

1 **Maternally expressed, paternally imprinted, embryonic non-coding RNA are expressed in**
2 **osteosarcoma, Ewing sarcoma and spindle cell sarcoma**

3

4 Darrell Green¹, Archana Singh², Jasmine Sanghera³, Lee Jeys⁴, Vaiyapuri Sumathi⁵, Tamas
5 Dalmay³, William D Fraser^{1, 6*}

6

7 1. Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, NR4
8 7TJ, United Kingdom

9 2. John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

10 3. School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich,
11 NR4 7TJ, United Kingdom

12 4. Department of Orthopaedic Oncology, The Royal Orthopaedic Hospital, Birmingham, B31
13 2AP, United Kingdom

14 5. Department of Musculoskeletal Pathology, The Royal Orthopaedic Hospital, Birmingham, B31
15 2AP, United Kingdom

16 6. Department of Clinical Biochemistry, Norfolk and Norwich University Hospital, Norwich
17 Research Park, Norwich, NR4 7UY, United Kingdom

18

19 ***Corresponding author:** Prof William D Fraser, Norwich Medical School, University of East
20 Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. Tel: 01603 597174 Email:
21 W.Fraser@uea.ac.uk

22

23 **Conflict of interest:** No conflict to declare

24

25 **Word count of main body of text:** 1,496

26

27

28

29

30 Sir:

31

32 In a human embryo it takes 8 weeks after fertilisation for the skeleton to begin to form, one of the
33 last organs to develop before becoming a foetus. Mesenchymal progenitors, derived from neural
34 crest cells, differentiate into chondrocytes where the skeleton is generated as a mostly cartilage
35 template. Other mesenchymal progenitors envelop the template, activate runt related transcription
36 factor 2 (RUNX2) and bone morphogenetic protein 2 (BMP2) and differentiate into osteoblasts,
37 where an osteoid matrix is secreted and subsequently mineralised to become bone.¹ During
38 development and up to late adolescence, cellular proliferation enabling skeletal growth is restricted
39 to the metaphysis and epiphyseal line or “growth plate”. It is in the growth plate of long bones
40 where most bone cancers develop, hence the predominantly childhood incidence of the cancer.
41 Primitive mesenchymal cells undergo transformation to form a heterogeneous group of bone
42 malignancies. The most common type of bone cancer in children is osteosarcoma, mostly initiated
43 by tumour protein p53 (*TP53*) mutations. The second most common type of bone cancer in children
44 is Ewing sarcoma, mostly initiated by a EWS RNA binding protein 1-Fli-1 proto-oncogene, ETS
45 transcription factor (*EWSR1-FLI1*) fusion. There are an average of 160 and 55 new cases of
46 osteosarcoma and Ewing sarcoma, respectively, every year in the UK. Five-year survival for both
47 cancer types is 50% when diagnosed early. Five-year survival is 15% when lung metastases are
48 present at diagnosis. Treatment progress for bone cancer is poor when compared to other cancers
49 such as breast where there is a twenty-year survival of 70%. Bone cancer requires extensive and
50 sometimes disabling multimodal treatment. Chemotherapy for osteosarcoma includes
51 methotrexate, cisplatin and doxorubicin, which were developed in the 1940’s and 1970’s.
52 Chemotherapy for Ewing sarcoma includes vincristine, ifosfamide and etoposide, which were
53 developed in the 1960’s and 1980’s. If the tumour responds well to chemotherapy and
54 radiotherapy, wide area resection or amputation is performed. New understanding of bone cancer
55 biology leading to better diagnosis and better treatments is required.

56

57 Transcriptomic analysis of bone cancer is lacking. Different RNA populations within cells are
58 generally classified as coding and non-coding, i.e. whether they have protein coding potential.

59 Messenger RNA (mRNA) molecules contain a start codon “AUG” encoding methionine at the
60 beginning of an open reading frame. Non-coding RNA lack protein coding ability and usually exist
61 within the cell without a start codon. Over 70% of known non-coding RNA are long non-coding
62 RNA (lncRNA) that are classed by their >200 nucleotide (nt) length. LncRNA similarly to mRNA
63 are transcribed by RNA polymerase II, have a 5' cap and are polyadenylated. LncRNAs have a
64 large diversity of roles including regulation of chromatin dynamics, enforcing imprinting and as
65 microRNA inhibitors by acting as a microRNA “sponge”. LncRNAs are further classified based on
66 their genomic localisation. Intergenic lncRNAs are named for their production from loci in between
67 genes. Intronic lncRNAs are named for their production from mRNA introns. Sense lncRNA are
68 named for their production from the sense strand of protein coding genes that overlap with an
69 exon/intron. Antisense lncRNA are named for their production from the antisense strand of protein
70 coding genes that overlap with an exon/intron. Another elusive class of non-coding RNA is the
71 small nucleolar RNAs (snoRNAs). SnoRNAs are 60-170 nt in length and are classed as C/D box
72 snoRNAs and H/ACA box snoRNAs. C/D box snoRNAs guide 2'-O-methylation of ribosomal and
73 transfer RNA. H/ACA box snoRNAs guide pseudouridylation mostly in transfer RNAs. The majority
74 of snoRNA are intronic. There is a recent interest in lncRNA and snoRNA and their role in cancer
75 biology. We took a next generation sequencing and bioinformatics approach to evaluate lncRNA
76 and snoRNA expression in bone cancer.

77

78 We extracted RNA using the miRCURY RNA isolation kit (Exiqon) from two tissue specimens of
79 osteoblastic osteosarcoma (patient ages 15 and 19, OS1 and OS2 respectively). OS1 had
80 undergone treatment with cisplatin and doxorubicin prior to surgery (Figure 1A&E). OS2 had
81 undergone treatment with methotrexate, cisplatin and doxorubicin prior to surgery (Figure 1B&F).
82 We extracted RNA from one tissue specimen of Ewing sarcoma (patient age 6, ES) where the
83 patient had undergone nine alternating cycles of vincristine, doxorubicin and cyclophosphamide in
84 one cycle and ifosfamide and etoposide in another cycle prior to surgery (Figure 1C&G). We
85 extracted RNA from one tissue specimen of a spindle cell sarcoma of bone (patient age 17, SCS)
86 where the patient had not undergone systemic treatment (Figure 1G&H). We used publically
87 available data for 4 control samples, which were obtained from long bone tissue derived from

88 surgical reconstruction procedures (patient ages 11-13).² Cancer RNA was stored at -80 °C. We
89 generated cDNA libraries using the SENSE mRNA library prep kit (Lexogen). We performed 150
90 bp paired end sequencing on the HiSeq 4000 Ultra High Throughput Sequencing System (Illumina)
91 at the Earlham Institute, Norwich Research Park.

92
93 Raw fastq files were converted to fasta format. Adapter sequences and reads <20 nt were trimmed
94 using Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore). Trimmed reads
95 were aligned to the human genome (v.38) using HISAT2.³ The latest set of human non-coding
96 RNAs were download from GENCODE (v.28) and Ensembl (v.92). Count matrix for lncRNAs and
97 snoRNAs was created using Kallisto.⁴ LncRNA and snoRNA expression was compared between
98 bone cancer and controls using the DESeq2 package available in R. We selected differentially
99 expressed lncRNA and snoRNA according to log₂ fold change ≥2, *p* value <0.05 and false
100 discovery rate <5%. Hierarchical cluster analysis and principle component analysis investigating
101 77% of variance shows distinct grouping between cancer and controls (Figure 1I&J).

102
103 We report a *H19* transcript, *H19-203*, is highly expressed in bone cancer (Table 1). *H19* is a
104 maternally expressed, paternally imprinted, embryonic lncRNA.⁵ *H19* is a key mediator of sonic
105 hedgehog (*SHH*) signalling in osteoblastic osteosarcoma.⁶ *SHH* ligand is a major embryonic
106 morphogen and is later required for adult stem cell division. *H19* is reciprocally imprinted and
107 regulated with its neighbouring gene insulin like growth factor 2 (*IGF2*).⁵ We find a *MEG3* transcript,
108 *MEG3-224*, to have a low expression in OS1 and ES (44 reads and 91 reads, respectively,
109 normalised data in GEO). *MEG3-224* expression was high in OS2 and SCS (2,091 reads and
110 1,616 reads, respectively, normalised data in GEO) (Table 1). *MEG3* is a maternally expressed,
111 paternally imprinted lncRNA with low expression correlating to a poor prognosis in osteosarcoma
112 and other cancers.⁷ *XIST-201* and *XIST-204*, products of X inactive specific transcript (*XIST*), were
113 highly expressed in our cohort with OS1 and ES showing the highest expression in the cohort
114 (Table 1). *XIST*, located on the X chromosome, has previously been reported as a poor prognostic
115 marker in solid tumours.⁸ We also found high expression of *HCP5-204* that has previously been
116 linked to cancer progression via sponging of miR-22, miR-186 and miR-216a.⁹ *MALAT1-203* and

117 *MALAT1-214*, products of metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*),
118 were downregulated in this study (Table 1).

119

120 We report 4 dysregulated snoRNAs in our bone cancer samples (Table 1). *SNORD68*, recruited
121 by DExD box helicase 21 (*DDX21*) to enable 2'-O-methylation of residue U428 of the 18S
122 ribosomal RNA (rRNA) sequence, was upregulated in each cancer type (Table 1). *SNORD3B-1*
123 and *SNORD3B-2*, close paralogues of U3 snoRNA, guide site specific cleavage of rRNA during
124 pre-rRNA processing.¹⁰ We report *SNORD3B-1* and *SNORD3B-2* are upregulated in bone cancer
125 (Table 1). *SNORD58B* is reported to guide the 2'-O-methylation of residue G4198 of the 28S
126 rRNA.¹¹ We found *SNORD58B* was downregulated in our bone cancer samples (Table 1).

127

128 RNA therapeutics are on the horizon. LncRNA and snoRNA are of recent interest because of their
129 elusive roles in regulating gene expression and epitranscriptomic modification of pre-RNAs. RNA
130 is also a specific biomarker, which is especially helpful in providing a robust diagnosis in rare and
131 heterogeneous cancers. Bone cancers are historically difficult to diagnose and sub classify prior
132 to surgery, which can delay the appropriate choice of neoadjuvant chemotherapeutic agents. *H19*
133 expression after birth is linked to Beckwith Wiedemann Syndrome, which increases the likelihood
134 of childhood cancer but not adult cancer. We find that *H19-203* may be a specific biomarker for
135 osteosarcomas involving *SHH* signalling. Patients may benefit from receiving targeted *SHH*
136 inhibitors sonidegib or vismodegib that are currently used to treat basal cell carcinoma. We also
137 find low *MEG3-224* and high *XIST-201/XIST-204* may be markers of poor prognosis and lower
138 overall survival in patients, which we detect in 2 out of 4 patients. Upregulation of snoRNAs is
139 consistent with the increased proliferative behaviour of cancer cells. *SNORD68*, *SNORD3B-1* and
140 *SNORD3B-2* may be useful biomarkers in the future. Previous research has shown the *EWSR1-*
141 *FLI1* chimeric transcript in Ewing sarcoma is sensitive to snoRNA loss of function due to changes
142 in splicing, demonstrating a potential target for intervention in Ewing sarcoma cells through
143 snoRNA activity.¹²

144

145 A limitation of this study is the size of the cohort studied. Bone cancer is rare and donation to tissue
146 banks is scarce. Our data highlights the value of being able to provide a specific tissue diagnosis
147 in addition to identifying regulatory transcriptomic molecules that could be exploited for targeted
148 therapy.

149

150 **DATA AVAILABILITY**

151

152 The cancer data presented in this study is publicly available on GEO under the accession
153 GSE113916. The control data used in this study is publically available on GEO under the accession
154 GSE55282.

155

156 **FUNDING**

157

158 This study was funded by The Humane Research Trust.

159

160 **ACKNOWLEDGEMENTS**

161

162 Approval to obtain and study human tissue was granted by the Faculty of Medicine and Health
163 Sciences Research Ethics Committee (2015/16 100 HT). We thank Lucy Bishton, Natalie Jackson,
164 Dionne Wortley, Kulvinder Jill and Karen Joynes (The Royal Orthopaedic Hospital) for supporting
165 collection of patient samples.

166

167 **REFERENCES**

168

- 169 1. Green, D., Dalmay, T. & Fraser, W.D. Role of miR-140 in embryonic bone development
170 and cancer. *Clin Sci (Lond)* **129**, 863-73 (2015).
- 171 2. Rojas-Pena, M.L. *et al.* Characterization of distinct classes of differential gene expression
172 in osteoblast cultures from non-syndromic craniosynostosis bone. *J Genomics* **2**, 121-30
173 (2014).

- 174 3. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. & Salzberg, S.L. Transcript-level expression
175 analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**,
176 1650-67 (2016).
- 177 4. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
178 quantification. *Nat Biotechnol* **34**, 525-7 (2016).
- 179 5. Raveh, E., Matouk, I.J., Gilon, M. & Hochberg, A. The H19 Long non-coding RNA in cancer
180 initiation, progression and metastasis - a proposed unifying theory. *Mol Cancer* **14**, 184
181 (2015).
- 182 6. Chan, L.H. *et al.* Hedgehog signaling induces osteosarcoma development through Yap1
183 and H19 overexpression. *Oncogene* **33**, 4857-66 (2014).
- 184 7. Tian, Z.Z., Guo, X.J., Zhao, Y.M. & Fang, Y. Decreased expression of long non-coding
185 RNA MEG3 acts as a potential predictor biomarker in progression and poor prognosis of
186 osteosarcoma. *Int J Clin Exp Pathol* **8**, 15138-42 (2015).
- 187 8. Mao, H. *et al.* Prognostic role of long non-coding RNA XIST expression in patients with
188 solid tumors: a meta-analysis. *Cancer Cell Int* **18**, 34 (2018).
- 189 9. Liang, L. *et al.* LncRNA HCP5 promotes follicular thyroid carcinoma progression via
190 miRNAs sponge. *Cell Death Dis* **9**, 372 (2018).
- 191 10. Clery, A., Senty-Segault, V., Leclerc, F., Raue, H.A. & Branlant, C. Analysis of sequence
192 and structural features that identify the B/C motif of U3 small nucleolar RNA as the
193 recognition site for the Snu13p-Rrp9p protein pair. *Mol Cell Biol* **27**, 1191-206 (2007).
- 194 11. Kiss-Laszlo, Z., Henry, Y., Bachelier, J.P., Caizergues-Ferrer, M. & Kiss, T. Site-specific
195 ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell*
196 **85**, 1077-88 (1996).
- 197 12. Grohar, P.J. *et al.* Functional Genomic Screening Reveals Splicing of the EWS-FLI1
198 Fusion Transcript as a Vulnerability in Ewing Sarcoma. *Cell Rep* **14**, 598-610 (2016).

199

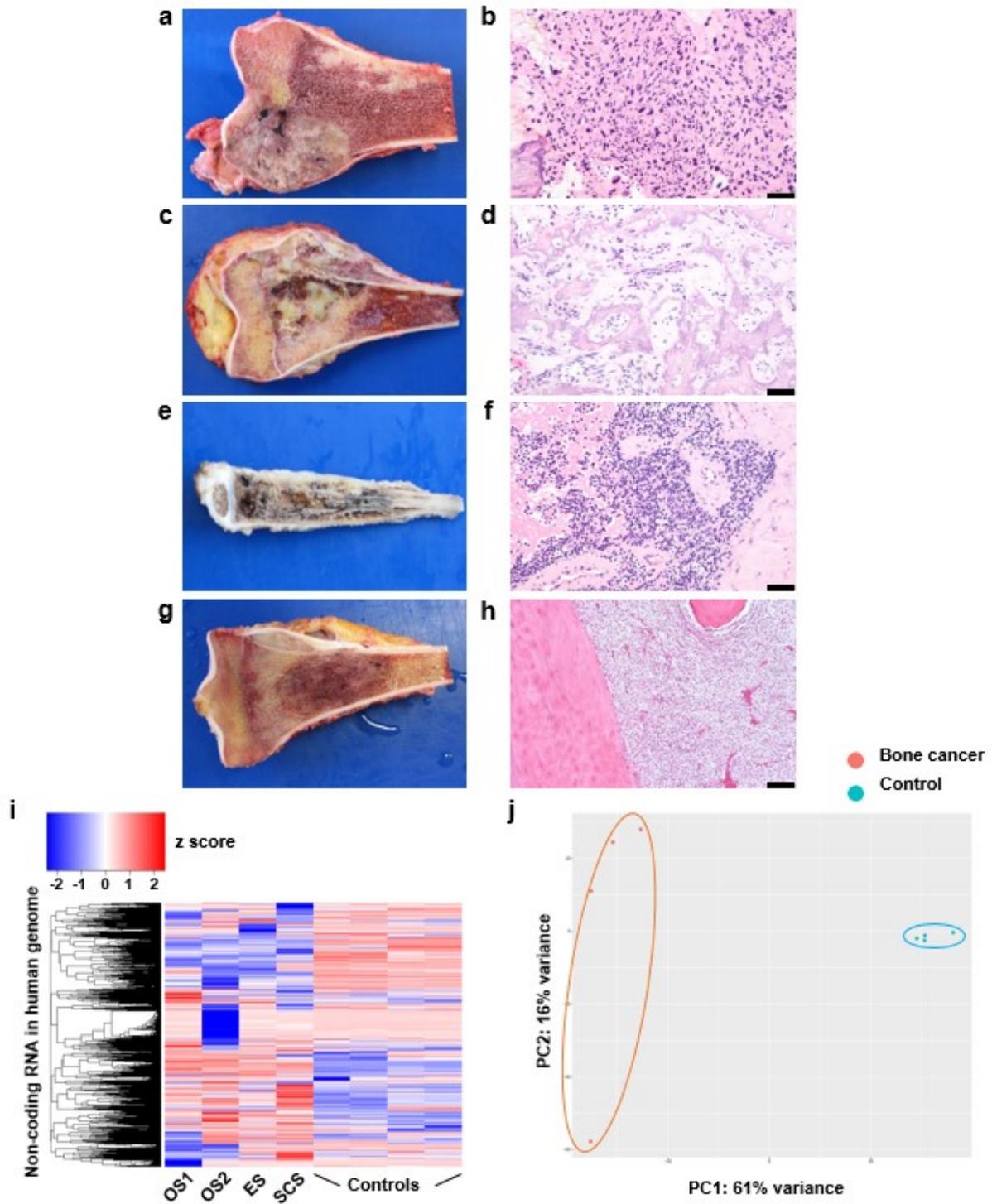
200

201

202

203 **FIGURE 1**

204



205

206

207 **Figure 1.** Macroscopic and microscopic photographs of resected tumours used in this study. In
208 the microscopic images, the black scale bar is 100 μ M (a) osteoblastic osteosarcoma was
209 diagnosed at biopsy. Patient underwent treatment with cisplatin and doxorubicin. On resection, a

210 partly cream/grey focally haemorrhagic tumour is observed (**b**) H & E stain of **a** shows a 5%
211 response to chemotherapy (minimal response). Examination is consistent with osteoblastic
212 osteosarcoma (**c**) osteoblastic osteosarcoma was diagnosed at biopsy. Patient underwent
213 treatment with methotrexate, cisplatin and doxorubicin. On resection, a focally gelatinous and
214 haemorrhagic tumour is observed (**d**) H & E stain of **b** shows a 95% response to chemotherapy
215 (excellent response). Examination is consistent with osteoblastic osteosarcoma (**e**) Ewing sarcoma
216 was diagnosed at biopsy. Patient underwent nine alternating cycles of vincristine, doxorubicin and
217 cyclophosphamide (cycle 1) and ifosfamide and etoposide (cycle 2). On resection, a lesion with
218 extensive involvement of the medulla and cortex with periosteal elevation is observed (**f**) H & E
219 stain of **e** shows a monomorphic population of small round blue cells with hyperchromatic and
220 indistinct nuclei. Cells were positive for CD99 and S100. Cells were negative for CD3, CD20, CD45,
221 CKMNF116, EMA, AE1/3, SMA, desmin and FLI1. *EWSR1* gene rearrangement observed in 56/72
222 nuclei. Examination is consistent with Ewing sarcoma (**g**) high grade sarcoma was diagnosed at
223 biopsy. On resection, a pale cream/grey and destructive lesion with permeation into underlying
224 bone is observed (**h**) H & E stain of **g** shows an infiltrative pattern of growth within the bone and
225 extraosseous extension. Tumour cells were negative for CD30, CD31, CD34, CD45, CD99,
226 CKMNF116, EMA, AE1/3, SMA, desmin, S100, HMB45, STAT6, TLE1, Melan-A, CCNB3 and
227 ALK1. There is no *EWSR1-NR4A3* gene rearrangement. Examination is consistent with spindle
228 cell sarcoma of bone (**i**) heat map based hierarchical cluster analysis of differentially expressed
229 non-coding RNAs (y-axis) between bone cancer, OS1, OS2, ES, SCS and controls (x-axis). Z
230 score refers to high (in red) and low (in blue) non-coding RNA expression using normalised values
231 when compared to the mean of total sequencing reads (**j**) a biplot principle component analysis
232 shows two distinct clusters along the PC1 axis that correspond to the bone cancer samples (in
233 orange) and control samples (in blue).

234

235

236

237

238 **Table 1.** The 10 most upregulated and 10 most downregulated lncRNA in our bone cancer cohort
 239 when compared to control samples. We also show 4 dysregulated snoRNAs. Differentially
 240 expressed lncRNA and snoRNA were identified according to log₂ fold change ≥2, *p* value <0.05
 241 and FDR <5%. The full data set is publically available on GEO under accession GSE113916.
 242

Transcript ID	Human Genome Annotation	Type	Dysregulation (up / down)	Log ₂ Fold Change
ENST00000617687.1	<i>AC244100.2-202</i>	antisense lncRNA	up	13.22
ENST00000414790.6	<i>H19-203</i>	intronic lncRNA	up	12.9
ENST00000535913.2	<i>SLC12A5-AS1</i>	antisense lncRNA	up	12.69
ENST00000554639.5	<i>MEG3-224</i>	intergenic lncRNA	up	12.45
ENST00000429829.5	<i>XIST-204</i>	intergenic lncRNA	up	11.28
ENST00000534150.5	<i>AP000757.1-201</i>	antisense lncRNA	up	10.97
ENST00000541196.2	<i>HCP5-204</i>	sense lncRNA	up	10.7
ENST00000404665.3	<i>TMEM51-AS1</i>	antisense lncRNA	up	10.5
ENST00000416330.1	<i>XIST-201</i>	intergenic lncRNA	up	10.43
ENST00000618234.4	<i>AL034397.3-201</i>	antisense lncRNA	up	10.4
ENST00000618925.1	<i>MALAT1-214</i>	intergenic lncRNA	down	10.85
ENST00000510859.5	<i>PAX8-AS1-209</i>	antisense lncRNA	down	9.94
ENST00000587245.2	<i>PRMT5-AS1-203</i>	antisense lncRNA	down	9.88
ENST00000544868.2	<i>MALAT1-203</i>	intergenic lncRNA	down	9.5
ENST00000637700.1	<i>MIR100HG-224</i>	intergenic lncRNA	down	9.35
ENST00000468186.5	<i>THUMPD3-AS1-202</i>	antisense lncRNA	down	9.12
ENST00000412059.5	<i>GAS5-201</i>	intronic lncRNA	down	7.44
ENST00000620594.1	<i>ZFAS1-209</i>	antisense lncRNA	down	7.2
ENST00000512932.5	<i>THAP9-AS1-211</i>	antisense lncRNA	down	6.53
ENST00000534782.3	<i>MIR100HG-222</i>	intergenic lncRNA	down	6.5
ENST00000363214.1	<i>SNORD68</i>	C/D box snoRNA	up	4.02

ENST00000577988.2	<i>SNORD3B-1</i>	C/D box snoRNA	up	2.86
ENST00000571722.3	<i>SNORD3B-2</i>	C/D box snoRNA	up	2.68
ENST00000607313.1	<i>SNORD58B</i>	C/D box snoRNA	down	2.28
