Long non-coding RNAs (LncRNAs), viral oncogenomics, and aberrant splicing events: therapeutics implications

Rodney Hull¹, Zukile Mbita², Zodwa Dlamini¹

¹SA-MRC/UP Precision Prevention & Novel Drug Targets for HIV-Associated Cancers Extramural Unit, Pan African Cancer Research Institute, University of Pretoria Hatfield, 0028, South Africa; ²Department of Biochemistry, Microbiology and Biochemistry, University of Limpopo, Sovenga 0727, South Africa

Received October 2, 2020; Accepted December 21, 2020; Epub March 1, 2021; Published March 15, 2021

Abstract: It has been estimated that worldwide up to 10% of all human cancers are the result of viral infection, with 7.2% of all cancers in the developed world having a viral aetiology. In contrast, 22.9% of infections in the developing world are the result of viral infections. This number increases to 30% in Sub-Saharan Africa. The ability of viral infections to induce the transformation of normal cells into cancerous cells is well documented. These viruses are mainly Hepatitis B and C viruses, Epstein Barr virus, Human papillomavirus and Human Cytomegalovirus. They can induce the transformation of normal cells into cancer cells and this may be the underlying cause of carcinogenesis in many different types of cancer. These include liver cancer, lymphoma, nasopharyngeal cancer, cervical cancer, gastric cancer and even glioblastoma. Long non-coding RNAs (LncRNAs) can function by regulating the expression of their target genes by controlling the stability of the target mRNAs or by blocking translation of the target mRNA. They can control transcription by regulating the recruitment of transcription factors or chromatin modification complexes. Finally, LncRNAs can control the phosphorylation, acetylation, and ubiquitination of proteins at the post-translation level. Thus, altering protein localisation, function, folding, stability and ultimately expression. In addition to these functions, LncRNA also regulate alternate pre-mRNA splicing in ways that contribute to the formation of tumours. This mainly involves the interaction of LncRNAs with splicing factors, which alters their activity and function. The ability of LncRNAs to regulate the stability, expression and function of tumour suppressor proteins is important in the development and progression of cancers. LncRNAs also regulate viral replication and latency, leading to carcinogenesis. These factors all make LncRNAs ideal targets for the development of biomarker arrays that can be based on secreted LncRNAs leading to the development of affordable non-invasive biomarker tests for the stage specific diagnosis of tumours. These LncRNAs can also serve as targets for the development of new anticancer drug treatments.

Keywords: Non-coding RNAs, hepatitis C virus, hepatitis B virus, epstein barr virus, human papilloma virus, hepatocellular carcinoma, nasopharyngeal cancer, gastric cancer, head and neck cancer, lymphoma, glioma

Introduction

In 1909, Dr. Francis Peyton Rous extracted tumour cells from a chicken and grafted these cells into other chickens. These chickens were then found to also develop tumours [1]. The transforming factor was eventually identified as the RNA virus, the Rous sarcoma virus. This discovery led to Dr. Rous receiving a Nobel prize for medicine in 1966 [2]. It is estimated that up to 10% of all human cancers are caused by infection with oncogenic viruses [3]. The number of cancer cases that are attributable to viral infections is much higher in developing nations where 22.9% of all cancers are the result of infection with oncogenic viruses, compared to 7.2% in the developed world [4]. However, in general, only a relatively small percentage of individuals infected with an oncovirus develop cancer. An indication of the number of cancer cases caused by oncoviruses contribute to the total number of cancer cases is given in Figure 1. This percentage is even higher in Sub-Saharan Africa where up to 30% of all cancers are due to infections with oncogenic viruses [5]. These include the avian Rous sarcoma virus, cottontail rabbit papillomavirus, mouse mammary tumour virus [4], adenovirus and simian virus 40 (SV40), Epstein-Barr virus (EBV), human T-cell lymphoma virus-1 (HTLV-1), high
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Figure 1. Prevalence of different cancers with a viral aetiology. The cancers that can potentially be caused by viral infection, accounted for approximately 23% of all cancer cases in 2018.

Table 1. Oncoviruses and their associate cancers

<table>
<thead>
<tr>
<th>Virus</th>
<th>Taxonomy</th>
<th>DNA/RNA</th>
<th>Oncogenes</th>
<th>Tumour type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenovirus 12, 18</td>
<td>Adenoviridae</td>
<td>DNA</td>
<td>E1A, E1B</td>
<td>Solid tumours in rodents</td>
</tr>
<tr>
<td>BK virus</td>
<td>Papovaviridae</td>
<td>DNA</td>
<td>T antigens</td>
<td>Solid tumours in primates and rodents</td>
</tr>
<tr>
<td>EBV</td>
<td>Herpesviridae</td>
<td>DNA</td>
<td>LMP-1, BARF-1</td>
<td>Burkitt's lymphoma Hodgkin's lymphoma, stomach cancer nasopharyngeal carcinoma</td>
</tr>
<tr>
<td>HBV</td>
<td>Hepadnaviridae</td>
<td>DNA</td>
<td>HBx</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>Hepatitis C</td>
<td>Flavaviridae</td>
<td>RNA</td>
<td>-</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>HPV 16, 18, 21, 45</td>
<td>Papillomaviridae</td>
<td>DNA</td>
<td>E6, E7</td>
<td>Cervical, anal, oral cancers</td>
</tr>
<tr>
<td>hTCLV1</td>
<td>Retroviridae</td>
<td>RNA</td>
<td>Tax</td>
<td>Adult T-cell leukemia lymphoma</td>
</tr>
<tr>
<td>KHSV</td>
<td>Herpesviridae</td>
<td>DNA</td>
<td>vGPCR</td>
<td>Kaposi's sarcoma</td>
</tr>
<tr>
<td>Merkel cell polyomaviridea</td>
<td>Papovaviridae</td>
<td>DNA</td>
<td>T antigen</td>
<td>Merkel cell carcinoma</td>
</tr>
<tr>
<td>Human cytomegalovirus (HCMV)</td>
<td>Cytomegaloviridae</td>
<td></td>
<td>DNA</td>
<td>Liver cancer</td>
</tr>
</tbody>
</table>

Table 1. Oncoviruses and their associate cancers

risk human papillomaviruses (HPVs), Kaposi’s sarcoma-associated herpes virus (KSHV), Merkel cell polyomavirus (MCV) [6]. Some of these viruses as well as details concerning their taxonomy, the nature of their genetic material, the oncogenes they express and their related cancers are given in Table 1. The way in which these viruses are able to initiate cancer development was first discovered by Harold Varmus and Michael Bishop in 1976. They identified the viral oncogene v-src. Further characterisation of other oncogenes led to the discovery of p53, which is an antigen-associated protein for viral oncogenes such as SV40 [7]. Viral
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Oncogenes manipulate the infected cells' machinery in order for the virus to replicate and survive in the host. The expressed proteins then lead to the transformation of the cells. Alternate mRNA splicing is required for the effective expression of viral oncogenes. This process allows for the generation of a more diversified viral transcriptome [8]. In addition to viral oncogenes that code for proteins, these viruses encode multiple non-coding RNAs (ncRNAs).

NcRNAs are functional RNA molecules that are not translated into proteins, but are transcribed from DNA, and these include housekeeping (transfer RNAs, ribosomal RNAs, small nuclear RNAs, small nucleolar RNAs, telomerase RNAs) and regulatory noncoding RNAs. These regulatory non-coding RNAs include both small non-coding and long noncoding RNAs. One of the common types of noncoding RNAs are long non-coding RNAs (lncRNAs). LncRNAs are non-coding transcripts that are more than 200 nucleotides long and Next-generation sequencing (NGS) technology has revealed that lncRNAs are involved in differentiation, proliferation, and cell death [9, 10]. As a result, these molecules play a role in cancer development and progression [11]. Viral infections lead to changes in the host transcriptome including the expression of lncRNAs. Infected cells may also express viral lncRNAs [12].

Long non-coding RNAs

The majority of lncRNAs are transcribed by RNA polymerase II, and they are 5'-capped, spliced, and polyadenylated to form a structure similar to mRNA. Multiple large-scale projects have been undertaken to identify lncRNAs and have resulted in the identification over 20,000 lncRNAs [13]. LncRNAs function by interacting with either DNA or RNA or proteins, influencing the formation and function of the secondary and tertiary structures of these molecules. When they are localised within the nucleus, they can guide chromatin-modifying-complexes or transcription factors. In the cytosol, they can control the stability of mRNA or compete with endogenous mRNA for access to the protein expression machinery [14].

lncRNAs can be classed by their mode of action into four separate categories or by their genomic origin into seven categories. In terms of genomic location, the seven groups are as follows: i) Intergenic lncRNAs are coded for by intergenic regions between protein coding genes, ii) Sense or intronic lncRNAs are encoded by intronic regions in protein coding genes, iii) Antisense lncRNAs are encoded by the opposite strand, iv) Bi-directional lncRNAs are transcribed in the promoter regions of protein coding genes, v) Enhancer lncRNAs are transcribed from the enhancer regions [15], vi) Circular RNAs result from the joining of mRNAs or ncRNAs to other RNA molecules at 3'upstream or 5'downstream splice sites forming a looped circular RNA [16], and finally, vii) pseudogenes that can originate from gene duplication and may lose the ability to code for proteins due to mutations [17].

In terms of mode of action, lncRNAs can be broadly divided into four categories as follows: i) Decoy lncRNAs bind protein or RNAs, resulting in the negative regulation of protein expression, ii) These include Guide lncRNAs direct protein localisation by binding to proteins, iii) Signal lncRNAs interact with transcription factors or chromatin modifying enzymes, resulting in the regulation of transcription, and signalling pathways, iv) The final class are scaffold lncRNAs that act as an organising structure where molecules can bind and interact with each other more easily [18, 19]. These broad categories group the many ways that lncRNA can control gene expression (Figure 2), through epigenetic silencing, splicing regulation, sequestering miRNAs, protein interaction, and genetic variation [18, 20, 21]. Since lncRNAs are involved in many important biological processes, their aberrant expressions result in the development of diseases, especially cancer (Figure 2) [22, 23]. Non-coding RNAs have been shown to play an important role in tumour viruses, where non-coding RNAs might utilize non-coding RNAs to manipulate gene expression in an infected cell [24].

The role played by lncRNAs in different cancers

Hepatocellular carcinoma (HCC): The development of cancer is a multistep process resulting from chronic inflammation, DNA damage, chromosomal instability, epigenetic modifications, senescence, and telomerase reactivation. These processes can also promote carcinogenesis in hepatocytes through genetic dam-
Regulation of gene expression by LncRNAs. LncRNAs can regulate gene expression in a number of ways. Firstly, through chromatin remodelling, by recruiting chromatin modifying complexes to specific genomic loci. LncRNAs can also regulate transcription by either recruiting transcription factors to their target gene promoters or by blocking transcription factors from binding to target gene promoters. Decoy lncRNAs bind miRNAs acting as a miRNA sponge sequestering them and altering protein expression. They can also be processed to produce small-interfering RNAs (siRNAs) leading to degradation of target mRNA. Finally, they can bind directly to mRNA and regulate mRNA stability, or competitively bind to mRNA to improve mRNA stability. Guide lncRNAs direct protein localisation. LncRNAs also bind to proteins and control protein phosphorylation, acetylation, and ubiquitination at the post-translation level. Alternative splicing can be regulated by lncRNA which interact with splicing factors to regulate splicing.
The expression of different lncRNAs in HCC depends on the stage of HCC and whether the cancer was caused by HCV infection [29]. HCV infection is a major cause HCC [30], as a result of the virus contributing to the development and progression of fibrosis [31]. HCV is a positive-strand RNA virus that does not integrate into the host cells genome. The viral genome consists of a single reading frame that codes for a large 3000 amino acid protein that is then processed into 10 smaller proteins [28]. The 10 smaller proteins include structural, core, E1, and E2, and non-structural, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B, proteins that play important roles in the viral life cycle and interact with the components of the host cell to regulate cell signalling, transcription, cell proliferation, apoptosis, vesicular trafficking, and translational regulation [32, 33]. These proteins can also promote carcinogenesis in hepatocytes by inducing genetic damage and epigenetic dysregulation [25]. By comparing lncRNAs in healthy liver tissue, preneoplastic lesions and HCC, seven lncRNAs were identified with de-regulated expressions. These include LINC01419, BC014579, BC014579, AK021443, RP11-401P9.4, RP11-304 L19.5, AF070632 and CTB-167B5.2 [34]. In early HCC stages, LINC01419 transcripts are found to be highly expressed. The expression levels of this lncRNA were even higher in HCV-related HCC. The levels of both the AK021443 and AF070632 lncRNAs change as the cancer progresses. In advanced stage HCC, AK021443 levels were found to be increasing, while the expression of AF070632 decreases. LncRNAs such as AK021443, whose level decreases as HCC progresses were found to be responsible for metabolic processes not required for cancer cells. i.e. the Krebs cycle and oxidative phosphorylation, while the lncRNAs LINC01419 and AF070632 were found to be highly expressed in HCC, and are involved in the regulation of cell cycle genes [34].

Prader Willi/Angelman region RNA 5 (PAR5), and the lncRNA hypoxia-inducing factor a (aHIF) are two lncRNAs that are downregulated in HCV-related HCC. At the same time, the lncRNA human downregulated expression by HBx (hDREH) was up-regulated in HCV-related HCC [35, 36]. The lncRNA, Hox transcript antisense intergenic RNA (HOTAIR) is highly expressed in HCC and is associated with elevated incidence of metastases and therefore, poor prognosis in [37]. Through the control of histone methylation, HOTAIR regulates the expression of homeobox D (HOXD), transcription factors that control developmental pathways. Increased levels of HOTAIR also lead to the overexpression of HER2 by binding to and sequestering the miRNA that controls the expression of HER2 and regulating the stability of HER2 mRNA [38]. Two other lncRNAs have been implicated in the development and progression of HCC, WRAP53 and EGOT. The expression of the tumour suppressor TP53 is controlled through the action of the lncRNA WD repeat containing antisense to TP53 (WRAP53) [36]. Finally the urothelial carcinoma associated-1 (UCA1) binds to miR-203, sequestering it, leading to increased expression of the target of this miRNA, the transcription factor Snail2 [36]. The long non-coding RNA EGOT is expressed during HCV infection and acts against the hosts’ antiviral response resulting in increased viral replication. The expression of the miR-33a-5p results in the formation of an EGOT/miR-33a-5p/ HMGA2 complex and acts to inhibit EGOT function [39]. The chromatin interaction protein, High mobility group protein A2 (HMGA2), binds to the chromatin to change its structure to alter the interaction of transcription factors with DNA and changing gene transcription, thereby contributing to tumorigenesis. The expression of HMGA2 is inhibited by the miRNA 33-a-5p [40]. EGOT and miR-33-a-5p share a binding site on HMGA2. EGOT thereby blocks HMGA2 activation, promoting cancer development and progression [41]. Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) is another
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

pro-oncogenic IncRNA that is highly expressed in HCV positive HCC [42].

In addition to HCV, HBV is another aetiological factor in liver cancer development. The IncRNA, highly up regulated in Liver Cancer (HULC) is expressed at high levels in HCC and activates HBV replication. The IncRNA achieves this by decreasing the expression of the cell cycle regulation gene, APOBEC3B, and by HULC up-regulating microRNA-539. HULC also increases the stability of HBV covalently closed circular DNA through this inhibition of APOBEC3B [43]. The HBV X protein alters the expression of lncRNAs, which can assist in the development and progression of HCC. One of these deregulated lncRNAs is the Semaphorin 6A antisense RNA 1 (SEMA6A-AS1). In HBV positive HCC, the levels of SEMA6A-AS1 are reduced and this reduced expression of SEMA6A-AS1 inversely correlates with tumour stage, metastasis, and mortality. This low level of SEMA6A-AS1 expression is also a good indicator of poor treatment outcome in HBV-associated HCC [44].

Lymphoma: EBV infection of human primary resting B lymphocytes (RBLs) results in these cells expressing viral latency genes, which alter gene expression. This can result in the development of cancer through the transformation of RBLs to lymphoblastoid cells. Once an individual is infected with EBV, the virus is present in a latent form in B lymphocytes for the rest of the individual's life. The presence of latent viral infection in the lymphocytes can lead to the development of cancer such as Burkitt's lymphoma and Hodgkin's lymphoma [45]. During this transformation, the expression of a number of lncRNAs changes such as CYTOR and NORAD that play an important role in cell growth. The increased expression of these RNAs is directly associated with cell transformation [46]. NORAD is also involved in stimulating DNA replication and repair by binding to proteins involved in this process [47]. The lncRNAs 7SL, H19, and H19-AS play a role in the development of many cancers including lymphomas [48]. Infected lymphocytes release nine lncRNAs into the extracellular spaces through the use of exosomes. These include 7SL, H19, H19 upstream conserved 1&2, H19 antisense, HAR1B, HOXA6as, NDM29, SNHG5, and Tsix. Of these, only H19 and H19-AS were up-regulated [49].

In addition to altering the expression of host IncRNA, EBV encodes its own ncRNAs that can influence the development and progression of cancer, named EBV-encoded small RNAs (EBERs). Some of these RNAs accomplish this by binding to host proteins, forming ribonucleoprotein (RNP).

Nasopharyngeal carcinoma: EBV infection is a common risk factor for many different types of lymphoid cancer. This includes nasopharyngeal carcinoma (NPC). The viral proteins associated with the development of NPC are encoded for by the viral RNA transcript, Bam-HI A rightward transcripts (BARTs). This transcript also codes for viral microRNAs and IncRNAs. These IncRNAs affect the expression of cell adhesion, oxidoreductase activity, inflammation, and immunity genes [50]. One of these genes, IKAROS family zinc finger 3 (IKZF3/Aiolos), plays a role in lymphocyte development and cell attachment. BART IncRNA induces the expression of this gene. It is also likely that the expression of BART IncRNA interacts with the hosts’ chromatin remodelling machinery to control and maintain EBV latency [50].

Genome wide association studies and NGS have revealed that the expression of 2192 lncRNAs change in the development and progression of NPC. This results in the upregulation of 62 genes that are controlled through these lncRNAs [51]. These include CD44 (Hyaluronan/CD44) and interleukin 1 receptor associated kinase 1 (IRAK1), IRAK is involved in EBV infection while CD44 is involved in cancer progression [51]. LncRNAs that were downregulated in NPC include NR2F2 antisense RNA1 (NR2F21), IncRNA-family with sequence similarity 95 member C (FAM95C), and long intergenic nonprotein coding RNA 1106 (LINC01106). Those that were upregulated include IncRNA-CH507-513H4.6, IncRNA-THAP9 antisense RNA 1 (THAP9-AS1), IncRNACH507-513H4.3 and IncRNA-RP4-794H19.1 [51].

EBV expresses many miRNAs, which can interact with host lncRNAs to regulate their function. NPC associated gene 7 (NAG7) is a lncRNA that acts as a tumour suppressor gene by inhibiting cell cycle progression and promoting apoptosis. EBV expresses a ncRNA called EBER1, which negatively regulates NAG7 promoting metastasis [52]. EBV-miR-BART6-3p binds to lncRNA LOC553103 leading to a decrease in
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Figure 3. Genomic organization of the human papillomavirus type 16. ORFs deduced from the DNA sequence are designated E1 to E7, and L1 and L2. The non-coding region (NCR, also known as a long control region) is also shown. AE and AL indicate early and late polyadenylation sites. The functions of these genes are given by the color coded boxes.

the expression of this, tumour suppressor, IncRNA [53].

All these previously discussed IncRNAs act as tumour suppressors and their expression negatively correlates with NPC progression. However, some circulating IncRNAs found in the blood of NPC patients are associated with a poor prognosis. These include the IncRNAs MALAT1, APAF1-AS1, and AL359062 [54].

Head and neck cancer: Head and neck squamous cell carcinoma (HNSCC) are the sixth most common cancer worldwide. It has a poor prognosis and does not respond well to treatment, being associated with high relapse rates. Recent studies have reported that ncRNAs might play critical role in the development and progression of HNSCC [55]. Other than tobacco and alcohol, some of the biggest risk factors for the development of this cancer is HPV infection [56]. HPV-related HNSCC has a different molecular basis underlying the progression of the disease compared to non-HPV related HNSCC and therefore, has different treatment outcomes as well.

A comparison between HPV positive and HPV negative HNSCC revealed that there are 132 IncRNAs that appeared in the HPV positive group and not the HPV negative group [57]. The four IncRNAs HOTAIR, PROM1, CCAT1, and MUC19 are found in HPV positive HNSCC and are associated with a decreased number of myeloid-derived suppressor cells (MDSCs). These immune cells are normally found in high numbers in individuals suffering from infection or diagnosed with cancer. They play a role in tumour immunity suppression and have been observed to be elevated in HNSCC [58], with even higher numbers of MDSCs in HPV positive HNSCC. This implies that one of the ways in which HPV can promote the formation of HNSCC is by manipulating the immune system [59]. Since the four IncRNAs mentioned above seem to inhibit this process, they may be targets for the HPV E7 oncoprotein. The main function of the E7 protein in HPV infection is to prevent cell death [57].

Cervical cancer: Cervical cancer is the third most common cancer worldwide and the fourth leading cause of cancer-associated mortality in women. The most important risk factor for cervical cancer is infection with HPV. This infection is possibly the initial cause of carcinogenesis in cervical cancer. The ability of HPV to induce cervical cancer relies on the viral oncoproteins E6 and E7 (Figure 3) to inactivate tumour suppressor proteins such as TP53and Rb. The expression of the viral proteins inhibits the tumour suppressor proteins such as TP53 by forming a complex with the E3 ubiquitin ligase UBE3A. This E6/E7/UBE3A complex ubiquitinates TP53, leading to its proteasomal degradation of TP53 [60, 61]. The expression of HPV-16 E6 also leads to changes in the expression of IncRNAs. The IncRNA, Damage-induced (DINO), is expressed in response to TP53 sig-
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Figure 4. The function of the IncRNA DINO as anti-oncogenic. Following DNA damage, the IncRNA Damage-induced (DINO), is expressed. DINO stabilises TP53 and disrupts the E6/E7/ubiquitin ligase complex which ubiquitinates TP53, leading to its degradation.

nalling and DNA damage. DINO functions to stabilise TP53 and is able to inhibit the viral protein induced degradation of TP53 by inhibiting the formation of the E6/E7/UBE3A complex. This stabilisation of TP53 results in an increase in the transcription of the targets of TP53 (Figure 4). In HPV positive cervical cancer, the expression of DINO is downregulated. The degradation of TP53 resulting from the complex formed by the E3 ligase and viral protein is thought to be the result of the decreased expression of DINO [62].

Some of the other IncRNAs whose expression is disrupted by HPV infection include GAS5, H19, and FAM83H-AS1 [63]. FAM83H-AS1 expression was increased following E6 expression and led to increased cellular proliferation and decreased apoptosis. Increased levels of FAM83H-AS1 is associated with a poor prognosis [63]. NGS analysis identified 19 IncRNAs, 3 novel IncRNAs were found to be differentially expressed in three cervical cancer patients infected with HPV16. All of these IncRNAs play roles in the carcinogenesis and development of cervical cancer [64].

Gastric cancer: Gastric cancer is most common in South America (ASR 13.8) and Asia (ASR 9.5) [65]. The mortality rate amongst gastric cancer patients is high. Risk factors for gastric cancer include H. pylori infection, which leads to the formation of ulcers, gut inflammation, pernicious anaemia and the presence of polyps in the stomach [66]. Infection with EBV is also thought to be related to the development and progression of gastric cancer in up to 10% of gastric cancer patients [67]. Gastric cancer associated with EBV infection is known as EBV-associated gastric carcinoma (EBVaGC). EBV infection leads to changes in IncRNA expression, that may contribute to the development and progression of gastric cancer.

The expression of the IncRNA Small nucleolar RNA host gene 8 (SNHG8) is increased in gastric cancer cells infected with EBV [68]. The level of expression is a good indicator of tumour
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Figure 5. Types of brain cancer by incidence. Malignant brain tumours are named after the tissue from which they develop. Gliomas are the most common type of malignant brain tumour.

stage with increased expression indicating a more advanced stage of cancer progression. This lncRNA influences the expression of multiple genes such as TRIM28 (a mediator of gene silencing), EIF4A2 (RNA helicase involved in mRNA translation), NAP1L1 (promoter of cell proliferation), PLD3 (Regulator inflammatory cytokine responses), RPL18A (a ribosome unit), and TRPM7 (a regulator of apoptosis and necrosis) [68]. The downregulation of this lncRNA inhibits cell growth and causes cell cycle arrest and increases apoptosis [69]. RNAs that regulate other RNAs by competing for binding to miRNAs that target these other RNAs, are known as competing endogenous RNAs (ceRNAs). Two lncRNAs were identified as playing a role in the EBV associated gastric cancer by acting as ceRNAs. These lncRNAs, RP5-1039K5.19 and TP73-AS1, regulate signalling pathways [70]. The pro-oncogene Guanine nucleotide-binding protein beta polypeptide 2-like promotes cell growth and metastasis. The expression of this protein is promoted by the lncRNA OR3A4 in virus-associated gastric cancer [71].

Brain cancers: Malignant brain tumours can be classified according to the tissue or cell type that gives rise to them (Figure 5). Gliomas start in the glial cells of the brain or spine. Malignant gliomas are highly invasive and have a high mortality rate [72]. Risk factors that contribute to the development of gliomas include exposure to ionizing radiation, polymorphisms in genes involved in DNA repair and cell cycle regulation as well as infection by retroviruses such as cytomegalovirus reviewed in [73]. Human cytomegalovirus (HCMV) is a member of the herpes virus family that is associated with the development of gliomas. The virus expresses more than 200 proteins, which can be divided into three groups based on the stage during which they are expressed. These three categories are immediate early (IE), early (E), and late (L) [74]. HCMV genomic sequences as well as HCMV proteins have been isolated from gliomas. A high HCMV viral load in brain cancer patients is associated with lower survival times and poor treatment outcomes [75].

Latent infection also involved the expression of 646 lncRNAs (208 known lncRNAs and 438
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

novel lncRNAs) and decreased expression of 424 (140 known and 284 novel) [76]. Whole transcriptome sequencing of HCMV infected and uninfected CD14 (+) and CD34 (+) cells, led to the identification of two lncRNAs in the infected cells. These lncRNAs, named RNA4.9 and RNA2.7, enable the virus to exist in a latent state within these cells. It is now known that RNA4.9 is involved in the repression of host gene transcription [77] and is involved in promoting viral replication [78]. Another IncRNA, RNA1.2, is involved in the inhibition of the hosts’ immune response to the virus. RNA1.2 prevents the expression of IL-6 by inhibiting the NF-κB pathway [78].

The expression of IncRNAs also differ between glioma tissues and normal brain. The IncRNA FGDS-AS1 was overexpressed in glioma cells and is associated with the viability, migration, and invasion of glioma cells [79]. The expression of the IncRNA colon cancer-associated transcript-1 (CCAT1) is higher than in normal tissue. It functions as a competitor for the binding of miR-181b to target proteins such as fibroblast growth factor receptor 3 (FGFR3) and platelet-derived growth factor receptor (PDGFR). This leads to the decreased expression of these proteins. In this way, CCAT1 increases the expression of these growth factors leading to cancer progression [80]. The levels of the IncRNA LOC728196 is higher in glioma cells than in normal nerve tissue, where it prevents the inhibition of transcription factor 7 (TCF7). This transcription factor is involved in excessive activation of the Wnt/β-catenin signalling pathway [81].

The immune response to glioma and viral infection can be modulated by the IncRNA, LINCO0152. The exact mechanism behind this is unknown. Another nine IncRNAs are also related to the immune response surrounding glioma cells. These include Phosphoglucomutases 5-antisense RNA 1 (PGM5-AS1), ST20-antisense RNA 1 (ST20-AS1), ankyrin repeat and PH domain 2-antisense RNA 1 (AGAP2-AS1), MIR155 host gene (MIR155HG), SNHG8, LINCO0937, TUG1, MAPK activated protein kinase 5-antisense RNA 1 (MAPKAPK5-AS1) and HLA complex group 18 (HCG18). Additionally, the IncRNA TUG1 is also increased in glioma cells [82, 83].

Alternate splicing and IncRNA in various cancers: Alternate splicing is the mechanism whereby the pre-mRNA encoded by a single gene is processed into multiple transcripts. These can give rise to multiple protein isoforms which can have slightly different functions or can act in a completely antagonistic manner to the other protein isoforms. The splicing process is controlled by splicing factor proteins. These proteins are RNA binding proteins (RBPs) that interact with RNA through one or more RNA binding domains, these include domains such as the RNA recognition motif (RRM), the hnRNP K homology domain (KH) or the DEAD box helicase domain [84]. There are two major families of splicing factors, the serine arginine splicing factors (SR proteins) and the heterogeneous nuclear ribonucleoproteins (hnRNPs). The splicing of mRNA can be affected through the function of IncRNAs in one of two ways. The IncRNAs can bind to splicing factors and either influence the expression of specific isoforms or by forming nuclear paraspeckles to regulate gene expression.

The IncRNA, IncRNA-p21, is transcribed from the p21 locus. Like p21 the IncRNA is involved in regulating cell proliferation, apoptosis and DNA damage response [85]. Through its binding to hnRNPK, IncRNA-p21 is able to promote TP53-mediated expression of p21 and other YP53 targets by promoting the expression of TP53 transcripts [86]. SR proteins are phosphorylated/dephosphorylated in order to properly splice pre-mRNA. Hyperphosphorylation of SR proteins changes the binding affinity to target pre-mRNA and guiding the selection of splice sites [87]. The phosphorylation state also influences the movement of the SR protein. LncRNA regulates the phosphorylation state of the SR proteins and therefore regulates alternative splicing. Partially dephosphorylated SR proteins support the first steps of the trans-esterification reactions. Additionally, intranuclear trafficking of SR proteins between nuclear speckles and transcription sites is dependent on their phosphorylation status [88]. The IncRNA MALAT1 was found to regulate pre-mRNA alternate splicing by interacting with SR-splicing factors. It regulates the phosphorylation of these splicing factors, thereby changing the subcellular distribution of these SR proteins. This results in a change in the set of pre-mRNAs that the SR protein is able to interact with and regulate the splicing of [89].

LncRNAs such as MALAT1 are also able to alter splicing by hijacking splicing factors leading to
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Figure 6. The function of NEAT1 in the splicing of PPARy. NEAT1 interacts with and phosphorylates the splicing factor, SRp40. This splicing factor controls the pre-mRNA splicing of PPARy. Phosphorylation of the splicing factor leads to changes in the isoform that results from pre-mRNA splicing.

the selection of specific isoforms. For example, Polypyrimidine tract binding protein (PTBP2) is a proto-oncogene whose splicing is controlled by the splicing factor, SFPQ. SFPQ achieves this by binding to PTBP2 mRNA and altering the splicing of PTBP2 to favor an anti-tumour isoform. MALAT1 binds SFPQ and disrupts the PTBP2/SFPQ splicing complex from forming. This prevents SFPQ from altering the splicing of PTBP2 to favor anti tumourogenic isoforms (Figure 7) [90].

LncRNAs interact with RNA-binding proteins to form paraspeckles. These loci are distributed throughout the cell and are involved in gene expression. The IncRNA Nuclear Enriched Abundant Transcript 1 (NEAT1) has at least five known splice variants and is an essential component of these speckle. Of these five, the short variant is NEAT1-1 (3.7-kb) and the long variant is NEAT1-2 (23.7-kb) [91]. NEAT1 is also involved in the sequestration of related proteins. It is known to regulate the splicing of PPARy mRNA that results in the production of the two isoforms, PPARy1 and PPARy2. These proteins are involved in the differentiation of adipocytes. SRp40 is involved in PPARy splicing. SRp40 associates with NEAT1, which leads to the alteration in the splicing isoform of PPARy favoured by SRp40. Different levels of NEAT1 and SRp40 have different effects on the levels of these isoforms. With the level of NEAT1 expression changing the relative levels of PPARy isoforms (Figure 6) [92].

LncRNAs in cancer treatment

LncRNAs as biomarkers: Currently, the vast majority of diagnostic tests rely on the symptoms and detecting physiological changes as cancer progresses. The adoption of biochemical markers, known as biomarkers, that are related to the development and progression of the disease offers distinct advantages. These include price, speed, and decreased invasiveness. For instance, HCC is normally only diagnosed following the onset of symptoms, at which point the cancer is advanced and treatment options are limited. These advanced stages are also associated with a poor prognosis. The large number of lncRNAs whose expression is altered during viral associated carcinogenesis (Table 2) makes them an attractive choice for biomarkers, which can be used to screen individuals in a regular non-invasive screening program. The selection of specific lncRNAs may be useful for diagnosing HCV or HBV specific HCC, which may influence treatment choices. Two lncRNAs that may be used for this purpose in HCV-related HCC include UCA1 and WRAP53. Not only can these biomarkers be used to diagnose HCV-related HCC, but in the case of UCA, the expression level is also associated with the size of the tumour and its ability to invade other...
tissues and therefore, tumour stage [29]. This makes UCA an ideal IncRNA biomarker, since it is capable to diagnose and stage the cancer. The ideal IncRNAs to be used as a biomarker would be a panel of differentially expressed circulating IncRNAs [93].

Apart from using IncRNAs to diagnose and stage cancer, IncRNAs can also be used as a prognostic marker to predict patient survival. A circulating IncRNA that shows promise as a prognostic biomarker for HCC is the IncRNA Small Nucleolar RNA Host Gene 15 (SNHG15). Increased expression of this IncRNA results in decreased HCC patient survival [94], making it an ideal prognostic biomarker. Many of the IncRNAs that were previously discussed as useful diagnostic biomarkers would also be useful prognostic biomarkers. Other IncRNAs that may prove useful are the IncRNAs, MALAT1 and HEIH that have also been studied for their potential use as diagnostic biomarkers for HCV-related HCC. HEIH expression can be monitored in the blood [95]. Many alternatively expressed IncRNAs in EBV-associated lymphoma are exported from the cell. The release of some of these IncRNAs from the cells using exosomes is important for diagnostic purposes. These secreted IncRNAs allow for detection of many of these IncRNAs using blood tests [49].

**LncRNAs as drug targets:** The importance of IncRNAs in cancer development and progression makes them the ideal targets for the development of new treatments. By inhibiting the activity of these IncRNAs through the use of antagonists can inhibit the pro-carcinogenic activity of these IncRNA. This has been shown to be an effective potential strategy in multiple cancers such as HCC [96]. There are numerous methods whereby IncRNA can be therapeutically targeted. One of the methods that can be used to interfere with the activity of IncRNAs is through the use of small interfering RNAs to inhibit the expression of these pro-oncogenic IncRNAs. By silencing IncRNA using RNAi to knockdown expression of the target IncRNA, positive anti-cancer effects have been observed. Examples of silencing the expression of target IncRNAs include the IncRNA Linc00974, which is highly expressed in HCC [97, 98]. As a result, targeting Linc00974 with siRNA resulted in the inhibition of cell proliferation and invasion, and increased apoptosis in the tumour [98]. Another technique is to target the regulatory element of the IncRNA [97].

---

**Figure 7.** The role of MALAT1 in the regulation of alternative splicing. A. MALAT1 controls the phosphorylation of SRSF1 proteins in the nucleus, including the MALAT1-interacting SRSF1. Phosphorylated SRSF1 is accumulated in nuclear speckles (NS). Dephosphorylated SRSF1 promotes its interaction with mRNAs and transport into the cytoplasm. Here, the spliced mRNA is either translated by the ribosome or degradation by the non-sense mediated decay (NMD) machinery. B. MALAT1 binds to (hijacks) SFPQ, disrupting the SFPQ/PTBP2 complex. This results in the pro-oncogenic isoform of PTBP2, promoting tumour growth.
In order to effectively use these as an anticancer treatment, first the intracellular localization of cancer-related lncRNAs should be inhibited [99]. One of these solutions is the use of anti-
sense oligonucleotides (ASOs). These can enter the nucleus and induce RNase H-dependent degradation or sterically block lncRNA activity. ASOs can also be modified to increase their stability. The lncRNA MALAT1 has been experimentally targeted using ASO to decrease its levels. This was performed in a mouse model where it reduced the growth and metastasis of tumours [100]. The delivery of these ASOs is complicated as they have low levels of intracellular uptake and low stability in the body. New techniques of delivering these ASOs include the development of biocompatible nanoparticles [101]. Another technique involves blocking the binding of lncRNAs to protein targets and this applies to the lncRNAs that interact with chromatin remodelling complexes, transcription factors or those that act as protein chaperones by binding to protein targets [101].

Conclusions

Perturbations to the gene regulation network is an important cause for the development and progression of cancer. For a long time, studies have focused on changes in gene control that occur at the transcription and expression levels, neglecting the control of gene expression that occurs at the post-transcriptional level. This includes changes in mRNA splicing, and the control of gene expression by ncRNAs, such as lncRNAs and miRNAs. Unlike splicing and miRNAs, lncRNAs interact with DNA, RNA, and protein, meaning they act to regulate gene expression at the level of transcription, post transcription and expression.

The role of lncRNAs in enhancing viral infection-mediated carcinogenesis process provides an important insight into the ability of infection by these viruses to act as causative agents in the development of these cancers. This becomes especially important in cancers such as cervical cancer, where viral infection by HPV is the most important risk factor in cervical carcinoma. The ability of these viral infections to alter the expression patterns of IncRNA and facilitate tumourigenesis has led to them being considered as probable biomarkers for the future diagnosis or prognosis of these cancers. Since lncRNAs expression is altered at multiple stages of cancer development and with cancer progression means that specific patterns of lncRNA expression can be used to not only diagnose cancer, but stratify patients based on the stage of cancer and their need for treatment. The identification of changes in the expression of lncRNAs, has been advanced by technologies such as NGS and bioinformatics. These advances not only make the study of lncRNA expression changes easier but would help in the establishment of specific lncRNAs as biomarkers.

Acknowledgements

The authors would like to thank the South African Medical Research Council (SAMRC) for funding this research.

Disclosure of conflict of interest

None.

Address correspondence to: Zodwa Dlamini, SAMRC/UP Precision Prevention & Novel Drug Targets for HIV-Associated Cancers Extramural Unit, Pan African Cancer Research Institute, University of Pretoria Hatfield, 0028, South Africa. E-mail: Zodwa.Dlamini@up.ac.za

References

[8] Epstein MA, Henle G, Achong BG and Barr YM. Morphological and biological studies on a virus
Long Non-Coding RNAs in viral associated cancers: implications for therapeutics


[35] Zhang DY, Zou XJ, Cao CH, Zhang T, Lei L, Qi XL, Liu L and Wu DH. Identification and functional characterization of long non-coding RNA MIR22HG as a tumor suppressor for hepato-
Long Non-Coding RNAs in viral associated cancers; implications for therapeutics

Am J Cancer Res 2021;11(3):866-883


as an intermediate in HPV16 promoting myeloid-derived suppressor cell recruitment of head and neck squamous cell carcinoma. Oncotarget 2017; 8: 42061-42075.


noncoding RNA LINCO0152 is a potential prognostic biomarker in patients with high-grade glioma. CNS Neurosci Ther 2018; 24: 957-966.


[87] Cao W, Jamison SF and Garcia-Bianco MA. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. RNA 1997; 3: 1456-1467.


