Long non-coding RNAs (lncRNAs) have emerged as prolific regulators of gene expression. lncRNAs are RNA transcripts which do not code for proteins like “conventional” genes. lncRNA was once presumed to be non-functional genomic noise and biologically irrelevant. Recent work, however, has shown that lncRNAs are spatiotemporal ‘master regulators’ of the genome. Unlike double stranded DNA, single-stranded lncRNA folds internally to assume complex structures which allows it to recruit protein complexes such as Polycomb Repressive Complex 2 (PRC2) and repress genes. This lncRNA regulation was first shown in X inactivating specific transcript (Xist) in mammalian females, which inactivates one of two X chromosomes to prevent X gene and subsequent protein double-dose. Such whole-chromosome inactivation may also be applied in treatments for chromosome disorders such as Down’s syndrome. HOX transcript antisense RNA (HOTAIR) broadens the extent of lncRNA gene regulation, controlling hundreds of genes around the genome. HOTAIR’s widespread control has implications in cancer, as HOTAIR regulates tumour suppressor genes. lncRNA control also regulates immune system responses and initiated pathogenic infection. lncRNA regulation provides ‘fine control’ of genes, and a full understanding of lncRNA may improve diagnostic and therapeutic approaches to disease in the future.
Introduction

Most of the human genome is transcribed into RNA, but never translated into protein (Carninici et al. 2005, Flicek et al. 2014). This non-translated DNA is not, as was once thought ‘junk’ or functionless. Instead, it produces a diverse array of regulatory non-coding RNAs (Martin & Chang 2012, Brosnan & Voinnet, 2009). Long non-coding RNAs (lncRNAs) are those longer than 100 nucleotides that do not code for a protein. Instead, lncRNAs have functions ‘distinct’ from protein coding (Fitzgerald & Caffrey 2014). They assemble into structures and work as guides and modulators of protein complexes which regulate when and at which chromosomal point genes are expressed (Mercer and Mattick 2013). One such complex is the histone modifying protein complex Polycomb Repressive Complex 2 (PRC2). Histones are the proteins which package and organise DNA and PRC2 modifications to histones can repress gene expression. lncRNAs evolved as PRC2’s spatiotemporal directors (Lee 2012), ‘master’ overseers of genomic regulation (Nie et al. 2012). An LNCipedia database has been assembled for these ‘master’ regulators, cataloguing their characteristics. (Volders et al. 2012), (Volders et al. 2015), (Nie et al. 2012). The evolution of lncRNA explains some of the ‘fine control’ of gene expression (Luco 2013).

One of the first notable demonstrations of lncRNA genome regulation was the X inactivating specific transcript (Xist) (Brown et al. 1991a), (Brown et al. 1991b), (Borsani et al. 1991), (Brockdorff et al. 1991a). In mammalian females, one of the two X (sex) chromosomes are silenced during development to prevent double-dosage of X chromosome genes. Xist is a conserved lncRNA (Brockdorff et al. 1991) transcribed from the X chromosome to be silenced, which co-ordinates the repression of the entire chromosome in X inactivation (Brockdorff et al. 1992), guiding PRC2 to repress X genes.

Xist provided the initial evidence of lncRNA’s regulatory significance. Genomic studies have shown that such RNAs are widespread within the genome (Bertone P. et al. 2004), (Bernstein et al. 2006). With Xist demonstrating such powerful control over entire chromosomes, research into its application in silencing extra chromosomes in Down’s syndrome has begun (Jiang et al. 2013).

Investigation into other lncRNAs that could regulate gene expression led to the discovery of HOX transcript antisense RNA (HOTAIR) lncRNA, regulating hundreds of genes in a wide regulatory network (Rinn et al. 2007). HOTAIR regulates many genes involved in cancer development and metastasis (Gupta et al. 2010). lncRNAs also activate host immune system responses as well as allow pathogens to initiate infection (Carpenter et al. 2013). Biological complexity arises not from sheer quantity of genes, but the finer control of when and where they are expressed. lncRNAs have emerged as critical regulators of this control and have the potential to affect all areas of gene expression (Necsulea et al. 2014), (Kogo et al. 2011).
IncRNA Lessons in X-Inactivation

Analysis of complete mammalian genome has shown an abundance of non-coding RNA (Carninci et al. 2005). IncRNA relevance was first shown in the IncRNA orchestration of the X inactivation centre (Xic). Mammalian females have two X chromosomes, whereas males have one. To ensure equal X gene expression in males and females, approximately 1000 X-linked genes (Brown et al. 1991a) on one of the two, randomly selected X chromosomes are repressed in females. Such dosage compensation was first observed by Mary Lyons (Lyons 1961). The inactivated chromosome is turned into a compact “Barr Body” incapable of gene expression (Walker et al. 1991).

Protein complexes catalyse X-inactivation, but require IncRNAs to be directed to their targets. IncRNAs recruit and coordinate the activity of the repressive PRC2 protein complex to “turn off” one X chromosome. PRC2 is a multi-subunit protein complex (Margueron et al 2011) which, by adding repressive (methyl) marks to the histone proteins (package and organise DNA), can inhibit gene expression (Clapier & Cairns 2009). In the Xic, at least seven IncRNAs coordinate the actions of PRC2 to control X chromosome inactivation (Lee 2009). Xic demonstrates the ability of IncRNA to orchestrate regulation of gene expression (Figure 1).

This ability of RNA to form complex structures is essential to their ability to recruit proteins and guide them to control gene expression (Sharp et al. 2009). Xist is a 17 kilobyte IncRNA transcribed exclusively from the inactive X chromosome and it does not code for protein translation like conventional transcriptional RNA (Brown et al. 1992). Instead, it folds into a complex secondary structure which allows Xist docking to the repressive PRC2 complex. It loads PRC2 with a distinct structural motif, Repeat A (RepA), a tetra-loop loading platform for this repressive complex (Zhao et al 2008), (Duszczyk et al. 2011). This IncRNA folding is significant as complex folded structures are hallmarks of functional biological molecules. Once loaded to the chromosome by Xist, PRC2 adds methyl groups to the histone proteins to epigenetically repress expression of X genes. The bases in IncRNAs, unlike those in double stranded DNA, can fold in on each other and form stable structures, such as the tetra-loop RepA.
Figure 1. X-Chromosome Inactivation. The X-Inactivation centre (Xic) is present on both the active X chromosome (Xa) and the inactive X chromosome (Xi). Xic encodes for Xist and when expressed Xist RNA binds PRC2. PRC2 facilitates the initiation of inactivation of the X chromosome destined to be inactivated; Xi, through the direction of Xist RNA. Xist then propagates and through maintained interactions with PCR2 keeps Xi inactive.

Xist is the master lncRNA, initiating the process by spreading across the entire 3-D structure of the inactive chromosome (Clemson et al. 1996). It guides the repressive PRC2 which modifies the histone proteins packaging the DNA. Xist performs this job locally repressing the chromosome from which it is transcribed and regulation of the inactivation is provided by other lncRNAs.

The 40 kilobyte lncRNA Tsix is transcribed from and negatively regulates Xist on the active X chromosome, allowing gene expression (Lee et al. 1999). Xist is activated on the inactive chromosome by another lncRNA, Jpx. Jpx activates Xist allowing it to repress the inactive chromosome (Tian et al. 2010). Tsix and Jpx act as lncRNA ‘switches’ with opposing controls over Xist on either chromosome (Figure 2).
Figure 2. X-Inactivation centre (Xic) gene expression on Xi and Xa. Jpx and Tsix are the positive and negative regulators of Xist repression on either chromosome. Xa will produce Tsix. Jpx activates Xist on Xi. The RepA motif from Xist recruits the repressive PRC2 complex. The RepA motif then binds PRC2 and together with Xist will catalyse the repression of the Xi chromosome. Tsix on the other hand prevents PRC2 binding Xist and thus ensures the Xa remains active.

The designation of chromosomes as either ‘active’ or ‘inactive’ X’s is also controlled by lncRNA and is thought to be clonally maintained. Initially identical chromosomes become active or inactive after making physical contact. The “choice” of which X chromosome is to be silenced is made by Xite. Xite is an RNA element which enhances Tsix on the active chromosome alone leaving Xist to repress the inactive chromosome (Ogawa & Lee 2003).

Xic is illustrative of extensive, chromosome-specific lncRNA regulation. lncRNAs either repress chromosomes (Xist, Jpx) or activate them (Tsix). The chromosomes ‘choose’ which lncRNAs they will express by communicating through physical contact between Tsix and Xite RNA (Xu et al. 2006), (Lee 2009), (Ogawa & Lee, 2003), while ultimately repression of inactive X chromosome is performed by proteins (PRC2), spatiotemporal control of these proteins is lncRNA driven.
Applying Xist Lessons to Trisomy Disorders

Trisomy disorders develop in patients with three copies of any chromosome instead of the usual two. Down’s syndrome (DS) is a chromosomal disease caused by trisomy of chromosome 21 (Chr21). Xist’s ability to repress an entire chromosome could be applied in possible chromosomal therapies to turn off the supranumery DS chromosome. Jiang et al. (2013) tested this approach, applying Xist to cells derived from DS patients. Xist RNA “territories” were established in 85% of cells. 95% of Chr21 genes were repressed, bringing gene expression levels closer to normal, two-chromosome cells. This chromosome inactivation was maintained after three weeks, similar to inactive X chromosomes. This demonstrates that Xist lncRNA can silence extra chromosome in DS cells as it does to the inactive X. Most notably, Xist inactivation of Chr21 introduces the tentative possibility of corrections, or at least therapeutic options, for chromosomal disorders involving lncRNAs in the future (Disteche 2013).

HOTAIR broadens IncRNA influence

X inactivation’s demonstration of powerful lncRNA genetic regulation led to research into other functional IncRNAs controlling gene expression. Functional IncRNA is identified by demonstrating interactions with regulatory complexes. The RIP-seq technique developed by Zhao et al (2010) identified thousands of IncRNAs which bind to and control PRC2, thus repressing genes. Such widespread IncRNA regulation was previously suggested by Khalil et al. (2009). With thousands of IncRNAs guiding and modulating protein complexes, they have since been dubbed genomic ‘master regulators’ (Nie et al. 2012).

One such master regulator is the IncRNA HOTAIR identified by Rinn et al (2007) and Woo & Kingston (2007) regulating thousands of genes across the genome in an expansive regulatory network (Lee et al 2012). HOTAIR folds into a more elaborate structure than Xist, acting as a scaffold for multiple protein complexes which control gene expression (Tsai et al 2010). Further studies of HOTAIR identified a specific 89-nucleotide binding site for PRC2 (Wu et al. 2013). A precise structural analysis of HOTAIR revealed multiple motifs (helical sections, terminal loops, internal loops, and junctions) in four domains, some binding PRC2 (Somarowthu et al. 2015). Such folding was previously thought characteristic of proteins but IncRNAs can self-assemble into similarly complex structures. Structure is essential to function in biomolecules and is another demonstration of IncRNA biological significance.
IncRNA – Implications In Disease

IncRNAs are widespread gene expression controllers and biological complexity comes not from increasing the number of genes, but in precise control of when, where, and for how long they are expressed. This control is important in organising immune responses to pathogens and its dysregulation is involved in carcinogenesis.

Immune function and immunopathology

IncRNAs are important regulators of immune system genes controlling both pathogenic and host responses (Yu et al. 2015) and many immune genes are X-linked. IncRNAs activate host immune responses to pathogens by controlling expression of hundreds of immune system genes (Fitzgerald & Caffrey 2014), (Heward & Lindsay 2014). Immunity IncRNAs include ‘Nettoie Salmonella pas Theiler’s’ (NeST), which activates Interferon-γ (IFN-γ), a cytokine involved in defence against pathogenic infection (Gomez et al. 2013), (Baccala et al. 2005), (Hertzog et al. 2011). Toll-like receptors (TLRs) recognise pathogen-associated molecular patterns (PAMPs), molecules characteristic of pathogenic micro-organisms, helping to initiate inflammatory responses (Janeway & Medzhitov 2002). TLR4, after recognising Gram negative bacterial component lipopolysacharride, induces Cox-2 IncRNA expression. Cox-2 regulates hundreds of immune genes, repressing some and activating others to coordinate immune response (Guttman et al 2009), (Carpenter et al. 2013), (Li & Rana 2014).

TNFα and hnRNPL related immunoregulatory LincRNA (THRIL) activates tumour necrosis factor α (TNF-α) as well as other genes involved in the immune response (IL-8, CSF1, & CCL1) (Li et al. 2014). Genes encoding IL-8 and CCL5 are also activated by another IncRNA, nuclear enriched abundant transcript 1 (NEAT1) (Imamura et al. 2014).

Pathogens can exploit host IncRNA to infect their cells. HIV-1 viruses upregulate host cell NEAT1 to increase viral replication (Zhang et al. 2013), (Atianand & Fitzgerald 2014). A second host RNA, noncoding repressor of Nuclear Factor of T-Cells [NFAT] (NRON), is upregulated by HIV to control viral activity at specific stages during its life cycle (Imam et al 2015). The range of host IncRNAs exploited by HIV are reviewed by Lazar et al. (2016)

Pathogens express their own IncRNA during infectious attack. Kaposi’s sarcoma-associated herpes virus (KSHV) expresses polyadenylated nuclear (PAN) RNA, which enhances viral activity and inhibits host immune response during infection. (Rossetto & Pari 2011). Human cytomegalovirus (HCMV) uses the IncRNA β-27 to prevent apoptosis in HCMV-infected cells, keeping them alive and protecting the virus, to permit persistent infection (Zhang & Jeang 2013), (Tycowski et al. 2015). Through their regulation of gene expression IncRNAs control both pathogenic infection and the host immune response, however notably the pathogens themselves can also use IncRNAs to evade the immune system.
Cancer

As HOTAIR regulates hundreds of genes, including tumour suppressors, loss of its control results in cancer development and metastasis (Lee et al. 2006) (Zhao et al. 2010), (Esteller 2011), (Wapinski & Chang 2011). Understanding HOTAIR’s role in cancer may improve diagnosis and provide therapeutic targets (Zhang et al. 2014).

HOTAIR expression is significantly increased in breast cancer epithelial cells (Gupta et al. 2010). Experimental overexpression of HOTAIR guides PRC2 to repress 854 genes including tumour suppressors such as PCDH and JAM2, inducing breast cancer development (Gupta et al. 2010), (Novak et al. 2008), (Naik et al. 2008). HOTAIR repression of tumour suppressors removes the safeguards against breast cancer. This control over tumour suppression also applies to other cancer types. Overexpression of HOTAIR increases metastatic and invasive capability of colorectal cancers through inhibition of genes which suppress tumour growth such as cadherin, which normally maintains cellular adhesion, preventing metastasis (Jeannes et al. 2008), (Berx & van Roy 2009). HOTAIR overexpression is also associated with hepatocellular carcinoma, upregulating MMP-9 and VEGF, genes which promote metastasis (Geng et al. 2011). In gastric cancers, HOTAIR overexpression results in dysregulation of metastasis-associated genes (ICAM-1, MMP1, MMP3 & MMP9) (Xu et al. 2013), (Emadi-Andani et al. 2014), (Endo et al. 2013). As a result of HOTAIR’s wide regulatory reach, many tumour-related genes become dysregulated in HOTAIR overexpression, leading to cancer (Cai et al. 2014).

Conclusion

The fine spatiotemporal control that IncRNAs provide to the genome demonstrates the regulatory significance of IncRNA. IncRNA ‘master’ regulation is an elaborate, widespread mechanism for controlling when and where genes are expressed. IncRNAs in the X-inactivation centre allows chromosomes to communicate with each other, establish correct expression profiles (Tsix, Xite vs Jpx, Xist), and repress the X chromosome appropriately. Xist chromosome inactivation can be applied to extra chromosomes in Trisomy conditions such as Down’s syndrome, potentially implicating chromosomal therapies for this disorder. HOTAIR extends the influence of IncRNA to hundreds of genes across the genome (Lee, 2012). The dysfunction of HOTAIR regulation leads to cancer because IncRNAs control wide regulatory networks, which include tumour suppressors. IncRNA regulation is also used by host immune system responses as well as pathogens infection.

IncRNA gene regulation remains poorly understood. Chromosome inactivation and IncRNA regulation of cancer and immunity are interesting, however many precise details of IncRNA function remain unclear. Further investigation will undoubtedly reveal more uncharted non-coding RNA as only 25 years have passed since Xist’s characterisation. With improved sequencing and structural studies, the hidden complexity of genomic silencing may become understood.
Acknowledgments

The author would like to thank Ms. A. Worrall, Mr. D. O’Driscoll, Mr. A. Rammos, Mr. P. Fox and Prof. R. McLoughlin for their edits, advice, and guidance.

References


