Investigation of the cellular significance of long non-coding RNA NEAT1 and paraspeckles

Li, Ruohan

Harry Perkins Institute of Medical Research
School of Medicine and Pharmacology
School of Anatomy, Physiology and Human Biology

This thesis is submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy

The University of Western Australia

2016
Abstract

The majority of the human genome is transcribed into RNA, a large component of which consists of long non-coding RNAs (lncRNAs). For those lncRNAs that have been well-studied, a wide variety of mechanisms in gene regulation have been revealed. NEAT1 (nuclear paraspeckle assembly transcript 1) is a lncRNA essential for the formation and maintenance of subnuclear structures termed paraspeckles. Paraspeckles are RNA-protein complexes that function by sequestering specific nuclear proteins and mRNAs to prevent them trafficking to their functional destinations, resulting in downstream effects on gene expression. Ablation of NEAT1 and paraspeckles is important for various aspects of female fertility in mice, as well as the ability of cultured cells to cope with various types of cellular stresses. However, not much is known of the importance of NEAT1 and paraspeckles at the cellular level. Here, the main focus is to address this by systematically examining the regulation of NEAT1 expression, as well as the effect of NEAT1 on cell function.

In examining the upstream pathways that regulate NEAT1 transcription I found that as cells transition from pluripotency to a differentiated state, epigenetic silencing of NEAT1 is gradually lost, replaced with an epigenetic signature indicative of a highly active state of transcription. This explains why NEAT1 is expressed in differentiated, but not pluripotent cells. I also found that transcription factors associated with the human NEAT1 promoter are involved in a wide range of cellular processes, thus presenting a possible mechanism NEAT1 expression is generally both ubiquitous and dynamic. Under different stress conditions including serum starvation and the acidosis of culture media, NEAT1 levels and paraspeckle abundance was robustly induced in a subset of cell lines that were examined. In U2OS (human osteosarcoma cells), the upregulation of NEAT1 appeared to be primarily caused by direct transcriptional activation, mediated, at least in part by the action of p53 on the NEAT1 promoter. These findings link the transcriptional regulation of NEAT1 to the well-characterized p53 stress response pathway, and helps explain why NEAT1 expression is highly dynamic under conditions that are potentially stressful to cells.
The second part of this thesis investigates the effect of NEAT1 expression on cell biology. CRISPR-Cas9 technology and homologous recombination were used to disrupt the NEAT1 promoter, thereby generating the first human cell lines that are permanently devoid of functional NEAT1 and paraspeckles. Transcriptomic analysis of these NEAT1/-/- U2OS cell lines showed widespread changes in gene expression, many of these changes predictive of modulation of cell mobility, proliferation and cell death pathways. These predictions generally fit the observations for cells grown under normal conditions, where NEAT1/-/- cell lines showed a small reduction in their migration ability, but exhibited a significantly altered response to stress.

Finally, the possibility that NEAT1 may also function with paraspeckle-independent mechanisms was also investigated. The NEAT1 gene produces two isoforms of RNA: the 3,700 nucleotide long NEAT1_1 and the 23,000 nucleotide long NEAT1_2, which completely overlap at their 5’ ends. Both RNAs are found within paraspeckles, but only NEAT1_2 is essential and capable of building paraspeckles, whereas NEAT1_1 is dispensable. I hypothesized that one function of NEAT1_2/paraspeckles might be to sequester this shorter NEAT1_1 isoform that may have paraspeckle-independent roles. Utilizing the CRISPR-Cas9 approach and homologous recombination, isoform-specific NEAT1_2/-/- U2OS cell lines were made. These cells are devoid of paraspeckles, as they cannot generate NEAT1_2, but have intact NEAT1_1 RNA molecules that are distributed around the nucleus in foci termed ‘microspeckles’. Transcriptomic analysis comparing the wild-type, NEAT1/-/- and NEAT1_2/-/- cells showed that gene expression differed markedly, suggesting paraspeckle-independent roles for NEAT1_1. Interestingly, the expression of hundreds of genes were found to fluctuate in a way that fits both the existing model of molecular sequestration by paraspeckles, as well as the hypothesis that NEAT1_2/paraspeckles could have a role as a negative regulator of NEAT1_1. These data will open up new areas of investigation into the activity of NEAT1_1 and the functional inter-relationship between the two different isoforms of this abundant IncRNA.

Taken together, these studies not only provided valuable insight into the stress-mediate
induction and function of NEAT1 and paraspeckles in a cellular context, but also developed a new understanding of how NEAT1 isoforms and paraspeckles are related to each other in regulating gene expression.
# Table of Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Declaration</td>
<td>8</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>9</td>
</tr>
<tr>
<td>Contributors</td>
<td>11</td>
</tr>
<tr>
<td>List of Figures and Tables</td>
<td>12</td>
</tr>
<tr>
<td>Supplementary Information</td>
<td>15</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>16</td>
</tr>
</tbody>
</table>

## Chapter 1: Introduction

1.1 Chapter overview                          | 20   |
1.2 A brief overview of non-coding RNAs in the human genome | 21   |
1.3 Long non-coding RNA                        | 22   |
    1.3.1 Definition and characteristics         | 22   |
    1.3.2 LncRNAs function in all levels of the regulation of gene expression | 27   |
      1.3.2.1 HOTAIR and epigenetic silencing      | 27   |
      1.3.2.2 LincRNA-p21 in transcriptional and translational control | 28   |
      1.3.2.3 Sno-LncRNA and post-transcriptional regulation | 30   |
      1.3.2.4 BACE1-AS in mRNA stabilization      | 31   |
      1.3.2.5 Airn in genomic imprinting          | 32   |
      1.3.2.6 XIST in X-chromosome inactivation   | 33   |
    1.3.3 LncRNA and subnuclear structures       | 34   |
      1.3.3.1 A LncRNA that is enriched within a subnuclear structure, but not | 34   |
          as a structural component: MALAT1 and nuclear speckles  |
      1.3.3.2 LncRNA that forms structural scaffolds for subnuclear bodies: | 37   |
          hsr-ω and Omega Speckles                        |
1.4 LncRNA NEAT1 and paraspeckles            | 41   |
    1.4.1 Discovery of paraspeckles              | 41   |
    1.4.2 Discovery of NEAT1 in paraspeckles     | 41   |
    1.4.3 Proteins in paraspeckles               | 45   |
    1.4.4 Assembly and structure of paraspeckles | 47   |
    1.4.5 Functions of paraspeckles              | 51   |
      1.4.5.1 Nuclear retention of specific mRNAs | 51   |
      1.4.5.2 Subnuclear sequestration of paraspeckle proteins | 54   |
      1.4.5.3 Physiological functions of paraspeckles | 54   |
1.4.6 Expression of NEAT1                    | 55   |
    1.4.6.1 Association with differentiation in vitro | 56   |
    1.4.6.2 Relevance to physiological conditions | 57   |
1.4.7 Unsolved questions relating to NEAT1 and paraspeckles | 59   |
1.5 Hypothesis and aims                      | 61   |

## Chapter 2: Materials and Methods

64
Chapter 3: Manuscript

The long non-coding RNA NEAT1 is induced by acidosis and influences the expression of many genes with diverse cellular functions through both paraspeckle-dependent and independent actions.

3.1 Abstract
3.2 Introduction
3.3 Results
  3.3.1 Acidosis induces NEAT1 expression and formation of paraspeckles
  3.3.2 p53 is involved in the induction of NEAT1 in acidosis
  3.3.3 CRISPR-mediared generation of NEAT1 knockout U2OS cells
  3.3.4 NEAT1 alters expression of genes controlling cell movement and cell death
  3.3.5 NEAT1_1 and NEAT1_2/paraspeckles co-regulate specific gene sets
3.4 Discussion
  3.4.1 Paraspeckles and omega speckles – convergent evolution?
3.4.2 Transcriptional upregulation of NEAT1 by p53 and other factors

3.4.3 Physiological role for NEAT1 in cancer

3.4.4 Physiological role for NEAT1 in normal animals

3.4.5 NEAT1 mechanism of action

3.5 Experimental procedures

3.6 References

Chapter 4: Regulation of NEAT1 expression

4.1 Introduction

4.2 Results

4.2.1 Data mining indicated more repressive epigenetic markers of NEAT1 in stem cells

4.2.2 Transcriptional regulation of NEAT1 could be very complex

4.2.3 Transcriptional activation of NEAT1 could be very strong

4.2.4 NEAT1 can be induced by acidosis and serum starvation depending on the cell line

4.3 Discussion

Chapter 5: Using genome engineering to create permanent isoform-specific NEAT1 knockout cell lines

5.1 Introduction

5.2 Results

5.2.1 Knocking out human NEAT1 using CRISPR-Cas9 system

Part I: Design and preparation of materials

Part II: Transfection and sorting

Part III: Screening for homozygous NEAT1 knockout colonies

Part IV: Verification of NEAT1 knockout and summary

5.2.2 Preliminary RNA-seq and differential expression analysis of NEAT1 knockout colonies

5.2.3 Generation of U2OS cells that were NEAT1_2 knockout but still retained NEAT1_1

5.3 Discussion

Chapter 6: General discussion

6.1 NEAT1 and paraspeckles – current research status

6.2 Findings, problems, and future directions

6.2.1 Results relating to events upstream of NEAT1 expression

6.2.2 Results relating to events downstream of NEAT1 expression

6.2.3 Further verification of the differential gene expression analysis

6.2.4 NEAT1_2 might stabilize interactions between paraspeckle components

6.2.5 NEAT1_1 and NEAT1_2 might have opposite roles in gene regulation

6.2.6 How NEAT1_1 regulates gene expression is still unclear
| 6.2.7 | The function of paraspeckles might be more prominent under certain conditions | 204 |
| 6.2.8 | The function of NEAT1 and paraspeckle might expand beyond transcriptional control | 205 |
| 6.3 Final remarks | | 207 |

**Appendices**

- Appendix 1: Published work and contributions 209
- Appendix 2: Sequences of cloned human *NEAT1* promoter and enhancers in Chapter 4 253
- Appendix 3: Sequences of homology arms for homologous recombination repairing of CRISPR-mediated NEAT1 knockouts 262
- Appendix 4: Supplementary figures and tables for chapter 4 and 5 265
- Appendix 5: References 281
My contribution in these publications, as well as copies of these publications, are listed in the Appendix 1 at the end of this thesis. They do not contribute to the conclusions obtained from this Ph.D project.
Acknowledgements

So here it comes to the big thank you section.

Looking back over the past few years, as I have grown four years older, I have also started to appreciate what it takes to become a scientist. I think it is about focus, commitment, sheer will to challenge the boundary of human knowledge, and also to learn to pursue my own research interest with practical considerations. There were many people, at different stages, who came across my journey of the Ph.D and left some mark on it. Dr. Archa Fox, my principle supervisor, has offered me an incredible amount of freedom to pursue what I wanted to do. I know this can be very annoying for supervisors sometimes; therefore I see this as a real luxury that many of my friends did not enjoy. My father has also influenced me a lot. Although he is not a biologist, his clinical experience has taught me to balance the applied science and basic research equally, which influenced my project a great deal. My other two supervisors, Professor Alan Harvey and Associate Professor Stuart Hodgetts, and of course other scientists I met during conferences, have provided me with important feedback for my Ph.D along the way. I have to stress again that this feedback is so important as it has made my scientific training considerably more rigorous, which I believe is very, very important for us young people in this field. Lastly, my grandfather has consistently provided me with the most critical guidance for my life, more than anyone else. He often says ‘always know which problem is the main problem, and which ones are secondary’. This teaching is so correct in research.

Furthermore, a few friends of mine have to be highlighted. Dr. Ramesh Ram promoted himself from my food buddy to my master of bioinformatics towards the end of my Ph.D. For me, his technical assistance was like offering berries to a thirsty man, because the handling of complex bioinformatics data is incredibly hard to learn for a person like me who has never received computer science training. Without his guidance, it would have taken so much longer for my Ph.D to finish. Dr. Luciano Martelotto, despite his short stay in the lab, also provided me with important technical advice over many experiments that I found difficult to deal with. To the two research assistants, Pei Wen Lee and Agata
Sadowska, I am very grateful for the help with dealing with so many cell lines. For all these friends of mine, you have made my life during this Ph.D much easier.

Lastly, financial aid from the University of Western Australia, the Harry Perkins Institute of Medical Research, Dr. Archa Fox, and NHMRC research funding has also been critical. I have learnt that research is expensive, and research in Australia is even more expensive.
Contributors

People listed below have made some contribution towards the completion of the project and this thesis.

Ruohan Li:
All wet-lab experiments, all data analysis, and writing up of all chapters of the thesis.

Dr. Ramash Ram:
Provided guidance for the bioinformatic analyses presented in Chapter 3 and 5, but did not carry out the analyses.

Dr. Archa Fox (principle supervisor):
Provided assistance with the experimental design and analysis for all results presented in this thesis. Edited all chapters, but most extensively for Chapter 3.

Assoc. Prof. Stuart Hodgetts (co-supervisor) and
Winthrop Prof. Alan Harvey (co-supervisor):
Provided and assistance with the experimental design and analysis for results presented in Chapter 3 and 4.

Ciara Duffy:
Provided proof-reading service for all chapters except Chapter 3.
List of Figures and Tables

Contents

Chapter 1:

Figure 1.1: lncRNAs are involved in a wide range of cellular functions related with regulating gene expression, through binding with different macromolecules.

Figure 1.2: LincRNA-p21 functionality.

Figure 1.3: MALAT1 localization.

Figure 1.4: hsr-omega and omega speckles.

Figure 1.5: Paraspeckles and human NEAT1 gene structure.

Figure 1.6: The assembly and structure of paraspeckles.

Figure 1.7: Paraspeckles play regulatory roles in gene expression through protein sequestration and mRNA nuclear retention.

Chapter 2:

Table 2.1: Suppliers of services.

Table 2.2: Suppliers of reagents and kits.

Table 2.3: Suppliers of antibodies and probes.

Table 2.4: Suppliers of instruments.

Table 2.5: PCR and sequencing primers.

Table 2.6: qRT-PCR primers.

Table 2.7: CRISPR gRNA oligonucleotides.

Chapter 3:

Figure 3.1: Acidosis induces NEAT1 expression and formation of paraspeckles in U2OS cells.

Figure 3.2: p53 is involved in the transcriptional upregulation of NEAT1 following the acidosis of U2OS cells.
**Figure 3.3:** CRISPR-mediated genome engineering is successful in generating U2OS cell lines with ~100 fold permanent depletion of NEAT1 RNA, or isoform-specific depletion in NEAT1_2 RNA.

**Figure 3.4:** RNA-seq and differential gene expression analysis of NEAT1-depleted cell lines predicts alterations in key phenotypic features.

**Figure 3.5:** Total NEAT1⁻/⁻ cells have impaired response to stress, and subtly reduced migration ability, but similar cellular proliferation rates to wild-type cells.

**Figure 3.6:** Subsets of genes are differentially regulated by the two NEAT1 isoforms indicating that the majority of genes are regulated by both isoforms in an equal but opposite manner, or are only sensitive to NEAT1_2/paraspeckes.

**Chapter 4:**

**Figure 4.1:** Epigenetic marks upstream of the human NEAT1 gene in different cell lines.

**Figure 4.2:** UCSC genome browser tracks of human NEAT1 upstream region.

**Figure 4.3:** Cloning and analysis of human NEAT1 upstream regulatory region.

**Figure 4.4:** NEAT1 was upregulated significantly in differentiated cell lines as shown by ENCODE RNA-seq tracks of human and mouse NEAT1 expression.

**Figure 4.5:** RT-qPCR of endogenous total NEAT1 and NEAT1_2 expression under different treatments.

**Chapter 5:**

**Figure 5.1:** Design of CRISPR-Cas9 mediated NEAT1 knockout cell line.

**Figure 5.2:** Workflow of the CRISPR NEAT1 knockout experiments.

**Figure 5.3:** PCR screening of NEAT1⁻/⁻ clonal cell lines.
**Figure 5.4:** Direct observations of NEAT1 levels and morphology of the NEAT1 knockout colonies.

**Figure 5.5:** RNA-seq data normalized by RUVg indicated multiple cellular processes that could be affected by the NEAT1 knockout.

**Figure 5.6:** The design, screening and verification of the CRISPR strategy to create NEAT1_2 knockout cell lines.

**Chapter 6:**

**Figure 6.1:** An updated model of molecular sequestration mediated by NEAT1_2/paraspeckles proposed at the end of this study.
Supplementary Information

Chapter 3:
Figure S3.1: Strategy of using PCR to identify homozygous NEAT1⁺⁺ colonies.

Figure S3.2: RLE (relative log expression) plots and PCA (principle component analysis) plots of three types DE analyses.

Table S3.1: NEAT1 transcription factors.

Table S3.2: DE genes in NEAT1_2 KO vs WT.

Table S3.3: DE genes in Total NEAT1 KO vs WT.

Table S3.4: DE genes in Total NEAT1 KO vs NEAT1_2 KO.

Table S3.5: Prediction on affected cellular functions.

Table S3.6: Genes only regulated by NEAT1_1.

Table S3.7: Genes regulated by NEAT1 isoforms synergistically.

Table S3.8: Genes only regulated by NEAT1_2.

Table S3.9: Genes regulated by both NEAT1_1 and NEAT1_2 in opposite ways.

Table S3.10: Oligonucleotides and antibodies.

Chapter 4:

Table S4.1: NEAT1 transcription factors.

Chapter 5:

Figure S5.1: Highlighted cellular functions and molecular networks in DE analysis for Total NEAT1 KO vs WT.

Table S5.1: DE genes in Total NEAT1 KO vs WT.

Supplementary information is not printed due to large size, except Table S4.1 and Figure S5.1, which are attached in Appendix 4. All other supplementary materials can be found in the USB flash drive attached.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'-UTR</td>
<td>3'-untranslated region</td>
</tr>
<tr>
<td>ADARB2</td>
<td>Adenosine deaminase, RNA-Specific, B2</td>
</tr>
<tr>
<td>ASOs</td>
<td>Antisense oligonucleotides</td>
</tr>
<tr>
<td><em>C. elegans</em></td>
<td><em>Caenorhabditis elegans</em></td>
</tr>
<tr>
<td>Cas9</td>
<td>CRISPR associated protein 9</td>
</tr>
<tr>
<td>CGI</td>
<td>CpG island</td>
</tr>
<tr>
<td>CHART-MS</td>
<td>Capture hybridization analysis of RNA targets coupled with mass spectrometry</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered regularly interspaced short palindromic repeats</td>
</tr>
<tr>
<td>crRNA</td>
<td>CRISPR RNA</td>
</tr>
<tr>
<td>DAVID</td>
<td>Database for annotation, visualization and integrated discovery</td>
</tr>
<tr>
<td>DBHS</td>
<td>Drosophila behavior/human splicing</td>
</tr>
<tr>
<td>DE</td>
<td>Differential expression</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's modified eagle medium</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DSB</td>
<td>Double stranded DNA breaks</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>ENCODE</td>
<td>The Encyclopedia of DNA Elements</td>
</tr>
<tr>
<td>ESCs</td>
<td>Embryonic stem cell lines</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal bovine serum</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>gRNA</td>
<td>Guide RNA</td>
</tr>
<tr>
<td>hDF</td>
<td>Human dermal fibroblast</td>
</tr>
<tr>
<td>HDR</td>
<td>Homology-directed repair</td>
</tr>
<tr>
<td>HeLa</td>
<td>Henrietta Lacks</td>
</tr>
<tr>
<td>hMSC</td>
<td>Human mesenchymal stem cells</td>
</tr>
<tr>
<td>HNRNPK /</td>
<td>Heterogeneous nuclear ribonucleoprotein K</td>
</tr>
<tr>
<td>hnRNP-K</td>
<td></td>
</tr>
<tr>
<td>hsr-ω</td>
<td>Heat Shock RNA omega</td>
</tr>
<tr>
<td>IL8</td>
<td>Interleukin 8</td>
</tr>
<tr>
<td>indels</td>
<td>Insertions and deletions</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity pathway analysis</td>
</tr>
<tr>
<td>iPSCs</td>
<td>Induced pluripotent stem cells</td>
</tr>
<tr>
<td>KD</td>
<td>Knockdown</td>
</tr>
<tr>
<td>KEGG</td>
<td>Kyoto encyclopedia of genes and genomes</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>IncRNA</td>
<td>Long non-coding RNA</td>
</tr>
<tr>
<td>MALAT1</td>
<td>Metastasis associated lung adenocarcinoma transcript 1</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple cloning sites</td>
</tr>
<tr>
<td>MEMα</td>
<td>Minimum essential medium α</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNAs</td>
</tr>
<tr>
<td>NEAT1</td>
<td>Nuclear paraspeckle assembly transcript 1</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Non-homologous end joining</td>
</tr>
<tr>
<td>NONO</td>
<td>Non-POU domain containing, octamer-binding</td>
</tr>
<tr>
<td>nSB</td>
<td>Nuclear stress body</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotides</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frames</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>PANTHER</td>
<td>Protein analysis through evolutionary relationships</td>
</tr>
<tr>
<td>PC</td>
<td>Principle component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>poly(A)</td>
<td>Polyadenylation</td>
</tr>
<tr>
<td>PSPC1</td>
<td>Paraspeckle component 1</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA-induced silencing complex</td>
</tr>
<tr>
<td>RLE</td>
<td>Relative log expression</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>RRM</td>
<td>RNA recognition motifs</td>
</tr>
<tr>
<td>RUV</td>
<td>Removal of unwanted variation</td>
</tr>
<tr>
<td>SatIII</td>
<td>Satellite III</td>
</tr>
<tr>
<td>SC35</td>
<td>Splicing component, 35kDa</td>
</tr>
<tr>
<td>SFPQ</td>
<td>Splicing factor proline/glutamine-rich</td>
</tr>
<tr>
<td>shRNA</td>
<td>Small hairpin RNA</td>
</tr>
<tr>
<td>SSB</td>
<td>Single stranded DNA breaks</td>
</tr>
<tr>
<td>SV40</td>
<td>Simian virus 40</td>
</tr>
<tr>
<td>TALEN</td>
<td>Transcription activator-like effector nucleases</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TFBS</td>
<td>Transcription factor binding site</td>
</tr>
<tr>
<td>TncRNA</td>
<td>Trophoblast non-coding RNA transcript</td>
</tr>
<tr>
<td>tracrRNA</td>
<td>Trans-activating crRNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
<tr>
<td>TTS</td>
<td>Transcriptional termination signal</td>
</tr>
<tr>
<td>XCI</td>
<td>X-chromosome inactivation</td>
</tr>
<tr>
<td>XIST</td>
<td>Xi-specific transcript</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow fluorescent protein</td>
</tr>
<tr>
<td>ZFN</td>
<td>Zinc-finger nucleases</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction
1.1 Chapter overview

Research into noncoding RNAs (ncRNAs) has experienced a massive expansion since the Human Genome Project revealed conclusively that the limited number of genes alone could not explain our complexity. Among various types of ncRNAs, long non-coding RNAs (lncRNAs) are perhaps the most heterogeneous and enigmatic group. These lncRNAs are hard to define and characterize as a whole, as they vary greatly in their evolutionary history, mechanisms and functions. In this introductory chapter, I will first focus on some specific lncRNAs, using them as examples to illustrate various characteristics of this group. This forms the background to the biology of the NEAT1 (nuclear paraspeckle assembly transcript 1) lncRNA and paraspeckles, which is the main focus of this project.

NEAT1 is one of the most abundant lncRNAs found in the nucleus of a variety of mammalian cells. Differentiated, epithelial and secretory cells often upregulate and highly express NEAT1 and paraspeckles (Mercer et al., 2010; Nakagawa et al., 2011; Sunwoo et al., 2009). The main role of NEAT1 forms subnuclear RNA-protein bodies termed ‘paraspeckles’ (Clemson et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009). Paraspeckles function, at least in part, as a sequestration center for nuclear protein and mRNAs. However, the understanding of the physiological role of paraspeckles, and many other attributes of NEAT1, is limited. This project aims to understand the physiological significance of NEAT1 and paraspeckles at the cellular level, by investigating the induction of NEAT1 and the consequences of its expression. It is hoped that the result of this approach will not only improve our understanding of the physiological role of these subnuclear structures, but also reveal more information about the functional relationship between NEAT1 and paraspeckles.
1.2 A brief overview of non-coding RNAs in the human genome

One of the greatest scientific achievements of the past 100 years is the completion of the Human Genome Project. Now more than a decade after its completion, its outcome has greatly accelerated scientific advancement in all aspects of human biology. The Human Genome Project demonstrated an important and interesting finding: that of the 3.2 billion base pairs of the human genome, there are only approximately 20,000 protein-coding genes. The genome of \textit{C. elegans} (\textit{Caenorhabditis elegans}) carries a similar amount of protein-coding genes as the human genome, even though the latter genome is 32-fold larger (Hillier et al., 2005; Stein, 2004). The vast difference between the complexities of these two organisms is therefore not explained by the similar number of protein-coding genes. Thus knowledge of the complete sequence of human DNA is only the first step to understanding the complexity of human genome biology.

Traditionally, genomic DNA segments were classified into transcribed genes (mostly for proteins), regulatory regions of genes (promoters/enhancers), and junk DNA (transposons/pseudogenes), this latter class with no obvious function. Given the development of faster and cheaper sequencing technologies, along with advanced computational hardware and software, we are now capable of deciphering this extremely complex regulatory system. The ENCODE Project, known as the Encyclopedia of DNA Elements, launched by the US National Human Genome Research Institute in 2003, has greatly enhanced our understanding of the information that our genome is carrying. This multinational project aimed to identify an atlas of functional elements in the human genome. By applying a large number of new technologies to different cell types and factors, it claimed that 80% of the human genome is associated with at least one biochemical RNA and/or chromatin associated event in at least one cell type (ENCODE 2012). It is important to note that there are many criticisms centered on the accuracy of this estimation, especially from the viewpoint of comparative genomics. For example, a recent publication suggests that only 8.2% of the human genome is functional (Rands et al., 2014). This study proposed that the ENCODE approach was flawed as it defined functionality based on the choice of experiments and details such as $P$-value cut-offs.
(Doolittle, 2013; Graur et al., 2013; Rands et al., 2014). Nevertheless, modern genomic research has re-classified a large amount of genomic DNA into either regulatory DNA or genes for non-coding RNAs (Pennisi, 2007).

Many types of ncRNA have been identified, including regulatory and non-regulatory subtypes. In the latter group, the most well-characterized examples are transfer RNA (tRNA) and ribosomal RNA (rRNA), both of which are crucial for protein synthesis, small nuclear RNA (snRNA), essential for pre-mRNA splicing, and small nucleolar RNAs (snoRNAs) that facilitate the chemical modification of other ncRNAs such as tRNAs and rRNAs. For regulatory RNAs, examples include small interfering RNA (siRNA), micro RNA (miRNA), and piwi-interacting RNA (piRNA). Finally, catalytic RNAs such as ribozymes display similar functions to enzymes. Added to the regulatory ncRNA group are the lncRNAs, which are commonly and vaguely regarded as ncRNAs longer than 200 nucleotides in length.

### 1.3 Long non-coding RNAs

#### 1.3.1 Definition and characteristics

LncRNAs are arbitrarily defined as non-protein-coding transcripts that are longer than 200 nucleotides (nt), which is primarily due to the earlier convenient technical cut-off in RNA purification protocols that excludes small RNAs (Kapranov et al., 2007). LncRNAs also lack open reading frames of greater than 50-100 residues (Wilusz et al., 2009). As these two criteria are not stringent, a large number of transcribed ncRNAs can be fitted into this category. As a result, lncRNAs range in size, varying from a few hundred nucleotides to more than 100,000 nt. The human KCNQ1OT1 (KCNQ1 overlapping transcript 1) lncRNA is encoded by a gene spanning 471kb on chromosome 7, with transcripts estimated to be either 91,000 nt or 121,000 nt in length (Pandey et al., 2008; Redrup et al., 2009). Yet on the other end of the spectrum, the human lncRNA LNC-SH2B3-1:3 is only 209 nt, manually annotated by the HAVANA (Human and Vertebrate Analysis and Annotation) Project, which is perhaps the shortest human
More than 14,000 lncRNA genes have been located in the human genome and this may be an underestimated figure (Derrien et al., 2012; Harrow et al., 2012; Kapusta and Feschotte, 2014; Necsulea et al., 2014; Ulitsky and Bartel, 2013). Still, the number of potential lncRNAs genes across different species appears to be less than protein coding genes (reviewed in Kapusta and Feschotte, 2014), but some authors acknowledge the number of lncRNA increases rapidly as organisms becomes more advanced and complex. This trend is indeed observed at great evolutionary distances, such as between unicellular organisms versus multicellular organisms, or vertebrates versus invertebrates (Liu et al., 2013a). However, due to fewer number of genes and other non-coding DNA materials, the genome is also significantly more compact in lower organisms than in higher ones across large evolutionary distances (Taft and Mattick, 2003; Taft et al., 2007). Therefore, it would be interesting for future studies to examine whether organisms with similar sized genome but with great developmental complexity would show significantly different numbers of lncRNA genes.

The position of genes encoding lncRNAs can vary in many different ways with respect to protein-coding genes: either overlapping, antisense, in between (intergenic, also known as long intergenic non-coding RNA/lincRNA), or within introns of coding genes. At the genomic level, many lncRNA genes do not appear significantly different from protein coding genes. In the majority of cases, lncRNAs are transcribed by RNA polymerase II, and the transcripts produced are capped, spliced, and polyadenylated (reviewed in Harrow et al., 2012). LncRNA genes also show a bias towards a dual-exonic structure (reviewed in Derrien et al., 2012).

Many lncRNAs contain small open reading frames (ORFs) even though they are classified as ‘non-coding’. Some of these ORFs are translated into functional short polypeptides with important physiological roles. For example, the lncRNA myoregulin, involved in muscle development encodes both a functional lncRNA and produces short peptides, and the lncRNA pri (polished rice) in Drosophila, also produces four so-called lncRNA.
'pri' peptides with functional relevance in embryogenesis (Anderson et al., 2015; Kondo et al., 2010; Slavoff et al., 2013; Tupy et al., 2005). However, only a small proportion of the lncRNAs that carry ORFs are eventually translated, as most lncRNA ORFs have non-AUG start codons, which makes them less efficient for translation (Niazi and Valadkhan, 2012). In addition, the location of ORFs in many lncRNA transcripts is such that nonsense-mediated decay would be triggered if the short ORF were used (Niazi and Valadkhan, 2012).

The sequence conservation of lncRNA genes is highly variable. Firstly, some functionally characterized lncRNAs show strong evolutionary conserved regions, particularly in mammals. For example, Gomafu/Miat (Myocardial infarction associated transcript), is a lncRNA regulated by POU5F1/OCT4 (POU class 5 homeobox 1) in mouse embryonic stem cells and the nervous system, that displays high conservation across mammals, but not with other non-mammalians (Rapicavoli et al., 2010; Sheik Mohamed et al., 2010; Tsuiji et al., 2011; Ulitsky et al., 2011). On the other hand, there are also cases of extremely poor sequence conservation between lncRNA orthologues. In these cases the lncRNA may function solely as a transcriptional bookmark, irrespective of lncRNA sequence. For example, the transcription of the lncRNA that is immediately downstream of the HIS enhancer of the human GH (growth hormone) gene was thought to improve the accessibility of the HIS enhancer, thus increasing the transcriptional activity of GH (reviewed in Vance and Ponting, 2014; Yoon et al., 2012). Interestingly, when an unrelated bacteriophage RNA was used to replace the lncRNA, the effect was also reproduced, indicating a sequence-unspecific manner of lncRNA action (Yoon et al., 2012). Furthermore, similar to protein coding genes, exons and splice sites of lncRNA tend to have stronger sequence conservation than introns, which implies functional lncRNA are mostly processed transcripts (Chodroff et al., 2010; Haerty and Ponting, 2013; Kapusta and Feschotte, 2014; Ponjavic et al., 2007). In generally, evidence of purifying selection is still more obvious in protein-coding exons, untranslated regions, and genes encoding small noncoding RNAs such as tRNA or miRNA, than that of lncRNAs. This may be due to protein-coding genes being under the pressure of preserving polypeptide information, whereas lncRNAs only need to preserve key secondary structures (He et al.,
As molecular actions of lncRNA can vary greatly, it is possible that each lncRNA may have different selective pressures for different functional aspects. For some lncRNAs the critical part may be the small regulatory sequences that they contain (Duret et al., 2006; Marques and Ponting, 2009), for others, their genomic location, length and orientation could be more important (Cabili et al., 2011; Ponting et al., 2009; Ulitsky et al., 2011). Interestingly, promoters of lncRNA have a comparable level of conservation to mRNA promoters, suggesting the expression and regulation of lncRNAs may be more important than their primary sequence (Guttman et al., 2009). Overall, as the field of lncRNA evolution still in its infancy, it is difficult to generalize the evolutionary history and origins of lncRNA.

LncRNAs are clearly very different from other regulatory ncRNAs such as miRNA and piRNA, which both have consistent mechanisms within each family (reviewed in Ameres and Zamore, 2013; Meister, 2013). Large-scale sequencing and bioinformatic analysis have had an important role to play in identifying lncRNAs, but the bigger picture of lncRNA importance will only be clearer after sufficient molecular and functional knowledge of each lncRNA is obtained. To date, there are only a small number of lncRNAs that have been studied extensively at the molecular level, especially compared to other ncRNAs and proteins. Nevertheless, a common theme has emerged that lncRNAs can regulate gene expression at many different levels. Figure 1.1 summarizes how lncRNAs achieve this by interacting with different binding partners.
Figure 1.1: IncRNAs are involved in a wide range of cellular functions related with regulating gene expression, through binding with different macromolecules.
1.3.2 LncRNAs regulate all levels of gene expression

The regulation of gene expression takes place at many levels. At the level of transcription, regulatory mechanisms influence epigenetic factors, chromatin remodeling and transcription factor binding to the DNA/chromatin. Co-transcriptional splicing, nuclear export and the binding of accessory factors influence the stability of the mRNA, and the translational efficiency of different mRNAs are also influenced by many mechanisms. This complex network of regulatory systems combines to deliver precise temporal and spatial control of gene expression. Proteins are undoubtedly playing a dominant role in this regulatory system, however lncRNAs have also been found to influence many aspects of gene regulation. Some examples of different lncRNAs that have been implicated in different steps in the control of gene regulation are outlined below.

1.3.2.1 HOTAIR and epigenetic silencing

HOTAIR (HOX transcript antisense RNA) is a 2.2kb transcript antisense to the HOXC (homeobox c cluster) locus on human chromosome 12 (Rinn et al., 2007). This lncRNA can act as a scaffold to bring together two different protein complexes, each tethered to a different part of the RNA. The 5’ end of HOTAIR associates with the PRC2 (polycomb repressive complex 2) complex, which trimethylates H3K27 (histone 3, lysine 27) resulting in a repressive histone modification generally associated with transcriptional silencing. The 3’ region of HOTAIR is bound by the LSD1 (lysine-specific demethylase 1) / CoREST (REST corepressor 1) / REST (RE1-silencing transcription factor) complex, responsible for demethylating H3K4me2 (histone 3 lysine 4 di-methylation) and H3K4me1 (histone 3 lysine 4 methylation) (Tsai et al., 2010), thereby removing active histone marks. Thus, both ends of the HOTAIR RNA orchestrate a repressive effect on the transcription of the locus that is targeted by this RNA protein complex. Knockdown of HOTAIR decreases H3K27me3 and PRC2 occupancy across the HOXD (Homeobox D Cluster) locus (Rinn et al., 2007). In mice, this caused de-repression of hundreds of genes, and resulted in homeotic transformation of the spine (L6 lumbar vertebrae transformed into S1 sacral vertebrae) and malformation of the metacarpal-carpal bones (Li et al., 2013).
The other side of HOTAIR as a developmental regulator is increased expression in cancer, where it has been implicated in driving cancer progression, particularly in breast cancer metastasis where it is associated with poor prognosis (Gupta et al., 2010). Increased activity of HOTAIR reduces the expression of several metastasis suppressor genes, eventually leading to increased expression of genes that drive metastasis (Gupta et al., 2010). HOTAIR is also significantly elevated in lung cancer where it regulates signaling pathways that are pivotal to differentiation, proliferation, and invasion (Liu et al., 2013b; Liu et al., 2013c; Nakagawa et al., 2013; Ono et al., 2014; Zhao et al., 2014; Zhuang et al., 2013). While these studies have suggested that HOTAIR could be a promising therapeutic target with diagnostic potential, the crucial mechanistic information lacking is how the HOTAIR-protein complex is recruited to specific target genes and how this process is controlled.

1.3.2.2 LincRNA-p21 in transcriptional and translational control

LincRNA-p21 is a 3,100nt transcript located ~15kb upstream of the cell cycle regulator gene p21 (Cdkn1a) in the intergenic region on mouse chromosome 17 and human chromosome 6 (HUARTe et al., 2010). LincRNA-p21 interacts with hnRNP-K (heterogeneous nuclear ribonucleoprotein K), a heterogeneous nuclear ribonucleoprotein, through its 5’ terminal region (Huarte et al., 2010). Characterization of the LincRNA-p21 knockout mouse revealed that the primary function of the LincRNA-p21:hnRNP-K complex was to act in cis to downregulate transcription of the neighboring protein-coding p21 gene (Dimitrova et al., 2014). Additional roles for the LincRNA-p21 complex are in repressing other p53 target genes (Huarte et al., 2010). As p53 is a transcriptional activator of LincRNA-p21, this suggested LincRNA-p21 participates in a regulatory loop for p53 targets (Huarte et al., 2010). Conversely, overexpression of LincRNA-p21 was found to promote apoptosis in a p53-dependent manner, suggesting LincRNA-p21 and p53 have synergistic roles (Huarte et al., 2010) (Figure 1.2). Interestingly, the LincRNA-p21 knockout mice are
Figure 1.2: LincRNA-p21 functionality. (A) LincRNA-p21 function in transcriptional regulation of target genes: induced by p53, acting with hnRNP-K to repress gene expression (Huart et al., 2010). (B) Function in mRNA destabilization: LincRNA-p21 is degraded in the presence of HuR, but binds to polysomes to repress translation in the absence of HuR (Yoon et al., 2012).
viable with no obvious defects, a phenomenon common to several other lncRNAs, suggesting fine-tuning of gene expression in response to environmental triggers. A specific role for lncRNAs in regulating gene expression under stress induced conditions is a theme that will be further explored in this thesis.

Revealing the diversity of lncRNA action, another study showed that LincRNA-p21 can function as a translational repressor. When found in the cytoplasm, LincRNA-p21 associates with specific mRNAs such as the mRNA encoding the Jun B Proto-Oncogene and β-catenin. LincRNA-p21 associates with the mRNAs enriched on polysomes, and is linked to suppression of mRNA translation (Yoon et al., 2012). On the other hand, LincRNA-p21 is targeted by HuR (Human antigen R), which recruits the let-7/AGO2 (Argonaute RISC Catalytic Component 2) microRNA/protein complex to degrade LincRNA-p21 (Yoon et al., 2012). From a physiological perspective, there is some evidence that increased expression of LincRNA-p21 can sensitize colorectal cancer cells to radiotherapy, which would fit with a pro-apoptotic role in a p53 induced network (Wang et al., 2014). Overall, however, whilst this lncRNA has been very well characterized at the molecular level, the physiological significance is yet to be fully elucidated.

1.3.2.3 Sno-lncRNA and post-transcriptional regulation
Sno-lncRNAs (small nucleolar lncRNA) were identified through the deep sequencing of non-polyadenylated transcripts from HeLa and human embryonic stem cells (Yang et al., 2011). These abundant polyA(-) lncRNAs are derived from introns that are spliced out of pre-mRNA, but are stabilized, unlike the vast majority of introns that are rapidly degraded after debranching (Yin et al., 2012). This stabilization is thought to be a result of snoRNA (small nucleolar RNA) sequences found at both ends of the sno-lncRNA transcripts. The snoRNA processing machinery protects the sno-lncRNAs from degradation, resulting in their accumulation. The processing machinery of these snoRNAs is the same machinery involved in snoRNP (small nucleolar ribonucleoprotein) biogenesis, which depends on the interaction between the C/D boxes of the snoRNAs with the snoRNP proteins 15.5K and Fibrillarin (Yin et al., 2012). The imprinted region
on human 15q11-q13, contains five sno-lncRNAs expressed at high levels. Interestingly, all five sno-lncRNAs localize to one or two closely positioned subnuclear sites, which have been identified near the locus of their genes.

RNA interference of all five of these highly expressed sno-lncRNAs had little effect on global gene expression, but did alter FOX2 (RNA binding protein fox-1 homolog 2) mediated splicing of neighbouring genes (Yin et al., 2012). The cell-type specific splicing regulator FOX2 regulates alternative polyadenylation, mRNA stability, localization and translation (Nakahata and Kawamoto, 2005; Shi et al., 2009; Underwood et al., 2005; Wang et al., 2008). The five sno-lncRNAs contain multiple binding sites for FOX2 (Yeo 2009), and co-localisation was confirmed by fluorescent microscopy (Yin et al., 2012). Thus the sno-lncRNAs might act as a molecular sink for FOX2, thereby controlling FOX2-regulated splicing. It is unclear how ubiquitous this post-transcriptional effect is in other organisms and cell types, although the concept of lncRNAs acting as a molecular sink, or ‘sponge’ for RNA binding proteins is becoming more common, and will be further expanded in the sections on paraspeckles below.

1.3.2.4 BACE1-AS in mRNA stabilization

Perfect base-pairing of an lncRNA with an mRNA has been implicated in mRNA stabilization, by antagonizing binding of the RISC (RNA-induced silencing complex) at a miRNA target site. BACE1-AS, a conserved ~2,000 nt antisense transcript of the BACE1 (β-secretase-1) gene, shows evidence of such function (Engström et al., 2006; Faghihi et al., 2008). BACE1 is a crucial enzyme in the pathophysiology of Alzheimer’s disease. The sequential cleavage of amyloid precursor protein (APP) mediated by BACE1 leads to the production of polypeptide Aβ 1-40 and Aβ 1-42, which are the main components of the amyloid plaques found in the brain of Alzheimer patients (reviewed in Nussbaum et al., 2013). BACE1-AS has 104 nucleotides of full complementarity to exon 6 of human BACE1 mRNA. This complementarity results in the formation of an RNA-duplex thereby increasing the stability of the BACE1 mRNA as shown by RNase protection assays (Faghihi et al., 2008). The stabilization was validated by knockdown of BACE1-AS, resulting in decreased levels of BACE1 mRNA, protein, and Aβ 1-40 and
Aβ 1-42 (Faghihi et al., 2008). A follow-up study elucidated a miR-485-5p miRNA site in BACE1 exon 6: BACE1-AS binding antagonized the RISC complex binding to the miRNA site, thereby stabilizing the mRNA (Faghihi et al., 2010). This mechanism may help explain increased BACE-1 expression in Alzheimer’s, as BACE1-AS can be significantly up-regulated in different brain regions of Alzheimer’s patients, coupled with miR-485-5p downregulation (Faghihi et al., 2010).

1.3.2.5 Airn in genomic imprinting

Several lncRNAs have been discovered to act in the imprinting mechanism, which is essentially controlling expression of certain genes based on the parent-of-origin of the chromosome of a given allele. Airn (antisense Igf2r RNA noncoding) is a 108kb poly(A)+ ncRNA located on mouse chromosome 17, antisense to the Igf2R (insulin-like growth factor 2 receptor) gene. The promoter of Airn lies within the second intron of the Igf2R gene, which is a 3.7kb imprinting control element (ICE) (Lyle et al., 2000). ICE is methylated exclusively on the maternal inherited allele, resulting in only paternal restricted expression of the Airn lncRNA. In mice, Airn is responsible for the cis imprinted expression of Igf2R, Slc22a2 and Slc22a3 (Solute Carrier Family 22, member 2/3), three genes that are located in proximity to the Airn/Igf2R locus. As a portion of the Airn transcript overlaps with the first two exons and the promoter of Igf2R on the paternal allele, its transcription is able to interfere with the action of RNA Polymerase II on the Igf2R gene (Latos et al., 2012; Sleutels et al., 2002). The interference can be released if the Airn is truncated so that it no longer overlaps with the Igf2R promoter (Latos et al., 2012). It is slightly less straightforward how Airn silences the other two neighboring genes. Once expressed, Airn lncRNA spreads throughout this local chromatin environment, as demonstrated by RNA-FISH (fluorescent in situ hybridization) (Nagano et al., 2008). Once bound to the locus, Airn recruits EHMT2 (euchromatin histone-lysine N-methyltransferase 2) resulting in repressive H3K9me3 methylation (Nagano et al., 2008). In addition, transcription of Airn may block the binding of transcriptional activators that promote activating chromatin loops involving the Slc22a2 and Slc22a3 loci (Pauler et al., 2012).
1.3.2.6 XIST in X-chromosome inactivation

The inactivation of the X-chromosome relies on the action of an lncRNA called XIST (Xi-specific transcript). XCI (X-chromosome inactivation) ensures that one X per diploid genome remains active, with the remainder, in females, subject to inactivation (Monkhorst et al., 2008). XCI is critical for mammalian development. Failure of XCI is associated with various forms of intellectual disabilities, infertility, autoimmunity conditions and can even be embryonically lethal (Marahrens et al., 1997; Powell, 1999).

The precise mechanism of X-inactivation is not completely clear, but it is a complex process involving both proteins and ncRNAs. The control of XCI is via the X-inactivation center (Xic), a genomic region on the X-chromosome containing cis-mediated genetic signals required to initiate and maintain XCI (reviewed in Rastan and Brown, 1990). Many ncRNAs are produced within the Xic, with the predominant and best characterized being XIST. XIST is a nuclear lncRNA, ~17,000 nt in length that is spliced and polyadenylated (Brockdorff et al., 1992; Brown et al., 1992). XIST contains up to six regions of randomly arrayed repetitive sequence, with several regions of sequence conserved between mouse and human (Brockdorff et al., 1992; Brown et al., 1992; Nesterova et al., 2001). RNA FISH has demonstrated that XIST coats the entire inactive X-chromosome (Xi) in a structure called a Barr body (Brown et al., 1991).

The coating of XIST along the X chromosome is required for the accumulation of several heterochromatic histone marks over gene dense regions (Costanzi et al., 2000; Kohlmaier et al., 2004; Mak et al., 2002; Plath et al., 2003; Silva et al., 2010). However, it is likely these repressive histone modifications consolidate the XCI effect, rather than initiate it (Kalantry and Magnuson, 2006). It has been shown that despite being early and crucial to XCI in mice, XIST is upstream of the key decision-making mechanism of XCI in rabbits and humans (Okamoto et al., 2011; Wutz and Jaenisch, 2000). Therefore, it is possible that some crucial co-activators of XCI are yet to be discovered. Nevertheless, progress has been made on understanding how XIST coats the X Chromosome. The nuclear matrix protein HNRNPU (Heterogeneous nuclear ribonucleoprotein U), may tether XIST to regions of the inactive X-chromosome (Clemson et al., 1996; Hasegawa
et al., 2010; Helbig and Fackelmayer, 2003; Pullirsch et al., 2010). Additionally, Long Interspersed Nuclear Elements (LINE) on the X-chromosome may boost the ability of XIST to spread (Chow et al., 2010; Fujita et al., 2010; Lyon, 1998; Popova et al., 2006; Tang et al., 2010). Another protein SATB1 (Special AT-Rich Sequence Binding Protein 1), commonly found in nuclear matrix attachment regions, may facilitate spreading through higher-order chromatin organization (Tattermusch and Brockdorff, 2011). The transcription factor YY1-dependent tethering of XIST to its site of synthesis is also a crucial step prior to coating of the inactive X (Jeon and Lee, 2011). Therefore, although the XIST lncRNA coating of the X chromosome is an important part of XCI, it is not yet the complete picture of the XCI mechanism.

Thus, lncRNAs have been identified in almost all aspects of the regulation of gene expression. This diversity of function, particularly in contrast to other ncRNA types, represents both a challenge and an opportunity in lncRNA research. One area that has proved useful for yielding clues to lncRNA function has been to study the sub-cellular localization of a given molecule, in particular, its nuclear organization.

### 1.3.3 LncRNA and subnuclear structures

In a highly organized organelle like the eukaryotic cell nucleus, macromolecules, including lncRNAs, often locate to definable domains that are distinguishable at the ultra-structural level with high-power microscopy. In general, these subnuclear structures have roles in controlling gene expression at both the transcriptional and post-transcriptional levels. Some lncRNAs can reside within a subnuclear structure as part of their function, while some others can act as the seeding molecule for these structures to form around. This section will present an example of each scenario.

#### 1.3.3.1 A lncRNA that is enriched within a subnuclear structure, but not as a structural component: MALAT1 and nuclear speckles

The gene for the lncRNA MALAT1 (metastasis associated lung adenocarcinoma transcript 1) is located on human chromosome 11, producing an infrequently spliced
transcript of approximately 8,000 nt (Hutchinson et al., 2007; Ji et al., 2003). The MALAT1 gene is highly conserved in vertebrates, containing minimal repetitive elements (Ulitsky et al., 2011; Zhang et al., 2012). MALAT1 was first identified as a prognostic marker for metastasis and low survival rates in the early-stages of non-small cell lung cancer (Ji et al., 2003). MALAT1 is among the top five most highly expressed lncRNAs in a wide range of normal adult tissues and cancer tissues (Gibb et al., 2011; Hutchinson et al., 2007).

MALAT1 is a nuclear-enriched lncRNA, and is likely retained in the nucleus by virtue of a small triple-helical structure located at the 3’ end of the transcript. This structure is a type of nuclear retention element, which stabilizes RNA that is not polyadenylated (Brown et al., 2014; Brown et al., 2012). A small tRNA-like structure downstream of the ENE-like element, termed the mascRNA (MALAT1-associated small cytoplasmic RNA), is cleaved by RNase P and exported to the cytoplasm. The function of mascRNA has not yet been characterized (Brown et al., 2012; Wilusz et al., 2008; Wilusz et al., 2012).

RNA FISH against MALAT1 overlaps with immunofluorescence for splicing proteins such as the protein SC35 (Splicing component, 35kDa) (Hutchinson et al., 2007) (Figure 1.3). Splicing proteins are markers for the nuclear domain termed ‘nuclear speckles’, or ‘splicing speckles’ (reviewed in Fu, 1995; Spector and Lamond, 2011). Nuclear speckles are thought to act as a reservoir for splicing factors, and also are involved in the post-translational modification of these factors. Nuclear speckles also frequently localize adjacent to active sites of transcription, potentially for increased efficiency of co-transcriptional splicing (reviewed in Lamond and Spector, 2003). The localization of MALAT1 to nuclear speckles requires the interaction between MALAT1 and a number of mRNA processing factors in the nuclear speckles (Miyagawa et al., 2012; Tripathi et al., 2010). Interestingly, loss of MALAT1 does not disrupt nuclear speckles (Eissmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012), but does reduce the relative enrichment of a number of pre-mRNA splicing factors and SR proteins (serine/arginine-rich splicing factors) to nuclear speckles compared to the nuclear pool of these factor
Figure 1.3: MALAT1 localization. (A) MALAT1 RNA (red) is found associated with SC35 (green) in Tig1 cells. (B) MALAT1 RNA (red) concentrates in domains in about 50% of the cell population, overlapping with SRM-300 speckles (green). (C,D,E) MALAT1 RNA (red in C) preferentially associates with polyadenylated RNA (green in D); overlapping fluorescence shown in (E). (Courtesy: (Hutchinson et al., 2007), scale bar is not provided)
It is postulated that the assembly of nuclear speckles results in MALAT1 recruitment, which in turn facilitates the recruitment of more components.

The physiological role of MALAT1 remains uncertain. Knockdown of MALAT1 in HeLa cells results in alteration of the phosphorylation state of SR splicing factors, with a flow-on effect in altering alternative splicing and a decrease in cell viability (Tripathi et al., 2010). Others have shown that loss of MALAT1 in lung cancer cells did not seem to alter the alternative splicing but rather changed the expression of a number of metastasis-associated transcripts (Gutschner et al., 2013). In support of a pro-proliferative role, the injection of A549 lung cancer cells without MALAT1 into the tail vein of nude mice showed an 80% reduction of metastases, compared to wild type cell xenografts (Gutschner et al., 2013). In animal studies, Malat1 knockout mice are viable, with no obvious changes to both the pattern of global gene expression and alternative splicing (Eißmann et al., 2012; Nakagawa et al., 2012; Zhang et al., 2012). However, a small reduction in expression of the neighboring Neat1 gene was observed in one of the Malat1 knockout mouse studies (Zhang et al., 2012). The mice without MALAT1 in this study also showed no defects in cell cycle progression, a finding which is contradictory to the cell line data (Tripathi et al., 2010; Xu et al., 2011). Thus, there is no consensus for a clear role for MALAT1 beyond cancer progression, in terms of how MALAT1 might influence the general physiology of cells. This is despite MALAT1 being the most abundant and widely expressed lncRNA in hundreds of primary and transformed cell types (with the exception of ribosomal RNAs) (FANTOM, 2014).

1.3.3.2 LncRNA that forms structural scaffolds for subnuclear bodies: hsr-ω and Omega Speckles

In Drosophila there is a well-studied lncRNA-induced subnuclear structure termed ‘omega speckles’ (Figure 1.4). The nucleating molecule for these structures is an lncRNA termed ‘Heat Shock RNA omega (hsr-ω)’. Although Hsr-ω is not the only lncRNA capable of forming a subnuclear body, it is an excellent example to demonstrate the mode of action and potential significance of these bodies.
Figure 1.4: *hsr-omega* and omega speckles.  (A) right arm of chromosome 3 (3R) showing *hsr-ω* gene (93D) and its structure.  The three known Hsr-ω transcripts, including the nuclear Hsr-ω-n and cytoplasmic Hsr-ω-c are shown (Courtesy: (Collinge et al., 2008).  (B) Nucleoplasmic omega speckles (red, arrow) seen after *in situ* hybridization of Hsr-ω 280bp repeat riboprobe with RNA in Malpighian tubule polytene nucleus (blue, chromatin).  The arrowhead indicates the *hsr-ω* locus on the chromosome.  Immunostaining of control (B) and heat shocked (C) Malpighian tubule nuclei with Hrb87F antibody (red).  Hrb87F follows the same speckled pattern as the Hsr-ω-n in control cells, but after heat shock it shows near exclusive localization at the 93D site (D) (Courtesy: (Lakhotia, 2011)).
The hsr-ω gene locus is conserved among Drosophila species, but has not been found in other types of organisms. In D. melanogaster, the hsr-ω gene contains two short exons (~475 and 700bp) separated by a 700bp intron, followed by a long stretch of tandem repeats, each of 280bp. The stretch of repeats can reach up to 5-15kb (Jolly and Lakhotia, 2006) (Figure 1.4). The overall gene may span 10-20kb, producing two major transcripts and one precursor transcript. The 2,000 nt cytoplasmic transcript (Hsr-w-c) is more abundant than the longer (Hsr-ω-n) transcript, which spans the entire length of the gene (Bendena et al., 1991; Garbe et al., 1986; Ryseck et al., 1987). Both transcripts are polyadenylated. The hsr-ω gene is active in all cell types and at various developmental stages of Drosophila, and can be one of the most active loci under heat shock or other stresses (Bendena et al., 1991; Mutsuddi M and Lakhotia, 1995; Prasanth et al., 2000; Tapadia and Lakhotia, 1997) (Figure 1.4).

The long Hsr-ω-n transcript is the most well studied RNA of the hsr-ω group. Hsr-ω-n has a rapid turnover in the nucleus under normal conditions, but is rapidly upregulated and accumulates with increased stability when the overall nuclear transcription is inhibited with stress (Bendena et al., 1989; Hogan et al., 1995; Lakhotia and Sharma, 1995). Hsr-ω-n was found colocalized with a variety of heterogeneous nuclear ribonucleoproteins (hnRNPs), forming a variable number of ‘omega speckles’ (Lakhotia et al., 1999; Prasanth et al., 2000). These omega speckles do not form without the active transcription of the hsr-ω gene, and as Hsr-ω-n is the only nuclear-retained transcript of this gene, it is thought the transcription of Hsr-ω-n is the indispensable factor in the formation of omega speckles (Prasanth et al., 2000). Omega speckles are found both next to the locus of hsr-ω, and away from the locus (Lakhotia et al., 1999; Mao et al., 2011; Prasanth et al., 2000). Of note is that under normal conditions the omega speckle proteins are present in several minor speckles distributed in the nucleus, and at other nucleoplasmic locations that are usually transcriptionally active (Lakhotia et al., 1999; Prasanth et al., 2000). However, under stressful conditions, these minor sites rapidly disappear and the omega speckles proximal to the gene locus enlarge. This stress-induced enlargement is accompanied by the translocation of omega speckle proteins, such as Hrb87f (the Drosophila homolog of human HNRNPA1) and Hrb57a (the Drosophila homolog of human HNRNPA1 and HNRNPF).
homolog of human HNRNPK), from their chromatin binding sites to the enlarged omega speckles, followed by a reduction of transcriptional activity at their previous binding sites (Buchenau et al., 1997; Dangli and Bautz, 1983; Hovemann et al., 1991; Lakhotia et al., 1999; Prasanth et al., 2000; Samuels et al., 1994; Zu et al., 1998). These data suggested a potential sequestration model of omega speckles in regulating the trafficking and availability of hnRNPs and other omega speckle proteins in the nucleus (Prasanth et al., 2000). In this model, it was proposed that Hsr-ω-n sequesters various proteins either to protect the cell, or as a temporary storage of factors into omega speckles, so that they can quickly resume normal function after the stress has passed (Jolly and Lakhotia, 2006; Lakhotia et al., 1999; Prasanth et al., 2000).

Is this mechanism of physiological significance? For many of the mammalian IncRNAs mentioned thus far, although their molecular functions have been well studied, knockdown studies in mice often failed to show any significant effect on development or normal growth. This is certainly not the case for hsr-ω. Flies that are hsr-ω null are frequently embryonic lethal. The small proportion of mutants that hatch are very weak and lacking omega speckles, thereby suggesting a critical role for omega speckles in the embryonic development of Drosophila (Prasanth et al., 2000). The over-abundance of Hsr-ω-n also seems to be pathogenic, leading to sequestration of hnRNPs to affect RNA metabolism, and causing polyglutamine induced neurodegeneration (Mallik and Lakhotia, 2009; Sengupta and Lakhotia, 2006), or male sterility if in the cyst cells of the testis (Rajendra et al., 2001).

Because of the diverse molecular functions of IncRNAs, it is hard to predict the function of individual IncRNAs simply based on their sequence and expression. It appears each IncRNA has individual mechanistic and functional characteristics, however there are still some surprising similarities shared between IncRNAs, sometimes even across significantly different species. This PhD project will focus on the functionality of the IncRNA NEAT1 and the subnuclear structure termed paraspeckles. Many aspects of this IncRNA resemble what we have learnt from Hsr-ω-n and omega speckles, yet NEAT1 is a mammalian IncRNA and Hsr-ω-n is from Drosophila, with no obvious sequence identity.
The following sections describe background information to illustrate why NEAT1 is an interesting lncRNA, and the reasons for why we would want to study its function.

1.4 LncRNA NEAT1 and Paraspeckles

1.4.1 Discovery of paraspeckles
Paraspeckles are mammalian subnuclear bodies that form around the NEAT1 (nuclear paraspeckle abundant transcript 1) lncRNA. Paraspeckles were first described as Interchromatin Granule Associated Zones, electron dense structures distinct from other nuclear bodies observed with the electron microscope in cultured cells (Visa et al., 1993). However, it was in 2002 that a clear marker protein, PSPC1, or Paraspeckle Component 1, was found, and the term ‘paraspeckles’ was used to describe the subnuclear foci in which PSPC1 was enriched (Fox et al., 2002) (Figure 1.5). Paraspeckles are so-named because they are observed in the interchromatin spaces near to, yet distinct from, the nuclear speckles that are enriched in splicing factors (Fox et al., 2002). Additional paraspeckle proteins have since been identified, and these include the DBHS (Drosophila behaviour human splicing) proteins related to PSPC1 (paraspeckles component 1), SFPQ (splicing factor proline/glutamine-rich) and NONO (non-POU domain containing, octamer-binding), as well as a host of other RNA- or DNA-binding proteins (Bond and Fox, 2009; Fox et al., 2005; Naganuma et al., 2012). It is important to note that paraspeckle proteins are found within the nucleoplasmic milieu, as well as inside paraspeckles (Fox et al., 2002). Paraspeckles vary in size from 300nm to up to 1 µm, and their numbers range from 0 to around 20 per cell nucleus. These numbers vary greatly within cell populations and between cell types (reviewed in Bond and Fox, 2009).

1.4.2 Discovery of NEAT1 in paraspeckles
In the years following the identification of paraspeckles, several clues suggested that RNA would likely be crucial to both paraspeckle structure and formation. Firstly, the paraspeckle proteins were all known RNA-binding proteins, and several only localize to paraspeckles via key RNA Recognition Motifs (RRM). Secondly, paraspeckles were
Figure 1.5: Paraspeckles and human *NEAT1* gene structure. (A) Fluorescence micrographs showing HeLa cells expressing YFP-Paraspeckle protein PSPC1 (green), and RNA-FISH of NEAT1 (red). Detection of the colocalization of paraspeckle protein and NEAT1 RNA allows accurate determination of paraspeckles, and reduction of NEAT1 with anti-sense oligonucleotides results in the loss of paraspeckles (Courtesy: (Wilusz et al., 2009)). (B) UCSC genome browser tracks of the human *NEAT1* gene, in order of: gene, conservation, CpG island, repeating elements and expression level in HeLa cells. Locations of ENE (nuclear retention element) and MenRNA (tRNA-like structure) are labelled in grey and blue respectively.
sensitive to degradation by RNase treatment. Thirdly, paraspeckles only formed in newly divided cells once RNA polymerase II transcription was well established. Finally, paraspeckles were disassembled by inhibition of RNA Polymerase II transcription (Dye and Patton, 2001; Fox et al., 2005; Fox et al., 2002).

In 2009, three groups reported that paraspeckles were formed around the NEAT1 IncRNA, and that NEAT1 was an essential structural component of paraspeckles (Clemson et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009). NEAT1 is a mammalian-specific gene located on human chromosome 11q13, and mouse 19qA (Hutchinson et al., 2007; Sasaki et al., 2009). Sequence conservation analysis of the NEAT1 gene suggests that it might have arisen by the duplication of the MALAT1 gene (Stadler, 2010). Given that NEAT1 is restricted to mammalian genomes, yet MALAT1 is found within all vertebrates, this may have occurred relatively recently in evolution. Indeed, the NEAT1 and MALAT1 gene loci are syntenically conserved in mammals, and are found 70kb apart in the human genome.

The NEAT1 gene promoter triggers transcription of two major isoforms of RNA that overlap completely at their 5’-end, yet have very different 3’-ends (Hutchinson et al., 2007; Sasaki et al., 2009; Sunwoo et al., 2009) (Figure 1.5). Both transcripts are predominantly nuclear. The shorter canonically polyadenylated isoform (3,700 nt in human, 3,100 nt in mouse), is termed NEAT1_1 (originally termed MENε). The longer non-canonically polyadenylated isoform, 23,000 nt in human and 20,500 nt in mouse, is termed NEAT1_2 (originally MENβ). Similar to the 3’ end of MALAT1, the 3’ end of NEAT1_2 also has a small triple-helical nuclear retention element (Brown et al., 2014; Brown et al., 2012). In addition, also similar to MALAT1, NEAT1_2 carries an unusual tRNA-like structure named MenRNA at its 3’-end that is recognized and cleaved by RNase P. Following cleavage, NEAT1_2 contains a 3’-end with a short genomically encoded poly(A)-rich sequence (Sunwoo et al., 2009). NEAT1_2 has high level of mammalian conservation at the 5’ and 3’ ends of the transcript, but carries numerous non-conserved repetitive regions in the middle of the transcript. Thus, the NEAT1_1 transcript may be considered as a conserved 5’-end of NEAT1_2 (Figure 1.5).
It is relevant here to make a comment about the nomenclature of the \textit{NEAT1} gene. I have chosen to use the uppercase term ‘NEAT1’ when discussing any RNA produced from the human or murine \textit{NEAT1} gene; ‘NEAT1\_1’ for the shorter RNA transcript, ‘NEAT1\_2’ for the longer transcript. In mouse, in line with mouse convention, the gene name is referred to in lower case as \textit{Neat1}.

Paraspeckles form in close proximity to the \textit{NEAT1} gene locus (Clemson et al., 2009). Several different studies and approaches have been used to show that transcription of NEAT1 triggers paraspeckle formation (Clemson et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009). Spector and colleagues generated a cell line with inducible expression of an array of MS2-tagged \textit{NEAT1} loci. Once induced, fluorescently tagged MS2 binding protein was used to visualize nascent NEAT1 transcripts. Recruitment of DBHS and other paraspeckle proteins to the nascent NEAT1 transcript, forming paraspeckles, was then observed (Mao et al., 2011). This study also demonstrated that paraspeckle formation and maintenance depended on the active transcription of NEAT1, rather than the pre-existing NEAT1 transcripts in the nucleus (Mao et al., 2011). In a variety of cultured cell lines, both NEAT1 isoforms clearly display the characteristic punctate localization typical of paraspeckles, co-localizing and co-purifying with DBHS proteins (Clemson et al., 2009; Mao et al., 2011; Sasaki et al., 2009; Sunwoo et al., 2009) (Hutchinson et al., 2007; Sreenivasa Murthy and Rangarajan, 2010). The two isoforms of NEAT1 have very different abundance in cells. In all cell lines examined so far, regardless of the detection technique used, NEAT1\_1 is consistently at least four times as abundant as the NEAT1\_2 isoform, and typically, much more than four times (Figure 1.5B, (ENCODE, 2012; Nakagawa et al., 2011; Sunwoo et al., 2009). Giving this, techniques that aim to detect the 5’ end of NEAT1, which is shared by both NEAT1\_1 and NEAT1\_2, are primarily detecting the NEAT1\_1 isoform. In contrast, it is possible to specifically amplify NEAT1\_2, by targeting the 3’end of the NEAT1 RNA. In the literature, RNA-FISH of both the 5’ and 3’ end of NEAT1 have been shown to locate with paraspeckle proteins, suggesting both isoforms can be physically seen in paraspeckles (Naganuma et al., 2012; Nakagawa et al., 2011).
While both isoforms are found in paraspeckles, it is now generally accepted that the two major NEAT1 isoforms play different roles in the structural assembly of paraspeckles. More specifically, the transcription of the less abundant NEAT1_2 transcript, instead of the more abundant NEAT1_1 (Figure 1.5B), is the critical molecule for the assembly and maintenance of paraspeckles. This was first suggested by experiments showing that siRNA specifically targeting NEAT1_2 was sufficient to ablate paraspeckles (Sasaki et al., 2009). The conclusive experiments were performed with MEF (mouse embryonic fibroblast) cells from NEAT1 knockout mice, which are devoid of paraspeckles. By transiently over-expressing either NEAT1_1, or NEAT1_2 in these MEFs, it was observed that only the overexpression of NEAT1_2 could restore the appearance of paraspeckles (Naganuma et al., 2012). This discovery thus leaves an unsolved question as to why NEAT1_1 is localized to paraspeckles, given it is not essential for paraspeckle formation.

1.4.3 Proteins in paraspeckles

NEAT1_2 is the seeding molecule for the formation of paraspeckles, but many proteins are also crucial to paraspeckle formation. As previously mentioned, paraspeckles were first discovered based on the punctate localization of DBHS (Drosophila behavior human splicing) proteins, and hence these proteins are frequently used as paraspeckle markers (Fox et al., 2002). DBHS proteins all carry RRM (RNA recognition motif) RNA-binding domains that are required for their localization to paraspeckles (reviewed in Bond and Fox, 2009). More recently, DBHS proteins were shown to heterodimerize and form long extended oligomers that are crucial to paraspeckle formation (Lee et al., 2015; Passon et al., 2012). Among the three DBHS proteins, NONO (Non-POU Domain-Containing, Octamer-Binding) and SFPQ (Splicing Factor Proline/Phosphate-Rich) are essential for the formation and maintenance of paraspeckles, as siRNA knockdown of either protein results in the loss of paraspeckles in HeLa cells (Sasaki et al., 2009). Interestingly, despite being the original paraspeckle marker protein, PSPC1 (Paraspeckle Component 1) is dispensable for paraspeckle formation (Fox et al., 2002; Sasaki et al., 2009). It is important to note that immobilizing DBHS proteins to chromatin could not effectively recruit NEAT1 to form de novo paraspeckles, again consolidating the notion that the transcription of NEAT1 must come first before the binding of paraspeckle proteins.
DBHS proteins are considered multifunctional nuclear proteins. They can bind double-stranded DNA, double- and single-stranded RNA, and their functions are closely related to transcription, including transcription initiation (Dong et al., 1993; Yang et al., 1997; Yang et al., 1993), co-activation (Amelio et al., 2007; Kuwahara et al., 2006), co-repression (Dong et al., 2009; Mathur et al., 2001), constitutive and alternative splicing (Ito et al., 2008; Kameoka et al., 2004; Patton et al., 1993; Peng et al., 2002), and transcriptional termination (Kaneko et al., 2007). DBHS proteins can also facilitate the repair of double stranded DNA damage (Bladen et al., 2005; Ha et al., 2011; Krietsch et al., 2012; Li et al., 2014; Morozumi et al., 2009; Rajesh et al., 2011; Salton et al., 2010; Udayakumar et al., 2003).

In addition to DBHS proteins, several other proteins were recently found to be capable of paraspeckle localization. In a systematic high-throughput microscopy study that examined the localization of more than 18,000 human proteins in HeLa cells, almost 40 new paraspeckle proteins were identified (Naganuma et al., 2012). Almost all of these proteins have some relationship with transcriptional or post-transcriptional gene regulation and contain RNA or DNA binding domains. By systematically knocking down each new paraspeckle protein in turn and measuring the effects on paraspeckle formation and NEAT1 stability the importance of each of these proteins in paraspeckle formation was determined, leading to the finding that there are seven proteins essential for the formation of paraspeckles, or maintaining the stability of NEAT1 (Naganuma et al., 2012; Sasaki et al., 2009). These essential proteins include the DBHS proteins NONO and SFPQ, RBM14 (RNA Binding Motif Protein 14), HNRNPK (Heterogeneous nuclear ribonucleoprotein K), DAZAP1 (DAZ associated protein 1), FUS (fused in sarcoma), and HNRNPH3 (heterogeneous nuclear ribonucleoprotein H3) (Naganuma et al., 2012).

The protein-protein interactions within paraspeckles are also important for normal paraspeckle formation. Mutations causing loss and gain of function to FUS can trigger
disruption of paraspeckle assembly, and lead to abnormal aggregates of inclusions of NONO (Naganuma et al., 2012) (Shelkovnikova et al., 2014). Meanwhile, it has been implicated that DBHS proteins might be able to form long oligomers through the interaction between the extended coiled coils of DBHS dimers, which was crucial for the formation of paraspeckles as its disruption would also compromise paraspeckle formation (Lee et al., 2015; Passon et al., 2012).

1.4.4 Assembly and structure of paraspeckles

We now have a reasonable model for paraspeckle formation. The proposed process starts with the transcription of NEAT1_1. In the majority of instances, the poly(A) signal at the end of NEAT1_1 is recognized by CPSF6 (Cleavage and polyadenylation Specific Factor 6) and NUDT21 (Nudix type motif 21), that bind NEAT1_1 and promote cleavage efficiently to mediate canonical polyadenylation (Naganuma et al., 2012). Thus, NEAT1_1 is produced and NEAT1_2 is not transcribed. In rare occasions, the cleavage and polyadenylation of NEAT1_1 can be compromised which leads to the transcription of the longer NEAT1_2 transcript. Naganuma and colleagues showed that the paraspeckle protein HNRNPK competes with NUDT21 and CPSF6 to occasionally disrupt their action on NEAT1_1 (Naganuma et al., 2012). This model was confirmed through siRNA knockdown of CPSF6, which resulted in the increased prevalence of NEAT1_2 and paraspeckles (Naganuma et al., 2012). On the other hand, strong overall NEAT1 transcription also seems to be able to induce NEAT1_2 production more rapidly (Hirose et al., 2014). Nevertheless, once NEAT1_2 transcription takes place, the presumption is that many binding sites for RNA-binding proteins are created, recruiting the paraspeckle proteins. Together, the proteins and NEAT1_2 build up the paraspeckle itself (Figure 1.6). It is still a mystery as to how NEAT1_1 associates with the structure, but it is likely through RNA-protein or RNA-RNA interactions.

The spatial organization of NEAT1 in paraspeckles has been determined using electron microscopy and in situ hybridization of probes targeted to different sequences within NEAT1_2 (Souquere et al., 2010). These results have confirmed the importance of NEAT1_2 to paraspeckle structure. This method determined that the 5’ and 3’ ends of
NEAT1_2 (and, by inference, NEAT1_1) were found at the periphery of paraspeckles, whereas the center of NEAT1_2 is located within the core. In contrast, the DBHS protein NONO, is found throughout the paraspeckle (Figure 1.6A). It appears the length of the NEAT1_2 transcript is important for the structural assembly of paraspeckles, as the human NEAT1_2 lacking 10kb sequence at the 3’ end is not capable of supporting paraspeckle formation (Sasaki et al., 2009) (Figure 1.6B,C). These studies have elucidated paraspeckle structure. The next section addresses what is known about paraspeckle function.
Figure 1.6: The assembly and structure of paraspeckles. (A) When the NEAT1 gene is actively transcribed, the majority of the transcripts form NEAT1_1 due to the typical cleavage and polyadenylation signals, and only a small proportion of transcripts escape cleavage and are transcribed into NEAT1_2 transcripts, which act as the structural backbone of paraspeckles. Nascent paraspeckles always form near
to the NEAT1 gene locus (not drawn to scale). (B) EM-coupled in situ hybridization with probes targeting different regions of NEAT1 revealed an organized spatial distribution of NEAT1, where both the 5' and 3' ends locate at the periphery of paraspeckles, with the middle region inside paraspeckles. Note that the 5’ probe binds to both NEAT1_1 and NEAT1_2. (C) EM-immunostaining the paraspeckle protein NONO is located throughout the entire structure. (D) RNA-FISH of NEAT1 under fluorescent microscopy showing punctate appearance of paraspeckles, red: NEAT1, blue: DAPI, scale bar = 10μm. (B) and (C) are EM images courtesy (Souquere et al., 2010), scale bar = 100nm.
1.4.5 Functions of paraspeckles

1.4.5.1 Nuclear retention of specific mRNAs

Not long after the discovery of paraspeckles, the first functional insights were reported. In a study investigating nuclear-retained poly(A⁺) RNAs, a specific mRNA was identified as partially co-localized in paraspeckles (Prasanth et al., 2005). This mRNA was an alternative isoform of a typical gene, with a long 3’-untranslated region (3’UTR) containing inverted Alu repeat structures (Prasanth et al., 2005). These inverted repeats could form adenosine-to-inosine edited dsRNA regions, known substrates of DBHS proteins (Zhang and Carmichael, 2001). It was speculated that these dsRNA regions resulted in DBHS protein binding, and subsequent paraspeckle targeting and nuclear retention of this RNA. Specific signaling molecules such as interferon gamma resulted in the RNA translocating to the cytoplasm, along with the cleavage of the inverted repeats, and a concomitant increase in translation (Prasanth et al., 2005). Several years later, following the discovery of NEAT1 in paraspeckles, it was demonstrated that knockdown of NEAT1 altered the nuclear retention of these inverted Alu repeat RNAs (Chen and Carmichael, 2009). It is not yet known how the dsRNA regions are cleaved, but many aspects of this nuclear retention mechanism could also be applied to other genes with inverted repeats in their 3’-UTRs, including LIN28 (Lin-28 homolog A), NICN1 (nicolin 1), and APOBEC3G (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G) (Chen and Carmichael, 2009; Mao et al., 2011) (Figure 1.7A). However, it has also been found that some other genes with similar A-to-I edited inverted Alu repeats in their 3’UTR are not subject to nuclear retention at all, but are exported to the cytoplasm and in this case the RNAs are translationally repressed (Capshew et al., 2012; Fitzpatrick and Huang, 2012). The repression appears to be mediated by cytoplasmic stress granules, which can form in response to heat shock stress (Capshew et al., 2012; Fitzpatrick and Huang, 2012). Because of this dual translational repression mechanism, it is still not known how widely the paraspeckle nuclear retention mechanism of mRNAs is utilized by the cells.
Figure 1.7: Paraspeckles play regulatory roles in gene expression through protein sequestration and mRNA nuclear retention. (A) mRNA nuclear retention model. Paraspeckles bind to inverted repeats of certain mRNA under normal conditions. When cells are stressed, inverted repeats are cleaved, and mRNAs are exported to the cytoplasm, leading to increased translation. 

(B) NEAT1 Level effects. High NEAT1 levels lead to more Paraspeckles, which enhance the nuclear retention of mRNAs, resulting in higher TF target gene expression and increased gene of interest expression. Low NEAT1 levels lead to fewer Paraspeckles, reducing mRNA nuclear retention, and lower gene of interest expression.
are first released into the nucleus and then exported into the cytoplasm for translation. (B) Protein sequestration model. As the size of the paraspeckles change with the transcriptional level of NEAT1, certain transcription factors relocate to paraspeckles from the nuclear pool, and hence their downstream target genes will be affected. (Images courtesy (Hirose et al., 2014).
1.4.5.2 Subnuclear sequestration of paraspeckle proteins

Recently it has been suggested that an additional molecular function for paraspeckles could be the subnuclear sequestration of paraspeckle proteins, leading to depletion of these factors from the nuclear milieu, and thereby the reduction of their influence. Two recent reports showed how an increase in paraspeckle size results in the quantitative relocation of the essential paraspeckle protein SFPQ from its target sites in chromatin to paraspeckles, leading to a change in the regulation of SFPQ target genes (Hirose et al., 2014; Imamura et al., 2014) (Figure 1.7B). In these reports IL8 (interleukin 8) and ADARB2 (adenosine deaminase, RNA-specific, B2) were the key SFPQ target genes influenced. Interestingly, SFPQ acts as a transcriptional repressor for IL8, and an activator for ADARB2. Thus far, SFPQ is the only paraspeckle protein known to be subject to this sequestration mechanism, however, given the presence of more than 40 proteins within paraspeckles, there are likely to be many other target genes of these other proteins that may be influenced by paraspeckles. The sub-nuclear sequestration of nuclear proteins has been hypothesized as a function for other nuclear bodies that also rely on a structural lncRNA component for their assembly. For example, the Satellite III (SatIII) lncRNAs are thought to protect cells from unnecessary or harmful transcriptional, splicing, or translational events following heat shock by sequestering active transcriptional machinery, as well as nucleating various splicing factors and tRNAs into nuclear stress bodies (nSB) (Biamonti and Vourc’h, 2010; Lindquist, 1986; Metz et al., 2004). Additionally, omega speckles, as described above, also are thought to act through sequestration.

Both mRNA nuclear retention and the sequestration of proteins can be regarded as a single model for paraspeckle function: the sequestration of molecules, either preventing their transport to the cytoplasm, or preventing their normal nuclear distribution. By controlling the abundance of paraspeckles, cells may therefore be able to regulate those paraspeckle retained proteins and mRNAs.

1.4.5.3 Physiological functions for paraspeckles

Despite the relatively detailed understanding of the molecular assembly and functional
modes of paraspeckles, the physiological effects of NEAT1 knockout in animal models has not been clearly defined until very recently. The generation of *Neat1*<sup>−/−</sup> mice in 2011 provided a useful tool to allow further investigation into these physiological effects. The very initial observation of the phenotype was that *Neat1*<sup>−/−</sup> mice were viable and fertile under laboratory conditions, showing no apparent phenotypes except for the disappearance of paraspeckles (Nakagawa et al., 2011). This was a surprising result, because NEAT1 is such an abundant transcript, with a high level of conservation, at least for NEAT1_1, among mammals.

Indeed, two recent reports have now illustrated the importance of NEAT1 in animal physiology. The first study showed defects in *Neat1*<sup>−/−</sup> mice in the formation of the corpus luteum, an important structure supporting pregnancy. In mice this structure has highly variable expression of NEAT1 and paraspeckles (Nakagawa et al., 2014). The corpus luteum is normally involved in the production of progesterone, signaling the maintenance of the thick lining of the uterus that will provide a rich blood supply for zygotes to develop. However, in *Neat1*<sup>−/−</sup> mice this tissue does not form in a certain proportion of knockout animals, thus leading to failure of pregnancy due to insufficient secretion of progesterone (Nakagawa et al., 2014). In the second report, a lack of NEAT1 and paraspeckles leads to aberrant mammary gland branching morphogenesis and lobular-alveolar development, resulting in lactation defects, collectively causing the malnutrition of pups, and therefore starvation (Standaert et al., 2014). In both reports, the authors emphasized that the defects observed within the development of the corpus luteum and mammary gland tissues occurred stochastically from knockout mouse to mouse. This implies that the general function of NEAT1 and paraspeckles might be cell type- or condition-specific, leaving their importance subject to change based on the internal and/or external environmental effects upon the tissue.

1.4.6 Expression of NEAT1

NEAT1 has been frequently reported to be one of the most dynamically regulated IncRNA in a variety of experimental systems. The expression of NEAT1 often fluctuates in different diseases, cellular states, or with the introduction of biochemical stimulants.
The location of the *NEAT1* locus was initially published in 1997, as a genomic region with linkage analysis for Multiple Endocrine Neoplasia Type 1 cancer syndrome (Guru et al., 1997). Subsequently, a small 0.5kb poly(A+) *NEAT1* isoform named TncRNA (Trophoblast non-coding RNA transcript, alternatively known as TSU, located at the 3’ end of *NEAT1_1*), was found to bind to STAT1 (signal transducer and activator of transcription 1) in the cytoplasm to inhibit its nuclear translocation, and hence suppressed the MHC (Major Histocompatibility Complex) class II expression induced by IFN-γ in human trophoblasts (Peyman, 1999; Peyman, 2001). TncRNA also suppresses the promoter of CIITA (Class II transactivator) and MHC class II expression in murine B-lymphocytes (Geirsson et al., 2004; Geirsson et al., 2003a). Together, these data suggested a role for TncRNA in dampening the immune response of maternal lymphocytes against cells expressing paternal antigens, thus promoting the survival of the placenta and fetus (Peyman, 1999). Interestingly, a connection between these observations and the now considerable literature on paraspeckles and the other, longer, *NEAT1* isoforms, is yet to be made.

With the discovery of the widespread and abundant *NEAT1_1* isoform in 2007 (Hutchinson et al., 2007), the opportunity then arose to interrogate many microarray studies, since probes against *NEAT1_1* were included on many standard microarrays. This, coupled with some studies examining paraspeckle prevalence, has allowed us to obtain a clearer picture of the pattern of *NEAT1* expression.

**1.4.6.1 Association with differentiation in vitro**

There are cell types and tissues that express low, or undetectable levels of *NEAT1*, but can nevertheless be stimulated, resulting in upregulation of *NEAT1*. Embryonic and fetal tissues of the mouse generally do not express *NEAT1* (Nakagawa et al., 2011). Human Embryonic Stem Cell lines (ESCs) also do not express *NEAT1*, however *NEAT1* becomes upregulated as ESCs are differentiated into trophoblasts by treatment with BMP4 (bone morphogenetic protein 4) (Chen and Carmichael, 2009). *NEAT1* is also dramatically upregulated in C2C12 (murine myoblast) cells are differentiated into myotubes via serum starvation (Sunwoo et al., 2009). In this case, the upregulation of
both NEAT1 RNA levels and paraspeckle prevalence was observed. Similar observations were made when neuronal/oligodendrocyte progenitor cells were induced to mature after manipulating the concentration of certain molecules in the culture media (Mercer et al., 2010). NEAT1 upregulation was also observed during all-trans retinoic acid (ATRA)-induced NB4 cell differentiation. Interestingly, the reduction of NEAT1 by small interfering RNA (siRNA) was found to hamper this differentiation (Zeng et al., 2014).

A systematic study of NEAT1 and paraspeckle appearance in mouse was conducted by Nakagawa et al., who carried out in situ hybridization on mouse tissues using probes targeted to either 5’ end of NEAT1, common to both isoforms, or to the unique 3’ end of NEAT1_2 (Nakagawa et al., 2011). They reported that the majority of tissues within the adult mouse showed evidence of NEAT1 expression when using the probe to the 5’ end whereas embryonic and fetal tissues showed very low, or undetectable levels of NEAT1. Interestingly, the NEAT1_2 specific probes showed expression only within restricted highly differentiated, and often secretory tissues such as adult salivary, adrenal and Harderian glands, and the stomach epithelial lining (Nakagawa et al., 2011). This intriguing observation has suggested a possible paraspeckle independent role for NEAT1_1, which will be discussed further below.

1.4.6.2 Relevance to physiological conditions

The expression of NEAT1 is also altered within a number of physiological conditions and diseases. NEAT1 is amongst the top five most abundant lncRNAs in a wide range of cancer tissues (Gibb et al., 2011). NEAT1 was upregulated in the nucleus accumbens within post-mortem brain tissues of heroin abusers (Michelhaugh et al., 2011). In babies with intrauterine growth restriction (a multifactorial disease defined by an inability of the fetus to reach its growth potential), the placental NEAT1 expression was found almost four fold higher than normal subjects (Gremlich et al., 2014). Infection can also induce NEAT1. For example, NEAT1 is upregulated in the brains of mice after Japanese encephalitis virus infection (Saha et al., 2006). HIV1 infection also upregulates NEAT1, in what is thought to be a cell defense mechanism, as NEAT1 knockdown increased viral
production (Zhang et al., 2013). Increasing NEAT1 levels are also observed within early-stage decaying motor neurons in the spinal cord of ALS (amyotrophic lateral sclerosis) patients (Ito et al., 2011; Nishimoto et al., 2013). It is possible there is a strong connection between paraspeckles and ALS, as several paraspeckle proteins are known to have mutations that cause ALS, including FUS, and several of these proteins show increased binding to NEAT1 in post-mortem brain samples of patients with the related condition, FTLD (frontotemporal lobar degeneration) (Tollervey et al., 2011).

On the other hand, some stresses seem to downregulate NEAT1. NEAT1 was downregulated moderately (two-fold) in muscle biopsies from Duchenne muscular dystrophy patients, but was upregulated two-fold after subjects performed endurance training (Timmons et al., 2005). NEAT1 is downregulated 2-3 fold upon nicotine stress in MCF-10A normal breast epithelial cells (Bavarva et al., 2013). Nicotine is reported to induce oxidative stress in cultured cells. Furthermore, NEAT1 expression is also significantly repressed in acute promyelocytic leukemia samples compared with those from healthy donors (Zeng et al., 2014).

Knockdown studies have also generated some interesting results in relation to NEAT1 expression. NEAT1 siRNA knockdown resulted in increased apoptosis, decreased cell viability, and morphological changes of the giant cells in the Burkitt’s lymphoma BJAB cell line (Halford, 2013). Another study showed that conditioned medium harvested from human adipose derived stem cells in which NEAT1 was knocked down was impaired in promoting neural cell repair. This suggests that NEAT1 knockdown cells have altered secretion of growth factors beneficial for neuronal regeneration (Tajiri et al., 2014).

In summary, in mammals, NEAT1 expression is widespread, potentially regulated by complex mechanisms, and may be affected by many different environmental triggers. It is unclear how much NEAT1 expression data relates to paraspeckle prevalence and function, with only some studies examining this relationship.
1.4.7 Unsolved questions relating to NEAT1 and paraspeckles.

After a decade of research, the bigger picture of NEAT1 and paraspeckle function is gradually becoming clear. NEAT1 is widely expressed in differentiated tissues, and can be further induced under many conditions. The strong induction sometimes leads to the transcription of the NEAT1_2 isoform, and hence causes the formation of paraspeckles. Paraspeckles are now defined as the RNA-protein complex that is primarily built on NEAT1_2 transcript collaboratively with several essential paraspeckle proteins. The formation of paraspeckles then may sequester mRNAs or nuclear proteins, and hence regulate their location and downstream biological activities. This mechanism of regulating gene expression might generate benefits for the organism. However, one obvious problem in this model is that the link between the molecular sequestration function of NEAT1 and paraspeckles, and the physiological significance of their existence shown by the high conservation of NEAT1 among mammals and phenotypes of Neat1-/- mice, has never been elucidated. In other words, there is still no systematic understanding of why cells need paraspeckles and NEAT1 in various situations, and how they benefit the cells and organism through their molecular functions. The aim of this Ph.D project is to address this issue. Following this logic, if the ‘systematic understanding’ is the key, then there are at least three aspects that need more investigation.

Firstly, the transcriptional regulation of NEAT1 expression is still not fully understood. As aforementioned, NEAT1 is widely expressed in a variety of tissues, and is highly dynamic and inducible with many environmental triggers. This generally reflects a lack of tissue specific expression, suggesting that either the molecular determinants of NEAT1 expression are controlled by many factors at the same time, or are controlled by factors that are shared between large numbers of cells or tissues. Therefore, understanding the major upstream molecular pathways that control NEAT1 expression under specific conditions will be important, as these can link the expression of NEAT1 to various triggers.

Secondly, it is important to determine the extent of influence of the sequestration model to gene expression and cell physiology. Only two clear examples of paraspeckle target genes are known (IL8 and ADARB2, as a result of SFPQ sequestration), but the full extent
of paraspeckle/NEAT1 target genes and the exact molecular pathways that NEAT1 and paraspeckles modulate through sequestration have never been examined systematically. With this information, it should be possible to predict which cellular functions or processes could be affected the most, and link these to the phenotypes exhibited by the Neat1−/− mice and diseases with NEAT1 involvement.

Lastly, the relationship between the two NEAT1 isoforms, and how this relates to paraspeckles, remains unknown. If the main function of NEAT1 transcription is to make paraspeckles, a process dependent on NEAT1_2, then what is the function of the NEAT1_1 isoform? Further, when looking at the expression of the NEAT1_2 isoform specifically, there are a very limited number of tissues in which it is expressed, even though the majority of tissues express NEAT1_1 very abundantly. This makes the relationship between NEAT1_1 and NEAT1_2 very confusing. If NEAT1 is expressed at low levels, then cells will not make NEAT1_2 or paraspeckles, resulting in free NEAT1_1 with unknown function; however if NEAT1 expression is high, then cells will make both NEAT1_1 and NEAT1_2, thereby forming paraspeckles. The result of paraspeckle formation is that NEAT1_1 concentrates into paraspeckles rather than floating freely in the nuclear pool. Therefore, in cells which have no paraspeckles, yet NEAT1_1 is still expressed, what is the function of NEAT1_1?

Once these three questions are addressed, we will be in a much better position to understand the significance of NEAT1 and paraspeckles, how they are connected to general cell regulatory systems and their importance to the whole organism if there is any.
1.5 Hypothesis and aims

The goal of this project is to understand the role of NEAT1 and paraspeckles systematically. The three aims to address this are:

**Aim 1: To understand the upstream regulatory system that controls NEAT1 expression.**
- I hypothesize that the upstream regulatory system of NEAT1 would involve factors from many different molecular pathways.

**Aim 2: To identify the downstream molecular pathways that are controlled by NEAT1 and paraspeckles.**
- I hypothesize that these downstream molecular pathways may promote cell proliferation, viability or stress resistance, hence offering cells advantages under certain conditions.

**Aim 3: To identify if NEAT1_1 holds functions outside of paraspeckles.**
- I hypothesize that NEAT1_1 may regulate gene expression in the absence of NEAT_2 or paraspeckles, and the presence of paraspeckles could regulate NEAT1_1 function through sequestration, with a similar mechanism to the sequestration of paraspeckle proteins.

To address the aims, a number of experiments and analyses were performed:

**For Aim 1:**
1. NEAT1 promoter analysis was conducted utilizing publicly available databases, and used bioinformatics tools to characterize the regulatory system that controlled human NEAT1 expression.

2. Relative transcriptional strength of *NEAT1* promoter and enhancers was determined.
3. Conditions that consistently upregulated NEAT1 were identified, and then at least one influential transcription factor that mediated the induction was determined.

**For Aim 2:**
1. Stable human cancer cell lines were generated that were not able to produce NEAT1.

2. Transcriptomic differences between these cells and wild type cells with focus on global gene expression were identified, and then major pathways or cellular functions that were affected by the NEAT1 knockout were examined.

3. The affected cellular functions in response to NEAT1 knockout were validated via *in vitro* cell assays.

**For Aim 3:**
1. Stable human cancer cell lines that were not able to produce NEAT1_2 isoform, yet still had intact NEAT1_1 isoform at normal expression level, were generated.

2. Differential gene expression analysis on wild type cells, total NEAT1 knockout cells, and NEAT1_2 specific knockout cells were conducted to search for evidence that supports the paraspeckle-independent functions of NEAT1_1.

**Significance of this project:**
The outcome of this project is to obtain a much better understanding of the role of NEAT1 and paraspeckles in cell biology. This study will provide fundamental knowledge on how NEAT1 expression and paraspeckles are related with multiple diseases and developmental conditions, in order to establish a basis for the future discovery of any therapeutic use of NEAT1 and paraspeckles. These studies are also important more broadly within the emerging field of lncRNA biology, which still requires more understanding on how RNA biology is involved with the complex regulatory networks we have already established with proteins.
**Structure of the results chapters:**

Chapter 3 is presented in manuscript format, ready for submission. This chapter presents the main story of this project, describing how NEAT1 induction and function are related to cell biology, and indications of the functional relationship between NEAT1_1 and NEAT1_2/paraspeckles. This chapter therefore contains answers for all the aims above, and is the essence of the entire PhD project.

Chapter 4 and 5 are presented in a typical thesis format. Chapter 4 covers a more thorough exploration of the understanding of the regulation of NEAT1 expression. Chapter 5 covers more details on how the unprecedented human stable NEAT1 knockout cells were generated using the CRISPR-Cas9 technology.
Chapter 2
Materials and Methods
2.1 Database and software

**Databases**

Data tracks for NEAT1 promoter analysis and gene structure are available from UCSC genome browser: [http://genome.ucsc.edu/](http://genome.ucsc.edu/)

UniProt protein ID and gene ontology was obtained from the UniProt database: [http://www.uniprot.org/](http://www.uniprot.org/)

The DNA methylation status of cells expressing NEAT1 was obtained from the Human induced pluripotent stem cell methylome browser: [http://neomorph.salk.edu/ips_methylomes/browser.html](http://neomorph.salk.edu/ips_methylomes/browser.html)

DAVID, The Database for Annotation, Visualization and Integrated Discovery provided the gene ontology analysis: [http://david.abcc.ncifcrf.gov/](http://david.abcc.ncifcrf.gov/)

Ingenuity was used for network and pathway analysis: [http://www.ingenuity.com/](http://www.ingenuity.com/)

**Algorithms and software**

p53MH software was used for p53 DNA binding site prediction: [http://www.jurgott.org/linkage/p53MH.htm](http://www.jurgott.org/linkage/p53MH.htm)

This software helps to generate a score for the likelihood of p53 binding to a DNA sequence (Hoh et al., 2002).

CRISPR Design Tool: [http://crispr.mit.edu/](http://crispr.mit.edu/)

This tool helps to calculate number of off-target sites of each potential CRISPR guide RNA that can be used for a provided genomic sequence (Hsu et al., 2013).

pDRAW32: [www.acaclone.com](http://www.acaclone.com)

was the standard software for plasmid construction and DNA analysis in this thesis.
EMBOSS Needle Pairwise Sequence Alignment Tool is the standard tool available online for sequence alignment:

http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=emboss_needle&context=nucleotide

2.2 Suppliers

See Table 2.1 for suppliers of services.
See Table 2.2 for suppliers of reagents and kits.
See Table 2.3 for suppliers of antibodies and probes.
See Table 2.4 for suppliers of instruments
Table 2.1: Suppliers of services

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian Genome Research Facility (AGRF; Melbourne, Australia)</td>
<td>RNA sequencing by Illumina HiSeq 2000</td>
</tr>
<tr>
<td>Geneworks (Thebarton, Australia)</td>
<td>Oligonucleotide synthesis</td>
</tr>
<tr>
<td>Integrated DNA Technologies (IDT; Coralville, USA)</td>
<td>Oligonucleotide synthesis</td>
</tr>
<tr>
<td>Macrogen (Seoul, South Korea)</td>
<td>Sanger DNA sequencing</td>
</tr>
</tbody>
</table>
### Table 2.2 Suppliers of reagents and kits

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Reagents and kits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amresco (Solon, USA)</td>
<td>LB Broth (Lysogeny broth) (Miller)</td>
</tr>
<tr>
<td>Becton Dickinson Biosciences (BD Biosciences; Franklin Lakes, USA)</td>
<td>BD Falcon 12 x 75 mm Tube with Cell Strainer Cap</td>
</tr>
<tr>
<td>Bioline (London, UK)</td>
<td>TRIsure™, SensiMix™ SYBR No-ROX kit, Pfu DNA polymerase, MyTaq™ DNA polymerases</td>
</tr>
<tr>
<td>BioRad Scientific (Hercules, USA)</td>
<td>Precision Plus Protein™ Dual Colour Standards</td>
</tr>
<tr>
<td>GE Healthcare Life Sciences (Pittsburgh, USA)</td>
<td>Whatman® 3MM Chromatography papers</td>
</tr>
<tr>
<td>Life Technologies (Carlsbad, USA)</td>
<td>PureLink® Genomic DNA Mini Kit, NuPAGE® MOPS SDS Running Buffer(20x), NuPAGE® 4-12% Bis-Tris Gel, RNase ZAP®, GlycoBlue, ERCC RNA Spike-in Mix, Dulbecco's Modified Eagle Medium (DMEM) (4.5g/L D-glucose, +L-Glutamine, +110mg/L Sodium Pyruvate, +Phenol Red), Minimum Essential Medium α (MEMα) (GlutaMAX™, +Phenol Red), Opti-MEM™, TrypLE™ Express, Fetal Bovin Serum (FBS), Penicillin-Streptomycin, Dulbecco's Phosphate-Buffered Saline (DPBS), Silencer® Select siRNAs, Lipofectamine® 2000, Lipofectamine® RNAiMAX, Lipofectamine® LTX, UltraPure® Agarose</td>
</tr>
<tr>
<td>Macherey-Nagel (Düren, Germany)</td>
<td>NucleoBond® Xtra Midi Plus DNA purification kit</td>
</tr>
<tr>
<td>Merck Millipore (Darmstadt, Germany)</td>
<td>Immobilon-P membrane PVDF 0.45μm, Milli-Q Ultrapure water</td>
</tr>
<tr>
<td>Supplier</td>
<td>Products</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>MP Biomedicals (Santa Ana, USA)</td>
<td>LB-Agar Medium</td>
</tr>
<tr>
<td>New England Biolabs (NEB; Ipswich, USA)</td>
<td>Restriction Enzymes (high fidelity version), Calf intestinal alkaline phosphatase (CIP), dNTPs Deoxynucleotide Solution Set, Phusion® High-Fidelity DNA polymerase, Gibson Assembly® Cloning Kit with NEB 5-alpha Competent E.coli, Quick-load™ 1kb/100bp DNA ladder</td>
</tr>
<tr>
<td>Promega (Madison, USA)</td>
<td>Wizard® Plus SV Miniprep DNA Purification System, Wizard® SV Gel and PCR Clean-Up System, Dual-Luciferase® Reporter Assay System, T4 DNA Ligase and Buffer</td>
</tr>
<tr>
<td>Qiagen (Venlo, Netherlands)</td>
<td>QuantiTect® Reverse Transcription Kit</td>
</tr>
<tr>
<td>Sigma-Aldrich (St.Louis, USA)</td>
<td>Z-Leu-Leu-Leu-al (MG132) proteasome inhibitor, RNase A, Triton™-X100, Tween® 20, DAPI</td>
</tr>
<tr>
<td>Takara Bio, Clontech (Japan)</td>
<td>In-Fusion™ HD Cloning Kit with Stellar™ Competent Cells</td>
</tr>
<tr>
<td>Vector Laboratories (Burlingame, USA)</td>
<td>VectaShield® Mounting Medium</td>
</tr>
</tbody>
</table>

**Table 2.2 Suppliers of reagents and kits (continued).**
<table>
<thead>
<tr>
<th>Supplier</th>
<th>Antibodies and probes</th>
<th>Working dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abcam (Cambridge, UK)</td>
<td>Mouse monoclonal [AC-15] anti-β-actin antibody</td>
<td>1/15000</td>
</tr>
<tr>
<td>Biosearch Technologies (Petaluma, USA)</td>
<td>Stellaris® RNA FISH probes for human NEAT1 and NEAT1_v2</td>
<td>1/100</td>
</tr>
<tr>
<td>Jackson Laboratory (Bar Harbor, USA)</td>
<td>FITC-conjugated AffiniPure® Donkey anti-mouse IgG (H+L)</td>
<td>1/250</td>
</tr>
<tr>
<td>LI-COR Biosciences (Lincoln, USA)</td>
<td>Donkey anti-mouse IRDye® 800CW (western blot secondary antibody)</td>
<td>1/10000</td>
</tr>
<tr>
<td>Life Technologies (Carlsbad, USA)</td>
<td>AlexaFluor 647 goat anti-mouse IgG(H+L) F(ab’)&lt;sub&gt;2&lt;/sub&gt; fragment</td>
<td>1/250</td>
</tr>
<tr>
<td>Santa Cruz Biotechnology (Dallas, USA)</td>
<td>Mouse monoclonal anti-p53 (DO-1) IgG&lt;sub&gt;2a&lt;/sub&gt;</td>
<td>1/100</td>
</tr>
<tr>
<td>Sigma-Aldrich (St.Louis, USA)</td>
<td>Mouse monoclonal anti-SC-35 antibody</td>
<td>1/500</td>
</tr>
<tr>
<td>(Souquere et al., 2010)</td>
<td>Mouse monoclonal anti-NONO antibody</td>
<td>1/500</td>
</tr>
</tbody>
</table>
### Table 2.4: Suppliers of instruments

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Instruments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACEA Biosciences (San Diego, USA)</td>
<td>E-plate 16 (xCELLigence system)</td>
</tr>
<tr>
<td>Becton Dickinson Biosciences (BD Biosciences; Franklin Lakes, USA)</td>
<td>FACS Aria™ III Cell Sorter</td>
</tr>
<tr>
<td>BMG Labtech (Offenburg, Germany)</td>
<td>FLUOstar Optima Microplate Reader</td>
</tr>
<tr>
<td>Eppendorf (Hamburg, Germany)</td>
<td>Mastercycler® Nexus Gradient (PCR cycler)</td>
</tr>
<tr>
<td>GE Healthcare Life Sciences (Pittsburgh, USA)</td>
<td>DeltaVision® Elite Imagine System</td>
</tr>
<tr>
<td>LI-COR Biosciences (Lincoln, USA)</td>
<td>Odyssey® Infrared Imaging System</td>
</tr>
<tr>
<td>Life Technologies (Carlsbad, USA)</td>
<td>XCell SureLock® Mini-Cell and XCell II™ Blot Module</td>
</tr>
<tr>
<td>Merck Millipore (Darmstadt, Germany)</td>
<td>Milli-Q Integral Water Purification System</td>
</tr>
<tr>
<td>Nikon (Tokyo, Japan)</td>
<td>Eclipse TS100 Inverted Routine Microscope, Eclipse Ti-E Inverted Microscope System</td>
</tr>
<tr>
<td>Olympus (Tokyo, Japan)</td>
<td>IX71 Inverted Microscope</td>
</tr>
<tr>
<td>Qiagen (Venlo, Netherlands)</td>
<td>Rotor-Gene Q Real-time PCR cycler</td>
</tr>
<tr>
<td>Roche (Basel, Switzerland)</td>
<td>RTCA DP Analyzer (xCELLigence system)</td>
</tr>
<tr>
<td>Thermo Fisher Scientific (Waltham, USA)</td>
<td>NanoDrop 1000 Spectrophotometer</td>
</tr>
</tbody>
</table>
2.3 Oligonucleotides

See Table 2.5 for PCR and sequencing primers.

See Table 2.6 for qRT-PCR primers.

See Table 2.7 for CRISPR gRNA oligonucleotides.
<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'<em>hNEAT1pr</em>-911_nest</td>
<td>A708</td>
<td>GCTATGTCAGATACTGCTTTCCGCAC</td>
</tr>
<tr>
<td>5'<em>hNEAT1pr</em>-2295_nest</td>
<td>A707</td>
<td>AGGAGACTTGGTCTCCTTTTGAGCAGG</td>
</tr>
<tr>
<td>5'<em>hNEAT1pr</em>-3231_nest</td>
<td>A706</td>
<td>TGACCTGGCTTCAGAAAGATGTGG</td>
</tr>
<tr>
<td>3'_hNEAT1pr+387_nest</td>
<td>A709</td>
<td>GCCCTTCTCTGACTCTCCACC</td>
</tr>
<tr>
<td>3'<em>hNEAT1pr</em>-2543 EcoRV</td>
<td>A723</td>
<td>AACCACATGGGATATCTCTCTGCTCTTAAG</td>
</tr>
<tr>
<td>5'_hNEAT1pr_KpnI_01</td>
<td>A657</td>
<td>ACTACT GTTACC GCACAATATCTTTGTTTACATTATTTGC</td>
</tr>
<tr>
<td>5'_hNEAT1pr_KpnI_02</td>
<td>A658</td>
<td>ACTACT GTTACC GAACCTGCGCTCTTAGCAATCTC</td>
</tr>
<tr>
<td>5'<em>hNEATpr</em>-2988_KpnI</td>
<td>A722</td>
<td>ACTACT GTTACC CCAATGTGTTGTTGAAGCTATAG</td>
</tr>
<tr>
<td>5'_hNEAT1pr_KpnI_04</td>
<td>A660</td>
<td>ACTACT GTTACC GTGATGGGGCCACCCGGAC</td>
</tr>
<tr>
<td>5'_hNEAT1pr_Kpn05_Infusion</td>
<td>A777</td>
<td>TTCCCCATCGCTGACCTACAGGTAAAATGCAGAAAGAGAGAG</td>
</tr>
<tr>
<td>3'_hNEAT1pr_NheI_01</td>
<td>A662</td>
<td>ACTACT GCTAGC CACTCTCCTCCCCACAAC</td>
</tr>
<tr>
<td>3'<em>hNEAT1pr</em>-5400_Infusion</td>
<td>A778</td>
<td>GGTGGCCCATCACGGTACC AGCACCAGTTGACTGCTCTTGAAATC</td>
</tr>
<tr>
<td>5'_hNEAT1_p53mutation_v1</td>
<td>A871</td>
<td>CAGGCGCCAGGGAGTAAACCTGGTCTTTACCACATCACCC</td>
</tr>
<tr>
<td>3'_hNEAT1_p53mutation_v1</td>
<td>A872</td>
<td>GGTGGTATGTGTTGAAACCCAGGTCTTACTCCCTGCGCCTG</td>
</tr>
<tr>
<td>RVprimer3</td>
<td>A685</td>
<td>CTAGCAAAAATAGCGTGTCCC</td>
</tr>
<tr>
<td>GLprimer2</td>
<td>A686</td>
<td>CTTATGTTTTTGCGGTCTTCCA</td>
</tr>
<tr>
<td>3'_hNEAT1pr_check 01</td>
<td>A687</td>
<td>TCTCAGAGTATGTTGTG</td>
</tr>
<tr>
<td>5'_hNEAT1pr_SEQ_01</td>
<td>A663</td>
<td>GTAAGCACAGCCCGAGGC</td>
</tr>
<tr>
<td>5'_hNEAT1pr_SEQ_02</td>
<td>A664</td>
<td>GCCGACAGCAATCCCGG</td>
</tr>
<tr>
<td>5'_NEAT1pr_fullarm_net_v1</td>
<td>A932</td>
<td>ATGTTGTAACCCACTAGCTCTGGAGCG</td>
</tr>
<tr>
<td>3'_NEAT1pr_fullarm_net_v1</td>
<td>A933</td>
<td>CGAGCTAATGGTCTCCTGAGACAGACCAGGC</td>
</tr>
<tr>
<td>5'_NEAT1pr_L_arm_v1</td>
<td>A927</td>
<td>ACCGCGATCAGTTGTTGACCTAGCAGCCCGACCCAGG</td>
</tr>
<tr>
<td>3'_NEAT1pr_L_arm_v1</td>
<td>A928</td>
<td>GACCCGTAATGGTACTATGCTCAGCAGCCGTCGTTAGCG</td>
</tr>
<tr>
<td>5'_NEAT1pr_R_arm_v1_Mlu</td>
<td>A958</td>
<td>CAAACTCATCAATGTTCTTAAATGTGCAAGTCCAGCCGGAGTCTAGC</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------------------</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1pr_R_arm_v1_Mlu</td>
<td>A959 AATATTAACGCTTACAATTAAAGTTCCACAAGACCAGGCCTCATCCCCAG</td>
<td></td>
</tr>
<tr>
<td>5'NEAT1pr_R_arm_v2_Mlu</td>
<td>A960 CAAACTCATCAATGTATCTTAACGACAGGGAGGGATGCAGCC</td>
<td></td>
</tr>
<tr>
<td>5'NEAT1m_fullarm_net_v1</td>
<td>A938 CCCTGCTTGGTTTCTGAGCTGCGTCTATTG</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1m_fullarm_net_v1</td>
<td>A939 ACCAGGAGACAGAGGCAGTTTGCAGCC</td>
<td></td>
</tr>
<tr>
<td>5'NEAT1m_L_arm_v1</td>
<td>A934 ACCGCCATGCATTAGTTATGTGTTCTGAGCTGCGTCTATTGATTTG</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1m_L_arm_v1</td>
<td>A935 GACCCGTAATTGATTACTATTGCTATGGAAGGGAATGTGATAGGG</td>
<td></td>
</tr>
<tr>
<td>5'NEAT1m_R_arm_v1_Mlu</td>
<td>A961 CAAAACATCAATGTATCTTTAACTTGGATGATCTTTCTATTTGATAG</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1m_R_arm_v1_Mlu</td>
<td>A962 AATATTAACGCTTACAATTTATCCAGATTCACGTGATCC</td>
<td></td>
</tr>
<tr>
<td>5'pEYFP-C1_AseI_seq_v1</td>
<td>A954 TCACATGTTCTTTCTCG</td>
<td></td>
</tr>
<tr>
<td>3'pEYFP-C1_AseI_seq_v1</td>
<td>A955 GTTATGTAACGGGAACTCC</td>
<td></td>
</tr>
<tr>
<td>5'pEYFP-C1_MluI_seq_v1</td>
<td>A956 CACCTGATTCTAGTTGTG</td>
<td></td>
</tr>
<tr>
<td>3'pEYFP-C1_MluI_seq_v1</td>
<td>A957 CCGATTTCGCGCTATTTG</td>
<td></td>
</tr>
<tr>
<td>5'FP_downstream_seq_v1</td>
<td>A963 CAACCACCTACCTGAGCTACAGTCCAG</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1_prv2_seq_v1</td>
<td>A964 TACTGTCTCCGGCTTACGATAGCC</td>
<td></td>
</tr>
<tr>
<td>3'NEAT1_mv1_seq_v1</td>
<td>A965 AGACACAGTTGCAAAGGACGGCAACC</td>
<td></td>
</tr>
<tr>
<td>5' U6 promoter_seq</td>
<td>A459 CATATGCTTACCGTAAC</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: PCR and sequencing primers (continued).
### Table 2.6: qRT-PCR primers.

<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
<th>Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLuc_F</td>
<td>A878</td>
<td>AGGTTGCTCCCGCTGAAT</td>
<td>Detect firefly luciferase mRNA</td>
</tr>
<tr>
<td>FLuc_R</td>
<td>A879</td>
<td>CATCGTCTTTCCGTGCTCA</td>
<td></td>
</tr>
<tr>
<td>mNEAT1_F</td>
<td>A851</td>
<td>TTGGGACAGTGAGCAGTGAGG</td>
<td>Detect total mouse NEAT1 lncRNA, both v1 and v2 isoforms</td>
</tr>
<tr>
<td>mNEAT1_R</td>
<td>A852</td>
<td>TCAAGTCCACAGCAGACAGCA</td>
<td></td>
</tr>
<tr>
<td>mNEAT1_v2_F</td>
<td>A655</td>
<td>GCTCTGGGACCTTCGTGACTCT</td>
<td>Specifically detect mouse NEAT1_v2 lncRNA</td>
</tr>
<tr>
<td>mNEAT1_v2_R</td>
<td>A656</td>
<td>CTGCCCTGTTGGAAATGTAAGG</td>
<td></td>
</tr>
<tr>
<td>hNEAT1_F</td>
<td>A273</td>
<td>GTGGCTGTGGAGTCCGAAT</td>
<td>Detect total human NEAT1 lncRNA, both v1 and v2 isoforms</td>
</tr>
<tr>
<td>hNEAT1_R</td>
<td>A274</td>
<td>TAAACAAACACGGTCCCATGA</td>
<td></td>
</tr>
<tr>
<td>hNEAT1_v2_F</td>
<td>A351</td>
<td>GTCTTTCCATCCACCTACGTCTATT</td>
<td>Specifically detect human NEAT1_v2 lncRNA</td>
</tr>
<tr>
<td>hNEAT1_v2_R</td>
<td>A352</td>
<td>GTACTCTGTGATGGGTAGTCAAGTCAG</td>
<td></td>
</tr>
<tr>
<td>hGAPDH_F</td>
<td>A320</td>
<td>ATGGGGAAGGTTAAGGTCGG</td>
<td>Detect human GAPDH mRNA</td>
</tr>
<tr>
<td>hGAPDH_R</td>
<td>A321</td>
<td>GGGGTCAATGGATGGCAACAAATA</td>
<td></td>
</tr>
<tr>
<td>U6_F</td>
<td>A432</td>
<td>CTCGCTCAGCAGCACA</td>
<td>Detect both human and mouse U6 mRNA</td>
</tr>
<tr>
<td>U6_R</td>
<td>A433</td>
<td>AACGCTTACGAATTTGCGT</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2.7: CRISPR gRNA oligonucleotides

<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
<th>Sequence</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’gRNA NEAT1pr_v1</td>
<td>A921</td>
<td>CACCGGCTATAAAAGCAAAGTGG</td>
<td>For total NEAT1 knockout</td>
</tr>
<tr>
<td>3’gRNA NEAT1pr_v1</td>
<td>A922</td>
<td>AAACCAACTTTGGCTTTCTATTAGCC</td>
<td></td>
</tr>
<tr>
<td>5’gRNA NEAT1pr_v2</td>
<td>A923</td>
<td>CACCGGTCAGCGGAGGATTCAGGA</td>
<td>For total NEAT1 knockout</td>
</tr>
<tr>
<td>3’gRNA NEAT1pr_v2</td>
<td>A924</td>
<td>AAACCTGCTAATCGGCTGGTGGACC</td>
<td></td>
</tr>
<tr>
<td>5’gRNA NEAT1m_v1</td>
<td>A925</td>
<td>CACCGATGCTCTCTCATAACGACTT</td>
<td>For NEAT1_v2 specific knockout</td>
</tr>
<tr>
<td>3’gRNA NEAT1m_v1</td>
<td>A926</td>
<td>AAACAAAGTGCTTTATAGGAAAGCAATC</td>
<td></td>
</tr>
</tbody>
</table>
2.4 Recipes of buffers and reagents

All buffers, reagents and experiments were prepared using Milli-Q Ultrapure water, unless stated otherwise.

<table>
<thead>
<tr>
<th>Name</th>
<th>Recipe</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEPC water</td>
<td>Add 0.1% DEPC to water, mix and sit at room temperature overnight. Then autoclave.</td>
</tr>
<tr>
<td>DMEM culture media</td>
<td>DMEM, 10% FBS, 50 units/mL Penicillin, and 50 μg/mL Streptomycin.</td>
</tr>
<tr>
<td>Acidic culture media</td>
<td>Culture media adjusted to pH 6.2-6.4 using 1M HCl.</td>
</tr>
<tr>
<td>Serum starved culture media</td>
<td>Cultured media supplemented with 0.1% FBS, 50 units/ml Penicillin and 50μg/ml Streptomycin.</td>
</tr>
<tr>
<td>FACS buffer</td>
<td>DPBS, 2% FBS.</td>
</tr>
<tr>
<td>Phosphate buffered saline 1x (PBS)</td>
<td>NaCl 137mM, KCl 2.7mM, Na₂HPO₄ 10mM, KH₂PO₄ 1.8mM, dissolved in water, adjust pH to 7.2.</td>
</tr>
<tr>
<td>PBS-T (1x)</td>
<td>0.2% Tween-20 in PBS.</td>
</tr>
<tr>
<td>Fixative solution (Immunofluorescence)</td>
<td>4% paraformaldehyde, 20mM NaOH, dissolved in water.</td>
</tr>
<tr>
<td>Fixation buffer (RNA-FISH)</td>
<td>3.7% formaldehyde, dissolved in PBS.</td>
</tr>
<tr>
<td>Hybridization buffer (RNA-FISH)</td>
<td>1g dextran sulphate, 1mL 20X SSC, 1mL deionized formamide, dissolved in 10mL water.</td>
</tr>
<tr>
<td>Wash buffer (RNA-FISH)</td>
<td>5mL 20X SSC, 5ml deionized formamide, dissolved in 50mL water.</td>
</tr>
<tr>
<td>SDS sample loading buffer (1x)</td>
<td>50mM Tris-HCl pH 6.8, 2% SDS, 10% Glycerol, 1% β-mercaptoethanol, 12.5mM EDTA, 0.02% Bromophenol blue, dissolved in water.</td>
</tr>
<tr>
<td>SSC (20x)</td>
<td>3M NaCl, 0.3M Trisodium citrate, dissolved in water, pH adjusted to 7.0.</td>
</tr>
<tr>
<td>TAE buffer (1x)</td>
<td>40mM Tris, 20mM acetic acid, 1mM EDTA, dissolved in water.</td>
</tr>
<tr>
<td>TE buffer</td>
<td>10mM Tris, bring to pH 8.0 with HCl, 1mM EDTA.</td>
</tr>
<tr>
<td>Tris-Glycine buffer (25x)</td>
<td>18.2g Tris base, 90g Glycine in 500ml water. Autoclaved.</td>
</tr>
<tr>
<td>Transfer buffer (1x)</td>
<td>25x Tris-Glycine buffer 20ml, methanol 80ml, mixed in 400ml water.</td>
</tr>
<tr>
<td>Blocking and blotting buffer (1x)</td>
<td>5% skim milk in PBS-T.</td>
</tr>
</tbody>
</table>
2.5 General cloning

2.5.1 Polymerase chain reactions

For amplifying the human NEAT1 promoter, the original human DNA template was obtained from a healthy Australian male, a kind gift from Professor Grant Morahan. Polymerase chain reactions (PCR) were performed using various commercially available DNA polymerases based on individual cases. Conditions and thermocycles were carried out according to the manufacturer’s instructions. The experimental setup conditions corresponding to each DNA polymerase are listed below:

<table>
<thead>
<tr>
<th>Phusion® DNA polymerase (See Table 2.2)</th>
<th>For each reaction of 20μl</th>
<th>Thermocycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA</td>
<td>10-30ng</td>
<td></td>
</tr>
<tr>
<td>5x GC or 5x HF buffer</td>
<td>4μl</td>
<td></td>
</tr>
<tr>
<td>dNTPs (10mM each)</td>
<td>0.4μl</td>
<td></td>
</tr>
<tr>
<td>forward primer (10μM)</td>
<td>0.4μl</td>
<td>98°C</td>
</tr>
<tr>
<td>reverse primer (10μM)</td>
<td>0.4μl</td>
<td>98°C 10 secs</td>
</tr>
<tr>
<td>DMSO (if needed)</td>
<td>1μl</td>
<td>50-68°C 30 secs</td>
</tr>
<tr>
<td>Phusion DNA polymerase</td>
<td>0.2μl</td>
<td>72°C 0.5-5 mins</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 20μl</td>
<td>72°C 5 mins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4°C ∞</td>
</tr>
</tbody>
</table>
**Pfu DNA polymerase (see Table 2.2)**

<table>
<thead>
<tr>
<th></th>
<th>For each reaction of 20μl</th>
<th>Thermocycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA</td>
<td>10-30ng</td>
<td></td>
</tr>
<tr>
<td>10x <em>Pfu</em> buffer</td>
<td>2μl</td>
<td></td>
</tr>
<tr>
<td>dNTPs (10μM each)</td>
<td>0.4μl</td>
<td></td>
</tr>
<tr>
<td>forward primer (10μM)</td>
<td>0.4μl</td>
<td>98°C 3 mins</td>
</tr>
<tr>
<td>reverse primer (10μM)</td>
<td>0.4μl</td>
<td>95°C 30 secs</td>
</tr>
<tr>
<td>DMSO (if needed)</td>
<td>1μl</td>
<td>50-68°C 30 secs</td>
</tr>
<tr>
<td><em>Pfu</em> DNA polymerase</td>
<td>0.1μl</td>
<td>72°C 0.5-5 mins</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 20μl</td>
<td>72°C 10 mins</td>
</tr>
</tbody>
</table>

**MyTaq™ DNA polymerase (see Table 2.2)**

<table>
<thead>
<tr>
<th></th>
<th>For each reaction of 20μl</th>
<th>Thermocycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA</td>
<td>10-30ng</td>
<td></td>
</tr>
<tr>
<td>5x MyTaq buffer</td>
<td>4μl</td>
<td></td>
</tr>
<tr>
<td>forward primer (10μM)</td>
<td>0.4μl</td>
<td>98°C 3 mins</td>
</tr>
<tr>
<td>reverse primer (10μM)</td>
<td>0.4μl</td>
<td>95°C 30 secs</td>
</tr>
<tr>
<td>MyTaq DNA polymerase</td>
<td>0.1μl</td>
<td>50-68°C 30 secs</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 20μl</td>
<td>72°C 0.5-5 mins</td>
</tr>
</tbody>
</table>

If the PCR products were used for downstream cloning, then they were purified using Wizard® SV Gel and PCR Clean-Up System (see Table 2.2) according to the manufacturer’s instructions.
2.5.2 Restriction digestion

Restriction digestion was performed using restriction enzymes (see Table 2.1) according to the manufacturer’s instructions. The reaction consisted of at least 1μg of DNA, 1μl of enzyme in 1x digestion buffer, 1μl of CIP (if required), in a 30μl final volume with water (see Table 2.2). Reactions were incubated at corresponding temperatures for at least 2 hours, and then heat inactivated if possible. Digested DNA fragments were later separated through electrophoresis, and the desired fragment was gel extracted using Wizard® SV Gel and PCR Clean-Up System (see Table 2.2) according to the manufacturer’s instructions.

2.5.3 Ligation reactions

Ligation reactions were performed using T4 DNA ligase (see Table 2.1) according to the manufacturer’s instructions. The reaction consisted of a molar ratio of approximately 5:1 of insert to cut vector, 3 units of T4 DNA ligase in 1x ligase buffer, in a 10μl final volume with water. Ligation reactions were carried out at 16°C overnight.

2.5.4 Gibson and In-Fusion cloning

Gibson and In-Fusion cloning (see Table 2.2) was performed using kits under the instructions provided. In short, the first step was to use primers with overlapping sequences to the target vector of inserts to PCR amplify inserts. Purified insert and vector were assembled directionally in a reaction consisting of a molar ratio of approximately 5:1 of insert to cut vector, 1x of enzyme and buffer pre-mix, in a 10μl final volume. The reaction was mixed well and incubated at 50°C for 15 minutes, and then transformed into designated chemical competent E.coli cells provided according to the manufacturer’s instructions.

2.5.5 Site-directed mutagenesis

Site-directed mutagenesis was performed in the classical workflow. In brief, two primers incorporating the desired mutation were designed to reverse complement each other to enable amplification of the whole vector in a single PCR reaction. The PCR was carried out using Phusion DNA Polymerase (see Method 2.4.1). The resulting
product was then digested with DpnI restriction enzyme to remove the previous vector template. This final product was then transformed into DH5α *E.coli* (see Method 2.5.6).

### 2.5.6 Transformation

A chemical transformation protocol was used to transform competent cells with plasmids. 5μl ligation reaction was added to a 50μl aliquot of thawed *E.coli* DH5α chemical competent cells, and mixed gently. After incubating on ice for 20 minutes, the tube was heat shocked at 42°C for 1 minute, and then was immediately put on ice for another 5 minutes. Lastly, cells were recovered in 1ml of LB media at 37°C for 1 hour, and then plated on LB agar plates with antibiotic selection overnight.

### 2.5.7 Plasmid preparation

For cloning and sequencing grade plasmid preparations, the Wizard® Plus SV Miniprep DNA Purification System was used according to the manufacturer’s instructions (see Table 2.2). For transfection grade plasmid preparation, NucleoBond® Xtra Midi Plus DNA purification kit was used according to the manufacturer’s instructions (see Table 2.2). Both plasmid preparation procedures essentially follow the classical ‘miniprep’ procedure.

### 2.5.8 Nucleic acid quantification

The DNA or RNA concentration, 260:230nm ratio, and 260:280nm ratio were measured using the NanoDrop 1000 Spectrophotometer according to the manufacturer’s instructions (see Table 2.2).

### 2.5.9 Sequencing

DNA and sequencing primers were sent to Macrogen (see Table 2.1) for Sanger sequencing using the standard protocol.

### 2.5.10 Oligonucleotides annealing

When required, such as for the cloning of CRISPR guide RNA sequence, two complementary oligonucleotides were annealed to form a double stranded DNA to enable
ligation into another vector. To anneal the two, oligonucleotides were mixed at 10μM each in TE buffer in a 100μl volume (see Table 2.8). The tube containing the mixture was then placed into a beaker containing higher than 95°C hot water. This beaker was then placed inside a sealed polystyrene box to allow the temperature to decrease at a very slow rate overnight. The resulting annealed oligonucleotides were then used at a further 1:250 dilution for cloning.

### 2.6 Cell culture and treatments

HeLa cells are part of the Fox lab laboratory stock, and, as with many other HeLa lines, are thought to be significantly different to the original ATCC HeLa stock. Neuro2A cells were a kind gift from Professor John Mattick (Garvan Institute, NSW). U2OS cells were a kind gift from Associate Professor Evan Ingley (Harry Perkins Institute, WA). Human Dermal Fibroblasts (hDF) and human Mesenchymal Stem Cells (hMSC) were a kind gift from Associate Professor Stuart Hodgetts (UWA).

HeLa, Neuro2A, U2OS and hDF cells were cultured in DMEM supplemented with 10% FBS, penicillin and streptomycin (see Table 2.2 and 2.8), cultured at 37°C and 5% CO2. hMSCs were cultured in MEMα medium supplemented with GlutaMAX, 10% FBS, penicillin and streptomycin (see Table 2.2 and 2.8). When cells were 90% confluent they were trypsinized using TrypLE Express, and 1/10th of the resultant cell suspension placed into a new flask. All experiments were conducted on cells within 15 passages of thawing a vial of frozen cells. Mycoplasma contamination was checked periodically and not detected throughout the project.

Media containing treatment (serum starvation, acidic pH, or KCl) (see Table 2.8) were put onto the cells that were seeded for 24 hours. Cells were always seeded at 2-2.5x10^5 cells per each well of a 6-well plate, or equivalent concentration in other plate formats.
2.7 Transfection

Cells for transfection were seeded in antibiotic free media at $2.5 \times 10^5$ cells per well of a 6-well plate. The following day, different reagents were used depending on the purpose of the transfection.

For transient plasmid transfection, plasmids were transfected using Lipofectamine LTX with PLUS reagent according to the manufacturer’s instructions (see Table 2.2). Briefly, $1.2 \mu$g of DNA was mixed with $1.25 \mu$l PLUS reagent in $250 \mu$l Opti-MEM serum free media for five minutes at room temperature (see Table 2.2). Meanwhile, $3.2 \mu$l of Lipofectamine LTX was mixed with $250 \mu$l of Opti-MEM. The two mixes were combined to create a final complex and left at room temperature for another 20 minutes before being added to the cells. Next, cells were incubated with normal culture conditions for 4 to 6 hours, then media was removed, cells were washed with DPBS once and then normal culture media added.

Lipofectamine 2000 was used for plasmid transfection in CRISPR genomic engineering experiments ($1.2 \mu$g of DNA in each well of a 6-well plate), and Lipofectamine RNAiMAX was used for all siRNA knockdown experiments ($33 \text{nM}$ of final siRNA concentration was used). Other parts of the protocol were similar to that of Lipofectamine LTX (same amount of DNA as Lipofectamine 2000), and were all carried out according to the manufacturer’s instructions (see Table 2.2).

2.8 Luciferase reporter assay

pGL3-basic and pGL3-control Firefly luciferase plasmids were provided by Promega (Madison, USA). pGL3-basic contains a firefly luciferase gene that is to be controlled by a cloned DNA sequence, whereas pGL3-control is a control plasmid with a constitutively active SV40 promoter driven luciferase gene. pcDNA3.1 based, CMV promoter driven, Renilla control plasmid was a kind gift from Associate Professor Kevin Pfleger (Harry Perkins Institute, WA). Luciferase reporter assays were conducted with
Dual-Luciferase Reporter Assay System (see Table 2.2), using the FLUOstar Optima microplate reader (see Table 2.4), both according to the manufacturer’s instructions. In brief, cells were transfected with the Firefly and Renilla luciferase expressing vectors. 24 hours later the cells were lysed using a lysis buffer provided. 20μl of lysate was then transferred into a 96-well white non-transparent plate. 50μl of substrate for Firefly luciferase was then dispensed by the microplate reader into the well, and a continuous reading of luminescence for 10 seconds was obtained, and averaged for a final reading. Once finished, another 50μl of Renilla luciferase substrate was dispensed into the same well. This substrate would strongly inhibit the luminescence from the Firefly luciferase while allowing the Renilla luciferase to luminesce. A continuous 3 seconds reading was obtained, and then averaged as the final reading.

In the dual luciferase assay system, Renilla luciferase was used as a control for transfection efficiency. To this end, readings from the Firefly luciferase were divided by the reading from the Renilla luciferase for normalization. This ratio was then compared between tested promoter sequences for obtaining a general concept of the transcriptional activity.

\[
\text{Relative promoter activity} = \frac{\text{Firefly intensity (promoter A)}}{\text{Renilla intensity (promoter A)}} = \frac{\text{Firefly intensity (promoter B)}}{\text{Renilla intensity (promoter B)}}
\]

2.9 RNA extraction

Total RNA of samples was extracted using TRIsure™ reagent, according to the manufacturer’s instructions (see Table 2.2). In brief, cell pellets were first lysed using TRIsure reagent. Chloroform was then added and mixed. The mix was then centrifuged at 12,000g for 15 minutes to separate it into three liquid layers. The top aqueous phase was collected and transferred into a new tube containing 1μl of GlycoBlue (see Table 2.2) to help visualize RNA in later steps. The extracted top phase was then added with isopropanol to allow RNA to precipitate. This mixture was then centrifuged at ≥ 12000g for 15 minutes to pellet the precipitated RNA. The supernatant was
removed, and the pellet washed twice with 70% ethanol. As much ethanol was removed as possible, and the RNA pellet was air dried in a 55°C heat block for 10 minutes. Once the pellet was completely dried, DEPC water was added to dissolve the RNA, quantified using a NanoDrop 1000 Spectrophotometer, and then stored at -80°C for further use. All RNA handling was carried out after using RNase ZAP® on gloves to prevent unwanted RNases contaminating samples.

2.10 Quantitative real-time PCR

Extracted RNA was converted into cDNA using the QuantiTect® Reverse Transcription Kit according to the manufacturer’s instructions. In short, 100ng of total RNA was used in this process. Genomic DNA in the RNA samples was removed under a genomic DNA wipe-out step, which consisted of adding RNA into a buffered enzyme mix for incubation at 42°C for 5 minutes. This mix was then treated with reverse transcriptase provided, incubated at 42°C for 15 minutes and then heat inactivated at 95°C for 3 minutes. cDNA was stored at -20°C for further use.

Quantitative real-time PCR was performed using the SensiMix™ SYBR No-ROX kit and Rotor-Gene Q Real-time PCR cycler. Each qPCR reaction consists of:

<table>
<thead>
<tr>
<th></th>
<th>For each reaction of 20μl</th>
<th>Thermocycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template cDNA</td>
<td>1-2μl</td>
<td>95°C 15 mins</td>
</tr>
<tr>
<td>2x SensiMix buffer and enzymes</td>
<td>10μl</td>
<td>95°C 30 secs</td>
</tr>
<tr>
<td>forward primer (10μM)</td>
<td>0.4μl</td>
<td>55°C 30 secs</td>
</tr>
<tr>
<td>reverse primer (10μM)</td>
<td>0.4μl</td>
<td>72°C 30 secs</td>
</tr>
<tr>
<td>water</td>
<td>Up to 20μl</td>
<td>72-85 °C* 20 secs</td>
</tr>
</tbody>
</table>

* An extra fluorescence acquisition step was added to reduce the interference from the detection of unspecific PCR products. The temperature of this step is typically 3°C lower than the melting temperature of the specific product. At this temperature, none of
the unspecific products that have a lower melting temperature than this value will be detected, but the specific product is not affected, thus dramatically increasing the precision and specificity of the fluorescence detected from the desired PCR products. For more information, see: http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/General_Information/qpcr_technical_guide.pdf

Cycle threshold (Ct) values of each product were determined under a constant normalized fluorescence threshold of 0.075. Relative expression of genes was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). U6 mRNA was typically used as a housekeeping gene to control for variation in the cDNA concentration between samples.

### 2.11 Immunofluorescence

Cells grown on coverslips were fixed with fixative solution consisting of 4% paraformaldehyde in PBS, for 10 minutes at room temperature (see Table 2.8). Fixed cells were then washed with PBS thoroughly (see Table 2.8), then incubated in PBS with 1% Triton for 15 minutes and rocked to permeabilize cell membranes. Permeabilized cells were washed with PBS thoroughly and blotted with primary antibody against desired proteins (see Table 2.3) in PBST for 1 hour in a humidified chamber. Blotted coverslips were then washed 3 times with PBST, and then followed by blotting with the secondary antibody (see Table 2.3) in the same manner. After this, cells were washed thoroughly with PBS, and then counterstained with 1:15000 DAPI dissolved in water for 1 minute. Coverslips were then mounted onto glass slides using VectaShield® Mounting Medium to prolong the fluorescence and reduce the effect of photobleaching (see Table 2.2). Fluorescent cellular images were visualized using either an Eclipse Ti-E Inverted Microscope System or DeltaVision® Elite Imagine System (see Table 2.4). Images were taken as Z-stacks at 0.2μm intervals, followed by deconvolution and maximum projection under the default settings by the operation software.
2.12 RNA fluorescent in situ hybridization (RNA-FISH)

Stellaris® RNA FISH probes were used to visualize the localization of lncRNA NEAT1 according to the manufacturer’s instructions. In brief, cells that were seeded on coverslips were washed with PBS and fixed using fixation solution (see Table 2.8). Fixed cells were then permeabilized overnight in 70% ethanol at 4°C. To hybridize cells with RNA probes, coverslips were first washed with the wash buffer (see Table 2.8), and then incubated upside down with the hybridization buffer (see Table 2.8) containing probes on the bottom of a humidified chamber. The incubation took place at 37°C overnight in the dark. Coverslips were then washed with wash buffer thoroughly the next day, and then counterstained with DAPI. These coverslips ready for mounting were stored temporarily in 2X SSC (see Table 2.8), before mounting in VectaShield® Mounting Medium, and then immediately imaged using the DeltaVision® Elite Imaging System. Z-stack images were taken at 0.2μm intervals, deconvolved and maximum intensity projections performed.

To combine RNA-FISH with immunofluorescence staining, primary antibody blotting was employed after the coverslips were washed thoroughly following the overnight hybridization. The rest of the protocol was identical to what has been described in section 2.10.

2.13 Fluorescence-activated cell sorting (FACS)

Fluorescent cells were sorted using the FACSARia™ III Cell Sorter (see Table 2.4). Cells cultured in either a 75cm² or 175 cm² flask were typsinized and resuspended in 10-15mL culture media. Cells were then centrifuged at 1500g for 3 minutes and resuspended in 1-2mL FACS buffer to maintain their viability during the FACS (see Table 2.8). Next, cells were passed through a 35μm nylon mesh filtering cap into tubes so that the maximum amount of singlet cells could be obtained in the cell suspension for FACS. During the FACS, singlet fluorescent cells were analyzed based on their fluorescent intensity on 530nm, and all identifiable ones were collected into a new tube containing
culture media, and seeded back into corresponding culture dishes depending on the number of cells recovered. Cells were cultured normally once they had reached confluency (see 2.5 Cell culture and treatment).

### 2.14 Genomic DNA extraction
The PureLink® Genomic DNA Mini Kit was used to extract genomic DNA from cultured cells according to the manufacturer’s instructions (see Table 2.2). As few as 1x10⁵ cells could be used as starting material for this procedure. The concentration of final genomic DNA solution was quantified by NanoDrop 1000 Spectrophotometer.

### 2.15 Cell viability assay
Cell viability was measured based on the contact area cells had with the surface they were seeded on. Trypsinized cells were seeded into an E-plate 16 (see Table 2.4) at 1x10⁴ per well in 200μL of desired media. Each well was also surrounded by water to reduce evaporation. Before cells started to attach to the bottom of the wells, the plate was quickly loaded onto the RTCA DP Analyzer (xCELLigence system) (see Table 2.4), and the measurement was taken by the software provided by the instrument. Cell index, which indicates the surface contact area, was measured and recorded every 10 minutes for 72 hours routinely. Four technical replicates were included for each cell line assessed in each experiment. Outlier wells which were potentially due to the ‘edge effect’ of the E-plate 16 were removed from further analysis.

### 2.16 Western blotting
Cells were trypsinized, resuspended in PBS, and centrifuged at 3000g for 3 minutes, followed by removal of the supernatant. Protein samples were prepared by adding 1X SDS sample buffer directly onto the cell pellets. Cells were lysed after vigorous mixing and heated in 95°C for 15 minutes. They were then further pipetted up and down
vigorously using 100uL pipette tips to break up any large genomic DNA fragments to increase the fluidity of the sample. For polyacrylamide gel electrophoresis, protein samples and Precision Plus Protein™ Dual Color Standards (see Table 2.2) were loaded into a NuPAGE® 4-12% Bis-Tris Gel (see Table 2.2) which was assembled in XCell SureLock® Mini-Cell module (see Table 2.4), and run at 180V for 1 hour. NuPAGE® MOPS SDS Running Buffer was used to buffer the electrophoresis in this process (see Table 2.2). Next, separated proteins in the gel were transferred onto Immobilon-P membrane PVDF 0.45μm, which was activated by a brief methanol wash followed by washes with water and transfer buffer. The transfer procedure was carried out using the XCell II™ Blot Module (see Table 2.4), and run at 25V for 2 hours inside the transfer buffer (see Table 2.8). The membrane transferred with proteins was then blocked with blocking and blotting buffer (see Table 2.8) at 4°C and rocked overnight to reduce potential unspecific binding of antibodies in later steps. For primary and secondary antibody blotting, antibodies were diluted in blocking and blotting buffer (see Table 2.8) at desired concentrations (see Table 2.3). The blotting solution was then left on the membrane at 4°C and rocked for at least two hours before the membrane was washed for 3 times for 5 minutes per wash, using 1X PBS-T. After the secondary antibodies blotting and washes, the membrane was stored in 1X PBS, and visualized using the Odyssey® Infrared Imaging System according to the manufacturer’s instructions (see Table 2.4).

2.17 CRISPR/Cas9 mediated targeted genomic engineering

CRISPR (Clustered regularly interspaced short palindromic repeats) as sequence-specific guides for S. pyogenes Cas9 endonuclease have been developed into mammalian systems to efficiently create double strand DNA breaks at desired genomic locations. In this project, protocols and materials from Feng Zhang’s laboratory were employed (Cong et al., 2013).

Step 1: Design and construction of repair template for homologous recombination

In order to incorporate a foreign sequence into a genomic locus using a homologous recombination approach, a repair template that consisted of two homologous arms with
the foreign sequence in the middle was designed. Two approximately 500bp homologous arms were designed for each repair template used. In this thesis, the foreign DNA to be inserted into the target genome was a fluorescent protein expression cassette consisting of a CMV promoter, a yellow fluorescent protein gene, and a SV40 terminator. This expression cassette was previously constructed in the laboratory, based on a vector backbone of a pEGFP-C1 vector but with artificial mutations to change the color of the original green fluorescent protein to yellow (Takara Bio, Clontech). The left and right homologous arms were cloned into AseI and MluI sites respectively using Gibson cloning (see Method 2.4.4). The actual repair sequences are supplied in Appendix 3. The plasmids for promoter disruption, as well as NEAT1_2 disruption, were sequenced using A954+A955 (left arm), and A956+A957 (right arm) sequencing primers (see Table 2.5), and prepared using NucleoBond® Xtra Midi Plus DNA purification kit for transfection.

**Step 2: Design and construction of targeting sequence**

To select the double strand break location in the genomic DNA, a short 100bp sequence that covered the desired genomic region was first collected from UCSC genome browser (Method 2.1), and then put into the CRISPR Design Tool: [http://crispr.mit.edu/](http://crispr.mit.edu/) (Hsu et al., 2013) to generate a list of guide sequences that could be used for targeting Cas9 in the CRISPR system. The algorithm analyzed all possible 20nt sequences in both DNA strands that have an immediately adjacent NGG or NAG (PAM sequence) in their 3’ end, and calculated their number of off-target binding and their property (intronic vs. exonic). The design tool gave a score for each sequence based on the combination of two, and sequences with an off-target score greater than 70 were taken into the next round of selection. In this round, guide sequences were first aligned in EMBOSS Needle (see Method 2.1) to the repair template designed in Step 1, and only those with at least two mismatches in the last 6nt in the 3’end were chosen for further consideration. This was to ensure that they would not direct Cas9 to cut the repair template in the transfection afterwards (Hsu et al., 2013). Next, the remaining ones were screened based on the distance between the potential cut sites (3bp to the 5’ end of the PAM sequence) to both nearest borders of the two homologous arms adjacent to the YFP expression cassette in
the repair template. Generally an equal distance to homologous arms on both sides was desired, and this distance was controlled within 50bp, or as close as possible.

As instructed by the protocol of cloning guide sequences into the desired plasmid (see: http://www.genome-engineering.org/crispr/wp-content/uploads/2014/05/CRISPR-Reagent-Description-Rev20140509.pdf), which was provided by Zhang’s laboratory, guide sequences had BbsI restriction site overhangs incorporated onto each end to facilitate their cloning, as well as a ‘G’ in the 5’ end of the start of the guide sequence for its transcription. These single strand DNA oligonucleotides were then synthesized by Integrative DNA Technologies (see Table 2.1, 2.7) using a standard desalting procedure. Oligos were then annealed into double strand DNA before being ligated into the designated BbsI restriction site into the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (see Method 2.4.2, .2.4.3, 2.4.9), which was a gift from Feng Zhang (Addgene plasmid # 42230) (Cong et al., 2013). The ligated product was then transformed into DH5α E.coli (see Method 2.4.5) and sequenced using the A459 sequencing primer (see Table 2.5) with the services provided by Macrogen Inc. Colonies with the correct sequence were then cultured, and the plasmids were extracted using NucleoBond® Xtra Midi Plus DNA purification kit for transfection.

**Step 3: Transfection**

Plasmid containing guide sequence and Cas9, and plasmid containing the repair sequence were co-transfected using Lipofectamine 2000 in a 1:1 ratio into the cells. See Method 2.6 for more details.

**Step 4: Selection of homozygous engineered cells**

Transfected cells were cultured in normal media until they reached at least 85% confluency in a 175cm² culture flask, and then underwent FACS for enrichment of fluorescent cells (see Method 2.12). Sorted fluorescent cells were cultured until confluent, and then FACS was repeated. After 3 repeats, or when the overall percentage of fluorescent cells had reached at least 80%, single fluorescent cells were FACS sorted, seeded into 96-well plates and cultured in normal conditions. Fluorescent colonies
which had reached confluency in the 96-well plates were trypsinized and transferred into 12-well plates. After reaching confluency, the genomic DNA of each fluorescent colony was extracted (see Method 2.13). PCR using MyTaq DNA polymerase (see Method 2.4.1) was conducted to verify whether the fluorescence carried by the cells was due to either a homo- or heterozygous genomic engineering event, or only due to random insertion of the repair template into the genome. Primer sets A963-A964, A708-A709, A963-A965, and A938-A939 were used in this process (see Table 2.5). qPCR and RNA-FISH were later employed to check the efficiency and effect of the NEAT1 knockout (see section 2.8, 2.9, and 2.11). qPCR primer sets can be found in Table 2.6.

2.18 Whole transcriptomic RNA sequencing and differential expression analysis

Total RNA was prepared from cultured cells using TRIsure (Bioline). Ribosomal RNA was depleted in all samples before library preparation and sequencing, which was carried out by the Australian Genome Research Facility (AGRF). Library was constructed using Illumina TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero- Gold Kit for the removal of ribosomal RNA. The whole transcriptome (ie. Not just limited to poly(A)+ RNA) was sequenced on the Illumina HiSeq 2000 platform to generate a minimum of 20 million 50bp single-end reads for each sample, and all samples were sequenced in a single lane to prevent batch-effects. Sequenced read files were uploaded onto the Galaxy platform (http://galaxyproject.org), converted to FASTQ format by FASTQ Groomer (Blankenberg et al., 2010), and then mapped to the human Hg19 reference genome using Tophat2 (version 0.6) with the default settings (Kim et al., 2013). Read alignments were processed by the htseq-count python script using the ‘union’ and ‘reverse stranded’ mode for gene level read counting (Anders et al., 2015). For differential gene expression analysis, non-expressed genes were filtered out by setting a requirement of more than 10 reads in at least three samples for each gene. Next, the gene expression counts were normalized using the RUVg-empirical method (Risso et al., 2014). The top 5000 most consistently differentially expressed genes in each DE
analysis were used to calculate normalization factors of all genes. RLE (relative log expression) and PCA (principle component analysis) of DE analyses were generated by R for testing statistical features. A list of differentially expressed genes was computed by EdgeR using a cut-off of FDR (False Discovery Rate) ≤ 0.05. DE genes were subsequently imported for Ingenuity pathway analysis (Ingenuity Systems, www.ingenuity.com) using standard settings for predicting potential pathways and functions being affected by the NEAT1 KO. Data exploration on relative log expression, principle component analysis, and mean-difference plots etc. were generated by R and the EDASeq R package.

2.19 Gene ontology and network analysis

Ingenuity Pathway Analysis (IPA)
Lists of differentially expressed genes were uploaded into the IPA. Gene ID, log2 fold changes, and FDR values were included in these gene lists. Default stringency settings were used for all analyses, which were limited to experimentally observed data in the human database.

DAVID gene ontology analysis
Gene lists were uploaded and analyzed using the whole human genome as background. Gene ontology, pathways, and tissue expression data were collected all using default statistical cut-offs, in which Benjamini FDR ≤ 0.1 is considered as statistically significant enrichment.

2.20 Statistics
Student t-tests were performed on replicated experiments. P≤0.05 was considered statistically significant.
Chapter 3

The long non-coding RNA NEAT1 is induced by acidosis and influences the expression of many genes with diverse cellular functions through both paraspeckle-dependent and independent actions

(Manuscript to be submitted)
3.1 Abstract

The long non-coding RNA NEAT1 (nuclear paraspeckle assembly transcript 1) is an abundant mammalian RNA that is frequently induced in stressful conditions or diseases. The main function of NEAT1 is in forming subnuclear bodies called ‘paraspeckles’ that regulate gene expression by the sequestration of specific nuclear proteins and RNAs. The NEAT1 gene locus produces two different RNAs, yet only the longer, NEAT1_2 isoform is required for forming paraspeckles, thus that the biological significance of the shorter, and more abundant NEAT1_1 isoform is not known. Here we demonstrate that both isoforms of NEAT1 are significantly upregulated by acidosis stress in U2OS osteosarcoma cells, and show that the transcriptional activation of the NEAT1 promoter by p53 is a key factor in this process. To better characterize the role of NEAT1, and to determine a potential role for NEAT1_1, we used CRISPR-cas9 genome editing to create the first human cancer cell lines with permanent reduction of either both NEAT1 isoforms, or, just the longer NEAT1_2 isoform, leaving NEAT1_1 intact. As expected, cells devoid of NEAT1_2 did not contain paraspeckles. In the NEAT1_2^−/− cells, the NEAT1_1 is redistributed from paraspeckles to small ‘microspeckles’ throughout the nucleus that may be the sites of a paraspeckle-independent activity of NEAT1_1. Transcriptomic analysis of the cell lines revealed many thousands of genes differentially expressed, with enrichment for cell migration and movement gene ontologies, mirrored in phenotypic observations of the total NEAT1 knockout cells showing subtle impaired migration, but significantly increased sensitivity to acidosis stress. Careful comparison of differentially expressed genes in total NEAT1^−/− cells compared to NEAT1_2^−/− cells revealed subsets of genes regulated in a reciprocal manner by the NEAT1 isoforms. These data support the model in which NEAT1_1 carrying out paraspeckle-independent functions that are negatively regulated by NEAT1_2/paraspeckles, likely via sequestration of NEAT1_1. This is yet another example of a complex gene regulatory mechanism carried out by a long noncoding RNA that likely impacts mammalian physiology in a number of contexts, particularly in cancer.
3.2 Introduction

Large scale transcriptomic and epigenetic analyses have systematically demonstrated pervasive transcription and regulatory potential for a considerable amount of our genome (ENCODE, 2012). In many cases this regulatory potential takes the form of various noncoding RNAs, amongst which long non-coding RNAs (lncRNA) are one of the least understood and most diverse groups. LncRNAs are nevertheless emerging as critical functional regulators in the molecular biology of the cell and are often misregulated in disease (reviewed in Esteller, 2011; Wapinski and Chang, 2011). Relative to the total number of lncRNAs, there are still only a subset functionally characterized, and even in this set of lncRNAs, many questions remain, particularly relating to their mechanism of action in controlling gene expression and their physiological relevance (reviewed in Morris and Mattick, 2014).

The majority of lncRNAs are expressed at low levels, with highly tissue-specific and even developmental stage specific expression (reviewed in Mattick, 2011; Mercer et al., 2009; Mercer et al., 2008). In contrast, a few well-characterized lncRNAs appear to be ubiquitous and abundant, suggesting a more generic biological role. One such lncRNA is the mammalian-specific NEAT1 (nuclear paraspeckle abundant transcript 1). Importantly, NEAT1 is one of the few lncRNA known to have a robust phenotype when knocked out in mice. Mice lacking NEAT1 are born viable (Nakagawa et al., 2011), but female mice have defects in corpus luteum formation leading to lack of pregnancy and significantly reduced fertility, as well as defects in mammary gland epithelial cell proliferation leading to reduced lactation and pup malnutrition (Nakagawa et al., 2014; Standaert et al., 2014).

NEAT1 has the potential to be expressed in many tissues and cell lines, with the exception of embryonic stem cells and induced pluripotent stem cells (Chen and Carmichael, 2009). NEAT1 is also induced in a variety of disease states and is often associated with cellular stress, including hypoxia and cancers, viral infection, drug abuse, intra-uterine growth retardation and neurodegeneration (Chakravarty et al., 2014; Choudhry et al., 2014;
The induction of NEAT1 is driven by transcriptional upregulation at the NEAT1 promoter, with oestrogen receptor α (ERα) and hypoxia-inducible factor II (HIF-2α) as the only factors thus far described as being involved (Chakravarty et al., 2014; Choudhry et al., 2014).

This growing body of data suggesting physiological importance for NEAT1 has lent greater weight to molecular studies into this lncRNA. NEAT1 is now known as the essential RNA backbone for subnuclear bodies termed paraspeckles (Clemson et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009). Paraspeckles are irregularly shaped mammalian nuclear bodies found in close proximity to nuclear speckles in many cell types (Fox et al., 2002; Hutchinson et al., 2007). In addition to NEAT1, there are approximately forty different nuclear RNA- and DNA-binding proteins found within paraspeckles, and these act in concert with NEAT1 to establish the paraspeckle (Dettwiler et al., 2004; Fox et al., 2002; Hata et al., 2008; Naganuma et al., 2012; Prasanth et al., 2005). Once formed, paraspeckles regulate gene expression by two mechanisms. Firstly, through the nuclear retention of specifically edited RNA species, thus modulating their translation (Chen and Carmichael, 2009; Chen et al., 2008; Choudhry et al., 2014; Prasanth et al., 2005). Secondly, through the subnuclear sequestration of paraspeckle proteins, thus changing the availability of those proteins in the nuclear pool (Hirose et al., 2014; Imamura et al., 2014).

NEAT1 function is thus intimately linked to paraspeckle formation, but many aspect of this lncRNA are not understood. NEAT1 is transcribed as two major isoforms that overlap completely at the 5’ end of the gene: the shorter, polyadenylated NEAT1_1 (also called MENepsilon, 3.7kb in human) and the longer NEAT1_2 isoform (also called MENbeta, 23kb in human) (Sasaki et al., 2009; Sunwoo et al., 2009). Both NEAT1 isoforms can be found in paraspeckles, however, only the longer NEAT1_2 isoform is essential for the formation of paraspeckles (Mao et al., 2011; Naganuma et al., 2012). This is despite NEAT1_2 having lower abundance, lower levels of overall sequence conservation, and less wide-spread expression than NEAT1_1 (Sasaki et al., 2009;
In contrast to the role of NEAT1_2 in forming paraspeckles, little is known about the role of NEAT1_1 in cells. NEAT1_1 has similar properties to some paraspeckle proteins: it is found in paraspeckles, and yet does not appear to play any part in paraspeckle formation or function. In addition, NEAT1_1 is the more abundant isoform and can also be present in cells devoid of paraspeckles that do not express NEAT1_2. Interestingly, we and others recently found a pro-proliferative and oncogenic role for NEAT1_1, as well as unexpected binding of NEAT1_1 with active histone marks, and localization to the transcriptional start sites and termination sites of active genes (Chakravarty et al., 2014; West et al., 2014). This raises the possibility that NEAT1_1 may have paraspeckle-independent roles, and that paraspeckles might also sequester the NEAT1_1 transcript and regulate possible paraspeckle-independent functions of this isoform.

Here, we first established a pathophysiological stressful condition that can induce NEAT1 and paraspeckles, and then identified p53 as a key transcription factor in this induction. We next created cell lines that cannot produce NEAT1 using CRISPR engineering, and observed reduced viability with stress. We also used CRISPR to create U2OS cells that specifically cannot produce NEAT1_2 and paraspeckles, but still have abundant NEAT1_1. A transcriptomic comparison between the wild-type and these two different NEAT1 knockout cell lines was conducted to help us understand how gene regulation could be affected by each of the two NEAT1 isoforms. We found that NEAT1 target genes largely exist in 2 classes: either those solely regulated by NEAT1_2/paraspeckles, or those regulated in an opposite manner by the two NEAT1 isoforms.
3.3 Results

3.3.1 Acidosis induces NEAT1 expression and formation of paraspeckles

In order to probe the role of NEAT1 and paraspeckles under stress conditions, we sought cell culture models and stresses in which NEAT1 levels were increased and paraspeckles were induced. U2OS osteosarcoma cell line, which is highly proliferative and is a frequently used cancer model for studying nuclear architecture, was chosen for this study. Using the U2OS cells, we discovered that a slightly acidic culture media dramatically increased NEAT1 RNA levels and the prevalence of paraspeckles (Figure 3.1). We found that levels of both NEAT1_1 and NEAT1_2 increased significantly with acidosis (using HCl to bring the pH of the medium to 6.2-6.4). The most dramatic increase of approximately 20-fold was observed with the longer NEAT1_2 isoform at 48h acidosis treatment, with total NEAT1 levels (as measured using primers that detect both NEAT1_1 and NEAT1_2) increasing 4-fold, on average (Figure 3.1A,B). Reflecting how naturally dynamic NEAT1 levels are in culture, a significant, but much more subtle, increase in NEAT1 levels was also observed with KCl incubation, here used as a control for chloride ion levels (Figure 3.1A,B). We and others have shown that increased NEAT1_2 levels correspond to enlarged and more abundant paraspeckles (Hirose et al., 2014; Imamura et al., 2014), and we confirmed that observation for acidosis in U2OS by immunofluorescence (IF) of NONO, an essential paraspeckle protein, coupled with RNA-FISH against NEAT1_2, revealing an increase in the prevalence of paraspeckles marked by the co-localization of both signals (Figure 3.1C). Taken together, these data indicate that acidosis can induce NEAT1, leading to the formation of abundant paraspeckles.
Figure 3.1: Acidosis induces NEAT1 expression and formation of paraspeckles in U2OS cells. RT-qPCR analysis of (A) total NEAT1 RNA levels and (B) NEAT1_2 levels under three different culture conditions: control media, KCl (media with extra KCl), and acidosis (media with pH adjusted to 6.2-6.4 by adding extra HCl), with a 48h incubation. RT-qPCR results were normalized to both GAPDH and U6 levels, and were taken from three independent replicates. (***P<0.005, Student’s t-test) (C) Immunofluorescence of the paraspeckle protein NONO (red), RNA in situ hybridization with probes against NEAT1_2 (green), and DAPI counterstaining (blue), showing dramatically increased NEAT1_2 fluorescence and prevalence of paraspeckles in acidosis-treated U2OS cells. Paraspeckles are indicated by the colocalization of NONO and NEAT1_2. Scar bar = 10μm
3.3.2 p53 is involved in the induction of NEAT1 in acidosis

We reasoned that identification of the key transcription factors controlling NEAT1 expression could help us link NEAT1 to known signaling pathways. However, the transcriptional control of NEAT1 is likely very complex. Harnessing ENCODE data on transcription factor binding sites we found over 80 different transcription factors (TFs) reported to bind to the promoter region of the human *NEAT1* gene. Some influential factors that are highly connected with other TFs also have multiple binding sites across a 6.4kb region upstream of the transcription start site of NEAT1 (Figure 3.2A, Supplementary Table S3.1 – see Top Level TFs). Interestingly, several genome-wide analyses have identified p53 binding to the NEAT1 promoter using chromatin immunoprecipitation (Botcheva et al., 2011; Smeenk et al., 2008). Given the significance of p53 as a ‘master’ transcription factor likely to have a strong influence on the expression of target genes (Gerstein et al., 2012), as well as its role in stress response pathways (reviewed in Levav-Cohen et al., 2014), we investigated the role of p53 in NEAT1 activation.

We first cloned approximately 3.0kb of the human *NEAT1* promoter and inserted it upstream of a luciferase reporter gene. This section contains four major regulatory DNA elements saturated with transcription factor binding sites identified by ENCODE ChIP-seq (Figure 3.2A). The p53 binding site that was identified in previous literatures (Botcheva et al., 2011; Smeenk et al., 2008) was then located, and mutated in a way that can disrupt any potential binding of p53 (Figure 3.2A). Both wild-type and mutant vectors were individually transfected into U2OS cells, followed by control and acidosis treatment and analysis of relative luciferase mRNA levels (Figure 3.2B). Luciferase mRNA, instead of protein activity, was chosen as the measurement to reduce potential variations on luciferase translation under the influence of acidic stress. We found that acidosis resulted in upregulation of the reporter in both cases, however when the p53 binding site was mutated this increase was significantly diminished (Figure 3.2B,C). The induction of endogenous NEAT1 by p53 was next investigated, after confirming that p53 protein levels are indeed increased with acidosis in U2OS cells, and that siRNA knockdown of p53 is effective (Figure 3.2D). When NEAT1 levels were examined...
under these conditions, we observed that siRNA knockdown of p53 reduced the induction of endogenous NEAT1 with acidosis by approximately 50% (Figure 3.2E,F). Taken together, these data confirmed that p53 directly activates the transcription of NEAT1 with acidosis, although it is likely acting in concert with a multitude of transcription factors to activate NEAT1 in this context.
**Figure 3.2:** p53 is involved in the transcriptional upregulation of NEAT1 following the acidosis of U2OS cells. p53 is involved in the transcriptional upregulation of NEAT1 following acidosis treatment of U2OS cells. (A) ChIP-seq tracks (ENCOD, UCSC genome browser [www.genome.ucsc.edu](http://www.genome.ucsc.edu)) of 161 transcription factors in the promoter region of the human NEAT1 gene. 3kb of the NEAT1 promoter was
cloned into a luciferase reporter vector and a potential p53 TFBS mutated as indicated. (B) A slight, but
significant, reduction of NEAT1 induction was detected by RT-qPCR analysis of luciferase mRNA
controlled by the NEAT1 promoter with or without the p53 TFBS mutation. Vectors were transfected
separately into U2OS cells, followed by 68 hours of acidosis until the total RNA was extracted. Luciferase
under the control of SV40 promoter was used as normalization standard. The graph presented is a
representative experiments of three biological replicates. (C) The average fold change of the induction of
luciferase mRNA controlled by the NEAT1 promoter with or without the mutated p53 TFBS. The graph
was generated using the average of three biological replicates. (D) Western blotting of p53 confirms the
accumulation of p53 protein with acidosis treatment in U2OS cells, and that the siRNA against
p53 effectively inhibited this accumulation. (E) RT-qPCR showing that the total NEAT1 levels and (F)
NEAT1_2 levels both increased in acidosis, but to a lesser extent when p53 was knocked down. All RT-
qPCR results were normalized to U6 levels, and were taken from three independent replicates. (*P<0.05,
***P<0.005, Student’s t-test)
3.3.3 CRISPR-mediated generation of NEAT1 knockout U2OS cells

To better understand the role of NEAT1 in gene regulation, we generated the first human NEAT1 knockout cell lines. The CRISPR-Cas9 system was used to generate U2OS clonal cell lines with more than 98% permanent reduction in the expression of all NEAT1 isoforms (Figure 3.3A,B). The strategy involved knock-in of a YFP (yellow fluorescent protein) expression cassette into the genomic position between the TATA box and the transcriptional start site (TSS) of NEAT1, to delete the TATA box and thus disrupt initiation of NEAT1 transcription (Figure 3.3A). After three rounds of FACS selection, individual fluorescent cells were seeded into 96-well plates, clonal lines generated and screened for negative, heterozygous or homozygous gene editing. The resulting homozygous U2OS NEAT1<sup>-/-</sup> cell lines had an almost 100-fold reduction in both isoforms of NEAT1 RNA, albeit with some trace levels of NEAT1 remaining (Figure 3.3B). As expected, paraspeckles were absent in these lines, marked by the loss of co-localization of the 5’ end of NEAT1 (green, detecting both the NEAT1_1 and NEAT1_2 isoforms, Figure 3.3C) and NONO (red, Figure 3.3C).

Given that NEAT1_1 is much more abundant than NEAT1_2, yet is dispensable for paraspeckle formation, and also recent evidence of an oncogenic role for NEAT1_1 (Chakravarty et al., 2014), cell lines were created that could separate the two NEAT1 isoforms functionally. We used the same YFP knock-in strategy, but instead targeted the genomic region closely downstream of the transcription termination site (TTS) of the NEAT1_1 isoform (a polyadenylation signal), to selectively disrupt NEAT1_2, leaving NEAT1_1 intact (Figure 3.3D). We reasoned that the TTS in the YFP cassette would lead to a significant reduction of the transcription of NEAT1_2, and that the insertion of the cassette sequence would alter the NEAT1_2 primary sequence/structure in the event of any transcription reading through the TSS. This strategy was successful in achieving approximately 70% reduction of NEAT1_2 levels in homozygous NEAT1_2<sup>-/-</sup> cell lines (Figure 3.3E). NEAT1_1 levels were intact, and even increased modestly in some clones (Figure 3.3E). As expected, the homozygous NEAT1_2<sup>-/-</sup> cell lines were indeed devoid of paraspeckles both by definition and by the diffused localization of NONO (Figure 3.3F, using FISH probes targeted to the NEAT1_2 isoform). Thus, even with
residual detectable NEAT1_2, this is a functional ‘knock out’ of paraspeckle formation, rather than a ‘knock down’. It is thus also reasonable to assume the same functional ‘knock out’ is also achieved for NEAT1−/− cells.

Examining the localization of the NEAT1_1 transcript in NEAT1_2−/− cells yielded an interesting observation: using FISH probes targeted to NEAT1_1 we observed nuclear puncta that are smaller and more uniformly distributed throughout the nucleus than typical paraspeckles, with an absence of NONO co-localization (Figure 3.3G). In contrast, these NEAT1_1 ‘microspeckles’ associate closely with the nuclear speckle protein SC35 (Figure 3.3H). In fact, a previous study has already showed sign that NEAT1 may localize in the peripheral of SC35 domains (Hutchinson et al., 2007). These puncta may represent the sites for a paraspeckle-independent role for NEAT1_1. Thus we have successfully generated the first human stable total NEAT1−/− cell lines, and have separated the two NEAT1 isoforms using CRISPR-Cas9 genome engineering technology. Both of these cell lines display a significant reduction in their respective targeted NEAT1 isoforms, and do not have any paraspeckles.
Figure 3.3: CRISPR-mediated genome engineering is successful in generating U2OS cell lines with ~100-fold permanent depletion of NEAT1 RNA, or isoform-specific depletion in NEAT1_2 RNA. (A) A schematic of the total NEAT1 knockout strategy (see below for higher magnification). Two CRISPR-Cas9 cuts (cut v1 and cut v2, but only one is shown) were designed to target Cas9 to cut the DNA close to the transcription start site (TSS) of NEAT1. Cuts were to be repaired by homologous recombination resulting in the insertion of a YFP expression cassette. The TATA box (orange) of the NEAT1 promoter was removed during this process. Two repair templates were constructed separately for each CRISPR-Cas9 cut, and transfected individually with the corresponding vector carrying the gRNA and CRISPR-Cas9. (B) NEAT1 reduction efficiency as measured by qRT-PCR. Three homozygous clonal cell lines are used for illustration purposes (col.17 and 71 made with guide v1, and col.24 from guide v2) showing at least a 98% reduction in both total and NEAT1_2 transcripts. (C) Immunofluorescence of NONO (red) and RNA-FISH (green) with probes targeted to the 5’ end of NEAT1 showed complete loss of NEAT1 RNA and paraspeckle signal as a result of total NEAT1 depletion by CRISPR. (D) A schematic of the NEAT1_2 knockout strategy. One CRISPR-Cas9 cut was employed to cut the genomic DNA closely downstream of the transcription termination site of NEAT1_1 (see below for higher magnification). The homologous recombination would result in insertion of the YFP expression cassette. (E) NEAT1_2 knockout efficiency as measured by RT-qPCR. Both homozygous colonies for illustration purposes (col.44, and 85) have demonstrated at least 70% reduction in NEAT1_2 transcripts, yet total NEAT1 levels either had no change or were slightly increased. (F) Immunofluorescence of NONO (red) and RNA-FISH (green) with probes for NEAT1_2 showed complete loss of NEAT1_2 RNA and paraspeckle signal as a result of NEAT1_2 specific depletion by CRISPR. (G) In NEAT1_2 knockout cells, NONO immunofluorescence (red) and NEAT1 FISH with probes against the 5’ end of NEAT1 showed that NEAT1_1 no longer overlaps with NONO. (H) In NEAT1_2 knockout cells, SC35 immuno-fluorescence (red) and NEAT1 FISH with probes against the 5’ end of NEAT1 showed co-localization. qRT-PCR results were normalized to U6 levels. (***P<0.005, Student’s t-test) All scale bars = 10μm.
A

HR repair template

Left arm  CMV + YFP + SV40(T)  Right arm

NEAT1 promoter  TATA

CRISPR-Cas9 cut, v1/v2

NEAT1_1 transcript
NEAT1_2 transcript

both NEAT1 isoforms were abolished

Total NEAT1 KO
(total NEAT1^−/− cells)

D

HR repair template

Left arm  CMV + YFP + SV40(T)  Right arm

NEAT1 promoter  TATA

CRISPR-Cas9 cut

NEAT1_1 transcript
NEAT1_2 transcript

only NEAT1_2 was abolished, NEAT1_1 was still intact and transcribed

NEAT1_2 KO
(NEAT1_2^−/− cells)
3.3.4 NEAT1 alters expression of genes controlling cell movement and cell death

To determine the molecular consequences of NEAT1 loss, RNA sequencing and differential expression (DE) analysis was performed for U2OS wild-type cells (three individually passaged populations), total NEAT1\(^{−/−}\) cell lines (2 clonal lines made with one guide RNA -gRNA, and one line made with a different gRNA), and NEAT1\(_{2}^{−/−}\) cell lines (three clonal lines made with the same gRNA). Reads were normalized using the RUVg method (Risso et al., 2014), and the effect of normalization on each type of DE analysis was assessed using relative log expression plots and principle component analysis plots (Supplementary Figure S3.2). A large numbers of DE genes were revealed in both types of NEAT1 depleted cell line compared to WT, indicating that loss of both NEAT1\(_{1}\) and NEAT1\(_{2}\) (total NEAT1\(^{−/−}\) cells, 947 DE genes at FDR\(≤0.05\) and Log2 fold change $≤ -1$ and $≥1$), or only loss of NEAT1\(_{2}\)/paraspeckles (NEAT1\(_{2}^{−/−}\) cells, 2568 DE genes at FDR\(≤0.05\) and Log2 fold change $≤ -1$ and $≥1$) results in widespread alteration of gene expression (Figure 3.4A,B) (Supplementary Table S3.2-S3.3). Interestingly, the two types of NEAT1-depleted cell lines had significantly altered gene expression profiles, suggesting different roles for paraspeckles and NEAT1\(_{1}\) (807 DE genes at FDR\(≤0.05\) and Log2 fold change $≤ -1$ and $≥1$) (Figure 3.4C) (Supplementary Table S3.4).

Analyzing the lists of differentially expressed genes revealed the gene ontology terms of ‘cellular movement’ and ‘cellular growth and proliferation’ in all three DE comparisons ($p$-values $≤1.14E-02$). The gene ontology term ‘cell death and survival’ was enriched in the NEAT1\(_{2}^{−/−}\) vs. WT comparison and the total NEAT1\(^{−/−}\) vs. NEAT1\(_{2}^{−/−}\) analyses ($p$-values $≤ 1.23E-02$, Supplementary Table S3.2-S3.4). To generate a prediction of the level of perturbation of these cellular processes, activation z-scores were generated, followed by manual prediction based on the sign and number of each z-score (Summarized in Figure 3.4, see details in Supplementary Table S3.5). These z-scores predicted that total NEAT1\(^{−/−}\) cells may have reduced cellular movement compared to WT, whilst NEAT1\(_{2}^{−/−}\) cells may display greater cell movement; and that both types of NEAT1-depleted cells could exhibit greater activation propensity for cell death than WT cells. To test these predictions, we measured cell proliferation and migration for total NEAT1\(^{−/−}\) cells and WT cells with the xCelligence system that measures electrical
impedance across the surface of a culture well to quantify cell growth (Figure 3.5A,B). Both WT and NEAT1\(-/-\) cell lines proliferated in a similar fashion (Figure 3.5A), but the total NEAT1\(-/-\) cells had a subtle, compromised migration phenotype (Figure 3.5B). We next tested the viability of the lines under acidosis conditions, and observed a significantly increased sensitivity to the acidic media for the total NEAT1\(-/-\) cell lines compared to WT lines (Figure 3.5C), thus supporting the pathway analysis predictions. Taken together, these data indicate widespread roles for NEAT1 in altering expression of genes involved in cell viability and movement, in line with new roles for NEAT1 in cancer progression. Interestingly, these pathways appear to be regulated by NEAT1 even under unstressed conditions, with the effect on cell viability only becoming apparent under stress conditions.
Figure 3.4: RNA-seq and differential gene expression analysis of NEAT1-depleted cell lines predicts alterations in key phenotypic features. Mean-difference plots (MD-plots) of RNA-seq DE analysis (A) NEAT1_2⁻ cells compared with wild-type cells. (B) Total NEAT1⁻ cells compared with wild-type cells. (C) Total NEAT1⁻ cells compared with NEAT1_2⁻ cells. The red dots represent DE genes that have
FDR (false discovery rate) ≤ 0.05. Blue lines represent a logFC (log2 based fold change) cut-off of 1. LogCPM represents log2 based expression level in ‘counts per million’. ‘NEAT1’ labelled dots represent primarily the NEAT1_1 isoform, but due to the overlap in sequence, cannot exclude NEAT1_2. NEAT1_2 represents only the NEAT1_2 isoform. Bioinformatic prediction was performed with Ingenuity Pathway Analysis available on: http://www.ingenuity.com/, and more details of the prediction can be found in Supplementary Table S3.2-S3.5.
Figure 3.5: Total NEAT1−/− cells have impaired response to stress, and subtly reduced migration ability, but similar cellular proliferation rates to wild-type cells. The cell-surface contact area (cell index) was measured by the xCELLigence system for (A) proliferation analysis under normal culture conditions, (B) migration of cells to 10% FCS containing media away from serum-free media, and
(C) viability of cells when seeded under acidosis treatment. Graphs are representative data of three biological replicates, plotted as mean of three technical replicates ± SD.
3.3.5 NEAT1_1 and NEAT1_2/paraspeckles co-regulate specific gene sets

The three DE gene analyses also allowed us to determine if the same genes might be affected in different ways in NEAT1^−/− cells compared to NEAT1_2^−/− cells, thereby revealing some regulatory interplay between the two isoforms of NEAT1. We first determined how many genes were solely regulated by NEAT1_1 alone, with no regulatory influence from NEAT1_2/paraspeckles. Genes in this category were unchanged in NEAT1_2^−/− cells compared to WT, but were then subsequently altered when NEAT1_1 was downregulated in total NEAT1^−/− cells (Figure 3.6A). A surprisingly small number of transcripts were found that fit this description (11 genes inhibited by NEAT1_1, and 17 genes activated by NEAT1_1, yet both unaffected by NEAT1_2, Figure 3.6A, Supplementary Table S3.6). This suggested that NEAT1 target genes are generally sensitive to either NEAT1_2/paraspeckles, or, are affected by an interplay between both isoforms. To support this argument, many hundreds of genes were detected that are only sensitive to the presence of NEAT1_2/paraspeckles, but not NEAT1_1. Genes in this category were up- or down-regulated in NEAT1_2^−/− cells compared to WT, with no subsequent further change in expression levels when NEAT1_1 was lost in total NEAT1^−/− cells (352 genes inhibited and 103 genes activated, Supplementary Table S3.8, Figure 3.6B,C).

Finally, if was of interest to see if we could find genes regulated by both isoforms, but in an opposite manner. In this case genes were differentially expressed in NEAT1_2^−/− cells compared to WT, but these genes were then restored to the original WT expression levels in total NEAT1^−/− cells (Figure 3.6B,C). This pattern can be rationalized thus: in WT cells, the two NEAT1 isoforms both exert an equal, but opposite, influence on the target gene. When NEAT1_2 is lost, NEAT1_1 then exists unregulated, leading to altered gene expression of its target genes. When both NEAT1 isoforms are lost in the total NEAT1^−/− cells, the presence of both the activator and the repressor are removed, leading to a neutral change compared to WT. Based on this supposition, 333 genes were inhibited by NEAT1_2 and activated by NEAT1_1 to the same extent, and 103 genes were activated by NEAT1_2 and inhibited by NEAT1_1 to the same extent (Supplementary Table S3.9, Figure 3.6B,C). There was a final group of genes that were regulated by both isoforms,
but not to the same extent. Of these, the majority were more sensitive to NEAT1_2 levels than NEAT1_1 levels (115 genes strongly regulated by NEAT1_2 and mildly regulated by NEAT1_1, Supplementary Table S3.9, Figure 3.6B,C). In contrast, only 3 genes were found to be targets of both isoforms acting synergistically (Supplementary Table S3.7, Figure 3.6B,C), suggesting that the predominant regulation is antagonistic between the two isoforms. This finding explains why there was a much more dramatic change in gene expression in the NEAT1_2⁻/⁻ cells than the total NEAT1⁻/⁻ cells, compared to WT (Figure 3.4). The reason is that the total NEAT1⁻/⁻ cells experienced the loss of two opposing regulators – the effects of which essentially cancelled each other out. These data support the hypothesis of a potential paraspeckle-independent role for the NEAT1_1 isoform, and propose a new role for paraspeckles as a negative regulator of this activity.
<table>
<thead>
<tr>
<th>Expression level in WT</th>
<th>Expression level in NEAT1_2</th>
<th>Expression levels in Total NEAT1</th>
<th># genes counted</th>
<th>Interpretation: How ‘gene x’ is regulated?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>unchanged</td>
<td></td>
<td>11</td>
<td>Inhibited by NEAT1_1, but not affected by NEAT1_2.*</td>
</tr>
<tr>
<td>WT</td>
<td></td>
<td></td>
<td></td>
<td>Not affected by both NEAT1 isoforms.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>Activated by NEAT1_1, but not affected by NEAT1_2.*</td>
</tr>
<tr>
<td>Baseline</td>
<td>increased</td>
<td></td>
<td>1</td>
<td>Inhibited by both NEAT1 isoforms synergistically.**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>352</td>
<td>Inhibited by NEAT1_2, and not affected by NEAT1_1.***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>102</td>
<td>Inhibited by NEAT1_2, activated by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEAT1_2 is more powerful.****</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>333</td>
<td>Inhibited by NEAT1_2, activated by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEAT1_1 is more powerful.****</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>Activated by NEAT1_2, activated by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEAT1_1 is more powerful.****</td>
</tr>
<tr>
<td>Baseline</td>
<td>decreased</td>
<td></td>
<td>1</td>
<td>Activated by NEAT1_2, inhibited by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>111</td>
<td>Activated by NEAT1_2, inhibited by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEAT1_1 is more powerful.****</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>Activated by NEAT1_2, inhibited by NEAT1_1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NEAT1_2 is more powerful.****</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>103</td>
<td>Activated by NEAT1_2, and not affected by NEAT1_1.***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>Activated by both NEAT1 isoforms synergistically.**</td>
</tr>
</tbody>
</table>
Figure 3.6: Subsets of genes are differentially regulated by the two NEAT1 isoforms indicating that the majority of genes are regulated by both isoforms in an equal, but opposite manner, or are only sensitive to NEAT1_2/paraspeckles. (A) Example of changes in the expression level of a gene with the same expression level in WT and NEAT1_2^-/- cells, followed by three possible ways of changing after both NEAT1_1 and NEAT1_2 were depleted. (B) Example of changes in the expression level of a gene that was upregulated compared to WT cells after NEAT1_2 depletion, followed by five different ways this gene could change with total NEAT1^-/-. (C) The opposite situation to (B), except the original gene is downregulated in NEAT1_2 depleted cells compared to WT cells. In each case the number of genes that fit each situation is listed. * indicates genes that are only affected by the level of NEAT1_1, but not NEAT1_2. ** indicates genes affected by both NEAT1_1 and NEAT1_2 synergistically. *** Indicates genes only affected by NEAT1_2 levels, but not NEAT1_1. **** Indicates genes affected by both NEAT1_1 and NEAT1_2, but in the opposing manner.
3.4 Discussion

In this study we delineated acidosis in U2OS cells as a stress scenario that induces NEAT1 and paraspeckles, and determined that p53 is playing a key role in orchestrating this increase by transcriptionally activating NEAT1. We then created knockout cell lines of the different NEAT1 isoforms, confirming NEAT1_2/paraspeckles can regulate expression of many genes, as well as revealing a potential paraspeckle independent role for NEAT1_1, and a new role for paraspeckles as a regulator of this NEAT1_1 function. These regulatory mechanisms appear to be used by oncogenic cells to modulate migration and cell survival under stress.

3.4.1 Paraspeckles and omega speckles – convergent evolution?

Although NEAT1 is mammalian specific, a similar stress-induced mechanism of lncRNA action may exist back to invertebrates. The Drosophila-specific heat shock RNA omega (hsr-ω) locus produces two major transcripts, a shorter, abundant, cytoplasmic form (Hsr-ω-c), and a longer, less abundant, nuclear transcript (Hsr-ω-n) (Bendena et al., 1991; Garbe et al., 1986; Ryseck et al., 1987). Like NEAT1_1 and NEAT1_2, the Hsr-ω transcripts overlap at their 5’ ends (Bendena et al., 1991; Garbe et al., 1986; Jolly and Lakhotia, 2006; Lakhotia and Sharma, 1995). With heat shock, the nuclear Hsr-ω-n isoform is rapidly upregulated and accumulates within enlarged omega speckles (Bendena et al., 1989; Hogan et al., 1995; Lakhotia and Sharma, 1995). Like paraspeckles, many nuclear proteins become enriched in the omega speckles, such as Hrb87f (the Drosophila homolog of HNRNPA1, a paraspeckle protein) and Hrb57a (the Drosophila homolog of HNRNPK, an essential paraspeckle protein) (Buchenau et al., 1997; Dangli and Bautz, 1983; Dangli et al., 1983; Hovemann et al., 1991; Lakhotia et al., 1999; Naganuma et al., 2012; Prasanth et al., 2000; Samuels et al., 1994; Zu et al., 1998). It has been speculated that omega speckles act as a temporary storage site for nuclear proteins, so that they can quickly resume normal function after the stress has been relieved (Jolly and Lakhotia, 2006; Lakhotia et al., 1999; Prasanth et al., 2000). Similarly, the mammalian lncRNA Satellite III which nucleates proteins to form nuclear stress bodies, was also postulated to act as a storage site for nuclear proteins under stress.
(Biamonti and Vourc’h, 2010; Lindquist, 1986; Metz et al., 2004). Interestingly, NEAT1 and hsr-ω do not have sequence similarity, but nevertheless share stress induction, recruitment of the same proteins, and even pattern of isoform production, suggesting a possible case of convergent evolution for the formation of omega speckles and paraspeckles, and strengthening arguments for the physiological importance of such structures.

3.4.2 Transcriptional upregulation of NEAT1 by p53 and other factors

In our study, we found that extracellular acidosis treatment is a strong inducer of the expression of NEAT1 and paraspeckles in U2OS cells. The data support our and other studies showing that increased NEAT1 levels are due to increased transcription of NEAT1; however we cannot exclude the possibility that NEAT1 can additionally be upregulated through increased RNA stability or reduced degradation. In a previous study, we showed that the proteasome inhibitor MG132 induces NEAT1 primarily through transcriptional activation (Hirose et al., 2014). Interestingly, MG132 is also known to inhibit p53 degradation (Zhang et al., 2011). We added weight to the notion of p53 regulating NEAT1, showing it acts as a transcriptional activator of NEAT1, and is one of several factors involved in the upregulation of NEAT1 induced by acidic stress. p53, well-known for its tumor suppressive function by mediating cell apoptosis and cell cycle arrest (reviewed in Zilfou and Lowe, 2009), also binds to a large number of lncRNA genes and potentially regulates their expression (Huart et al., 2010; Zhang et al., 2014). Thus, NEAT1, added to this list of lncRNAs, joins an already extensive regulatory network supported by p53. Beyond p53, the regulation of NEAT1 transcription is extremely complex, with ENCODE ChIP-seq data revealing a large number of transcription factor binding sites in the promoter and gene body of NEAT1 (Supplementary Table S3.1). Thus the transcriptional regulation of NEAT1 is likely a dynamic equilibrium of the interactions between many different molecular pathways. This complex regulatory network controlling NEAT1 is also an explanation for the ubiquitous and dynamic NEAT1 expression observed in tissues and cell lines (Gibb et al., 2011; Nakagawa et al., 2011).
3.4.3 Physiological role for NEAT1 in cancer
Several studies have indicated that, at the cellular level, NEAT1 is associated with pro-survival, pro-proliferative, and pro-migration functions in cancer cells. In breast cancer, NEAT1 is induced with hypoxia, leading to increased proliferation, clonogenic survival and reduced apoptosis of breast cancer cells under hypoxic, but not normoxic, conditions (Choudhry et al., 2014). In prostate cancer, high levels of NEAT1 are associated with the aggressive, castrate-resistant subtype. Forced over-expression of NEAT1_1 in prostate cancer cell lines led to increased tumor size in mouse xenograft studies, and increased invasion and colony formation attributes in vitro, with shRNA targeting NEAT1 producing the opposite effects in these assays (Chakravarty et al., 2014). The total NEAT1^-/- cells in our study also showed reduced migration and reduced viability under acidosis treatment. This observation matches the finding of a previous study, in which murine embryonic fibroblasts (MEFs) from NEAT1^-/- mice also showed reduced tolerance to proteasome inhibitors, compared to wild-type MEFs (Hirose et al., 2014). These results suggest that NEAT1 is not essential for maintaining the steady state of transformed cells, but is more important when cells are subjected to stress, allowing cells a window of increased survival. In this context it would be useful in the future to use the lines generated in this study for xenografts in mice. Although our data did not suggest significant changes of cell proliferation and movement under normal conditions, it is possible that a more dramatic difference could be seen when these cells are transplanted into an animal and exposed to a more physiologically relevant environment.

3.4.4 Physiological roles for NEAT1 in normal animals
In mice, selected cell types with high NEAT1 expression are highly proliferative and associated with secretion. For example, NEAT1 is highly expressed in the corpus luteum, a structure transiently formed by ovarian epithelial cells to become a secretory organ in pregnancy, as well as in mammary epithelium, the target of dramatic remodelling in mammary gland differentiation prior to lactation (Nakagawa et al., 2014; Standaert et al., 2014). Although NEAT1^-/- mice can survive with no gross abnormalities (Nakagawa et al., 2011), the formation of both the corpus luteum and lactating tissues are impaired, causing failures in secreting enough progesterone and lactation. Interestingly, both
these defects result in significant challenges to the reproductive capacity of the mice. Thus the pro-proliferative property of NEAT1 in transformed cells could be linked with its importance for supporting the formation of highly plastic and dynamic tissues that are remodelled in adult animals.

3.4.5 NEAT1 mechanism of action

NEAT1 expression has been largely synonymous with paraspeckle formation in the literature. The observation that the NEAT1_2−/− cell lines had a very different transcriptomic landscape to both total NEAT1−/− and wild-type cells suggests different roles for the two NEAT1 isoforms. By analyzing DE genes between the three types of cell lines, we identified different examples of gene subsets regulated by NEAT1 isoforms, either directly or indirectly (Figure 3.6). The two main groups of genes are those co-regulated by both NEAT1 isoforms in opposite ways, as well as genes only regulated by NEAT1_2 but not NEAT1_1. This latter group of genes are paraspeckle target genes, and this concept concurs with that accepted in the literature that the principal role of NEAT1 is in the formation of paraspeckles. The former group of genes is somewhat more novel, as it postulates distinct and opposing roles for NEAT1_1 and NEAT1_2/paraspeckles. Interestingly, the levels of NEAT1_2/paraspeckles are intimately linked to the levels of NEAT1_1: they only increase when NEAT1_1 increases. This is typical of a negative regulator whose levels increase only in instances where the target molecule has increased.

Thus, our data suggest that gene regulation by the NEAT1 IncRNA can be carried out at several levels. The first mechanism involves NEAT1_1 acting locally at the chromatin of target genes (Chakravarty et al., 2014; West et al., 2014). This may be the predominant mechanism, as mouse studies have shown widespread expression of NEAT1_1, but not NEAT1_2, in the majority of cells in a widespread range of tissues (Nakagawa et al., 2011). The mechanism for targeting NEAT1_1 to chromatin is not known, but CHART-seq (a method for identifying chromatin associated with a given RNA) with probes against NEAT1_1 suggests a relationship with the transcription start sites of active gene loci (West et al., 2014). We, in collaboration with others, have shown
using RNA immunoprecipitation and proteomics, that NEAT1_1 associates with histone H3 (Chakravarty et al., 2014). It is possible this chromatin-associated NEAT1_1 mechanism could be similar to what has been proposed for several other lncRNAs that act as epigenetic regulators, such as HOTAIR (Tsai et al., 2010). The numerous ‘micro speckles’ observed for NEAT1_1 in NEAT1_2^−/− cells may represent the sites for this activity. Future experiments with RNA FISH against NEAT1_1, coupled with DNA FISH against the gene loci identified by the CHART experiments, should be used to determine if these sites indeed colocalize. The second mechanism of NEAT1 action is to produce the longer NEAT1_2 isoform in some cells and scenarios, including with stress. In mice, under normal conditions, the expression of NEAT1_2 is restricted to only a subset of cells in any given tissue type, as well as to specialized structures such as the corpus luteum. Once transcribed, the function of NEAT1_2 is to form paraspeckles, and in doing so, sequester both paraspeckle proteins, and also NEAT1_1. The effect of this sequestration is modulation of expression of both NEAT1_1 and paraspeckle protein target genes. Future studies to test this hypothesis could include over-expression of NEAT1_1 in cells both with and without paraspeckles to determine changes in expression of putative NEAT1_1 target genes. It would be advantageous to be able to quantitate the NEAT1_1 molecules inside and outside paraspeckles, but this is confounded by the complete overlap of NEAT1_1 and NEAT1_2 at their 5’ ends.

In conclusion, we identified NEAT1 as a p53 responsive gene under acidic stresses, and combined novel methods to demonstrate the possible functional relationship between the two isoforms of NEAT1 in regulating gene expression. Overall, this study provides an important global view of the importance of NEAT1 isoforms in cell biology so that future research can be focused on their molecular mechanisms and their respective biological roles.
3.5 Experimental procedures

Cell culture and treatments

U2OS cells were grown in high glucose DMEM (Invitrogen) and supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. For acidosis treatment, we used 1M HCl to adjust the pH of media to 6.2-6.4, which is equivalent to approximately 330μL of 1M HCl per 10mL of culture media. The same amount of 1M KCl was used to provide control for chloride ions. All cells were cultured in a 37°C incubator supplied with 5% CO₂.

NEAT1 promoter DNA cloning

A 3.0kb section of the human NEAT1 promoter was amplified from healthy human genomic DNA using Pu polymerase (Bioline) using nested PCR. Human genomic DNA was a kind gift from Grant Morahan’s laboratory. The first PCR product was amplified using the primer set: Fwd 5’-TGACCTGGCTTCAGAACAAGATGTGGG-3’, Rev 5’-GCCTATTCCTCCTGACTCTCCACC-3’. The DNA product was gel purified and then used as the template DNA for the second PCR, using primers: Fwd 5’-ACTACTGGTACCCCAATGTGGTGGTGAAGCCTATAG-3’, Rev 5’-ACTACTGCTAGC CACTTCCTCCCCACAAC-3’. The product was then purified and digested by KpnI and NheI restriction enzymes (NEB), and then subsequently ligated into pGL3-basic luciferase vector (Promega) that was also digested with the same two enzymes. The ligated products were transformed into DH5α E.coli for amplification.

Site-directed mutagenesis

The p53 binding site in the human NEAT1 promoter was mutated by site-directed mutagenesis. The pGL3-NEAT1pr-Luc vector was first amplified by Phusion polymerase (NEB), using primer set that carried the desired mutation: Fwd 5’-CAGGCGCCAGGGAGTAAAACCTGGGTATTACCACATCAACC-3’, Rev 5’-GGGGTGGTGTGGTAAACCCAGGTTTACTCCCTGCGCCCTTG-3’. The product was then digested with DpnI and transformed into DH5α E.coli for amplification.
**Promoter reporter assays**

The day prior to transfection, cells were seeded at $1 \times 10^5$ per well of 12-well. Plasmids carrying p53-normal and p53-mutated NEAT1 promoter were transfected using Lipofectamine LTX (Invitrogen) individually at 1.2μg per 12-well. 4 hours after transfection, the media was replaced with acidosis-treatment media. 68 hours after the acidosis treatment, cells were lysed in TRIsure (Bioline), and total RNA was extracted for qRT-PCR analysis of luciferase mRNA. The qPCR primers used are: Fwd 5’-AGGTGGCTCCCGCTGAAT -3’, Rev 5’- CATCGTCTTCCGCTGAT-3’.

**qRT-PCR**

RNA was first extracted using TRIsure (Bioline), and 100ng of extracted RNA was converted to cDNA using the QuantiTect Reverse Transcription Kit following the manufacturer’s instructions. Quantitative real-time PCR was performed using the SensiMix SYBR No-ROX kit and Rotor-Gene Q Real-time PCR cycler. Fluorescence was detected at three degrees lower than the Tm of the specific product, as determined by the melt curve to reduce the interference from unspecific products. The Cycle threshold (Ct) values of each product was determined under a constant normalized fluorescence threshold of 0.075. Relative expression of genes were calculated using $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). U6 and GAPDH mRNA were used as housekeeping gene to control for variation in cDNA concentration between samples. Primers used for different genes are as following: Total NEAT1 qPCR primer set: Fwd 5’-GTGGCCTGTGGAGTCCGGA-3’, Rev 5’- TAACAAACCACGGGCTA-3’. NEAT1_2 qPCR primer set: Fwd 5’-GTCTTTCCATCCACGTGAT-3’, Rev 5’-GTACTCTGTGGGTAGTGTCAG-3’. GAPDH qPCR primer set: Fwd 5’-ATGGGGAAGGTGAAGGTCGA-3’, Rev 5’-GGGTCATTGATGGCAACAATA-3’. U6 qPCR primer set: Fwd 5’-CTCGCTTCCGACAGCA-3’, Rev 5’-AACGCTTCAATTGCGT-3’.

**RNA interference**

Cells were seeded at $1 \times 10^5$ per well of a 12-well plate on the same day of transfection for a reverse transfection procedure. p53 knockdown was achieved by transfecting a mixed
pool of four SMARTpool siRNAs (Dharmacon) at 33nM final concentration using Lipofectamine RNAiMAX (Invitrogen).

**RNA-FISH and Immunofluorescence**

Cells grown on coverslips were fixed using 4% paraformaldehyde/PBS, and permeabilized by 70% ethanol overnight. Stellaris RNA-FISH probes (Biosearch Technologies) were used and the procedure was carried out according to the instruction provided. In brief, permeabilized cells were incubated with probes in buffer overnight at 37°C, and washed with buffer for 30 minutes afterwards. To further conduct immunofluorescence staining, cells were incubated with primary antibodies in PBST for 1 hour at 37°C, and then washed with PBST for 3 times of 5 minutes each. Cells were then incubated with secondary antibodies in PBST for 1 hour and 37°C, and then washed again. Finally, cells were incubated with DAPI (1:15000) in PBS for 2 minutes at room temperature, and then mounted onto slides with VectaShield (Vector Laboratories) mounting media. Fluorescence was imaged using the DeltaVision Elite Imaging System and its bundled software (GE Healthcare). All images taken were Z-stacks of 0.2μm sections, that were subsequently deconvolved and maximum intensity projections generated. The same exposure settings and post-processing parameters were used within each set of experiments. The antibodies and RNA-probes used were listed in Supplementary Table S3.10.

**Western blotting**

Total cell lysate was obtained by homogenizing cells in SDS sample buffer. Lysates were separated on a 4-12% Bis-Tris SDS-PAGE gel (Invitrogen) and then blotted onto a PVDF membrane (Merck Millipore). After blocking in skim milk powder, primary and secondary antibodies were incubated over the membrane with gentle rocking. Results were visualized using the Odyssey Infrared Imaging System (LI-COR Biosciences). Antibodies used are listed in Supplementary Table S3.10.

**CRISPR genomic engineering**

The px330 plasmid was obtained as a kind gift from Feng Zhang’s laboratory through
Addgene for the cloning of the guide RNAs (Cong et al., 2013). Guide RNAs were designed using a CRISPR design tool available on http://crispr.mit.edu/ (Hsu et al., 2013) (Supplementary Table S3.10). Oligonucleotides of guides were first annealed to form double stranded DNA before being ligated into the BbsI restriction site in the px330 plasmid. To construct repair templates for the homologous recombination, we utilized a pEGFP-C1 plasmid (Clontech), with mutations that turned the EGFP to a Venus YFP, as the backbone to carry the CMV-YFP-SV40(T) expression cassette. We amplified homologous sequences of approximately 500bp from the genomic DNA of a healthy human sample, and cloned each pair of homologous arms into the AseI and MluI sites that flank each side of the YFP expression cassette using Gibson cloning (New England Biolabs) (Supplementary Table S3.10). To transfect cells, px330 plasmids containing guide sequence and repair template plasmids were co-transfected using Lipofectamine 2000 (Invitrogen) at 1:1 ratio into the cells. Transfected cells were passaged for one week, and then sorted using FACS for the top 10% most fluorescent cells. A total of three to four sortings were carried out at 3-4 day intervals to increase the proportion of fluorescent cells in the population. Single cells obtained from the last sorting were seeded into 96-well plates and cultured for colony formation. Approximately 2 to 3 weeks after the single cell seeding, genomic DNA of each colony was extracted and subjected to a PCR-based method for screening whether the YFP was inserted into neither, one, or both of the desired loci (strategy shown in Supplementary Figure S1). qRT-PCR was used to verify the expression level of the target gene. Only homozygous NEAT1 deleted CRISPR clones were used for further experiments. Oligonucleotides of guide RNA, and primers for cloning repair templates are shown in Supplementary Table S10.

Whole transcriptomic RNA sequencing and analysis
A total of nine U2OS cell lines were subject to RNA-seq: three separately passaged wild-type U2OS cells, three total NEAT1\(^{-/-}\) colonies, and three NEAT1\(_2^{-/-}\) colonies had their transcriptome analyzed. Total RNA was prepared from cultured cells using TRIsure (Bioline). Ribosomal RNA was depleted in all samples before library preparation and sequencing, which was carried out by the Australian Genome Research Facility (AGRF). Library was constructed using Illumina TruSeq Stranded Total RNA Library Prep Kit
with Ribo-Zero- Gold Kit for the removal of ribosomal RNA. The whole transcriptome (ie. not just limited to poly(A)+ RNA) was sequenced on the Illumina HiSeq 2000 platform to generate a minimum of 20 million 50bp single-end reads for each sample, and all samples were sequenced in a single lane to prevent batch-effects. Sequenced read files were uploaded onto the Galaxy platform (http://galaxyproject.org), converted to FASTQ format by FASTQ Groomer (Blankenberg et al., 2010), and then mapped to the human Hg19 reference genome using Tophat2 (version 0.6) with the default settings (Kim et al., 2013). Read alignments were processed by the htsq-count python script using the ‘union’ and ‘reverse stranded’ mode for gene level read counting (Anders et al., 2015). For differential gene expression analysis, non-expressed genes were filtered out by setting a requirement of more than 10 reads in at least three samples for each gene. Next, the gene expression counts were normalized using the RUVg-empirical method (Risso et al., 2014). The top 5000 most consistently expressed genes generated across all samples in each DE analysis were used to calculate normalization factors of all genes. RLE (relative log expression) and PCA (principle component analysis) of DE analyses were generated by R for testing statistical features. A list of differentially expressed genes was computed by EdgeR using a cut-off of FDR (False Discovery Rate) ≤ 0.05. DE genes were subsequently imported for Ingenuity pathway analysis (Ingenuity Systems, www.ingenuity.com) using standard settings for predicting potential pathways and functions being affected by the NEAT1 KO. Data exploration on relative log expression, principle component analysis, and mean-difference plots etc. were generated by R and the EDASeq R package.

Real-time cell analysis (proliferation, viability and migration)

All real-time cell analyses were conducted using the xCELLigence system (Roche). This system measures the cell-surface contact area from an impedance-based signal. Cells (NEAT1 WT and KO) were cultured as previously described before experiments. For proliferation and viability analysis, E-16 xCELLigence plates (Roche) were used according to the manufacturer’s instruction. In brief, trypsinized cells were seeded at 1x10^4 cells in each well in 200μl, either in standard DMEM culture media, or in pH6.2-6.4 DMEM culture media. The plate was immediately inserted into the xCELLigence
apparatus inside an incubator, and the electrical impedance was measured across the bottom of the well in real time every 10 minutes for 72 hours. For the cell migration assay, CIM-16 xCELLigence plate (Roche) was used according to the manufacturer’s instruction. In brief, trypsinized cells were seeded at $8 \times 10^4$ cells per well in the upper chamber with 150μl of serum free media, and then migrated to the normal culture media containing 10% FBS in the lower chamber. Cells were allowed to settle for 1 hour at room temperature after seeding, and then the plate was inserted into the xCELLigence apparatus inside an incubator, and the electrical impedance was measured across the bottom of the well in real time every 10 minutes for 48 hours.
3.6 References


Chapter 4
Transcriptional regulation of NEAT1
4.1 Introduction

One of the most obvious features of NEAT1 expression is that it is either not expressed or very weakly expressed in pluripotent stem cells, but is abundantly expressed in any cells that are differentiated, regardless of cell type (Chen and Carmichael, 2009; Nakagawa et al., 2011). Despite this lack of tissue specificity, the actual expression level may still vary greatly from cell to cell, and from tissue to tissue. One very interesting observation is that NEAT1 is highly expressed particularly in the surface epithelium cell layer of the forestomach and esophagus of mice, or in glands such as adrenal glands, and Harderian glands, where epithelium and endothelium are the main tissues (Nakagawa et al., 2011). As a result, the NEAT1_2 transcript and hence paraspeckles are much more likely to be found in those tissues than in others. However, there is still no appreciation of how this observation can relate to the physiological function of NEAT1 and paraspeckles. One important task is to investigate the mechanism that drives NEAT1 expression.

The final expression level of any gene is the result of both transcriptional and post-transcriptional events, which contribute to the generation of the RNA and the stability of the nascent RNA respectively. In the case of NEAT1, the mechanism of the biogenesis of the NEAT1_1 and NEAT1_2 isoforms, which is a post-transcriptional event, is well understood. As reported by Naganuma et al. in 2012, the differential production of the two isoforms is primarily a result of a complex competition between several proteins for the cleavage and polyadenylation signal located at the 3’ end of the NEAT1_1 transcript (Naganuma et al., 2012). Due to the high efficiency of the polyadenylation, the vast majority of the nascent NEAT1 transcript thus becomes the NEAT1_1 isoform, rather than NEAT1_2. This result suggested that factors that block the polyadenylation of NEAT1_1 would increase the abundance of the NEAT1_2 isoform, thus making more paraspeckles. And indeed, one essential paraspeckle protein, HNRNPK, is involved in this process (Naganuma et al., 2012). The stability of the two NEAT1 isoforms is also different, but both are relatively unstable. In mouse 3T3 cells, that have abundant paraspeckles, NEAT1_2 appears to be more stable with a half-life of about 60 minutes,
whereas NEAT1_1 has a half-life of approximately 20 minutes (Clark et al., 2012). This high turnover of NEAT1 may contribute to the dynamic nature of NEAT1 expression, and significantly highlights the importance of the transcriptional activity of the \textit{NEAT1} gene. Thus the most efficient way for cells to make paraspeckles is to maintain strong transcription of the entire \textit{NEAT1} gene. This will offset the poor stability of both NEAT1 isoforms, and raise NEAT1_2 levels by virtue of simply increasing the overall number of transcripts being made, of which a constant small proportion takes the form of NEAT1_2. Eventually, when NEAT1_2 is abundant enough, paraspeckles will start to form (Mao et al., 2011; Naganuma et al., 2012). This mechanism fits the general observation that NEAT1_2 is not detected in tissues that do not already express high NEAT1 levels (for example, the spleen, bladder, and pancreas, (Nakagawa et al., 2011)). But more interestingly, the \textit{NEAT1} gene appears to be specifically evolved to allow this mechanism to work, because the entire gene, regardless of the isoform produced, is regulated with the same promoter with only one transcription start site (TSS) (Guru et al., 1997; Sunwoo et al., 2009). Therefore, understanding the transcriptional regulation of the \textit{NEAT1} promoter is a key step to determining the cellular significance of NEAT1 expression.

The transcriptional activity of a gene is dependent on the chromatin status at both the epigenetic level, as well as the binding of transcription factors to promoter and enhancer elements. At the epigenetic level, one can evaluate DNA methylation, the covalent addition of methyl groups to DNA nucleotides, and histone modification, the covalent modification of specific histone residues, associated with the gene and its regulatory DNA. Both DNA methylation and histone modifications are related to a wide range of biological outcomes.

In mammalian systems, the most extensively studied DNA methylation is at the fifth position of cytosine, namely 5-methylcytosine (5mC). Three conserved enzymes, DNA methyltransferases (DNMT1, 3A and 3B) are responsible for its deposition and maintenance, and are essential for normal development (Li et al., 1992; Okano et al., 1999). The most extensively studied 5mC is in the context of CpG dinucleotides, which are often described as a ‘silencing’ epigenetic mark that can be inherited through somatic
cell division (Holliday and Pugh, 1975; reviewed in Jones, 2012; Riggs, 1975). CpG islands are regions of DNA that shows high frequency of CpG sites, and are often associated with the promoter and transcription start site of genes (reviewed in Deaton and Bird, 2011). Abundant 5mC in these regions would result in a repressive effect on transcription (Holliday 1975, Riggs 1975). However, although most (60-80%) CpGs in mammalian genomes remain methylated during development, the CpG islands found at promoters of many housekeeping or developmentally regulated genes are constitutively hypomethylated (reviewed in Smith and Meissner, 2013). This pattern presents a landscape in which the ‘bulk genomic methylation patterns are static across tissues and throughout life, changing only in localized contexts as specific cellular processes are activated or shut down’ (Smith and Meissner, 2013). For methylated repressive CpG islands to enter into a permissive state that allows transcription, 5mC has to be actively removed. This process requires a number of proteins, such as the TET (ten-eleven translocation methylcytosine dioxygenases), AID (activation-induced cytidine deaminase), and TDG (thymine DNA glycosylase) enzymes (Bhutani et al., 2010; Cortazar et al., 2011; Cortellino et al.; Inoue and Zhang, 2011; Iqbal et al., 2011; Jones, 2012).

The state of CpG island methylation is closely associated with the status of histone modifications, because DNA methylation can be directly suppressed by the presence of active histone marks (Ooi et al., 2007). A large number of different histone modifications have been identified, and they are often considered as marks for either ‘active’ or ‘repressive’ chromatin states (reviewed in Bannister and Kouzarides, 2011). For example, histone acetylation and phosphorylation are ‘active marks’ that are considered to lead to a less compact chromatin structure. These negatively charged modifications can effectively reduce the positive charge of histones, reducing the electrostatic interactions between histones and negatively charged DNA. This eventually reduces the compactness of chromatin and facilitates DNA access by proteins such as transcription factors. On the other hand, methylation of histone tails, which are neutrally charged, can result in either ‘active’ or ‘repressive’ marks, depending on which amino acid residue is modified. For instance, methylation on lysine 4, 36 and 79 is associated with
transcriptional activation, and lysine 9, 27 and 20 is associated with transcriptional repression (reviewed in Kouzarides, 2007).

A more opened state of chromatin eventually allows the access of more proteins. Transcription factors require those spaces to finally mediate transcriptional events, which are involved with large protein complexes such as the RNA polymerase machinery. Transcription factors also have activatory and inhibitory properties, and sometimes one transcription factor can carry both properties at the same time and regulate gene expression differently for different genes, such as p53 (reviewed in Riley et al., 2008). However, it appears a limited number of transcription factors have led to the extremely complex patterns of gene expression in different cell types and conditions. To this end, it is accepted that transcription factors act in combination to regulate gene expression, which is perhaps the only economic way for cells to avoid having an overwhelming large number of transcription factors. Indeed, findings published in recent years from the ENCODE project have systematically demonstrated this possibility, and revealed patterns of co-association between transcription factors (Gerstein et al., 2012). Moreover, these studies also showed how transcription factors can be organized in hierarchical order based on their connectivity with other molecules, as well as their influence the target gene expression (Gerstein et al., 2012). From this, one can easily see that the regulation of genes, and the regulation of the regulatory mechanisms themselves, has intertwined into an incredibly complex network.

From a system biology point of view, all transcription factors are regulated by molecular pathways that react to different triggers. Pathways themselves could influence and interact with each other, and this eventually forms the biological system (reviewed in Barabasi and Oltvai, 2004). Fortunately, although these pathways are complex, they can still be characterized, often by analyzing the biological functions a subset of proteins or genes are associated with, one may find specific functions being enriched. These enriched functions may subsequently provide implications on firstly, which cellular functions these genes are related with, and secondly, how the downstream genes regulated by these genes are controlled by certain biological processes. It is interesting that co-
regulated genes often carry similar biological functions, and are regulated by a common set of transcription factors (Allocco et al., 2004).

In a biological system, transcriptional regulation is affected by epigenetics and transcription factors, and all these extend to the interaction between upstream molecular pathways controlling them. For a specific gene, understanding the key pathways behind its expression is perhaps more helpful in understanding how this gene is connected with its cellular functions, and may even imply the possible function this gene carries. For NEAT1, due to the lack of knowledge in general of its function at the cellular level, such upstream analysis may provide an important insight into how NEAT1 can be linked to the regulatory system of the cell. This understanding would eventually provide valuable information once the effect of NEAT1 expression is also well-understood, and hence a concept of the general function of the NEAT1 gene on the cellular level can be proposed.

Thus the aims of this chapter are to understand:

1. The epigenetic controls involved in the regulation of NEAT1.
2. The pathways controlling NEAT1 expression, and how they contribute to the dynamic regulation of NEAT1.
3. The mechanism that leads to the ubiquitous expression of NEAT1, although expression is especially high in only a few tissues.

Based on the existing knowledge, my hypothesis is that the regulation of NEAT1 will be complex and will likely comprise a limited number of pathways that are either shared by many different cell types, or many pathways that each respond to different triggers.
4.2 Results

4.2.1 Data mining indicated more repressive epigenetic markers of \textit{NEAT1} in stem cells

The literature suggests that \textit{NEAT1} is not expressed in human embryonic stem cells, but is expressed when these cells differentiate (Chen and Carmichael, 2009). To pinpoint if epigenetic changes were involved in this process, data from \url{http://neomorph.salk.edu/ips_methylomes/browser.html} was utilized (Lister et al., 2009). The data consists of genome-wide, single-base-resolution maps of methylated cytosines along with several histone modifications that were obtained from human embryonic stem cells, human induced pluripotent stem cells (iPSCs), fetal fibroblasts, and a few other differentiated cell lines (Lister et al., 2009). To identify the epigenetic state of \textit{NEAT1}, an 8kb upstream region of the human \textit{NEAT1} gene was located and the methylcytosine (mC) state of three cell lines of each differentiation level were examined: [pluripotent stem cells: (1) H1 embryonic stem cells, (2) IMR90-iPSCs fetal lung fibroblast-driven iPSCs (induced pluripotent stem cells), and (3) FF-iPSCs foreskin fibroblast induced iPSCs] vs. [Differentiated cells: (1) ADS adipocyte stem cells, (2) IMR90 fetal lung fibroblast, and (3) FF foreskin fibroblast] (Figure 4.1A). As shown in Figure 4.1, all three stem cell lines displayed significant enrichment of mC from approximately 6kb upstream of the \textit{NEAT1} transcription start site (TSS) into the 5’ end of the gene body. The methylation enrichment is, however, greatly reduced in all three differentiated cell types. This is particularly obvious in the gene body of \textit{NEAT1}, marked in red, which is a GC-rich region (shown as CpG island:17 in Figure 4.2).

In addition to methylcytosines, histone modifications also showed differential enrichment that favoured transcriptional activation in differentiated cells (Figure 4.1B). Comparison of the state of four different histone modifications in H1 embryonic stem cell and IMR90 fetal fibroblast was performed. H3K4me1, which was often found to be associated with enhancers, showed increased prevalence particularly from -2kb to -6kb upstream of the \textit{NEAT1} TSS in IMR90. H3K9ac, which was often found to associate
with transcriptional initiation and open chromatin, showed significant enrichment around the TSS of NEAT1 in IMR90. H3K36me3, which often marks elongation of RNA polymerase II, was found to be significantly enriched downstream of the TSS of NEAT1. Finally, H3K27me3, which often associates with repressed chromatin, was found to generally lose signals across the entire region in differentiated cells. Together, these published data indicated that as cells become more differentiated, a series of epigenetic changes takes place around the TSS of the *NEAT1* gene, making it more permissive to transcriptional activation.
Figure 4.1: Epigenetic marks upstream of the human *NEAT1* gene in different cell lines. (A) Comparison of methylcytosines of the genomes of three stem cell lines (H1 embryonic stem cell, IMR90 fetal lung fibroblast-derived iPSC, and foreskin fibroblast-derived iPSC, line 6.9), and three more differentiated cell lines (ADS adipocyte stem cells, IMR90 fetal lung fibroblast and FF foreskin fibroblast). (B) Comparison of four types of histone modifications between H1 embryonic stem cells and IMR90: H3K4me1 (associated with
enhancers), H3K9ac (associated with transcriptional initiation and open chromatin), H3K36me3 (marks regions of RNA polymerase II elongation) and K3K27me3 (associated with repressed chromatin regions). All tracks are available publically on http://neomorph.salk.edu/ips_methylomes/browser.html (Lister et al., 2011).
4.2.2 The transcriptional regulation of NEAT1 is controlled by multiple pathways

To further pinpoint the regulatory mechanism controlling the expression of NEAT1, the ENCODE TFBS (transcription factor binding site) track, which provides a large amount of ChIP-Seq data on the genome-wide binding sites of more than 160 transcription factors in multiple cell lines, was thoroughly examined (Figure 4.2). Firstly, the TFBS detected both nearby and upstream of the TSS of the human \textit{NEAT1} gene formed six clusters, named here cluster A to F. These clusters are significantly more conserved among vertebrates than other TFBS-depleted regions in-between. The entire region of six clusters spans from the TSS of NEAT1 to around 6.5kb upstream, before entering the upstream gene named \textit{FRMD8} (FERM Domain Containing 8). As the CTCF (CCCTC-binding factor) transcription factor is well-known for its ability to insulate the transcriptional regulatory mechanisms of one gene from other genes nearby, which is perhaps mediated by looping of genomic DNA (reviewed in Dunn and Davie, 2003; Phillips and Corces, 2009), the CTCF site is highlighted in Figure 4.2 as an indication of the boundary of \textit{NEAT1} upstream regulatory regions. Notably, a large number of TFBS can be seen in cluster A and B immediately adjacent to the TSS of NEAT1, which is also the location of the CpG island (the green bar). Manual counting of the TFBS in the entire upstream regulatory region revealed 82 transcription factors, with many having multiple binding sites across the region in various clusters (Appendix 4, Supplementary Table S4.1A).

In order to understand the cellular schemes upstream of these TFs, gene ontology and functional annotation analysis was performed on the DAVID (database for annotation, visualization and integrated discovery) database (Huang et al., 2008; Huang et al., 2009) (Supplementary Table S4.1B in Appendix 4). This revealed a large number of cellular and molecular functions that these TFs are associated and enriched with. Firstly, these TFs are a mixture of transcriptional activators (36.6%) and repressors (22%). Secondly, a large range of biological processes were enriched with strong statistical significance (FDR (False discovery rate) ≤ 0.1), however some of them are contradictory in their functions, such as the ‘regulation of cell proliferation’ vs. ‘regulation of apoptosis’, or ‘regulation of cell cycle’ vs. ‘regulation of cell differentiation’. Thirdly, pathway
analysis also showed enrichment of a number of pathways that vary greatly in cellular functions, thus making it hard to judge which is more biologically significant. All pathways that are potentially enriched were shown in Table S4.1B in Appendix 4, even if some of them did not meet the arbitrary FDR cut-off of 0.1. Note that the top enrichment of the PANTHER (protein analysis through evolutionary relationships) and KEGG (Kyoto encyclopaedia of genes and genomes) databases are ‘Pathways in cancer’ and ‘PDGF signaling pathway’ respectively, both indicating relationships with highly proliferative cells. Finally, the tissue specificity of these potential TFs has shown a significant enrichment in the epithelium, associated with 36.6% of TFs analyzed.

Taken together, the GO and pathway enrichment analysis based on the ENCODE TFBS data of the human NEAT1 gene showed a potentially highly complex regulatory network behind the expression of NEAT1. There is some enrichment in proliferative pathways and tissues, but contradictory cellular processes have also been found to regulate NEAT1 expression, meaning NEAT1 expression is potentially a balanced result of cross-talk between them.
Figure 4.2: UCSC genome browser tracks of the human \textit{NEAT1} upstream region, see \url{http://genome.ucsc.edu/}. Tracks shown from top to bottom are: 100 vertebrate conservation, UCSC genes, CpG islands, ENCODE transcription factor binding sites (TFBS, 161 factors). Color depth of the TFBS indicates the proportion of the cell lines examined that the site was detected (the darker the color, the higher the proportion). Broadly, six regions of TFBS were named from A to F. The arrow indicates the TFBS of CTCF, which marks the boundary of the upstream regulatory region of NEAT1.
4.2.3 The transcription NEAT1 is highly active

Next, I investigated the role that each TFBS cluster may contribute to NEAT1 transcription, as well as to obtain a general idea how strong NEAT1 is expressed comparing to a standard ubiquitously expressed SV40 promoter. Based on the coordinates of each cluster, the six clusters were cloned in incremental order from A to F as connected DNA fragments, resulting in five different lengths, into a luciferase reporter vector pGL3-basic (Figure 4.3A).

The cloning of the five constructs required different strategies (Figure 4.3A). Construct ‘A-B’, ‘A-C’, and ‘A-D’ were individually amplified from human genomic DNA through nested PCR. In these cases, a slightly larger DNA fragment that contained the desired sequence was first amplified from genomic DNA. This product then served as a template for the second PCR for the desired sequence inside. This strategy was found to effectively improve the specificity of the final PCR product, and help avoid difficult PCR amplifications of GC-rich sequences directly from genomic DNA. Due to difficulties in cloning cluster ‘A’ alone, cluster ‘A’ and ‘B’ had to be merged into the same construct. Finally, each of these constructs was then ligated into the pGL3-basic vector, using KpnI and Ndel restriction sites, ahead of the promoter-less luciferase gene in the multiple cloning sites (MCS) region. For cloning of cluster ‘A-E’, a unique EcoRV restriction site in construct ‘A-D’ was used to ligate the PCR product of cluster ‘E’. Finally, cluster ‘F’ was amplified from the genomic DNA and integrated into the ‘A-E’ vector at its unique KpnI site using In-Fusion cloning. Note that cluster A, which has a CpG island and TATA box, was viewed as the promoter of NEAT1 with strong confidence (Figure 4.2, Chapter 3), therefore must be included in all constructs and cloned into the pGL3-basic vector.

Next, to analyze the transcriptional regulatory property of each cluster, these cloned luciferase vectors were each transfected into HeLa human cervical cancer cells (Figure 4.3B). HeLa cells were chosen because NEAT1 has been highly expressed and it is also the cell lines that the first discovery of paraspeckles was made (Fox et al., 2005). A pGL3-control plasmid, which uses a SV40 promoter to control the luciferase gene, was
also included for a relative comparison to indicate how strong the NEAT1 transcriptional activity is. The results showed that cluster ‘A-B’, although containing a large number of TFBS, could not drive the expression of the luciferase gene to a greater extent than the SV40 promoter. However, once cluster ‘C’ was included, a significant increase in luciferase expression could be seen, suggesting effective transcriptional enhancement caused by cluster ‘C’. Clusters ‘A-D’ showed a slightly lower activity than ‘A-C’, but this was not significant, perhaps suggesting a weak repressive role carried by cluster ‘D’. Clusters ‘A-E’ and ‘A-F’ showed a similar level of transcriptional activation to cluster ‘A-C’, despite the inclusion of a much greater number of TFBS in the two larger constructs. Taken together, the transcriptional activity driven by the entire human NEAT1 promoter/enhancer region could reach a minimum of 2-fold above the SV40 promoter activation in HeLa cells, indicating an overall strong potential for transcriptional activation.
Figure 4.3: Cloning and analysis of human NEAT1 upstream regulatory region.  (A) Cloning of regulatory regions.  Primers used for cloning are represented as code names (see Table 2.5, Appendix 2.2). Red codes represent primers used for generating initial products of nested PCR.  Black codes represent primers used for generating actual cloned DNA fragments.  Restriction sites used for the cloning of each fragment are labelled in the graph on each end of the fragment.  (B) Luciferase reporter analysis of human
NEAT1 upstream regulatory regions of various length in HeLa cells. Asterisks on each bar represent statistical likelihood of difference when comparing each bar to the bar of Cluster ‘A-B’. Asterisks between bars represent statistical likelihood of difference when comparing between the two connected bars. The data is represented as mean±SD of three independently repeated experiments. Student t-test: *p-value ≤ 0.05, **p-value ≤ 0.01, ***p-value ≤ 0.005, n.s: not significant.
4.2.4 NEAT1 is induced by acidosis and serum starvation in specific cell lines

To further understand the signaling pathways leading to transcriptional upregulation of NEAT1, I examined cell culture conditions in which NEAT1 was robustly upregulated. The ENCODE RNA-Seq data had already provided both human and mouse genome-wide transcriptomic information for different cell lines, with or without various treatments. Examination of these data demonstrated that NEAT1 was frequently upregulated when cells were differentiated by serum deprivation treatment (Figure 4.4). For example, after 7 days of differentiation of human LHCN myoblasts into LHCND myotubes through serum starvation, both isoforms of NEAT1_1 and NEAT1_2 were greatly upregulated (Figure 4.4A) (Zhu et al., 2007). Similarly, 60-hours of induced differentiation of mouse C2C12 myoblasts to myocytes with serum deprivation also results in increased NEAT1 expression of both isoforms (Figure 4.4B). Furthermore, similar NEAT1 induction can also be found in a completely different type of cellular transition, such as from 10T1/2 fibroblasts to fibrocytes, also induced by serum deprivation. In each of these cases it is difficult to discern whether it was serum-reduced media, differentiation or stress associated with the treatment that was the primary cause for the induction of NEAT1.

Interestingly, in an accidental bacterial contamination of murine Neuro2A neuroblastoma cell culture, I noticed an obvious increase in the prevalence of paraspeckles (data not shown), which is commonly regarded as linked to NEAT1 upregulation (Hirose et al., 2014; Mao et al., 2011; Sunwoo et al., 2009). The contamination caused a marked reduction of the pH of the media, and cell death, before the remaining cells were imaged for paraspeckles. This was intriguing as the paraspeckle induction may have been related to several factors at the same time: firstly, the drop in media pH as a result of the growth of microbes; secondly, the rapid reduction of nutrients in the media; thirdly, the presence of toxic microbial molecules; and lastly, the infection of microbes into the cultured cells. Among the four, the easiest one to test was whether the reduction of media pH could induce NEAT1. Media with different acidic pH was tested (data not shown), and it was found that pH 6.2-6.5 was able to induce NEAT1 significantly by 48 hours, reaching 3.5-fold when counting both isoforms of NEAT1 together (Figure 4.5A). In addition, neuro2A cells are known to be able to differentiate, at least morphologically
as determined by neurite outgrowth, via serum starvation (Evangelopoulos et al., 2005). This treatment was also shown to induce NEAT1, to an even greater extent than acidosis, reaching almost 8-fold by 48 hours (Figure 4.5A).

In order to determine whether acidosis and serum starvation are able to induce NEAT1 in a broad number of cell types, three other human cell lines were also tested. Here, HeLa cells showed significant NEAT1 induction only under acidosis, but not with serum starvation (Figure 4.5B). Both primary human dermal fibroblasts (hDF) and human mesenchymal stem cells (hMSCs) failed to show NEAT1 induction of more than ± 2 fold by both treatments (Figure 4.5C, D). For acidosis, none of the cell lines presented here had upregulated NEAT1 to the same level as U2OS cells, which was at least 4-fold for total NEAT1 level and around 20-fold for the NEAT1_2 isoform (see Chapter 3). Taken together, these data indicated that the treatments capable of inducing NEAT1 may be very cell line specific and treatment specific simultaneously, and the extent of induction also has to be measured empirically.
Figure 4.4: NEAT1 was upregulated significantly in differentiated cell lines as shown by ENCODE RNA-seq tracks of human and mouse NEAT1 expression. (A) Human LHCN (human skeletal myoblasts) and LHCND (LHCH differentiated to myotubes) cell lines. Myogenic differentiation was initiated by switching cells to serum-free medium. Samples were taken after 7 days treatment. (B) Mouse C2C12 myogenic differentiation and 10T fibroblast-to-fibrocyte differentiation initiated by 2% serum deprivation. Samples were taken at 60 hours after treatment. Data available on http://genome.ucsc.edu.
Figure 4.5: RT-qPCR of endogenous total NEAT1 and NEAT1_2 expression after different treatments. Acidosis (pH 6.2-6.5) and serum starvation of serum-free media. (A) Mouse neuro2A neuroblastoma cell line, (B) human HeLa cervical cancer cell line, (C) human dermal fibroblast primary cell line and (D) human mesenchymal stem cells, treated with acidosis and serum starvation. NEAT1 expression was normalized to U6 RNA. Data is represented as mean ± SD of the representative experiments of three biological replicates in each case.
4.3 Discussion

This is the first study to systematically examine the transcriptional regulation of the human \textit{NEAT1} gene. Although previous reports, published during the course of this PhD project, characterized the function of specific transcription factors, such as HIF-2α and ER in regulating \textit{NEAT1} expression, none of these addressed \textit{NEAT1} transcriptional regulation at a systems biology level (Chakravarty et al., 2014; Choudhry et al., 2014). In the present study, by utilizing publicly available data, a glimpse of the complexity of the \textit{NEAT1} regulatory mechanism can be readily obtained. Firstly, the suppression of \textit{NEAT1} expression in pluripotent stem cells and embryonic stem cells correlates with epigenetic control via both DNA methylation and repressive histone modifications (Figure 4.1). However, in cells that are differentiated and expressing \textit{NEAT1}, these epigenetic locks are no longer present. Secondly, the large numbers of transcription factors that are detected by the ENCODE TFBS studies have shown functional enrichment in a wide variety of signaling pathways and biological processes (Figure 4.2, Supplementary Table S4.1 in Appendix 4). Some enriched cellular processes are even contradictory to each other. This observation highlights the possibility that \textit{NEAT1} expression could be more of a reflection of overall cellular activities underlining the biological and environmental conditions.

\textit{NEAT1} expression is reported to respond dynamically to a change of environmental conditions, as well as being ubiquitously expressed (Chakravarty et al., 2014; Choudhry et al., 2014; Gibb et al., 2011; Gremlich et al., 2014; Hirose et al., 2014; Mao et al., 2011; Nakagawa et al., 2011; Sunwoo et al., 2009; Zhang et al., 2013). Our combined knowledge of \textit{NEAT1} transcriptional regulatory components fits with the notion that the transcription factors that control \textit{NEAT1} are either shared by a large range of tissues, and/or, that each subset of tissues uses a different subset of transcription factors. \textit{NEAT1} would not be as dynamically expressed as is commonly observed if only a few housekeeping pathways were controlling \textit{NEAT1} expression. Moreover, \textit{NEAT1} would not be so ubiquitously expressed if only a few pathways with specific functions were controlling \textit{NEAT1} expression. Interestingly, a large number of TFs bound to the \textit{NEAT1}
promoter are expressed in epithelial tissue (Supplementary Table S4.1B in Appendix 4), providing evidence as to why NEAT1 can be particularly abundant in those tissues, with a notable example as the surface layer of cells in the mouse forestomach (Nakagawa et al., 2011).

Identifying the influential enhancer or specific transcription factors that govern NEAT1 expression is required to understand how this regulatory system functions. Chapter 3 outlined how an influential transcription factor, p53, was identified to play an activatory role in NEAT1 expression. In this chapter, a large scale promoter/enhancer analysis of NEAT1 was performed. Promoter analysis is typically a procedure for identifying important activating and inhibitory TFBS that regulate the transcription of a gene. Traditionally due to the lack of information about the TFBS locations at the genomic scale, the approach was to clone a relatively small region (~1kb) of the suspected regulatory sequence and then create deletions 10-50bp at a time and monitor either loss or enhancement of transcription of the reporter gene. Clearly, such a strategy is redundant after the release of the ENCODE TFBS data, provided the gene of interest is expressed in one of the cell types included in the ENCODE study. Knowing the location of each TFBS cluster in the NEAT1 gene promoter, one can specifically clone them and investigate their influence on gene expression. In this study, six clusters of TFBS were analyzed and each displayed different levels of transcriptional activation. Cluster A and B have features of a traditional promoter and did not display strong transcriptional activity. Cluster C displayed strong activatory properties, and cluster D was slightly inhibitory. Finally cluster E and F did not enhance transcription beyond what was already shown by cluster C. This suggests the NEAT1 transcriptional activity is not simply controlled by the number of transcription factors involved, but also which ones of them are involved. Note that although cluster E and F could not enhance reporter gene expression as strongly as cluster C, the conservation of these regions is still much higher than the surrounding DNA, indicating some value that could not be revealed in this system, but may nevertheless exist (Figure 4.2). This suggested that the importance of E and F may not be in enhancing NEAT1 transcription, but rather conferring fine-control of NEAT1 expression, perhaps in a tissue-specific manner (tissue specific enhancer usage has been
reviewed in Ong and Corces, 2011. It will be important in the future to repeat these experiments in additional lines to HeLa cells, to determine the extent of cell-type influences. In addition, detection of chromatin structure using 3C, 4C, hi-C methods may help us identify distal enhancers and chromosomal looping which could also play important roles in regulating NEAT1 expression.

The inconsistent induction of NEAT1 by different treatments in different cell lines again indicates the dynamic nature of NEAT1 expression. Serum starvation was used as a treatment to induce cellular differentiation, and it was shown to induce NEAT1 in at least two types of muscular differentiation models in human and mouse ((Sunwoo et al., 2009), Figure 4.1). Serum starvation is also a well-known treatment for inducing cell cycle arrest and is frequently used for synchronization of the cell cycle (Cooper, 2003; Kues et al., 2000). In response to serum starvation, neuro2A cells develop elongated neurites and stop cell proliferation (Evangelopoulos et al., 2005). Interestingly, previous studies has shown p53 activation under serum starvation (Shi et al., 2012; Tamir and Bengal, 1998). This certainly raises a possibility that the cluster C enhancer region of NEAT1 in human, which was bound by p53 and found important in driving NEAT1 upregulation under acidosis in chapter 3, could also be responsible in driving NEAT1 upregulation in serum starvation in tested cells in this chapter.

Acidosis of media, caused by adding extra hydrochloric acid, was identified as a novel environmental inducer of NEAT1 in this study. Due to the strong induction of NEAT1 under acidosis, this may have been the primary factor in the upregulation of NEAT1 in the accidental bacterial contamination of Neuro2A cells. Interestingly, an acidic environment is highly physiologically relevant in both the mammalian reproductive tract (reviewed in Suarez and Pacey, 2006), pertinent with regards to the phenotype of the NEAT1 knockout mouse, as well as the stomach, pertinent due to high NEAT1 expression epithelial cell layer of the stomach in wild-type mice (Nakagawa et al., 2011). Acidosis is also associated with the tumor microenvironment, arising as a result of the lactic acid secreted by the tumor cells due to general hypoxic conditions. The acidic environment is thought to promote degradation of the extracellular matrix, thereby allowing greater
cell migration and invasion (Estrella et al., 2013; Kato et al., 2013). However, it is not yet clear how NEAT1 upregulation with acidosis in U2OS and HeLa cells is directly linked to these physiological effects of acidosis in a whole animal, or tumor, system.

It is also interesting to consider how acidosis and serum starvation induced NEAT1 upregulation dramatically in some cell types, but not in others: Neuro2A cells were responsive to both treatments; HeLa cells were only responsive to acidosis, but not serum starvation; and finally human dermal fibroblast and mesenchymal stem cells were not sensitive to either treatment. As continual acidosis of pH6.2-6.5 is sufficient to eventually induce extensive cell death in all cell types examined (data not provided), this suggests the acidic environment has exerted a common stress on all cell types. Thus each of these cell lines shared a common stimulation, yet responded differently, with one major difference being the change in the expression level of NEAT1.

One explanation for this is that cells may have more capability of upregulating NEAT1 when its expression is relatively weak in the original unstressed state. A high ‘basal’ expression of NEAT1 is perhaps closer to a maximum expression limit, hence having less potential for further induction. According to our and previous studies, neuro2A cells do not have many paraspeckles, which is an indirect indication of lower basal NEAT1 expression (Clark et al., 2012). However, each of the other human cell lines tested have significantly more paraspeckles than neuro2A cells even in the unstressed state (data not provided, (Fox et al., 2002; Hirose et al., 2014)). As strong paraspeckle prevalence is normally associated with high NEAT1 expression, this is a good indication that the human cell lines tested in this study already express NEAT1 at a high level. To confirm this, and to figure out the maximum NEAT1 expression level for each cell type, absolute NEAT1 RNA levels would need to be determined in both the normal and induced state for each cell type. However, this concept does not explain why HeLa cells induce NEAT1 under acidosis, but not serum starvation. Another possibility is that the acidosis and serum starvation treatments are fundamentally very different stresses, triggering different biochemical responses. For example, extracellular acidification would result in the accumulation of intracellular nitric oxide, but the same would not happen for serum
starvation (Capellini et al., 2013). On the other hand, the p53 pathway can be activated by both conditions (Ohtsubo et al., 1997; Shi et al., 2012). Furthermore, even if different cell lines could activate the same pathways under different conditions, the actual functional status of the pathways might vary. For example, neuro2A cells have a wild-type p53, yet HeLa cells have a mis-regulated p53 that interacts with E6 viral protein and undergoes rapid degradation, hence is not functional (Hengstermann et al., 2001; Hoppe-Seyler and Butz, 1993). Future experiments in primary cell lines should be able to address the inconsistent state of p53, but the regulation of NEAT1 expression could still vary greatly between cell types due to the intrinsic biological differences in their regulatory networks.

In summary, this chapter has demonstrated the complexity of the upstream regulatory mechanism that controls NEAT1 expression. This complexity forms the basis for the dynamic and yet ubiquitous nature of NEAT1 expression.

(see Appendix 5 for references)
Chapter 5
Using genome engineering to create permanent isoform-specific NEAT1 knockout cell lines
5.1 Introduction

Knockdown (KD) or knockout (KO) is an extremely effective approach for investigating the function of a gene. The most widely used method is the transient transfection of either small-interfering RNA (siRNAs), or antisense oligos (ASOs) into cells, resulting in transient RNA interference and hence degradation of the RNA transcripts of the target gene. In NEAT1 and paraspeckle research, transient KD of NEAT1 has been used to show the essential role of NEAT1, especially the NEAT1_2 isoform, in the formation and maintenance of paraspeckles (Clemson et al., 2009; Mao et al., 2011; Naganuma et al., 2012; Sasaki et al., 2009; Sunwoo et al., 2009). In mouse models, generation of stable NEAT1 KO mice was created by a knock-in of a lacZ-poly(A) reporter gene 5’ of the transcription start site of the mouse NEAT1 gene (Nakagawa et al., 2011). These mice have provided valuable observations of the phenotypic role of NEAT1 in vivo, especially within the reproductive system of females (Nakagawa et al., 2011; Nakagawa et al., 2014; Standaert et al., 2014).

However, prior to this project, no human stable NEAT1 KO cell lines had been generated. Stable gene KO has a number of advantages over transient approaches. Firstly, as a stable system, the KO effect is permanent, making it possible to study the long-term effects of the KO of the gene of interest. Secondly, when achieved with genomic editing, the efficiency of KO can often be close to be 100%, which is rarely achieved by any transient approach especially on highly abundant transcripts. Thirdly, the cost of creating stable gene KO might be higher initially, but as the effect of KO is permanent once achieved, over the long-term the cost is lower than a transient KD system. In the case of NEAT1, its strong transcription and high turn-over rate might be particularly ill-suited for transient methods to obtain long-lasting KD. To solve this problem, a stable system which allows long term observation of the effect of NEAT1 removal was needed.

There are a number of ways to achieving stable expression inhibition by targeting different stages of gene expression. At the RNA level, a shRNA expression cassette (small hairpin RNA) can be introduced into a random location in the genome of target
cells. Constitutively expressed shRNA will be first processed into siRNA by DICER (Bernstein et al., 2001; Billy et al., 2001; Doi et al., 2003; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001; Knight and Bass, 2001; Myers et al., 2003; Paddison et al., 2002; Provost et al., 2002; Zhang et al., 2002), and then used for RISC/AGO2 (RNA-induced silencing complex) mediated degradation of the RNA target (Martinez et al., 2002). Interestingly, siRNA-mediated KD is efficient at degrading nuclear-restricted RNAs, despite being thought to act primarily in the cytoplasm, perhaps reflecting new observations of RNAi machinery within the nucleus (Gagnon et al., 2014).

At the DNA level, the development of genomic editing tools has allowed a much wider range of manipulation to be performed with great precision, such as creating mutations for loss-of-function, insertion and deletion of DNA of all sizes at a given genomic location. Genomic editing tools are essentially molecular complexes that induce either double or single stranded DNA breaks (DSBs, SSBs) at a targeted location. When cells detected DNA damage, they utilize either the homology-directed repair (HDR) pathway or the non-homologous end joining (NHEJ) pathway to re-join the broken DNA strands (reviewed in Smith et al., 2007). The HDR pathway uses homologous sequences as a repair template to prime repair synthesis for both DSBs and SSBs. The process is very complex and not entirely understood, but the end-product is an accurate ‘transfer’ of the sequence information presented in the repair template into the damaged DNA. Therefore, if the repair templates do not faithfully copy the DNA sequences present before the damage, then the HDR pathway can transfer modified information non-reciprocally from the donor DNA to the recipient DNA, a process termed gene conversion. Based on this phenomenon, precise genomic editing can be achieved through the HDR pathway, with a repair template introducing the desired modifications. Providing these exogenous artificial repair templates in abundance, will increase the chance of their use, instead of the endogenous sister chromatid to promote the repair. In contrast, the NHEJ pathway is more prone to error when used to correct DSBs. It typically involves alignment of one or a few complementary bases to direct repair, leading to small deletions and sometimes insertions (indels) (reviewed in Bétermier et al., 2014). Thus, the NHEJ pathway is useful for introducing random indels to disrupt the protein-coding reading
Together, these repair pathways provide an opportunity for different genetic modifications to be incorporated into the genomic DNA. However, the bottleneck in genome editing has historically been inducing the DNA damage with precision and efficiency.

CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats, CRISPR associated protein 9) technology has rapidly gained popularity as one of the most easy-to-use genomic editing tools. Unlike other options such as TALENs (transcription activator-like effector nucleases) or ZFNs (zinc-finger nucleases), which rely on the recognition of carefully designed DNA-binding proteins towards target DNA as guides for the attached nuclease, the CRISPR-Cas9 system relies on the complementary binding between a designed guide RNA and the target DNA (reviewed in Kim and Kim, 2014). CRISPR is a widely observed feature of the ‘adaptive immune system’ in natural bacteria and archaea, to fight against viral infections (reviewed in Horvath and Barrangou, 2010). Essentially, it is a group of RNA-interference systems that uses complementarity between RNA and foreign DNA as guides for DNA nuclease mediated cleavage. The name CRISPR derives from the way RNA sequences are stored in the bacterial and archaeal genome. Prokaryotes, under constant exposure to exogenous DNA via transduction, conjugation, and transformation, have developed the CRISPR system to recognize and distinguish foreign DNA from their own. Exposure to the exogenous DNA, such as by infection of phage, will trigger bacteria to use a number of Cas (CRISPR associated) proteins to immunize them by incorporating fragments of the foreign DNA into the CRISPR repeats (Barrangou et al., 2007; Beloglazova et al., 2008; Wiedenheft et al., 2009). These CRISPR repeats are essentially arrays of short DNA from the exogenous source, termed ‘Spacer’, separated by repeats in the bacterial genome. To contribute to the immune defense, CRISPR repeats are transcribed into a long CRISPR transcript (pre-crRNA), and then subsequently processed into small RNA molecules that correspond to a spacer flanked by two partial repeats (crRNAs) (Brouns et al., 2008; Carte et al., 2008; Hale et al., 2008). CrRNAs are then incorporated into effector complexes, which consist of a number of Cas nucleases, to degrade invading DNA (reviewed in Terns and Terns, 2011). There are many pathways of CRISPR-activation, but one of them requires a
trans-activating crRNA (tracrRNA) to forming a 24-nt complementarity with the repeat region of pre-crRNA, followed by RNaseIII cleavage to generate mature crRNA (Deltcheva et al., 2011). This pathway requires a very small number of components, with only the expression of tracrRNA, pre-crRNA, RNaseIII, and Cas9 nuclease required for the cleavage of foreign DNA both in eukaryotic cells (Gasiunas et al., 2012; Jinek et al., 2012) and in prokaryotic cells (Magadán et al., 2012; Sapranaukas et al., 2011).

The current CRISPR-Cas9 system for genomic editing, which is derived from the *Streptococcus pyogenes* type II CRISPR system, contains the very same four components aforementioned. Other than RNaseIII, which is provided by the host cell, the other three components have to be transfected into the cell. To avoid complications of co-transfection of each component, all three components have now been integrated into a single plasmid (Cong et al., 2013). By changing the sequence of the pre-crRNA, which is also known as the guide RNA (gRNA), one can theoretically target the Cas9 nuclease to any location in the genome of the host cell. Both the tracrRNA sequence and Cas9 are consistent in all experiments in this system, which makes it versatile because the only thing required for the user to change is the sequence of gRNA. This is significantly easier and faster than altering protein coding sequences in ZFN and TALEN approaches. The only complication is that a ‘PAM’ (protospacer adjacent motif) sequence in the host genome is required at the 3’ end of the gRNA. The PAM sequence is normally a ‘NGG’ trinucleotide that presents in the host genomic DNA, and is essential for the Cas9 to recognize before mediating the DNA break (Jinek et al., 2012; Mojica et al., 2009; Shah et al., 2013). The PAM site, together with the sequence up to 11 bp 3’ of the gRNA, contributes to the specificity of the CRISPR-Cas9 system (Cong et al., 2013; Jinek et al., 2012; Sapranaukas et al., 2011). Therefore, mismatches in the gRNA in this region and lack of PAM will considerably hinder the effectiveness of this system.

The great versatility of the CRISPR-Cas9 system suggested it as a valuable tool to consider creating stable NEAT1 KO cell lines. However, at the start of the experimental work for this chapter in mid-2013, there were no reports on the application of this system to create stable KO of lncRNAs. Fortunately, the ZFN system had already been
successfully used to KO the lncRNA MALAT1 in human cell lines, achieving a greater than 1000-fold reduction in MALAT1 RNA (Gutschner et al., 2011). That work provided important guidance for the application of CRISPR-Cas9 in this current study. Specifically, a ZFN was used to target and create a DSB in the space between the TATA box and the transcription start site (TSS) of the endogenous MALAT1 gene. Once that was completed, the authors took advantage of the HDR pathway, and knocked-in a GFP expression cassette with various polyadenylation signals. The insertion of a transcriptional termination signal (TTS) 5’ of the MALAT1 TSS had a powerful inhibitory effect on MALAT1 transcription. Considering MALAT1 and NEAT1 are neighboring lncRNA genes, which are both highly abundant and have TATA boxes in their promoter regions, I decided to use the CRISPR-Cas9 system to adopt the same knock-in strategy for NEAT1 KO.
5.2 Results

5.2.1 Knocking out human NEAT1 using the CRISPR-Cas9 system

Part I: Design and preparation of materials

To achieve long-term knockout of NEAT1, the CRISPR-Cas9 system and HDR were used to permanently alter the original gene structure of NEAT1. The approach was designed to ensure NEAT1 was no longer being transcribed, and, even if it were transcribed, it would no longer carry its original sequence and structure. The end result was predicted to be both a loss-of-expression and loss-of-function of NEAT1. I therefore adapted the strategy that was described in (Gutschner et al., 2011), to integrate a reporter gene expression cassette in-between the transcriptional start site of a lncRNA gene and its upstream promoter elements. In my adaptation of this method, a YFP (yellow fluorescent protein) expression cassette, controlled by a CMV (cytomegalovirus) promoter and SV40 (simian virus 40) terminator, was to be integrated into that location of the human NEAT1 gene through HDR (Figure 5.1A). The fluorescent reporter was for downstream selection, and its targeted integration would also greatly extend the distance between the transcriptional start site (TSS) of NEAT1 and its upstream transcriptional regulatory elements. The insertion of the SV40 terminator before the original TSS of NEAT1 is also critical to prevent any transcriptional machinery from transcribing NEAT1. Furthermore, in case of any failed SV40 termination, the new NEAT1 transcript would carry an additional YFP mRNA at its 5’ end, and thus would have an altered three dimensional structure and loss-of-function. The HDR would ensure precise delivery of the sequence carried in the repair template to be found in the broken genomic sequence after the repairing. In this case, the TATA box of the original NEAT1 promoter was to be lost after the integration because it was not included in the left homology arm of the repair template DNA (Figure 5.1A). This design would act as a secondary block to prevent any potential effect of the upstream NEAT1 enhancers on the transcription initiated by this TATA box.

In order for the desired integration to take place with high efficiency, a targeted dsDNA break needed to be induced at the desired site in the genomic DNA. To induce the break,
Figure 5.1: Design of CRISPR-Cas9 mediated NEAT1 knockout cell line. (A) A graphic representation of the NEAT1 knockout strategy. Two CRISPR-Cas9 induced double stranded breaks (cut v1 and v2) were to be generated close to the transcription start site (TSS) of NEAT1, each cut site was made by an individual gRNA, in a separate transfection. Each cut site was repaired resulting in an insertion of the YFP expression cassette through HDR. The TA TA box (red) of the NEAT1 promoter is expected to be removed after this process. (B) Detail of the gRNAs that correspond to each cut. The top five most likely unspecific targets were identified by http://crispr.mit.edu/ (Hsu et al., 2013), and the mismatches to the target sequence are highlighted. (C) The cloning of the gRNAs into the px330 plasmid, which allows the gRNA to fuse with the tracrRNA in a chimeric structure. The gRNA DNA oligos were first annealed into dsDNA, and then ligated into the BbsI sites in the vector. (D) Cloning of the homologous repair template. Each homology arm was inserted into a restriction site that flanks an expression cassette of YFP, controlled by a CMV promoter and SV40 terminator.
I designed the CRISPR-Cas9 system to generate a cut immediately next to the TSS of NEAT1 (Figure 5.1A). The CRISPR-Cas9 system relies on a 20-ribonucleotide guide RNA (gRNA) to target Cas9 to induce double stranded breaks in DNA. Certain constraints are imposed due to the requirement of a PAM site, but even so, there are numerous possible gRNAs for any given experiment. Therefore, the most optimal gRNAs need to be selected based on certain criteria. The most important parameter to consider is that each gRNA will display various degrees of non-specific binding to unwanted loci in the target genome. This was evaluated using a CRISPR gRNA design tool (http://crispr.mit.edu, (Hsu et al., 2013)) to determine potential off-target sites of each gRNA that can be possibly used (Figure 5.1B). To choose the best gRNAs, three main criteria were considered: the lowest possible non-specificity score (as generated by the software based on the contributions of mismatch location and density in the gRNA), the identity of the off-target sites (coding vs non-coding region), and the absence of binding sites within the repair template.

For the NEAT1 promoter targeting experiment, two gRNAs were chosen that had different potential off-target activities. Although neither gRNA had potential off-target effects within annotated genes, I reasoned that using these two gRNAs in individual experiments would control for non-specificity in other regions of the genome. Figure 5.1A shows the cut sites in the human NEAT1 gene targeted by each gRNA. Figure 5.1B shows the top five off-target sites of each gRNA, and the location of sequence mismatches.

The gRNAs were cloned into the px330 vector (Figure 5.1C). This plasmid fused gRNA with the tracrRNA, thus creating a chimeric RNA that functions at a higher efficiency. The px330 plasmid also contains the expression cassette for Cas9 (Cong et al., 2013). To construct repair templates for homologous recombination, a peYFP-C1 vector backbone that carries a YFP expression cassette was used (Figure 5.1D). Homology arms on each side of the cut site were PCR amplified from the genomic DNA of a healthy human, which was a kind gift from the laboratory of Professor Grant Morahan (Harry Perkins Institute of Medical Research, WA). The left and right arms were separately
cloned into the *Ase*I and *Mlu*I sites flanking the YFP expression cassette through Gibson cloning. The sequence of the arms matched exactly to the nucleotides in the genomic DNA flanking the cut sites, with the exception of the TATA box deletion (Figure 5.1A).

**Part II: Transfection and sorting**

For each CRISPR experiment, two plasmids were transfected in an equal ratio. The first plasmid was the px330 vector consisting of the gRNA, tracrRNA, and Cas9. The second plasmid was the YFP expression plasmid with arms homologous to the desired cut site for repairing the targeted DSBs induced by the first vector (Figure 5.2A). Unsurprisingly, the transfection caused a rapid transient expression of YFP in the vast majority of cells (data not shown), with the presumption that at the same time a transient window of CRISPR-mediated genome engineering and HDR would result in the targeted disruption of NEAT1 in some cells. Like any transfection of a reporter expression plasmid, the random integration of the YFP into the genome was likely to occur at a very low rate. In addition, due to the presence of homologous arms flanking the YFP expression cassette, targeted integration was also expected to occur at a very low rate without the CRISPR system. The introduction of the CRISPR induced cut was to specifically increase the rate of success for the desired genomic modification, but not the random integration of the repair template into the host genome.

A detailed workflow based on the application of one of the gRNAs (v2), is shown in Figure 5.2. The following key points should be noted:

1. **U2OS** cells (chosen as explained in Chapter 3, page 98) were transfected with an equal ratio of the two plasmids. The transient expression of YFP slowly diminished over the six days following transfection, at which point confluent cells were expanded into a T175 flask to increase the population of cells that remained fluorescent.

2. Once cells were 100% confluent, they were subject to FACS (sort 1). In this case, 5.3% of the total cell population, being those cells with highest YFP fluorescence, were sorted for further expansion. Some of these cells were expected to have integrated the YFP stably, either randomly into the genome, or into the NEAT1 promoter site.
Figure 5.2: Workflow of the CRISPR NEAT1 knockout experiments.  (A) A flowchart downstream of the transfection of cells using the CRISPR guides and repair templates.  FACS graphs are actual data gathered when using the gRNA_v2.  Optional PCR checking steps are highly recommended for early detection of successful recombination.  (B) A schematic representation of the PCR checking strategy for the detection of the successful recombination event, where one primer binds inside the inserted YFP (green
arrow) and the other primer binds in the genomic DNA (blue arrow) that is not covered by the homology arm. (C) The semi-quantitative PCR checks indicated an increasing population of cells that carry the correctly inserted YFP expression cassette. An equal amount of genomic DNA template (40ng) was used for each reaction.
3. Once the sort 1 cells became confluent in the large T175 flask, an important but optional step was to extract the genomic DNA of a small proportion of these cells, and to use PCR to verify if the targeted integration had occurred as expected. The strategy of the primer placement is shown in Figure 5.2B. In short, the PCR should only take place if the YFP is integrated into the desired genomic location. The more cells that carried this YFP insertion, the greater the abundance of the PCR product. This step gave very early confidence for the likelihood of the CRISPR experiment succeeding, an important factor prior to commitment to the downstream labor-intensive screening procedure.

4. After repeating the fluorescent sorting 3 times, the vast majority of cells left (76.2%) had been stably fluorescent for approximately two weeks. In the meantime, the PCR check yielded an increased amount of specifically amplified product (Figure 5.2C). This suggested the proportion of cells that carried the targeted integration had been amplified.

5. As there was no guarantee that all cells that were fluorescent at this point were carrying stable fluorescence, an optional extra passage was carried out to further reduce the population of cells that were still transiently expressing the YFP. Once the cells were confluent again, three and half weeks post transfection, a diluted cell suspension was seeded into 96-well plates for isolating single cell colonies, each then subject to downstream screening procedures.

**Part III: Screening for homozygous NEAT1 knockout colonies**

Colonies formed from single cells were cultured and expanded until enough genomic DNA could be extracted. This took approximately three to four weeks. As colonies grew at different speeds, and to avoid taking colonies that were too slow to grow and work with, I used four weeks as the cut-off time for PCR screening of cells with homozygous genomic modifications. Genomic DNA was extracted when colonies were confluent in a 12-well plate, and the remaining cells were passaged until the screening was completed.

In order to screen a large number of colonies and to accurately determine whether the
YFP expression cassette had been either randomly integrated into the genome, or targeted to one or both NEAT1 loci (i.e. heterozygous, or homozygous) a PCR based method was used. This PCR method essentially used the results from two separate PCRs (Figure 5.3A). The first PCR primer set (primer set 1, primer code A963 + A964) tested whether the YFP had been integrated into the desired genomic location. The second primer set (primer set 2, primer code A708+A709) was used to determine whether the original genomic DNA had expanded in size due to the targeted insertion of extra DNA (Figure 5.3B). In a random integration event primer set 1 would not yield any product, and the original size of the genomic region would show a 1.2kb band by primer set 2. In a heterozygous line, as the desired YFP integration is detectable by primer set 1, but the original genomic size is also detectable by set 2, both primer set 1 and 2 yielded 1.2-1.3kb products at the same time. In a homozygous line, as both the desired YFP integration sites were detectable by primer set 1, and all copies of the original genomic region were expanded due to the integration, primer set 2 generated a large 3kb band instead of the smaller one. On occasion, due to the PCR inefficiency in the expanded genomic region, primer set 2 PCR did not yield any product.

Part IV: Verification of NEAT1 knockout and summary

The two gRNAs used displayed various degrees of efficiency, with gRNA_v1 having a greater number of desired integration events and fewer random insertions (Figure 5.3C). However, gRNA_v1 was also more likely to have off-target activities than gRNA_v2 (Figure 5.1B).

Homozygous colonies were also found to have variation between them. Their NEAT1 KO efficiency, measured by qRT-PCR, was variable (Figure 5.4A). All of the homozygous clones, except colony 53, displayed at least 90% reduction in both the total NEAT1 level and NEAT1_2 level, with colony 24 reaching more than 99% reduction. Examining RNA from the ‘sort 3’ mixed population, revealed that the reduction of NEAT1 globally had not even reached 50% (Figure 5.4A), even though the genomic PCR indicated the presence of many cells with the correctly integrated YFP (Figure 5.2C). Thus, the homozygous single clone screening was a necessary step for significant NEAT1
Figure 5.3: PCR screening of NEAT1\(^{-/-}\) clonal cell lines. (A) A schematic representation of the PCR screening strategy. Two primer sets were used in individual PCR reactions, performed simultaneously, to allow accurate determination of the recombination events. (B) An example of how different combinations of PCR products reveal how YFP is integrated into the genome. Random integration is when the YFP expression cassette is randomly integrated into one or more unknown sites in the genome. Heterozygous or homozygous represent either one or both targeted NEAT1 loci having taken up the YFP in the desired way. (C) A summary of the number of integration scenarios that have been determined for each gRNA when generating U2OS cell lines.
Figure 5.4: Direct observations of NEAT1 levels and morphology of the NEAT1 knockout colonies. (A) RT-qPCR of the expression of total NEAT1 and NEAT1_2 of selected homozygous NEAT1 knockout colonies, as well as the mixed population of cells after the third FACS (sort 3). Both total NEAT1 and NEAT1_2 were greatly reduced in the vast majority of colonies. Representative data is shown. The morphology of colonies of the NEAT1 knockout U2OS cells are of two distinctive types: compact (B) vs sparse (C).
reduction. Colony 53 is unusual in that the reduction in total NEAT1 transcript levels was not matched by a decrease in the NEAT1_2 transcript. However, despite seemingly incomplete KO, as stated earlier, there is also no guarantee that the original sequence of NEAT1_1 and NEAT1_2 still remained intact following YFP insertion. Nevertheless, to prevent any potential problems, colony 53 was not included in any of the downstream analysis in later sections, or Chapter 3.

Secondly, the morphology of the homozygous colonies varied greatly. In general, two types of colonies were observed from the homozygous NEAT1 KO colonies. Colony 24, 69 and 71 all shared a similar ‘compact’ colony morphology (Figure 5.4B) whereas other clones showed a more sparse appearance (Figure 5.4C). Initially, this was speculated to be a potential outcome of the total NEAT1 KO, but after careful examination of the wild-type U2OS cells, the same two populations were also identified. Thus, if U2OS cells started as a heterogeneous population, with potentially different phenotypes for each sub-population, then single cell clone selection would inevitably amplify these cell-to-cell differences. To overcome this, non-fluorescent single cell clones were grown after the transfection of CRISPR materials and the subsequent three FACS procedures, to generate control wild-type U2OS cells of both the compact and sparse colony morphology. These cells were checked by qRT-PCR to make sure NEAT1 was still expressed normally. These lines were then used as a control for all transfection, passaging, and FACS induced cellular changes.

5.2.2 Preliminary RNA-seq and differential expression analysis of NEAT1 knockout colonies

Isolated total NEAT1\(^{-/-}\) clones, despite obvious differences in their morphology, did not show any other apparent abnormalities in growth (data not shown). In order to quickly pinpoint the physiological aspects of the cells that might be affected by the stable NEAT1 reduction, I decided to infer it from transcriptomic changes. Specifically, it was expected that a bioinformatic analysis of the transcriptomic changes apparent in the NEAT1\(^{-/-}\) cells might help generate predictions for the physiology affected in the knockout
cells. Thus, immediately after the isolation of NEAT1\textsuperscript{-/-} clones from CRISPR gRNA v1 and v2, a relatively ‘crude’ total RNA sequencing experiment was performed. This involved three NEAT1\textsuperscript{-/-} clones of different morphologies compared to the heterogeneous cultured wild-type U2OS cells as a reference transcriptome. Total RNA was extracted from the original wild-type cells (from three independently cultured U2OS passages), and also from three homozygous NEAT1\textsuperscript{-/-} clones named colony 17 (sparse morphology), 24 (compact morphology) and 71 (compact morphology).

Once the total RNA was extracted from the six samples, ERCC spike-in RNA controls were added to each sample to facilitate normalization procedures downstream of the sequencing. Prior to the library preparation, ribosomal RNA was depleted in all samples to allow enrichment for all other RNA species. Approximately 30 million single-end 100bp reads were generated for each sample and mapped back to the hg19 human genome using TopHat2 with default settings (as available on Galaxy platform) (Kim et al., 2013). For counting reads, the HTseq package under the Python platform was used. This software was chosen as its output files were highly compatible with one of the most recently published normalization methods, RUV (removal of unwanted variation) (Risso et al., 2014). Among the variants of RUV methods, RUVg was used for the normalization of counted reads. RUVg took the first five thousand most consistently expressed genes in two conditions (WT vs KO) to calculate the normalization factors for all six samples. In this process, the ERCC controls were taken as a part of these 5000 genes along with others. Normalized read counts of each sample were first analyzed for their statistical reliability before being subject to differential expression analysis by the edgeR package (Figure 5.5A-D). Firstly, the relative log expression (RLE) plots detect the median of all expressed genes in each sample, and were expected to be identical for all samples with small error bars if the two conditions were not expected to have a difference in global gene expression. Figure 5.5A and B shows that very similar RLE plots can be seen on both unnormalized and normalized samples, but the normalized samples have smaller error bars for sample KO1 (colony 24) and KO2 (colony 17) in particular. This indicated that the overall expression of all samples was not biased. Secondly, a principle component analysis (PCA) detected the two conditions
Figure 5.5: RNA-seq data normalized by RUVg indicated multiple cellular processes that could be affected by the NEAT1 knockout. Three separately cultured wild-type U2OS mixed morphology cell populations (WT1,2,3) were compared with three homozygous NEAT1−/− colonies (KO1=colony 24,
KO2=colony 17, KO3=colony 71). (A,B) Boxplots of relative log expression (RLE) for un-normalized and RUVg-normalized counts revealed similar distribution. (The bottom and top of the box indicate the first and third quartiles respectively; the inside line indicates the median; the whiskers are located at 1.5 the inter-quartile range above and below the box). (C,D) Principle component analysis (PCA) scatterplots of the first two principle components for RUVg-normalized counts (log scale, centered) revealed clear segregation of the three WT samples from the NEAT1^−/− samples. (E) Mean-difference plot (MD-plot) for RUVg-normalized counts (log2 scale) revealed a large number of differentially expressed genes (red) to be around 0 on the log2 counts per million (CPM) axis. A positive log2FC value would indicate upregulated expression in KO samples. (F) A pie chart showing around 90% of DE genes were downregulated in the NEAT1^−/− cells compared to WT. Ingenuity Pathway Analysis (IPA) of DE genes suggested the top five most likely affected molecular and cellular functions (G), and associated network functions (H) based on the enrichment of the number of DE genes in each category and their p-value in the DE analysis.
(WT vs KO) could be separated mathematically by looking at the first two dimensions (principle components PC1 and PC2) of the data. Ideally, one should be able to separate the two groups of samples based on PC1 vs PC2 by placing a straight line on the plot to separate the two groups. Before normalization, this line could not be drawn: Although the WT samples formed a cluster with the PC2 value above 0, the data of the KO samples scattered all around the WT samples (Figure 5.5C). After normalization, although neither PC1 nor PC2 alone could be used to separate the two conditions, the two principle components combined still led to the separation of the two groups, indicated by the dashed line (Figure 5.5D). However, this is not a strong mathematical separation between the two groups, meaning the variation between the two groups was not significant.

Following the normalization of read counts, differential expression (DE) analysis was performed using edgeR. The DE genes with a false discovery rate (FDR) of \( \leq 0.1 \) and log fold change of either \( \leq -1 \) or \( \geq 1 \), were considered for further analysis. \( \text{NEAT1} \), now as the control gene for the DE analysis, was also counted as one of the most downregulated DE genes and was specifically noted on the graph. A total of 188 DE genes were counted, and almost 90% of them were downregulated (Figure 5.5F, Supplementary Table S5.1A,B).

To determine the physiological aspects and cellular functions that could be affected by the transcriptomic changes induced by the NEAT1 KO, gene network analysis was performed using Ingenuity Pathway Analysis (IPA, www.ingenuity.com). Among all outputs of the IPA, only the analysis on the potentially affected ‘molecular and cellular functions’ and ‘associated networks’ were relevant as this experiment was only conducted on normally cultured U2OS cells (Figure 5.5G, H). Cellular movement was the top hit in both analyses in the affected functions and networks. 29 out of the total 188 DE genes were found to be associated with functions related to regulating cellular movement (Figure 5.5G, Supplementary Table S5.1C, Supplementary Figure S5.1A in Appendix 4). Among all subcategories under cellular movement, cellular migration carried the most significant \( p \)-value of 6.71E-06, with a predicted activation \( z \)-score of -1.995, which indicated potentially downregulated function of cell movement in the total NEAT1\(^{-/-}\) cells.
Interestingly, when examining the DE genes associated with the networks of ‘cellular movement, cell-to-cell signaling and interaction, tissue development’, and ‘cellular development, cellular growth and proliferation, embryonic development’, most of the DE genes were peripheral to the network, that is they were not the central ‘hub’ genes (Figure 5.5H, Supplementary Figure S5.1B,C in Appendix 4). The top five potentially affected molecular functions (Figure 5.5G) were all covered in the two associated networks aforementioned (Figure 5.5H).

Given the weaknesses in the experiment described here, largely due to the inability to effectively group the samples in the PCA analyses, the RNA-seq experiments were repeated and better controlled with matched morphology lines. In addition, the next RNA-seq experiments included the NEAT1_2\(^{-/-}\) lines, as described below. The final RNA-seq analyses were used for the results as outlined in Chapter 3.

5.2.3 Generation of U2OS cells that were NEAT1_2 knockout but still retained NEAT1_1

The successful utilization of CRISPR genomic editing to KO overall NEAT1 expression led me to speculate if it was possible to use a similar strategy to specifically KO the NEAT1_2 isoform, whilst keeping the shorter NEAT1_1 isoform intact and expressed. Creating such a stable cell line would greatly facilitate the determination of specific functions for the NEAT1_1 and NEAT1_2 isoforms.

To this end, a YFP expression cassette was designed for insertion into the genomic site that is within 100bp downstream of the polyadenylation signal of the NEAT1_1 isoform (Figure 5.6A). Similar to the previous total NEAT1 KO experiment, the cut was to be induced by CRISPR-Cas9 and the integration was carried out through HDR. The rationale of the cut site was that once NEAT1_1 was terminated by the Poly(A), any residual ongoing transcription for making NEAT1_2 would then run into the YFP expression cassette. No matter if the YFP transcription was initiated either through its own CMV promoter, or the leakage of the NEAT1_1 polyadenylation, these events should
Figure 5.6: The design, screening and verification of the CRISPR strategy to create NEAT1_2 knockout cell lines. (A) A schematic representation of the NEAT1_2 knockout strategy. One CRISPR-Cas9 cut was designed to create a dsDNA break just downstream of the end of the transcription of NEAT1_1. A repair template with homologous arms flanking the cut site and a YFP expression cassette was constructed and transfected with the vector carrying the gRNA and CRISPR-Cas9. (B) Details of the gRNA and the
top-five potential off-target sites. (C) The PCR screening strategy for selecting homozygous colonies. Two primer sets were used in combination for accurate determination of YFP insertion. (D) Summary of all colonies screened and how each PCR result corresponded to different conclusions. (E) RT-qPCR of total NEAT1 and NEAT1_2 of all the homozygous colonies, which indicated that NEAT1_1 sometimes increased expression when NEAT1_2 was consistently and greatly reduced. Representative data is shown.
be all eventually terminated by the SV40 terminator at the end of the insert. This would severely disrupt NEAT1_2 transcription, and even if the SV40 terminator was not 100% efficient, the final transcript would still not be the same as the original NEAT1_2. Thus the function of forming paraspeckles was unlikely to be retained, resulting in a functional NEAT1_2 KO.

Similar to the total NEAT1 KO experiments, ideally multiple gRNAs should be used to act as both backup and controls for off-target activities, however due to the time limitations of this Ph.D project, only one gRNA was created. The off-target effects of this gRNA was calculated and summarized, and care was taken to make sure that no potential off-target effect carried a score higher than 1.0 (Figure 5.6B). For the repair template, 0.5kb homology arms were included on both ends of the YFP expression cassette. After constructing the plasmids, the two were transfected in an equal ratio into the wild-type U2OS cells. To screen single cell clones with homozygous NEAT1_2−/−, the same primer placement strategy was used as for screening total NEAT1 KOs (A963+A965 as primer set 1, A938+A939 as primer set 2) (Figure 5.6C). A total of 26 colonies were screened, and 13 colonies were found to carry at least one edited chromosome; 5 clones had both NEAT1 loci edited (Figure 5.6D). Four of these five homozygous clones were kept, three of them with a compact colony morphology – col.7, 44 and 85 (data not shown). RT-qPCR verification of NEAT1_2−/− indicated that although the NEAT1_2 expression level was significantly reduced by a minimum of 75% for all clones, the total NEAT1 level tended to be slightly upregulated to a maximum of 2.6 fold (Figure 5.6E). Note that in this case, the total NEAT1 level primarily consisted of NEAT1_1 transcripts. This means that an unchanged total NEAT1 level, such as for colony 44, is equivalent to the upregulation of NEAT1_1 isoform to compensate for the NEAT1_2 loss.
5.3 Discussion

The core component of this chapter is the creation of various NEAT1 stable KO cell lines. Using the CRISPR-Cas9 system, both total NEAT1\(^{-/-}\) cells and NEAT1\(_2\)^{-/-} U2OS cells were generated successfully. There are a number of observations that are worth highlighting. Firstly, the total NEAT1 KO strategy reduced both isoforms of NEAT1 with a similar efficiency, which supports the notion that the NEAT1 isoforms are primarily co-regulated at the transcriptional level. Secondly, NEAT1\(_2\) KO cell lines had a variable tendency for increased NEAT1\(_1\) expression. This suggests potential for a feedback mechanism on the transcription of NEAT1 that might be responsive to NEAT1\(_2\)/paraspeckle levels. Another possibility is that both transcriptional and post-transcriptional mechanisms might contribute to the upregulation of NEAT1\(_1\) at the same time. The last observation is that, whilst all of the homozygous KO clones showed a high NEAT1 reduction based on RT-qPCR data, variation was apparent in some clones. For example, the NEAT1\(_2\)^{-/-} colony 50 cell line had more than 95% reduction, whilst other clones had only around 75%. A more extreme case is the NEAT1\(^{-/-}\) colony 53 cell line, which had approximately 90% reduction of total NEAT1 levels, but no reduction in NEAT1\(_2\) levels. It is difficult to interpret how this might arise when taking into account the design of the YFP knock-in strategy. The results of colony 53 might reflect a mechanism that we are not yet aware of in the recombination machinery of the cells, which perhaps shuffle and rearrange the structure of the gene on rare occasions. Ultimately sequencing of the gDNA of the NEAT1 gene would be beneficial to explain this result. Alternatively, this observation may be due to technical errors in the RT-qPCR. To be on the safe side, these unexpected clones were not included in downstream analyses to avoid potential artifacts, but are certainly worth investigating and might hold value for unexpected new discoveries.

Despite the overall success of NEAT1 KO in this project, genomic engineering in any project usually leads to concerns regarding the potential non-specifically induced DSBs in the target genome. In normal cells, a DSB of genomic DNA will be resolved using either the HDR pathway or the NHEJ pathway. Homologous recombination uses the
sister chromosome to repair the damaged chromosome, resulting in perfect repair of the original sequence. NHEJ is prone to introduce indels, which is potentially deleterious to the genome especially in the coding regions. In the present study, DSBs were supplied with abundant repair plasmids for efficient HDR, however non-specific DNA breaks had to be repaired by either pathway, potentially producing undesired off-target effects.

At present, there is no gene KD or KO approach, either stable or transient that is completely immune to specificity issue. The CRISPR-Cas9 system is not different from any other methods on this, although efforts have been taken to minimize the possibility of off-targets. Firstly, although all gRNAs may cause non-specific targeting, bioinformatics tools have been developed to examine their binding sites in the host genome. This then makes it easier for users to decide on the gRNAs to use to avoid harmful off-target activity as much as possible. Secondly, although the actual efficiency of each gRNA in both on-target and off-target sites have to be determined empirically, off-target sites will have reduced efficiency due to the presence of mismatches. Moreover, even if a gRNA were to induce all of its potential non-specific targeting events, there are still no published data showing that all of these events will occur with the same probability. Therefore, occasionally it might be helpful to relax criteria in off-target activity in exchange for gRNAs that will induce better placed DSBs. In situations where only one gRNA is used, such as the NEAT1_2 KO experiment in this study, there is still no guarantee that the selected homozygous clones will have experienced the same set of non-specific Cas9 targeting in each clone. In other words, each individual NEAT1_2^{+/−} clone, even if made with the same gRNA, may still undergo off-target effects to a different extent. Based on this reason, single cell clones will significantly amplify the cell-to-cell difference, and the use of multiple gRNAs in the hope of controlling potential off-target effects might not be particularly helpful. However, using multiple gRNAs is still important if the CRISPR system will be applied to generate a heterogeneous population of KO cells (akin to an siRNA experiment), which is in effect the total sum of all possible off-target effects that might influence the final downstream readout.
In order to definitively determine the non-specific activity for each gRNA used in the experiment, genome-wide sequencing of the host DNA could be performed to detect unwanted indels. Alternatively, sequencing of PCR amplicons surrounding each of the bioinformatically predicted off-target locations may also be performed. However these are labor-intensive and expensive methods. The most economical approach is still to improve the design and selection of gRNAs to avoid the potential deleterious off-target effects in the first place. Fortunately, the more recent development of the D10A mutant Cas9 protein, the nickase that only induces single strand DNA break, has allowed much higher specificity to be achieved (Ran et al., 2013). In this system, two nickases are paired to flank the loci of interest leading to site-specific DSB. Non-specific single strand breaks will still take place, but as this form of DNA damage will have to be repaired by either HDR or BER (high-fidelity base excision repair, similar to straight ligation) (reviewed in Dianov and Hübscher, 2013), it thus effectively avoids the NHEJ pathway and substantially reduces the potential for introduction of indels to the off-target sites.

A rescue experiment, in which the gene that is knocked out/down is overexpressed, followed by determination of rescue of the effect the gene is associated with, is still the gold standard for the verification of the true effect of the gene of interest. Due to time limitations, this project did not contain rescue experiments, however this would clearly be important for future work. One complication is that thus far, the cDNA encoding the 23kb human NEAT1_2 isoform has not been cloned. However, it might be intriguing to investigate if the mouse NEAT1_2 equivalent (that has been cloned by collaborator Tetsuro Hirose) might serve the same purpose due to a conserved role in nucleating mouse paraspeckles (Imamura et al., 2014). Alternatively, a publicly available BAC (bacterial artificial chromosome) clone carrying the full length human NEAT1 gene may be used as a template for the PCR cloning of the human NEAT1_2 isoform.

From a technical point of view, the knock-in of the current reporter expression cassette to disrupt gene structure is not the only strategy one may consider for creating gene KO. There are at least the following four ways one may consider:
1. Introducing an open reading frame shift. This approach only works when the target gene encodes protein, and will not work for an ncRNA. This involves creating a DSB in the protein-coding exon without supplying an artificial repair template for HDR, forcing cells to use NHEJ for the repair. The major weakness of this approach is that it lacks an easy-to-use selection marker for the introduction of indels, and hence one has to verify by western blotting to show that the protein is indeed disrupted. In order to reduce the number of colonies for western blotting screening, a solution would be to integrate a selection marker into the plasmid that carries the CRISPR-Cas9 construct, such as the px459 plasmid provided by the Zhang laboratory (Ran et al., 2013). Then by selecting the cells that transiently expresses the marker, one would expect a higher rate of successfully engineered cells.

2. Creating a deletion of the entire gene. This approach is straightforward, complete, and can be conducted through both HDR and NHEJ, with the option of incorporating the selection marker in the process. However, the main concern is that deletion of a large region of DNA might also remove critical regulatory DNA, which may affect other nearby or distant genes. In the gene body of human NEAT1 a large number of transcription factor binding sites (TFBS) have been detected. Although it is probable that these TFBS mainly influence NEAT1 transcription, one can still not exclude the possible influence of these sites on other genes. Therefore, it is preferable to retain the DNA of the target gene within the genome.

3. Use CRISPR to guide artificial transcription factors to modulate the transcription of a gene, either activating or repressing it. Targeting of transcriptional activation domains such as the VP64 transactivation domain may be used to artificially upregulate endogenous gene expression (Maeder et al., 2013; Perez-Pinera et al., 2013). Further, the catalytically inactive mutant Cas9 might be used to repress target gene expression (Gilbert et al., 2013; Qi et al., 2013). The problem with this approach is that it is a transient system unless the entire expression system is permanently integrated into the genome in a random location.
4. The final approach, and the one used in this project, is a well-balanced approach that can be used to KO both protein-coding and non-coding RNA genes. Its efficiency is not lower than other approaches, as essentially it is still a CRISPR mediated method. It offers a selection marker for screening for successful modification (although it would still require verification), and also achieves targeted genomic modification with minimal excessive deletions.

One confusing aspect of this work is the apparent difference between the transcriptomic results in this chapter and the results presented in chapter 3. In this chapter, the global effect of total NEAT1 KO is the downregulation of around 180 genes with only a few being upregulated. Whereas in chapter 3, the global effect of total NEAT1 KO is the upregulation of a large number of genes with only a very small proportion being downregulated. These two results contradict each other. One possibility is that as these were two separate RNA-seq experiments, batch-to-batch differences in the library preparation might be a contributing factor. Another possibility, and the one that I favor, is that the major difference between the two total NEAT1 KO transcriptomic analyses is that in chapter 3, all chosen cells for RNA-seq, either WT or total NEAT1−/−, had a compact colony morphology. Whereas in this chapter, the WT cells were a mixed population of various morphologies, and among the three total NEAT1−/− cells chosen, two of them had compact morphology and one had sparse morphology. If I conservatively assume each colony carries different molecular characteristics regardless of their colony morphology, then gathering cells with smaller differences in morphology might be the only way to reduce the variation between cell clones being sequenced. Molecular characteristics inevitably affect individual gene expression. In differential gene expression analysis, if all possible cell types are mixed and their variations averaged within each sample group, then the estimation of the expression level of individual genes in each group will be less accurate due to the noise generated from the variation between samples. The smaller the differences between samples within each group/condition, the less the noise. Thus, if the effect of total NEAT1 KO on cells is subtle, then the experimental setup in chapter 3 will reveal this effect easier because less background noise is generated. However on the other hand, if the effect of total NEAT1 KO is large and obvious, it should still stand
out in the results described in the present chapter. This is because the transcriptomic differences that are obvious within the two heterogeneous cell groups are more likely to be consistently differentially expressed when two relatively more homogeneous sub-populations are compared. Indeed, when comparing the Ingenuity Pathway Analysis results of two DE analyses (Total NEAT1\textsuperscript{-/-} vs WT) from this chapter and chapter 3, cell movement was predicted to be more downregulated in the total NEAT1\textsuperscript{-/-} cells in both chapters, thus suggesting a relatively stronger effect of NEAT1 depletion in this cellular function. When looking closely at the actual DE genes that are related to cell movement in this chapter (Supplementary Table S5.1(C)), several can also be found in the Chapter 3 analysis (Chapter 3 Supplementary Table S3), and they all show the same orientation of differential expression. This is not likely to be a coincidence, and can provide confidence in the accuracy of the DE analysis in this chapter, and the analysis above. Finally, a more confident approach is to perhaps generate a lot more homozygous KO clones for each cell line, all with various colony morphologies, and then perform DE analyses and to selectively look at the most consistent DE genes.

In summary, CRISPR technology does have its limitations, but its advantages in its ease of use outweigh any disadvantages. The low threshold of usage and cost-effective output has made it a great technology to consider when stable gene KO is to be performed. In this project, stable NEAT1 knockout U2OS cells were obtained with relatively high efficiency. Even more importantly these same reagents can be applied to make many other human cell lines that are able to be transfected and subject to FACS. This potentially includes pluripotent stem cells if the effect of NEAT1 expression in differentiation were to be investigated in the future. Thus the success achieved in this chapter goes beyond this project towards the generation of many future human NEAT1\textsuperscript{-/-} cell lines.

(see Appendix 5 for references)
Chapter 6
General discussion
6.1 NEAT1 and paraspeckles – current research status

Long non-coding RNAs have highly variable modes of actions. The fact that the number of lncRNAs increases as the organism becomes more complex strongly implies their importance to complex eukaryotes such as humans (Taft and Mattick, 2003). The lncRNA NEAT1 is a relatively well-studied case among the approximately one hundred lncRNAs that have been studied to date. However, the cellular significance of NEAT1, since its initial discovery in 1997 (Guru et al., 1997), was not thoroughly investigated prior to the start of this Ph.D project. One of the most important features of NEAT1 is that both of its two major isoforms, the 3.7kb NEAT1_1 isoform and 23kb NEAT1_2 isoform in human, are located in paraspeckles, but yet only the NEAT1_2 isoform is crucial for the formation of these subnuclear structures (Clemson et al., 2009; Mao et al., 2011; Naganuma et al., 2012; Sasaki et al., 2009; Sunwoo et al., 2009). This observation suggested the function of NEAT1_1, both inside and outside of paraspeckles, was worthy of investigation.

When considering all isoforms of NEAT1 and paraspeckles together as a single entity, their combined biological significance is gradually becoming clearer. NEAT1/paraspeckle knockout mice have exhibited significantly compromised reproductive performance, such as reduced lactation and increased difficulty in maintaining pregnancy (Nakagawa et al., 2014; Standaert et al., 2014). These results implied that NEAT1 is likely to be related to tissue development in the adult female, but it is still not clear how these phenotypes are produced. Before these mouse data, significant progress had been achieved in understanding the molecular function of paraspeckles, which is essentially the sequestration of proteins and mRNAs (Chen and Carmichael, 2009; Hirose et al., 2014; Imamura et al., 2014; Prasanth et al., 2005). However, as only NEAT1_2, but not NEAT1_1, seeds paraspeckle formation, paraspeckle functions may only represent the NEAT1_2 isoform itself, rather than both NEAT1 isoforms.

The NEAT1_1 isoform, being more abundant, more conserved in mammals, and more
ubiquitously expressed (Nakagawa et al., 2011), still has not had its own molecular function identified. In addition to NEAT1_1, there are also other minor NEAT1 isoforms, such as tncRNA, that influence gene expression by acting on the promoter of certain types of genes in distinct biological contexts (Geirsson et al., 2004; Geirsson et al., 2003a; Geirsson et al., 2003b; Peyman, 1999; Peyman, 2001), but those works were never followed-up to reveal their relationship with paraspeckles or other NEAT1 isoforms. In contrast, a few studies have, albeit inadvertently, investigated the cellular significance of NEAT1_1, and the results are consistent. Overexpression of NEAT1_1 in prostate cancer cell lines led to increased proliferation and invasion in cell culture and xenografts (Chakravarty et al., 2014). These properties of proliferation and invasion are also reduced when NEAT1 is knocked down, with an increase in apoptosis and decreased cell viability, especially under stress (Chakravarty et al., 2014; Halford, 2013; Hirose et al., 2014). However, in all these studies the knockdown of NEAT1_1 results in knocking down both isoforms of NEAT1, thus also removing paraspeckles due to the loss of NEAT1_2. It is not possible to target NEAT1_1 without also targeting NEAT1_2, as they overlap completely. This makes it difficult to interpret those studies in the context of my results that have found a distinct role for NEAT1_1. In addition, the overexpression of NEAT1_1 has been shown to increase the size of paraspeckles (Sunwoo et al., 2009), it hence becomes difficult to tether out whether the phenotypes of NEAT1_1 overexpression is due to the NEAT1_1 itself or due to the increased size of paraspeckle. These problems are essentially because no transient transfection is able to selectively ablate NEAT1_1, whilst keeping NEAT1_2 normally expressed.

The core aim of this project was to deepen our understanding of the biological significance of NEAT1 and paraspeckles, and also to look for evidence that may indicate whether NEAT1_1 functions outside paraspeckles. Eventually, this knowledge may facilitate our understanding on how NEAT1_2/paraspeckle functions are related with the functions of NEAT1_1. To achieve these goals, I employed a systematic approach. The first step was to understand the mechanism that controls NEAT1 transcription, which is both highly active and dynamic (Gibb et al., 2011). The second step was to understand the impact of NEAT1 expression, and thus also the induction of paraspeckles, at the
cellular level. The third step was to gather evidence that may support a paraspeckle-independent role for NEAT1_1. This was based on the hypothesis that NEAT1_1 might have its own functions beyond simply residing within paraspeckles. Further, this role may be regulated by the transcription of the NEAT1_2 isoform, thus forming paraspeckles, resulting in sequestration of NEAT1_1 in a similar way to how other proteins and mRNAs are sequestered. These three steps formed an integrative approach that covered both the upstream and downstream of NEAT1 expression, and hence, generates a systematic understanding of the biological importance of it.

6.2 Findings, problems, and future directions

6.2.1 Results relating to events upstream of NEAT1 expression

Data-mining of the published information on the epigenetic state and transcription factor binding sites (TFBS) in the promoter of the human NEAT1 gene has revealed a highly complex regulatory system controlling NEAT1 expression. Examining methylation of the promoter, I found that as NEAT1 expression increases when cells differentiate from the pluripotent state, epigenetic ‘locks’ on the NEAT1 promoter are released. Examination of the pathways that participate in the transcriptional regulation of NEAT1 revealed a wide range of biological processes, including cellular proliferation, apoptosis, stress, differentiation, development and many others. The implication is that the ubiquitous and dynamic nature of NEAT1 expression is associated with the activation of multiple pathways; perplexingly, some of these pathways have seemingly opposite biological consequences for the cell. Amongst a myriad of transcription factors, p53 was demonstrated to act directly on the NEAT1 promoter to activate its expression specifically when cells are subjected to the stress induced by the pH drop of the culture media in U2OS cells. Thus the p53 stress-responsive pathway is an influential pathway among the many that control NEAT1 expression under this particular condition, and also links NEAT1 expression with the stress-response state of the cells. This result is consistent with other stress-related published data. In 2009, Guttman et al showed mouse Neat1 was significantly induced by p53 and harboured a p53 binding motif in its promoter (Guttman et al., 2009). Very recently, the DNA-damage response pathway
mediated by p53 was also demonstrated to transcriptionally induce human NEAT1 (Blume et al., 2015). In my studies, the acidosis and serum starvation treatments employed did not result in NEAT1 induction across all different cell lines tested, suggesting intrinsic differences in the transcriptional regulatory system for NEAT1 in these different lines. This also suggests that stresses used in other studies to induce NEAT1 may not be universally effective. Due to this reason, the conclusion for this part of the thesis was that strong NEAT1 expression might not be a good candidate as a universal stress-marker, but does imply the active state of multiple pathways within the cells.

6.2.2 Results relating to events downstream of NEAT1 expression

Using the CRISPR-Cas9 system coupled with homologous recombination, NEAT1 expression was successfully knocked out (KO) by at least 98%, with paraspeckles completed abolished. This has achieved a more substantive NEAT1 depletion than any previous study involving knockdown of NEAT1 in human cell lines. As the genome editing disrupted the structure of the NEAT1 gene, the resultant NEAT1 transcripts would not have their original sequence, nor their three-dimensional structure, likely resulting in loss-of-function for any remaining RNA. The total NEAT1 heterozygous cells have demonstrated slightly reduced proliferation and migration ability, but also shown significantly more sensitivity and weaker viability under the acidosis stress. Differential gene expression analysis from the RNA sequencing data of total NEAT1 heterozygous cells and wild-type (WT) cells yielded bioinformatic predictions that did not conflict with these phenotypic observations. Lastly, a wide range of molecular pathways and hundreds of genes appeared to be affected by the loss of NEAT1 heterozygous. This resembles the analysis of the upstream regulatory pathways of NEAT1 in general, which is also widely associated with many different cellular functions.

Taking advantage of the genomic editing technology, a NEAT1_2 isoform specific KO cell line was created. This creation significantly facilitated the endeavor to identify the role of each major isoform of NEAT1 in cell biology. By conducting a three-way transcriptomic comparison between the WT, total NEAT1 heterozygous, and NEAT1_2 heterozygous cells, the
change of expression of hundreds of genes (>300) was consistent with both the current sequestration model of paraspeckle functions, as well as the hypothesized model that the function of NEAT1_1 outside paraspeckles can be negatively regulated by these nuclear structures. Further, the data show that many genes seem only to be affected by the loss of NEAT1_2/paraspeckles, but not NEAT1_1. These genes could be indirect targets that are only affected by the sequestration mechanism postulated for paraspeckle function. A much smaller number of genes (less than or around 50) were also identified to support the notion that genes could be affected solely by NEAT1_1, or synergistically affected by both isoforms. Currently there is no supporting evidence at the molecular level for these observations, and the observed number of genes for each case is also relatively small, suggesting these relationships may not be significant.

6.2.3 Further verification of the differential gene expression analysis
Differential expression analysis is an essential step when comparing two transcriptomes. Knowing that DE analysis is highly dependent on the reliability of the basic mathematic and statistic handling of the data, a novel normalization method (RUV, removal of unwanted variation) (Risso 2014) and stringent statistical cut-off have been employed to increase the rate of true-positives. In addition, exploratory plots such as RLE (relative log expression) and PCA (principle component analysis) have also been used to determine the mathematical reliability of the data, each showing positive results. Finally, the level of NEAT1 isoforms in KO cells compared to WT, showed consistent results between the data from qRT-PCR and RNAseq, although there was a numerical difference, due to the use of different normalization and detection methods. However, the large number of DE genes was still surprising, with \( p \)-values of many DE genes being extraordinarily low.

To gain more confidence in the accuracy of these analyses, it would be beneficial to use qRT-PCR for verification of candidate genes. Additionally, performing rescue experiments using exogenous expression of the gene being knocked down/out and determining if the gene expression change could be reverted would be useful.

The three-way pairwise transcriptomic comparisons in chapter three is very dependent on the accuracy of DE analyses. However, by setting up relatively stringent cut-offs, the
results do seem to agree with the current model of paraspeckle function. Certainly, more verification will be required in the future, and these could take the form of rescue experiments and additional qPCR, as stated above.

6.2.4 NEAT1_2 might stabilize interactions between paraspeckle components

The creation of NEAT1_2\textsuperscript{-/-} cells also improved our understanding of how paraspeckle components are organized together. Using RNA-FISH and immunofluorescence on these cells, I showed that the NEAT1_1 transcripts were unable to co-localize with the essential paraspeckle protein NONO. This observation seems counter-intuitive when taking into account some other published links between NEAT1_1 and NONO: firstly, NEAT1_1 can interact with NONO \textit{in vitro} (Sreenivasa Murthy and Rangarajan, 2010); and secondly, tethering NEAT1_1 to chromatin using MS2 tagging technology could recruit paraspeckle proteins, including NONO, to form \textit{de novo} paraspeckles (Shevtsov and Dundr, 2011). These observations indeed suggested that NONO can bind to NEAT1_1. However, my data would suggest that for the binding to actually occur in the cell it may depend on having a local high concentration of NEAT1_1. This high concentration of NEAT1_1 clearly occurs in artificial \textit{in vitro} RNA-protein interaction assays, but also occurs in MS2 tagging experiments, where NEAT1_1 was forced to concentrate in a confined space through physical tethering. If this physical tethering is the key to the initial steps of paraspeckle assembly, then a similar endogenous system that stabilizes NEAT1_1 and paraspeckle proteins in a specific location in the nucleus might also result in the same consequence. The most likely candidate of this endogenous system is the transcription of NEAT1_2, as its transcript is very long, and is also bound by essential paraspeckle proteins such as NONO and SFPQ ((Hirose et al., 2014; Mao et al., 2011; Naganuma et al., 2012), and unpublished data from the Fox laboratory).

6.2.5 NEAT1_1 and NEAT1_2 might have opposite roles in gene regulation

As a result of this thesis, our understanding of both the regulation and biological outcome of NEAT1 expression has been greatly enhanced. NEAT1 and paraspeckles seem to be involved with house-keeping functions, lacking the specificity towards any particular biological processes. This finding still does not explain how these cellular processes
together contribute to the phenotypes presented for the NEAT1/paraspeckle KO mice. However, published data do support a generally pro-proliferative role played by NEAT1 and paraspeckles within cells. On one hand, the global development of Neat1-/- mice does not seem to be affected, and they are viable and able to grow into adults (Nakagawa et al., 2011). On the other hand, both the development of specific tissues that require rapid proliferation in short periods of time, and the growth of NEAT1 KD tumor cell xenografts, do seem to be dramatically affected (Chakravarty et al., 2014; Nakagawa et al., 2014; Standaert et al., 2014). In addition, NEAT1_1 overexpression has also been associated with poorer outcome of prostate cancer patients with increased incidence of metastasis and biochemical recurrence, and NEAT1_1 overexpression in vitro showed enhanced proliferation and invasion of cancer cells (Chakravarty et al., 2014). Thus these observations propose a pro-proliferative, pro-migration role played specifically by the NEAT1_1 isoform.

Knowing that NEAT1_2 and paraspeckles may act as negative regulator of NEAT1_1 function, this led to the question of whether this function of NEAT1_2 and paraspeckles could contribute to any phenotypes of to cells. With stress such as acidosis, NEAT1_2-/- cells may have greater cell viability compared to total NEAT1-/- cells, perhaps even higher than WT cells. This could be detected through the same viability experiment as conducted in Figure 3.5. However, even if this prediction was confirmed, it is unlikely that NEAT1_2 could be constantly effective at dampening proliferation by NEAT1_1. This is because NEAT1_2 is only observed in tissues where the overall NEAT1 expression is already high. In contrast, in normal cells with lower NEAT1 expression, NEAT1_2 and paraspeckles are not readily observed (Nakagawa et al., 2014). Therefore, a more likely situation in normal cells within the body is that NEAT1_2 could act as a temporary negative regulator of NEAT1_1, only when cells express too much NEAT1_1 (Figure 6.1). This theory fits the observation that (1) both NEAT1 isoforms are co-regulated by the same promoter (Hutchinson et al., 2007), and (2) that although NEAT1_1 is readily detected in almost any cell type (with the exception of pluripotent stem cells), NEAT1_2/paraspeckles are always much less common (Nakagawa et al., 2011). Interestingly, in transformed cell lines, and, indeed, many cultured cell lines, NEAT1
levels seem to be higher than in tissues, suggesting both NEAT1 isoforms are upregulated, as well as perhaps more steady regulation of NEAT1_1 by NEAT1_2.

6.2.6 How NEAT1_1 regulates gene expression is still unclear

The molecular mechanism of how NEAT1_1 generates its influence on gene expression will require more investigation in the future. Some published data have already suggested physical links between NEAT1_1 and active histone marks, as well as the transcriptional start and termination sites of active genes (Chakravarty et al., 2014; West et al., 2014). These data imply a role for NEAT1_1 in activating transcription. However, as these findings were not based on a system without NEAT1_2, these observations may be influenced by negative regulation via the NEAT1_2 isoform. It would be interesting to see if those associations between NEAT1_1 and chromatin were still observed in the NEAT1_2/- cells generated in this study. Nevertheless, data from this study do support this pro-transcriptional role of NEAT1_1, marked by more than 300 genes that were upregulated in NEAT1_2/- cells, and then downregulated again in total NEAT1/- cells. Interestingly, my data also shows a negative role for NEAT1_1 on the mRNA level of some genes (more than 100), which has not been previously shown. Here, it is very important to note that the regulatory effect of NEAT1_1 might not be direct (e.g. with the transcriptional machinery at those genes), but could be indirect, by influencing one or more mediators that have either positive or negative effect on their downstream genes. If this is the case, then NEAT1_1 would appear to be playing different roles for different genes, akin to a transcription factor.

In the future, it might be worthwhile to look for the common characteristics shared between the genes that are affected by NEAT1_1 outside paraspeckles. These genes might have similar biological functions, or share common regulatory mechanisms or gene/transcript structures that would explain the NEAT1_1 regulation. For example, there may be common transcription factors, similar types of TATA box employed, characteristic histone modifications or polyadenylation signal in mRNA. In particular with the search for shared transcription factors, one could use the DAVID database to identify shared transcription factors acting on the group of genes reciprocally regulated.
Figure 6.1: An updated model of molecular sequestration mediated by NEAT1_2/paraspeckles proposed at the end of this study. (A) Due to environmental and molecular triggers, NEAT1_1, the predominant isoform, is dynamically regulated and exerts a certain degree of influence on global gene expression. However, under specific conditions that greatly upregulate NEAT1, NEAT1_2 will be produced, along with more NEAT1_1 (B). (C) As NEAT1_2 induction results in the formation of paraspeckles, NEAT1_1 will rapidly relocate to paraspeckles away from its original destination, resulting in negative regulation of NEAT1_1 and a flow on effect on NEAT1_1 dependent gene expression. (D) Meanwhile, other molecules that carry their own functions will be also sequestered by paraspeckles, hence resulting in further, solely paraspeckle-dependent changes in the gene expression.
by two the NEAT1 isoforms, and then examine if these transcription factors are affected by any form of NEAT1 knockout. A different approach to this would be looking for whether knocking down specific essential paraspeckle proteins would reproduce the differential expression of certain genes in the same way as NEAT1 knockouts. A similar type of approach has been used in a previous report when investigating the role of paraspeckle proteins in controlling the stability of NEAT1 lncRNA (Naganuma et al., 2012).

6.2.7 The function of paraspeckles might be more prominent under certain conditions

Another confusing issue from the 3-way transcriptomic analysis (Figure 3.6, Supplementary Table S3.6-3.9) is that for many genes the influence of one NEAT1 isoform is equal, but opposite, to the influence of the other NEAT1 isoform. In other words, as long as both NEAT1 isoforms are expressed normally, or neither is expressed at all, then these target genes will not have altered expression level compared to WT cells. This appears to place NEAT1 and paraspeckles in a redundant position in regulating the expression of these many genes, and results in new question such as why these genes are regulated by NEAT1 and paraspeckles in the first place. To perhaps explain this it is important to note that RNA-seq analyses in this project were based on cells cultured under normal conditions, not stress. Thus, by measuring the expression of genes in response to NEAT1 loss, it primarily revealed clues as to the functional relationship between NEAT1_1 and NEAT1_2, rather than the biological significance of this relationship. By repeating the RNA-seq experiments using a stress by acidosis more biological insight may be revealed. Others have investigated the differences in the induction of genes in both WT cells and cells with transient siRNA-mediated NEAT1 loss under stressful conditions, identifying genes that are regulated specifically by paraspeckle formation (Imamura et al., 2014). It should also be highlighted that other similar subnuclear bodies that are built on lncRNAs, such as omega speckles in Drosophila and nuclear stress bodies in mammalian cells, also respond to many stresses by being rapidly induced, becoming enlarged and then sequestering proteins (reviewed in Lakhotia, 2011). These surprising similarities with paraspeckles might imply a more significant role for paraspeckles when cells are under stress than in normal conditions.
6.2.8 The function of NEAT1 and paraspeckles might extend beyond transcriptional control

Finally, but most importantly, this study only measured transcriptomic changes following NEAT1 knockout in cell lines. Alternative splicing was also not examined, as splicing differences cannot be shown in the current pipeline of gene-level differential expression analysis. This was due to the existing evidence indicating a higher likelihood of NEAT1 and paraspeckles being involved with transcriptional regulation (Chakravarty et al., 2014; West et al., 2014). However, the importance of paraspeckles and NEAT1 at the cellular level are still relatively unknown, ideally one should also expand the investigation into other –omics sectors. In particular, proteomics should be performed, due to the sequestration function of mRNAs by paraspeckles (Chen and Carmichael, 2009; Choudhry et al., 2014; Prasanth et al., 2005). Further, extending the transcriptomics, the sequestration of mRNAs may also be examined by measuring the nuclear vs. cytoplasm distribution of all mRNAs in WT and NEAT1 KO cells.

One point of interest supporting an alternative splicing analysis is the association between NEAT1 and SC35 that may reveal a closer association between nuclear speckles and paraspeckles than at first thought. I showed that in the absence of NEAT1_2, NEAT1_1 is distributed throughout the nucleus and in many cases overlaps with the signal from SC35 protein (Figure 3.3). SC35 is a splicing factor that is co-localised with the lncRNA MALAT1 in nuclear speckles, bodies considered as a reservoir for splicing factors (reviewed in Lamond and Spector, 2003). This relationship for NEAT1_1 in nuclear speckles presents a possibility that NEAT1 and/or paraspeckles might regulate some specific aspects of post-transcriptional processing. Interestingly, SC35 is not the only protein that has dual association with NEAT1 and MALAT1. In a recent report by West et al. a number of proteins in paraspeckles and nuclear speckles were identified by CHART-MS (capture hybridization analysis of RNA targets coupled with mass spectrometry) bound by both IncRNAs (West et al., 2014). Common to both RNA complexes they found some proteins, previously thought to be ‘paraspeckle’ proteins, including SFPQ, FUS, RBM14, PSPC1, CPSF7 (cleavage and polyadenylation specific factor 7, 59kDa), and HNRNPH3 (heterogeneous nuclear ribonucleoprotein H3), as well
as other proteins previously thought to be ‘nuclear speckle’ proteins such as SRSF5 (serine/arginine-rich splicing factor 5), CSTF2 (cleavage stimulation factor 64 kDa subunit), and SAFB (scaffold attachment factor B). Whilst these results need to be interpreted with caution, as mass spectrometry identification of nuclear RNA binding proteins on abundant nuclear RNAs will always yield false positives, these observations could indicate a complex network of NEAT1 and MALAT1 binding proteins, perhaps modulated through different localisations within paraspeckles or nuclear speckles. In addition, these observations could also be a reflection of the redundancy in paraspeckle and nuclear speckle functions. Both NEAT1 and MALAT1 are very abundant in the majority of cells, yet the KO of either one in mice is not very dramatic, with subtle phenotypes in the case of MALAT1. Interestingly, as mentioned in Chapter 1, NEAT1 may have arisen by duplication of the MALAT1 gene (Stadler, 2010). To this end, it would be intriguing to see whether the simultaneous KO of both NEAT1 and MALAT1 would cause a more dramatic phenotype than the single IncRNA knockouts. From there, one may be able to reveal the functional relationship between these two closely associated subnuclear structures and IncRNAs.
6.3 Final remarks

The ultimate aim of this thesis was to reach a more systematic understanding of the biological role of NEAT1 and paraspeckles in cell biology. Utilizing bioinformatic resources, molecular biology approaches, and novel applications of genome editing, our understanding of the general function of NEAT1 and paraspeckles is now suggesting a more house-keeping role particularly associated with gene regulation under cellular stress. However, further investigations are still required to reveal the actual molecular mechanisms NEAT1 and paraspeckles are regulating. The previously unclear functional relationship between NEAT1_1, NEAT1_2 and paraspeckles was finally examined systematically in this study. In terms of paraspeckle function, the evidence supports not only the existing sequestration model of proteins and mRNAs, but also the possible sequestration of NEAT1_1 transcripts. Tracking these observations back to the phenotypes of NEAT1 knockout mice, and the relevant cancer models, will clearly be important and fruitful future investigations.
Appendices
Appendix 1: Published work and contributions


I generated the human NEAT1 promoter-driven luciferase reporter plasmid that was controlled by the entire 6.4kb human NEAT1 promoter/enhancer sequence. This construct was used to demonstrate the function of an oestrogen receptor binding site in the NEAT1 promoter and demonstrated how NEAT1 upregulation was associated with prostate cancer progression (see Figure 2F,G of the paper attached below).
The oestrogen receptor alpha-regulated IncRNA NEAT1 is a critical modulator of prostate cancer

Dimple Chakravarty1,2, Andrea Sboner1,2,3, Sujit S. Naik4, Eugenia Giannopoulou5,6, Ruohan Li7, Sven Henning6, Juan Miguel Mosquera1,2, Jonathan Pauwels1, Kyung Park1, Myriam Kossai1,2, Theresa Y. MacDonald1, Jacqueline Fontugne1,2, Nicholas Erho1, Israel A. Vergara5, Mercede Ghadessi7, Elai Devici7, Robert B. Jenkins10, Nallazivam Palansamy11,12, Zhengming Chen14, Shinichi Nakagawa15, Tetsuro Hirose15, Neil H. Bander7,8, Himisha Beltran4,13, Archa H. Fox9, Olivier Elemento10,14 & Mark A. Rubin1,2

The androgen receptor (AR) plays a central role in establishing an oncogenic cascade that drives prostate cancer progression. Some prostate cancers escape androgen dependence and are often associated with an aggressive phenotype. The oestrogen receptor alpha (ERα) is expressed in prostate cancers, independent of AR status. However, the role of ERα remains elusive. Using a combination of chromatin immunoprecipitation (ChIP) and RNA sequencing data, we identified an ERα-specific non-coding transcriptome signature. Among putatively ERα-regulated intergenic long non-coding RNAs (lncRNAs), we identified nuclear enriched abundant transcript 1 (NEAT1) as the most significantly overexpressed IncRNA in prostate cancer. Analysis of two large clinical cohorts also revealed that NEAT1 expression is associated with prostate cancer progression. Prostate cancer cells expressing high levels of NEAT1 were recalcitrant to androgen or AR antagonists. Finally, we provide evidence that NEAT1 drives oncogenic growth by altering the epigenetic landscape of target gene promoters to favour transcription.
Serotonin receptors are key transducers of hormone signalling and control a wide spectrum of tissue-specific functions that control the homeostasis of reproductive organs. Aberrant or deregulated expressions of serotonin nuclear receptors are often associated with cancer progression and have been a major target for therapeutic intervention. The androgen receptor (AR) plays a central role in the progression of prostate cancer. Androgen ablation is highly effective in treating metastatic prostate cancer, although resistance inevitably develops leading to castrate-resistant prostate cancer (CRPC). Most cases of CRPC remain dependent on AR signalling, which has led to the clinical development and recent approval of potent AR-targeted therapies for CRPC (that is, abiraterone and enzalutamide). However, similar to first-generation anti-androgen therapies, patients develop resistance to these second-generation hormonal therapies. How CRPC tumours bypass AR signalling is emerging as a significant area of investigation. Many view co-targeting therapies as an important next step to managing the inevitable emergence of resistance to single-agent treatments, but critical to co-targeting is the identification of other biological pathways that drive disease progression and the development of strategies that can target these pathways. In CRPC, cross-talk between oestrogen- and androgen-signalling pathways may present an opportunity for clinical intervention. Oestrogen receptor (ER) signalling through ERs increases with prostate cancer progression and can drive important oncogenic events, including TPRM5/SIS-ERG expression. Although ERς signalling has been extensively studied in breast cancer, our understanding of the potential impact of this nuclear receptor on prostate physiology is less clear. Nevertheless, the connection is a particularly intriguing concept for the physiological homeostasis of reproductive tissue, as the oestrogen receptor (ER) is present in all stages of prostate cancer from the earliest precursor lesions to advanced disease. Thus, we provide experimental evidence to support this hypothesis and demonstrate a functional specialization and distinct genomic role of the nuclear receptor in prostate cancer. Here, we present a comprehensive study of ERs expression and function in prostate cancer to gain a better understanding of their role in promoting disease progression and management. We show that ERs is recruited to both coding and non-coding regions of the prostate genome and establishes a network of non-coding regulatory RNAs.

We identified novel enriched and abundant transcript 1 (NEAT1) long non-coding RNA (lncRNA) as a potential target of ERs and as an important factor for maintenance of prostate cancer cells. NEAT1 functions as a transcriptional regulator and contributes to a cancer-favourable transcriptome, thereby promoting tumorigenesis in experimental animal models. Our analysis of the transcriptional role of NEAT1 identified functions beyond its previously characterized role in maintaining the integrity of subnuclear organelles called paraspeckles. We demonstrate that NEAT1 is recruited to the chromatin of well-characterised prostate cancer genes and contributes to an epigenetic landscape. Analysis of two large clinical cohorts nominated NEAT1 as a novel biomarker of disease progression. Given its significance within the ERs signalling pathway, we propose that targeting NEAT1 might represent a novel and important therapeutic strategy for the treatment of prostate cancer.

Results

ERs in transcriptional regulation of prostate cancer.

To elucidate the role of ERs in prostate cancer, we analysed ERs protein and transcript levels in a panel of prostate cancer cell lines (n = 5) and in a cohort of matched benign prostate tissue (n = 14) and prostate adenocarcinoma (PCa) (n = 14), respectively. We observed that ERs was upregulated (P < 0.03) in prostate tumours compared with benign tissues (Fig. 1a). To determine the clinical relevance of ERs in prostate cancer, we performed immunohistochemistry using a tissue microarray composed of tissue cores from 64 samples of benign prostate tissue, 16 high-grade prostate intraprostatic neoplasia, 292 PCa, and 42 neuroendocrine prostate cancer (NEPC). Representative photomicrographs are depicted in Supplementary Fig. 1a. Although benign prostate had only low expression levels of ERs, ERs was detected in adenocarcinoma and the adjacent high-grade prostate intraprostatic neoplasia through focal nuclear and cytoplasmic staining (Supplementary Fig. 1a). ERs was overexpressed in a significant number of prostate cancer cohorts. It was also found to be overexpressed in prostate cancers with high Gleason score (Gle) compared with those with low Gle as well as in those with tumour recurrence when analysed via the Oncorex® database (Fig. 1b). Analysis of subcellular distribution in prostate cancer cell lines revealed significant nuclear distribution of ERs in all cell lines tested (Supplementary Fig. 1b). ERs protein levels were similar in both AR-positive LNCaP and VCaP cells (Fig. 1a, inset). We used parental VCaP and the AR-negative prostate cancer cell line NCI-H660 as model cell lines to further explore and delineate the specific contributions of ERs to prostate cancer. A ligand-dependent modulation of invasive potential was observed in VCaP cells on oestrogen (E2) treatment (Fig. 1c). These results suggest that a functionally relevant, ligand-dependent ERs signalling pathway is active in prostate cancer cell lines.

To further understand the impact of ERs, we generated VCaP cells that overexpressed the receptor (VacEP-ER) and confirmed this was confirmed by western blotting (Supplementary Fig. 1c). VCaP ERs exhibited significantly higher invasive potential than VCaP parental or NCI-H660 cells (Fig. 1e). Intriguingly, the noted effects of ERs overexpression were independent of AR status, as experimental silencing of AR in VCaP ERs cells did not completely abrogate the increased invasive potential of these cells (Fig. 1c). These data suggest that prostate cancer cells can use alternative nuclear receptors, including ERs, to propagate, and understanding these mechanisms will help discern the complete spectrum of key regulators of prostate cancer progression.

Studies have established ERs’s dominant role in transcriptional regulation of target genes in breast cancer25-28. Likewise, high nuclear levels of ERs in prostate cancer suggest a role in disease progression. We therefore analysed the association of ERs with chromatin in prostate cancer by ChIP-seq in VCaP cells. We identified a total of 8340 ERs-bound genomic regions, including 6491 previously identified ERs-bound genomic regions, implying that ERs are capable of binding to regions that have not been previously identified. We performed whole-genome RNA-seq analysis to identify differentially expressed genes in VCaP ERs cells compared with parental VCaP cells. We identified a total of 2452 upregulated and 3619 downregulated genes in VCaP ERs cells compared with parental VCaP cells. We then performed enrichment analysis of the differentially expressed genes in VCaP ERs cells to identify biological pathways that are affected by ERs overexpression. We identified a total of 21 enriched pathways, including pathways related to prostate cancer, cell proliferation, and cell survival.

Using publicly available data sets29, we found that 38% of the intergenic ERs-binding sites in the prostate cancer genome (from VCaP ERs and NCI-H660 cell lines) overlapped with the active histone mark trimethylated lysine 4 of histone H3 (H3K4me3) and trimethylated lysine 36 of histone H3 (H3K36me3).
Figure 1 | Efrs play a distinct role in prostate cancer. (a) Efrs are upregulated in prostate cancer compared with matched benign controls. Waterfall plots depict the qRT–PCR expression levels of Efrs mRNA in an independent cohort of benign (n = 14) and PCA (n = 14). (b) The expression of Efrs in different prostate cancer cell lines determined by western blotting and compared with NCCt, a breast cancer cell line. (c) Analysis of Efrs expression in Oncomine public data sets of normal versus prostate cancer and advanced disease. (d) Invasion of CCAfP and VCAfP Efrs cells analyzed by 48 h post treatment with vehicle control or 10 μM of the presence of control or ARRNA. Results are expressed as the mean ± SD of three independent experiments. Student’s t-test was performed for comparisons. *P < 0.05; **P < 0.01. (e) Recruitment of endogenous Efrs to target gene chromatin was analyzed in CCAfP cells with or without EZ treatment. Results are expressed as the percentage of input of two independent experiments. Error bars represent the range of data. (f) Identification of differentially expressed Efrs regulated InoRNAs that are differentially expressed between benign versus prostate cancer and prostate cancer versus NEPC. (g) Box plots show expression levels of the top three Efrs-regulated InoRNAs from 26 benign and 40 PCA cases, with lidogram depicting their chromosomal position. Waterfall plots depict the qRT–PCR expression levels on an independent cohort of benign (n = 14) and PCA (n = 14) of the three nominated InoRNAs NEAT1, NEAT2, and FIBR9.99.

(P < 1.0e− 7). On the other hand, 30.7% of those sites overlapped with histone marks typical of inactive chromatin, such as trimethylated lysine 9 of histone H3 or trimethylated lysine 27 of histone H3. To prioritize experimental validation of Efrs targets, we ranked the peaks according to the average P-value determined by the peak-calling algorithm ChIPSeeGer25 and selected the highest ranking peaks for further analysis. We assessed recruitment of endogenous Efrs to the top 11 binding
sites in parental VCaP cells (Fig. 1d), providing an experimental validation of the CHIP-seq data. A significantly higher upregulation of NEAT1 IncRNA was evident at the binding sites compared with control IgG.

Given the enhanced recruitment of ERs to intergenic regions of the genome, we next investigated the likelihood that ERs might influence transcriptional output and thereby the repertoire of non-coding RNA in the context of prostate cancer. We thus analyzed the abundance of non-coding transcripts in RNA-seq data derived from a cohort of 72 prostate tissues, which included 36 benign prostate samples, 49 PCA and 7 NEPC (Supplementary Methods). We identified 1,314 and 1,399 intergenic IncRNAs out of 12,143 known IncRNAs (see Supplementary Methods). We identified 31,674 and 26,483 intergenic IncRNAs in benign and PCA, and between PCA and NEPC, respectively (false discovery rate <0.01). We identified 410 intergenic IncRNAs putatively regulated by ERs (Fig. 1c and Supplementary Dataset 2). An analysis of AR-binding sites identified 98 IncRNAs that have an AR-binding site within the promoter. This supported the view that ERs might significantly influence the non-coding transcriptome in prostate cancer. Using the RNA-seq data on VCaP and VCaP ERα cells lines to validate the expression levels of the two differentially expressed ERα-regulated IncRNAs, we selected six potential candidate IncRNAs that had higher expression in VCaP ERα compared with VCaP. We used quantitative real-time PCR (qRT-PCR) to validate expression for these six ERα-regulated IncRNAs in VCaP and VCaP ERα-expressing cell lines (Supplementary Fig. 1E). Expression of the six IncRNAs was further determined in a cohort of 28 matched benign and prostate cancer samples, confirming upregulation of these three nominated IncRNAs in prostate cancer compared with benign prostate (Fig. 1f). Taken together, these analyses indicate that ERs is a transcriptional regulator of the non-coding transcriptome in prostate cancer.

Among the putatively ERα-regulated intergenic IncRNAs, we identified NEAT1 as the most significantly upregulated IncRNA in prostate cancer. In our patient cohort, IncRNA was overexpressed in 73 samples (73/75) (Fig. 1f and Supplementary Dataset 2). The NEAT1 gene is located on chromosome 11q13.1 and produces two RNA isoforms that are completely different at the 5′-end. The shorter isoform (hereafter abbreviated as NEAT1-NEAT1.1) is 3.7kb in length and more abundant than the longer, 23.4kb isoform (NEAT1.2; NEFP, a Spermatozoa marker) (Fig. 1f). IncRNAs from subnuclear bodies called paraspeckles and their role in prostate cancer remains unknown.

E2Rα-regulated NEAT1 IncRNA is upregulated in prostate cancer. In the Oncomine database, we observed significant overexpression of NEAT1 IncRNA in several prostate cancer data sets (normal versus cancer) and aggressive prostate cancer (Supplementary Fig. 2A). We first confirmed that amplification of chromosome 11q (where NEAT1 resides) was not seen across 189 adenocarcinoma cases, eliminating chromosome 11q13.1 amplification as an explanation for high NEAT1 expression (Supplementary Fig. 2A, C). The expression of NEAT1 was twofold in radical prostatectomy cohorts with long-term clinical follow-up from the Mayo Clinic 22,43,134. We measured using Affymetrix HuEx microarrays (see Methods) Supplementary Table 1 contains the patient characteristics of the data sets. NEAT1's expression was masked in the 95th percentile of all genes on the microarray (Fig. 2b). We determined levels of NEAT1 by RNA-Seq in situ hybridization (ISH) in a tissue microarray that included 16 benign prostate tissues, 21 PCA, 12 PCa with neuroendocrine differentiation and 7 NEPC cases. NEAT1 was found to be highly expressed in prostate cancer compared with that in benign tissue (Supplementary Fig. 2b).

We observed that in a panel of prostate cancer cell lines, ERα overexpression and E2 treatment upregulated NEAT1 transcript levels in a time-dependent manner (Fig. 2c). In DU145, an ERα-negative cell line 22, E2/ERα signalling was intact (Fig. 2c), supporting an ERα-independent phenomenon. Following E2 overexpression, we also recorded an increase in expression of the long isoform NEAT1.2 (Supplementary Fig. 2c). This was not surprising as both isoforms of NEAT1 are driven by the same promoter 43. The preferential upregulation and increase in the NEAT1 long form alone is not well understood and is not further addressed in this study. Interestingly, knockdown of ERβ did not alter NEAT1 levels. Therefore, NEAT1 regulation is specific for ERα (Supplementary Fig. 2d).

NEAT1 was originally identified localized to subnuclear organelles called paraspeckles. The function of paraspeckles in various cell contexts is unclear, however, and NEAT1 expression has been proposed to act as a stress response to the maintenance of paraspeckles. NEAT1 overexpression may play a role in the maintenance of paraspeckles. NEAT1 overexpression may play a role in the maintenance of paraspeckles. NEAT1 overexpression may play a role in the maintenance of paraspeckles.
gene body (Fig. 3d). A recent study revealed that bivalent
H3K4Me0 and H3K27Me3 marks are indicators of functional
cis-regulatory loci from the non-coding genome. ERα
recruitment to specific regions of the NEAT1 promoter was
independently validated by ChIP in VCaP, VCaP ERα and
NCl-H1299 cells (Fig. 2c and Supplementary Fig. 2g) using
specific primers encompassing ERα-binding sites in the NEAT1
promoter. We found that a functional oestrogen/ERα signalling
pathway was active in VCaP cells, as determined by reporter-
based oestrogen response element (ERE) luciferase assays in VCaP
cells, with ERα and AR overexpression, and E2 or ICI182
pre-treatment, respectively, for 48h (Fig. 2f). To further test whether
ERα is required for NEAT1 transcriptional activation, we
generated luciferase promoter reporter constructs with both
ERα-binding sites upstream of the luciferase-coding region.
Luciferase reporter assays in VCaP cells confirmed that NEAT1
promoter activity was upregulated in an ERs-dependent manner and further enhanced with E2 treatment (Fig. 2g).

ERs and NEAT1 regulate several prostate cancer genes. We next sought to understand the physiological role of NEAT1 and determine the downstream targets of the ERs-NEAT1 axis in prostate cancer. We were particularly interested in identifying genes significantly deregulated in prostate cancer and positively correlated with ERs and NEAT1 expression. Transcriptome sequencing of VCaP and VCaP ERs cells and pairwise comparison revealed 588 genes to be upregulated in VCaP ERs cells (log2-fold change > 2) (Supplementary Dataset 3 and Fig. 3a). We performed a comparative analysis of this 588 gene signature using Oncomine concept analysis. We focused on data sets from prostate cancer studies that included both prostate tumour and benign prostate tissues. The analysis revealed that the ERs gene signature was significantly upregulated in a number of prostate cancer data sets, but was downregulated in other non-prostate data sets, indicating that ERs regulates prostate cancer-specific genes (Fig. 3b and Supplementary Dataset 4).

To validate whether ERs targets identified by in silico analysis are dependent on cellular levels of ERs, we experimentally silenced ERs in VCaP cells using a small interfering RNA (siRNA) approach and determined transcript levels of ten target genes using qRT-PCR. The target genes selected for validation were those genes that demonstrated the highest log2-fold difference in VCaP and VCaP ERs cells. Results indicated that messenger RNA levels of the target genes selected were dependent on ERs (Fig. 3c), suggesting a distinct contribution of ERs in determining the transcriptional program.

NEAT1 is a downstream target in the ERs signalling pathway. After determining an ERs signature, we next investigated the potential role of NEAT1. Interestingly, knockdown of NEAT1 compromised the expression of ERs target genes, suggesting that NEAT1 is not only a downstream target but also a mediator of ERs signalling in prostate cancer cells (Fig. 3d). To evaluate this further and to determine whether a functional synergy between ERs and NEAT1 pathways exists in prostate cancer cells, we performed RNA-seq of vector control and NEAT1-overexpressing VCaP cells to determine a NEAT1 signature. To achieve this, we limited our analysis to genes that were upregulated four-fold in NEAT1-expressing cells (Supplementary Dataset 5). Interestingly, the NEAT1 signature showed a strong correlation with the ERs signature (Fig. 3e). The top 1,000 genes of the NEAT1 signature revealed that this signature is upregulated in prostate cancer data sets when compared with other cancer data sets (Fig. 3e and Supplementary Dataset 4). Furthermore, the NEAT1 signature was also upregulated in all prostate cancer data sets (competing benign versus PCA, odds ratio > 2.0 and P < 1 x 10^-4) (Supplementary Fig. 3a).

We also queried Oncomine prostate data sets to identify genes whose mRNA levels correlate with those of NEAT1 (correlation coefficient > 0.5). We compared this gene list with the ERs signature genes from our analysis in Fig. 3a and identified 155 genes in common. These 155 genes were also found to be upregulated in all prostate cancer data sets compared with other cancer data sets (only normal versus cancer data sets were considered, odds ratio > 3 and P < 1 x 10^-7) (Supplementary Dataset 4 and Supplementary Fig. 3b).

To determine whether the genes identified by in silico analysis are indeed influenced by NEAT1, we silenced NEAT1 in VCaP cells and determined transcript levels of potential target genes using qRT-PCR. We selected the top ten genes that were significantly correlated to NEAT1 expression across all prostate cancer concepts. As expected, mRNA levels of these selected target genes were indeed dependent on NEAT1, further confirming a definite role of NEAT1 in the transcriptional programme (Fig. 4a). In addition to cell lines, we also determined transcript levels of these ERs-NEAT1 signature-selected genes in a small patient cohort (n = 29) of 13 matched benign and PCA, respectively. We observed that relative mRNA levels of these NEAT1-ERs signature-selected genes revealed significant upregulation in prostate cancer (Fig. 4b). We computed the log2-fold change of expression levels using the 13 paired tumour/benign samples for NEAT1 and for these selected genes. We then correlated the fold change values and observed a moderate-to-strong correlation between NEAT1 and the associated genes in clinical samples (Fig. 4c). Among these seven genes, prostate-specific membrane antigen (PSMA) and alpha-methylacyl-CoA racemase (AMACR) are well-known diagnostic and, in the case of PSMA, prognostic markers of prostate cancer progression. Furthermore, knocking down ERs did not alter expression of NEAT1 signature genes in LNCaP, PC3, VCaP and NCI-H669 cells (Supplementary Fig. 3c), suggesting a non-redundant regulatory role for ERs.

NEAT1 and chromatin regulation. To study the potential role of NEAT1 in regulation of target genes in vivo, we performed baculovirus reporter assays using PSMA-Ac as a candidate NEAT1 target. NEAT1 induced robust activation of the PSMA promoter in PC3 cells (Fig. 5a) and VCaP cells (Fig. 5b). These results prompted us to investigate whether NEAT1 is recruited to chromatin of target genes. We used the chromatin isolation by RNA purification (ChIRP) approach to pull down endogenous NEAT1 from VCaP cells. Analysis of the ChIRP data revealed that NEAT1 is recruited to the PSMA promoter, but not the downstream exon 1 (Fig. 5c). In addition to PSMA, we also tested NEAT1 recruitment to other target genes described in Figs 3c and 4a, and observed that in addition to PSMA, NEAT1 was also recruited to the promoter region of GJB1 (Supplementary Fig. 4a). This suggests that NEAT1 transcriptionally regulates a...
compendium of genes known to be involved in prostate cancer progression. We hypothesized that NEAT1 might contribute to gene transcription by interacting with chromatin-modifying proteins and/or interacting with histones. Several recent studies support the view that lncRNAs recruit chromatin-modifying machinery. To test this hypothesis, we analyzed the chromatin landscape at the PSMA promoter and observed that NEAT1, and not NEAT2, facilitated gene transcription by promoting an active chromatin state (Fig. 5d). Overexpression of NEAT1 significantly increased active chromatin marks at the PSMA promoter (that is, H3K4Me3 and H3AcK9). Of note, ERα was not significantly recruited to the PSMA promoter when expressed alone. Overexpression of NEAT1 resulted in subsequent recruitment of NEAT1 and ERα to the PSMA...
Figure 4. | NEAT1-Era signature is upregulated in prostate cancer. (a) Relative mRNA levels of genes nominated from analysis in Fig. 3(a) analyzed using qRT-PCR in normal VCaP cells transfected with scrambled (Sc) and NEAT1 siRNA (NR), respectively, with and without Era (0.01 mM) treatment. Results are expressed as the mean ± s.d. calculated from three independent experiments. Student’s t-test was performed for comparisons (relative mRNA levels of target gene expression) between − Era and + Era conditions for scrambled siRNA and NEAT1 siRNA transfections. A representative example is shown for AORR1 and PSA expression. *P < 0.05 and **P < 0.01 were considered statistically significant. (b) Validation of expression of the top target NEAT1 Era signature genes in a small matched patient cohort of 12 benign and 12 PCA, n = 24. Results are expressed as the relative mRNA levels lum/sym/benign from two independent experiments. Error bars represent the range of data. (c) Heatmap shows the Spearman’s correlation results from b.
Figure 5 | NEAT1 is a transcriptional regulator. (a,b) Promoter luciferase reporter assays show that NEAT1 activates PSMA promoter in PC3 and VCaP cells. Cells were co-transfected with empty vector or PSMA-luc and Renilla-luc reporter genes alone or with NEAT1, NEAT1 + E2, and NEAT1 + A. Luciferase activity was measured 48 h post treatment with E2 (10 nM) or R1881 (10 nM). Results are expressed as the mean ± s.d. calculated from three independent experiments. Student’s t-test was performed for comparisons relative to untreated cells. E2 and R1881 treatments were considered statistically significant. (c) Quantitative analysis of NEAT1 ChIP in VCaP cells with or without E2 treatment. ChIP was performed with specific primers for the PSMA promoter. Results are expressed as the percentage of input calculated from two independent experiments. Error bars represent the range of data. Results were reproducible between representative experiments. (d) Analysis of the chromatin landscape at the PSMA promoter performed by ChIP in VCaP cells alone or treated with NEAT1, E2, NEAT1 + E2, NEAT1, E2 + NEAT1, 1 URNA and NEAT1 + E2 + NEAT1, 1 URNA with and without E2 treatment. qPCR was performed with specific primers for the PSMA promoter. Results are expressed as the percentage of input calculated from two independent experiments. Error bars represent the range of data. Results were reproducible between representative experiments. (e) NEAT1 binds to Histone H3. 20 mer-biotinylated Histone H3 and NR 024840 antisense probes were used to immunoprecipitate NEAT1 and NR 024840 from nuclear lysates of VCaP cells using streptavidin magnetic beads. Immunoprecipitates from Streptavidin-IP were analysed on 15% gel and probed for Histone H3. NEAT1 is shown to also bind with active histone H3 modifications including H3K9 and H3K4Me3.
promoter. These studies indicate that although NEAT1 may function as a shaperone for ER and other chromatin-modifying machinery at other promoters, binding of ER and/or recruitment to NEAT1 targets is not necessary for transcriptional activation.

As our data suggests that NEAT1 overexpression favors a chromatin landscape for active transcription, we investigated whether NEAT1 could directly interact with nucleosomal histones. Nuclear lysates from VCaP cells were used in an immunoprecipitation experiment with streptavidin beads coupled with either scrambled, antisense NEAT1, or antisense NR_024490 (another ERβ IncRNA target) oligonucleotides. NEAT1 was found to specifically associate with histone H3 (Fig. 6a, right panel) and the specificity of this binding is apparent when comparing lanes 7 and 9, which represent Streptavidin-IP using scrambled biotinylated oligos and Streptavidin-IP using antiserene-NEAT1 oligos and nuclear lysates from NEAT1 shRNA-treated cells, respectively. As an additional negative control, we used scrambled and specific antisense oligos for a different lncRNA, NR_024490, another ERβ target. The results indicate that NEAT1 can associate with chromatin via a specific interaction with histone H3. We also determined association of NEAT1 with active histone H3 modifications, including H3K4me3 and H3K4me1 (Fig. 6a, right panel). Similar association patterns were seen for NEAT1 in NCI-H660 cells (Supplementary Fig. 4b). To complement this finding, we performed RNA immunoprecipitation experiments using VCaP ERα cells using anti-histone H3 and anti-SNRP70 (positive control) as the immunoprecipitation antibody. qRTPCR showed robust binding of NEAT1 to histone H3 (Supplementary Fig. 6c). The positive control U1 small nuclear RNA showed high enrichment in the immunoprecipitate with SNRP70. To further confirm the specificity of NEAT1 binding to histone H3, we performed a streptavidin-biotin pull-down assay in VCaP and VCaP ERα cells with and without E2 (Supplementary Fig. 4d). These data suggest that NEAT1 can directly interact with the histone H3 component of chromatin.

NEAT1 promotes prostate tumorigenesis. To better understand the physiological role of NEAT1 in the context of ER in prostate cancer, we first determined the levels of NEAT1 in VCaP cells (Supplementary Fig. 5a). Further, we generated stable VCaP and VCaP ERα cell lines that overexpress NEAT1 (Supplementary Fig. 5b). We also knocked down NEAT1 in VCaP cells using VCaP ERα-expressing cells by stably expressing NEAT1 shRNA targeting two different regions of NEAT1 and non-targeting shRNA (Supplementary Fig. 5c). Although overexpression of NEAT1 significantly increased proliferation and cell invasion, knockdown of NEAT1 significantly decreased proliferation and the invasive properties of the cells (Fig. 6a,b). Soft agar assays were performed in both VCaP and VCaP NEAT1 cells. Colonies were monitored over a period of 21 days. Overexpression of NEAT1 resulted in a significantly higher number of viable colonies (Fig. 6c). Colony-forming assays performed in NEAT1 clones in VCaP cells with and without E2 demonstrated that E2 treatment in NEAT1-overexpressing cells significantly increased the number of colonies (Fig. 6d). These in vitro assays establish an oncogenic role for NEAT1.

To further validate the oncogenic role of NEAT1, we extended our studies to an in vivo model system. We performed xenograft studies in NOD-SCID mice. The mice were treated with time-release estrogen pellets. They were divided into two groups and one group was implanted subcutaneously with VCaP ERα cells expressing control shRNA luciferase reporter, and the other group with VCaP ERα cells expressing NEAT1 shRNA luciferase reporter. The mice from both groups were imaged weekly for luciferase activity and Fig. 6e shows the bioluminescent signals at day 7 and day 35. The tumour growth was monitored weekly for 45 days and was found to be significantly lower in the NEAT1 shRNA-expressing group compared with the control group (Fig. 6f). The tumours were excised and weighed, and the NEAT1 shRNA group had significantly smaller tumours (Supplementary Fig. 6c). We confirmed the efficacy of the shRNA in vivo by measuring the NEAT1 and ERα levels in the tumours (Supplementary Fig. 6c).

To further substantiate our hypothesis that NEAT1 plays a role in tumorigenesis, we repeated the experiment in athymic nude mice using VCaP control and VCaP NEAT1-overexpressing cells, as well as NCI-H660 and NCI-H660 NEAT1-overexpressing cells. In both these experiments, a significantly higher tumour growth was seen in the NEAT1-overexpressing cells (Fig. 6g) (and Supplementary Fig. 6a) further confirming its oncogenic potential. qRT-PCR analysis confirmed an increased expression of the NEAT1 signature genes in VCaP NEAT1 xenografts compared with control VCaP xenograft tissue (Supplementary Fig. 6f).

NEAT1 is associated with therapeutic resistance. The study presented so far shows that ERβ establishes an oncogenic cascade and that NEAT1 functions as a downstream mediator of ERβ signalling. The ERβ-NEAT1 axis is functional both in AR-positive and negative cell lines, and drives prostate carcinogenesis. We hypothesized that targeting NEAT1 using mechanisms that can constrain ERβ might represent a novel therapeutic strategy in prostate tumours that are resistant to anti-androgen therapy. To test this hypothesis in vitro, we evaluated the effect of anti-androgens and anti-androgens on NEAT1 levels in prostate cancer cell lines. As shown in Fig. 7a, NEAT1 expression is constrained when cells are treated with the ERα antagonists ICI 182,780 (ICI) and 4-hydroxytamoxifen (4OHT) in combination with E2. Intriguingly, treatment of ICI and 4OHT alone for longer periods can enhance NEAT1 expression (Fig. 7a,b). We observed similar results with AR antagonists enzalutamide and bicalutamide (Fig. 7c,d). These results provide compelling evidence to evaluate NEAT1 levels in advanced CRPC cases. RNA-fluorescent ISH analysis of benign and advanced prostate tumours, including CRPC and NEPC tumour tissue samples, illustrated significantly upregulated NEAT1 levels in advanced prostate cancer, with enhanced focal staining throughout the tumour tissue (Fig. 7c). We also screened nine cases of benign prostate, seven PCa and seven CRPC (Supplementary Table 2) for NEAT1 and ERβ expression by qRT-PCR (Fig. 7b), and both NEAT1 and ERβ levels were significantly higher in the CRPCs. We determined the correlation between NEAT1 and ERβ expression by estimating the Pearson’s correlation coefficient R. The results indicate a strong positive correlation: \( R = 0.86 \) (\( p = 0.001 \)). Taken together, these results present a novel role for the non-coding transcriptome in cancer-favourable adaptations.

NEAT1 is associated with aggressive prostate cancer. Given the importance of NEAT1 in promoting tumorigenesis both in vitro and in vivo, we sought to determine the relationship between NEAT1 levels and prostate cancer clinical outcomes in 594 patients from two radical prostatectomy cohorts with long-term clinical follow-up from the Mayo Clinic Comprehensive Cancer Center between 1987 and 2001 for clinically localized prostate cancer.

We assessed the prognostic potential of NEAT1 expression using several statistical measures and correlating it with biochemical recurrence (BCR) and metastasis (M0), prostate
Figure 6 | NEAT1 is a driver of oncogenic cascade. (a) Cell proliferation assays were performed in VCAp vector control, NEAT1-overexpressing cells and also in scrambled and NEAT1 knockout cells with or without E2 treatment (10 nM) at 24 and 48 h time points. Results are expressed as the mean ± s.d. calculated from three independent experiments. Student’s t-test was performed for comparisons between two conditions for vector control, NEAT1-1, and NEAT1-2 and E2 conditions for sRNA-scrambled, VCAp-shRNA1 and shRNA2 transfections. *P<0.01 was considered statistically significant. (b) Quantitative bar chart for depicting percentage of cells invaded at the completion of invasion assay performed in VCAp vector control, NEAT1-overexpressing cells and also in scrambled and NEAT1 knockout cells with or without E2 treatment (10 nM). Results are expressed as the mean ± s.d. of three independent experiments. **P<0.01. Student’s t-test. (c) Soft agar assays were performed with VCAp control and NEAT1-overexpressing cells. Quantitative bar plot analysis of colony colonies at 21 days are shown. Results are expressed as the mean ± s.d. of three independent experiments. ***P<0.001. Student’s t-test. (d) Colony-forming assay were performed in VCAp vector control, NEAT1-overexpressing cells with or without E2 treatment (10 nM). The right panel depicts the number of colonies at 21 days. Results are expressed as the mean ± s.d. calculated from three independent experiments. **P<0.01 and **P<0.001. Student’s t-test. (e) VCAp-E6a cells expressing can shRNA-bas, can shRNA2 (top panel) and shRNA1 (bottom panel) injected mice in shown. **P<0.01. Growth curve for the tumours monitored up to 45 days. Results are expressed as the mean ± s.d. calculated from three independent experiments. **P<0.01, Student’s t-test. (f) VCAp and NCI-H660 vector control and NEAT1-overexpressing cells were injected subcutaneously into the flank of male NOD-SCID mice. Bioluminescence imaging on Day 7 and Day 35 in the VCAp-E6a scrambled shRNA (top panel) and VCAp-E6a-NEAT1 shRNA (bottom panel) injected mice is shown. **P<0.01. Growth curve for the tumours monitored up to 45 days. Results are expressed as the mean ± s.d. calculated from three independent experiments. **P<0.01. Student’s t-test. (g) VCAp and NCI-H660 vector control and NEAT1-overexpressing cells were injected subcutaneously into the flank of male NOD-SCID mice. Bioluminescence imaging monitored the tumour growth. Growth curve for the tumours monitored up to 45 days is shown. VCAp (g) and NCI-H660 (h).
Figure 7 | NEAT1 in therapy resistance. (a) NEAT1 expression in VCaP cells treated with E2 (10 nM) at different time points alone or E2 + ICI (10 nM + 10 μM) or ICI (10 μM) alone. Results are expressed as the mean ± s.d. of three independent experiments. *P<0.05, **P<0.01, Student's t-test. (b) NEAT1 expression in VCaP cells treated with E2, E2 + 4OH-T (10 nM), E2 + 4OH-T (10 nM) and 4OH-T (10 nM) alone for 48 h. (c) NEAT1 expression in VCaP cells treated with or without E2 (10 nM) or E2 + Enzalutamide (100 μM) at different time points. Results are expressed as the mean ± s.d. of three independent experiments. *P<0.05 and **P<0.01, Student's t-test. (d) qRT-PCR analysis of NEAT1, G6BP and TRPM8 in LNCaP and VCaP control cells, with bisulfite treatment (10 μM) alone or in combination with E2 (10 nM) for 48 h. Results are expressed as the mean ± s.d. of three independent experiments. Results were reproducible between representative experiments. (e) Representative image for RNA ISH of NEAT1 in benign, localized PCa and in advanced disease (top panel). Quantification for the RNA ISH signals shown in the bottom panel using RNA ISH software. (f) Scatter plot showing the correlation between E2x and NEAT1 expression by qRT-PCR in nine cases of benign prostate, seven PCa and seven CRPC. Pearson's correlation coefficient: R = 0.86 (P-value = 1.9e–07).

Figure 8 | NEAT1 overexpression is associated with aggressive prostate cancer. (a) ROC curves showing (a) BCR-free survival and (b) MET-free survival for NEAT1 low and high expression groups of samples from the Mayo case-control data set (n=715). The cut points to define high and low NEAT1 expression were selected using patients from the Mayo nested case-control data set (n=393) by maximizing the product of the sensitivity and specificity for each endpoint. The number of patients at risk for each group is shown beneath the plot.

The BCR, MET, PCaM and GS > 7 endpoints (Fig. 9), NEAT1 was significantly prognostic for aggregating high-risk from low-risk patients for each of the endpoints (P<0.05). Further multivariate analysis adjusting for adjacent radiation and hormone treatment, in addition to the other clinicopathological variables assessed, also demonstrates that NEAT1 was significantly prognostic for BCR, MET and GS > 7, supporting NEAT1 as a prognostic biomarker for aggressive prostate cancer independent of common clinical and pathologic variables (Supplementary Fig. 8). Overall, these results show that NEAT1 is significantly prognostic for several clinically relevant endpoints.

Discussion

The tissue-specific role of ERα in breast and other gynaecological malignancies is well understood. Interestingly, ERα is expressed in all prostate cancers, including those that lack AR expression, which is absent in normal prostate epithelium. Studies from our and other laboratories have examined the relevance of ERα in prostate cancer19,23–25. E2-mediated regulation of oncogenic TMPRSS2-ERG fusion and oestrogen regulation of the ERα-EGFR gene, which confirms aggressive behaviour of prostate cancer, are noted examples that suggest a functional ERα-signalling pathway exists in prostate cancer. From a clinical perspective, the association of a polymorphism in ERα with prostate cancer with a favourable GS or cancers of late onset has also been reported26. These initial observations prompted us to evaluate whether re-expression of ERα and the establishment of an alternate nuclear receptor-signalling axis (that is, ERα versus AR) in prostate cancer cells could represent an adaptive mechanism to evade AR-directed therapies.

Analysis of global ERα recruitment in prostate cancer cells using a CHIP-seq approach revealed that ERα is preferentially recruited to intergenic regions of the prostate genome. Comparison of binding profiles with transcriptome sequencing data
suggested that ERs drives expression of non-coding transcripts. These results led us to analyse the functional consequences of ERs recruitment to non-coding regions. From a large compendium of ERs-regulated non-coding transcripts, we selected NEAT1 for a detailed biochemical and in vivo evaluation, based on an in vivo approach that demonstrated a strong association of NEAT1 with prostate cancer progression. We show that ERs transcriptionally regulates NEAT1. NEAT1 is recruited to the promoter of several key target genes and induces an active chromatin state favourable for transcription. Our studies indicate that ERs does not function as a molecular chaperone to guide NEAT1 to target chromatin; rather, we suspect that a complex presence of chromatin-acting proteins interacts with and guides NEAT1 to promote targets. Interestingly, both ERs and NEAT1 signalling were refractory to AR inhibitors and the lack of AR or ERs, thus indicating a functional specialization of the ERs-NEAT1 axis for prostate cancer progression. Furthermore, introduction of cells overexpressing NEAT1 could clearly induce prostate cancer progression in experimental animal models.

The current study opens up a new arena of alternative mechanisms of tumorigenesis by ERs in prostate cancer. We show that ERs regulates NEAT1 IncRNA with distinct chromatin regulatory functions. Large-scale bioinformatic analysis of SAGE libraries has identified NEAT1 as one of the differentially regulated IncRNAs between some types of cancer and normal tissue. However, its possible role in promoting tumorigenesis has never been explored. We show here that NEAT1 regulates expression of prostate cancer genes by altering the epigenetic landscape at target gene promoters to favour transcription. A closer examination of NEAT1 revealed a previously uncharacterized role in recognition of modified histones. We have not tested whether NEAT1 is a reader of multiple histone H3 post-translational modifications (acetylation, methylation and so on) and our laboratory is actively pursuing this intriguing question. NEAT1 expression independently was sufficient to activate prostate cancer genes in an AR-independent manner. Further, our results confirmed an oncogenic role for NEAT1 in an experimental animal model of prostate cancer and in cell culture models.

Molecular sieving of the net non-coding transcriptome using comprehensive bioinformatic approaches and wet-lab validation over a decade has indicated that the non-coding transcriptome has a regulatory role beyond the speculated ‘transcriptional noise’ and a direct influence on the coding transcriptome and biologic host-system. We observed that several IncRNAs such as NEAT1 respond to cellular cues and signal transduction in a manner reminiscent of the coding transcriptome. Thus far, the literature on NEAT1 has focused on its architectural role in forming subnuclear parapedes. Our results indicate a role for NEAT1 beyond that of parapedes. It would be interesting in the future to reconcile how the formation of parapedes in the inter-chromosomal space ties in with the role of NEAT1 in activating gene expression at promoters. Our lab continues to pursue some of these unanswered questions to better understand the role of NEAT1.

Our identification of an ERs-NEAT1 axis illustrates a mechanism whereby prostate cancer cells may develop therapeutic resistance through positive selection of an alternate nuclear receptor signalling pathway in the absence of AR or during androgen ablation therapy (Fig. 10). However, we cannot exclude the presence of other NEAT1-interacting chromatin factors. This is the subject of ongoing investigation.

From a clinical perspective, our studies indicate for the first time that NEAT1 is significantly prognostic for several clinically relevant endpoints. In prostatectomy specimen from two large cohorts, high NEAT1 expression was associated with a significant increase in both biochemical and metastatic recurrence rates compared to those with low NEAT1 expression.

In summary, this study provides important insights into a unique mechanism of ERs regulation in prostate cancer and identifies NEAT1 as a novel prognostic marker and potential therapeutic target in this disease. Although our studies have identified a previously unexplored function of ERs in regulating IncRNAs, it is also the first of its kind to demonstrate transcriptional regulation of IncRNAs by an alternative steroid receptor in prostate cancer. We propose that NEAT1 is directly involved in modulation of the phenotype of a leading disease. Combinatorial targeting of NEAT1 and AR may represent a unique therapeutic regimen within a subset of patients with advanced prostate cancer.

**Methods**

**Cell culture and treatments.** IncRNA and PGC1α cells were grown in RPMI 1640 (Invitrogen) and supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. SW756 and Ad5.16 cells were grown in Keratinocyte-Serum-Free Medium (Invitrogen, 17050-046). VCaP and DU145 cells were grown in DMEM (Gibco) and supplemented with 10% FBS with 1% penicillin-streptomycin. CLEVER-1060 cells were grown in RPMI 1640 supplemented with 10% FBS and 1% penicillin-streptomycin. 10 nmol/l L-arginine, 0.04 mg/ml transferrin, 20 nmol/l sodium selenite, 20 nmol/l hydrocortisone, 10 nmol/l β-estradiol, 5% FBS, 1% penicillin-streptomycin and an extra 1 nmol/l of 5-bromoindole (as a final concentration of 50 μM). For cell treatments in some experiments, we used 10-100μM β-estradiol (Sigma Aldrich), 5% Lentinus Edulis, 10 μM Tocopherol (Sigma Aldrich), 1-10 nmol/l (Sigma Aldrich), and 1-10 μM (Sigma Aldrich), 10-100 μM (Sigma Aldrich), and 1-10 μM (Sigma Aldrich).

**Plasmids, siRNAs and transfection.** Plasmids, pcDNA 3.1, pDNA3.1-ERS, pDNA 3.1 AI, plasmid GFP, plasmid-NEAT1 siRNA-GERF (set of four, sequences provided in Supplementary Table S3), Lenti-expressing pLV-Tet-Neo vector, pcDNA-NEAT1 were used. siRNA for ERS, ERE, AI, NEAT1, and NEAT1_2 were used, and the sequence is provided in Supplementary Table S3. For the mammalian expression vectors, Lentiexpressor 3000 (Invitrogen) and Lentiviruses infection were...
Figure 10 | Model for NEAT1 function in prostate cancer. Functional Ets signaling in prostate cancer modulates expression of the lncRNA NEAT1. Prostate epithelial cells positive for NEAT1 have an oncogenic advantage and are resistant to androgen withdrawal or androgen ablation therapy. NEAT1, a histone interacting lncRNA and transcriptional regulator, is recruited to promoters of several prostate cancer-specific genes. NEAT1 can modulate the epigenetic landscape of target promoters and maintains expression of AR-dependent and independent genes. The selection of alternate nuclear receptor signaling is a novel hallmark of prostate cancer progression.

Identification of Ets-regulated lncRNAs. A set of known lncRNAs was generated from various data sources: RefSeq, GENCODE, v. 17, lncRNA.org and lncRNAdb. (see Supplementary Material) and those that were at least 200 bp long were selected, resulting in 2,441 lncRNAs. These lncRNAs were discounted according to their potential of being regulated by Ets by using Ets-binding site information from CisBP-Net experimental data. Moreover, several histone marks were considered to provide evidence of transcription, including H3K4me3 and H3K27ac (details in Supplementary Material).

Differential expression analysis. To test the experimental validation of lncRNAs, a high-throughput expression analysis was performed in the expression values determined by paired-end RNA sequencing of 79 samples (25 benign, 20 PCA and 34 MPCA). A pairwise Wilcoxon test was performed and all P-values were corrected for multiple hypothesis testing using Benjamin-Hochberg (details in supplementary material).

Ets and NEAT1 signature via Oncogene concept analysis. RNA sequencing data for PCA and VGP lncRNA-expressing cells, as well as some control and NEAT1-overexpressing VGP cells (details in Supplementary Methods). The expression of the genes was computed and those genes with a high FDR change ≥ 2 were selected. Results are reported in Supplemenary Tables 3 and 5. Five hundred and eight-eight genes were found to be upregulated in VGP lncRNA cells. A custom concept of this gene list was generated in Oncogene (Supplemental Figure 4). Similarly, genes from the VGP lncRNA group with a high FDR change ≥ 2 were selected and a custom concept was built in Oncogene using the top 1,000 genes from NEAT1 signature (Supplemental Figure 4). The significantly associated tumor-versus normal concept with odds ratio > 2.0 and P < 1 × 10⁻⁶ was considered tumor-versus normal analysis was determined. The resulting concepts and associations are represented through a concept network using Cytoscape version 2.8.2. Each node represents a concept to which the signature is associated as a > 5:1 odds ratio for lncRNA signature and > 2:1 odds ratio for NEAT1 signature. Node size represents the concept size, that is, the number of genes in each concept. Red and green colors represent correlations with over- or underexpressed genes in the concept, respectively, and edge thickness represents the odd ratio of the association between concepts, ranging from 3.4 to 26.9 and 1.2 to 18.7 for Ets and NEAT1 signatures, respectively. The border color of each node represents the tumor type. The layout of the network is based on the single-weighted spring-embedded algorithm.

Luciferase reporter assays. For lncRNA luciferase assays, VGP cells were transiently transfected with the NEAT1 or NEAT1\_fl (control) together with lncRNA luciferase reporter constructs, as well as internal control construct 2 (RL) harboring the luciferase luciferase gene. VGP cells were also transfected with empty vector or NEAT1prom. G5 + 200 ng luciferase reporter construct alone or with Ets, as well as an internal control construct 2 (RL) harboring the luciferase gene. To determine the PMSA reporter activity, 20T and PC3 cells were co-transfected with empty vector or PMSA luciferase reporter construct alone or with NEAT1, NEAT1\_fl or NEAT1 - AR. Twenty-four hours post-transfection, the media was changed to 5% charcoal-stripped media and the cells were then treated with 500 ng/mL of E2 (10M) or R1881 (10M) for 48 h. At 48 h, cells were [3H]thymidine labeled and luciferase activities were measured using the dual luciferase system (Promega). Presented and normalized with luciferase activity.

RNA IHS for NEAT1. RNA IHS for NEAT1 was performed on formalin-fixed, paraffin-embedded tissues from 117 cases of PCA and seven CRPC cases using custom IHC and FISH probes designed by Advanced Cell Diagnostics. Briefly, the single-color chromogen detection assay was performed at a specially designed oligonucleotide probe that probe sequences specific hybridization, recognizes both the specific target RNA sequence and the signal amplification system. Unique target probe oligonucleotides were designed to hybridize to targets to the target RNA. Once hybridization to other sequences is completed, binding of the probe RNA to the reporter probe is ensured by probe RNA association, which is then catalyzed through sequential hybridization. Signal amplification occurs at target sites bound by probe RNA association. Specificity of target binding by single probes does not result in signal amplification. All steps of RNA RNA IHS staining of the slides were performed manually, optimized in tissue microarrays. Briefly, formalin-fixed, paraffin-embedded tissue sections (3 μm) were deparaffinized in xylene and subjected through a series of ethanol. The deparaffinized sections were incubated with 0.3% hydrogen peroxide at room temperature (HT) for 10 min to block endogenous peroxidase. Sections were then blocked in 1 x
The statistical analysis was performed using...


My contribution to this work was performing experiments aimed at proving that cells treated with the proteasome inhibitor MG132 upregulate NEAT1 through the activation of the $NEAT1$ promoter (Figure 3E in the paper attached below). More specifically, a luciferase reporter vector containing $NEAT1$ promoter (generated from Chapter 4) was transfected into the cells. The cells were then stressed with MG132, RNA was extracted so that the mRNA level of luciferase expression could be measured by qRT-PCR and the data assembled into the figure.
NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies

Tetsuro Hirose1,*, Giorgio Vincinetti1, Akie Tanigawa1, Takao Nagasuma2, RuoHua Li3, Hiroshi Kimura4, Takahide Yokoi5, Shinichi Nakagawa5, Marianne Bénard6, Archa H. Fox1, and Gérard Pierron2

1Institute for Genetic Medicine, Hokkaido University, Sapporo 060-0815, Japan; 2Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0004, Japan; 3Centre National de la Recherche Scientifique, UMR 7122, Institut Gustave Roussy, Villejuif 94805, France; 4Western Australian Institute for Medical Research, Center for Medical Research, University of Western Australia, Perth, Western Australia 6001, Australia; 5Graduate School of Frontier Biosciences, Osaka University, Suita 565-0871, Japan; 6Mitsui, Kokubunji, Tokyo 185-0001, Japan; 7RIKEN Advanced Institute, Wako 351-0195, Japan; 8Centre National de la Recherche Scientifique, FRE 3402, UPMC Université Paris 06, Paris 75252, France

ABSTRACT Paraspeckles are subnuclear structures formed around nuclear paraspeckle assembly transcript 1 (NEAT1)/MEN5B long noncoding RNA (lncRNA). Here we show that paraspeckles become dramatically enlarged after proteasome inhibition. This enlargement is mainly caused by NEAT1 transcriptional up-regulation rather than accumulation of undegraded paraspeckle proteins. Of interest, however, using confocal microscopy, we find that key paraspeckle proteins become effectively depleted from the nucleoplasm by 50% when paraspeckle assembly is enhanced, suggesting a sequestration mechanism. We also perform microarrays from NEAT1-knockdown cells and find that NEAT1 represses transcription of several genes, including the RNA-specific adenosine deaminase B2 (ADARB2) gene. In contrast, the NEAT1-binding paraspeckle protein splicing factor proline/glutamine-rich (SFPO) is required for ADARB2 transcription. This leads us to hypothesize that ADARB2 expression is controlled by NEAT1-dependent sequestration of SFPO. Accordingly, we find that ADARB2 expression is strongly reduced upon enhanced SFPO sequestration by proteasome inhibition, with concomitant reduction in SFPO binding to the ADAR2 promoter. Finally, NEAT1+ fibroblasts are more sensitive to proteasome inhibition, which triggers cell death, suggesting that paraspeckles/NEAT1 attenuates the cell death pathway. These data further confirm that paraspeckles are stress-responsive nuclear bodies and provide a model in which induced NEAT1 controls target gene transcription by protein sequestration into paraspeckles.

This article was published online ahead of print in M&BC in Press (http://www.nature.com/mib; doi:10.1038/mi.2013.95) on October 28, 2013.

Address correspondence to: Tetsuro Hirose, Hirose@gen.hokudai.ac.jp; Archa Fox, fox.004@neu.edu, Archa Fox (archa Fox@neu.edu). Gérard Pierron (gerard.pierron@igfr.fr)

Abbreviations used: ADAR2, adenosine deaminase RNA-specific P2; AOS, antisense oligonucleotides; EJ, early type line; LNA, long noncoding RNA; NEAT, nuclear paraspeckle assembly transcript; NEAT1, nuclear paraspeckle assembly transcript 1; qRT-PCR, quantitative reverse transcription-PCR; RT-PCR, RNA polymerase II; SFPO, splicing factor proline/glutamine-rich.

© 2014 Hirose et al. This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0/). ASCB, The American Society for Cell Biology, and Molecular Biology of the Cell® are registered trademarks of The American Society for Cell Biology.

INTRODUCTION

The nucleus of mammalian cells is highly organized; it is composed of distinct nuclear bodies—membrane-less organelles containing specific proteins or RNAs characteristic of particular nuclear processes (Spector, 2006). Paraspeckles are nuclear bodies detected in mammalian cells as a variable number of foci found in close proximity to nuclear speckles (Vieira et al., 1995; Fox et al., 2002). Paraspeckles were initially defined as foci enriched in characteristic RNA-binding proteins, including the three mammalian Drosophila melanogaster behavior and human splicing (DBHS) proteins, PSF1, NONO (g54anb), and splicing factor proline/glutamine-rich (SFPO; PSF, Fox et al., 2002, 2005; Prasanth et al., 2005), Beyond their localization in
paraspeckles, the DBHS proteins have been implicated in numerous nuclear processes, including transcriptional control, RNA processing, and DNA repair (Sho-Tal and Zipori, 2002; Dong et al., 2007; Kaneoke et al., 2007; Bond and Fox 2009; Li et al., 2009; Hoyd and Lynch, 2010).

A number of relatively abundant long noncoding RNAs (lncRNAs) have been found to localize specifically to paraspeckles (Clouston et al., 1996; Hutchinson et al., 2007; Sone et al., 2007; Sasaki et al., 2009; Sunwoo et al., 2009; Tippat et al., 2010; Zheng et al., 2010; Xiong et al., 2011). Prominent among these is the nuclear paraspeckle assembly transcript 1 (NEAT1) lncRNA, which is found to localize specifically to paraspeckles, where it forms an essential structural component (Chen and Camathias, 2009; Clouston et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009; Tippat et al., 2010). Prominent among these is the nuclear paraspeckle assembly transcript 1 (NEAT1) lncRNA, which is found to localize specifically to paraspeckles, where it forms an essential structural component (Chen and Camathias, 2009; Clouston et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009; Tippat et al., 2010). Prominent among these is the nuclear paraspeckle assembly transcript 1 (NEAT1) lncRNA, which is found to localize specifically to paraspeckles, where it forms an essential structural component (Chen and Camathias, 2009; Clouston et al., 2009; Sasaki et al., 2009; Sunwoo et al., 2009; Tippat et al., 2010).

RESULTS

Proteasome inhibition induces grossly enlarged paraspeckles

To uncover the role of paraspeckles, we searched for conditions in which paraspeckle function might be enhanced. We found that paraspeckles were markedly enlarged when cells were treated with proteasome inhibitors. By a combination of RNA fluorescence in situ hybridization (FISH) and immunostaining to simultaneously detect NEAT1 and one of the paraspeckle marker proteins, PSCP1, we observed grossly enlarged paraspeckles in HeLa cells treated with MG132 for 6 or 14 h (Figure 1A and Supplemental Figure S1). In all of the enlarged paraspeckles observed, the NEAT1 and PSCP1 signals completely overlapped.

Electron microscopic (EM) studies revealed that these enlarged paraspeckles correspond to clusters of typical paraspeckles (Figure 1B). Even after 17 h of MG132 treatment, as in Figure 1B, paraspeckle ultrastructure was unchanged as determined by EM, but their frequency was markedly increased. When HeLa cells were treated with the structurally different proteasome inhibitor bortezomib (100 nM, 17 h), a similar increase of paraspeckle frequency was observed (Figure 1B), indicating a true dependence of paraspeckle inhibition in this phenomenon. Paraspeckle clustering, as seen in the EM studies, suggested elongated and twisted structures cut several times by thin sectioning. Thus we measured the long and short axes (L x S) of 120 paraspeckles in EM sections from control and MG132-treated HeLa cells (examples shown in Figure 1C). Plotting the values by increasing L x S (Figure 1D) illustrates a significant MG132-dependent paraspeckle elongation (mean length of 635 nm for MG132-treated paraspeckles vs. 464 nm for control, p < 0.001), whereas the constant Sx values indicate a similar constrained diameter in control and treated cells (mean 320 ± 36 and 312 ± 41 nm, respectively). Thus the enlargement of paraspeckles that we observed was due to significant elongation.

By immunogold EM-EM, we analyzed the distribution of proteins that accumulate upon proteasome inhibition and found no indication of enrichment in paraspeckles. After MG132 or bortezomib treatment, dense cytoplasmic and nuclear aggregates were conspicuous. Cytoplasmic aggregates formed around controls and were heavily labeled with an ubiquitin antibody, indicative of aggregates (Supplemental Figure S2). Nuclear aggregates, highly enriched in SUMOylated proteins, were always found closely associated with the nucleolus (Figure 2A) but, strikingly, did not overlap with paraspeckles (Figure 2A). Finally, the aggregates condensed into large, dense bodies concentrating ubiquitin and SUMO-1 and SUMO-2/3 at their periphery (Figure 2B). These MG132-induced nuclear bodies were surrounded by a thin layer of peroxisomal proteins. Thus formation of elongated paraspeckles and the accumulation of proteins by proteasome inhibition are compartmentalized, unrelated events.

Transcriptional up-regulation of NEAT1 causes elongated paraspeckle formation in proteasome-inhibited cells

In a search for factors controlling elongated paraspeckle formation, we determined by Western blotting that seven PSCPs, each essential for paraspeckle formation, did not increase with MG132 treatment (Figure 2C). By I-EM, we determined that endogenous NONO, CPSF6, and SF3A2 were similarly distributed and similarly abundant within the paraspeckles of control (dimethyl sulfoxide [DMSO]) and MG132-treated HeLa cells (illustrated in Figure 2D and quantified in Figure 2E). Taken together, these data indicate that elongated paraspeckles are unlikely to result from an unusual

170 | T. Hove et al. | Molecular Biology of the Cell
FIGURE 1: Proteasome inhibition results in paraspeckle elongation. (A) Paraspeckle enlargement upon proteasome inhibition. HeLa cells were treated with DMSO or 5 μM MG132 for 17 h. The paraspeckles were detected by RNA-FISH of NEAT1 IncRNAs (magenta) combined with immunofluorescence of PSPC1 (green). Scale bar, 10 μm. (B) Paraspeckle ultrastructure upon proteasome inhibition. HeLa cells treated with DMSO, 5 μM MG132, or 100 nM bortezomib for 17 h were analyzed by transmission EM after Epon embedding. Left, large fields; bar, 2 μm. Right, enlargement of dashed rectangles; bar, 0.5 μm. Red arrows, paraspeckle clusters; IG, interchromatin granules. (C) Size measurements of paraspeckle in control and MG132-treated HeLa cells (5 μM, 17 h). Short and long axes (Sx and Lx) were defined for 120 paraspeckles in each sample from Epon-embedded HeLa cells. Bar, 0.5 μm. (D) Dimensions were plotted by increasing Lx values, showing similar Sx values in both samples, whereas Lx values were augmented in the MG132-treated cells.

accumulation or repositioning of PSFs. Next, we investigated whether levels of the essential paraspeckle RNA, NEAT1, were affected by proteasome inhibition. RNAase protection assays and quantitative reverse transcription-PCR (qRT-PCR) measurements of NEAT1_1 and NEAT1_2 levels revealed that both isoforms significantly increased upon MG132 treatment (Figure 3, A and B), showing greater than eightfold increase in NEAT1_1 after 17 h MG132. RNAase protection assays showed that the kinetics of up-regulation is faster in NEAT1_2 than NEAT1_1 (Figure 3, A and B). To investigate whether this increase was transcriptional or posttranscriptional, we quantified the newly synthesized nascent NEAT1 mRNA. For capturing nascent RNAs, HeLa cells were pulse labeled with 5-ethyl-uridine (EU) for 1 h, and the EU-incorporated RNAs were biotinylated and purified with streptavidin-conjugated beads. qRT-PCR of the captured RNAs revealed that nascent NEAT1 RNA levels were similar to total RNA (steady-state levels) in control and MG132-treated cells (Figure 3C), indicating that increased NEAT1 was resulting from new transcripts. Chromatin immunoprecipitation (ChIP) with anti-RNA polymerase II (RNPAP1) phosphorylated at serine 5 in the carboxy-terminal domain (phospho-CTD-ser5) showed MG132 treatment resulted in a significant increase of RNPAP phospho-CTD-ser5 bound within the NEAT1 promoter but not the glyceroldehyde-3-phosphate dehydrogenase (GAPDH) promoter (Figure 3D). Further, a luciferase reporter gene driven by the human NEAT1 promoter was significantly activated by MG132 treatment, whereas a control SV40 promoter was not (Figure 3E).

Finally, by in situ hybridization of the EML, we determined that the known distinct location of each NEAT1 isoform within paraspeckles (Bouquelet et al., 2010) was maintained upon MG132 stimulation. As shown in Supplemental Figure S3, NEAT1_1 and the indistinguishable 5’ end of NEAT1_2 were, with the 3’ end of NEAT1_2, restricted to the periphery of the paraspeckle. In contrast, the internal sequence of NEAT1_2 (labeled as D1 in Supplemental Figure S3) was located within paraspeckles in control and MG132-treated HeLa cells. Collectively, these data show that paraspeckle elongation upon proteasome inhibition is driven by transcriptional up-regulation of the NEAT1 gene rather than paraspeckle protein accumulation or a change in ultrastructural organization of key paraspeckle components.

SFPQ and NONO are sequestered within elongated paraspeckles upon proteasome inhibition.

We next examined the relative amounts of paraspeckle proteins trapped within these MG132-induced elongated paraspeckles. Labeling densities of NONO, SFPQ, and CPSF6, as measured by I-EM in Figure 2, were comparable in paraspeckles of control and MG132-treated HeLa cells. However, normalizing the fourfold to fivefold increased frequency of paraspeckles on thin sections of proteasome-inhibited cells (see Materials and Methods), our results imply that four to five times more of NONO, CPSF6, and SFPQ are contained within the MG132-induced elongated paraspeckles. We next asked whether this increased inclusion within the enlarged paraspeckles is sufficient to deplete a significant proportion of these proteins from the nucleoplasm. By I-EM, we compared labeling densities of NONO or SFPQ in cytoplasmic, nucleoplasmic, and paraspeckle areas in control and MG132-treated cells. Our results indicate that in MG132-treated cells the nucleoplasmic pools of SFPQ and NONO were depleted by roughly 50%.
FIGURE 2: Proteasome inhibition does not result in accumulation of ubiquitinated proteins or reorganization of protein components within paraspckles. (A) EM detection of ubiquitinated and SUMOylated proteins accumulating upon proteasome inhibition (thin sections of Lowry-embedded HeLa cells; MG132 treated 5 μM, 17 h). Left, dense aggregations (black arrow) of nuclear ubiquitinated proteins are formed in the vicinity of the nucleolus (No). In contrast, the expanded paraspckles (red arrows) do not contain significant amount of ubiquitin. Right, SUMOylated proteins also accumulate in the ubiquitin-positive nuclear aggregations. Bars, 0.5 μm. (B) Some nuclear sections, MG132-induced protein aggregates form well-defined nuclear bodies, highly enriched at their periphery in ubiquitin- and SUMO-conjugated proteins and surrounded by a thin layer of PML protein. Bars, 0.5 μm. (C) The PSP levels were largely constant upon MG132 treatment. Seven essential PSPs (defined as those proteins whose knockdown in HeLa cells result in loss of paraspckles) were detected in cells treated with DMSO or MG132 for 6 or 17 h by Western blotting. Category 1A proteins (1A) are required for both paraspckle integrity and NEAT1, 2 accumulation, whereas category 1B proteins (1B) do not affect NEAT1, 2 levels but are required for paraspckles. α-tubulin is the control. (D) Abundance and distribution of NONO, CPSF6, and SFPQ in control and MG132-ambient paraspckles was determined by IEM. Red arrows indicate highly labeled paraspckles. Bars, 0.5 μm. (E) Labelling densities (gold particles/μm²) of the three PSPs were quantified from the images shown in D, displaying a constant level of protein per unit surface of paraspckle. At least 20 paraspckles or 500 gold particles were counted for each sample. Bars, 0.5 μm.

compared with control values (Figure 4). In contrast, relatively constant labeling values were obtained within the cytoplasmic and paraspckle compartments. We conclude that NEAT1 up-regulation by proteasome inhibition leads to NONO and SFPQ sequestration within enlarged paraspckles. Given that the nucloplasmic pools of SFPQ and NONO were affected by 50%, in a manner akin to haploinsufficiency, we anticipated that NONO and SFPQ sequestration within paraspckles was likely to have pronounced effects on transcription.

Identification of the adenosine deaminase RNA-specific B2 gene as a paraspckle-target gene

To complement these investigations into paraspckle enlargement, we then examined paraspckle loss. Specific depletion of NEAT1...
FIGURE 3: Proteasome inhibition activates NEAT1 IncRNA synthesis. (A, B) Proteasome inhibition results in elevation of NEAT1 IncRNA levels detected by qRT-PCR (A) and RNase protection assay (RPA, B). Quantitation of the RNase-protected bands is shown at the bottom. (C) Proteasome inhibition elevates the level of the newly synthesized nascent NEAT1 IncRNAs. Total and EU labeled nascent RNAs prepared from the cells treated with DMSO or MG132 for 6 h were used for quantitation of NEAT1. For qRT-PCR, two primer sets corresponding to the NEAT1.1/1.2-overlapping region (NEAT1.1/1.2) and the NEAT1.2-specific region (NEAT1.2) were used. (D) Elevated binding of RNP11 phospho CTD serine 5 at the NEAT1 promoter in MG132-treated cells. Chromatin immunoprecipitation with RNP11 phospho CTD serine 5 antibody was carried out in cells treated with DMSO or MG132 for 6 h. Levels of communoprecipitated DNA fragments were quantified by qPCR with primer sets that span promoter regions of NEAT1 and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as control. **p < 0.01, *p < 0.05. (E) The NEAT1 promoter is activated by MG132. The graph shows qPCR of luciferase cDNA, relative to GAPDH, normalized to the control (untreated sample). RNA was extracted from MG132- and DMSO-treated HeLa cells transiently transfected 48 h before with plasmids encoding luciferase driven by the promoters as indicated. Error bars are SD as a result of two biological replicates.

with a chimeric antiserum oligonucleotide (ASC) leads to disintegration of paraspeckles (Sasaki et al., 2009). To identify paraspeckle-target genes, we carried out microarray analyses using total RNA samples from HeLa cells treated with control (GFP) or NEAT1 ASC (ASC2) for 6, 12, and 24 h (Figure S4A). A modest total of 51 genes overrepresented or underrepresented greater than twofold upon NEAT1 reduction were detected from these three time points (Supplemental Table S1). We confirmed this result with qRT-PCR for several genes up-regulated by NEAT1 depletion (Figure S5B) and found that they were reproducibly up-regulated in cells treated with different NEAT1 ASCs (ASC1#12 and ASC1#17; Supplemental Figure S4A) shows the efficacy of the ASCs on NEAT1 levels. Given the intriguing link between one of these genes—the adenosine deaminase RNA-specific B2 (ADARB2) gene function in RNA editing—and the previous observation that
parasporides were involved in retaining edited RNA, we chose ADA3R2 for detailed characterization of a parasporide-target gene. First, Northern blot hybridization detected four ADA3R2 mRNA isoforms (5.5–8.5 kb) generated by different polyadenylation sites in HeLa cells, all of which increased upon NEAT1 IncRNA silencing (siNEAT1) with two different ASOs (lanes 12 and 17 in Figure 5C), suggesting elevation of ADA3R2 transcription rather than altered posttranscriptional regulation at heterogeneous 3′ ends. To investigate this possibility, we quantified the level of nascent ADA3R2 mRNA in control and siNEAT1 cells. The pulse-labeling experiment (as in Figure 3C) revealed that nascent ADA3R2 mRNA was elevated similar to that of total ADA3R2 mRNAs (Figure 5D). The stability of ADA3R2 mRNA (measured by degradation after actinomycin D treatment) was unaffected in the presence and absence of NEAT1 (EG5 in Figure 5E). We also determined that the subcellular localization of ADA3R2 mRNA was not altered when NEAT1 was knocked down (Supplemental Figure 5F). These results argue that ADA3R2 is a parasporide-target gene, transcriptionally up-regulated when NEAT1 levels and parasporide integrity are lowered.

The parasporide protein SPFQ is required for transcription of ADA3R2.

In a search for a link between NEAT1 levels, parasporide assembly, and ADA3R2 gene transcription, we independently silenced each of the 32 PSpSs that are expressed in HeLa cells, using validated small interfering RNA (siRNA) in which >70% knockdown is achieved (Naganuma et al., 2012). qRT-PCR of ADA3R2 mRNA levels revealed that SPFQ and, to a lesser extent, HRNPH1 were required for ADA3R2 gene expression (uSPFQ, ADA3R2 −/−/−; SPFQ, ADA3R2 −/−/− of control; siHRNPH1, −20%; Figure 6A). This effect was not observed in control GAPDH mRNA (Supplemental Figure 5E). To identify the stage of ADA3R2 gene expression that is suppressed by SPFQ and HRNPH1, we captured newly synthesized mRNAs as in Figure 5D in control, uSPFQ, and siHRNPH1 cells. Nascent ADA3R2 mRNA levels were markedly reduced in both uSPFQ and siHRNPH1 cells (Figure 6B). Finally, Northern blotting showed that all isoforms of ADA3R2 mRNA were strongly reduced upon SPFQ elimination (Figure 6C). Taken together, these results suggest that SPFQ and HRNPH1 have an activating role in ADA3R2 gene transcription.

NEAT1 IncRNA sequesters SPFQ into parasporides away from the ADA3R2 gene promoter.

ADA3R2 thus appears to be a parasporide-target gene negatively regulated by NEAT1 but positively regulated by the parasporide components SPFQ and HRNPH. We anticipated that NEAT1 up-regulation by proteasome inhibition and the ensuing sequestration of SPFQ within enlarged parasporides would suppress and strongly repress ADA3R2 gene expression. Accordingly, qRT-PCR showed that a 17h treatment with MG132 or bortezomib led to 10- to 20-fold-down ADA3R2 RNA levels (Figure 7A). The down-regulation of ADA3R2 gene expression likely occurs at the transcription level, since there is a comparable reduction of ADA3R2 mRNA observed after 6h MG132 treatment in both total and nascent RNA (Figure 7B). To determine whether down-regulation of ADA3R2 is coupled to enlargement of parasporides, we measured ADA3R2 mRNA levels in control (uSPFQ) and siNEAT1 cells treated with MG132. The suppression of ADA3R2 gene expression was significantly milder in siNEAT1 cells (Figure 7C), indicating that ADA3R2 gene transcription by proteasome inhibition is, at least in part, NEAT1 dependent.

These data led us to propose a model to explain the negative effect of NEAT1 on ADA3R2 levels: namely that NEAT1 sequesters SPFQ within parasporides, thereby depleting SPFQ from the nucleoplasm, limiting the availability of SPFQ to the ADA3R2 gene (Figure S4A). To test this assumption, we looked at binding of SPFQ to both NEAT1 RNA and the ADA3R2 gene locus. First, by RNA immunoprecipitation (RIP), we observed an increased amount of SPFQ bound to NEAT1 after MG132 treatment (Figure 7B); despite SPFQ levels remaining constant (Figure 8, B and C), although this is perhaps not surprising, given the overall increased NEAT1 levels with MG132 (e.g., Figure 3A). Of interest, the increased association of SPFQ and NEAT1 was more enhanced for a NEAT1-specific region (NEAT1_241 and NEAT1_242, 3.5-fold) than in the NEAT1_111-2 overlapping region (NEAT1_111 and NEAT1_242, 2-fold; Figure S4A), which is relevant because NEAT1_2 and not NEAT1_1 is the critical isoform for parasporide formation. However, immunofluorescence and electron microscopic in situ hybridization (EM-SH) show that both NEAT1 isoforms, even when up-regulated by MG132 treatment, were strongly enriched within the parasporides (Supplemental Figure 5B). Next, SPFQ association with the ADA3R2 gene locus was investigated by ChIP. As shown in Figure 8C, SPFQ specifically associated with the ADA3R2 promoter region spanning 127 nt upstream of the transcription start site (−127) in control cells; however, the association markedly dropped in MG132-treated cells (Figure 8C). These results argue that SPFQ is depleted from the ADA3R2 promoter region and nucleoplasm upon MG132
FIGURE 5: Intact paraspeckles act to suppress transcription of the ADARB2 gene. (A) Experimental strategy to identify the genes controlled by intact paraspeckles. Total RNAs were prepared from HeLa cells treated with either GFP ASO or NEAT1 ASO #12 for microarray analysis. The cells were harvested for RNA preparation 6, 12, and 24 h after ASO administration. The list of the target gene candidates is shown in Supplemental Table S1. (B) qRT-PCR validation identified five genes that were reproducibly up-regulated by two NEAT1 ASO #12 and #17 treatments. (C) Northern blot analysis of ADARB2 mRNAs in NEAT1-silenced cells (5NEAT1). PolyA+ RNAs prepared from HeLa cells treated with ASCs (green fluorescent protein [GFP] as a control, #12 and #17 for knockdown of NEAT1 IncRNA) were separated by electrophoresis on 1% agarose gels. The schematics of four putative ADARB2-mRNA isoforms (A~D) are shown. GAPDH mRNA is the loading control. (D) NEAT1 IncRNA elimination elevates the level of the nascent ADARB2 mRNA. Total RNA and pulse-labeled RNA with EU for 1 h (nascent RNA) were used as template for qRT-PCR to quantify ADARB2 mRNAs. (E) ADARB2 mRNA stability is unaffected upon NEAT1 IncRNA elimination. The mRNA decay curves (as measured by qRTPCR) in HeLa cells treated with either GFP (solid lines) or #12 ASO (dashed lines) after actinomycin D treatment are shown. The two primers sets to detect the coding region of ADARB2 mRNA (CDS) and the 3’ UTR were used.
FIGURE 6: Two PSPs are required for transcription of the ADAR82 gene. (A) Identification of the PSPs required for ADAR82 gene expression. qRT-PCR to monitor ADAR82 mRNA level was carried out using RNA samples obtained from Hela cells in which each of the PSPs was reduced by RNAi (Nagamura et al., 2012). Two primer sites (CDS and UTR) were used for ADAR82 qRT-PCR. The ADAR82 mRNA level in the cells treated with control siRNA was adjusted to 1.0. The control experiment to monitor GAPDH mRNAs is shown in Supplemental Figure S6. (B) Knockdown of SFPQ and HNRNPH1 down-regulates synthesis of nascent ADAR82 mRNA. Total RNA and EU pulse-labeled RNA (nascent RNA) from control (NC) and siRNA-treated cells were used for quantification of ADAR82 mRNAs by qRT-PCR. SFPQ and HNRNPH1 were knocked down with siRNAs #14 or #19 and H1#1, respectively. (C) Northern blot analysis of ADAR82 mRNAs in control (NC) and SFPQ knock-down (#14 and #19) cells. ADAR82 isoforms are labeled as in Figure 5C.

treatment, in parallel with elevated NEAT1 binding within enlarged paraspeckles.

To determine whether this sequestration mechanism alters expression of other loci, we examined the four other selected genes (Figure 5B), which, like ADAR82, were enhanced by NEAT1 KD. Like ADAR82, these four genes were repressed to some extent by MG132, and these were dependent on SFPQ for their expression, as shown by SFPQ silencing (Figure 9). This shows that the mechanisms of SFPQ sequestration by paraspeckle-elongation is likely to be of general significance. The full extent of genes regulated by PSP sequestration upon proteasome inhibition is yet to be measured but is likely important, considering the diversity of transcription factors that, beyond NONO and SFPQ, are possibly trapped within enlarged paraspeckles.

Possible role of paraspeckles in response to proteasome inhibition

In a search for cellular functions altered by enlarged paraspeckle formation in response to proteasome inhibition, we used MEFs prepared from Neat1+/− knockout mice (Neat1+/−/−, Nakagawa et al., 2011). First, we confirmed that Neat1 IncRNA was also up-regulated by MG132 treatment in MEFs (Figure 10A). In particular, Neat1_2 is significantly increased after 3 and 6 h of MG132 treatment. Next we monitored the growth of Neat1+/− MEFs or Neat1+/− MEFs, treated either with DMSO (control) or MG132 over 48 h. Cell growth was measured using the xCELLigence system (Roche), which derives a “cell index” correlated with the number of cells still attached to the dish (Figure 10B). These data show that primary MEFs from both Neat1+/− and Neat1+/− mice die with MG132
treatment, consistent with the induction of apoptosis under these conditions as reported previously (Jiang and Weak, 2005). Of interest, however, the MG132-treated Neat1\(^{-/-}\) MEFs consistently displayed an overall less severe effect compared with the Neat1\(^{+/-}\) MEFs (Figure 10B). We also noted that in the first 4 h after addition of MG132, the Neat1\(^{+/-}\) MEFs started to die very quickly, with an almost immediate difference in growth for the control cells compared with the MG132-treated cells (Figure 10C; compare black and green plots). In contrast, the Neat1\(^{-/-}\) MEFs had indistinguishable growth between the two treatments over the first 1–2 h of the experiment, after which they too began to die in a dose-dependent manner (Figure 10C; also Supplementary Figure S7). The effect, although transient, was consistent between experiments and different batches of MEFs (Supplemental Figure S7 shows the combined data from three biological replicates). The same pattern of more severe and quicker cell death in the NEAT1\(^{-/-}\) MEFs was also observed with different doses of inhibition (Figure 10D). These results prompted us to determine whether the ADAR2B gene is implicated in NEAT1\(^{-/-}\) cell-death sensitivity, but we could not detect ADAR2B expression in MEFs, even when NEAT1 was deleted. Thus it is likely that other genes and mechanisms are at work for a NEAT1 pro-survival role upon proteasome inhibition in MEFs. Together these cell analyses suggest an increased sensitivity of Neat1\(^{-/-}\) MEFs to proteasome inhibitors, which is particularly evident in the first 2.5 h of treatment, and provide an indication that proteasome inhibitor-induced up-regulation of NEAT1 plays a role in early apoptotic defense.

**DISCUSSION**

Since their discovery in 2002, the function of paraspeckles has remained largely elusive. Here we show that paraspeckles are stress-inducible structures that modulate gene expression through sequestration of transcriptional regulators (Figure 8).

**Protein sequestration within proteasome inhibition-induced elongated paraspeckles**

Paraspeckles are markedly elongated as a result of NEAT1 transcriptional up-regulation upon proteasome inhibition. Our luciferase-reporter assay showed that NEAT1 promoter activity is increased in these conditions, and by EM we showed that elongated paraspeckles neither were enriched in ubiquitin-conjugated proteins nor did they have an increased density of P-S-F (although they do sequester a greater proportion of the total pool of P-S-F due to their increased length). Thus we attribute paraspeckles enlargement to increased NEAT1 levels, consistent with reported NEAT1 up-regulation and paraspeckle enlargement taking place during in vitro myo-tube differentiation (Suzuki et al., 2007) and prolonged expression of a tagged NEAT1 transgene (Mao et al., 2017). Our stepwise analysis further shows that NEAT1 synthesis is a rate-limiting step for paraspeckle formation. We confirm that paraspeckle width is fixed in HeLa cells, likely reflecting the molecular size of NEAT1, as we suggested previously (Souquere et al., 2010), and we further show that paraspeckle length is correlated with the amount of NEAT1 transcribed within the cell. It remains to be determined how de novo paraspeckles reach their normal optimal length, presumably as part of their assembly process at the NEAT1 locus. Proteasome inhibition may lead to a delay in the timing of detachment of paraspeckles from the chromatin, thus resulting in the marked elongation of paraspeckles.

NONO and SFQ were previously shown to be NEAT1-associated proteins (Gusak et al., 2009). Here we show by EM that both proteins are sequestered within elongated paraspeckles in which the up-regulated NEAT1 transcripts accumulate upon proteasome inhibition. The intensity of this sequestration process is such that the nucleoplasmic pool of these P-S-F is reduced by a factor of two. Given the variety of P-S-F other than NONO and SFQ possibly trapped within the paraspeckles in this setting, it is likely that the stress-induced sequestration mechanism has a profound influence on gene expression.
FIGURE 8: Proteasome inhibition induces SFQ protein sequestration in the paraspeckles, which causes removal of SFQ from the ADAR2 gene promoter. (A) Model of paraspeckle function: NEAT1 lncRNA expression dictates paraspeckle size and shape. NEAT1 sequences the PPs such as SFQ (small black shapes) in paraspeckles. The paraspeckle-unbound SFQ is free to function as a transcriptional regulator of the several paraspeckle target genes such as ADAR2 through interaction with other proteins. (B) MS132 treatment results in increased association of NEAT1 RNA with SFQ. (C) Chromatin immunoprecipitations (ChIP) with control (IgG) and anti-SFQ (aSFQ) were carried out from control (DMSO) and MS132-treated HeLa cells (MS132: top). The levels of immunoprecipitated NEAT1 lncRNAs were monitored by qPCR with isoform-specific primer sets (NEAT111, NEAT112, NEAT121, and NEAT131) and normalized by immunoprecipitated SFQ levels as shown at the bottom, in which immunoprecipitated SFQ was quantified by Western blotting. a-Tubulin is the control. (D) qPCR analysis of SFQ promoter in MS132-treated cells: Chromatin immunoprecipitations (ChIP) with aSFQ antibody was carried out in cells treated with DMSO or MS132 for 12 h. The levels of immunoprecipitated RNA fragments were quantified by qPCR with the primer sets shown below, which span different segments of the ADAR2 gene locus. **p < 0.01, *p < 0.05.

Identification of paraspeckle-target genes and regulation by SFQ.

Using microarrays, we show that paraspeckle disintegration by NEAT1 ASO in contrast has only minor effects on the HeLa cell transcriptome. This suggests that the many multifunctional genes and RNA-binding proteins that are concentrated within paraspeckles as mentioned earlier. However, it is important to note that loss of "normal" (i.e., non-stress-induced) paraspeckles may not have the same levels of gene regulatory effects as the gain of enhanced paraspeckles by proteasome inhibition. We can reach this conclusion because a fivefold enrichment of paraspeckles with proteasome inhibition resulted in a 50% reduction in nuclear/psammic SFQ and NONO. Hence, one could speculate that freeing SFQ and NONO by NEAT1 ASO from "normal" paraspeckles would result in
FIGURE 9: Other parapalce-ke-target genes are regulated with the same molecular mechanism as for the regulation of ADA162 gene expression. (A) Proteasome inhibition down-regulates the parapalce-ke-target genes. The mRNA levels were quantified as in Figure 7A. (B) The influence of siPcG RNAi on mRNA accumulation of the parapalce-ke-target genes was monitored by qRT-PCR. RNAi was carried out as in Figure 5B. **p < 0.01.

FIGURE 10: NEAT1 IncRNA acts to attenuate MCI132 induced apoptosis. (A) Induction of Neat1 expression by MCI132 treatment in mouse embryonic fibroblasts. The Neat1 levels were quantified with two primer sets (Neat1 and Neat1.2) in MEFs treated with DMSO (−) or MCI132 (+) for 0, 3, and 6 h. (B) Neat1 knockdown MEFs display greater sensitivity to MCI132 in real-time growth assays. Neat1−/− and Neat1−/− MEFs were analyzed in triplicate in a real-time growth assay over 40 h (Roche xCELLigence system), following the addition at time zero of 1 μM MCI132 (blue for Neat1−/−, green for Neat1−/−) or DMSO (red for Neat1−/−, black for Neat1−/−) to the culture medium. Each 15-min time point is plotted. Error bars are SD. (C) Neat1 knockdown MEFs display a greater sensitivity to MCI132 in the first 2–3 h of incubation. Experiments were performed as in B, with 0.25, 0.5, and 1 μM mg132 added to both Neat1−/− and Neat1−/− MEFs. The 15-min time points in the first 4 h are plotted. (D) Incubation of Neat1−/− and Neat1−/− MEFs with bortezomib shows a similar sensitivity of the Neat1−/− MEFs to proteasome inhibition relative to Neat1−/− MEFs. Experiments were conducted as in B, with 7.5, 10, and 20 nM bortezomib added to cells and incubated for 24 h. Error bars are SD.
a mere 1% increase in the nucleoplasmic pool of these factors, which might be insufficient to alter gene expression measurably and globally. This result, however, is consistent with the fact that the NEAT1 knockdown mouse has no obvious phenotype in development and behavior (Nakagawa et al., 2011). From the latter observation, it was proposed that paracaspase function might be linked to stressful situations, including viral infections, cancer, and in vivo cell differentiation, and in vitro growth of primary cells (Nakagawa and Hirose, 2012). Indeed, a recent study shows that NEAT1 is significantly up-regulated upon HIV infection and may be part of a viral defense mechanism (Zheng et al., 2013). Our observations support the view that NEAT1 is induced by stress, showing that NEAT1 transcription and paracaspase formation are enhanced by proteasome inhibition.

Despite its limitations, the NEAT1 knockdown experiment led to the identification of several paracaspase-target genes. Among those, the ADAR2 gene was particularly intriguing as one of the three members of the ADAR family that, mostly through ADAR1 and to a lesser extent through ADAR1, is believed to function to inhibit the activity of the other ADARs in vitro (Chen et al., 2000). It has been argued that paracaspases are the retention site for ADAR RNA substrates, namely hyper-A-U-rich mRNAs (Prasanth et al., 2005; Chen and Camilo, 2009), raising the possibility that ADAR2 regulates nuclear ADAR activities and in turn nuclear retention of hyper-edited mRNAs in paracaspases. However, as a word of caution, although the ADAR2 gene may play a role in HeLa cells, it is not measurably expressed in MEFs, even when NEAT1 is knocked out, indicating that its role may be restricted to human and/or transformed cells.

Because the ADAR2 gene is transcriptionally regulated by both NEAT1 and SFPQ in untreated HeLa cells, it is well suited for monitoring paracaspase expression and its involvement in gene transcription. In parallel to the increase in SFPQ associated with NEAT1 within the elongated paracaspases, we measured a concomitant increase in SFPQ mRNA by qPCR. On the basis of these results, we propose a model of SFPQ sequestration by NEAT1 within paracaspases that ultimately regulates ADAR2 gene expression (Figure 6A). ADAR2 is poorly expressed when more SFPQ is available (in cells with MG132 treatment) and well expressed when more SFPQ is available in the nucleoplasm when paracaspase integrity is compromised (such as after NEAT1 knockdown).

There is a precedent for SFPQ behaving in a similar manner to regulate the mouse Rob3 gene, in that its binding at the promoter can be reduced by overexpression of a “competitor” retroelementic InsRNA (Wang et al., 2009). SFPQ possesses both RNA- and DNA-binding domains, which may be responsible for binding to NEAT1 (and other InsRNAs) and the promoter region of certain genes. It is interesting that of the three DBHS proteins, highly similar in sequences, SFPQ is the only one with a characterized DNA-binding domain, found within a unique N-terminal that is distinct from NONO and PSF. In contrast, all three DBHS proteins contain two similar RNA recognition motifs, which together fold to give a conserved domain structure (Passen et al., 2012).

The other four paracaspase-target genes described in this study include a mitochondrial iron transporter (FPIT5373), a translational initiation factor (IF4F), a Srb substrate that plays a pivotal role in inosadap spine formation in cancer cells (SHPXD2A), and a zinc finger homodomain protein (CVC10-2). Under basal conditions in HeLa cells, paracaspases likely act to subvert suppress the expression of all four genes, and this is reinforced by even further suppression with paracaspase enlargement upon proteasome inhibition. Furthermore, three of these four genes (SHPXD2A, CVC10-2, and IF4F) depend on SFPQ for their expression, whereas FPIT5373 is independent but could be controlled by sequestration of another SFPQ. It would be intriguing to investigate whether paracaspase disorganization and/or enlargement influences the regulatory pathways in which these target genes are involved.

Proteasome inhibition and apoptosis

We show that NEAT1 transcription and paracaspase assembly are responsive elements when protein degradation is impaired. Moreover, when these two closely linked processes were prevented as in the NEAT1−/− MEFs, the resistance of the cells to proteasome inhibitor-induced cell death was clearly transiently reduced. Thus NEAT1 and paracaspases are needed for the cells to survive when facing accumulation of undegraded proteins, at least in an initial phase of accumulation compatible with cell survival.

Our demonstration that NEAT1 is transcriptionally up-regulated by MG132 adds to the list of genes that have been shown to be induced by proteasome inhibition, including the proteasome subunit genes, which is known as the proteasome recovery pathway. This pathway is spearheaded by transcriptional activation by a variety of factors. In mammalian cells, the specific transcription factor NF-κ was identified to mediate this response (Reddy et al., 2010). In contrast, transcriptional induction of the cyclophilinA homologous-2 gene induced by proteasome inhibition requires reactive oxygen species-dependent protein kinases and transcription factors CCAAT/enhancer-binding protein α and its coactivator cAMP-response element-binding protein (CREB-binding protein) (Chen et al., 2009). In the case of the NEAT1 gene, future identification of the molecular network responsible for transcriptional activation in response to proteasome inhibition will be guided by analyses of the promoter that contains potential binding sites for various transcription factors, including CREB and nuclear factor xB (NFkB).

Prolonged proteasome inhibition triggers apoptosis, believed to be induced by several events, including inhibition of the phosphoinositide 3-kinase/Akt and NFκB pathways and activation of the p38/JNK1/2 pathway (Zanotto-Filho et al., 2012). It was also reported that failure of amino acid homeostasis caused cell death after proteasome inhibition (Brown et al., 2011). Not only NEAT1 induction and the subsequent enlargement of paracaspases is a futile side effect induced by proteasome inhibition, but instead is an element in the cellular response to stress. The higher sensitivity toward proteasome inhibition of NEAT1-deficient cells suggests an autoregulatory function mediated by paracaspase formation. Of interest, in adult mouse tissues, paracaspase formation occurs in restricted cell types, in particular within the epithelium of the gastrointestinal tract, where paracaspases are observed in differentiated cells at the tip of crypts, where apoptosis is occurring (Nakagawa et al., 2011). Further studies are needed to expand our understanding of the roles that NEAT1 and the paracaspases play in modulating entry to apoptosis in response to stress and how these are related to changes in gene expression profiles mediated by SFPQ sequestration.

MATERIALS AND METHODS

Cell culture

HeLa cells were cultured in DMEM/10% fetal calf serum (FCS) at 37°C with 5% CO2. MEFs from NEAT1−/− and NEAT1−/− mouse embryos (Nakagawa et al., 2011) were cultured in DMEM/F-12/10% FCS (Life Technologies, Carlsbad, CA) at 37°C with 5% CO2. The
cells were treated with proteasome inhibitors in culture medium at concentrations of 5 nM and 0.25, 0.5, and 1 μM for MG132 in HeLa and MEF, respectively, and concentrations of 100 nM and 7.5, 10, and 20 nM for bortezomib in HeLa and MEF, respectively. DMSO equivalent to the highest drug concentration was added as negative control treatment.

Cell fractionation
We followed the nuclear isolation protocol developed by the Lamond lab (www.lamondlab.com/Nuclearlysisprotocol.html). HeLa cells were prepared from 1 × 107 cells, then washed three times with phosphate-buffered saline (PBS), resuspended in 5 mL of buffer A (10 mM 4-2-hydroxyethyl-1-piperazineethanesulfonic acid, pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.2 mM dithiothreitol), and homogenized for 10 strokes using a Dounce homogenizer with a tight pestle. An aliquot of the homogenate was removed as a source for the total RNA. The rest of the pellet was centrifuged at 215 × g for 5 min at 4°C, yielding cytosolic (supernatant) fractions. The pellet was resuspended in 3 mL of solution B (0.25 M sucrose, 10 mM MgCl2), overlaid onto 3 mL of solution C (0.35 M sucrose, 0.5 mM MgCl2), and centrifuged at 14,000 × g for 10 min at 4°C. The supernatant was removed as the nuclear fraction. The rest of the supernatant was sonicated by repeating 5 × pulses 20 times, using a band sonicator equipped with a microprobe (US-200, TOMY Seiko, Tokyo, Japan) at 60% of the maximum output. Sonication was overlaid onto equal volumes of solution D (0.5 M sucrose, 0.5 mM MgCl2). After centrifugation, the border of the two layers was marked. The solution was then centrifuged at 5000 × g for 10 min at 4°C, resulting in two layers of supernatant (NP1 and NP2) and one pellet layer (Ns). The pellet was rinsed with solution D and recovered by centrifugation. The resulting pellet, together with previously obtained fractions, was subjected to RNA extraction.

RNA interference
HeLa cells were transfected with siRNAs at 33 nM (final concentration) by using Lipofectamine RNAiMAX according to the manufacturer’s instructions. HeLa nuclei were isolated and incubated for 48 h. Knockdown efficiencies were verified by qRT-PCR or Western blotting. Stealth siRNAs for PFS and the negative control were purchased from Invitrogen. The siRNA sequences used in this study are shown in Supplemental Table S3.

Capture of nascent RNAs
To capture nascent RNAs, 0.5 nM EU was added into the culture medium and was incorporated into the cells for 30 min. Total RNA was prepared with TRIzol reagent (Invitrogen). The EU-labeled RNAs were biotinylated and captured using the Click-it Nascent RNA Capture Kit (Life Technologies), in accordance with the manufacturer’s instructions. A 1-μg amount of EU-labeled RNA was biotinylated with 0.5 μM biotin azide in Click-iT reaction buffer. The biotinylated RNAs were precipitated with ethanol and resuspended in distilled water. The biotinylated RNAs mixed with Dyna-beads MyCone Streptavidin T1 magnetic beads in Click-iT RNA binding buffer and heated at 80°C for 5 min, followed by incubation at room temperature for 30 min while gently vortexing. The beads were immobilized using the DynaMag-2 magnet and washed with Click-iT wash buffer 2. The washed beads were resuspended in Click-iT wash buffer 2 and used for cDNA synthesis.

qRT-PCR and reporter gene assays
qRT-PCR was performed as described previously (Sasaki et al., 2009). Total RNA was prepared from cell culture using TRIzol reagent (Life Technologies). The total RNA (500 ng) or the nascent RNA was reverse transcribed using Quantitect reverse transcription kit (Qiagen, Venlo, Netherlands). The primers were designed by Primer3 software (www-genome.wi.mit.edu/ presses/software.html) and purchased from Invitrogen. Aliquots of cDNA were subjected to real-time PCR, performed using a LightCycler 480 SYBR Green I Master (Roche, Basel, Switzerland) according to the manufacturer’s protocol. Primers used are shown in Supplemental Table S5.

For luciferase reporter assays, the human NEAT1 promoter (8-kb coordinates of +4445 to +5589) was amplified from human genomic DNA and inserted into the pGL3 plasmid (Promega, Fitchburg, WI) with Kpnl and Xhol. pGL3-NEAT1 promoter and pGL3-SV40 promoter (Promega) were transfected into HeLa cells with Lipofectamine 2000 (Life Technologies) and 5 μM MG132 or equivalent DMSO added 24 h later. After a 17-h incubation, RNA was harvested and reverse transcribed, and levels of luciferase RNA and GAPDH were measured using qPCR.

Northern and Western blotting
For Northern blotting, total RNA was separated by electrophoresis in 1% agarose gel containing 2% formaldehyde, followed by blotting to a positively charged nylon membrane (Roche). The blotted RNAs were fixed to the membrane by ultraviolet irradiation. Antisense RNA probes were synthesized with the DIG Easy Hyb RNA labeling kit (Roche) and hybridized with the DIG Easy Hyb reagent overnight at 65°C. The membrane was washed, and the hybridized bands were detected with the DIG Wash and Block buffer set (Roche). For Western blotting, the total cell lysate was run on an 8% SDS–PAGE gel and then blotted on a polyvinylidene fluoride membrane. The antibodies used are shown in Supplemental Table S2.

RNase protection assay
Total RNA was prepared with TRIzol reagent (Life Technologies). The RNase protection assay was performed with the Rneo3 kit (Ambion, Austin, TX), according to the manufacturer’s protocol. A 3-μg amount of total RNA was hybridized with a 32P-labeled antisense RNA probe that was synthesized in vitro. After 48 h, the reaction was treated with Rneo3 ex solution, and RNase digestion reaction was incubated at 37°C for 1 h. RNA digestion products were separated on a 10% TBE gel and visualized with autoradiography.

ASO administration into cells
The antisense chimeric oligonucleotides used for knockdown experiments were phosphothioate modified at their backbones to increase their stability. Five terminal nucleotides from the 5’ and 3’ ends were substituted by 2′-O-methylribonucleotides. NEAT1 ASO #12 and #17 are targeted to +1422 and +410 in the region common to NEAT1_1 and NEAT1_2 isoforms. The sequences of the ASOs used in this study are shown in Supplemental Table S4. The transgenic HeLa cells (1 × 106) were infected with 300 μL of Solution R of the Coll Cell Nucleofector Kit R (Lonza, Basel, Switzerland) and then mixed with oligonucleotides (4 μM final concentration). Transfection was carried out in an electroporation cuvette using the Nucleofector instrument (Lonza). The transfected cells were transferred to fresh DMEM plus 10% fetal bovine serum and incubated at 37°C and 5% CO2 for 24 h, followed by harvesting cells for RNA preparation.

DNA microarray
HeLa cells were nucleofected with GFP ASO or NEAT1 ASO (#12) and incubated for 6, 12, and 24 h. Total RNA was then prepared and
labeled with Cy3. Samples were hybridized to a Human Oligo Microarray (G4112F, Agilent, Santa Clara, CA) according to the manufacturer's protocol. Arrays were scanned with a G2565BA Microarray Scanner System (Agilent), and the resulting data were analyzed using the GeneSpring GX software (Agilent). The raw data are available in Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/info/landing.html, accession number GSE49159).

Chromatin immunoprecipitation assay

Hela cells were fixed with 1% formaldehyde for 10 min, after which the cross-links were stopped by treatment with 125 mM glycine for 5 min. The fixed cells were lysed in cell lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.5% NP40) for 10 min on ice. The lysed cells were centrifuged to recover the nuclear pellet, which was suspended in nuclear lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS). The nuclear lysate was sonicated with BioCycler UCU-310 (Ultrasonics, Geneva, Belgium) to an average fragment size of 500 base pairs. After the cell debris was removed by centrifugation at 20,000 g for 10 min at 4°C, the supernatants were diluted with a 10-fold volume of dilution buffer (167 mM Tris-HCl, pH 8.0, 167 mM NaCl, 1.2 mM EDTA, pH 6.0, 1.1% Triton X-100). Chromatin was immunoprecipitated overnight with anti-RNAAPI or anti-SPPG antibodies conjugated to Dynabeads-mouse immunoglobulin G or Dynabeads-protein G (Life Technologies), respectively. The recovered beads were washed once with wash buffer 1 (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, pH 8.0, 1% Triton X-100, and 0.1% SDS), once with wash buffer 2 (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM EDTA, pH 8.0, 1% Triton X-100, and 0.1% SDS), and once with wash buffer 3 (20 mM Tris-HCl, pH 8.0, 500 mM UC0, 2 mM EDTA, pH 8.0, 1% Triton X-100, and 0.1% SDS). The captured chromatin was eluted and diluted in elution buffer (25 mM Tris-HCl, pH 8.0, 5 mM EDTA, 0.5% SDS, and 0.1 mg/ml protease K) at 65°C for 6 hr and then treated with 50 µg/ml RNase A at 37°C for 30 min. The precipitated DNAs were used for qPCR. The primers used are shown in Supplemental Table S5.

Immunoprecipitation

Hela cells (2 x 10^6 cells) were lysed with lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, and 1% Triton X-100) for 30 min on ice, and the cell extract (1 mg of protein) was used for immunoprecipitation with an antibody against SPPG conjugated to Dynabeads-protein G beads (Life Technologies) for 1 hr at room temperature. The IP products were washed four times with lysis buffer. Detailed information about the antibodies used is shown in Supplemental Table S5.

Immunofluorescence

Hela cells were fixed with 4% paraformaldehyde/PBS. Fixed cells were permeabilized with 0.2% Triton X-100/PBS for 5 min, rinsed, and blocked with 10% normal horse serum (Vector Laboratories, Burlingame, CA) in PBS for 1 hr. Primary antibodies were applied for 1 hr at room temperature or overnight at 4°C. The samples were washed three times with PBS/PBS (0.1% Tween-20) for 5 min each. Secondary antibodies were applied for 1 hr at room temperature. After washing, the slides were covered with Vectashield (Vector Laboratories) containing 4',6-diamidino-2-phenylindole (DAPI). Fluorescence images were visualized by microscopy at room temperature on a microscope (Fluoview FV1000; IX81; Olympus, Tokyo, Japan) equipped with U-Plan Apochromat 40×/0.95 objective lenses (Olympus). Fluoview FV10-ASW1.7 software (Olympus) was used for image acquisition and processing. All confocal images were transferred as high-resolution JPEG files. Figures were compiled using Photoshop (Adobe Systems, San Jose, CA), and the antibodies used are shown in Supplemental Table S2.

RNA fluorescence in situ hybridization

Cells were seeded onto a multichamber culture slide (Corning, Corning, NY) and fixed with 4% paraformaldehyde/PBS. The fixed cells were permeabilized with 0.5% Triton X-100/PBS for 5 min. RNA probes were prepared using a DIG/ITC RNA Labeling Kit (Roche Diagnostics) according to the manufacturer's protocol. Dehydrated slides were incubated for 16 hr at 56°C with a hybridization solution (2 M sodium-citrate SSC, 50% formamide, 1× Denhardt's salt, Sigma-Aldrich, St. Louis, MO), 10 mM EDTA, 100 µg/ml yeast RNA, 0.01% Tween-20, and 5 µg/ml streptomycin) containing digoxigenin- or fluorescein-sulfocyaninate–labeled RNA probes. The slides were washed twice with prewarmed wash buffer (2× SSC, 50% formamide, and 0.01% Tween-20) at 55°C for 30 min. Excess RNA probes were digested with 10 µg/ml RNAase A/NT2 (Roche Diagnostics) at pH 8.0, 10 mM EDTA, 500 mM NaCl, and 0.1% Tween-20 at 37°C for 1 hr. The slides were washed with buffer (2× SSC, 0.01% Tween-20) at 55°C for 30 min and twice with a second buffer (0.1× SSC, 0.01% Tween-20) at 55°C for 30 min. After washing, the slides were covered-slipped with Vectashield (Vector Laboratories) containing DAPI.

Electron microscopic studies

Conventional ultrastructural microscopy after Epon embedding was as in Souquere et al. (2009). Thin sections were analyzed with a FEI Tecnai Spirit, and digital images were taken with a 218 Megaswift charge-coupled device camera. Parapodopedia frequency was established by counting parapodopic sections within 10 squares of 200 mesh EM grids for samples of control (5.6 ± 2.8 parapodopic/square) and MS132-treated Hela cells (19 ± 6.1 parapodopedia/square) with similar cell densities (cell/pollens). When coupled to the 50% increase in the mean length of the parapodopes (as in Figure 1C), this indicates a 5.1-fold increase of parapodopic surface area after 17.4% 5 mM MS132 treatment. Immuno–electron microscopic studies were as in Souquere et al. (2010). Data were obtained on paraformaldehyde-fixed Hela cells, except for anti-FLI1 immunodetection, which was carried out on glutaraldehyde-fixed Hela cells. The primary antibodies used for immunodetection are listed in Supplemental Table S1. Anti-rabbit or anti-rat secondary antibodies, coupled to 10 nm gold particles, were from EBER (Cardiff, UK). For quantification of labeling densities, control and MS132-treated cells were processed in parallel from chemical fixation to Lumiodyl embedding and then to final incubation with gold particles. The gold particle count was determined by eye. Surface areas were determined with analySIS. Calculations and standard deviations were obtained with Excel (Microsoft, Redmond, WA). Electronic microscopy in situ hybridization were as in Souquere et al. (2010), except that duration of 15 µm of DUTP/biotinylation of DNA probes by nick translation was reduced to 30 min.

Monitoring cell death

Roche's xCELLigence System for real-time cell analysis (which measures impedance-based signals) was used to quantify cell proliferation. Different preparations of Neuro2A and Neuro117 MEFs, each from different individual knockout mice and wild-type littermates (Nakagawa et al., 2011), were cultured in DMEM/F-12 (Life Technologies), supplemented with 10% FCS (Life Technologies). On the day before the experiment, cells were detached with trypsin Xpress (Life Technologies), and 2.5 x 10^4 viable cells were plated in individual wells of an E-16 xCELLigence plate (xCELLigence, Roche, Germany) in a 100 µl volume of culture medium. Cells were left to settle for 24 hr in the incubator, followed by the addition of 100 µl of
prewared MG132, bortezomib, or DMSO (diluted in culture medium) to each well as appropriate, with each drug treatment on each cell type being carried out in triplicate. The plates were immediately inserted into the xCELLigence apparatus as an incubator, and electrical impedance was measured across the bottom of each well in real time every 15 min for 40 h according to the manufacturer's recommendations.

ACKNOWLEDGMENTS

We thank the members of the Hirose, Fox, and Piroon laboratories. G.V. was supported by an Erasmus exchange program with the University of Napoli (Naples, Italy). This research was supported by the Funding Program for Next-Generation World-Leading Researchers of the Japan Society for the Promotion of Science (to T.H.) and by grants from the New Energy and Industrial Technology Development Organization (to T.H.), the Takeda Science Foundation (to T.H.), the National Health and Medical Research Council of Australia (to A.H.P.), the Medical Research Foundation of Royal Perth Hospital (to A.H.P.), and the Centre National de la Recherche Scientifique and the Association pour la Recherche sur le Cancer (to G.P.).

REFERENCES


A book chapter co-authored with Ellen Fortini and Archa Fox reviewing the function of IncRNAs that are related to subnuclear structures. I wrote the first and third drafts of sections on NEAT1/paraspeckles, hsr-ω/omega speckles, and Satellite III RNA/nuclear stress bodies.
Long Non-coding RNAs and Nuclear Body Formation and Function

Ellen Fortini, Ruohan Li and Archa H. Fox

1 General Introduction

In the past decade we have made a quantum leap in our understanding of the genetics of complex organisms, with the discovery that the nonprotein coding regions of our genomes are transcribed into tens of thousands of long noncoding RNA (lncRNA) molecules. However, while we know of their identity, deciphering the functions of these lncRNAs has been, and is continuing to be a challenge. In this chapter we focus on one of the well-characterized functions of specific lncRNAs: to form subnuclear structures and/or influence the function of subnuclear bodies. These findings have been important to the field of lncRNA biology, as the ability to place specific lncRNAs within the context of known nuclear architecture has given many clues as to the roles of these lncRNAs, and has also affirmed their functional relevance. So, what do lncRNAs do in subnuclear bodies? The mechanisms range from dynamic induction of nuclear bodies to sequester or modify nuclear proteins involved in splicing and transcription, to lncRNA enrichment in subnuclear bodies directing the recruitment of gene loci to influence their transcriptional environment. The formation and enrichment of lncRNAs in subnuclear bodies has thus become one more example of the myriad different ways that lncRNAs regulate gene expression.

Here we discuss lncRNAs with defined nuclear localizations, and separate them into two classes (Fig. 1). First, there are the lncRNAs whose role is to form subnuclear bodies as essential structural scaffolds, these include mammalian...
NEAT1 in paraspeckles, primate Satellite III (SatIII) transcripts in nuclear stress bodies (nSBs), drosophila lur-o RNA in omega speckles, and mammalian neuronal MIAT in granular speckles. The second class of nuclear IncRNAs have been observed to localize to particular subnuclear sites, but are not essential for the nucleation or formation of the subnuclear structures they associate with. For these IncRNAs, their enrichment within subnuclear bodies may reflect an aspect of their function that is associated with nuclear organization. Examples here include MALAT1 in nuclear speckles, as well as TUG1 and (potentially) HOTAIR in polycomb bodies. In this chapter we focus on each of these well-studied examples, and describe the history, structure, and functions of the subnuclear bodies and their associated IncRNAs, to build up a picture of the insights being gained in this important nexus between IncRNA biology and nuclear organization.

2 LncRNAs that Form Structural Scaffolds for Subnuclear Bodies

In recent years, it has emerged that several types of subnuclear bodies are built on a IncRNA scaffold or backbone. A common theme seems to be that these IncRNAs nucleate the assembly of these bodies, in most cases by 'seeding' the bodies: recruiting abundant nuclear RNA-binding proteins to the site of IncRNA

Fig. 1 Nuclear bodies formed by, or associated with IncRNAs: a Under steady-state conditions, RNA FISH can be used to demonstrate that NEAT1 IncRNA is co-localised with paraspeckle markers, TUG1 resides within polycomb bodies, and MALAT1 is found in nuclear speckles. b Under stress, such as heat shock, specific IncRNAs are transcribed that nucleate additional subnuclear bodies. These include Satellite III IncRNA derived from pericentric heterochromatin in primates to form nuclear stress bodies, and lur-o RNA in Drosophila to form omega speckles.

Long Non-coding RNAs and Nuclear Body Formation and Function

2.1 NEAT1 and Paraspeckles

Paraspeckles are mammalian subnuclear bodies that form around the NEAT1 (Nuclear Paraspeckle Assembly Transcript) 1 IncRNA. Paraspeckles were first described as Interechromatin Granule Associated Zones, electron dense structures distinct from other nuclear bodies observed with the electron microscope in cultured cells (Visa et al. 1993). However, it was in 2002 that a clear marker protein, PSCP, or Paraspeckle protein 1, was found, and the term 'Paraspeckles' was coined to describe the subnuclear foci in which PSCP was enriched (Fox et al. 2002). Additional paraspeckle proteins have since been identified, and these include the DBHS (Drosophila Behaviour Human Splicing) proteins related to PSCP, SPPQ and NONO, as well as a host of other RNA-binding proteins (Bond and Fox 2009; Fox et al. 2005; Naganuma et al. 2012; Prasonth et al. 2005). It is important to note that paraspeckle proteins, while enriched in paraspeckles, are also generally diffusely distributed in the nucleoplasm (Fox et al. 2002).

In the years following their identification, several early clues also suggested that RNA would likely be crucial to both paraspeckle structure and function: first, the paraspeckle proteins were all known RNA-binding proteins, and several only localized to paraspeckles via key RNA recognition Motifs (RRM), second, paraspeckles were sensitive to RNome treatment, third, they only formed in newly divided cells once RNA Polymerase II transcription was well established, and finally, they were disassembled by inhibition of RNA Polymerase II transcription (Dye and Patton 2001; Fox et al. 2002, 2005).

In 2009, three groups reported that paraspeckles were formed around the NEAT1 IncRNA, and that NEAT1 was an essential structural component of paraspeckles (Clemson et al. 2009; Sasaki et al. 2009; Sunwoo et al. 2009). NEAT1 is a mammalian-specific gene located on human chromosome 11q13 and mouse 19qA (Hutchinson et al. 2007; Sasaki et al. 2009). The NEAT1 gene promoter triggers transcription of two major isoforms of RNA that overlap completely at their 5'-end, yet have very different 3'-ends (Hutchinson et al. 2007; Sasaki et al. 2009; Sunwoo et al. 2009). The shorter canoniclly polyadenylated isoform (3,700 nt in human, 3,100 nt in mouse), is termed NEAT1_v1 or MEN4. The longer isoform, 23,000 nt in human and 20,500 nt in mouse, is termed NEAT1_v2 or
245

E. Fortini et al.

MEG, and contains an unusual RNA-like structure at its 3′-end that is recognized and cleaved by RNase P, to produce a 3′-end with a short genomeically encoded poly(A)-rich sequence (Sunwoo et al. 2009). NEAT1_v2 is estimated to be at least fivefold less abundant than NEAT1_v1, and in many tissues and cell types, present at an even lower proportion (Sasaki et al. 2009; Sunwoo et al. 2009).

Transcription of NEAT1 IncRNA is the seed that triggers paraspeckle formation. This has been elegantly demonstrated with two main pieces of evidence: first, paraspeckles form in close proximity to the NEAT1 gene (Clemson et al. 2009), clustering near there, (although, once formed, they are capable of moving further afield), and, second, inducible NEAT1 expression is sufficient to nucleate the formation of paraspeckles (Mao et al. 2011). In another fascinating twist, Spector and colleagues showed that it is not enough to simply have NEAT1 in the nucleus for paraspeckles to form; instead, NEAT1 has to be actively transcribed (Mao et al. 2011). Interestingly, in a variety of cultured cell lines, both NEAT1 isoforms clearly display the characteristic punctate localization typical of paraspeckles, co-localizing and co-purifying with DHBS proteins (Clemson et al. 2009; Hutchinson et al. 2007; Mao et al. 2011; Sasaki et al. 2009; Sreekumar Murthy and Ranganathan 2010; Sunwoo et al. 2009). However, while both isoforms are found in paraspeckles, it is now generally accepted that transcription of the lower abundance NEAT1_v2, instead of the more abundant NEAT1_v1, is the critical factor for the assembly and maintenance of paraspeckles. These pieces of evidence all support this: siRNA specific for NEAT1_v2 is sufficient to ablate paraspeckles (Sasaki et al. 2009; Sunwoo et al. 2009), cells expressing endogenous NEAT1_v1, but not NEAT1_v2, lack paraspeckles; and overexpressed NEAT1_v3, but not NEAT1_v1 restores paraspeckles in NEAT1_−/− Murine Embryonic Fibroblasts (MEFs) (Naganuma et al. 2012; Sasaki et al. 2009).

In line with the importance of NEAT1_v2 in paraspeckle formation, we also know from electron microscopy analysis, that NEAT_v2 RNA extends throughout the core of a paraspeckle, whereas NEAT1_v1 is only found at the periphery (Souquere et al. 2010). In fact, our understanding of the spatial organization of NEAT1 within paraspeckles is unrivalled by any other IncRNA in nuclear organization (Fig. 2). There are also some additional observations that suggest NEAT1_v1 may play a greater role in paraspeckle formation when artificially tethered to the chromatin at high levels: when NEAT1_v1 is post-transcriptionally targeted to a specific genomic location this can also recruit paraspeckle proteins efficiently, presumably forming de novo paraspeckles (Shevlyov and Dandri 2011). While it is not known if these de novo paraspeckles are functional, these data raise the possibility that the function of NEAT1_v2 is to provide a binding platform for NEAT1_v1, for it to reach a local high concentration in order to allow paraspeckle proteins to associate with the RNA and form stable RNA-protein complexes (Nakagawa and Hirao 2012; Shevlyov and Dandri 2011).

While NEAT1 is essential for paraspeckle formation, so too are a number of paraspeckle proteins. For example, siRNA against the DHBS proteins SFQ and NONO results in paraspeckle disassembly and a reduced stability of NEAT1_v2 (Sasaki et al. 2009). However, it is important to note that paraspeckle proteins, while an essential factor for making paraspeckles, do not have the capacity to nucleate paraspeckle formation: immobilizing DHBS proteins to chromatin could not effectively recruit NEAT1 to form de novo paraspeckles (Mao et al. 2011), suggesting a sequential assembly of different components that starts with NEAT1 transcription. At present there are approximately 40 proteins identified that are enriched in paraspeckles, mostly having RNA or DNA binding domains. Many of those proteins are indispensable for the formation of paraspeckles, or maintaining the stability for NEAT1 (Naganuma et al. 2012; Sasaki et al. 2009). One area that is still largely unknown is the molecular details of paraspeckle protein interactions with NEAT1. Structural studies on the essential DHBS paraspeckle proteins have revealed a novel diner consisting of four RRM motifs held in a brace position by a coiled coil domain (Passian et al. 2012), however the RNA-binding modalities of these dimers are not yet known. In addition, the DHBS protein SFQ has been shown to interact with several other IncRNAs besides NEAT1 (Li et al. 2009; Takayama et al. 2013; Wu et al. 2013). Hirose and colleagues recently revealed that the function for some paraspeckle proteins is in establishing NEAT1_v2 production (Naganuma et al. 2012). They showed that the paraspeckle protein HNRNPU co-localizes with paraspeckle-associated RNA cleavage proteins to determine if NEAT1 transcripts are cleaved and polyadenylated after 3.7 kb, or if this process is prevented to allow transcription to continue and generate NEAT1_v2 (Naganuma et al. 2012). This mechanism suggests a constant competition for the production, stabilization, and degradation of NEAT1_v2, which is in turn closely linked to paraspeckle formation.

While we know a considerable amount about the formation, components, and structure of paraspeckles, we have a poorer understanding of paraspeckle function. Mice lacking NEAT1, devoid of paraspeckles, have no gross phenotype, indicating that their function is unlikely to be crucial for development (Nakagawa et al. 2011). Nakagawa and colleagues (2011) have thus far produced the most comprehensive mapping of NEAT1 expression in tissues, using in situ hybridization against NEAT1 on mouse tissues, and have found that while most cells express NEAT1_v1, NEAT1_v2 is only found in a distinct subpopulation of cells. In silico,
RNA-seq datasets show widespread and abundant NEAT1 expression in most of the cell lines and tissues examined (Gibb et al. 2011), as well as indicating dynamic regulation of NEAT1 in various models of cellular differentiation (Sunwoo et al. 2009). However, there are exceptions to the rule, and NEAT1 is expressed at extremely low levels in embryonic stem cells (Chen and Carmichael 2009; Gibb et al. 2011; Nakagawa et al. 2011).

In terms of the molecular function of parapseckes, the best evidence suggests that sequestering both RNA and protein components may be the route to influencing gene expression. In 2003, a specific nuclear retained mRNA was identified that partially co-localized in parapseckes (Prasanth et al. 2005). This mRNA contains a long 3'-untranslated region (UTR), with Adenosine to Inosine (A-to-I) edited inverted Alu repeats that are a binding site for the parapseckes proteins NONO and SFPQ (Prasanth et al. 2005; Zhang and Carmichael 2001). Specific stresses resulted in the edited RNA translocating to the cytoplasm, with a concomitant increase in translation (Prasanth et al. 2005). It has also been demonstrated that knockdown of NEAT1 alters the nuclear retention of these inverted Alu repeat RNAs (Chen and Carmichael 2009). Aspects of this nuclear retention mechanism could also be applied to other genes with inverted repeats in their 3'-UTRs, including Linc28, Ncrl, and ApoeMcG (Chen and Carmichael 2009; Mac et al. 2011); however, it has also been found that some other genes with A-to-I edited Alu repeats in their 3'-UTRs may undergo export to the cytoplasm where they are translationally repressed (Caspiew et al. 2012; Fitzpatrick and Huang 2012). This repression appears to be mediated by cytoplasmic stress granules, which can form under heat shock stress (Caspiew et al. 2012; Fitzpatrick and Huang 2012). Recently, it has been postulated that an additional molecular function for parapseckes could be the sequestration of parapseckes proteins such as SFPQ (Nakagawa and Hirose 2012). This is interesting as the sequestration of nuclear proteins is often seen in highly conserved and unique to primates, as SatIII elements are found on essential structural IncRNA components in their assembly.

2.2 Satellite III LncRNA and Nuclear Stress Bodies

Nuclear stress bodies (nSBs) are formed around stress-induced IncRNAs transcribed from SatIII repetitive pericentromeric heterochromatin. nSBs were first identified when heat shock responsive transcription factor (HSF1) was observed to accumulate in large foci at pericentromeric heterochromatin after heat shock, chemical, and hypotonic stresses (Denegri et al. 2001; Jolly et al. 1997; Mahi et al. 1989; Sarge et al. 1993). These accumulation sites were formed primarily on the 9q12 loci of human chromosome 9, but also chromosome 12 and 15, which contain long tandem repeats of SatIII DNA (Denegri et al. 2002; Jolly et al. 2002). The nSBs were sensitive to RNAse treatment, and also required ongoing RNA transcription for their maintenance, suggesting that RNA might play a structural role in their assembly (Chiodi et al. 2000; Weighardt et al. 1999). In 2002, Jolly and colleagues (2002) reported that HSF1 bound to the SatIII DNA and facilitated transcription of SatIII IncRNA. Indeed, under heat shock, these heterochromatic DNA regions shifted to euchromatin, marked by active histone modification marks, reinforcing the finding that the SatIII loci were becoming transcriptionally active following stress (Rizzi et al. 2006). Once transcribed, the SatIII nRNA transcripts remain locally associated with the chromatin, and are bound by a number of pre-mRNA processing factors to form the nSBs, including SF2/ASF, SRp30c, and 9G8, and small nuclear ribonucleoproteins (snRNPs) (Denegri et al. 2001; Jolly et al. 2004; Metz et al. 2004). Interestingly, HSF1, the transcription factor responsible for up-regulating the nRNAs, can also be found in nSBs (Shevtsov and Dunbd 2011).

SatIII IncRNAs can have a variable length from either 2,000–5,000 nt, to no more than 10,000 nt (Biamonti and Caceres 2009; Jolly et al. 2004; Rizzi et al. 2004). This variable length of RNA likely results from the repetitive SatIII sequence, the multiple transcription start sites inside the array of tandem repeats, or the close contact with these bound splicing factors which have found to cause splicing of the IncRNA (Metz et al. 2004; Valgardsson et al. 2005). The SatIII RNA is absolutely required for nSB formation: knockdown of SatIII IncRNA abolishes the recruitment of the protein splicing factors to the nSBs. However, Sat III knockdown does not prevent the initial accumulation of HSFI (Metz et al. 2004; Valgardsson et al. 2005). Recent studies have demonstrated that the immunoblotting of SatIII IncRNA, transcript artificially onto chromatin can recruit HSFI, SF2/ASF, and SF2/ASF to form de novo nSBs (Shevtsov and Dunbd 2011). Interestingly, heat shock resulting in the massive upregulation of SatIII IncRNA is accompanied by a global deacetylation of chromatin in the rest of the nucleus (Frisch et al. 2009).

The specific function of nSBs remains a matter for speculation. Whatever the function, it is possible it is highly complex and unique to primates, as SatIII elements appear late in evolution, being primate specific (Denegri et al. 2002; Jarmuz et al. 2007). One possible function for nSBs is that they sequester RNA-binding proteins and RNAs to prevent them from circulating freely or performing their normal functions under heat shock conditions. This might be in line with the global suppression of transcription, altered splicing functions, and suppression of translation after heat shock, (with the exception of ongoing expression and translation of the heat shock responsive genes) (Lindquist 1986). Heat shock proteins rarely have introns in their genes, and they undergo a dramatic increase in expression and translation following heat shock stress, without great reliance on splicing factors (Lindquist 1986). It is therefore interesting to ponder if mobilizing active transcriptional power to the production of SatIII IncRNA, and then trapping particular splicing factors and RNAs in the nSBs might aid cells to prevent unnecessary or even harmful transcriptional, splicing, or translational events following heat shock (Biamonti and Vourc'h 2010; Metz et al. 2004). As with many other nuclear bodies, there remain many unanswered questions about the functions of these structures.
In Drosophila there is a well-studied lncRNA-induced subnuclear structure termed ‘omega speckles’ that are nucleated by the Heat Shock RNA omega (har-ω or 93D). The har-ω gene locus is conserved among Drosophila species, but has not been found in other types of organisms. The har-ω gene contains two short exons (~475 and 700 bp in D. melanogaster) separated by a 700 bp intron, followed by a long stretch (5–15 kb) of short (280 bp in D. melanogaster) tandem repeats (Jolly and Lakhota 2006). The overall gene may span 10–20 kb long, and produces two major transcripts and one precursor transcript. The major cytoplasmic transcript, termed har-ωG, is less than 2,000 nt long, and contains the spliced exons with a polyadenylated 3’-end. The long nuclear transcript har-ωA spans the entire length of the gene, including the intron, and is also polyadenylated (Bendena et al. 1991; Garbe et al. 1986; Rusek et al. 1987). Therefore, har-ω could be considered a shorter spliced and overlapping version of har-ω. Har-ωA appears to have a 23–27 amino acids open reading frame, but its sequence is not conserved and product is undetectable (Bendena et al. 1991; Garbe et al. 1986; Lakhota and Sharma 1995; Rusek et al. 1987). The har-ω gene is active in all cell types and at various developmental stages of Drosophila, and can be one of the most active loci under heat shock or amide stresses (Bendena et al. 1991; Mutsuddi and Lakhota 1995; Prasanth et al. 2000; Tapadia and Lakhota 1997).

The long har-ωA transcript has been the most closely studied RNA of the har-ω group. Har-ωN has a rapid turnover in the nucleus under normal conditions, but under stress that might inhibit general nuclear transcription, it is rapidly upregulated and accumulates with increased stability (Bendena et al. 1989; Hogan et al. 1993; Lakhota and Sharma 1995). Har-ωN was found co-localized with a variety of hnRNPs, forming a variable number of ‘omega speckles’ (Lakhota et al. 1999; Prasanth et al. 2000). Without active transcription of har-ωA, omega speckles cannot form (Prasanth et al. 2000). Similar to paraseckles, omega speckles can be found both next to the locus of har-ω, or away from the locus (Lakhota et al. 1999; Mao et al. 2011; Prasanth et al. 2000). It is particularly important to note that in normal conditions, most of the omega speckle proteins are present in both omega speckles, and at other nucleoplasmic locations that are usually transcriptionally active (Lakhota et al. 1999; Prasanth et al. 2000). However, under stressful conditions, these minor sites rapidly disappear and the omega speckle proximal to the gene locus becomes enlarged. This stress-induced enlargement is accompanied by the translocation of omega speckle proteins, such as HRB87F (Drosophila orthologue of HNRNP1) and HRB57A (Drosophila orthologue of HNRNPK), from their chromatin binding sites to the enlarged omega speckles, followed by a reduction of transcriptional activity at their previous binding sites (Buchanau et al. 1997; Dengil and Boutz 1983; Dengil et al. 1983; Hovemann et al. 1991; Lakhota et al. 1999; Prasanth et al. 2000; Samuels et al. 1994; Zu et al. 1998). These data suggest a potential involvement of omega speckles in regulating the trafficking and availability of hnRNPs and other related RNA-binding proteins.

The expression hypothesis suggested for both paraseckles and nSBs, where the transcription of the nucleating lncRNAs results in the accumulation of proteins in those bodies, thus altering their original localization and function. This sequestration might be a protection, or a temporary storage mechanism for those proteins, so that they can quickly resume normal function after the stress has been relieved (Jolly and Lakhota 2006; Lakhota et al. 1999; Prasanth et al. 2000).

A major focus of omega speckle research in the past decade has been determining the physiological significance of har-ω. Flies that are har-ω null are mostly embryonic lethal, with some flies hatching that are very weak and lacking omega speckles, suggesting that har-ω has a critical role in the development of Drosophila and assembly of omega speckles (Prasanth et al. 2000). The overall expression level of har-ω also seems to be critical, as its overexpression in whole flies results in polyglutamine (poly-Q) induced neurodegeneration (Malik and Lakhota 2009; Sekiguda and Lakhota 2006), and its overexpression in the cyst cells of testis leads to male sterility (Rajendra et al. 2001). However, it is not clear yet if and how omega speckles are critically involved in causing the abnormal phenotypes resulted by the deletion or overexpression of har-ω.

The difference between nSBs, paraseckles, and omega speckles lies in the different lncRNA identities, induced under different conditions, to nucleate different sets of proteins. For example, Serine/Arginine (SR) proteins, which are frequently found in nSBs, are not found in hnRNP containing omega speckles (Jolly and Lakhota 2006). Intriguingly, SR proteins are generally considered as competitors of hnRNPs in pre-mRNA splicing, and yet both nSBs and omega speckles can be rapidly induced by heat shock stress (Jolly and Lakhota 2006). Another interesting connection is that the drosophila homologs of two paraseckle proteins, NONO and HNRNP4, were also shown to associate with har-ω, which might indicate conservation of functions shared by the two subnuclear bodies (Prasanth et al. 2000; Zimowsku and Paddy 2002). Finally, there is a similarity in gene structure, such that, as with the har-ωA transcripts, the paraseckle nucleating lncRNA NEAT1-v1 and v2 also share their 5’-end, with NEAT1-v2 and har-ω being longer and containing repetitive sequences.

2.4 MIAT and Gomafu Speckles

MIAT (Myocardial infarction associated transcript) lncRNA is also known as GOMAFU, or retinal noncoding RNA 2, however, here we will use the official HGNC (Human Gene Nomenclature Committee) symbol MIAT when referring to this lncRNA. MIAT was originally identified as an lncRNA differentially expressed during the development of mouse retina cells (Blackshaw et al. 2004; Ishii et al. 2006). MIAT is also widely expressed throughout the nervous system in development and adulthood (Sone et al. 2007). MIAT contains multiple spliced exons, with a final transcript size of approximately 10,000 nt, and is
Nuclear speckles (also known as splicing speckles) are distinct subnuclear domains that are defined by the co-localization of snRNPs and the pre-mRNA splicing factor SC-35 (Spector and Lamond 2011; Thiry 1995). There are 20–50 irregularly shaped nuclear speckles in a typical mammalian nucleus, located within the interchromatin space, and a large number of additional pre-mRNA splicing-related proteins are also enriched there (Miaz 1999).

A major function of nuclear speckles is acting as a reservoir for splicing proteins, rather than the site of actual splicing per se. This is supported by evidence that there is little active splicing occurring within the nuclear speckles (reviewed in Spector and Lamond 2011). Rather, it is thought that splicing happens in a co-transcriptional manner at transcription sites (Zhang et al. 1994). The key pre-mRNA splicing SR proteins are targeted in and out of nuclear speckles to transcription sites via their selective phosphorylation (Misteli 1998; Misteli et al. 1997). Another function of nuclear speckles relates to their frequent close proximity to highly expressed genes, suggesting that they are enhancing processing of the resulting transcripts.

In 2007 a nuclear speckle lncRNA, MALAT1 (Metastasis Associated Lung Adenocarcinoma Transcript 1, also known as NEAT2) was observed to co-localize with nuclear speckle marker proteins (Hutchinson et al. 2007). MALAT1 is an unspliced approximately 8,000nt lncRNA that exhibits broad tissue expression, and is associated with tumorigenesis and metastasis (Gutschner et al. 2013).

Interestingly, the MALAT1 gene is located in a syntenically conserved fashion in close proximity to the NEAT1 gene (11q13 in human and 19qA in mouse).

Although MALAT1 has clear co-localization with nuclear speckle markers, it is not essential for their formation. Nuclei of mice lacking MALAT1 still contain nuclear speckles (Eissmann et al. 2012; Nakagawa et al. 2012; Zhang et al. 2012), and siRNA against MALAT1 does not disrupt nuclear speckles in cultured cells (Clemson et al. 2009), although it can alter the recruitment of various nuclear speckle proteins to these domains by regulating the phosphorylation of SR proteins (Lin et al. 2011; Tripathi et al. 2010).

MALAT1 is targeted to nuclear speckles through interactions with various proteins: knockdown of RPS1, SRn160, or LRIP160, which are well-known mRNP processing factors, resulted in MALAT1 becoming diffusely distributed within the nucleoplasm (Miyagawa et al. 2012). There are contrasting reports indicating the importance of different regions of MALAT1 to nuclear speckle targeting: Tripathi et al. found that overexpression of any 2 kb fragment of MALAT1 resulted in its targeting to nuclear speckles (Tripathi et al. 2010), whereas Miyagawa et al. expressed smaller 1 kb fragments of MALAT1 and observed a more significant role for a region of MALAT1 towards its 3’-end that is predicted to form a binding site for key splicing proteins (Miyagawa et al. 2012).

Besides influencing splicing proteins, how else might MALAT1 exert its function on gene expression? An interesting study has shown that MALAT1 can
other lncRNA-enriched subnuclear structures (Yang et al. 2011). In this seminal study, the authors showed that in response to growth signals, MALAT1 participates in a gene activation program through binding unmethylated polycomb protein, to sequester polycomb-associated genes to the surface of nuclear speckles. In contrast, in a repressive environment, genes with an associated methylated polycomb protein are recruited to polycomb group (PcG) bodies through interaction with the TUG1 lncRNA (more of which below). This interplay between subnuclear localization sites and gene expression status gives an intriguing insight into the myriad ways that lncRNAs may be affecting gene expression through as yet undiscovered mechanisms.

Beyond these studies, other researchers have defined the mechanism that MALAT1 uses to enhance cellular proliferation, through its involvement in regulating the expression and/or pre-mRNA processing of oncogenic transcription factors, especially those that control mitotic progression (Tripathi et al. 2013). Given that an important role for MALAT1 in cell growth, proliferation, synaptogenesis and cancer is now well defined, it is fascinating that MALAT1 is not required for mouse development, as seen with the viability of the MALAT1 knockout mice with no gross phenotype (Bernard et al. 2010; Eissmann et al. 2012; Nakagawa et al. 2012; Tripathi et al. 2010; Zhang et al. 2012). It is interesting to speculate that there exist compensatory mechanisms in vivo to account for these effects. Indeed, recent work has indicated that either MALAT1, or SRSp1, can ‘seed’ nuclear speckles, suggesting they compensate for each other, and this may explain the intact nuclear speckles and unimpaired nuclear speckle function in MALAT1 knockout mice (Nakagawa et al. 2012).

3.2 TUG1 and Polycomb Bodies

Polycomb (PcG) bodies are defined as subnuclear foci enriched in the chromatin-associated polycomb group proteins (Pirrotta and Li 2012). PcG bodies vary in size, shape, and number from cell type to cell type, likely reflecting the gene activity of polycomb-regulated genes. It is generally thought that PcG bodies form near to localized clusters of PcG-regulated genes, or as a result of interaction with insulator proteins at PcG-regulated genes (reviewed in Pirrotta and Li 2012).

It has been speculated that PcG bodies may have some dependency on lncRNA for their formation or function, largely due to the growing number of reports indicating that individual lncRNAs can associate with PcG proteins. In this context, it is of interest that a recent report has identified TUG1 as a PcG localized lncRNA (Yang et al. 2011). TUG1 (Taurine upregulated gene J) is a conserved mammalian lncRNA that was first found up-regulated in mouse post natal retinal cells following taurine treatment, with evidence that it promotes proliferation through chromatin regulation (Young et al. 2005). TUG1 was subsequently observed in clear defined speckles in the nucleus and cytoplasm of several human and mouse TUG1 associates with a variety of proteins associated with transcriptional repression including the PcG proteins, and that TUG1 localized within PcG bodies. As indicated above, TUG1 is involved in directing the recruitment of gene loci to PcG bodies, via interactions with methylated PcG and its associated gene targets (Yang et al. 2011).

Another lncRNA with a potential involvement in PcG bodies is HOTAIR. The HOTAIR lncRNA acts as a scaffold to recruit chromatin-modifying complexes to their site of action (Wang and Chang 2011). HOTAIR is expressed from the HOXC locus and its mechanism of action includes recruiting the PcG protein PRC2 to multiple loci, playing crucial roles in development and cancer metastasis (Gupta et al. 2010; Kogo et al. 2011). In cancer cells, high HOTAIR expression is associated with increased metastasis, as it redirects chromatin-modifying complexes to suppress metastasis suppressor genes and pro-apoptotic factors (Tsai et al. 2010). RNA FISH against HOTAIR is human foreskin fibroblasts revealed a pattern of distinct foci found throughout the nucleus and cytoplasm, however, it is yet to be determined if these nuclear foci overlap PcG bodies, or represent distinct structures (Khaliq et al. 2009). It is likely that these foci could be colocalized with the gene loci regulated by HOTAIR, and the organization of HOTAIR into these bodies may enhance the efficiency of the regulation. It will be important in the future to determine the composition and role of these HOTAIR foci in the function of this important lncRNA.

4 Concluding Remarks

While there is only at present a handful of lncRNAs known to associate or form subnuclear bodies, these molecules have nevertheless provided a wealth of information about the mechanisms that lncRNAs can use when enriched in subnuclear bodies to alter gene expression (Fig. 3). It is also highly likely that the small number of lncRNAs described here may in fact represent the tip of the iceberg, in terms of the number of lncRNAs that will eventually emerge as associating or forming subnuclear structures. This is likely considering that most lncRNAs are found enriched in the nucleus and are tissue, developmental stage, or cell-type specific, and their localization, if indeed examined at all, are yet to be studied in the relevant cell type. Given this likelihood, it is with confidence that the efforts of researchers in the field of nuclear organization be redoubled to identify function for subnuclear structures, as this will continue to be important in increasing our understanding of lncRNAs that form them and localize to these bodies.
Fig. 3 Functions of IncRNAs in subnuclear bodies. a Several subnuclear bodies (paracombles, omega speckles, cSIs, and gomfu speckles), formed by IncRNAs, act to sequester nuclear proteins, thereby reducing their availability within the nucleoplasm and affecting transcriptional and alternative splicing regulation by these factors. These bodies may also be involved in retaining specific RNAs within the cell nucleus. b MALAT1 presence in nuclear speckles has been demonstrated to influence the phosphorylation of pre-mRNA splicing factors, thereby affecting alternative splicing in the cell. c TUG1 in polycomb bodies and MALAT1 in nuclear speckles can both bind Polycomb group protein PRC2 (although TUG1 binds the methylated PRC2), resulting in the recruitment of gene loci to active (nuclear speckles) or repressed (polycomb bodies) environments.

References


Long Non-coding RNAs and Nuclear Body Formation and Function


Appendix 2: Sequences of cloned human *NEAT1* promoter and enhancers in Chapter 4

Each cloned DNA fragment will have a *Kpn*1 site (GGTACC) on the 5’ end and *Nde*I site (GCTAGC) on the 3’ end.

Please see Table 2.5 for sequence of all primers mentioned below.

Cluster ‘A-B’

Sequenced using primer: A685, A686

```
1  GGTACCGCAC  AATATCTTGG  TTTTACATTA  TTTTGCAACG  GCCTCTTCCC
51  ACTTAATCCA  TCCTGAACAC  ACTTCTTGGA  TGTCTACC  CGGCAGGGAT
101  GGCATTCTTC  AGGAAACACG  TCCCCCGCC  AGGCCCCTGG  GAAGAGTCAA
151  AACCACGATG  CCTGCCCCTG  AAGCGCCCGC  GCGGCTCCAC  GGGGCTCCAT
201  GTTGTCACCC  ACTAGCTCCT  GGAGCTATC  AGCCCGAGC  AGGTTTCTCTCTCT
251  GGCAGAAGAA  ACCGCTGGTT  GGGTGCGGA  GGTTCCAG  TCCCCCGCA
301  GCAACCCCGCT  GGGCTCTCTG  GCGGGCCGCC  CCCCAACCAG  CGCCCGGGGC
351  CTGACGTCAT  GGGCGAGGCC  CAACTGGGAA  CACCGCGCGG  AAGATCGAGC
401  GGGCCCGCGG  CCGGGACGGG  CCGCTTCGCC  AGATCCCTCC  GGGGCCGCT
451  GGAATTTTTC  AGATGGCTTG  CCGGCCCTTC  TTTGGGCCCT  CGGCTGGGCG
501  TGGTGACCTCC  AGGCGAAGCC  GGGCGCTTCC  GTGCGGACAT  CAAGGGGGGC
551  CCCGCCGGGC  CACGCCGCCC  AGCCCTCTTC  GGGGAAAGGC  GTGCCCGGCG
601  CGACCTCAAC  AACATCGGGG  AAGAAAAGGG  GTCTTCTTCC  TCAATGGCAT
651  GCCCTCCCAA  ATGTCACCTG  GAGGCGTGGC  CGGATCCCA  AAGACCTGAC
701  TTAAAGAGAA  GGGGGGGACG  AATCTCGGCC  TTTGCTTGGG  GCCGCTGGTGG
751  AGACATGCGA  CGCCGCGGGA  GTCTTCCGG  GCAGGGTCGG  GAGGGACCTT
801  TTTTTGCGTC  GGAGTGGGAG  GAATGCTGCC  TTTGATGACA  GCCCGGCTGG
851  ACAGCGCTGA  GGGCTATAAA  AGCAAAAGTT  GTGGCAGATC  CAGCGGGAGT
901  TAGCGACAGG  GAGGGATGCG  CGCCTGGGTG  TAGTTCTGGG  GGAGAAGAGT
951  GCTAGC
```

Cluster ‘A-C’

Sequenced using primer: A685, A708, A686

```
1  GGTACCGAAC  AAATCTTGGG  TTTTACATTA  TTTTGCAACG  GCCCTCTTCCC
51  ACTTAATCCA  TCCTGAACAC  ACTTCTTGGA  TGTCTACC  CGGCAGGGAT
101  GGCATTCTTC  AGGAAACACG  TCCCCCGGC  AGGGCCCTGG  GAAGAGTCA
151  AACCACGATG  CCTGCCCCTG  AAGCGCCCGC  GCGGCTCCAC  GGGGCTCCAT
201  GTTGTCACCC  ACTAGCTCCT  GGAGCTATC  AGCCCGAGC  AGGTTTCTCTCTCTCT
251  GGCAGAAGAA  ACCGCTGGTT  GGGTGCGGA  GGTTCCAG  TCCCCCGCA
301  GCAACCCCGCT  GGGCTCTCTG  GCGGGCCGCC  CCCCAACCAG  CGCCCGGGGC
351  CTGACGTCAT  GGGCGAGGCC  CAACTGGGAA  CACCGCGCGG  AAGATCGAGC
401  GGGCCCGCGG  CCGGGACGGG  CCGCTTCGCC  AGATCCCTCC  GGGGCCGCT
451  GGAATTTTTC  AGATGGCTTG  CCGGCCCTTC  TTTGGGCCCT  CGGCTGGGCG
501  TGGTGACCTCC  AGGCGAAGCC  GGGCGCTTCC  GTGCGGACAT  CAAGGGGGGC
551  CCCGCCGGGC  CACGCCGCCC  AGCCCTCTTC  GGGGAAAGGC  GTGCCCGGCG
601  CGACCTCAAC  AACATCGGGG  AAGAAAAGGG  GTCTTCTTCC  TCAATGGCAT
651  GCCCTCCCAA  ATGTCACCTG  GAGGCGTGGC  CGGATCCCA  AAGACCTGAC
701  TTAAAGAGAA  GGGGGGGACG  AATCTCGGCC  TTTGCTTGGG  GCCGCTGGTGG
751  AGACATGCGA  CGCCGCGGGA  GTCTTCCGG  GCAGGGTCGG  GAGGGACCTT
801  TTTTTGCGTC  GGAGTGGGAG  GAATGCTGCC  TTTGATGACA  GCCCGGCTGG
851  ACAGCGCTGA  GGGCTATAAA  AGCAAAAGTT  GTGGCAGATC  CAGCGGGAGT
901  TAGCGACAGG  GAGGGATGCG  CGCCTGGGTG  TAGTTCTGGG  GGAGAAGAGT
951  GCTAGC
```
Cluster ‘A-D’

Sequenced using primer: A685, A658, A687, A707, A686

1 GGTACCCCAA TTAAACCTATA GTCCCCGCTA CTCCGGAGGC
51 TGAGGGCTGC TATCGACAG CTGCCGCAAG GCGGCGCGGC TGGCATCTGA
101 ATGAGGCAGG CCGGGCCTGG TGGCGGGCGG GAGGGCGCGG TGGCCTGGGC
151 AAAAAAGAA AAGAAGAGAA GGCGCTGAGT TACCGCAGTAA AAGTCGCGC
201 TCCAGACATC GCCAACAGC CCTCTCGGGT TGGCATCTTG TGGCATCAGC
251 GAGGAGGAGG TGGCCCTGG GAGGCGCCTA TGGGCTGAGC TGGGCTGAGC
301 GCTCCACCAG GGAATTAGAG AATGCTTCTG CTCAGAGGGA CCACTATGAC
Cluster ‘A-E’


1 GGTACCGTGA TGGGGCCACC GGGACAAGCA CAGTGCAGGC CATTCCCATC
51 AGCAAGGACC CAAGAAGGAG TGTTAGGCTC TGCCAGTGGG GGGGCGTGGC
101 GTGTCCTGAAG ACAGAGGGCT TTGCAAACCT AGGGGTGGAG GTAGTGGGAT
151 GTGCAAGCCTC GGCGCCCTTG CTACAGGAGG GTGGCAGGCC GATTCCCCCTC
201 GGTTGGGCTT GGGGAGGAGG CTGGAAAGGT TAGCACAGGC GAAAGTGTGG
251 GGCCTAGGAA AGCTGGGCTC GCTTATTTGG CCAAGGAGGG TGGGGAGGGT
301 TGAGGAGGAT CTAGGGGCTG CTGAGGGGGG AGGGGAGGCC AGGGGAGGAG
351 GGTATCGGGG CCCAGCCACG CATTCACACG GCTTCATGCT TGGGGGACA
401 GGAATGAGGA CACAGGCGCT CACGGTGAGA GGTGTTGGAAC GCGAGGCCA
451 CGTGACGCTCT GTGACAGGAT GGGGCTACAG GAGACAGAGG TAGCAGTAAG
501 CTGGTGAGGA GCAGGCGGAG CAGAGGAGGG GTAGTGGGAT CTGGGAGGCC
551 AGCTCAGGCA AGGCTCCAAAG GAGTGGGTGA GTGCTCAGGA GGACGGGGCA
601 CCAAGGCCGT GGGTAGAAGG CAGGGCCACC ACCAAGCTG CAGGAGGAGG
651 AACAGGGGACC AGTCCAGAGA CATCCGGATG CAGGGGGCTG ACACCCCAA
701 GTTCTTCTCT GTAAAGCACA GCCCGAGGCC TGGGGCTTGG CCCTGGGCAG
751 AGGAGGCCCT CTTTATGATC GCCACTGGG CAGTGGAGGA AAGGCGAGGC
801 CAGACCCCCAG CGGGGCTGAG GTGACGGTAC AATGCAAGA GTAGAAGGCC
851 CCCCTCTTGA AGGGCTGCTG TCCCTGAGC CCCCGCCAGC CCCAGCGCAG
901 TCTTCCCACT GCGTGGGACC ACCTTGGGAC ACCAATCTTTA GATCTTTCT
951 GCGGAGGAGG TCAGTGCTGC TGGGGGGGGA GGGGGGGGGG TGGGGGCTGG
1001 TGAGGAGGAGG CCCAGGACCT TGGGTAGGGA GGAAGGCCCC GTGGGGGCTG
1051 GGTGGAGGAGA GTGGCTGCTC TTAGGACGAG GAGGCTGGGC CATGTCAC
1101 CATCACTCAG CATCCCCCTT TAAAGCAGAG CAGGGGACAG CCCCCCCCCC
1151 CTGATCCTCG TGGCTTAACT GCGTCTACC AACGTGGTGG CAGTGGTCA
1201 CCCATCCCT TGGGGCCAG CAGGGCGATT CAGGCTCAG ATGAGCAGC
1251 CGGCTGCTTT TCTCTTGGCC TGGCTCCTCC ACCGGGCGCT CCAAGACAGC
1301 GGGCTGCTCA TGAGCTTTAT GGGCGTCCA GCTAGACGTTA CCTCCCTTCT
1351 CTGACTGCTAT CGAGAATTT ATTCAGCCAG CCTCGGAGGT CTCATGCTCC
1401 TTACAAACAA AAAGTGGCACA GTACAGTGCC AGGCGCACGC AATCCCGGCC
1451 AAAAGCAAGC GGAGCCCTCT TCCTCGAGGG ATCAAAGTGG GGCAGGCCTT
1501 TAGAGTGGTG GTGACGGGAG TGGCGTATGG TGACCTTAG GCCAGGAGCC
1551 TGAAGGAGCA GTGGAATGGG ACTAGGCAAG AACGAAGTGG AGAGTGGTCC
1601 AGGAAACACT ACACTGTCAG AATCAAGAGA TGGTGGGCCCC GAGGAGGCC
Cluster ‘A-F’


1 GTTACCATAC AGTAAATGC AGCAAAGAGA GAGGGCATGG GATGGAGCAG
51 GGACGTCAAC CCAAAAGCCTG CCTCTTTGAG CCTAAAGGGAC ACTGACAGTG
101 CTTGACCACAAGGAGGTGCAAC CCTCAGGACAG ACTCTTCCAT CCTCTTTCTC
151 TCTCCTTTTG GCAGAGGACT TTTAGAGCTA CCTGGCATTT TTCTCCCTGT
201 CCAAAATTCA CCAAAATTCA AAAGAAAGTG GGCAGACTAC CCAACTTCAC
251 CCAGAATGGG GGTGCAGGGC TGGGCCTGGG TCTGAGGACT TCAGTTTTAA
301 GCAGCAAGTC CAGTGCCGTG TGGTGGCCCG GCTCCTCTCG CCCCCACTGG
351 GCGGACAGCG TGGGTGGGGC GGAGCCTACT CCCATCTGCA GGTTGCTGGT
401 TCGGCTGCGG CCAGCAGGGG GCAGCTGACG TCACCCTTTTC CTCCTGCCGG
451 GCCGCGCCGTG GGTAGGTTG TGGGGGAGGA AGTGGCTAGC

258
| 2801 | GCCGGCGCTT TAGAGTTGTG GTCAAGGGGC TGGCTGAGTG GTGACCTTAG |
| 2851 | GGCAGAGCCG TAGAGGAGCA GTGGAAATGG ACTAGGACG AGCTAGGTTA |
| 2901 | AGAGTTTGCAG ACCACTGCTAG ATATCAAGAGA TGTATTGGGGG |
| 2951 | GAGGAGCACC TTCAAAATGGG GAAGCAGCTAG CTGGGCTTGG |
| 3001 | TTGGCATTAG GAGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 3051 | GCCCGTTGTA GAGGCTGGGG GGGCCAGCGG GACATGGGCTA |
| 3101 | CAGGCTGGGT AGAAGGGCGG CCCCCGACGG ACTGAGGTGG |
| 3151 | CCAGCCCGCA GCCGGTGGAG GATGGGGATG CGGCTGAGTG |
| 3201 | GCCTAGATTC GAGCCAGCTG GGGGGCGGTT CAGAGGACGGC |
| 3251 | GCTCCAGAAC TAGAGCGGCA CCAAGGACGG ACTGAGGTGG |
| 3301 | TCAAGGGACTA GTGAGGTTTG CAGAGGACGG ACTGAGGTGG |
| 3351 | AGAGGCCAAC TCAAAATGGG GAAGCAGCTAG CTGGGCTTGG |
| 3401 | TGGTGGAATC CCGGGTGACT TCTTGGGACT GCTGGGGCGC |
| 3451 | ACCACTGCAG TCCAGCTGGG CCGACACAGT GAGACTGCTA |
| 3501 | CAGGGAGCCC GCAGCCGGCCG GAGGCTGGGG GCTGGGTATG |
| 3551 | AGAGTTTGGC CACCACGCCG GAGGCTGGGG GCTGGGTATG |
| 3601 | TCAAGGGCTG GAGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 3651 | ATGTGGTAAA ACCCTGCTCT ACTAAAAATA CACAAAAAATA |
| 3701 | GCCAAATGGG TCTGGTGAAG CGAGGTGGTT CAGAGGACGGC |
| 3751 | CAGGAAATTC CTTTCTTGAC GAGGCTGGGG GCTGGGTATG |
| 3801 | ACCACTGCAG TCCAGCTGGG CCGACACAGT GAGACTGCTA |
| 3851 | AACAGAGTCG TGGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 3901 | CAGCCCCAGG GAGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 3951 | AGGAGTTTGGC CACCACGCCG GAGGCTGGGG GCTGGGTATG |
| 4001 | ACCCGGGATG GCTTGGCTGA CAGAGGACGG ACTGAGGTGG |
| 4051 | TTTGTGACTA CAGGAACTTG TCTGACCGGG CCCCCAGGTC |
| 4101 | AGCTGGGGAC AGGTATTTTT GAGGATATTTT CTGAGAGAGC |
| 4151 | CAGGGGGCTG GAGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 4201 | AGGCGAGGACA GAGATATCCA TGGGTGGTGC TACCCCGCGG |
| 4251 | TATAGACCTCT TCTGCTTGGG TGTGGTTGTT GAGAGAGAGC |
| 4301 | CAGCAATCAG GCTGACTTGG GAGGCTGGGG GCTGGGTATG |
| 4351 | GAGGGATTTG GCTGGGGGTT TCTGGGCTGG GCTGGGTATG |
| 4401 | CTGTTGCTAGC ATGAGGCTAT CAGAGGACGG ACTGAGGTGG |
| 4451 | TTTTATAGTA CAGGAACTTG TCTGACCGGG CCCCCAGGTC |
| 4501 | AGCTGGGGAC AGGTATTTTT GAGGATATTTT CTGAGAGAGC |
| 4551 | CAGGGGGCTG GAGGCTGGGG GGGCCAGCCT GATGGTCAGTT |
| 4601 | GCCTCAGTCT CCCAGTTAGC TGGAGATCTA GGGAGTAGTG |
| 4651 | TTTTTTTTTT TTTTTTTTTT ATTTTTTTTA GAGAGGCTAC |
| 4701 | TTCCAGGGCT GCAGGGCTAG CAGGGCTAGC CAGGGCTAGC |
| 4751 | GCCCTAAAAG CCAAGGTGAT CACTCTTTTT TTTGAAAAAT AAAAAATCC |
| 4801 | ATGTTGTGGT TCTGCTTGGG TGTGGTTGTT GAGAGAGAGC |
| 4851 | CGAGAGGCTC CAAATAAGAA AGGAGTTTAA CATCGAAACT |
| 4901 | AACAGACCTA CAGGGGGGAC GACCACCTTG TCTGATGCTG |
| 4951 | AGAAGAGCTT GGCTGACCTG ACTGAGGGCT ACGGTGAGAG |
| 5001 | ACTCCCTCAT TGGTGGTCTA GAGTGGCGAG'C CTTTTTGAAC CATTAGTCA |
| 5051 | GACAATCCAG TCTTCTCCACA CAGGGCTCTT AAGCCGACTT CCCCCACCC |
Appendix 3: Sequences of homology arms for homologous recombination repairing of CRISPR-mediated NEAT1 knockouts

Please see Table 2.5 for sequence of all primers mentioned below.

Homology arms for HR of total NEAT1 KO:

```
CGCGGCTCCACGGGCTCCATGGTGTACACCACACTAGCTCTCTGAGCTATAGCCGAGGGTTTCTGGGCCAGAGAACCCAGCGCGACACTCTGCTGCTACGCTCTCCAGGCTGACGCCTGGGTGTCCTCAGCAGAGACAGCGCCGTCCTGGGCTGGGAAGCTTGGCAAGGAGACTAGGTCTAGGGGGACCACAGTGGGGCAGGCTGCA
```

**TATAAA:** TATA box of human NEAT1 gene

Letters in **orange:** left homology arm

Letters in **green:** right homology arm for gRNA_v2

Letters in **blue:** extra sequence specifically in right homology arm for gRNA_v1 in addition to the right arm of gRNA_v2

Shaded by **green:** PAM sequences (NGG/NAG)

| sign: CRISPR-Cas9 cut sites

For cloning, the entire orange, blue and green region was first amplified from human genomic DNA using primer A932 and A933. This product served as an overall net PCR product to be used as template for all subsequent PCRs. For constructing HR repair template for DSB induced by gRNAs, the left arm was amplified using A927 and A928.
The right arm for DSB induced by gRNA_v1 was amplified using A958 and A959. The right arm for DSB induced by gRNA_v2 was amplified using A960 and A959. All constructs were sequenced using primer: A954, A955, A956, A957

Homology arms for HR of NEAT1_2 KO:

```
ATCTGAAAACCTGTTGGTTCTGAGCTGCTATTGAATTGGTAAAGT
AATACCAATGCTTTTTATCAATTTTCTCTCCCTCTTAAGTTTACCTTGA
CTTTGTAAAATACTGGTTATTATTTATCGTTGAGATTTCTCTGTCTCTG
GGTTCACTTTTTTTAAATATGGGACATTTTGAAATTTTAAAAATCATGGTTATTTTTATCGTTGGGATCTTTCTGTCTTCTG
TTAAACTGAAAACCTGTTGGTTCTGAGCTGCTATTGAATTGGTAAAGT
AATACCAATGCTTTTTATCAATTTTCTCTCCCTCTTAAGTTTACCTTGA
```

Letters in **orange**: left homology arm

Letters in **blue**: right homology arm

Shaded by **green**: PAM sequences (NGG)

| sign: CRISPR-Cas9 cut site

For cloning, the entire orange, blue and green region was first amplified from human genomic DNA using primer A938 and A939. This product served as an overall net PCR product to be used as template for all subsequent PCRs. For constructing HR repair template for DSB induced by gRNAs, the left arm was amplified using A934 and A935.
The right arm was amplified using A961 and A962. All constructs were sequenced using primer: A954, A955, A956, and A957.
## Appendix 4: Supplementary figures and tables for Chapter 4 and 5

### Supplementary Table S4.1: NEAT1 transcription factors.

**(A) Information of TFs that potentially regulate human \textit{NEAT1} gene**

<table>
<thead>
<tr>
<th>Uniprot ID</th>
<th>TF name</th>
<th>Comments</th>
<th>TF Network Level*</th>
<th>sites in cluster F**</th>
<th>sites in cluster E**</th>
<th>sites in cluster D**</th>
<th>sites in cluster C**</th>
<th>sites in cluster B**</th>
<th>sites in cluster A**</th>
</tr>
</thead>
<tbody>
<tr>
<td>P18847</td>
<td>ATF3</td>
<td>Binds the cAMP response element (CRE), a sequence present in many viral and cellular promoters.</td>
<td>Low</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16520</td>
<td>BATF</td>
<td>Controls the differentiation of lineage-specific cells in the immune system.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Q9H165</td>
<td>BCL11A</td>
<td>Functions as a myeloid and B-cell proto-oncogene. May play important roles in leukemogenesis and hematopoiesis.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>P20749</td>
<td>BCL3</td>
<td>Contributes to the regulation of cell proliferation, and transcriptional activation of NF-kappa-B target genes.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Q9NYF8</td>
<td>BCLAF1</td>
<td>Death-promoting transcriptional repressor.</td>
<td>Middle</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>014503</td>
<td>BHLHE40</td>
<td>Transcriptional repressor involved in the regulation of the circadian rhythm.</td>
<td>Low</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GeneID</td>
<td>Gene Symbol</td>
<td>Description</td>
<td>Expression</td>
<td>Pearson Corr</td>
<td>Odds Ratio</td>
<td>p-value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P38398</td>
<td>BRCA1</td>
<td>E3 ubiquitin-protein ligase that plays a central role in DNA repair by facilitating cellular responses to DNA damage.</td>
<td>Low</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>060583</td>
<td>CCNT2</td>
<td>Facilitate the transition from abortive to production elongation.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P17676</td>
<td>CEBPB</td>
<td>Regulates the expression of genes involved in immune and inflammatory responses.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P01100</td>
<td>FOS</td>
<td>Proto-oncogene, has an important role in signal transduction, cell proliferation and differentiation.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>014647</td>
<td>CHD2</td>
<td>Leads to chromatin remodeling.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P05412</td>
<td>JUN</td>
<td></td>
<td>TOP</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P01106</td>
<td>MYC</td>
<td>Activates the transcription of growth-related genes.</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P49711</td>
<td>CTCF</td>
<td>Involved in transcriptional regulation by binding to chromatin insulators and preventing interaction between promoter and nearby enhancers and silencers.</td>
<td>TOP</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q16254</td>
<td>E2F4</td>
<td>Found in the promoter region of a number of genes whose products are involved in cell cycle regulation or in DNA replication.</td>
<td>Low</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>075461</td>
<td>E2F6</td>
<td>Inhibitor of E2F-dependent transcription.</td>
<td>Low</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9UH73</td>
<td>EBF1</td>
<td></td>
<td>TOP</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P18146</td>
<td>EGR1</td>
<td>Activates the transcription of target genes whose products are required for mitogenesis and differentiation.</td>
<td>Low</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P32519</td>
<td>ELF1</td>
<td></td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P28324</td>
<td>ELK4</td>
<td></td>
<td>TOP</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P03372</td>
<td>ESR1 (ER)</td>
<td>Estrogen receptor, affects cellular proliferation and differentiation in target tissues.</td>
<td>–</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P14921</td>
<td>ETS1</td>
<td>Directly controls the expression of cytokine and chemokine genes in a wide variety of different cellular contexts. May control the differentiation, survival and proliferation of lymphoid cells. May also regulate angiogenesis through regulation of expression of genes controlling endothelial cell migration and invasion.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P15408</td>
<td>FOSL2</td>
<td>Controls osteoclast survival and size.</td>
<td>Low</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P55317</td>
<td>FOXA1</td>
<td>Involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.</td>
<td>Low</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q9Y261</td>
<td>FOXA2</td>
<td>Involved in embryonic development, establishment of tissue-specific gene expression and regulation of gene expression in differentiated tissues.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Gene</td>
<td>Function</td>
<td>Expression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q06546</td>
<td>GABPA</td>
<td>Necessary for the expression of the Adenovirus E4 gene.</td>
<td>TOP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P15976</td>
<td>GATA1</td>
<td>A general switch factor for erythroid development.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P23769</td>
<td>GATA2</td>
<td>Transcriptional activator which regulates endothelin-1 gene expression in endothelial cells.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P23771</td>
<td>GATA3</td>
<td>Required for the T-helper 2 (Th2) differentiation process following immune and inflammatory responses</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P04150</td>
<td>NR3C1 (GR)</td>
<td>Glucocorticoid receptor, affects inflammatory responses, cellular proliferation and differentiation in target tissues.</td>
<td>Middle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q00403</td>
<td>GTF2B (TFIIB)</td>
<td>General transcription factor.</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P35269</td>
<td>GTF2F1</td>
<td>General transcription initiation factor.</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q01094</td>
<td>E2F1</td>
<td>Involved in cell cycle regulation or in DNA replication.</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q92769</td>
<td>HDAC2</td>
<td>Histone deacetylase 2, responsible for the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4).</td>
<td>Top</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9Y5J3</td>
<td>HEY1</td>
<td>Downstream effector of Notch signaling which may be required for cardiovascular development.</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Protein</td>
<td>Description</td>
<td>Localization</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Q15651</td>
<td>HMGN3</td>
<td>Binds to nucleosomes, regulating chromatin structure.</td>
<td>TOP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P41235</td>
<td>HNF4A</td>
<td>May be essential for development of the liver, kidney and intestine.</td>
<td>Low</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q14511</td>
<td>HNF4G</td>
<td></td>
<td>Middle</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q00613</td>
<td>HSF1</td>
<td>Heat shock factor protein 1, specifically binds heat shock promoter elements (HSE) and activates transcription.</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q15306</td>
<td>IRF4</td>
<td>Binds to the interferon-stimulated response element</td>
<td>Low</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P17275</td>
<td>JUNB</td>
<td>Regulating gene activity following the primary growth factor response.</td>
<td>–</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P17535</td>
<td>JUND</td>
<td></td>
<td>Middle</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q13263</td>
<td>TRIM28 (KAP1)</td>
<td>Enhances transcriptional repression by coordinating the increase in H3K9me.</td>
<td>TOP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q9ULX9</td>
<td>MAFF</td>
<td>May be involved in the cellular stress response.</td>
<td>–</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q60675</td>
<td>MAFK</td>
<td>May repress transcription via the recruitment of a chromatin remodeling complex.</td>
<td>–</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>P61244</td>
<td>MAX</td>
<td>Binds specifically to the MEF2 element found in numerous muscle-specific genes. Mediates cellular functions not only in skeletal and cardiac muscle development,</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q7Z6C9</td>
<td>MEF2A</td>
<td></td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
but also in neuronal differentiation and survival.

<table>
<thead>
<tr>
<th>ID</th>
<th>Gene</th>
<th>Description</th>
<th>TOP</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q06413</td>
<td>MEF2C</td>
<td>Controls cardiac morphogenesis and myogenesis, and is also involved in vascular development. Plays an essential role in hippocampal-dependent learning and memory.</td>
<td>TOP</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P50539</td>
<td>MXI1</td>
<td>Binds with MAX to act as transcriptional repressor.</td>
<td>Low</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P18615</td>
<td>NELFE (RDBP)</td>
<td>Negative elongation factor E</td>
<td>Low</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q16621</td>
<td>NFE2</td>
<td>Essential for regulating erythroid and megakaryocytic maturation and differentiation.</td>
<td>Low</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q04206</td>
<td>RELA (NFkB)</td>
<td>Present in almost all cell types, and is the endpoint of a series of signal transduction events that are initiated by a vast array of stimuli related to many biological processes such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P22736</td>
<td>NR4A1</td>
<td>Orphan nuclear receptor.</td>
<td>–</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q13127</td>
<td>REST (NRSF)</td>
<td>Transcriptional repressor which binds neuron-restrictive silencer element (NRSE) and represses neuronal gene transcription in non-neuronal cells.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Accession</td>
<td>Description</td>
<td>Location</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Q09472</td>
<td>EP300 (P300) Functions as histone acetyltransferase and regulates transcription via chromatin remodeling.</td>
<td>TOP</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Q02548</td>
<td>PAX5 Play an important role in B-cell differentiation as well as neural development and spermatogenesis.</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P40426</td>
<td>PBX3 Pre-B-cell leukemia transcription factor 3.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P09086</td>
<td>POU2F2 Regulates transcription in a number of tissues in addition to activating immunoglobulin gene expression.</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P17947</td>
<td>SPI1 (PU.1) A transcriptional activator that may be specifically involved in the differentiation or activation of macrophages or B-cells.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>060216</td>
<td>RAD21 Involved in chromosome cohesion during cell cycle, in DNA repair, and in apoptosis.</td>
<td>TOP</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>P48382</td>
<td>RFX5 Activates transcription from class II MHC promoters.</td>
<td>Middle</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P19793</td>
<td>RXRA Receptor for retinoic acid.</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q96ST3</td>
<td>SIN3A Interacts with MXI1 to repress MYC responsive genes and antagonize MYC oncogenic activities.</td>
<td>Middle</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Q8N196</td>
<td>SIX5 Involved in regulation of organogenesis.</td>
<td>TOP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q9UQE7</td>
<td>SMC3 Central component of cohesin, a complex required for chromosome cohesion during the cell cycle.</td>
<td>TOP</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Prow</td>
<td>EntrezID</td>
<td>Symbol</td>
<td>Description</td>
<td>Location</td>
<td>1</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>----</td>
</tr>
<tr>
<td>P08047</td>
<td>SP1</td>
<td></td>
<td>Can activate or repress transcription in response to physiological and pathological stimuli, and regulates the expression of a large number of genes involved in a variety of processes such as cell growth, apoptosis, differentiation and immune responses.</td>
<td>Middle</td>
<td>1</td>
</tr>
<tr>
<td>P11831</td>
<td>SRF</td>
<td></td>
<td>Serum response factor.</td>
<td>TOP</td>
<td></td>
</tr>
<tr>
<td>P42224</td>
<td>STAT1</td>
<td></td>
<td>Mediates cellular responses to interferons.</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td>P52630</td>
<td>STAT2</td>
<td></td>
<td>Mediates signaling by type I IFNs (IFN-alpha and IFN-beta).</td>
<td>Middle</td>
<td>1</td>
</tr>
<tr>
<td>P40763</td>
<td>STAT3</td>
<td></td>
<td>Mediates cellular responses to interleukins.</td>
<td>TOP</td>
<td>1</td>
</tr>
<tr>
<td>P21675</td>
<td>TAF1</td>
<td></td>
<td>Basal transcription factor.</td>
<td>TOP</td>
<td>2</td>
</tr>
<tr>
<td>Q15545</td>
<td>TAF7</td>
<td></td>
<td>A component of the DNA-binding general transcription factor complex TFIID.</td>
<td>Middle</td>
<td>1</td>
</tr>
<tr>
<td>P20226</td>
<td>TBP</td>
<td></td>
<td>A component of the DNA-binding general transcription factor complex TFIID.</td>
<td>TOP</td>
<td>2</td>
</tr>
<tr>
<td>Q99081</td>
<td>TCF12</td>
<td></td>
<td>Involved in the initiation of neuronal differentiation.</td>
<td>TOP</td>
<td>1</td>
</tr>
<tr>
<td>Q9NQ80</td>
<td>TCF7L2 (TCF4)</td>
<td></td>
<td>Participates in the Wnt signaling pathway and modulates MYC expression.</td>
<td>Middle</td>
<td>1</td>
</tr>
<tr>
<td>P49116</td>
<td>NR2C2 (TR4)</td>
<td></td>
<td>Repressor of nuclear receptor signaling pathways such as retinoic acid receptor, retinoid X, vitamin D3 receptor, thyroid hormone receptor and estrogen receptor pathways.</td>
<td>Middle</td>
<td></td>
</tr>
<tr>
<td>Protein ID</td>
<td>Description</td>
<td>Regulatory Action</td>
<td>Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P22415 USF1</td>
<td>Found in a variety of viral and cellular promoters.</td>
<td>Middle</td>
<td>1 4 2 2 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q15853 USF2</td>
<td>Found in a variety of viral and cellular promoters.</td>
<td>Low</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q96S55 WHIP</td>
<td>Modulator for initiation or reinitiation events during DNA polymerase delta-mediated DNA synthesis.</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P25490 YY1</td>
<td>Exhibits positive and negative control on a large number of cellular and viral genes.</td>
<td>TOP</td>
<td>1 3 2 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q86T24 ZBTB33</td>
<td>Promote histone deacetylation and the formation of repressive chromatin structures.</td>
<td>Middle</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P37275 ZEB1</td>
<td>Positively regulates neuronal differentiation.</td>
<td>Low</td>
<td>1 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*According to the analysis of ENCODE transcription factor ChIP-Seq data by Gerstein et al., 2012, transcription factors can be broadly classified into a hierarchy order of Top, Middle, and Low ‘power’ in regulating gene expression. These three levels are based on the association of each TF at the genomic location, as well as the ncRNA, miRNA, and protein-protein interaction data. As a general trend, the higher the class, the stronger the influence on gene expression.

** See diagram next page for location information of transcriptional regulatory elements A to F.
Table S4.1(B) DAVID gene functional annotation enrichment analysis

**GO_Molecular Function_ALL**  (a selection of relevant terms)

<table>
<thead>
<tr>
<th>Term</th>
<th>TF Counts</th>
<th>%</th>
<th>P-value</th>
<th>Benjamini FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>transcription regulator activity</td>
<td>74</td>
<td>90.2</td>
<td>2.6E-64</td>
<td>4.1E-62</td>
</tr>
<tr>
<td>transcription activator activity</td>
<td>30</td>
<td>36.6</td>
<td>2.8E-25</td>
<td>6.3E-24</td>
</tr>
<tr>
<td>transcription repressor activity</td>
<td>18</td>
<td>22</td>
<td>6.8E-13</td>
<td>8.9E-12</td>
</tr>
</tbody>
</table>

**GO_Biological Process_ALL**  (a selection of relevant terms)

<table>
<thead>
<tr>
<th>Term</th>
<th>TF Counts</th>
<th>%</th>
<th>P-value</th>
<th>Benjamini FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>regulation of transcription</td>
<td>78</td>
<td>95.1</td>
<td>4.9E-53</td>
<td>6E-50</td>
</tr>
<tr>
<td>system development</td>
<td>38</td>
<td>46.3</td>
<td>4.6E-10</td>
<td>7.9E-09</td>
</tr>
<tr>
<td>organ development</td>
<td>32</td>
<td>39</td>
<td>1.6E-09</td>
<td>0.000000027</td>
</tr>
<tr>
<td>cell differentiation</td>
<td>28</td>
<td>34.1</td>
<td>1.5E-07</td>
<td>0.0000023</td>
</tr>
<tr>
<td>response to cytokine stimulus</td>
<td>7</td>
<td>8.5</td>
<td>0.0000055</td>
<td>0.000077</td>
</tr>
<tr>
<td>response to hormone stimulus</td>
<td>12</td>
<td>14.6</td>
<td>0.0000066</td>
<td>0.000092</td>
</tr>
<tr>
<td>response to mechanical stimulus</td>
<td>6</td>
<td>7.3</td>
<td>0.000016</td>
<td>0.00021</td>
</tr>
<tr>
<td>embryonic development</td>
<td>14</td>
<td>17.1</td>
<td>0.000017</td>
<td>0.00022</td>
</tr>
<tr>
<td>placenta development</td>
<td>6</td>
<td>7.3</td>
<td>0.000022</td>
<td>0.00028</td>
</tr>
<tr>
<td>regulation of cell proliferation</td>
<td>16</td>
<td>19.5</td>
<td>0.00003</td>
<td>0.00038</td>
</tr>
<tr>
<td>reproduction</td>
<td>15</td>
<td>18.3</td>
<td>0.000092</td>
<td>0.0011</td>
</tr>
<tr>
<td>regulation of cell cycle</td>
<td>10</td>
<td>12.2</td>
<td>0.0001</td>
<td>0.0012</td>
</tr>
<tr>
<td>regulation of apoptosis</td>
<td>15</td>
<td>18.3</td>
<td>0.00015</td>
<td>0.0017</td>
</tr>
<tr>
<td>response to stimulus</td>
<td>35</td>
<td>42.7</td>
<td>0.00038</td>
<td>0.0039</td>
</tr>
<tr>
<td>response to stress</td>
<td>21</td>
<td>25.6</td>
<td>0.00092</td>
<td>0.0088</td>
</tr>
</tbody>
</table>

**KEGG_Pathway**

<table>
<thead>
<tr>
<th>Term</th>
<th>TF Counts</th>
<th>%</th>
<th>P-value</th>
<th>Benjamini FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathways in cancer</td>
<td>14</td>
<td>17.1</td>
<td>3.3E-07</td>
<td>0.000017</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>10</td>
<td>12.2</td>
<td>0.00011</td>
<td>0.0029</td>
</tr>
<tr>
<td>Basal transcription factors</td>
<td>5</td>
<td>6.1</td>
<td>0.00012</td>
<td>0.002</td>
</tr>
<tr>
<td>Cell cycle</td>
<td>7</td>
<td>8.5</td>
<td>0.00028</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

**PANTHER_PATHWAY**

<table>
<thead>
<tr>
<th>Term</th>
<th>TF Counts</th>
<th>%</th>
<th>P-value</th>
<th>Benjamini FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDGF signaling pathway</td>
<td>12</td>
<td>14.6</td>
<td>0.0000014</td>
<td>0.000046</td>
</tr>
<tr>
<td>Interleukin signaling pathway</td>
<td>10</td>
<td>12.2</td>
<td>0.00027</td>
<td>0.0047</td>
</tr>
<tr>
<td>Oxidative stress response</td>
<td>6</td>
<td>7.3</td>
<td>0.00052</td>
<td>0.0059</td>
</tr>
<tr>
<td>TGF-beta signaling pathway</td>
<td>7</td>
<td>8.5</td>
<td>0.0052</td>
<td>0.043</td>
</tr>
<tr>
<td>Ras Pathway</td>
<td>5</td>
<td>6.1</td>
<td>0.015</td>
<td>0.098</td>
</tr>
<tr>
<td>JAK/STAT signaling pathway</td>
<td>3</td>
<td>3.7</td>
<td>0.023</td>
<td>0.12</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td>6</td>
<td>7.3</td>
<td>0.088</td>
<td>0.36</td>
</tr>
</tbody>
</table>

**UP_Tissue**

<table>
<thead>
<tr>
<th>Term</th>
<th>TF Counts</th>
<th>%</th>
<th>P-value</th>
<th>Benjamini FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epithelium</td>
<td>30</td>
<td>36.6</td>
<td>6.2E-07</td>
<td>0.000055</td>
</tr>
<tr>
<td>B-cell</td>
<td>5</td>
<td>6.1</td>
<td>0.03</td>
<td>0.74</td>
</tr>
<tr>
<td>Tissue Type</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Human endometrium</td>
<td>2</td>
<td>2.4</td>
<td>0.056</td>
<td>0.82</td>
</tr>
<tr>
<td>Skin</td>
<td>13</td>
<td>15.9</td>
<td>0.086</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Figure S5.1: Highlighted cellular functions and molecular networks in DE analysis for Total NEAT1 KO vs WT. (A) DE genes that are related to cell migration as identified by Ingenuity Pathway Analysis (Ingenuity Systems, www.ingenuity.com). Green: Downregulated. Red: Upregulated. The red dotted lines imply predicted activatory relationship. The blue dotted line imply predicted inhibitory relationship. The yellow dotted lines imply inconsistent finding with regards to the state of downstream molecule. The grey dotted lines imply unpredictable relationship.
Figure S5.1(B): The biochemical network of ‘Cellular movement, Cell-to-Cell signaling and Interaction, Tissue development’ was predicted as the top canonical pathway affected by the NEAT1 knockout by Ingenuity Pathway Analysis (Ingenuity Systems, www.ingenuity.com). Green: Downregulated. Red: Upregulated. Solid lines imply direct relationship between genes; dotted lines imply indirect interactions.
Figure S5.1(C): The network of ‘Cellular Development, Cellular Growth and Proliferation, Embryonic Development’ was predicted as the 3rd top canonical pathway affected by the NEAT1 knockout by Ingenuity Pathway Analysis (Ingenuity Systems, www.ingenuity.com). Green: Downregulated. Red: Upregulated. Solid lines imply direct relationship between genes; dotted lines imply indirect interactions.
Appendix 5: References


conserved repeats and is highly localized within the nucleus. *Cell.* 71:527-542.


