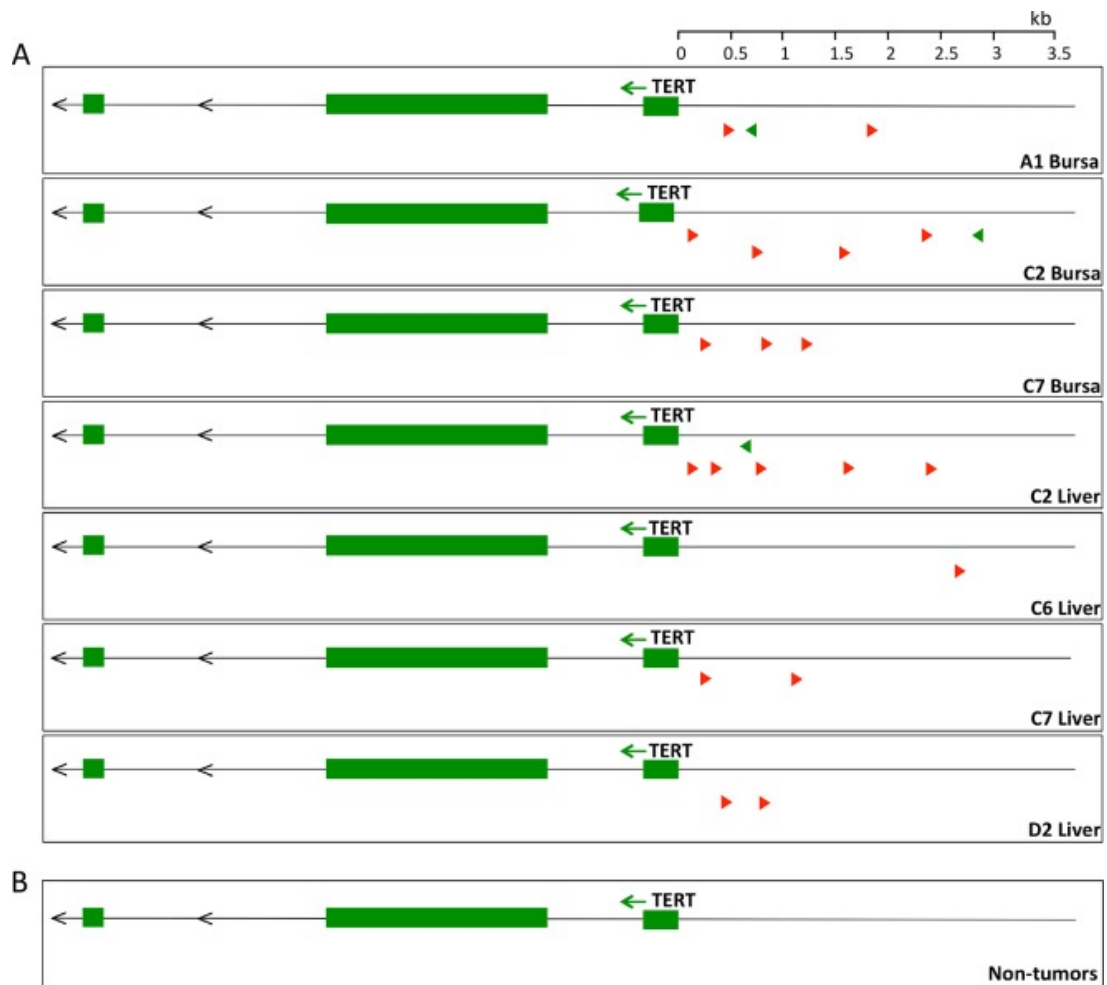


Figure 4.1 The TERT promoter region is a common site of ALV proviral integration in lymphomas. (A) Schematic of the most clonally expanded ALV integration sites near *TERT* in 7 tumors, shown with the first 3 exons of the *TERT* gene. The tumor names correspond to the bird and tissue from which the tumor was collected. All of the integrations clustered within 3.5 kb upstream of the *TERT* transcriptional start site. The integrations are predominantly in the opposite orientation (red) with respect to *TERT* transcription. (B) Schematic of integrations near *TERT* in 6 nontumor tissues from infected chickens.

Novel antisense (TAPAS) RNA is transcribed from *TERT* bidirectional promoter

In order to assess the effects of proviral integrations on host gene expression, deep sequencing of the transcriptome of selected ALV-induced lymphomas and normal bursa controls was performed. This analysis revealed a 9 kb transcribed region upstream and in the opposite transcriptional orientation of *TERT* in the normal bursa controls (**Figure 4.2A**). This suggests that the *TERT* promoter is bidirectional. With the use of TopHat bioinformatics tools, a number of putative introns were identified and confirmed by sequence analysis of exon junctions. This analysis suggested a 3.6 kb spliced transcript, containing 7 exons. RT-PCR studies confirmed 2.2 kb of this transcript containing exons 1 through 7 (**Figure 4.2A**). RT-PCR experiments were not able to amplify the last 1050 nt of exon 7. Two putative poly(A) sites were identified by 3'RACE at positions 1051 and 1114 of exon 7.

Strand-specific RNA-seq data indicates that the first exon of *TERT* and the associated bidirectional transcript overlap (**Figure 4.2A**). RT-PCR verified that at least the first 161 nt of *TERT* exon 1 are shared with TAPAS RNA. It is possible that more of exon 1 is shared between the two genes; however, this could not be confirmed by RT-PCR, likely due to the high GC content of *TERT* exon 1. Additionally, a number of alternatively spliced transcripts were detected, including some skipping exon 2 and others skipping both exons 2 and 3 (**Figure 4.2B**).

There is a small open reading frame (ORF) (258 nt) that spans exons 4 and 5 and two longer ORFs, of 408 and 375 nts, present in exon 7. However, one of these ORFs is located within the unverified region at the 3' end of the transcript, beyond the main transcription termination sites discussed above. Further, we observed that exon 7 is poorly conserved between most avian species (*data not shown*). Moreover, no protein domain homology was observed in any region of the transcript (Marchler-Bauer et al., 2015), implicating this transcript as a putative long non-coding RNA (lncRNA).

The recent release of the *Gallus gallus* 5 whole-genome assembly predicts an antisense transcript upstream of *TERT* (LOC107052651). The predicted transcript variant (XR_001465267.1) corresponds to exons 2 through 7 of TAPAS RNA (Figure 2a). This variant contains 643 nt more of exon 7 and does not share any sequence with *TERT* exon 1, unlike the transcript reported here. Another transcript variant with retention of an intron between our exons 2 and 3 is also predicted (XR_001465266.1). Further, the NCBI eukaryotic gene prediction tool, Gnomon, annotates the predicted transcript as a lncRNA (Accession No. NC_006089).

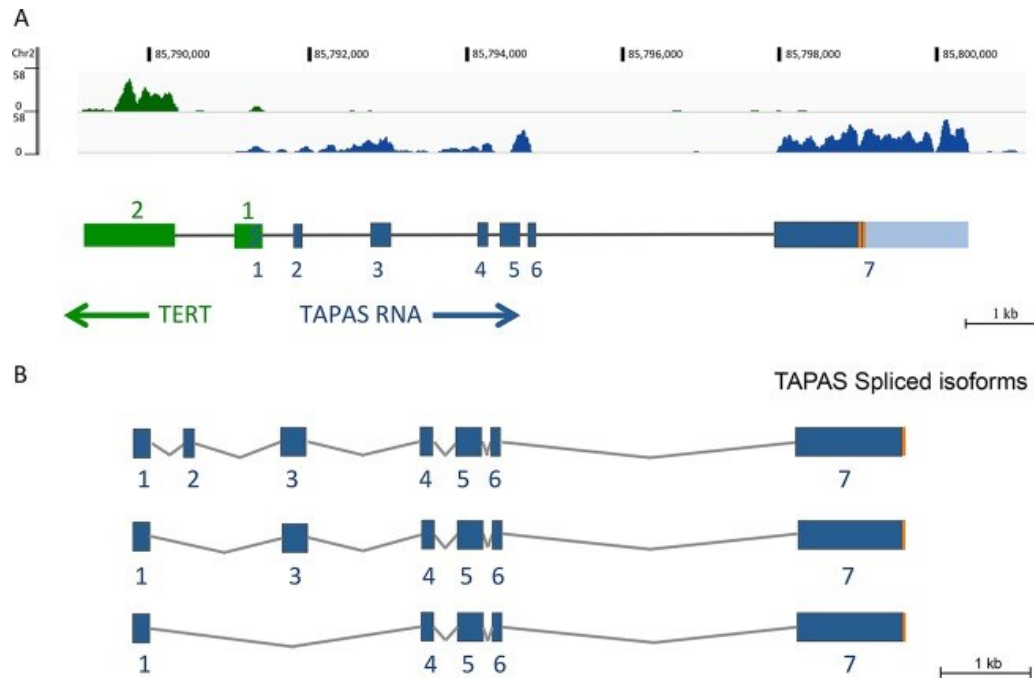


Figure 4.2 Schematic of the TAPAS gene. (A) Representative RNA-seq (Bedgraph) from normal bursa tissue showing normalized transcription coverage. Coverage on the plus (green) and minus (blue) strands is shown. The primary transcript observed by RNA-seq in normal chicken bursa is approximately 9 kb. The principal form of the spliced transcript is 3.6 kb and contains 7 exons. The confirmed region of the shared exon 1 is hatched in blue. A portion of exon 7 that could not be verified by RT-PCR is indicated in light blue. Transcripts confirmed by RT-PCR are 2 to 3 kb. The two main 3' ends identified are indicated by vertical orange lines and are located at nucleotides 1051 and 1114 of exon 7. (B) Multiple alternatively spliced variants of TAPAS RNA were also identified in normal bursa by RT-PCR.

TAPAS RNA expression is elevated in tumors with integrations in the *TERT* promoter

The predominance of proviral integrations in the opposite orientation of *TERT*, as well as the identification of a bidirectional transcript, suggested that the integrations might also be driving increased expression of TAPAS RNA. To test this hypothesis, we performed quantitative reverse-transcription PCR (qRT-PCR) to determine TAPAS RNA expression levels in tumors containing integrations in the *TERT* promoter region (**Figure 4.3A**). Normal liver has 148- and 5-fold less expression than normal bursa for TAPAS and *TERT* RNA, respectively. Compared to normal liver, tumors with integrations in the *TERT* promoter had significantly increased expression of the TAPAS RNA. Expression of the bidirectional TAPAS RNA was up-regulated approximately 250- to 3858-fold relative to normal liver tissue. In contrast, *TERT* was up-regulated 4- to 42-fold relative to normal liver tissue (**Figure 4.3B**). This suggests that the observed integrations in the *TERT* promoter are driving expression of a bidirectional lncRNA. These findings were also confirmed by RNA-seq analysis of liver tumors C6, C7 and D2 (*data not shown*).

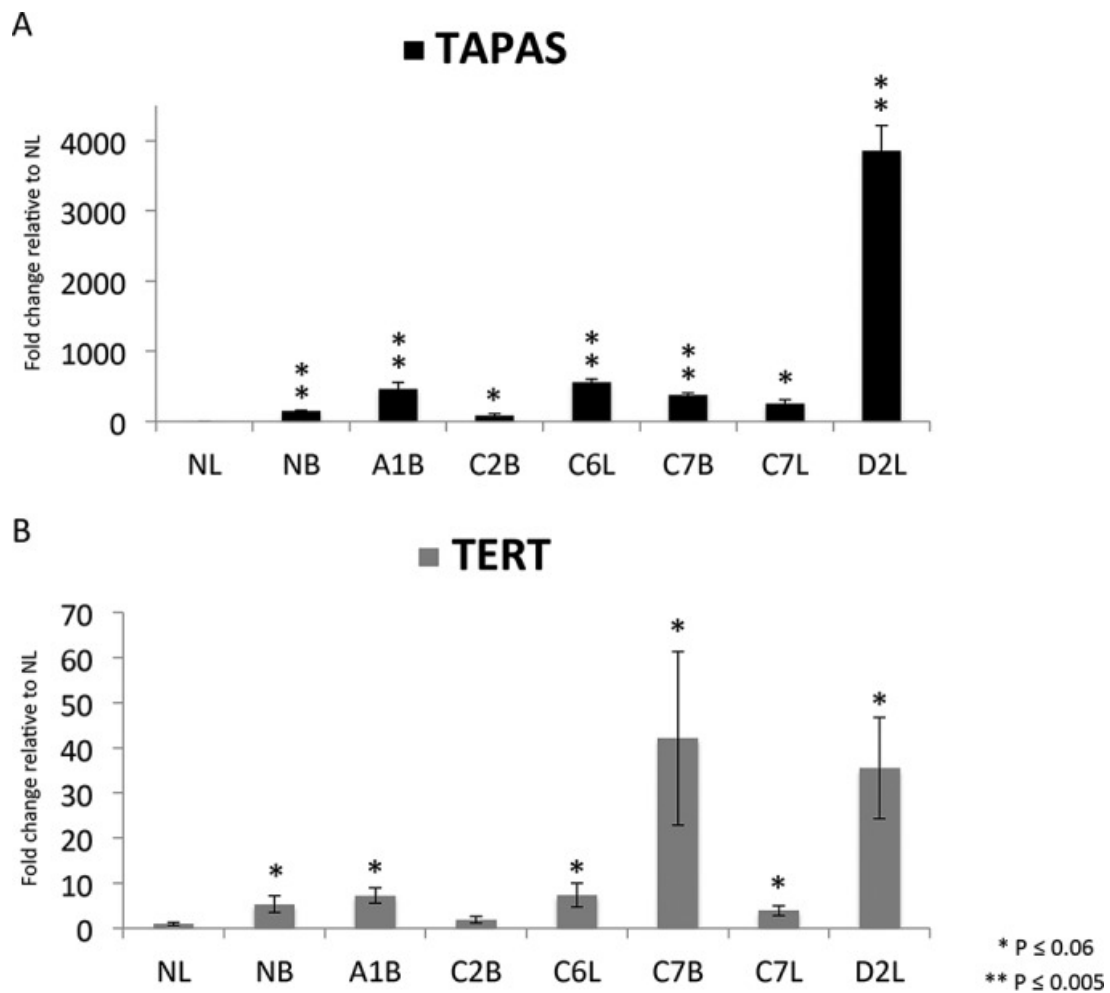


Figure 4.3 Expression of TAPAS RNA and *TERT* in ALV-induced B-cell lymphomas. qRT-PCR was performed to determine TAPAS RNA (A) and TERT (B) expression in seven ALV-induced tumors, as well as normal bursa (NB) and normal liver (NL) controls. Expression of both transcripts was significantly higher in 5 of the 6 tumors than in NL. The P values are representative of Bonferroni corrections for multiple comparisons. The error bars indicate standard deviations.

Viral transcripts splice into exon 4 of TAPAS RNA

Retroviruses can induce overexpression of host genes by multiple mechanisms (1,44). For instance, insertion of viral enhancer elements in the vicinity of host gene promoters can induce overexpression (Beemon and Rosenberg, 2012). Alternatively, the viral promoter can drive expression of the host gene directly, if both are in the same orientation, by readthrough of the viral poly(A) site (Kanter et al., 1988). If the promoter in the viral 5' LTR is driving expression, the viral RNA transcript can splice via the *gag* splice donor into the cellular mRNA (Kanter et al., 1988). Alternatively, if the promoter in the 3' LTR is used, read-through into the adjacent host genomic region occurs (Hayward et al., 1981). To determine the mechanism by which proviral integrations are affecting TAPAS RNA expression, we analyzed metastasized tumors with integrations in the same transcriptional orientation as the TAPAS RNA. We performed RT-PCR using LTR specific primers and primers within the TAPAS RNA exons to obtain and sequence viral TAPAS RNA fused transcripts (**Figure 4.4A**).

Provirus in tumor D2L use the 5' LTR to drive expression of TAPAS RNA. One splice variant used the canonical 5' viral splice donor site in *gag* (nucleotide 398). Transcripts were also detected in which an alternative splice donor site in the viral *gag* gene (nucleotide 857) spliced into exon 4 of TAPAS RNA. In tumor C7L, a provirus integrated in intron 2 also spliced into exon 4 of TAPAS RNA from the canonical 5' viral splice donor site.

Alternatively, transcripts in which the viral 3' LTR serves as the promoter were observed in tumor C6L. These transcripts contained 63 nucleotides of host DNA adjacent to the 3' LTR. It appears that a cryptic splice donor site present in this intronic region may be used to splice into the downstream exon 4. Sequencing of the provirus C6L revealed a large deletion that included the viral splice donor site in *gag* (**Figure 4.4B**); this would prevent its splicing into the TAPAS RNA if it initiated in the 5' LTR. However, this deletion would probably also prevent transcription initiation at the 5' LTR, as previously observed with ALV integrations in *MYC* (Goodenow and Hayward, 1987). A similar 3' LTR initiated transcript was observed from the C7L proviral integration in intron 1 of TAPAS RNA. This variant spliced from a cryptic splice donor site, 28 nucleotides downstream of the provirus, into exon 4 of the TAPAS RNA.

The majority of proviral integrations are located between exon 1 and exon 4 of the TAPAS gene (**Figure 4.4A**). However, regardless of the integration site location, all of the viral transcripts analyzed invariantly spliced into exon 4 of the TAPAS RNA. For most of the viral transcripts, this means that nearby exons are skipped and splicing is preferentially occurring to exon 4. While the 5' spliced viral leader sequence contains a bit of the ALV *gag* gene, the analyzed chimeric transcripts have a termination codon before the AUG of the ORF in exon 4 of the TAPAS RNA. Thus, no hybrid protein is predicted. Consistent splicing into exon 4 suggests the possible functional importance of this region of the TAPAS RNA.

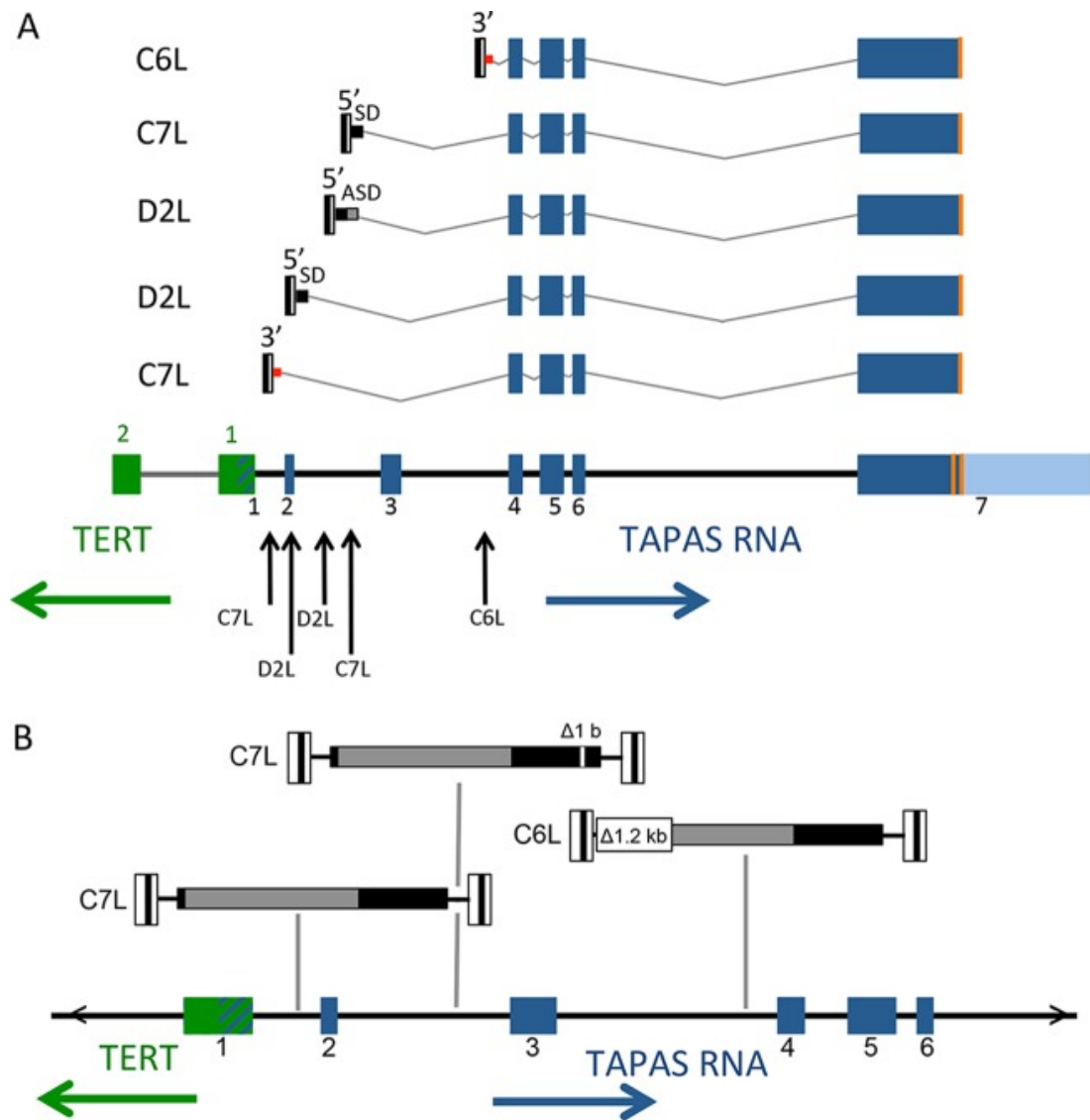


Figure 4.4 Viral RNAs splice into exon 4 of TAPAS RNA. (A) Splicing of viral transcripts was determined by RT-PCR of tumor RNA. All proviruses shown were in the opposite transcriptional orientation to *TERT*. The arrows indicate the genomic locations of proviral integration. Despite the presence of upstream exons, all the viral transcripts analyzed spliced into exon 4 of the TAPAS RNA either from a canonical *gag* splice donor (SD) or via alternate

splice donor (ASD) sites. Readthrough transcription from the 3' LTR is depicted by the red square. (B) Sequence analysis showing mutations in proviruses C6L and one of the C7L integrations. Three integrated proviruses, two in tumor C7L and one in tumor C6L, were sequenced. The viral LTRs are depicted as white boxes at the termini of the viral genome with the U3, R, and U5 direct repeats. *gag-pol* are depicted in gray and *env* in black. The 1.2-kb deletion in C6L removes the canonical viral splice donor and induces transcription from the 3' LTR. The exon positions of *TERT* and TAPAS RNA are shown for reference.

TAPAS RNA is expressed in normal chicken tissues and during development

To further characterize expression of the TAPAS RNA, we analyzed publicly available RNA-seq data sets from the SRA database (Leinonen et al., 2011). TAPAS RNA and *TERT* expression was measured in various chicken tissues of an 18 day embryo. In addition, transcriptome of total embryos were analyzed at different time points up to 12 days of development (SRA Accession no. ERX697750 and DRX001564 respectively) (Leinonen et al., 2011). Quantification of transcript expression was determined by calculating fragment count normalized to transcript length and total number of reads (FPKM). This analysis showed that the TAPAS RNA is expressed in some normal chick embryo tissues. TAPAS RNA is expressed at particularly high levels in bursa, testes and kidney and is undetectable in muscle and heart tissue (**Figure 4.5A**). TAPAS RNA expression was higher than *TERT* in bursa and testes but more comparable in other tissues and in total early embryos.

Furthermore, both TAPAS RNA and *TERT* expression is elevated early in chick development and progressively decreases with time (**Figure 4.5B**). Thus, the tissue specific and developmental expression of TAPAS RNA correlates with *TERT* expression. This suggests that the TAPAS RNA may have a role in regulating *TERT* expression or that the expression of the two transcripts is co-regulated.

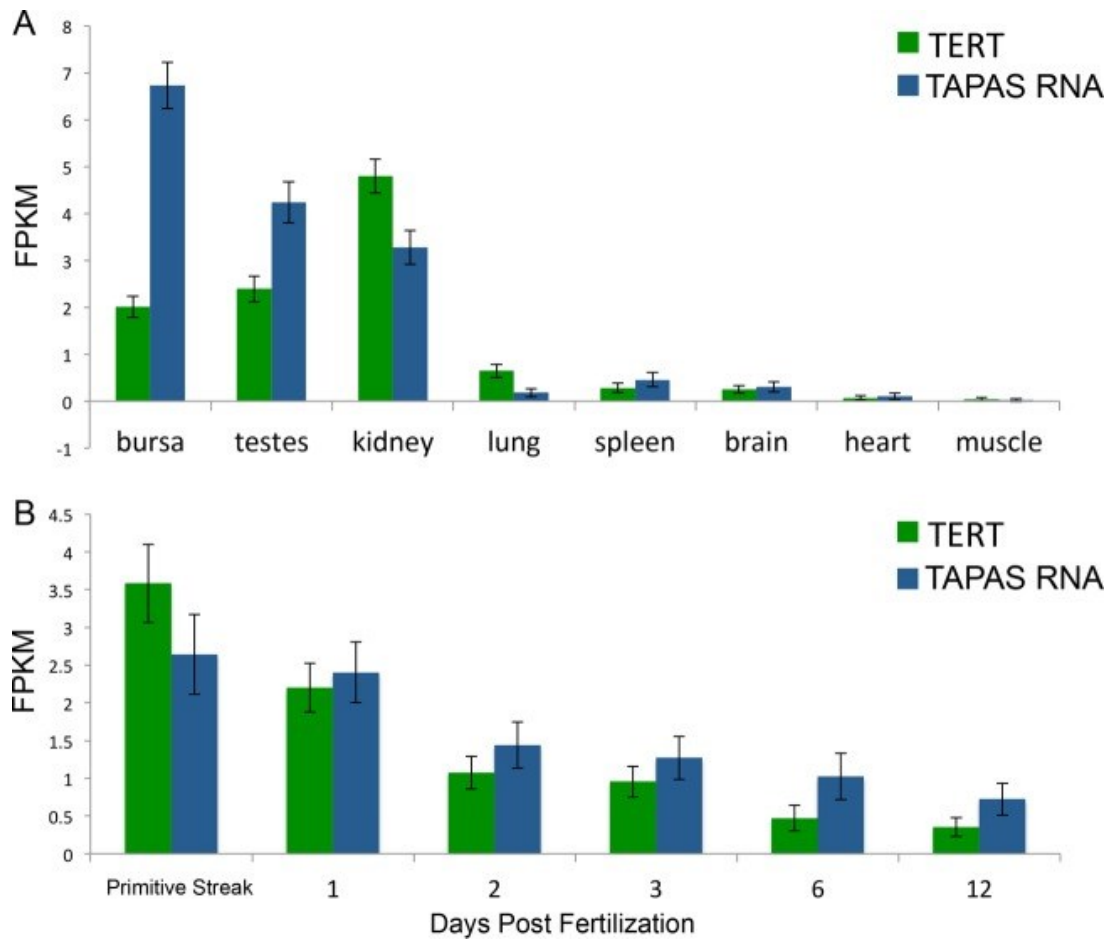


Figure 4.5 *TERT* and TAPAS RNA are expressed at comparable levels in adult tissues and during chick development. (A) RNA-seq data on expression of TAPAS RNA in normal 18-day chicken embryo tissues. TAPAS RNA is expressed in various tissues at levels similar to those of *TERT*. In bursa, TAPAS RNA expression is approximately 3-fold higher than that of *TERT* (Pearson correlation coefficient = 0.66; $P = 0.07$). (B) The expression of *TERT* and TAPAS RNA also seem to be correlated throughout development (Pearson correlation coefficient = 0.92; $P = 0.001$). The error bars indicate standard deviations.

TAPAS RNA is conserved

To determine if the novel TAPAS RNA is conserved, we performed phylogenetic analysis in multiple avian species. Exons 4, 5 and 6 were used for the analysis because this region was found in all alternatively spliced transcripts, as well as in all viral TAPAS RNA fused transcripts, suggesting it may have functional importance. Regions homologous to the TAPAS gene exons 4-6 and intervening introns, were identified in various avian genomes by BlastN, and the sequences were aligned by ClustalX. It was observed that this region is highly conserved at the sequence level in many, but not all, avian lineages (**Figure 4.6**). Additionally, based on transcriptome analysis, there exists a predicted lncRNA in the TERT promoter region in the most recent genome assemblies for chicken (LOC107052651), turkey (LOC104910189) and Japanese quail (LOC107309454) (*data not shown*). The chicken sequence shared 95% identity with the corresponding genomic region of the closely related black grouse (*Lyrurus tetrix*), and 72-76% identity with the genomes of various neoavian lineages. The turkey and chicken sequences share limited similarities in exons 4-6, but have significant sequence homology in exon 7. The only perching bird species in which this region was conserved was the most basal species of New Zealand wren (*Acanthisitta chloris*) suggesting the possibility that these sequences underwent rapid evolutionary changes in Passseriformes. The ORF in exons 4 and 5 was not conserved in most birds (*data not shown*). Further, we did not find any regions homologous to the chicken TAPAS RNA

in mammalian genomes at the sequence level. Exon 6 was the most conserved in avian species with 72-96% identity. The splice donor sites of exon 5 as well as the donor and acceptor sites of exon 6 are perfectly conserved in all species analyzed (*data not shown*).

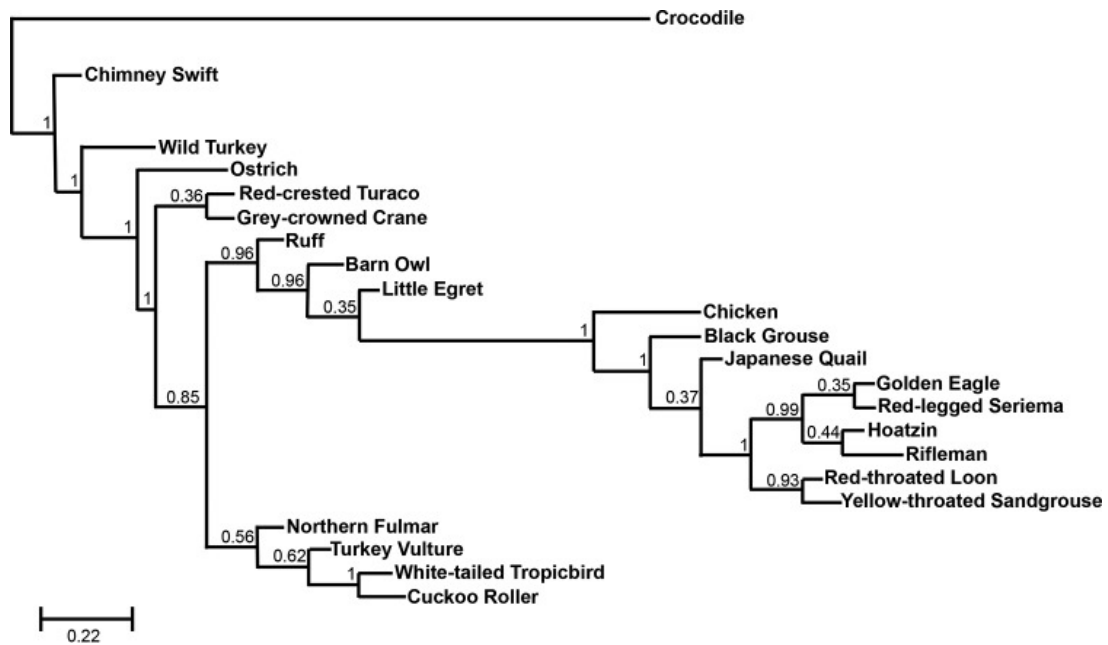


Figure 4.6 The TAPAS gene is conserved in avian species. Shown is phylogenetic analysis of exons 4 to 6 and the intervening introns of the TAPAS gene in several different avian species. Crocodile is depicted as an outgroup. The numbers at nodes of the tree are posterior probability values from a maximum-likelihood approach as described in Materials and Methods.

Discussion

In order to better understand tumor pathogenesis, we analyzed the distribution of ALV integration sites in chicken B-cell lymphomas using a high throughput sequencing strategy. We identified numerous clonally expanded proviral integrations in the *TERT* promoter region, associated with elevated *TERT* expression, suggesting these integrations may promote tumorigenesis (Justice et al., 2015b; Yang et al., 2007a). Previously, ALV subgroup J proviral integrations have been observed near *TERT* in chicken myeloid tumors (Li et al., 2014). Other studies have reported integrations of hepatitis B and human papilloma virus DNA in the *TERT* promoter region in human liver and cervical tumors, resulting in elevated *TERT* expression (Ferber et al., 2003). While sense and anti-sense retroviral activation have been previously reported (1,44), to the best of our knowledge this is the first report of retroviral activation of a bidirectional promoter-associated lncRNA. In this work we show that most of the proviral integrations in our tumors are in the antisense direction relative to *TERT* and are driving the over-expression of a novel lncRNA transcribed from a bidirectional promoter. The prevalence of integrations in the same orientation as the TAPAS RNA in multiple tumors suggests that over-expression of this transcript may have contributed to oncogenic transformation. We further characterize this novel transcript and show that in addition to being conserved amongst some avian lineages, expression of this transcript is correlated to *TERT* expression in various tissues and throughout development in chickens.

Deep sequencing of transcriptomes from various species, including humans, reveals the presence of a transcribed region upstream of *TERT* (*data not shown*). We found that in chickens this corresponds to an alternatively spliced, polyadenylated, antisense lncRNA, leading us to propose that the *TERT* promoter is bidirectional. Bidirectional promoters have been found to be pervasive in mammalian genomes with more than 10% of known protein coding genes being transcribed from such promoters (Orekhova and Rubtsov, 2013; Trinklein et al., 2004; Wakano et al., 2012). In humans, many lncRNAs are transcribed from bidirectional promoters. Approximately 65% of identified human lncRNAs originate within 2 kb of transcription start sites and almost all of these are transcribed in the antisense direction relative to the associated protein-coding gene (Sigova et al., 2013). Furthermore, bidirectional promoters are significantly more likely to regulate DNA repair genes and genes implicated in somatic cancers (Yang et al., 2007b).

The chicken *TERT* promoter has many features characteristic of bidirectional promoters such as the absence of a TATA box and high GC content. The most important feature of a bidirectional promoter seems to be the presence of binding sites for an Ets family transcription factor, GABP (Collins et al., 2007; Orekhova and Rubtsov, 2013). Interestingly, recurrent promoter mutations present in the human *TERT* promoter have been shown to introduce a de novo Ets binding site, which has been shown to drive *TERT* expression (Bell et al., 2015; Horn et al., 2013; Huang et al., 2013). It has

recently been shown that it is the specific Ets factor, GABP, that selectively binds the mutant *TERT* promoter (Bell et al., 2015). Thus, it is possible that because GABP selectively binds the promoter mutation, it may be inducing bidirectional transcription of an antisense transcript similar to what was observed in chickens. There is evidence for antisense transcription in the *TERT* promoter region in humans from deep sequencing data (*data not shown*). However, there is no conservation between this putative transcript and the chicken transcript. Further, RNA abundance in this region is much lower relative to *TERT* than what is observed in chickens.

Many lncRNAs are known to play a role in transcriptional regulation. For example, *XIST* acts in cis to recruit the polycomb repressive complex 2 (PRC2) to chromosome X, causing gene silencing (Zhao et al., 2008). *HOTAIR*, on the other hand, acts in trans to repress the expression of genes in the HoxD gene cluster (Gupta et al., 2010). It has been proposed that antisense lncRNAs transcribed from bidirectional promoters may be involved in regulation of the associated sense transcripts (Wakano et al., 2012; Wei et al., 2011). Such an arrangement may allow for tighter transcriptional regulation. Based on the correlation between *TERT* and lncRNA expression in normal tissues and tumors, we hypothesize that the lncRNA may be regulating *TERT* expression. However, it is also possible that *TERT* and lncRNA expression are only correlated due to mutual regulation by the same bidirectional promoter. This opens the possibility that the lncRNA may have an independent function apart from that of *TERT* regulation.

Given the important role of lncRNAs in transcriptional regulation, it comes as no surprise that many lncRNAs have been implicated in cancer and disease. With the advent of deep sequencing, many novel lncRNAs have been found to be associated with cancer pathogenesis. For example, *HOTAIR*, *MALAT1*, *PCAT-1*, *PCGEM1*, *TUC338* were reported to be oncogenes, while *GAS5*, *MEG3* and *PTENP1* were reported to be tumor suppressors (54). Since lncRNA functions range from cell growth to cancer development, they represent important biological players that merit further research.

In this study, we have identified a novel antisense lncRNA transcribed from the *TERT* promoter. We show that this lncRNA is up regulated in tumors and is implicated in tumorigenesis. Further characterization of the structural and functional motifs of this TAPAS RNA is required to better understand its mechanistic and functional role in cancer signaling. This will help elucidate possible gene regulatory mechanisms of lncRNAs, and may provide further insight into the role lncRNAs play in cancer pathogenesis.

Materials and Methods

Tumor induction. All of the B-cell lymphomas included in this study were rapid-onset lymphomas induced by either wild type (WT) or variants of LR-9 virus infections, in 10 day old chicken embryos, as described previously (Polony et al., 2003). LR-9 is an ALV subgroup A recombinant virus consisting of *gag*, *pol*, and *env* genes derived from UR2 associated virus and LTRs derived from ring-necked pheasant virus (Simon et al., 1987). Tumors were collected from primary bursal (B) tumors or metastasized liver (L) tumors. A1B was induced by Δ LR-9, with a deletion in the *gag* gene, causing increased splicing to downstream genes (Smith et al., 1997). Tumors C2B, C2L, C6L, C7B and C7L were induced by infection with LR-9 containing a silent mutation, G919A, which induces a higher incidence of rapid-onset lymphomas (Polony et al., 2003), probably due to increased readthrough and splicing to downstream genes (O'Sullivan et al., 2002). Tumor D2L was induced by WT LR-9 (Simon et al., 1987).

DNA and RNA isolation. Genomic DNA was prepared by standard proteinase K digestion followed by phenol-chloroform extraction as described previously (Justice et al., 2015a). RNA was extracted from tissue homogenates using RNA Bee extraction agent (Tel-Test, Inc, Friendswood, TX).

Reverse transcription, PCR amplification and Sanger sequencing. Total RNA was reverse transcribed with Maxima H minus reverse transcriptase (ThermoFisher Scientific), and either oligo(dT)18 primer and/or random hexamers following the manufacturers protocol. Tumor C2L was excluded from analysis due to poor RNA quality. All PCR primers were commercially synthesized (Integrated DNA Technologies, Inc.). 3'RACE was performed with the lock-docking oligo(dT) primer (Borson et al., 1992). PCR amplification was performed with Phusion® High-Fidelity DNA Polymerase following the manufacturer's instructions (New England Biolabs). Amplified fragments were gel purified using a gel extraction kit (Qiagen) and sequenced (Eurofins MWG Operon LLC).

Detection of splice variants and ALV-TAPAS RNA fusion transcripts.

Spliced variants were detected using primers in the antisense direction in *TERT* exon 1 and in exon 7 of TAPAS RNA (TGGCCTCGGCGTAGCAG, CAAATGGCTTGTCTGCATTTTCTTC). Fusion transcripts were amplified using a 5' primer positioned immediately before the viral splice donor or complementary to sequences in the R region of the viral LTR (TCAAGCATGGAAGCCGTCATAAAG, GCCATTTGACCATTCACCACATTG) and a set of 3' primers located throughout the TAPAS RNA gene sequence (CAAATGGCTTGTCTGCATTTTCTTC,

CCAAAGCCACGGCTTCCATGTTAGTATC,
TAAGGTGGAGAATAAGACATAATAATATGAGATGAG)

High throughput sequencing. DNA libraries for deep sequencing were prepared and analyzed as previously described (Justice et al., 2015a, 2015b). RNA-Seq libraries were prepared according to a previously published protocol (Nagalakshmi et al., 2010). The reads were aligned to the chicken genome (galGal4) using TopHat (Langmead et al., 2009a). Splice junctions enabled the determination of the orientation of spliced transcripts. Additional RNA-Seq data for analysis of tissue distribution and embryonic expression of TAPAS RNA and *TERT* in chickens were downloaded from the public sequence read archive (SRA) database (SRA Accession no. ERX697750 and DRX001564) (Leinonen et al., 2011). Abundance of transcripts (FPKM) was estimated and compared using Cufflinks (Cossio et al., 2012).

Quantitative PCR. Quantitative reverse transcription PCR was performed using iQ SYBR Green Supermix (BioRad) according to the manufacturer's protocol on a Biorad C1000 Thermal Cycler/CFX96 Real-Time System. Expression of TAPAS RNA in chicken tissue was measured using primers in exons 4 and 5 (CAGACTACTTTACCTCTTGACACAG, ATGGTGAGCCTTGTGTTGGC). *TERT* expression was measured using primers in exons 11 and 12 (AACATGAAATGCAAATTGACTGC, ACTGTCTGAAGGCTGTTGATCT). Expression was normalized to

expression of ribosomal protein L30 (RPL30) using exon junction primers (Yang et al., 2013). qPCR was performed in triplicate with each sample present in technical duplicates during each run. Results were normalized to normal bursa using the comparative C_T method.

Evolutionary analysis. The sequence encompassing exons 4-6 and intervening introns of chicken TAPAS gene was analyzed via BlastN against the whole genome shotgun database at the NCBI web site (<http://blast.ncbi.nlm.nih.gov/>). All high quality matches with an E-value lower than 10^{-40} were retrieved and further analyzed. Sequences were aligned and plotted via the maximum-likelihood method, by PhyML, utilizing the GTR substitution and aBayes branch support (Guindon et al., 2010).

**Chapter 5. A novel long non-coding RNA in the hTERT promoter region
regulates hTERT expression**

Summary

In this study we identify a novel antisense transcript in the human telomerase reverse transcriptase (hTERT) promoter region, suggesting that the hTERT promoter is bidirectional. This transcript, named hTERT antisense promoter-associated (hTAPAS) RNA, exhibits features of a long non-coding RNA (lncRNA). Through transcriptome analysis of several ENCODE cell lines and over 3,800 human cancer samples from The Cancer Genome Atlas, we determine that hTAPAS is a 1.6 kb long antisense RNA. hTAPAS transcription is initiated 167 nts upstream of the hTERT transcription start site, and it exists as both polyadenylated and non-polyadenylated transcripts in the nucleus and polyadenylated transcripts in the cytoplasm. Surprisingly, we observe that a large fraction of the hTERT polyadenylated transcript is localized in the nucleus, suggesting this might be an additional means of regulating the cellular abundance of hTERT protein. Similar to hTERT expression, hTAPAS exhibits poor expression in somatic cells and normal tissues from cancer patients. On the other hand, both hTAPAS and hTERT are expressed in immortalized B-cells and human embryonic stem cells. We show that hTAPAS expression inversely correlates with hTERT expression in many cancers, and its expression is not promoted by the hTERT promoter mutation (-66 nt from the hTERT transcriptional start site). Antisense-oligo mediated knockdown of hTAPAS results in an increase in hTERT expression. Conversely, ectopic overexpression of hTAPAS down regulates hTERT expression, suggesting a negative role in hTERT gene

regulation. These observations provide insights into the activation of hTAPAS as a novel player that negatively regulates hTERT expression and may be involved in telomere length homeostasis.

Introduction

Telomerase activity and telomere length have important implications in human disease and aging (Blasco, 2005); elevated telomerase activity has been detected in most human cancers (Garcia et al., 2007a; Kim et al., 1994; Shay and Wright, 2011). Expression of human telomerase reverse transcriptase (*hTERT*), the enzymatic component of telomerase, is tightly regulated at the transcriptional level through epigenetic modifications in the promoter region (Zhu et al., 2010), as well through alternative splicing (Cevik et al., 2015; Huang et al., 2013; Withers et al., 2012). Abundance of hTERT is a rate-limiting step in modulating telomerase activity (Cong et al., 2002). While normally expressed in the germline and stem cells, up-regulated hTERT is essential for the continual proliferation and long-term viability of cells in many cancers (Blasco, 2005; Shay and Wright, 2011). Recurrent mutations in the *hTERT* promoter region, at -66 or -88 nts relative to the *hTERT* transcriptional start site (TSS), are among the most common somatic mutations in many types of cancer, including melanomas, glioblastoma multiforme, hepatocellular carcinomas, and bladder cancers (Horn et al., 2013; Huang et al., 2013; Killela et al., 2013).

The *hTERT* promoter has many features characteristic of bidirectional promoters, such as the absence of a TATA box and high GC content (Renaud et al., 2003), as well as binding sites for the ETS transcription factor GABP (Collins et al., 2007; Orekhova and Rubtsov, 2013). Recent data demonstrate that the mutant *hTERT* promoter can be bound and activated

by GABP (Bell et al., 2015), suggesting it may induce bidirectional transcription of an antisense transcript.

We have previously reported retroviral activation of an antisense transcript upstream of *TERT* in chicken B-cell lymphomas, named *TERT* antisense promoter-associated (*TAPAS*) RNA (Nehyba et al., 2016). Truncated *TAPAS* RNA is up regulated in chicken B-cell lymphomas. Here, we identify and characterize a human *TAPAS* (*hTAPAS*) RNA transcript in many types of human cancer and in several cell lines, using bioinformatics and experimental analyses. The *hTAPAS* transcript spans approximately 1.6 kb and has a single unspliced exon. Further, the absence of any conserved large open reading frames (ORFs) with protein domain homology suggests that this transcript is a previously unidentified long non-coding RNA (lncRNA).

We observe that *hTAPAS* expression is inversely correlated with *hTERT* expression in human cancers in The Cancer Genome Atlas (TCGA).

Furthermore, we do not observe any activation of *hTAPAS* expression with the *hTERT* promoter mutation in a mini-gene construct expressed in HEK-293 cells. Knocking down of *hTAPAS* via antisense-oligonucleotides increases *hTERT* expression, and ectopic overexpression of *hTAPAS* down regulates *hTERT* expression. This suggests that *hTAPAS* is involved in negatively regulating *hTERT* expression. Our work confirms the existence of an antisense lncRNA, *hTAPAS*, upstream of *hTERT* that exhibits negative regulation of *hTERT* expression. Moreover, we observe that nearly half of

the *hTERT* transcript is localized in the nucleus, suggesting this might serve as an additional way of regulating the cellular abundance of hTERT protein.

Results

An antisense transcript is expressed upstream of *hTERT*

Deep sequencing of transcriptomes from chickens revealed an antisense transcript upstream of *TERT*, which is an alternatively spliced, polyadenylated, lncRNA named *TAPAS* (Nehyba et al., 2016); this suggests that the *TERT* promoter is bidirectional. To determine whether a similar transcript is expressed in humans, we first examined RNA sequencing data from the ENCODE Consortium (ENCODE Project Consortium et al., 2012). An antisense RNA in the *hTERT* promoter region was readily observed in two different human B-cell tumor lines (GM12878 and OCI-LY7) (**Figure 5.1 and 5.2**). Moderate expression levels of this transcript were also observed in a human embryonic stem cell line (H1-hESC), and to a lesser extent in a human hepatocellular carcinoma cell line (HepG2) (Figure S1). However, this transcript was not detected in HeLa (human cervical carcinoma) and K562 (human leukemic) cell lines (**Figure 5.2**). All of these cell lines expressed hTERT (**Figure 5.2**).

To better determine the 5' end of *hTAPAS*, we used 5' cap analysis gene expression (CAGE) data from the ENCODE Consortium for the GM12878 cell line (ENCODE Project Consortium et al., 2012). 5' CAGE data specifically identifies the 5' ends of capped transcripts, indicating TSSs. We observe a strong signal starting 167 nt upstream of the annotated *hTERT* TSS and in the opposite orientation (**Figure 5.1**). Taken together, the RNA-seq and 5' CAGE data from the ENCODE Consortium suggest that an

antisense transcript, hereafter referred to as the *hTERT* antisense promoter-associated (*hTAPAS*) RNA, was detected in the *hTERT* promoter region in human embryonic stem cells and in some but not all cancer cell lines. Interestingly, the *hTERT* promoter mutations found in several tumor types, at -66 or -88 nts relative to the *hTERT* TSS (Horn et al., 2013; Huang et al., 2013; Killela et al., 2013), lie in between the TSS for *hTERT* and *hTAPAS* (**Figure 5.1**).

From TCGA (Chang et al., 2013) we downloaded normal and tumor tissue RNA-seq read alignments of approximately 3800 samples from 8 different cancer types, (**Table 5.1**), and reconstructed transcripts de novo to identify annotated and novel transcripts (**Figure 5.3, 5.4A**). In addition to identifying *hTERT* transcript isoforms, we identified an approximately 1.6 kb long *hTAPAS* transcript at the locus predicted by the ENCODE data, in some tumors (**Figure 5.4A**). Low expression levels were observed for both *hTAPAS* and *hTERT* in corresponding normal tissues from cancer patients (**Figure 5.4B**).

Finally, to experimentally validate expression of this transcript, we performed RT-PCR, using RNA from HEK-293, HeLa, SNU-449 and SNU-475 cells and again identified an approximately 1.6 kb long *hTAPAS* RNA (**Figure 5.1**). Based on these results and the ENCODE Consortium RNA-seq and CAGE data, we predict that *hTAPAS* is about 1.6 kb long with a single unspliced exon and is transcribed starting at 167 nt upstream of the *hTERT* TSS.

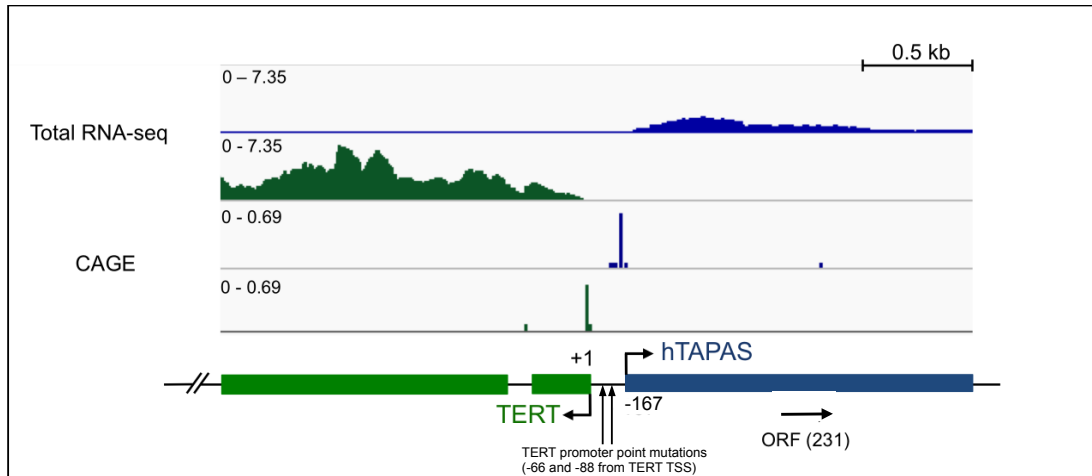


Figure 5.1 An antisense RNA, named *hTAPAS*, is expressed in the *hTERT* promoter region. Normalized and stranded RNA-seq (Bedgraph) transcription coverage for *hTERT* (green) and *hTAPAS* (blue) expression, as well as corresponding CAGE start sites, are depicted for the human B-cell cell line (GM12878). The schematic below denotes the first 2 exons of *hTERT* and an approximately 1.6 kb long *hTAPAS* gene, located 167 nts upstream of the *hTERT* transcriptional start site. Arrowheads represent sites of point mutations in the *hTERT* promoter, located 66 and 88 nts upstream of the *hTERT* transcriptional start site.

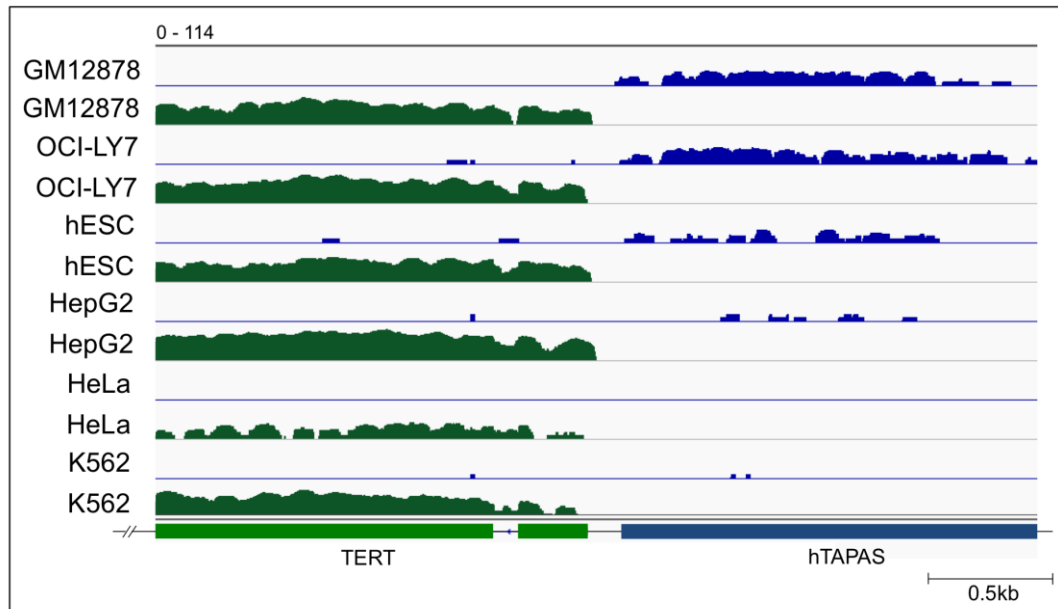


Figure 5.2 Normalized and stranded RNA-seq (Bedgraph) transcription coverage for *hTERT* (green) and *hTAPAS* (blue) expression are depicted for different ENCODE cell lines. All plus strand (blue) and minus strand (green) track signals are depicted on a log-scale (0 - 114). Represented data are from the ENCODE Consortium.

Cancer type	Metastatic	Normal Tissue	Primary Tumor	Recurrent Tumor
Bladder (BLCA)	0	19	414	0
Liver (LIHC)	0	50	374	0
Glioma (LGG)	0	0	512	18
Glioblastoma (GBM)	0	0	156	13
Breast (BRCA)	7	123	1104	0
Lymphoma (DLBC)	0	0	48	0
Prostate (PRAD)	1	50	502	0
Melanoma (SKCM)	368	0	103	0

Table 5.1 Number of RNA-seq samples from TCGA for each cancer type investigated for *hTERT* and *hTAPAS* expression.

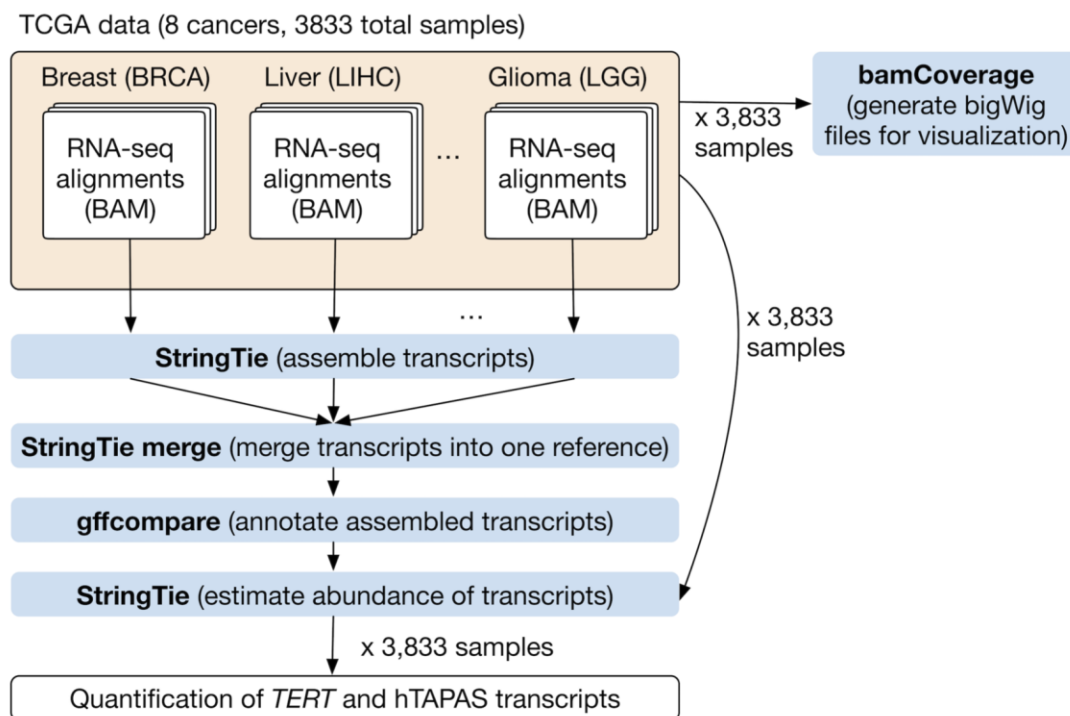


Figure 5.3 Pipeline of TCGA RNA-seq data analysis. Each STAR-aligned BAM file of RNA-seq data was downloaded for eight cancers (BRCA, BLCA, DLBC, GBM, LGG, LIHC, PRAD, SKCM) and used as input to StringTie without a reference transcript file to assemble transcripts. All assembled transcript files were merged using StringTie --merge with an hg38 reference GTF file of known transcripts to produce a single reference GTF file that was annotated using gffcompare. Then, each BAM file was used to quantify annotated transcripts using StringTie.

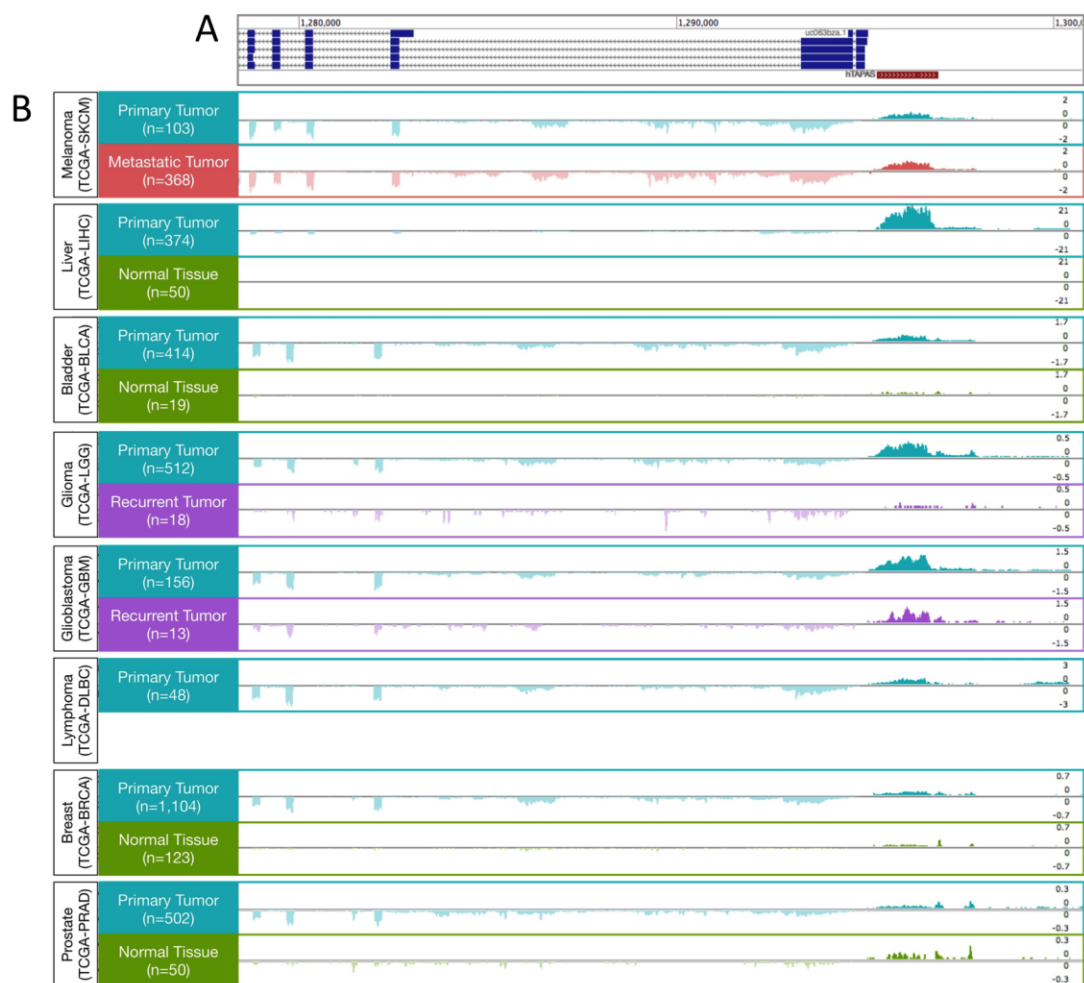


Figure 5.4 Gene models and expression of hTERT and hTAPAS in tumors. (A) Browser track with gene models shows UCSC hTERT gene models (dark blue) and StringTie-called hTAPAS gene model (red). (B) Expression of hTAPAS (plus strand, positive values) and hTERT (minus strand, negative values) in tumor and normal tissues of eight cancer types. Signal shown is FPKM summed over all samples.

***hTAPAS* exhibits features of a long non-coding RNA**

There is a small open reading frame (ORF) in the middle of *hTAPAS* RNA (231 nt) (**Figure 5.1**). However, this ORF is not conserved and lacks any protein domain homology, implicating this transcript as a lncRNA. Unlike *hTERT* exons, the *hTAPAS* transcribed locus does not exhibit evolutionarily conserved elements as determined by phastCons scores (**Figure 5.5**) (Siepel et al., 2005). Positions representing the *hTERT* exons had a mean phastCons score of 0.204, *hTAPAS* locus positions had a mean score of 0.048, and random positions had a mean score of 0.047. *hTAPAS* also shows evidence of faster evolution compared to neutral evolution rates as determined by phyloP scores (**Figure 5.5**) (Cooper et al., 2005). phyloP scores were aggregated for *hTAPAS*, *hTERT* exons, and random positions. Positions representing *hTERT* exons had a mean phyloP score of 0.089 (positive scores indicate conservation), the *hTAPAS* locus had a mean score of -0.081, and random positions had a mean score of -0.1236 (negative scores indicate faster evolution compared to a neutral evolution rate). Thus, similar to the previously reported chicken *TAPAS* transcript, *hTAPAS* lacks ORFs with known protein domain homology and exhibits no protein-coding potential based on phyloCSF scores (**Figure 5.6**) (Lin et al., 2011). Based on these features, we propose the *hTAPAS* transcript to be a lncRNA.

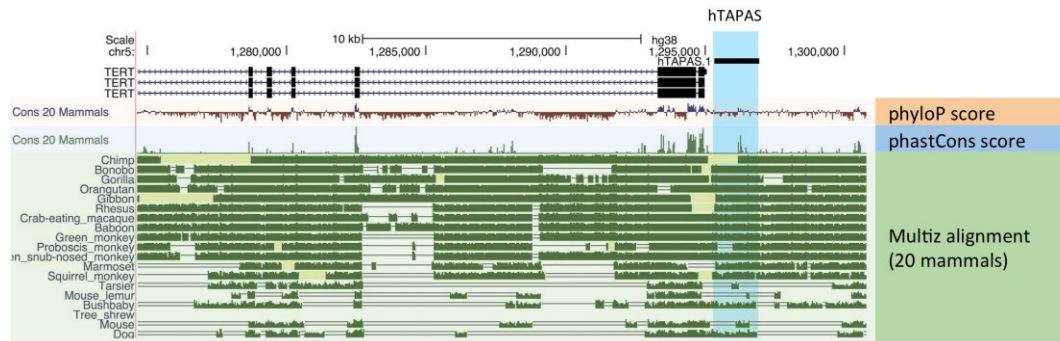


Figure 5.5 hTAPAS does not exhibit conservation. PhyloP (top track) and phastCons (middle track) scores are displayed across the hTERT and putative hTAPAS genomic locus. Positive phyloP scores indicate conservation, while negative phyloP scores indicate faster rates of evolution compared to the neutral rate. A phastCons score of 0 indicates no conservation, while scores closer to 1 indicate higher conservation. Multiz alignment across 20 mammalian genomes (bottom track) suggest conservation among primates, but not other mammals. The hTAPAS locus displays more negative phyloP scores, indicating faster evolution rates than neutral, and lower phastCons scores, indicating low conservation, compared to hTERT exons.

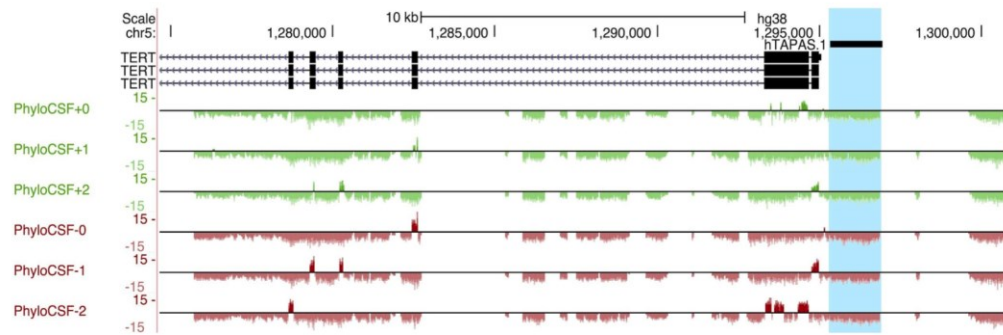


Figure 5.6 hTAPAS does not exhibit protein-coding potential. PhyloCSF (codon substitution frequencies) scores based on alignment of 58 mammalian genomes are displayed for all six open reading frames across the hTERT and putative hTAPAS genomic locus. Positive values indicate that the region is likely to represent a conserved coding region, while negative values represent low protein-coding probability. hTERT exons overlap regions of positive phyloCSF scores, while the hTAPAS locus does not.

Sub-cellular localization of *TAPAS* and *hTERT*

To determine the sub-cellular localization of *hTAPAS* RNA, RNA-seq data was analyzed for polyA⁺ and polyA⁻ RNAs from cytoplasmic and nuclear fractions. We found evidence for both polyadenylated (polyA⁺) and non-polyadenylated (polyA⁻) populations of *hTAPAS* in a human B-cell line (GM12878) (**Figure 5.7A**). The polyadenylated transcript was present in both nuclear and cytoplasmic fractions of the B-cell line. On the other hand, the non-polyadenylated (polyA⁻) transcript was absent from the cytoplasm. This suggests that a fraction of *hTAPAS* lncRNA is exported from the nucleus to the cytoplasm after being polyadenylated.

Surprisingly, a substantial fraction of the *hTERT* transcript was also observed in the nucleus of GM12878 cells (**Figure 5.7A**). A similar sub-cellular distribution of *hTERT* RNA was also seen in H1-hESC, HepG2, HeLa and K562 cell lines (**data not shown**). This was an unexpected observation since *hTERT* is a protein-coding gene and would be expected to be localized mostly in the cytoplasm. As expected, the control GAPDH mRNA was predominantly polyadenylated and cytoplasmic (**Figure 5.7A**).

We observed similar sub-cellular localization of *hTERT* and *hTAPAS* RNAs in both the nucleus and cytoplasm, after sub-cellular fractionation of HEK-293 cells and RNA quantitation by qRT-PCR (**Figure 5.7B**). Again, GAPDH mRNA was mainly observed in the cytoplasmic fractionation in this experiment (**Figure 5.7B**). Further, we tested the possibility that *hTAPAS* RNA might be sequestering *hTERT* RNA in the nucleus. However, knocking

down of *hTAPAS* via antisense oligonucleotides did not alter the cellular localization of *hTERT* in HEK-293 cells (**data not shown**).

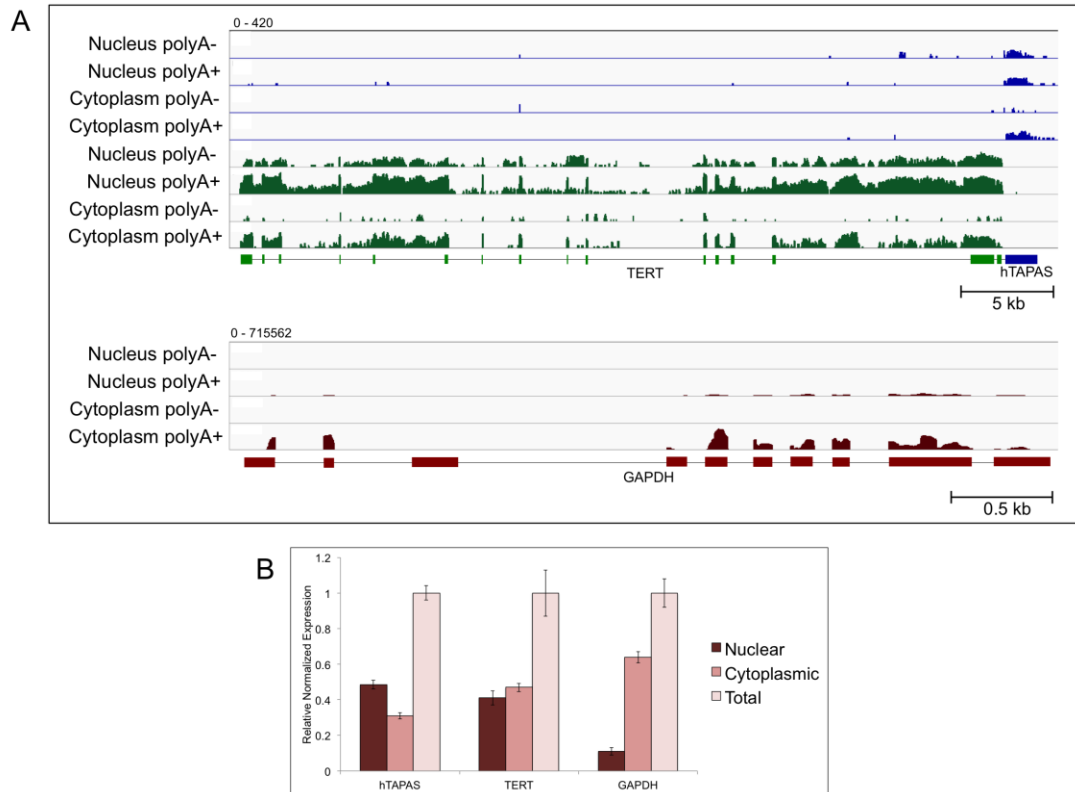


Figure 5.7 Sub-cellular localization of *hTERT* and *hTAPAS* RNAs (A) Normalized and stranded RNA-seq (Bedgraph) transcription coverage for *hTERT* (green) and *hTAPAS* (blue) expression are depicted for the polyadenylated (polyA+) and non-polyadenylated (polyA-) transcripts in the nuclear and cytoplasmic fractions, from the human B-cell cell line (GM12878). All plus strand (blue) and minus strand (green) track signals are depicted on a log-scale (0 - 420). Transcription coverage for corresponding transcripts of GAPDH (red) are depicted as a control, on a log scale (0 - 715562). Represented data are from the ENCODE Consortium. (B) Abundance of *hTERT*, *hTAPAS* and GAPDH transcripts in the nucleus, cytoplasm and total cellular fractions determined by qRT-PCR after sub-

cellular fractionation of HEK-293 cells. Expression levels in the nuclear and cytoplasmic fractions are plotted relative to total.

***hTAPAS* expression negatively correlates with *hTERT* expression in cancer patients**

Antisense lncRNAs transcribed from bidirectional promoters have been known to be involved in regulation of the associated sense transcripts (Wakano et al., 2012; Wei et al., 2011). We thus investigated the relationship between the expression of *hTAPAS* and *hTERT* in TCGA database (Chang et al., 2013). First, we analyzed *hTAPAS* and *hTERT* expression in primary tumor samples from eight different types of cancer by RNAseq (**Table 5.1**).

hTERT RNA detection ranged from 3 to 50% in the different types of primary tumors (**Figure 5.8A and 5.9**). Lymphomas had the highest percentage of patients expressing detectable levels of *hTERT*, followed by bladder and liver cancers, while prostate cancers had very few patients expressing detectable levels of *hTERT*. Tumor samples expressing detectable *hTAPAS* RNA ranged from 1% (prostate) to 20% (glioblastoma) (**Figure 5.8B**). In addition a few of the liver tumors expressed high levels of *hTAPAS* (**Figure 5.4B**).

In the primary tumor samples in which at least one of these transcripts was detected, fewer than 20% expressed both *hTAPAS* and *hTERT* RNAs (**Figure 5.9**). Co-expression of *hTERT* and *hTAPAS* RNAs was nearly undetectable in breast and prostate cancers (**Figure 5.8C and 5.9**). The highest levels of co-expression (4-8%) of *hTERT* and *hTAPAS* RNAs were observed in primary tumors of lymphomas, bladder and liver cancers (**Figure**

5.8C). This pattern is suggestive of an antagonistic relationship between *hTAPAS* and *hTERT* expression.

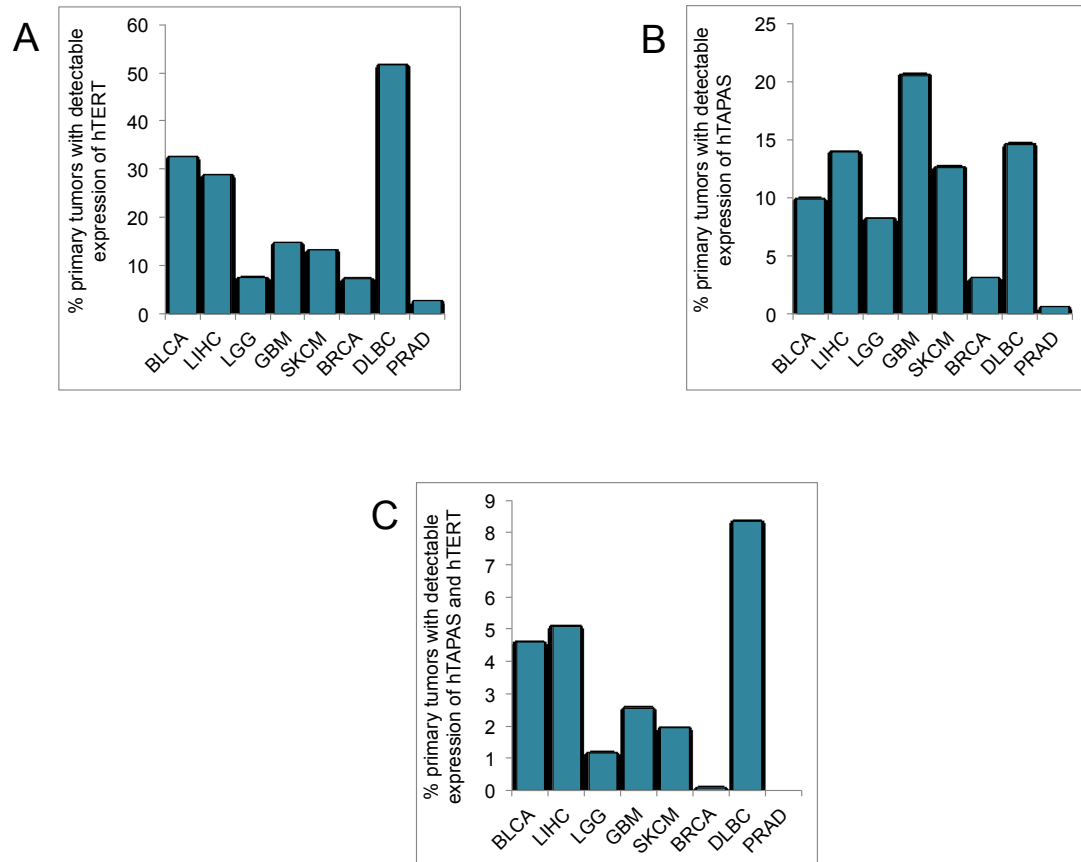


Figure 5.8 Proportion of primary tumor samples with detectable expression of (A) hTERT or (B) hTAPAS or (C) both, based on the TCGA RNA-seq data analysis.

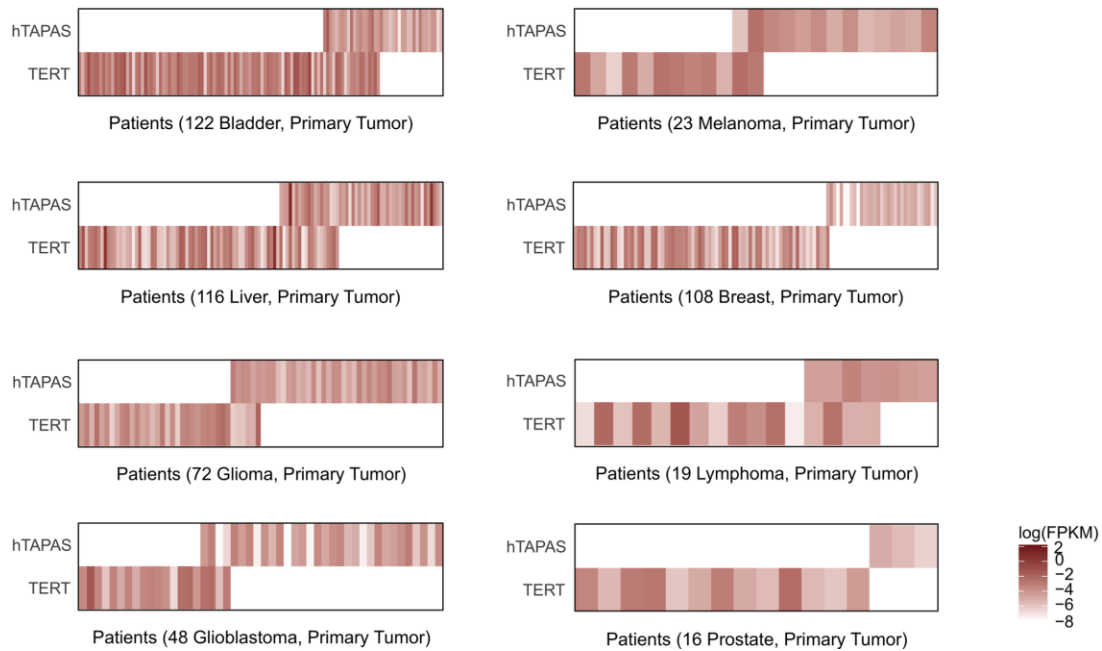


Figure 5.9 *hTERT* and *hTAPAS* expression are inversely correlated in primary tumors. RNA-seq data samples from TCGA, which have detectable expression levels for either *hTERT* or *hTAPAS* were analyzed in the 8 different cancers (melanomas, gliomas, glioblastomas, hepatocellular carcinomas, bladder, lymphoma, prostate and breast cancers). A majority of these tumor samples express either *hTERT* or *hTAPAS*, but not both. The heat-maps depict corresponding expression levels of *hTAPAS* and *hTERT* among individual patient samples from the eight different cancers. Expression is represented as log-transformed FPKM values (white indicates undetectable expression).

Knockdown of *hTAPAS* up regulates *hTERT* expression in HEK-293 cells

To test whether loss of *hTAPAS* RNA can alter *hTERT* expression, targeted knockdown of *hTAPAS* was performed in a human kidney (HEK-293) cell line, using oligonucleotides antisense to the *hTAPAS* transcript (570 to 594 nts from the *hTAPAS* TSS). Cells transfected with a scrambled oligonucleotide showed no significant change in either *hTAPAS* or *hTERT* expression; however, transfection with an *hTAPAS* antisense oligonucleotide (ASO) resulted in a 4-fold knockdown of *hTAPAS* and a significant 1.8-fold increase in *hTERT* expression (using primers in *hTERT* exons 13 and 14) (**Figure 5.10**). We also measured expression of the *hTERT* transcript with primers in the catalytic reverse transcriptase domain (exons 7 and 8). Similarly, we detected a 2.1-fold increase in expression of these *hTERT* transcripts. These results suggest that *hTAPAS* lncRNA functions to negatively regulate *hTERT* expression.

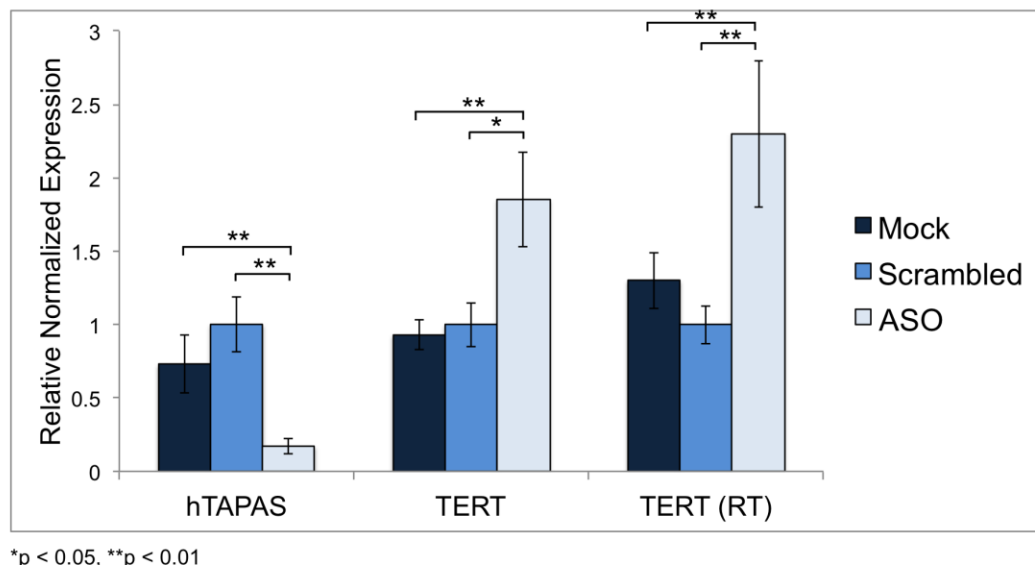


Figure 5.10 *hTAPAS* knockdown in HEK-293 cells results in increased *hTERT* expression. *hTERT* and *hTAPAS* expression was determined by qRT-PCR in HEK-293 cells transfected with mock, scrambled oligos, and antisense oligonucleotides (ASO) targeted to *hTAPAS*. Use of ASO resulted in a 4-fold knock down of *hTAPAS* expression, relative to scrambled. *hTERT* expression was analyzed using primers spanning its exons 13-14 or exons 7-8 (in the RT domain).

***hTAPAS* expression does not correlate with the *hTERT* promoter mutation**

The mutant *hTERT* promoter can be bound and activated by GABP (Bell et al., 2015), suggesting it may induce bidirectional transcription of an antisense transcript (Collins et al., 2007; Orekhova and Rubtsov, 2013). Therefore, we asked whether this *hTERT* promoter mutation affects *hTAPAS* expression. We first analyzed RNA-seq data from 8 hepatocellular carcinoma cell lines, 4 with and 4 without the *hTERT* promoter mutation located 66 nt upstream of the *hTERT* TSS (**Figure 5.11**) (Cevik et al., 2015). Consistent with previous reports (Bell et al. 2015), *hTERT* exhibited an average of approximately 2-fold higher expression in the cells with the promoter mutation, relative to the cells with the wild-type promoter (**Figure 5.11**). In contrast, no obvious difference in expression of *hTAPAS* was apparent between cell lines with the absence or presence of the point mutation (**Figure 5.11**). Furthermore, in cells with a wild-type *hTERT* promoter, *hTERT* and *hTAPAS* had similar expression levels; however, in cells with the presence of the point mutation, *hTERT* expression was approximately 2 to 4 fold higher than *hTAPAS* expression (**Figure 5.11**).

Next, we generated constructs to ectopically express the *hTERT* promoter region, with a luciferase reporter in lieu of *hTERT* (**Figure 5.12A**). We introduced the *hTERT* promoter point mutation (at -66 nt from the *TERT* TSS) into this construct to determine its influence on luciferase activity (as an indicator of *hTERT* expression) and *hTAPAS* expression in HEK-293 cells.

We observed that luciferase activity was up regulated approximately 2-fold by the presence of the point mutation (**Figure 5.12B**), consistent with previously reported effects of the point mutation on *hTERT* expression. However, *hTAPAS* expression was not altered by the presence of the point mutation (**Figure 5.12C**). Moreover, in the ENCODE cell line HepG2, which has an *hTERT* promoter mutation, we observed *hTERT* expression but poor *hTAPAS* expression levels (**Figure 5.2**). Taken together, our data suggest that the *hTERT* promoter mutation significantly increases expression of *hTERT* but not of *hTAPAS*.

As a consequence of transfecting these luciferase constructs into HEK-293 cells, *hTAPAS* was ectopically over-expressed approximately 70-fold relative to endogenous levels (**Figure 5.12C**). *hTAPAS* over-expression resulted in down regulation of endogenous *hTERT* expression approximately 2-fold in HEK-293 cells. (**Figure 5.12D**). This expression pattern is consistent with our observations of *hTAPAS* knockdown resulting in increased *hTERT* expression levels (**Figure 5.10**). Therefore, these observations support our proposal that *hTAPAS* is involved in negatively regulating *hTERT* expression.

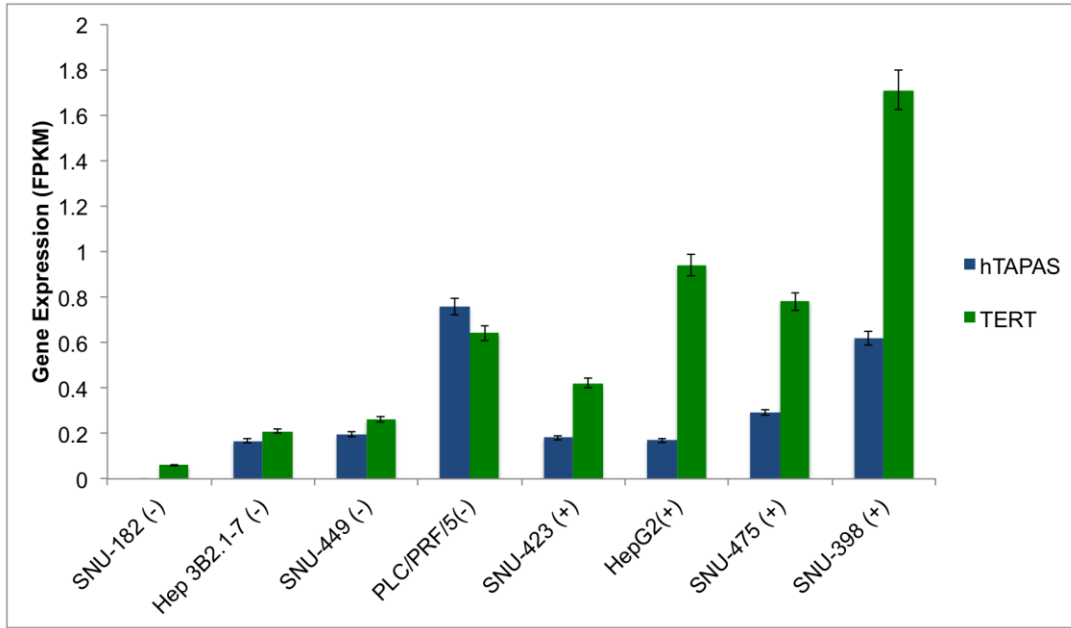


Figure 5.11 hTAPAS and hTERT expression (FPKM) in 8 different human liver cancer cell lines, with the absence (-) or presence (+) of hTERT promoter mutation (-66 nt from hTERT TSS).

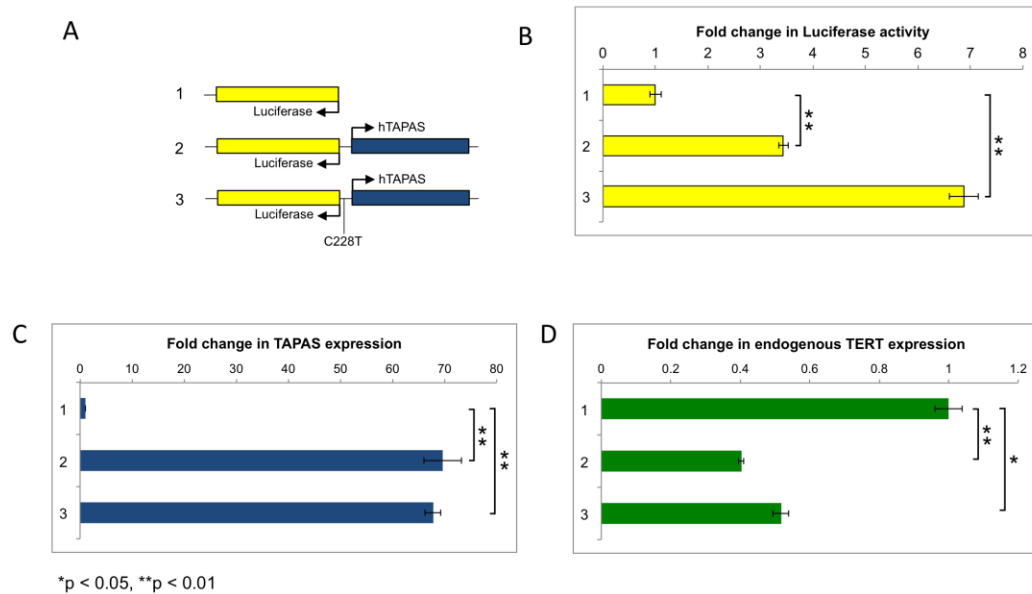


Figure 5.12 Presence of the hTERT promoter mutation does not alter hTAPAS expression. (A) Schematic representation of an empty control vector (1), and hTAPAS insert with a wild-type (2) or mutant (3) hTERT promoter region, in a luciferase reporter vector. (B) Fold-change in luciferase activity observed for the different constructs relative to empty vector. (C) Fold-change in hTAPAS or (D) endogenous hTERT expression levels observed for the different constructs relative to empty vector.

Discussion

Telomerase activity in somatic cells is limited by the availability of hTERT protein and thus expression of *hTERT* is tightly regulated (Lindsey et al., 2014). At the transcriptional level, *hTERT* is regulated by many transcription factors, such as c-MYC, ETS, and SP1/SP3, as well as through epigenetic modifications in the promoter region (Huang et al., 2013). Extensive alternative splicing of the *hTERT* transcript has also been shown to generate inactive and dominant negative *hTERT* variants that decrease telomerase activity (Cevik et al., 2015; Huang et al., 2013; Withers et al., 2012). In this work, we identify a novel antisense lncRNA in the *hTERT* promoter region, *hTAPAS*, which functions as an additional regulator of *hTERT* expression. We also observe that the *hTERT* polyadenylated transcript is predominantly localized in the nucleus, suggesting this might be an additional means of regulating the cellular abundance of catalytic hTERT by limiting its translation.

Many lncRNAs play a role in transcriptional regulation (Tsai et al., 2010). For example, XIST acts in cis to recruit the polycomb repressive complex 2 (PRC2) to chromosome X, causing gene silencing (Zhao et al., 2008), while HOTAIR acts in trans to repress the expression of genes in the HoxD gene cluster (Gupta et al., 2010). It has been proposed that antisense lncRNAs transcribed from bidirectional promoters may be involved in regulation of their associated sense transcripts (Gupta et al., 2010; Wakano et al., 2012). Such an arrangement may allow for tighter transcriptional

regulation. A similar regulatory module was observed for the *ANRIL* lncRNA and the antisense p15 gene (Kotake et al., 2011). We observe that *hTAPAS* over-expression can down-regulate *hTERT* expression in trans (**Figure 5.12**). This affect could be due to promoter-occlusion in the *hTERT* bidirectional promoter region. We also observe that knocking down *hTAPAS* up regulates *hTERT* expression (**Figure 5.10**). This therefore, suggests a role of the *hTAPAS* RNA in the transcriptional regulation of *hTERT*. *hTAPAS* might recruit epigenetic machinery to regulate *hTERT* expression, similar to other well-studied lncRNAs (Huarte, 2015). Furthermore, a fraction of polyadenylated *hTAPAS* transcripts are exported to the cytoplasm (**Figure 5.7**). This suggests, that in addition to negative regulation of *hTERT* expression, *hTAPAS* might be involved in other cellular functions in the nucleus and/or cytoplasm.

Shortening of telomeres limits the replicative potential of most primary human cells and serves a tumor-suppressive function. Conversely, telomerase expression in the germline, in somatic cells during early embryogenesis, and in cancer cells promotes telomere length homeostasis (Bodnar et al., 1998; Harley et al., 1990; Kim et al., 1994; Ulaner and Giudice, 1997; Vaziri and Benchimol, 1998; Wright et al., 1996). The preferential elongation of short telomeres and maintenance of telomere length homeostasis requires a low cellular concentration of telomerase (Cristofari and Lingner, 2006). This is achieved via expression of *hTERT* over a limited range, approximately 500 hTERT molecules per cell (Cristofari

and Lingner, 2006; Xi and Cech, 2014). Telomere length regulation and telomerase expression in multicellular organisms is thought to have evolved by opposing selective pressures to suppress tumor formation on one side, while not promoting premature cellular senescence in highly proliferative tissues on the other side. Similar to *hTERT* expression, *hTAPAS* expression is observed in the germline, cancer cell lines and tumor samples, and is absent in normal tissues. Since *hTAPAS* is not expressed in normal tissues of cancer patients, it does not appear to be involved in negatively regulating *hTERT* expression in somatic cells. Therefore, it is possible that as a regulator of *hTERT* expression, *hTAPAS* might facilitate the maintenance of *hTERT* expression within the narrow range required for telomere length homeostasis in cancer cells and stem cells.

We have previously reported the presence of numerous clonally expanded integrations of the avian leukosis virus (ALV) in the *TERT* promoter region in chicken B-cell lymphomas, associated with slightly elevated *TERT* expression (Justice et al., 2015b; Yang et al., 2007a). We have also shown that proviral integrations in these tumors are in the antisense orientation relative to *TERT* and are driving the over-expression of chicken *TAPAS* lncRNA (Nehyba et al., 2016). In this work we find evidence for a similar upstream antisense transcript in the *hTERT* promoter region. Similar to elevated expression levels of chicken *TAPAS* in B-cell lymphomas, we observe that *hTAPAS* also exhibits high expression levels in immortalized human B-cell lines (**Figure 5.1 and 5.2**). Interestingly, the *hTERT* promoter

region is also reported to be the most common integration site for hepatitis B virus integrations (HBV) in hepatocellular carcinomas (Buendia and Neuveut, 2015; Ding et al., 2012; Toh et al., 2013; Zhao et al., 2016). Approximately 20% of hepatocellular carcinomas have HBV integrations in the *hTERT* promoter region (Buendia and Neuveut, 2015). These integrations near *hTERT* are thought to confer an early clonal advantage during chronic HBV infection. Moreover, HBV integrations lead to up-regulated *hTERT* expression and are associated with poorer survival rates in infected patients with these integrations (Zhao et al., 2016). Of note, the frequency of *hTERT* promoter point mutations is significantly lower in hepatocellular carcinomas bearing HBV integrations in the *hTERT* promoter region (Buendia and Neuveut, 2015).

Interestingly, nearly all of the reported ALV and HBV integrations in the *hTERT* promoter region are present within the 5' end of chicken *TAPAS* and *hTAPAS*, respectively (**Figure 5.13**). The majority of the ALV integrations in the chicken *TERT* promoter region are in the same transcriptional orientation as *TAPAS* (Nehyba et al., 2016). In contrast, HBV integrations are in a mixed transcriptional orientation with respect to *hTAPAS* (Buendia and Neuveut, 2015; Ding et al., 2012; Toh et al., 2013; Zhao et al., 2016). The prevalence of viral integrations within *TAPAS* in multiple types of cancers in different organisms like chickens and humans suggests that these integrations are selected for during oncogenesis. These viral integrations might confer a proliferative advantage and make cells predisposed to

oncogenic transformation. Since *hTAPAS* negatively regulates *hTERT* expression and HBV integrations would likely disrupt *hTAPAS* expression, this could promote *hTERT* expression. These observations, therefore, suggest functional implications of *hTAPAS* in tumorigenesis.

Like human telomerase expression, chicken telomerase is also down regulated in most somatic tissues (Delany and Daniels, 2004). Furthermore, chicken telomeres shorten with age, and telomerase activity is important for oncogenesis (Delany et al., 2000). In contrast, the mouse TERT enzyme is active in normal somatic cells (Hackett and Greider, 2002). This discrepancy between humans and mice is important because telomerase activation is a critical step in the human oncogenic process, with aberrant telomerase activation seen in most human cancers (Garcia et al., 2007a; Shay and Wright, 2011). In contrast to the human and chicken *TERT* promoter regions, there is lack of any expression in the mouse *TERT* promoter region. This absence of a *TAPAS* transcript might be a means of facilitating increased TERT expression in mice cells. Therefore, chicken serves as an advantageous model over mouse, to study oncogenic events of *TERT* signaling and transcriptional regulation.

LncRNAs as well as bidirectional transcription have arisen as novel players in tumorigenesis, warranting a need for further research of their role in regulating cancer development. The mutant *hTERT* promoter can be bound and activated by GABP, an ETS transcription factor (Bell et al., 2015), which has been shown to induce bidirectional transcription in other loci (Collins et al., 2007; Orekhova and Rubtsov, 2013). However, we observe no

association of *hTAPAS* expression with the presence of a point mutation, in transfected HEK-293 cells or liver cancer cell lines (**Figure 5.11 and 5.12**). We observe *hTERT* and *hTAPAS* co-expression among 8-16% of cancers that have a high frequency of *hTERT* promoter mutations (liver cancers, bladder cancers, melanomas, gliomas and glioblastomas) (**Figure 5.9**). This suggests that even though there is no association of *hTAPAS* expression with the presence of a point mutation in cell lines, the same might not be true in tumor samples. Additionally, we observe *hTAPAS* expression in the absence of *hTERT* expression in TCGA samples (**Figure 5.9**), but not in cell lines (**Figure 5.1 and 5.2**). This variation in the *hTAPAS* expression profile between TCGA samples and cell lines could be due to the differences that exist in the genomic features and microenvironments between cell lines and tumor samples (Domcke et al., 2013; Goodspeed et al., 2016).

Reactivated *hTERT* expression is reported in 90% of all human cancers (Garcia et al., 2007a; Shay and Wright, 2011). However, through our analysis from the TCGA we detect *hTERT* expression in only 3-50% of analyzed samples (**Figure 5.8**). While previous studies have used qRT-PCR to observe *hTERT* transcript levels, our analysis involves the use of total RNA-seq data from primary tumors. The detection and quantification sensitivity of RNA-seq is dependent on the read depth. Even at an above average coverage of 100 million reads, RNA-seq suffers from greatly reduced quantification and detection sensitivity compared to qRT-PCR (Everaert et al., 2017). Therefore, the differences in our observations

compared to previous work can be explained by the method used for detection of *hTERT* expression.

Activation of *hTERT* expression is a crucial step in the progression of many cancers, and understanding the molecular mechanisms of such activation is important for understanding oncogenesis. Recent genome-wide analyses have highlighted that somatic point mutations in the *hTERT* promoter are among the most common mutations in human cancer and are known to drive *hTERT* expression (Chiba et al., 2015). The *hTERT* promoter region is also the site of viral integrations in both chicken (ALV) and human (HBV) tumors (**Figure 5.13**). We observe that in many tumors there is activation of a previously uncharacterized antisense transcript in the *hTERT* promoter region, which we call *hTAPAS*. This lncRNA negatively regulates *hTERT* expression and, thus, is implicated in oncogenesis. Further characterization of the structural and functional motifs of this lncRNA is required to better understand its mechanistic and functional role in cancer signaling and its role in cancer pathogenesis.

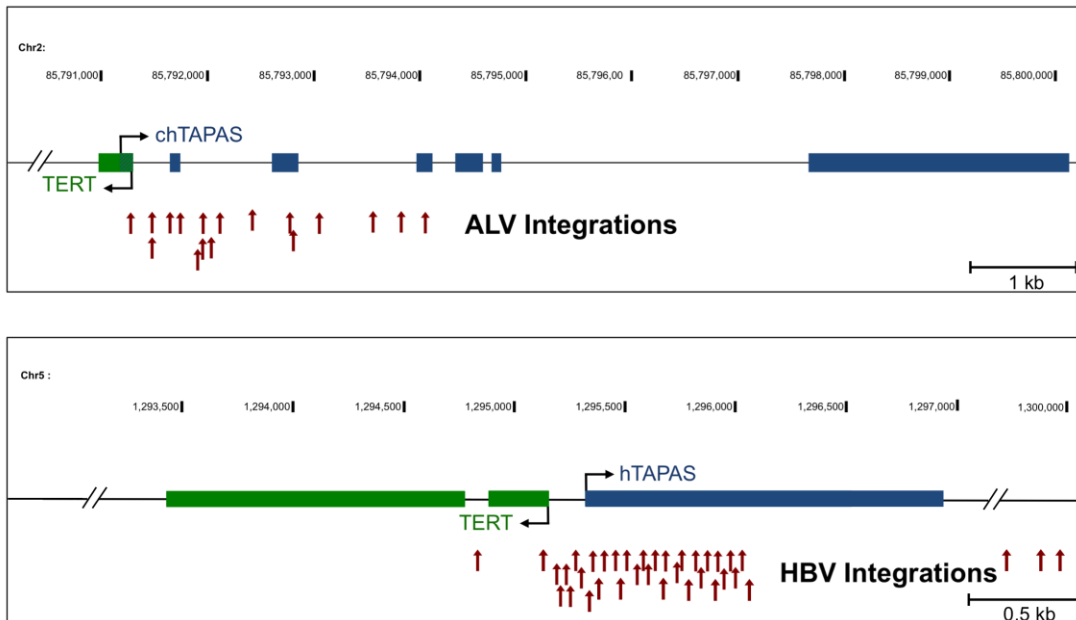


Figure 5.13 TAPAS is a viral integration hotspot in chicken and human tumors. The schematic represents the hTERT promoter region, including the transcription start sites for hTERT and TAPAS, in the (A) chicken and (B) human genomes. This region is depicted as a viral integration hotspot for the avian leukosis virus (ALV) and hepatitis B virus (HBV), respectively in tumors. Arrowheads represent the integration sites for ALV or HBV along the TAPAS gene.

Materials And Methods

Analysis of RNA-seq and CAGE data.

Total and cellular fractionated RNA-seq data for analysis of expression of *hTAPAS* and *hTERT* in human cells were downloaded from the ENCODE Consortium and TCGA (Chang et al., 2013; ENCODE Project Consortium et al., 2012). The ENCODE accession number for the total RNA-seq data analyzed for GM12878 is ENCSR889TRN, for OCI-LY7 is ENCSR001HHK, for H1-hESC is ENCSR537BCG, for HepG2 is ENCSR181ZGR, for HeLa-S3 is ENCSR552EGO, and for K562 is ENCSR000AEL. The RNA-seq data for sub-cellular fractions of GM12878 are available through the ENCODE RNA Dashboard (hg19) (ENCODE Project Consortium et al., 2012). Data available through the ENCODE Data Coordination Center, and TCGA and represent normalized, stranded RNA-seq signal. RNA-Seq libraries for liver cancer cell lines were prepared in duplicate using the TruSeq stranded mRNA library kit according to manufacturers directions and sequenced on the Illumina HiSeq platform. Two biological replicates were analyzed for each data set. The abundances of transcripts (fragments per kilobase per million [FPKM]) were estimated and compared using Cufflinks (Cossio et al., 2012). CAGE data for analysis of *hTAPAS* and *hTERT* TSSs in GM12878 cells were downloaded from the ENCODE Consortium (ENCODE accession number ENCSR000CKA).

Evolutionary conservation and coding potential analysis.

phastCons, phyloCSF and phyloP scores, based on multi-way alignment of 20 mammalian genomes, were aggregated over the *hTAPAS* locus (chr5:1295329-1296947, 1,619 positions), *hTERT* exons (NM_198253 exons annotated in hg38, 4,013 positions), and an intergenic sample of 10,000 positions from the *hTAPAS/hTERT* locus (chr5:1247464-1308783).

Fractionation Experiments in HEK-293 cells.

Crude sub-cellular fractionation of cultured HEK-293 cells was done as described previously (Holden and Horton, 2009).

RNA Isolation and qRT-PCR.

Total RNA was extracted using Trizol (Thermo Fisher) according to manufacturer's protocol. RNA was reverse transcribed with Maxima H minus reverse transcriptase (ThermoFisher Scientific) and oligo(dT)18 primer and/or random hexamers following the manufacturer's protocol. All the PCR primers were commercially synthesized (Integrated DNA Technologies, Inc.). Quantitative reverse transcription-PCR (qRT-PCR) was performed using iQ SYBR green Supermix (Bio-Rad) according to the manufacturer's protocol on a Bio-Rad C1000 thermal cycler/CFX96 Real-Time System. The expression of *hTAPAS* RNA was measured using primers X and X. *hTERT* expression was measured using primers GTGCTGCAGCTCCCATTTCAT and GCTTTCAGGATGGAGTAGCAGA. Expression was normalized to the

expression of GAPDH using exon junction primers AATCCCATCACCATCTTCCA and TGGACTCCACGACGTACTCA. Quantitative PCR (qPCR) was performed in triplicate, with each sample present in technical duplicate during each run. The results were normalized using the comparative threshold cycle (C_T) method.

***hTAPAS* knockdown experiments using ASO.**

Antisense oligonucleotides (ASO) with phosphorothioate bonds were generated by IDT. The ASO targeting *hTAPAS* binds 570 nt downstream from the transcription start site (GTGATTAAACAGATTTGGGGTGGTTG). A scrambled ASO control was generated with matched base composition (GGTACGATTATTTTCGGTTCGATTAGT). HEK-293 cells were transfected with 100 nM ASO using FuGene 6 (Promega) according to manufacturers protocol. Cells were then cultured for 48 hours before being collected for measurement of RNA expression. *hTAPAS* primers flanked the ASO binding site (TGAGCAACCACCCCAAATCT, TTTCCCACCCTTTCTCGACG), *hhTERT* primers spanned either the exon 13-14 junction (GTGCTGCAGCTCCCATTTTCAT, GCTTTCAGGATGGAGTAGCAGA) or the exon 7-8 junction (GCGTAGGAAGACGTCAAGA, ACAGTTCGTGGCTCACCTG). All expression was normalized to a housekeeping gene, GAPDH using primers AATCCCATCACCATCTTCCA and TGGACTCCACGACGTACTCA.

Generating *hTAPAS* constructs and performing Luciferase assays.

The *hTERT* promoter region, including *hTAPAS*, was cloned into the firefly luciferase construct pGL3-basic (Promega) using XhoI and Acc65I cloning sites. The *hTERT* promoter mutation (-66 nts from *hTERT* TSS) was introduced by a quick-change PCR as described previously (Liu and Naismith, 2008). The pRL-CMV construct (Promega) was cotransfected into HEK-293 cells as a transfection control. Forty-eight hours after transfection, HEK-293 cells were harvested and assayed for firefly and Renilla luciferase activities by using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's protocol.

Chapter 6. Future Directions

6.1 MET can be targeted for therapeutic intervention in development of hemangiomas

Uncontrolled MET signaling has been implicated in development of many human cancers including carcinomas, sarcomas, hematopoietic tumors and central nervous system tumors (Graveel et al., 2013). MET plays an important role in angiogenesis, a process that is crucial to the development of hemangiomas (Birchmeier et al., 2003). For example, activation of the HGF/SF-MET pathway is now understood to be a potent inducer of angiogenesis, specifically in endothelial cells – the same cell-type that composes hemangiomas (Abounader and Laterra, 2005). In addition, the HGF/SF-MET pathway can suppress TSP1, a negative regulator of angiogenesis, and can induce expression of VEGFA, a pro-angiogenic gene (Graveel et al., 2013). In this way the HGF/SF-MET pathway controls an “angiogenic switch”, turning on angiogenesis (Zhang et al., 2003). While numerous studies have shown the role of MET in tumorigenesis and metastasis, it has not been implicated in the induction and growth of hemangiomas.

Hemangiomas are vascular anomalies, which are benign tumors of vascular endothelial cells (Haggstrom et al., 2006; Mulliken and Glowacki, 1982). Nearly 10% of newborns in the US are diagnosed with hemangiomas, making them the most common tumors of infancy (Mabeta and Pepper, 2011). In severe cases hemangiomas can develop on internal organs like the airway, liver or intestines, and can lead to tissue damage. In rare events

associated to 1% of all liver cancers in humans, spontaneous rupture of hemangiomas can be lethal.

Our studies of establishing a role of MET in hemangiogenesis are novel. We implicate MET overexpression as a causal agent in the development of ALV-J and ALV-A induced hemangiomas in chickens (Justice et al., 2015a). Although viral induced tumors have been the subject of much study, no previous work has implicated MET in tumor induction by insertional mutagenesis. This suggests that MET may play a unique role in inducing hemangiomas rather than other types of viral induced tumors.

Human and chicken MET have significant functional domain homology, with over 90% sequence identity of the catalytic kinase domain. MET expression and signaling has been associated to early development in chick embryos. Moreover, patterns of expression for MET and ligand HGF/SF are found to be consistent with their known roles in human epithelial-mesenchymal transformation and angiogenesis. Data from our lab suggests that ALV-J induces oncogenesis by insertional mutagenesis, and integrations in the MET oncogene can drive MET overexpression, and contribute to development of chicken hemangiomas. Therefore, we hypothesize that MET plays a role in human hemangiogenesis.

There are wide arrays of known point mutations in MET protein, associated with its aberrant catalytic activity in its signaling cascade. The first activating mutations in MET were found in families with hereditary papillary renal carcinoma demonstrating the oncogenicity of MET. These included

missense mutations in the critical tyrosine residues Y1234 and Y1235 within the kinase domain. MET kinase domain mutations have also been identified in childhood hepatocellular carcinomas, gastric carcinomas, head and neck cancers, and squamous cell cancers. Numerous studies have helped identify the mechanisms by which mutational activation of MET can be oncogenic, and have shown it to be either ligand-dependent or ligand-independent activation (Graveel et al., 2013). Additionally, several groups have discovered multiple point mutations within the ligand binding domain and juxta-membrane domains in gastric, small-lung and other lung cancers (Graveel et al., 2013). Understanding mutations that activate the MET kinase activity is critical for development of effective cancer therapeutics.

In order to address our hypothesis we could identify any known or novel mutations in the MET cDNA from human adult and infantile hemangioma samples. This would provide us an insight of MET enzymatic activity and corresponding predicted changes in its signaling, in human hemangiomagenesis. Any changes in MET expression in human hemangiomas, can also be quantified via qPCR and immunoblotting methods.

Studying the role of MET in hemangiomagenesis provides a potential therapeutic target to treat clinically problematic hemangiomas. Since MET is a well studied oncogene, a large number of small molecule inhibitors and neutralizing antibodies are already available to counter its effect in tumorigenesis (Graveel et al., 2013). If future work establishes a molecular

role of MET in governing hemangioma induction, then it would provide a great basis for these therapeutic strategies in countering hemangiomagenesis.

6.2 Identification of cooperating gene players in ALV induced tumorigenesis

We identified *TERT* and *MYB* among the most clonally expanded integrations in the same tumors from several birds (Chapter 2). We also observed up regulated *MYB* expression in *TERT* tumors without any ALV integrations near or within *MYB*. This suggests a possible cooperation between MYB and TERT. MYB is involved in regulating cell differentiation and proliferation (Sandberg et al., 2005). On the other hand, TERT plays a role in regulating cellular senescence, immortalization and apoptosis (Autexier and Lue, 2006; Del Bufalo et al., 2005; Lee et al., 2004). Therefore, together these protein players might function to facilitate oncogenic transformation of cells in tumorigenesis. Additionally, a number of common transcription factor target gene networks were also identified as common integration hotspots in various tumors. These include the genes targeted by the E2F, EGR, WT1 and SP families of transcription factors.

These gene players might be involved in regulating oncogenic pathways such as differentiation, proliferation, apoptosis and phosphorylation. They might function together via cooperative gene networks to trigger oncogenic transformation resulting in tumor formation. Further

analysis via single cell sequencing would be useful to investigate this network of cooperativity between individual gene players in ALV induced tumorigenesis.

6.3 Functions of human TAPAS RNA

In this work we show that hTAPAS RNA negatively regulates TERT expression in HEK 293 cells (Chapter 5). Further work in additional human cell lines would help identify the cell-type specificity of this regulation. Moreover, the underlying mechanism of this gene regulation is not clear. We hypothesize that hTAPAS RNA might function as a scaffold to recruit different chromatin remodelers to the TERT promoter region to facilitate regulation of TERT expression. Additional experiments, such as RNA immunoprecipitation, would be required to validate this hypothesis, and identify the protein factors that bind to TAPAS RNA. This would help define a concrete mechanism by which hTAPAS acts on the TERT locus. Subsequent work could be done to observe histone modifications in the presence or absence of hTAPAS to determine how it alters the chromatin status at the TERT promoter.

Since ectopic over-expression of hTAPAS also down regulates TERT expression, hTAPAS might have other functions in trans. These could include the transcriptional regulation of additional genes. To further explore this hypothesis, we could perform deep sequencing of the transcriptome in

the presence or absence of hTAPAS expression, and identify global changes in gene expression.

6.4 Role of HBV integrations in altering hTAPAS

The hTERT promoter region is reported to be the most common integration site for HBV in hepatocellular carcinomas (Buendia and Neuveut, 2015; Ding et al., 2012; Toh et al., 2013; Zhao et al., 2016). Approximately 20% of HBV integrations in hepatocellular carcinomas occur in the hTERT promoter region (Buendia and Neuveut, 2015). These integrations near hTERT are thought to confer an early clonal advantage during chronic HBV infection. Moreover, HBV integrations lead to up-regulated hTERT expression and are associated with poorer survival rates in infected patients with these integrations (Zhao et al., 2016).

Nearly all of the reported HBV integrations occur within the 5' region of hTAPAS, and are in a mixed transcriptional orientation (Buendia and Neuveut, 2015; Ding et al., 2012; Toh et al., 2013; Zhao et al., 2016). The prevalence of viral integrations within TAPAS in multiple types of cancers in different organisms like chickens and humans suggests that these integrations are selected for during oncogenesis. These viral integrations might confer a proliferative advantage and make cells predisposed to oncogenic transformation. Since hTAPAS negatively regulates hTERT expression and HBV integrations would likely disrupt hTAPAS expression,

this could promote hTERT expression. These observations, therefore, suggest functional implications of hTAPAS in tumorigenesis.

In order to address this hypothesis, we could analyze tumor samples from human patients with HBV integrations in hTAPAS. Since the majority of the HBV integrations occur in the beginning of hTAPAS, this suggests the expression of a truncated transcript. We could perform RT-PCR to detect the presence of any viral fusion transcripts that may alter the expression levels or activity of hTAPAS RNA. If we observe such truncated transcripts, then it suggests that HBV integrations could be down regulating expression of functional hTAPAS in order to induce elevated hTERT expression. This might be required for oncogenic transformation of cells in hepatocellular carcinomagenesis.

References

- Abounader, R., and Latterra, J. (2005). Scatter factor/hepatocyte growth factor in brain tumor growth and angiogenesis. *Neuro. Oncol.* 7, 436–451.
- Adkins, H.B., Brojatsch, J., and Young, J.A. (2000). Identification and characterization of a shared TNFR-related receptor for subgroup B, D, and E avian leukosis viruses reveal cysteine residues required specifically for subgroup E viral entry. *J. Virol.* 74, 3572–3578.
- Atala, A. (2015). Re: TERT promoter mutations and telomerase reactivation in urothelial cancer. *J. Urol.* 194, 848.
- Autexier, C., and Lue, N.F. (2006). The Structure and Function of Telomerase Reverse Transcriptase. *Annu. Rev. Biochem.* 75, 493–517.
- Baba, T.W., and Humphries, E.H. (1985). Formation of a transformed follicle is necessary but not sufficient for development of an avian leukosis virus-induced lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* 82, 213–216.
- Baba, T.W., and Humphries, E.H. (1986). Selective integration of avian leukosis virus in different hematopoietic tissues. *Virology* 155, 557–566.
- Bai, J., Payne, L.N., and Skinner, M.A. (1995). HPRS-103 (exogenous avian leukosis virus, subgroup J) has an env gene related to those of endogenous elements EAV-0 and E51 and an E element found previously only in sarcoma viruses. *J. Virol.* 69, 779–784.

- Barnard, R.J.O., and Young, J.A.T. (2003). Alpharetrovirus envelope-receptor interactions. *Curr. Top. Microbiol. Immunol.* **281**, 107–136.
- Barr, S.D., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2005). Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* **79**, 12035–12044.
- Beemon, K., and Rosenberg, N. (2012). Mechanisms of oncogenesis by avian and murine retroviruses. In *Current Cancer Research*, E.S. Robertson, ed. (New York, NY: Springer), pp. 677–704.
- Bell, R.J.A., Rube, H.T., Kreig, A., Mancini, A., Fouse, S.F., Nagarajan, R.P., Choi, S., Hong, C., He, D., Pekmezci, M., et al. (2015). The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **348**, 1036–1039.
- Bellanger, J.M., Lazaro, J.B., Diriong, S., Fernandez, A., Lamb, N., and Debant, A. (1998). The two guanine nucleotide exchange factor domains of Trio link the Rac1 and the RhoA pathways in vivo. *Oncogene* **16**, 147–152.
- Benson, S.J., Ruis, B.L., Garbers, A.L., Fadly, A.M., and Conklin, K.F. (1998). Independent isolates of the emerging subgroup J avian leukosis virus derive from a common ancestor. *J. Virol.* **72**, 10301–10304.

Berry, C.C., Gillet, N.A., Melamed, A., Gormley, N., Bangham, C.R.M., and Bushman, F.D. (2012). Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* 28, 755–762.

Birchmeier, C., Birchmeier, W., Gherardi, E., and Vande Woude, G.F. (2003). Met, metastasis, motility and more. *Nat. Rev. Mol. Cell Biol.* 4, 915–925.

Black, A.R., Black, J.D., and Azizkhan-Clifford, J. (2001). Sp1 and Krüppel-like factor family of transcription factors in cell growth regulation and cancer. *J. Cell. Physiol.* 188, 143–160.

Blangy, A., Vignal, E., Schmidt, S., Debant, A., Gauthier-Rouvière, C., and Fort, P. (2000). TrioGEF1 controls Rac- and Cdc42-dependent cell structures through the direct activation of rhoG. *J. Cell Sci.* 113 (Pt 4, 729–739.

Blankenberg, D., Kuster, G. Von, Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010a). Galaxy: A web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* Chapter 19, Unit 19.10.1–21.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., Nekrutenko, A., and Team, G. (2010b). Manipulation of FASTQ data with galaxy. *Bioinformatics* 26, 1783–1785.

Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. *Nat Rev Genet* 6, 611–622.

Blei, F. (2005). Basic science and clinical aspects of vascular anomalies. *Curr. Opin. Pediatr.* 17, 501–509.

Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S., and Wright, W.E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science* 279, 349–352.

Bolisetty, M., Blomberg, J., Benachenhou, F., Sperber, G., and Beemon, K. (2012). Unexpected diversity and expression of avian endogenous retroviruses. *MBio* 3, e00344–12.

Borsani, G., DeGrandi, A., Ballabio, A., Bulfone, A., Bernard, L., Banfi, S., Gattuso, C., Mariani, M., Dixon, M., Donnai, D., et al. (1999). EYA4, a novel vertebrate gene related to *Drosophila* eyes absent. *Hum. Mol. Genet.* 8, 11–23.

Borson, N.D., Salo, W.L., and Drewes, L.R. (1992). A lock-docking oligo(dT) primer for 5' and 3' RACE PCR. *Genome Res.* 2, 144–148.

Brojatsch, J., Naughton, J., Rolls, M.M., Ziegler, K., and Young, J.A. (1996). CAR1, a TNFR-related protein, is a cellular receptor for cytopathic avian leukosis-sarcoma viruses and mediates apoptosis. *Cell* 87, 845–855.

Buendia, M.-A., and Neuveut, C. (2015). Hepatocellular Carcinoma. *Cold Spring Harb. Perspect. Med.* 5, a021444–a021444.

Del Bufalo, D., Rizzo, A., Trisciuoglio, D., Cardinali, G., Torrisi, M.R., Zangemeister-Wittke, U., Zupi, G., and Biroccio, A. (2005). Involvement of hTERT in apoptosis induced by interference with Bcl-2 expression and function. *Cell Death Differ.* 12, 1429–1438.

Burns, J.C., Friedmann, T., Driever, W., Burrascano, M., and Yee, J.-K. (1993). Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: Concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells (gene therapy/zebrafish). *Genetics* 90, 8033–8037.

Bushman, F.D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. *Science* 249, 1555–1558.

Bussolino, F., Di Renzo, M.F., Ziche, M., Bocchietto, E., Olivero, M., Naldini, L., Gaudino, G., Tamagnone, L., Coffey, A., and Comoglio, P.M. (1992). Hepatocyte growth factor is a potent angiogenic factor which stimulates endothelial cell motility and growth. *J. Cell Biol.* 119, 629–641.

Cepko, C. (2001). Large-scale preparation and concentration of retrovirus stocks. In *Curr Protoc Mol Biol*, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p. Unit9 12.

Cevik, D., Yildiz, G., and Ozturk, M. (2015). Common telomerase reverse transcriptase promoter mutations in hepatocellular carcinomas from different geographical locations. *World J. Gastroenterol.* 21, 311–317.

Chai, N., and Bates, P. (2006). Na⁺/H⁺ exchanger type 1 is a receptor for pathogenic subgroup J avian leukosis virus. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5531–5536.

Chang, H., and Delany, M.E. (2004). Karyotype stability of the DT40 chicken B cell line: macrochromosome variation and cytogenetic mosaicism. *Chromosome Res.* 12, 299–307.

Chang, K., Creighton, C.J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y.S.N., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120.

Chen, Y., Chen, L., and Yang, X. (2017). The versatile role of microRNA-30a in human cancer. *Cell Physiol Biochem* 41, 1616–1632.

Cheng, Z., Liu, J., Cui, Z., and Zhang, L. (2010). Tumors associated with avian leukosis virus subgroup J in layer hens during 2007 to 2009 in China. *J. Vet. Med. Sci.* 72, 1027–1033.

Chiba, K., Johnson, J.Z., Vogan, J.M., Wagner, T., Boyle, J.M., and Hockemeyer, D. (2015). Cancer-associated tert promoter mutations abrogate telomerase silencing. *Elife* 4, 1–20.

Clurman, B.E., and Hayward, W.S. (1989). Multiple proto-oncogene activations in avian leukosis virus-induced lymphomas: evidence for stage-specific events. *Mol. Cell. Biol.* 9, 2657–2664.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3260–3265.

Collins, P.J., Kobayashi, Y., Nguyen, L., Trinklein, N.D., and Myers, R.M. (2007). The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet.* 3, 2247–2255.

Cong, Y.-S., Wright, W.E., and Shay, J.W. (2002). Human telomerase and its regulation. *Microbiol. Mol. Biol. Rev.* 66, 407–425.

Cook, L.B., Melamed, A., Niederer, H., Valganon, M., Laydon, D., Foroni, L., Taylor, G.P., Matsuoka, M., and Bangham, C.R.M. (2014). The role of HTLV-1 clonality, proviral structure, and genomic integration site in adult T-cell leukemia/lymphoma. *Blood* 123, 3925–3931.

Cooper, G.M., Stone, E.A., Asimenos, G., NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15, 901–913.

Cooper, M.D., Payne, L.N., Dent, P.B., Burmester, B.R., and Good, R.A. (1968). Pathogenesis of avian lymphoid leukosis. i. histogenesis. *J. Natl. Cancer Inst.* 41, 373–389.

Cossio, M.L.T., Giesen, L.F., Araya, G., Pérez-Cotapos, M.L.S., Vergara, R.L., Manca, M., Tohme, R. a., Holmberg, S.D., Bressmann, T., Lirio, D.R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* XXXIII, 562.

Cristofari, G., and Lingner, J. (2006). Telomere length homeostasis requires that telomerase levels are limiting. *EMBO J.* 25, 565–574.

Cui, Z., Sun, S., Zhang, Z., and Meng, S. (2009). Simultaneous endemic infections with subgroup J avian leukosis virus and reticuloendotheliosis virus in commercial and local breeds of chickens. *Avian Pathol.* 38, 443–448.

Delany, M.E., and Daniels, L.M. (2004). The chicken telomerase reverse transcriptase (chTERT): Molecular and cytogenetic characterization with a comparative analysis. *Gene* 339, 61–69.

Delany, M.E., Krupkin, A.B., and Miller, M.M. (2000). Organization of telomere sequences in birds: evidence for arrays of extreme length and for in vivo shortening. *Cytogenet. Cell Genet.* *90*, 139–145.

Deniaud, E., Baguet, J., Mathieu, A.-L., S, G., Marvel, J., and Leverrier, Y. (2006). Overexpression of Sp1 transcription factor induces apoptosis. *Oncogene* *25*, 7096–7105.

Derse, D., Crise, B., Li, Y., Princier, G., Lum, N., Stewart, C., McGrath, C.F., Hughes, S.H., Munroe, D.J., and Wu, X. (2007). Human T-cell leukemia virus type 1 integration target sites in the human genome: comparison with those of other retroviruses. *J. Virol.* *81*, 6731–6741.

Ding, D., Lou, X., Hua, D., Yu, W., Li, L., Wang, J., Gao, F., Zhao, N., Ren, G., Li, L., et al. (2012). Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing–Based Approach. *PLoS Genet.* *8*, e1003065.

Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* *4*, 2126.

Donzé, O., and Spahr, P.F. (1992). Role of the open reading frames of Rous sarcoma virus leader RNA in translation and genome packaging. *EMBO J.* *11*, 3747–3757.

Dupuis-Maurin, V., Brinza, L., Baguet, J., Plantamura, E., Schicklin, S., Chambion, S., Macari, C., Tomkowiak, M., Deniaud, E., Leverrier, Y., et al. (2015). Overexpression of the transcription factor Sp1 activates the OAS-RNase L-RIG-I pathway. *PLoS One* 10, e0118551.

Einfeld, D., and Hunter, E. (1988). Oligomeric structure of a prototype retrovirus glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* 85, 8688–8692.

Ellermann, V., and Bang, O. (1908). Experimentelle Leukämie bei Hühnern. *Zentralbl. Bakteriol. Parasitenkd. Infect. Hyg. Abt. Orig.* 46, 595–609.

ENCODE Project Consortium, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Fietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.

Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67, 1211–1221.

Everaert, C., Luypaert, M., Maag, J.L. V, Cheng, Q.X., Dinger, M.E., Hellemans, J., and Mestdagh, P. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.* 7, 1559.

Fadly, A., and Nair, V. (2008). Leukosis/sarcoma Group. In Diseases of Poultry, D.E. Saif, Y.M., Fadly, A.M., Glisson, J.R., McDougald, L.R., Nolan, L.K., Swayne, ed. pp. 514–568.

Fadly, A.M., and Smith, E.J. (1999). Isolation and some characteristics of a subgroup J-like avian leukosis virus associated with myeloid leukosis in meat-type chickens in the United States. *Avian Dis.* 43, 391–400.

Faschinger, A., Rouault, F., Sollner, J., Lukas, A., Salmons, B., Günzburg, W.H., and Indik, S. (2008). Mouse mammary tumor virus integration site selection in human and mouse genomes. *J. Virol.* 82, 1360–1367.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* (N. Y). 39, 783–791.

Fenton, S.P., Reddy, M.R., and Bagust, T.J. (2005). Single and concurrent avian leukosis virus infections with avian leukosis virus-J and avian leukosis virus-A in Australian meat-type chickens. *Avian Pathol. J. W.V.P.A* 34, 48–54.

Ferber, M.J., Montoya, D.P., Yu, C., Aderca, I., McGee, a, Thorland, E.C., Nagorney, D.M., Gostout, B.S., Burgart, L.J., Boix, L., et al. (2003). Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* 22, 3813–3820.

Ficht, T.A., Chang, L.J., and Stoltzfus, C.M. (1984). Avian sarcoma virus gag and env gene structural protein precursors contain a common amino-terminal sequence. *Proc. Natl. Acad. Sci. U. S. A.* *81*, 362–366.

Gao, Y.-L., Qin, L.-T., Pan, W., Wang, Y.-Q., Qi, X.-L., Gao, H.-L., and Wang, X.-M. (2010). Avian Leukosis Virus Subgroup J in Layer Chickens, China. *Emerg. Infect. Dis.* *16*, 1637–1638.

Garcia, C.K., Wright, W.E., and Shay, J.W. (2007a). Human diseases of telomerase dysfunction: Insights into tissue aging. *Nucleic Acids Res.* *35*, 7406–7416.

Garcia, S., Dalès, J.-P., Charafe-Jauffret, E., Carpentier-Meunier, S., Andrac-Meyer, L., Jacquemier, J., Andonian, C., Lavaut, M.-N., Allasia, C., Bonnier, P., et al. (2007b). Poor prognosis in breast carcinomas correlates with increased expression of targetable CD146 and c-Met and with proteomic basal-like phenotype. *Hum. Pathol.* *38*, 830–841.

Gherardi, E., Birchmeier, W., Birchmeier, C., and Vande Woude, G. (2012). Targeting MET in cancer: rationale and progress. *Nat. Rev. Cancer* *12*, 89–103.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* *15*, 1451–1455.

Giles, K.E., and Beemon, K.L. (2005). Retroviral splicing suppressor sequesters a 3' splice site in a 50S aberrant splicing complex. *Mol. Cell. Biol.* **25**, 4397–4405.

Gillet, N.A., Malani, N., Melamed, A., Gormley, N., Carter, R., Bentley, D., Berry, C., Bushman, F.D., Taylor, G.P., and Bangham, C.R.M. (2011). The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* **117**, 3113–3122.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86.

Goodenow, M.M., and Hayward, W.S. (1987). 5' Long Terminal Repeats of myc-Associated Proviruses Appear Structurally Intact but Are Functionally Impaired in Tumors Induced by Avian Leukosis Viruses. *J. Virol.* **2489–2498**.

Goodspeed, A., Heiser, L.M., Gray, J.W., and Costello, J.C. (2016). Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Mol. Cancer Res.* **14**, 3–13.

Grant, D.S., Kleinman, H.K., Goldberg, I.D., Bhargava, M.M., Nickoloff, B.J., Kinsella, J.L., Polverini, P., and Rosen, E.M. (1993). Scatter factor induces blood vessel formation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1937–1941.

Graveel, C., Su, Y., Koeman, J., Wang, L.-M., Tessarollo, L., Fiscella, M., Birchmeier, C., Swiatek, P., Bronson, R., and Vande Woude, G. (2004). Activating Met mutations produce unique tumor profiles in mice with selective duplication of the mutant allele. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 17198–17203.

Graveel, C.R., Tolbert, D., and Vande Woude, G.F. (2013). MET: A critical player in tumorigenesis and therapeutic target. *Cold Spring Harb. Perspect. Biol.* *5*(7).

Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* *481*, 306–313.

Gudleski, N., Flanagan, J.M., Ryan, E.P., Bewley, M.C., and Parent, L.J. (2010). Directionality of nucleocytoplasmic transport of the retroviral gag protein depends on sequential binding of karyopherins and viral RNA. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 9358–9363.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* *59*, 307–321.

Gupta, R. a, Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA

HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.

Hackett, J.A., and Greider, C.W. (2002). Balancing instability: dual roles for telomerase and telomere dysfunction in tumorigenesis. *Oncogene* 21, 619–626.

Haggstrom, A.N., Drolet, B.A., Baselga, E., Chamlin, S.L., Garzon, M.C., Horii, K.A., Lucky, A.W., Mancini, A.J., Metry, D.W., Newell, B., et al. (2006). Prospective Study of Infantile Hemangiomas: Clinical Characteristics Predicting Complications and Treatment. *Pediatrics* 118, 882–887.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458–460.

Hayward, W.S., Neel, B.G., and Astrin, S.M. (1981). Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. *Nature* 290, 475–480.

Heidari, M., Zhang, H.M., and Sharif, S. (2008). Marek's disease virus induces Th-2 activity during cytolytic infection. *Viral Immunol.* 21, 203–214.

Heidenreich, B., Rachakonda, P.S., Hemminki, K., and Kumar, R. (2014). TERT promoter mutations in cancer development. *Curr. Opin. Genet. Dev.* 24, 30–37.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.

Holden, P., and Horton, W.A. (2009). Crude subcellular fractionation of cultured mammalian cell lines. *BMC Res. Notes* 2, 243.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., et al. (2013). TERT Promoter Mutations in Familial and Sporadic Melanoma. *Science* (80). 339, 959–961.

Hrdlickova, R., Nehyba, J., and Bose, H.R. (2012). Alternatively Spliced Telomerase Reverse Transcriptase Variants Lacking Telomerase Activity Stimulate Cell Proliferation. *Mol. Cell. Biol.* 32, 4283–4296.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G. V, Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. *Science* 339, 957–959.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. *Nat. Med.* 21, 1253–1261.

Hughes, S.H. (2004). The RCAS vector system. *Folia Biol. (Praha)*. 50, 107–119.

Hughes, S.H., Mutschler, A., Bishop, J.M., and Varmus, H.E. (1981). A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration. *Proc. Natl. Acad. Sci. U. S. A.* 78, 4299–4303.

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55, 447–458.

Jiang, L., Zeng, X., Hua, Y., Gao, Q., Fan, Z., Chai, H., Wang, Q., Qi, X., Wang, Y., Gao, H., et al. (2014). Genetic diversity and phylogenetic analysis of glycoprotein gp85 of avian leukosis virus subgroup J wild-bird isolates from Northeast China. *Arch. Virol.* 159, 1821–1826.

Jiang, W., Kanter, M.R., Dunkel, I., Ramsay, R.G., Beemon, K.L., and Hayward, W.S. (1997). Minimal truncation of the c-myc gene product in rapid-onset B-cell lymphoma. *J Virol* 71, 6526–6533.

Justice, J., Malhotra, S., Ruano, M., Li, Y., Zavala, G., Lee, N., Morgan, R., and Beemon, K. (2015a). The MET gene is a common integration target in avian leukosis virus subgroup J-induced chicken hemangiomas. *J. Virol.* 89, 4712–4719.

Justice, J.F., Morgan, R.W., and Beemon, K.L. (2015b). Common viral integration sites identified in avian leukosis virus- induced B-cell lymphomas. *MBio* 6, e01863–15.

Justice IV, J., and Beemon, K.L. (2013). Avian retroviral replication. *Curr. Opin. Virol.* 3, 664–669.

Kanter, M.R., Smith, R.E., and Hayward, W.S. (1988). Rapid induction of B-cell lymphomas: insertional activation of c-myc by avian leukosis virus. *J. Virol.* 62, 1423–1432.

Killela, P.J., Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz Jr., L.A., Friedman, A.H., Friedman, H., Gallia, G.L., Giovannella, B.C., et al. (2013). TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6021–6026.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011–2015.

Koh, C.M., Khattar, E., Leow, S.C., Liu, C.Y., Muller, J., Ang, W.X., Li, Y., Franzoso, G., Li, S., Guccione, E., et al. (2015). Telomerase regulates MYC-driven oncogenesis independent of its reverse transcriptase activity. *J. Clin. Invest.* 125, 2109–2122.

Kohonen, P., Nera, K.P., and Lassila, O. (2007). Avian model for B-cell immunology - New genomes and phylotranscriptomics. *Scand. J. Immunol.* 66, 113–121.

Koskela, K., Kohonen, P., Nieminen, P., Buerstedde, J.M., and Lassila, O. (2003). Insight into lymphoid development by gene expression profiling of avian B cells. *Immunogenetics* 55, 412–422.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene* 30, 1956–1962.

Kvaratskhelia, M., Sharma, A., Larue, R.C., Serrao, E., and Engelman, A. (2014). Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res.* 42, 10209–10225.

Landman, W.J.M., Post, J., Boonstra-Blom, A.G., Buyse, J., Elbers, A.R.W., and Koch, G. (2002). Effect of an in ovo infection with a Dutch avian leukosis virus subgroup J isolate on the growth and immunological performance of SPF broiler chickens. *Avian Pathol. J. W.V.P.A* 31, 59–72.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009a). 2C- Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009b). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

- Laskey, S.B., Pohlmeier, C.W., Bruner, K.M., and Siliciano, R.F. (2016). Evaluating clonal expansion of HIV-infected cells: optimization of PCR strategies to predict clonality. *PLoS Pathog.* 12, e1005689.
- LeBlanc, J.J., Uddowla, S., Abraham, B., Clatterbuck, S., and Beemon, K.L. (2007). Tap and Dbp5, but not Gag, are involved in DR-mediated nuclear export of unspliced Rous sarcoma virus RNA. *Virology* 363, 376–386.
- Lee, K.M., Choi, K.H., and Ouellette, M.M. (2004). Use of exogenous hTERT to immortalize primary human cells. *Cytotechnology* 45, 33–38.
- Leinonen, R., Sugawara, H., and Shumway, M. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21.
- Li, D., Qin, L., Gao, H., Yang, B., Liu, W., Qi, X., Wang, Y., Zeng, X., Liu, S., Wang, X., et al. (2013). Avian leukosis virus subgroup A and B infection in wild birds of Northeast China. *Vet. Microbiol.* 163, 257–263.
- Li, Y., Liu, X., Yang, Z., Xu, C., Liu, D., Qin, J., Dai, M., Hao, J., Feng, M., Huang, X., et al. (2014). The MYC, TERT, and ZIC1 genes are common targets of viral integration and transcriptional deregulation in avian leukosis virus subgroup J-induced myeloid leukosis. *J. Virol.* 88, 3182–3191.
- Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27, i275–i282.

Lindsey, J.C., Schwalbe, E.C., Potluri, S., Bailey, S., Williamson, D., and Clifford, S.C. (2014). TERT promoter mutation and aberrant hypermethylation are associated with elevated expression in medulloblastoma and characterise the majority of non-infant SHH subgroup tumours. *Acta Neuropathol.* 127, 307–309.

Liu, H., and Naismith, J.H. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* 8, 91.

Liu, Y., Emilion, G., Mungall, A.J., Dunham, I., Beck, S., Le Meuth-Metzinger, V.G., Shelling, A.N., Charnock, F.M., and Ganesan, T.S. (2002). Physical and transcript map of the region between D6S264 and D6S149 on chromosome 6q27, the minimal region of allele loss in sporadic epithelial ovarian cancer. *Oncogene* 21, 387–399.

Mabeta, P., and Pepper, M.S. (2011). Hemangiomas - current therapeutic strategies. *Int. J. Dev. Biol.* 55, 431–437.

Macville, M., Schröck, E., Padilla-Nash, H., Keck, C., Ghadimi, B.M., Zimonjic, D., Popescu, N., and Ried, T. (1999). Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res.* 59, 141–150.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 is essential for nuclear

and chromosomal targeting of HIV-1 integrase in human cells. *J. Biol. Chem.* 278, 33528–33539.

Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 345, 179–183.

Malhotra, S., Justice, J., Lee, N., Li, Y., Zavala, G., Ruano, M., Morgan, R., and Beemon, K. (2015). Complete genome sequence of an american avian leukosis virus subgroup j isolate that causes hemangiomas and myeloid leukosis. *Genome Announc.* 3, e01586–14.

Malkinson, M., Banet-Noach, C., Davidson, I., Fadly, A.M., and Witter, R.L. Comparison of serological and virological findings from subgroup J avian leukosis virus-infected neoplastic and non-neoplastic flocks in Israel. *Avian Pathol. J. W.V.P.A* 33, 281–287.

Malkinson, M., Banet-Noach, C., Davidson, I., Fadly, A.M., and Witter, R.L. (2004). Comparison of serological and virological findings from subgroup J avian leukosis virus-infected neoplastic and non-neoplastic flocks in Israel. *Avian Pathol. J. W.V.P.A* 33, 281–287.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* 43, D222–D226.

McNally, M.T., and Beemon, K. (1992). Intronic sequences and 3' splice sites control Rous sarcoma virus RNA splicing. *J. Virol.* 66, 6–11.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2, e234.

Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. *Neuropsychopharmacology* 38, 23–38.

Moustakas, A., Sonstegard, T.S., and Hackett, P.B. (1993). Alterations of the three short open reading frames in the Rous sarcoma virus leader RNA modulate viral replication and gene expression. *J. Virol.* 67, 4337–4349.

Mulliken, J.B., and Glowacki, J. (1982). Hemangiomas and vascular malformations in infants and children: a classification based on endothelial characteristics. *Plast. Reconstr. Surg.* 69, 412–422.

Nadaraia-Hoke, S., Bann, D. V, Lochmann, T.L., Gudleski-O'Regan, N., and Parent, L.J. (2013). Alterations in the MA and NC domains modulate phosphoinositide-dependent plasma membrane localization of the Rous sarcoma virus Gag protein. *J. Virol.* 87, 3609–3615.

Nagalakshmi, U., Waern, K., and Snyder, M. (2010). RNA-seq: A method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol. Chapter 4*, Unit 4.11.1–13.

Narezkina, A., Taganov, K.D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, M., Katz, R. a, and Skalka, A.M. (2004). Genome-wide analyses of avian sarcoma virus integration sites. *J. Virol.* 78, 11656–11663.

Neel, B.G., Hayward, W.S., Robinson, H.L., Fang, J., and Astrin, S.M. (1981). Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. *Cell* 23, 323–334.

Nehyba, J., Malhotra, S., Winans, S., O'Hare, T.H., Justice, J., and Beemon, K. (2016). Avian Leukosis Virus Activation of an Antisense RNA Upstream of TERT in B-Cell Lymphomas. *J Virol* 90, 9509–9517.

Neiman, P.E. (1994). Retrovirus-induced B cell neoplasia in the bursa of Fabricius. *Adv. Immunol.* 56, 467–484.

Neiman, P., Payne, L.N., and Weiss, R. (1980a). Viral DNA in bursal lymphomas induced by avian leukosis viruses. *J. Virol.* 34, 178–186.

Neiman, P., Payne, L.N., and Weiss, R.A. (1980b). Viral DNA in bursal lymphomas induced by avian leukosis viruses. *J. Virol.* 34, 178–186.

Neiman, P., Wolf, C., Enrietto, P.J., and Cooper, G.M. (1985). A retroviral myc gene induces preneoplastic transformation of lymphocytes in a bursal transplantation assay. *Proc Natl Acad Sci U S A* 82, 222–226.

Neiman, P.E., Grbić, J.J., Polony, T.S., Kimmel, R., Bowers, S.J., Delrow, J., and Beemon, K.L. (2003). Functional genomic analysis reveals distinct neoplastic phenotypes associated with c-myb mutation in the bursa of Fabricius. *Oncogene* 22, 1073–1086.

Nilsen, T.W., Maroney, P.A., Goodwin, R.G., Rottman, F.M., Crittenden, L.B., Raines, M.A., and Kung, H.J. (1985). c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor. *Cell* 41, 719–726.

O’Sullivan, C.T., Polony, T.S., Paca, R.E., and Beemon, K.L. (2002). Rous Sarcoma Virus Negative Regulator of Splicing Selectively Suppresses src mRNA Splicing and Promotes Polyadenylation. *Virology* 302, 405–412.

Orekhova, a S., and Rubtsov, P.M. (2013). Bidirectional promoters in the transcription of mammalian genomes. *Biochem. Biokhimiia* 78, 335–341.

Pan, W., Gao, Y., Sun, F., Qin, L., Liu, Z., Yun, B., Wang, Y., Qi, X., Gao, H., and Wang, X. (2011). Novel sequences of subgroup J avian leukosis viruses associated with hemangioma in Chinese layer hens. *Viol. J.* 8, 552.

Payne, G.S., Bishop, J.M., and Varmus, H.E. (1982). Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. *Nature* 295, 209–214.

Payne, L.N., Brown, S.R., Bumstead, N., Howes, K., Frazier, J.A., and Thouless, M.E. (1991). A novel subgroup of exogenous avian leukosis virus in chickens. *J. Gen. Virol.* 72, 801–807.

Payne, L.N., Gillespie, A.M., and Howes, K. (1992). Myeloid leukaemogenicity and transmission of the HPRS-103 strain of avian leukosis virus. *Leukemia* 6, 1167–1176.

Peghini, P.L., Iwamoto, M., Raffeld, M., Chen, Y.-J., Goebel, S.U., Serrano, J., and Jensen, R.T. (2002). Overexpression of epidermal growth factor and hepatocyte growth factor receptors in a proportion of gastrinomas correlates with aggressive growth and lower curability. *Clin. Cancer Res. An Off. J. Am. Assoc. Cancer Res.* 8, 2273–2285.

Pincetic, A., and Leis, J. (2009). The Mechanism of Budding of Retroviruses From Cell Membranes. *Adv. Virol.* 2009, 6239691–6239699.

Pisani, D.F., Cabane, C., Derijard, B., and Dechesne, C. (2004). The topoisomerase 1-interacting protein BTBD1 is essential for muscle cell differentiation. *Cell Death Differ* 11, 1157–1165.

Polager, S., and Ginsberg, D. (2009). p53 and E2f: partners in life and death. *Nat. Rev. Cancer* 9, 738–748.

Polony, T.S., Bowers, S.J., Neiman, P.E., and Beemon, K.L. (2003). Silent point mutation in an avian retrovirus RNA processing element promotes c-myc-associated short-latency lymphomas. *J. Virol.* 77, 9378–9387.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.

Reimand, J., Kull, M., Peterson, H., Hansen, J., and Vilo, J. (2007). G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* 35, W193–W200.

Renaud, S., Bosman, F.T., and Benhattar, J. (2003). Implication of the exon region in the regulation of the human telomerase reverse transcriptase gene promoter. *Biochem. Biophys. Res. Commun.* 300, 47–54.

Rous, P. (1911). A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS. *J. Exp. Med.* 13, 397–411.

Sacco, M.A., Flannery, D.M., Howes, K., and Venugopal, K. (2000). Avian endogenous retrovirus EAV-HP shares regions of identity with avian leukosis virus subgroup J and the avian retrotransposon ART-CH. *J. Virol.* 74, 1296–1306.

- Saebøe-Larssen, S., Fossberg, E., and Gaudernack, G. (2006). Characterization of novel alternative splicing sites in human telomerase reverse transcriptase (hTERT): analysis of expression and mutual correlation in mRNA isoforms from normal and tumour tissues. *BMC Mol. Biol.* 7, 26.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sandberg, M.L., Sutton, S.E., Pletcher, M.T., Wiltshire, T., Tarantino, L.M., Hogenesch, J.B., and Cooke, M.P. (2005). c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev. Cell* 8, 153–166.
- Scheifele, L.Z., Garbitt, R.A., Rhoads, J.D., and Parent, L.J. (2002). Nuclear entry and CRM1-dependent nuclear export of the Rous sarcoma virus Gag polyprotein. *Proc. Natl. Acad. Sci.* 99, 3944–3949.
- Scheifele, L.Z., Ryan, E.P., and Parent, L.J. (2005). Detailed mapping of the nuclear export signal in the Rous sarcoma virus Gag protein. *J. Virol.* 79, 8732–8741.
- Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521–529.

Shao, W., Shan, J., Kearney, M.F., Wu, X., Maldarelli, F., Mellors, J.W., Luke, B., Coffin, J.M., and Hughes, S.H. (2016). Retrovirus Integration Database (RID): a public database for retroviral insertion sites into host genomes. *Retrovirology* 13, 47.

Sharma, A., Larue, R.C., Plumb, M.R., Malani, N., Male, F., Slaughter, A., Kessl, J.J., Shkriabai, N., Coward, E., Aiyer, S.S., et al. (2013). BET proteins promote efficient murine leukemia virus integration at transcription start sites. *Proc. Natl. Acad. Sci.* 110, 12036–12041.

Shay, J.W., and Wright, W.E. (2011). Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* 21, 349–353.

Shi, M., Tian, M., Liu, C., Zhao, Y., Lin, Y., Zou, N., Liu, P., and Huang, Y. (2011). Sequence analysis for the complete proviral genome of subgroup J Avian Leukosis virus associated with hemangioma: a special 11 bp deletion was observed in U3 region of 3'UTR. *Viol. J.* 8, 158.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050.

Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C., et al. (2013).

Divergent transcription of long noncoding RNA / mRNA gene pairs in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 2876–2881.

Simon, M.C., Neckameyer, W.S., Hayward, W.S., and Smith, R.E. (1987). Genetic determinants of neoplastic diseases induced by a subgroup F avian leukosis virus. *J. Virol.* *61*, 1203–1212.

Singh, P.K., Plumb, M.R., Ferris, A.L., Iben, J.R., Wu, X., Fadel, H.J., Luke, B.T., Esnault, C., Poeschla, E.M., Hughes, S.H., et al. (2015). LEDGF/p75 interacts with mRNA splicing factors and targets HIV-1 integration to highly spliced genes. *Genes Dev.* *29*, 2287–2297.

Smith, L.M., Toye, A.A., Howes, K., Bumstead, N., Payne, L.N., and Venugopal, K. (1999). Novel endogenous retroviral sequences in the chicken genome closely related to HPRS-103 (subgroup J) avian leukosis virus. *J. Gen. Virol.* *80 (Pt 1)*, 261–268.

Smith, M.R., Smith, R.E., Dunkel, I., Hou, V., Beemon, K.L., and Hayward, W.S. (1997). Genetic determinant of rapid-onset B-cell lymphoma by avian leukosis virus. *J. Virol.* *71*, 6534–6540.

Srikantha, T., Landsman, D., and Bustin, M. (1990). A single copy gene for chicken chromosomal protein HMG-14b has evolutionarily conserved features, has lost one of its introns and codes for a rapidly evolving protein. *J. Mol. Biol.* *211*, 49–61.

Sukhatme, V.P., Cao, X., Chang, L.C., Tsai-Morris, C.H., Stamenkovich, D., Ferreira, P.C.P., Cohen, D.R., Edwards, S.A., Shows, T.B., Curran, T., et al. (1988). A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. *Cell* 53, 37–43.

Sung, H.-W., Kim, J.-H., Reddy, S., and Fadly, A. Isolation of subgroup J avian leukosis virus in Korea. *J. Vet. Sci.* 3, 71–74.

Sung, H.-W., Kim, J.-H., Reddy, S., and Fadly, A. (2002). Isolation of subgroup J avian leukosis virus in Korea. *J. Vet. Sci.* 3, 71–74.

Swanstrom, R., and Wills, J. (1997). Synthesis, Assembly, and Processing of Viral Proteins.

Swanstrom, R., Parker, R.C., Varmus, H.E., and Bishop, J.M. (1983). Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A.* 80, 2519–2523.

Takahashi, K., Mulliken, J.B., Kozakewich, H.P., Rogers, R.A., Folkman, J., and Ezekowitz, R.A. (1994). Cellular markers that distinguish the phases of hemangioma during infancy and childhood. *J. Clin. Invest.* 93, 2357–2364.

Tam, W., Ben-Yehuda, D., and Hayward, W.S. (1997). bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. *Mol. Cell. Biol.* 17, 1490–1502.

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 11030–11035.

Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* *30*, 2725–2729.

Tan, S.T., Velickovic, M., Ruger, B.M., and Davis, P.F. (2000). Cellular and extracellular markers of hemangioma. *Plast. Reconstr. Surg.* *106*, 529–538.

Taylor, H.A., and Delany, M.E. (2000). Ontogeny of telomerase in chicken: Impact of downregulation on pre- and postnatal telomere length in vivo. *Dev. Growth Differ.* *42*, 613–621.

Toh, S.T., Jin, Y., Liu, L., Wang, J., Babrzadeh, F., Gharizadeh, B., Ronaghi, M., Toh, H.C., Chow, P.K.-H., Chung, A.Y.-F., et al. (2013). Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis* *34*, 787–798.

Trinklein, N.T., Force Aldred, S., Hartman, S.J., Schroeder, D.I., Otilar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res.* *14*, 62–66.

Trobridge, G.D., Miller, D.G., Jacobs, M.A., Allen, J.M., Kiem, H.-P., Kaul, R., and Russell, D.W. (2006). Foamy virus vector integration sites in normal human cells. *Proc. Natl. Acad. Sci.* 103, 1498–1503.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693.

Ulaner, G.A., and Giudice, L.C. (1997). Developmental regulation of telomerase activity in human fetal tissues during gestation. *Mol. Hum. Reprod.* 3, 769–773.

Vaziri, H., and Benchimol, S. (1998). Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.* 8, 279–282.

Venugopal, K. (1999). Avian leukosis virus subgroup J: a rapidly evolving group of oncogenic retroviruses. *Res. Vet. Sci.* 67, 113–119.

Verderame, M.F., Nelle, T.D., and Wills, J.W. (1996). The membrane-binding domain of the Rous sarcoma virus Gag protein. *J. Virol.* 70, 2664–2668.

Wakano, C., Byun, J.S., Di, L.J., and Gardner, K. (2012). The dual lives of bidirectional promoters. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1819, 688–693.

Wang, Q., Gao, Y., Wang, Y., Qin, L., Qi, X., Qu, Y., Gao, H., and Wang, X. (2012). A 205-Nucleotide Deletion in the 3' Untranslated Region of Avian Leukosis Virus Subgroup J, Currently Emergent in China, Contributes to Its Pathogenicity. *J. Virol.* **86**, 12849–12860.

Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. *Trends Genet.* **27**, 267–276.

Weil, J.E., Hadjithomas, M., and Beemon, K.L. (2009). Structural characterization of the Rous sarcoma virus RNA stability element. *J. Virol.* **83**, 2119–2129.

Weill, J.-C., Weller, S., and Reynaud, C.-A. (2004). A bird's eye view on human B cells. *Semin. Immunol.* **16**, 277–281.

Weiss, R.A., and Vogt, P.K. (2011). 100 years of Rous sarcoma virus. *J. Exp. Med.* **208**, 2351–2355.

Werner, S., Hindmarsh, P., Napirei, M., Vogel-Bachmayr, K., and Wöhr, B.M. (2002). Subcellular localization and integration activities of rous sarcoma virus reverse transcriptase. *J. Virol.* **76**, 6205–6212.

Winans, S., Larue, R.C., Abraham, C.M., Shkriabai, N., Skopp, A., Winkler, D., Kvaratskhelia, M., and Beemon, K.L. (2017a). The FACT complex promotes avian leukosis virus DNA integration. *J. Virol.* **91**, JVI.00082–17.

- Winans, S., Flynn, A., Malhotra, S., Balagopal, V., and Beemon, K.L. (2017b). Integration of ALV into CTDSPL and CTDSPL2 genes in B-cell lymphomas promotes cell immortalization, migration and survival. *Oncotarget* 8, 57302–57315.
- Withers, J.B., Ashvetiya, T., and Beemon, K.L. (2012). Exclusion of Exon 2 Is a Common mRNA Splice Variant of Primate Telomerase Reverse Transcriptases. *PLoS One* 7, e48016.
- Withers-Ward, E.S., Kitamura, Y., Barnes, J.P., and Coffin, J.M. (1994). Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev.* 8, 1473–1487.
- Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W., and Shay, J.W. (1996). Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.* 18, 173–179.
- Wu, X., Li, Y., Crise, B., and Burgees, Shawn, M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* (80-.). 300, 1749–1751.
- Wu, X.L., Griffin, K.B., Garcia, M.D., Michal, J.J., Xiao, Q., Wright, R.W., and Jiang, Z. (2004). Census of orthologous genes and self-organizing maps of biologically relevant transcriptional patterns in chickens (*Gallus gallus*). *Gene* 340, 213–225.

Xi, L., and Cech, T.R. (2014). Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Res.* *42*, 8565–8577.

Xu, B., Dong, W., Yu, C., He, Z., Lv, Y., Sun, Y., Feng, X., Li, N., F. Lee, L., and Li, M. (2004). Occurrence of avian leukosis virus subgroup J in commercial layer flocks in China. *Avian Pathol.* *33*, 13–17.

Yang, F., Xian, R.R., Li, Y., Polony, T.S., and Beemon, K.L. (2007a). Telomerase reverse transcriptase expression elevated by avian leukosis virus integration in B cell lymphomas. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 18952–18957.

Yang, F., Lei, X., Rodriguez-Palacios, A., Tang, C., and Yue, H. (2013). Selection of reference genes for quantitative real-time PCR analysis in chicken embryo fibroblasts infected with avian leukosis virus subgroup J. *BMC Res. Notes* *6*, 402.

Yang, M.Q., Koehly, L.M., and Elnitski, L.L. (2007b). Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Comput. Biol.* *3*, 733–742.

Zhang, Y.-W., Su, Y., Volpert, O. V, and Vande Woude, G.F. (2003). Hepatocyte growth factor/scatter factor mediates angiogenesis through positive VEGF and negative thrombospondin 1 regulation. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 12718–12723.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.

Zhao, L.-H., Liu, X., Yan, H.-X., Li, W.-Y., Zeng, X., Yang, Y., Zhao, J., Liu, S.-P., Zhuang, X.-H., Lin, C., et al. (2016). Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat. Commun.* 7, 12992.

Zhu, J., Zhao, Y., and Wang, S. (2010). Chromatin and epigenetic regulation of the telomerase reverse transcriptase gene. *Protein Cell* 1, 22–32.

Zhu, S., Pan, W., Song, X., Liu, Y., Shao, X., Tang, Y., Liang, D., He, D., Wang, H., Liu, W., et al. (2012). The microRNA miR-23b suppresses IL-17-associated autoimmune inflammation by targeting TAB2, TAB3 and IKK- α . *Nat. Med.* 18, 1077–1086.

Curriculum Vitae

SANANDAN MALHOTRA
3400 N Charles St, Baltimore, MD - 21218, USA
smalhot7@jhu.edu +1(574) 387-1109

EDUCATION

Johns Hopkins University, Baltimore, USA

August 2013-Present
PhD Candidate, Biology
Cumulative GPA: 3.90

The Ohio State University, Columbus, USA

September 2009- March 2012
Bachelor of Molecular Genetics
Cumulative GPA: 3.99

AWARDS AND FELLOWSHIPS

- Dean's Teaching Fellowship for teaching an independent course on "Emerging Infectious Diseases", Fall 2017, Johns Hopkins University, USA
- Victor Corces award for excellence in teaching, Genetics Laboratory, Fall 2014, Johns Hopkins University, USA
- Morgan-Anfinsen fellowship for international graduate fellows, 2013, Johns Hopkins University, USA
- SOLAR Foundation Scholarship for excellence in undergraduate research, 2012, Ohio State University, USA
- Sharps Merit Scholarship for academic distinction, 2011-2012, Ohio State University, USA
- Dean's List, 2009-2012, Ohio State University, USA

RESEARCH EXPERIENCE

Graduate Student – Dr. Karen L. Beemon

Department of Biology, Johns Hopkins University, Baltimore, Maryland.

August 2013-Present

- Research in cancer biology using retroviruses as an insertional mutagenesis tool, to identify novel cancer genes.
- Role: assist with protocol development, experiment design and review; data collection, interpretation and analysis.

Research Assistant- Dr. Paul K. Herman, Department of Molecular Genetics The Ohio State University, Columbus, Ohio.

July 2012 – May 2013

- Research on molecular biology of P-bodies, ribonucleoprotein complexes using baker's yeast - *Saccharomyces cerevisiae* as a model system.
- Role: serving as a hands-on technician to provide accurate statistical analysis of research data and prepare reports in accordance with internal guidelines and lab requirements.

Undergraduate Student Researcher – Dr. Mark E. Drew, Center for Microbial Interface Biology The Ohio State University, Columbus, Ohio.

December 2010-March 2012

- Bachelor's honor thesis project: Research in parasite *Plasmodium falciparum*, in erythrocyte culture work to identify novel therapeutic targets for malaria treatment.
- Role: experiment designing and report analysis.

TEACHING EXPERIENCE

Johns Hopkins University

- Teaching Assistant, Genetics Laboratory, Fall 2014
- Teaching Assistant, Developmental Biology Laboratory, Spring 2015
- Teaching Assistant, Eukaryotic Molecular Biology, Fall 2015
- Teaching Assistant, Cancer Biology, Spring 2016 and Spring 2017
- Teaching Assistant, Genetics Lecture, Fall 2016
- Instructor, Emerging Infectious Diseases, Seminar Course, Fall 2017

CONFERENCE PRESENTATIONS

- **Malhotra S**, Justice J, Morgan R, and Beemon K. ALV-J insertional mutagenesis induces MET overexpression in chicken hemangiomas. Oral presentation at the Retroviruses Meeting, Cold Spring Harbor, NY, May 2015.
- **Malhotra S**, Winans S, Nehyba J, Justice J, and Beemon K. ALV integrations into TERT promoter-associated lncRNA in B-cell lymphomas. Poster presentation at the Retroviral Pathogenesis Meeting, Mülheim an der Ruhr, Germany, August 2015.

PEER-REVIEWED PUBLICATIONS

- **Malhotra S.**, Freeberg M., Winans S., Taylor J., Beemon K. (2017). A novel long non-coding RNA in the hTERT promoter region regulates hTERT expression. (manuscript in review).
- **Malhotra S.**, Winans S., Lam G., Justice J., Morgan R., Beemon K. (2017). Selection for avian leukosis virus integration sites determines the clonal progression of B-cell lymphomas. PLoS Pathog. 2017 Nov 3;13(11):e1006708.
- Winans S., Flynn A., **Malhotra S.**, Balagopal V., Beemon K. (2017). CTDSPL and CTDSPL2 identified as a common integration site in ALV-induced B-cell lymphomas. Oncotarget. 2017; 8:57302-57315.
- Nehyba J.*, **Malhotra S.***, Winans S.*, O'Hare T, Justice J. 4th, Beemon K. (2016). Avian leukosis virus activation of an antisense RNA upstream of TERT in B-cell lymphomas. J Virol. 90(20). (* indicates equal contribution).
- **Malhotra S**, Justice J 4th, Lee N, Li Y, Zavala G, Ruano M, Morgan R, Beemon K. (2015). Complete genome sequence of an American avian leukosis virus subgroup J isolate that causes hemangiomas and myeloid leukosis. Genome Announc. 9;3(2).
- Justice J 4th, **Malhotra S**, Ruano M, Li Y, Zavala G, Lee N, Morgan R, Beemon K. (2015). The MET gene is a common integration target in avian leukosis virus subgroup J-induced chicken hemangiomas. J Virol. 89(9).

LEADERSHIP AND OUTREACH

Stem Achievement in Baltimore Elementary Schools (SABES)
September 2015 – present

Maryland State High School Science Olympiad Volunteer, Baltimore and
College Park, USA. September 2015 – present

CERTIFICATIONS AND ACCREDITATIONS

- Peer reviewed scientific journals of Journal of Virology, Veterinary Microbiology, and Veterinary Research.
- National Ambassador for SOS Children's Villages of India 'Child Caring for Child' (2004-2014). □
- Ambassador for NCPEDP (National Center for Promotion of Employment for Disabled people) 'Help My Friends Learn' campaign to spread awareness on the cause of disability (2009-13).
- Certified and merited support for Cancer Patients Aid Association (2008-2012).

REFERENCES

Karen Beemon, PhD, Professor, Department of Biology, Johns Hopkins University, USA, KLB@jhu.edu +1(410) 516-7289

Beverly Wendland, PhD, Dean of Krieger School of Arts and Sciences, Johns Hopkins University, USA, bwendland@jhu.edu, +1(410) 516-4065

M. Andrew Hoyt, PhD, Professor, Department of Biology, Johns Hopkins University, USA, hoyt@jhu.edu +1(410) 516-7299

Paul Herman, PhD, Professor, Department of Molecular Genetics, The Ohio State University, USA, Herman.81@osu.edu +1(614) 688-5499

Mark Drew, PhD, Professor, Division of Infectious Diseases, The Ohio State University, USA, Mark.Drew@osumc.edu +1(614) 247-5337