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

Darrell Green[1], Archana Singh[2], Jasmine Sanghera[3], Lee Jeys[4], Vaiyapuri Sumathi[5], Tamas Dalmay[3], William D Fraser[1, 6]*

1. Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom
2. John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom
3. School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom
4. Department of Orthopaedic Oncology, The Royal Orthopaedic Hospital, Birmingham, B31 2AP, United Kingdom
5. Department of Musculoskeletal Pathology, The Royal Orthopaedic Hospital, Birmingham, B31 2AP, United Kingdom
6. Department of Clinical Biochemistry, Norfolk and Norwich University Hospital, Norwich Research Park, Norwich, NR4 7UY, United Kingdom

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

Conflict of interest: No conflict to declare

Word count of main body of text: 1,496

Sir:

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

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

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

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

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

Raw fastq files were converted to fasta format. Adapter sequences and reads <20 nt were trimmed using Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore). Trimmed reads were aligned to the human genome (v.38) using HISAT2.[3] The latest set of human non-coding RNAs were download from GENCODE (v.28) and Ensembl (v.92). Count matrix for lncRNAs and snoRNAs was created using Kallisto.[4] LncRNA and snoRNA expression was compared between bone cancer and controls using the DESeq2 package available in R. We selected differentially expressed lncRNA and snoRNA according to $\log_2$ fold change ≥2, *p* value <0.05 and false discovery rate <5%. Hierarchical cluster analysis and principle component analysis investigating 77% of variance shows distinct grouping between cancer and controls (Figure 1I&J).

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

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

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

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

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

DATA AVAILABILITY

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

FUNDING

This study was funded by The Humane Research Trust.

ACKNOWLEDGEMENTS

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

REFERENCES

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

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

FIGURE 1

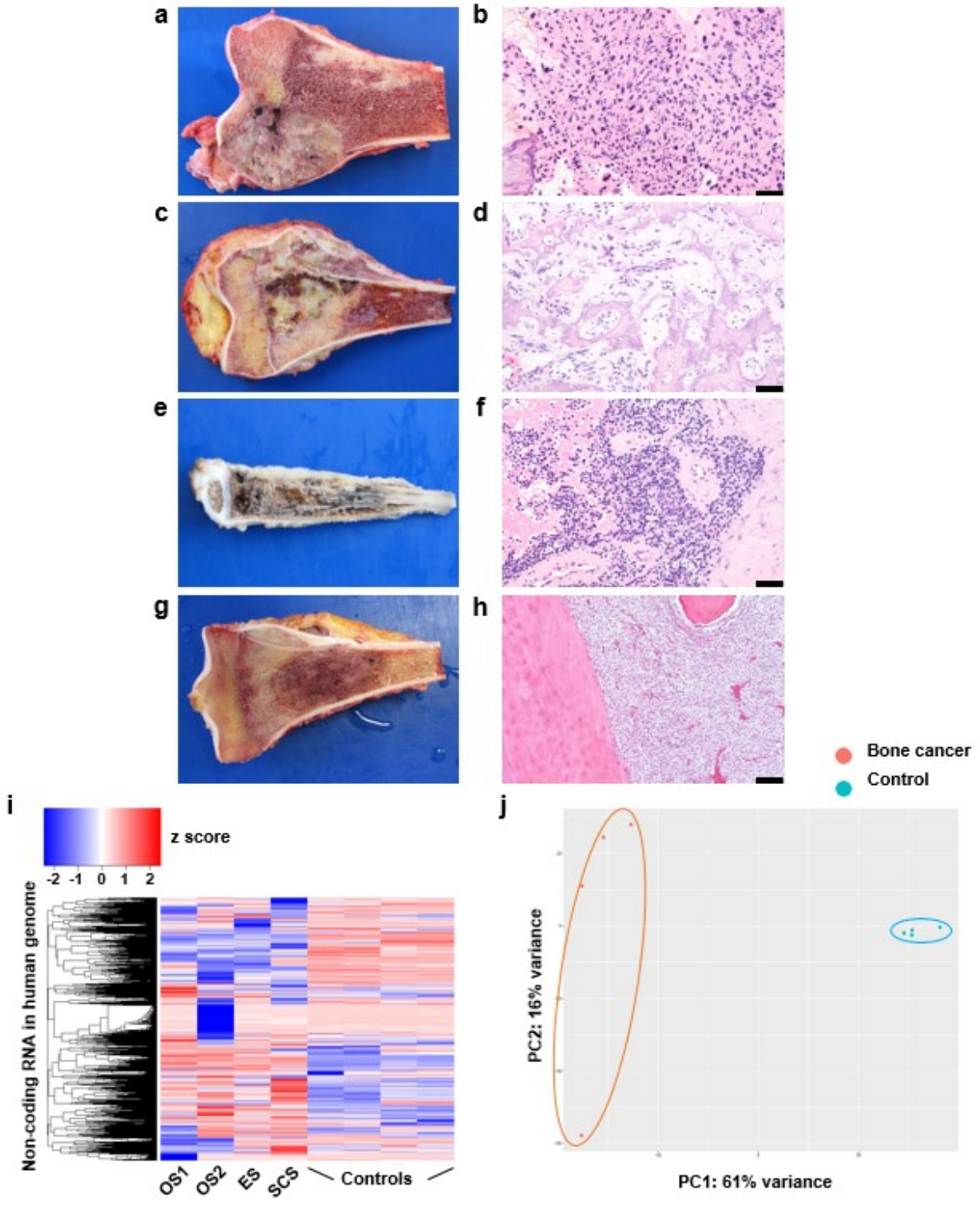

Figure 1. Macroscopic and microscopic photographs of resected tumours used in this study. In the microscopic images, the black scale bar is 100 μM (**a**) osteoblastic osteosarcoma was diagnosed at biopsy. Patient underwent treatment with cisplatin and doxorubicin. On resection, a

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

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 $\log_2$ 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	*AC244100.2-202*	antisense lncRNA	up	13.22
ENST00000414790.6	*H19-203*	intronic lncRNA	up	12.9
ENST00000535913.2	*SLC12A5-AS1*	antisense lncRNA	up	12.69
ENST00000554639.5	*MEG3-224*	intergenic lncRNA	up	12.45
ENST00000429829.5	*XIST-204*	intergenic lncRNA	up	11.28
ENST00000534150.5	*AP000757.1-201*	antisense lncRNA	up	10.97
ENST00000541196.2	*HCP5-204*	sense lncRNA	up	10.7
ENST00000404665.3	*TMEM51-AS1*	antisense lncRNA	up	10.5
ENST00000416330.1	*XIST-201*	intergenic lncRNA	up	10.43
ENST00000618234.4	*AL034397.3-201*	antisense lncRNA	up	10.4
ENST00000618925.1	*MALAT1-214*	intergenic lncRNA	down	10.85
ENST00000510859.5	*PAX8-AS1-209*	antisense lncRNA	down	9.94
ENST00000587245.2	*PRMT5-AS1-203*	antisense lncRNA	down	9.88
ENST00000544868.2	*MALAT1-203*	intergenic lncRNA	down	9.5
ENST00000637700.1	*MIR100HG-224*	intergenic lncRNA	down	9.35
ENST00000468186.5	*THUMPD3-AS1-202*	antisense lncRNA	down	9.12
ENST00000412059.5	*GAS5-201*	intronic lncRNA	down	7.44
ENST00000620594.1	*ZFAS1-209*	antisense lncRNA	down	7.2
ENST00000512932.5	*THAP9-AS1-211*	antisense lncRNA	down	6.53
ENST00000534782.3	*MIR100HG-222*	intergenic lncRNA	down	6.5
ENST00000363214.1	*SNORD68*	C/D box snoRNA	up	4.02

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