

Supplement to: “CpG islands’ clustering uncovers early development genes in human genome”

Vladimir N. Babenko^{1*}, Anton G. Bogomolov¹, Roman O. Babenko², Elvira R. Galieva²
and Yuriy L. Orlov^{1,2}

¹ Institute of Cytology and Genetics SB RAS, Lavrentyeva 10, 630090 Novosibirsk, Russia

bob@bionet.nsc.ru

² Novosibirsk State University, Pirogova 2, 630090 Novosibirsk, Russia

1. Megasatellites comprising CGIs as a monomer subunit.

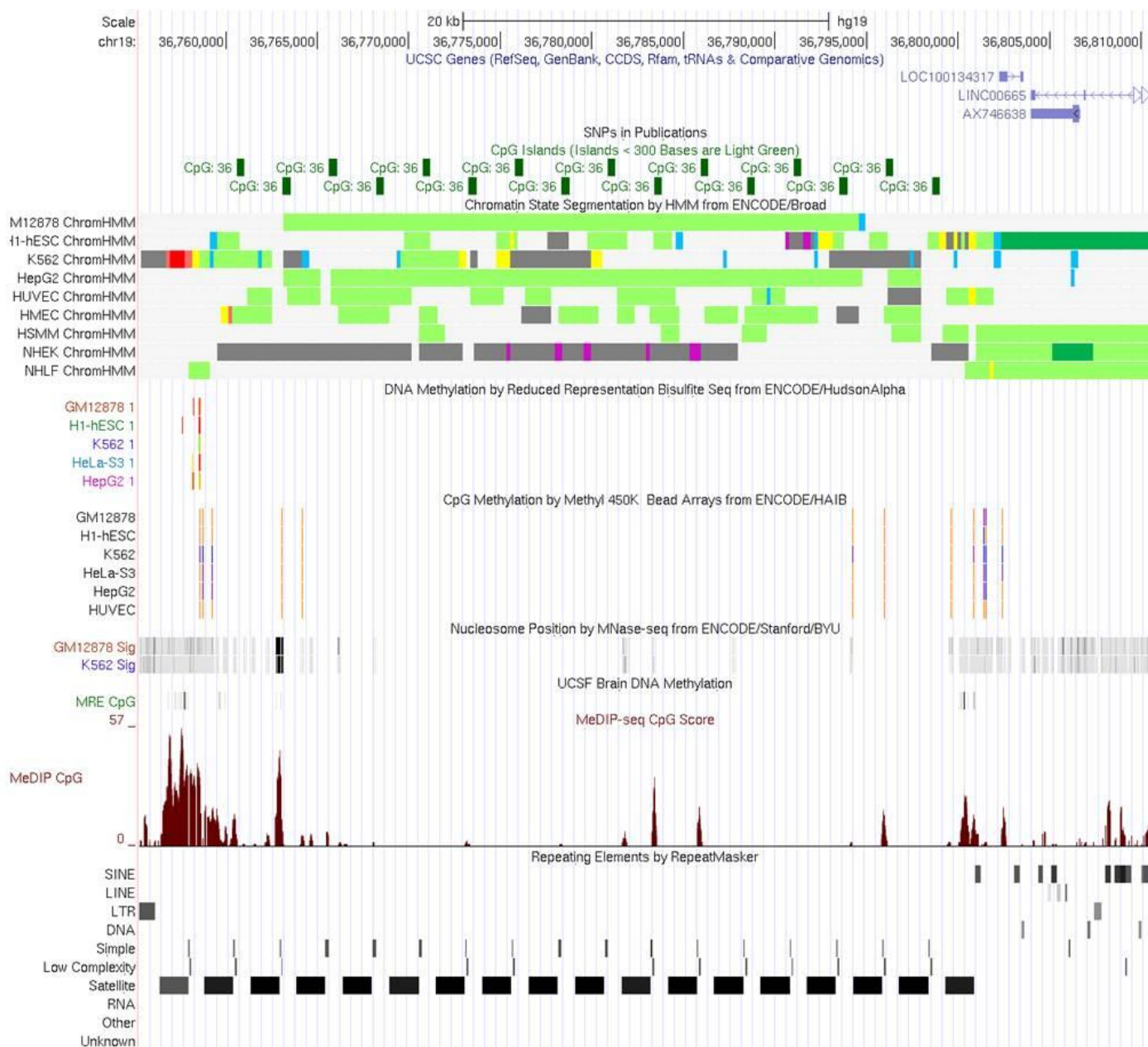


Fig. S1. Tandem structure of monomers comprising (SST1_SR_CGI) representing 19SST11 macrosatellite. (chr19:36,755,278-36,810,578). Intersatellite (19SST11-19SST12) region comprises 17 KRAB-ZNF genes with SST1_SR_CGI monomers performing as promoters (Table 4; Fig.1 in main text).

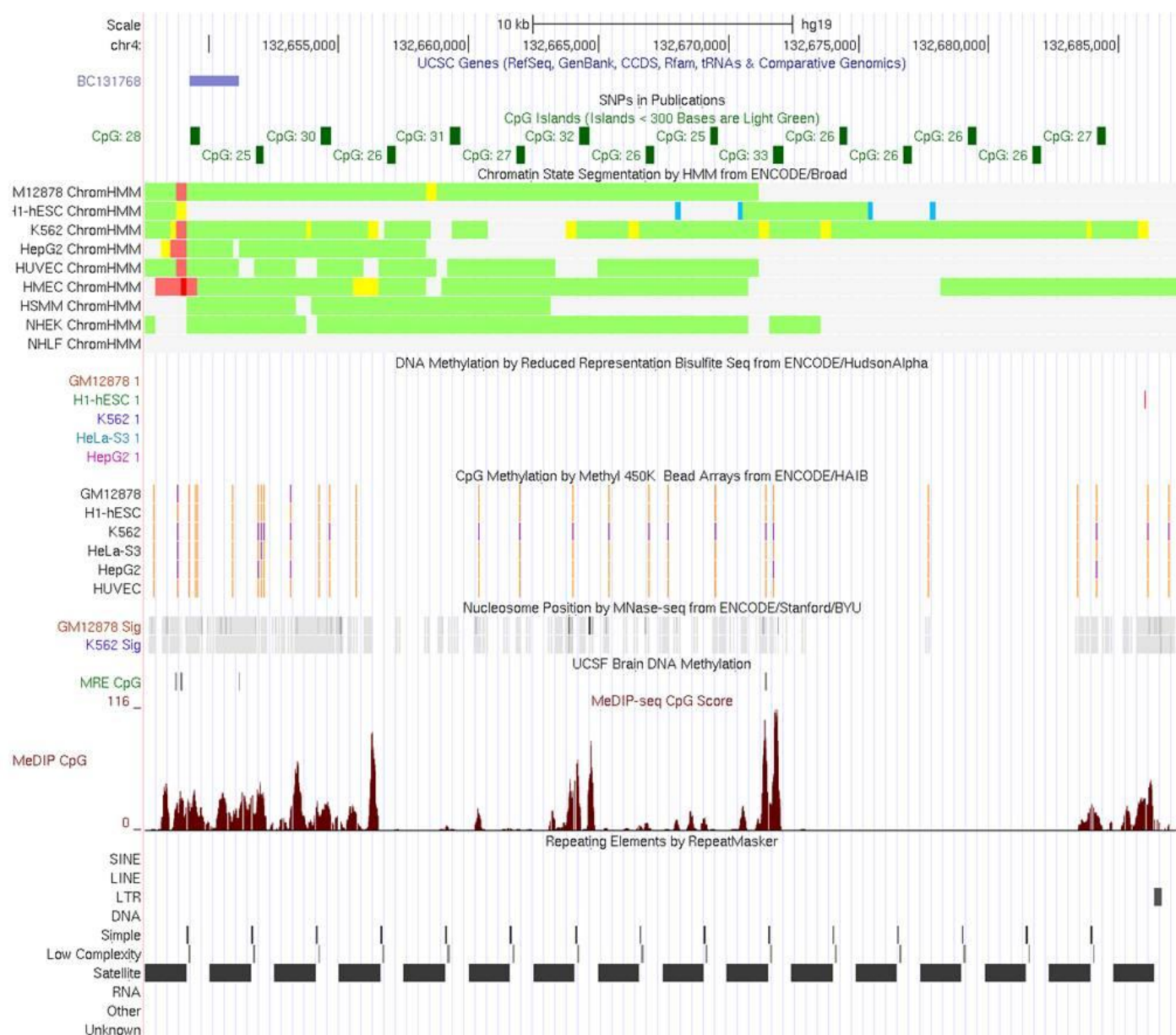


Fig. S2. Tandem structure of monomers comprising (SST1_SSR_CGI) representing 4SST11 mcarosatellite maintaining high identity to the 19SST11, 19SST12 megasatellites (Fig. S1).

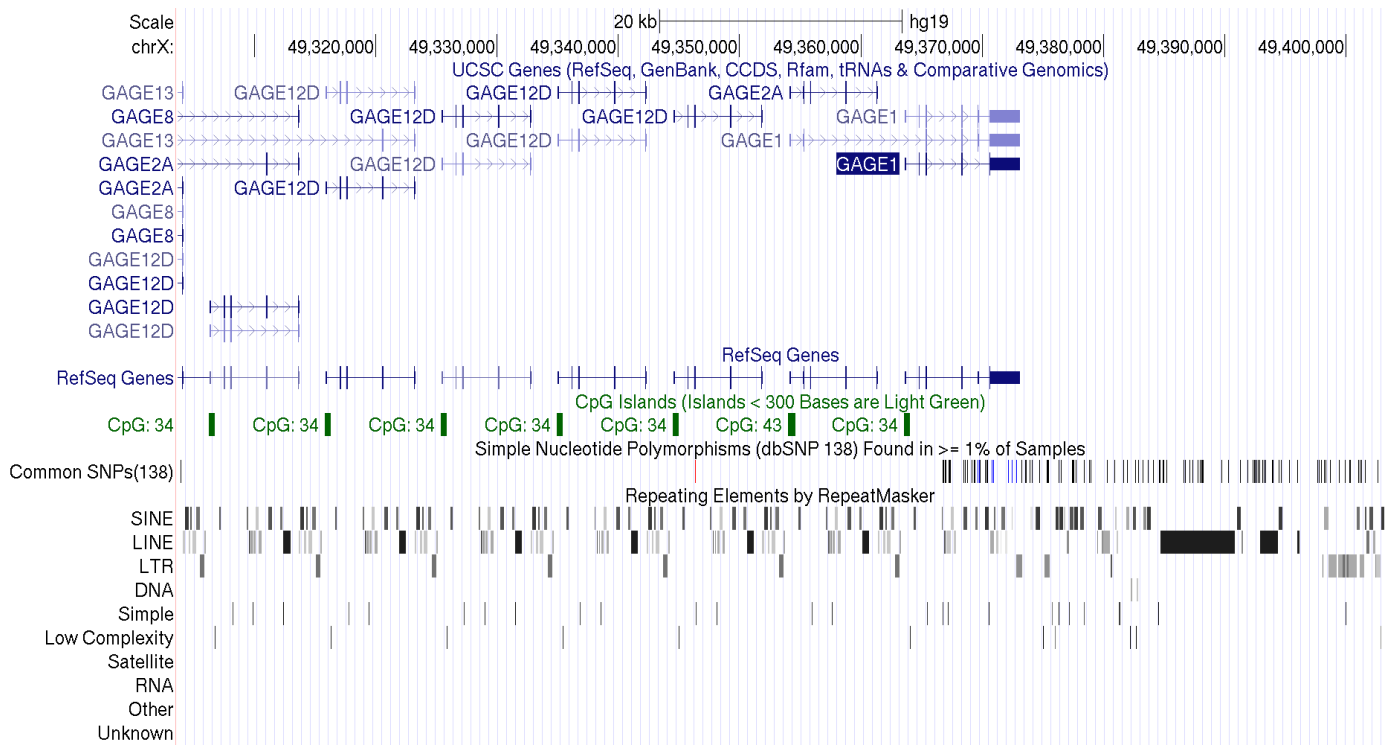


Figure S3. *Gage* genes cluster with tandem repeated CGIs as promoters at chrX (Warburton et al., 2008).

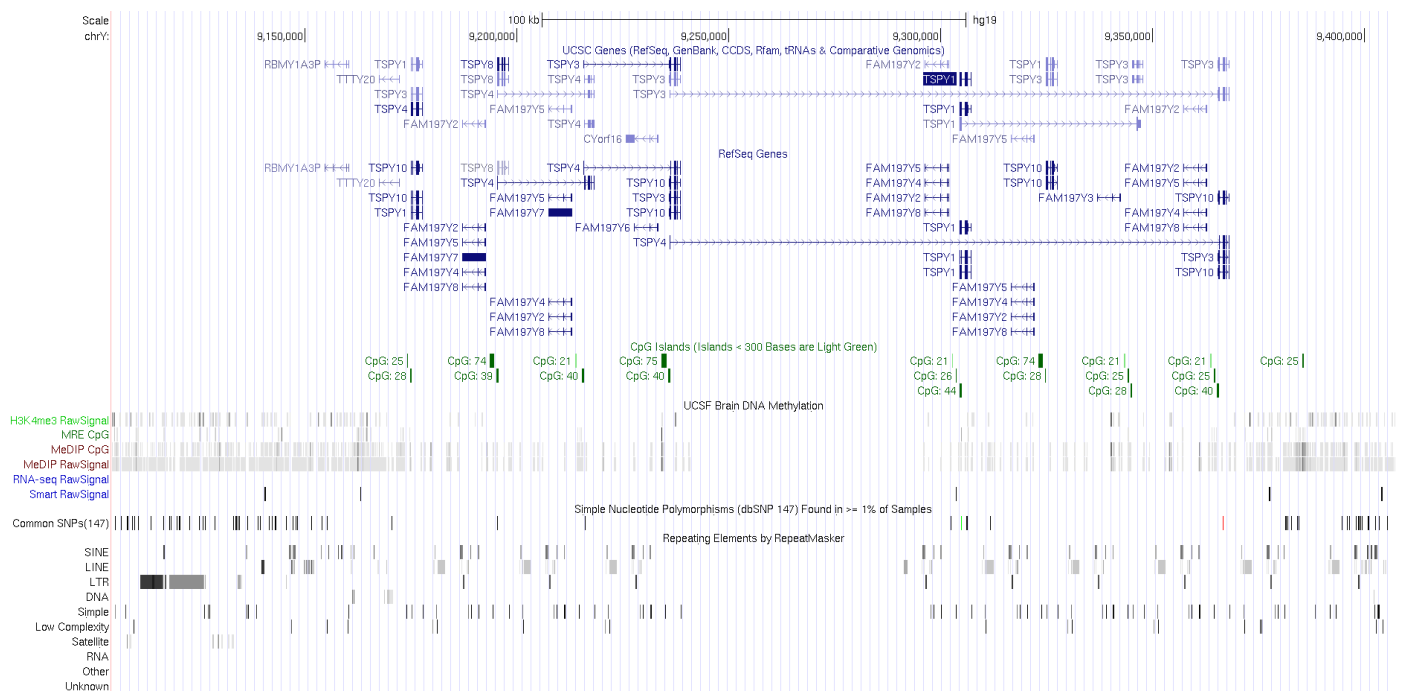


Figure S4. *TSPY* (testis specific protein on chromosome Y) genes cluster with tandem repeated CGIs bidirectional promoter segments (Warburton et al., 2008).

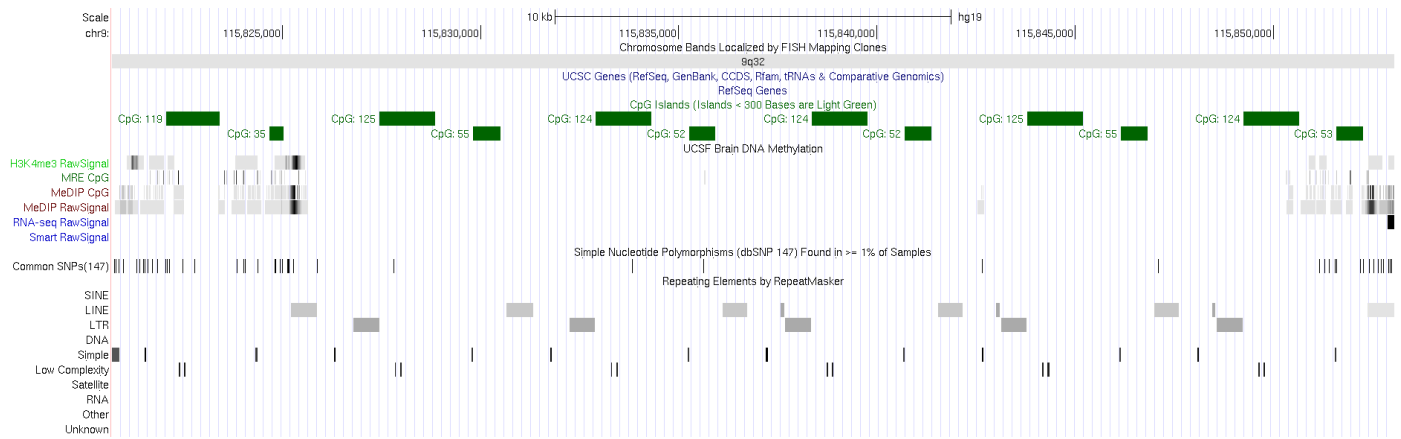


Figure S6. Non – coding satellite with 4 CGIs per 10kb on chr9 (L1MA7; Warburton et al., 2008).

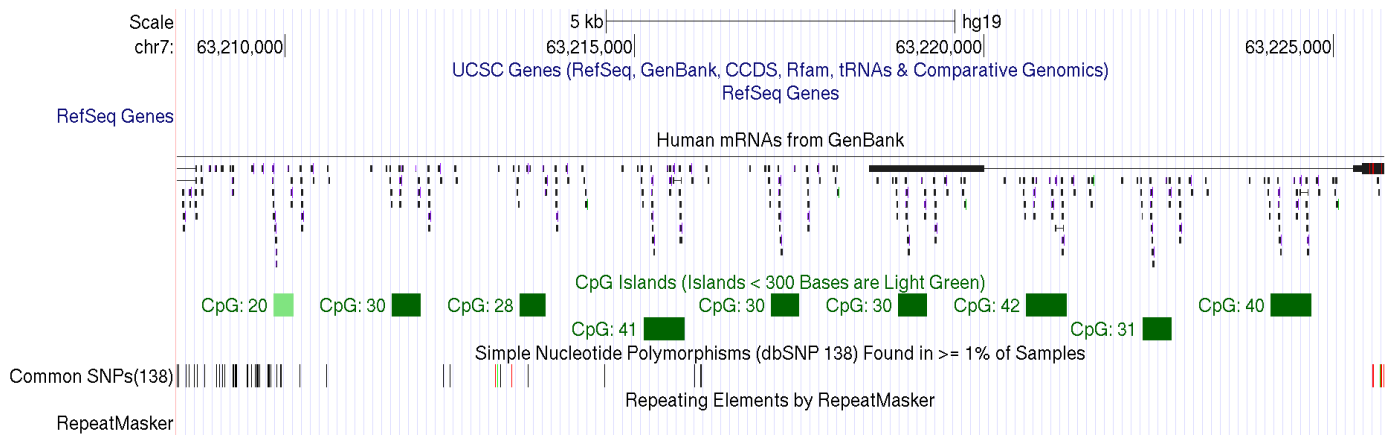


Fig. S7. 9 CGIs at chromosome 7 spanning 12.3kb of 1.8kb monomers (8) including CGIs (VNTRs; Warbuton et al., 2008).

2. Tandem organised merged CGIs more than 10kb.

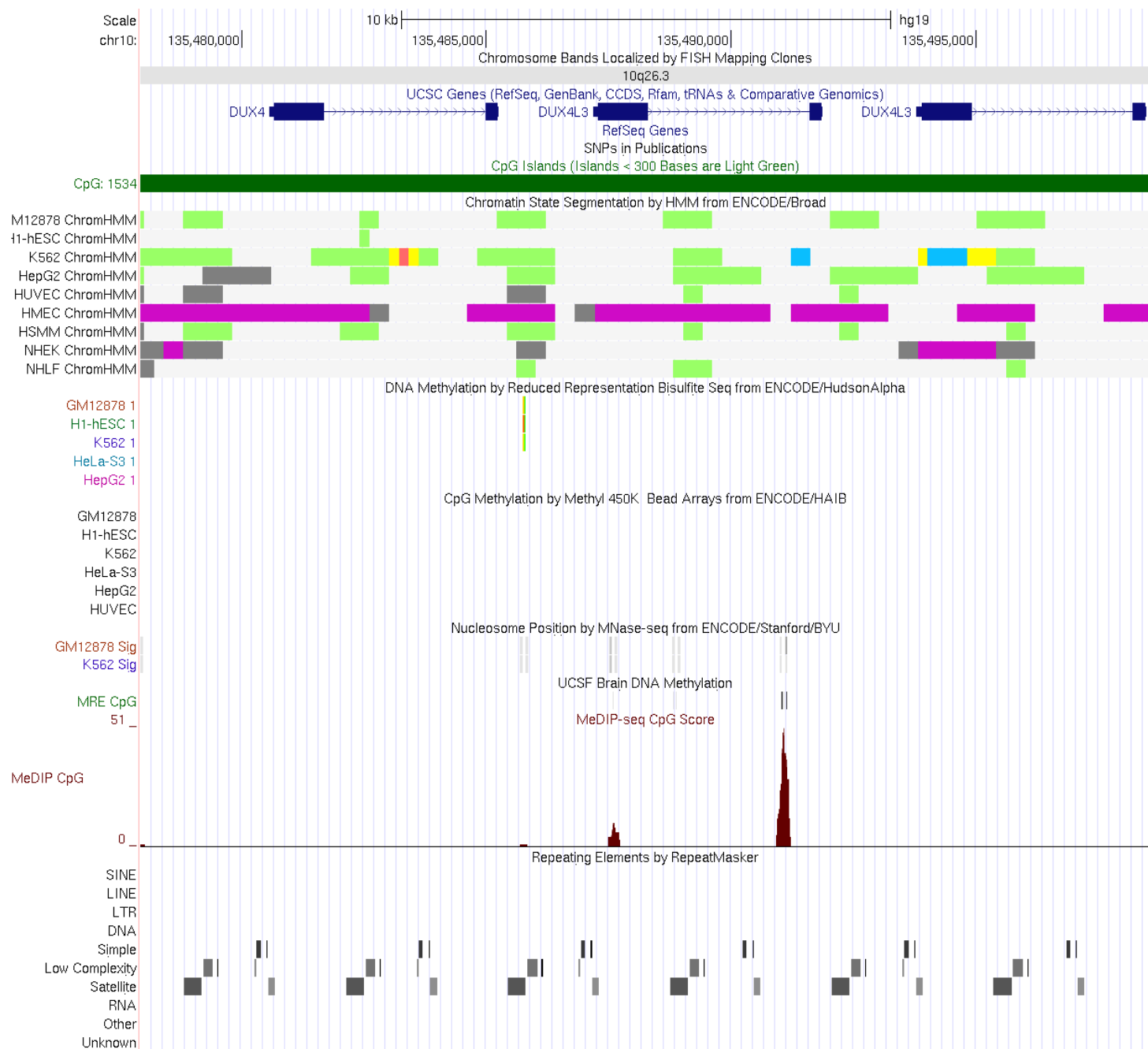


Fig. S8. DUX gene cluster on Chr10, pos 135Mb, telomere.



Fig. S9. Chr 1. miRNA cluster (SINE/simple repeats; 40kb).



Figure S10. Chr. 1. tRNA cluster (LINE, LTR, Simple, tRNA elements RMSK annotated repeats as a monomer)..

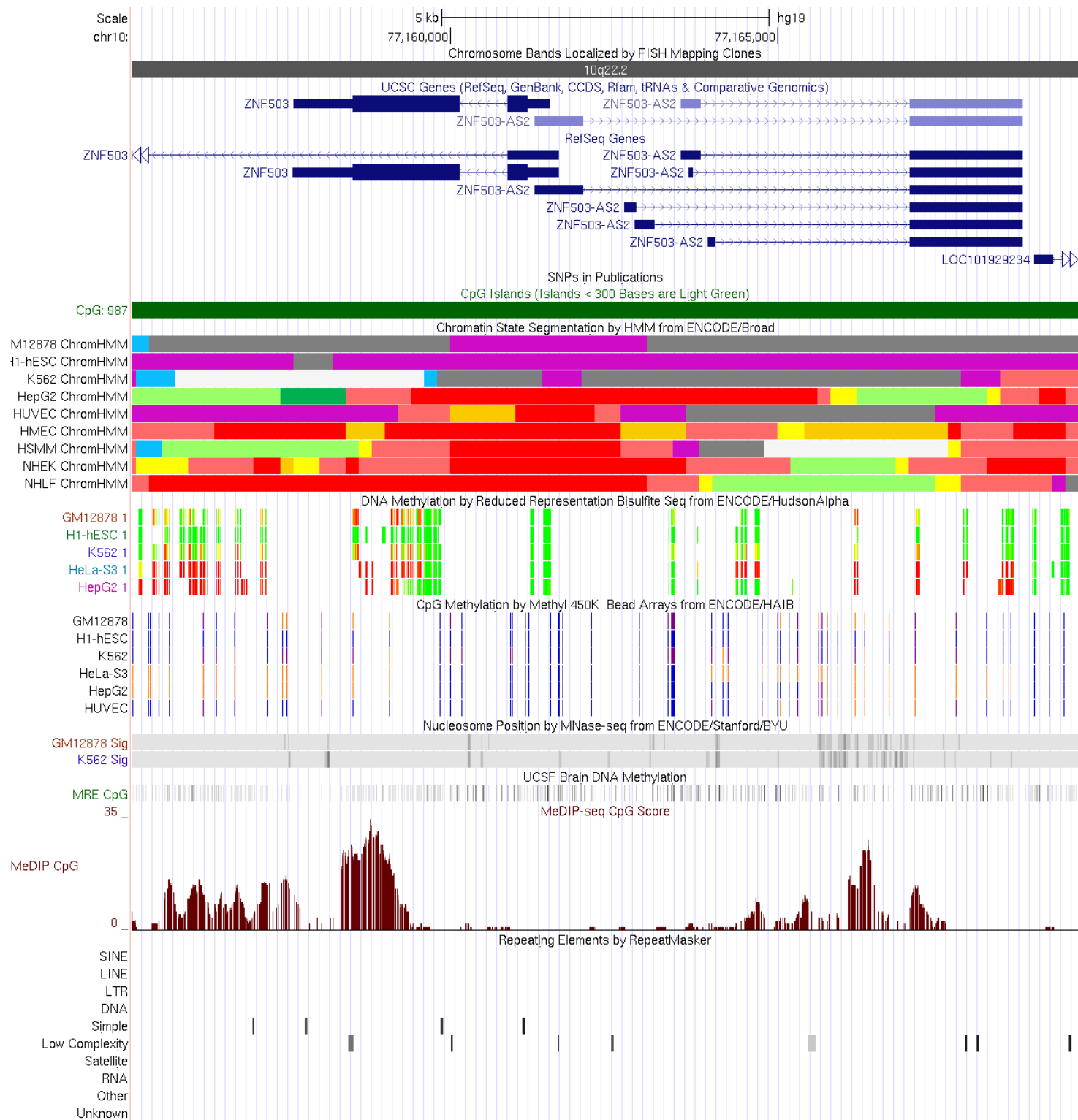


Figure S11. Chr. 10 ZNF and ZNF-AS processing (non-tandem). Heavy methylation of promoter segment in adult brain implies early embryonic development function of this gene.

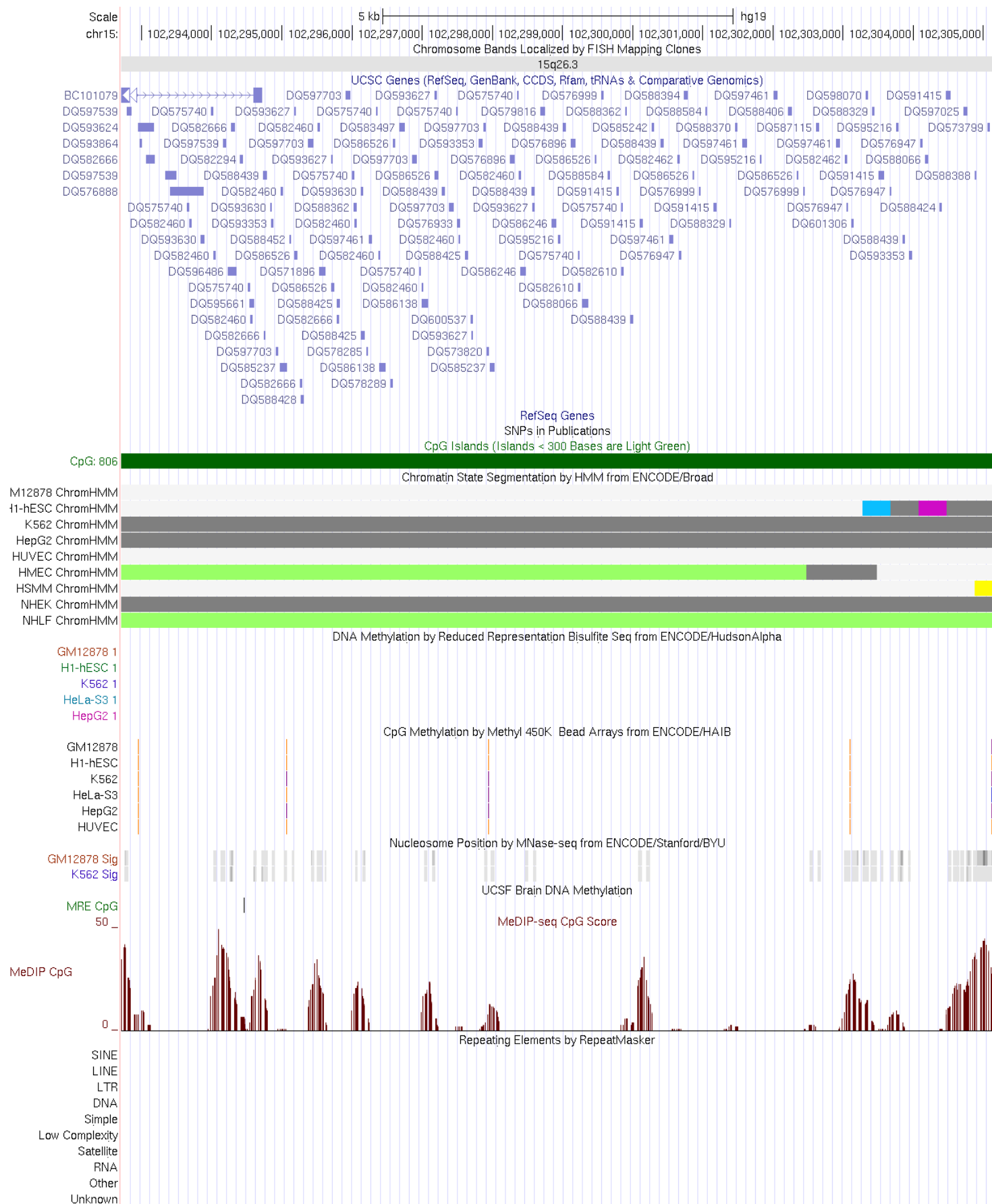


Figure S12. Chr15. piRNA cluster (non-tandem). Heavy methylation in adult brain implies early embryogenesis function of the cluster.

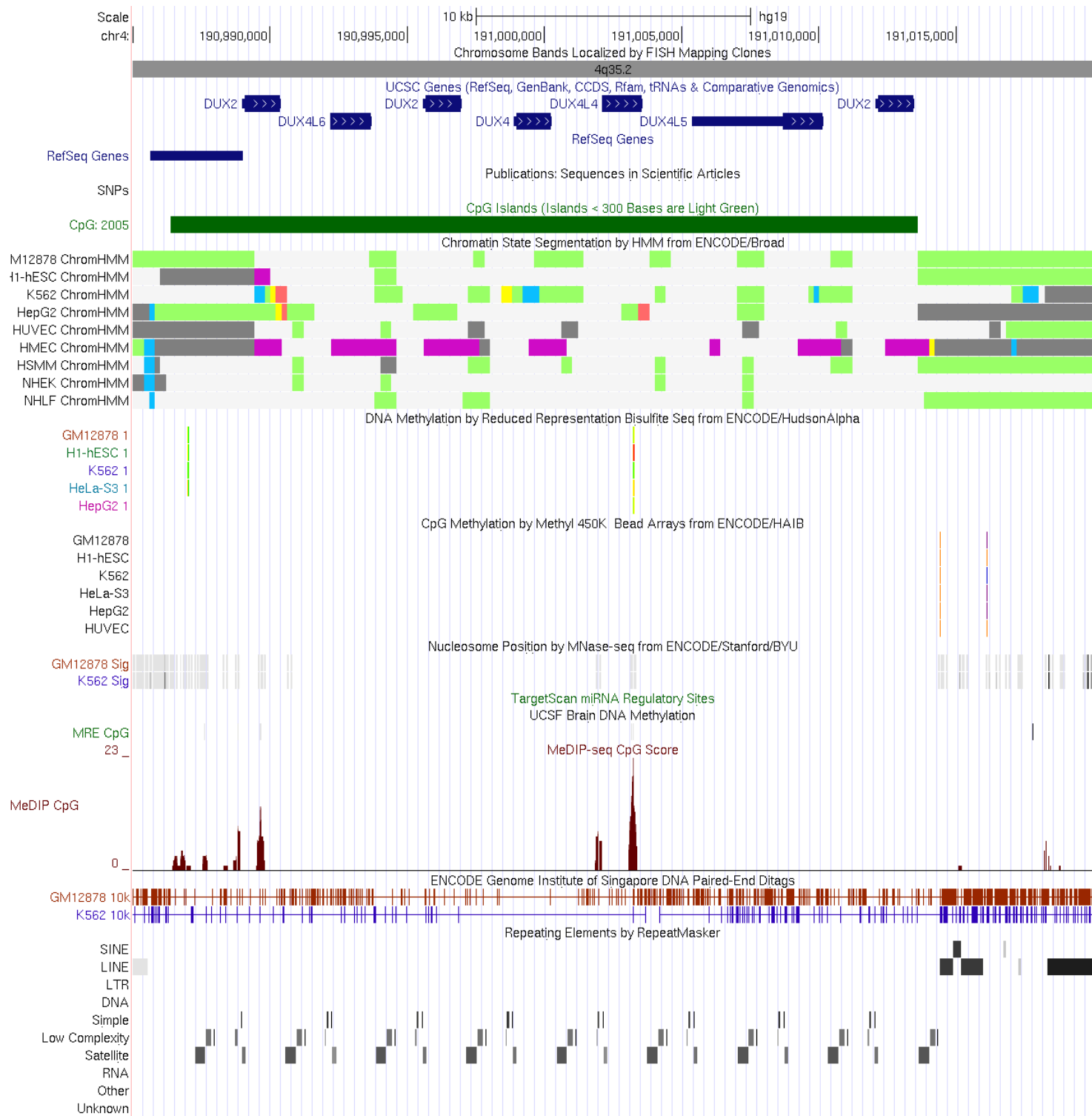


Figure S13. Chr.4. D4Z4 cluster (chr4:190,985,000-191,020,000) comprising DUX homeobox genes.

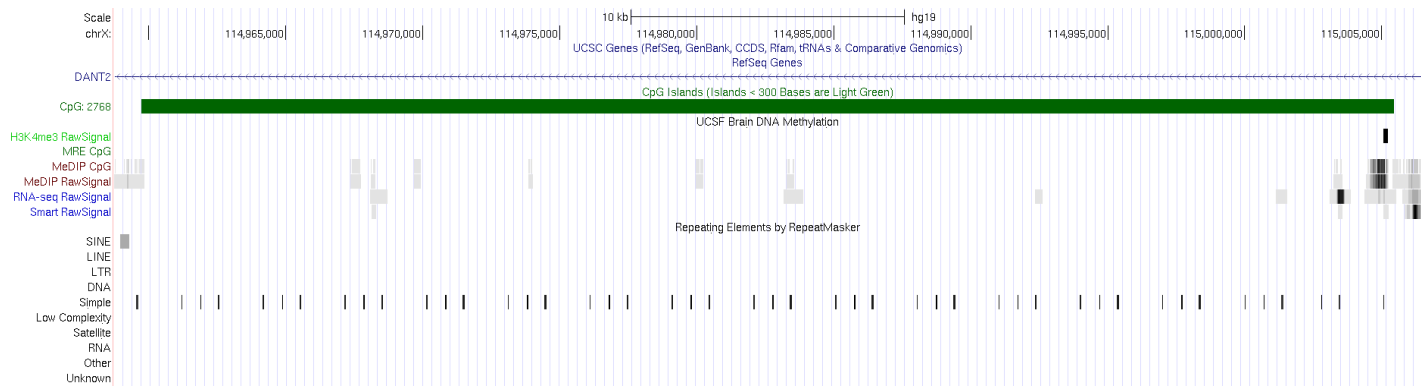


Figure S14. DXZ4 satellite within intron of DXZ4 associated non-coding transcript 2, distal (DANT2), transcript variant 2, long non-coding RNA.

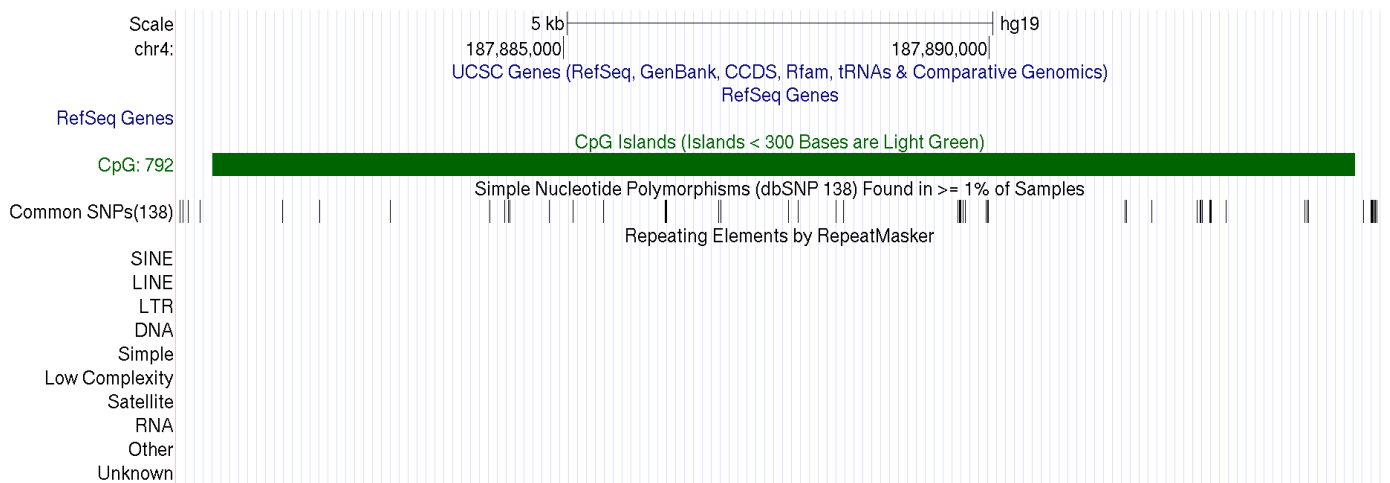


Figure S15. Tandem structured CGI on chromosome 4 comprising 13.4 kb of 59bp VNTR monomers (HOR structure; Warburton et al., 2008).

3 2 Tandem clusters not attributed to CGIs but early replication.



Figure S16. chr19 C19MC miRNA cluster (note hESC transcription and abundant methylation afterwards. It is reported this miRNA cluster works as maternal imprinted locus in placenta). It is early replicated loci on chromosome 19 and is flanked by ctcf insulator sites.

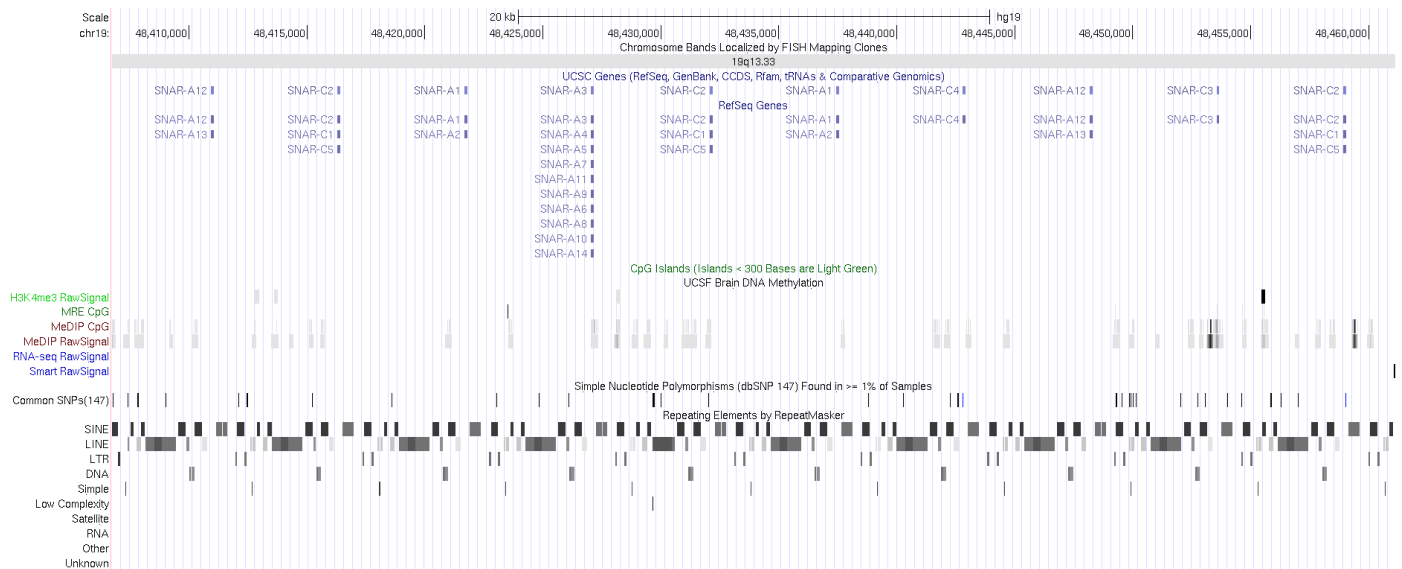


Figure S17. SNAR* non-coding RNA cluster emerged by monomers of SINE, LINE and Mer33 fusion (Mer33; chr19; Warburton et al., 2008). It is early replicated region of chr19 as well.

References

1. Darrow, E.M., Chadwick, B.P.: A novel tRNA variable number tandem repeat at human chromosome 1q23.3 is implicated as a boundary element based on conservation of a CTCF motif in mouse. *Nucleic Acids Res.*, Vol. 42, No. 10, 6421-35. (2014)
2. Das, S., Chadwick, B.P.: Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS One*, Vol. 11, No. 7, :e0160022. (2016)
3. Horakova A.H., Moseley S.C., McLaughlin C.R., Tremblay D.C., Chadwick B.P. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum. Mol. Genet.* 2012;21(20):4367-4377.
4. Warburton PE, Hasson D, Guillem F, Lescale C, Jin X, Abrusan G. Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics.* 2008 Nov 7;9:533. doi: 10.1186/1471-2164-9-533.