HOTAIR ancient sequence suggests regulatory roles both in *cis* and *trans*

Chirag Nepal¹, Yavor Hadzhiev², Sachin Pundhir¹, Piotr Mydel^{3,4}, Boris Lenhard⁵, Ferenc Müller², Jesper B Andersen¹

- 1. Biotech Research and Innovation Centre, Department of Health and Medical Sciences, University of Copenhagen, Ole Maaløes Vej 5, DK-2200, Copenhagen N, Denmark
- 2. Institute of Cancer and Genomics Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, B15 2TT, Birmingham, United Kingdom
- 3. Broegelmann Research Laboratory, Department of Clinical Science, University of Bergen, Bergen, Norway
- 4. Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, and Malopolska Centre of Biotechnology, Jagiellonian University, Karkow, Poland
- Institute of Clinical Sciences MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0NN, United Kingdom

Correspondence to: Chirag Nepal chirag.nepal@bric.ku.dk Biotech Research and Innovation Centre University of Copenhagen Ole Maaløes Vej 5, DK-2200 Copenhagen N Denmark

Keywords: HOTAIR, IncRNA, conservation, CNE, sequence complementarity, enhancer

ABSTRACT

HOTAIR is a long noncoding RNA transcribed between HOXC11 and HOXC12 in mammals. The proposed function(s) of HOTAIR lacks consensus as to whether it regulates HoxD cluster genes in *trans* or HoxC cluster genes in *cis*. We have identified a 32-nucleotide long conserved noncoding element (CNE) as HOTAIR ancient sequence which has a paralogous copy embedded in HOXD11 noncoding transcript. All vertebrates except teleosts have two copies of CNE and the paralogous CNEs exhibit sequence complementarity in the transcribed orientation. Moreover, paralogous CNEs underwent compensatory mutations suggesting they co-evolved and might hybridize. In both human and mouse, HOTAIR CNE exhibits characteristic features of a poised enhancer in HOTAIR-unexpressed stem cells and of an active enhancer in HOTAIR expressed cells. Tight correlation between the transcriptional activity of the CNE and HOTAIR promoter suggests HOTAIR transcription is crucial for enhancer activity. In HOTAIR-expressed cells, HOTAIR expression is positively correlated with HOXC11 in *cis* and negatively correlated with HOXD11 in *trans*, suggesting a dual modality of HOTAIR ancient sequence.

INTRODUCTION

Mammalian genomes are pervasively transcribed, giving rise to thousands of long noncoding RNAs (IncRNAs) (1,2). Only a handful of IncRNAs have well-characterized functions, which are attained through diverse mechanisms (chromatin regulation, alternative splicing, gene silencing, cis-regulation, trans-regulation) (3,4). Among these, HOTAIR is a prototypic model for IncRNAs regulating chromatin modifications. HOTAIR is an intergenic (between HOXC11 and HOXC12) IncRNA proposed to regulate HOXD cluster genes (i.e., HOXD8, HOXD9, HOXD10 and HOXD11) in trans by recruiting Polycomb Repressive Complex 2 (PRC2) (5). The proposed model has been questioned as PRC2 binding is promiscuous (6) and PRC2 is dispensable for HOTAIR-mediated transcriptional repression (7). Deletion of the entire mouse Hoxc cluster (including Hotair) showed little effect on gene expression and H3K27me3 levels at Hoxd genes (8). Specific deletion of Hotair produced a phenotype of homeotic transformation, skeletal malformation, global decrease in H3K27me3 levels and upregulation of posterior HoxD genes (i.e., Hoxd10, Hoxd11 and Hoxd13) (9). These observations are now challenged as specific knockouts of Hotair locus in vivo show neither homeotic transformation nor upregulation of HoxD genes but, instead, a significant change in Hoxc (especially Hoxc11 and Hoxc12) cluster genes, which argue in favor of a DNA-dependent effect of the Hotair deletion (10). The observed different regulatory mechanisms (cis versus trans) might be due to different tissues and developmental stages from different genetic backgrounds (11) and hence there is no consensus unifying model of HOTAIR regulation (12).

Some IncRNAs have enhancer-like functions (13), although it is unclear whether IncRNAs function through IncRNA transcripts or through the underlying genomic DNA with *cis*-regulatory functions. Genomic deletion of IncRNA also removes *cis*-regulatory DNA elements, thus confounding whether the observed phenotype is due to the act of transcription of a IncRNA transcript or underlying genomic DNA (14). A systematic analysis involving transcription blockage and perturbation of the *Lockd* IncRNA sequence showed that it regulates *Cdkn1b* transcription through an enhancer element, while the IncRNA transcript is dispensable for *Cdkn1b* expression (15). Deletion of 12 genomic loci encoding IncRNAs revealed five loci where the deletion showed effects on the general process of transcription and enhancer-like activity but no requirement for the IncRNA products themselves (16). Identification of accurate 5' start sites of IncRNAs revealed that a large majority of IncRNA have enhancer-like features (1). Others like *Haunt* IncRNA have dual roles where *Haunt* DNA encodes potential enhancers to activate HoxA genes and *Haunt* RNA prevents aberrant HoxA expression (17).

Here, we set to address whether *HOTAIR* regulates HOXC cluster genes in *cis* (10) or HOXD cluster genes in *trans* (5,9). HOXC and HOXD clusters are duplicated from the ancestral HOXC/D cluster, by the second round of whole genome duplication (WGD), thus we asked whether *HOTAIR* and HOXD cluster have retained some sequence from ancestral HOXC/D

cluster. We have identified a 32-nucleotide conserved noncoding element (CNE) as the *HOTAIR* ancestral sequence that is present across all jawed vertebrates (except teleost) and has a paralogous copy embedded within *HOXD11* noncoding transcript. Paralogous CNEs underwent compensatory mutations and exhibit sequence complementarity. *HOTAIR* CNE represents an active or poised enhancer in different cellular context. *HOTAIR* expression is positively correlated with *HOXC11* and negatively correlated with *HOXD11*, suggesting dual modality of *HOTAIR* CNE.

RESULTS

Identification of HOTAIR ancient sequence and its paralog in HoxD cluster

Hox clusters have multiple conserved noncoding elements (CNEs) that are highly conserved from human to fish (18,19); thus, we asked whether any region of *HOTAIR* (which resides on HOXC cluster) is conserved across vertebrates. We used human and zebrafish CNEs from ANCORA (19) (see Materials and Methods) and identified a 32-nucleotide long CNE that overlapped with *HOTAIR* intronic region (Fig. 1A), which is eight nucleotides away from the splice site and overlaps a CpG island (CGI). Identification of the homolog of CNE in zebrafish is intriguing because *HOTAIR* was proposed to have a *de novo* origin in marsupials (20) and a homolog(s) of *HOTAIR* RNA have not been reported in zebrafish (21-23). Thus, we mapped the orthogonal position of the CNE in zebrafish and identified its location between *hoxd11a* and *hoxd12a* (Fig. 1B), but not in the expected hoxc clusters (Fig. 1A).

As this CNE is conserved across all vertebrates (Supplementary Fig. S1A-B), we investigated whether the CNE is located in the HoxC (the capitalization "HoxC" is used to represent the HoxC cluster across multiple species) or HoxD cluster. We systematically mapped the CNE sequences across 37 organisms (34 vertebrates and 3 invertebrates) (Supplementary Table S1) and identified a homologous CNE in HoxD and HoxC clusters common to all jawed vertebrates except in teleosts and birds, but absent in lamprey (jawless vertebrate) and invertebrates (Fig. 1C). The homologous CNEs mapped between *HoxD11* and *HoxD12* in HoxD cluster and between *HoxC11* and *HoxC12* in HoxC cluster (Fig. 1C). The absence of HoxC CNE in birds is likely due to unassembled HoxC cluster (Supplementary Table S2). In contrast, teleosts have well-annotated *HoxC11* and *HoxC12* genes in the same cluster (Supplementary Table S2) but underwent an additional round of teleost-specific whole genome duplication (WGD) that might have resulted in lineage-specific loss. Thus, we conclude that all jawed vertebrates except teleosts have a homolog(s) of *HOTAIR* CNE in the HoxC and HoxD clusters.



Figure 1 Identification of the *HOTAIR* conserved noncoding element (CNE) and its homolog in HOXD cluster across vertebrates. (**A**) A genome browser view around *HOTAIR* locus showing CNE from ANCORA browser and UCSC PhyloP conservation track. The CNE highlighted in a rectangular box is located eight nucleotides away from the splice site. (**B**) Orthologs of *HOTAIR* CNE mapped to the zebrafish hoxd (between *hoxd11* and *hoxd12*) cluster. (**C**) Homology search of the CNE across 37 species identified homologous CNEs in only HoxC and HoxD clusters. Homologs of the CNE are undetected in jawless vertebrate and invertebrates. Homologs in HoxC and HoxD clusters are represented by blue and red, respectively. Empty boxes indicate absence of homologs. (**D**) Schematic representation for the proposed model on the origin of the CNE. The CNE might have *de novo* origin in ancestral HoxC/D cluster after the first round of whole genome duplication (1R-WGD), where the second round of whole genome duplication (2R-WGD) resulted two copies of the CNE at HoxC and HoxD clusters. Additional round of teleosts specific duplication might have resulted in loss of CNE from both HoxC clusters and from one of the HoxD cluster.

Two rounds of WGD at the root of vertebrates have resulted in four Hox clusters (24), where the second round of WGD resulted HoxC and HoxD clusters from the ancestral HoxC/D cluster (25). Two copies of the CNE in the basal group of jawed vertebrates, such as elephant shark (cartilaginous fish) and spotted gar (basal ray-finned fish; sister group of teleosts) (Fig. 1C), suggest that the CNE was already present in the ancestral HoxC/D cluster. Therefore, it is likely that the ancestral CNE might have had a *de novo* origin in jawed vertebrates after the first WGD (Fig. 1D). However, it is also possible that the CNE was present before the first WGD and then disappeared from ancestral HoxA/B cluster. To understand how CNE flanking

sequences evolved after WGD, we aligned flanking sequences within species and observed limited homology (for example, in human and elephant shark; Supplementary Fig. S2A) suggesting only CNEs were under selection. Then we separately aligned HoxC and HoxD CNE flanking sequences and observed independent conservation along each cluster (Supplementary Fig. S2B) and, in turn, identified a relatively long stretch of about 175 nucleotides of *HOTAIR* sequence that is conserved across vertebrates (except teleosts).

Paralogous HoxD CNE is transcribed from *HoxD11* noncoding transcript and exhibit sequence complementarity with *HOTAIR* CNE

Some CNEs are transcribed from noncoding transcripts (26); thus, we asked if human HOXD CNE is transcribed from coding/noncoding transcript. We intersected the genomic coordinates of the CNE with Ensembl genes and observed HOXD CNE is transcribed from an alternative noncoding isoform of *HOXD11* (referred to as *HOXD11* noncoding transcript) (Fig. 2A), which is a reported target gene that is silenced by *HOTAIR* in both human (5) and mouse (9). As both CNEs are embedded within their respective transcripts, we asked whether these CNEs exhibit sequence complementarity, similar to a miRNA seed and target sites. We aligned paralogous CNEs in the orientation of their respective transcripts and observed sequence complementarity (Fig. 2A), suggesting paralogous CNEs might hybridize.

To confirm whether the observed sequence complementarity in human is conserved in other vertebrates, we analyzed Ensembl-annotated genes and RNA-seq transcripts (22,27) across a subset of species and identified the HOTAIR and HoxD11 noncoding transcripts in mouse, ferrets, dog and coelacanth (Supplementary Fig. S3A-B). The HoxD11 noncoding transcript was detected in chicken but undetected in teleosts (zebrafish and tetraodon; Supplementary Fig. S3A-B). For other species, we inferred the putative orientation of the missing transcript (see Materials and Methods), as illustrated for chimp and painted turtle (Supplementary Fig. S3C-D). We aligned CNEs in the 5' -> 3' orientation and observed two interesting patterns. First, the sequence complementarity observed in human is conserved across vertebrates (Fig. 2B). This suggests that sequence complementarity between paralogous CNEs is an ancient feature that has been under selection pressure for more than 300 million years. This raises an important question as to whether the key function of these transcripts is to provide transcription of the CNE. Second, both CNEs underwent substitutions at specific positions that evolved in two separate waves in vertebrates and mammals (Fig. 2B). One of the substitutions resulted in gain of complementarity in mammals and the other substitution resulted in loss of complementarity in mammals (Fig. 2B). Collectively, the co-evolution of CNEs and retention of sequence complementarity in the transcribed orientation suggest that paralogous CNEs might interact and be co-regulated.



Figure 2 Sequence complementarity of the *HOTAIR* CNE and the paralogous HoxD CNE. (**A**) *HOTAIR* CNE and HOXD CNE are represented by rectangular blue and red bars, respectively. *HOTAIR* CNE and HOXD CNE are zoomed and show the genomic DNA sequences. *HOTAIR* is transcribed from negative strand and the CNE sequence is reverse transcribed to show in 5' to 3' orientation. Paralogous CNEs are complementary in the orientation of transcripts suggesting they might hybridize if co-expressed in the same cells. (**B**) Alignment of the *HOTAIR* CNE and HoxD CNE in transcribed orientation reveal sequence complementarity is conserved across vertebrates. Genetic substitution within paralogous CNEs occurred at specific positions that co-evolved in two successive waves in vertebrates and mammals resulting in gain of complementarity at one position and loss of complementarity at other position.

HOTAIR CNE reflects a poised enhancer in HOTAIR-unexpressed stem cells

CNEs are putative regulatory elements (28) where many of them function as enhancers (29); we thus asked whether CNEs have open chromatin state as in active regulatory elements

(30,31). We selected 29 cell lines (Supplementary Table S3; see Materials and Methods) from Roadmap Epigenome (30) and classified them as HOTAIR-expressed (N=10) and HOTAIR-unexpressed (N=19) groups based on RNA-seq expression levels (Supplementary Fig. S4A). Among HOTAIR-unexpressed cells, H1-hESC cell line is unique as the CNE is enriched for both repressive (H3K27me3) and open chromatin (as indicated by DNase hypersensitive sites (DHSs)) marks (Supplementary Fig. S4B), suggesting it to be in the poised state. We observed enriched DHS signals around the CNE in HOTAIR-expressed cells and HOTAIR-unexpressed stem cells, but not in HOTAIR-unexpressed differentiated cells (Fig. 3A). The paralogous HOXD CNE also has a similar chromatin state in corresponding cell lines (Supplementary Fig. S4C). Similar analysis on mouse cell lines (32) (see Materials and Methods) revealed that open chromatin state around the CNE is conserved in Hotair-expressed cells and Hotair-unexpressed stem cells (Fig. 3A; Supplementary Table S3). The chromatin state of Hotair CNE gradually becomes open during reprogramming of mouse embryonic fibroblasts (MEF) to induced pluripotent stem cells (iPSC) (33) (Supplementary Fig. S4D). We conclude that the chromatin state of CNE is open in HOTAIR-unexpressed stem cells and HOTAIR-expressed cells in both human and mouse.



Figure 3 The *HOTAIR* CNE represents a poised enhancer in *HOTAIR*-unexpressed stem cells in both human and mouse. (A) The *HOTAIR* CNE chromatin is open in *HOTAIR*-expressed and *HOTAIR*-unexpressed stem cells in human and mouse. Y-axis is normalized DNase I hypersensitive sites (DHS) coverage in reads per million (RPM). "N" denotes the number of cell lines. (B) The *HOTAIR* CNE is marked by p300, H3K4me1 and bimodal H3K27me3 peaks in H9-hESC (human) and iPSC (mouse) cell line that collectively define a poised enhancer. Y-axis is normalized coverage in reads per million (RPM). (C) A genome browser view around the *HOTAIR* locus with transcription factors, DHS, histone modifications and RNA-seq tracks from H1-hESC cell line. *HOTAIR* is unexpressed (as shown by RNA-seq track) and marked by broad H3K27me3 peak. The CNE is marked by depleted

H3K27me3 marks, reflecting a nucleosome-depleted region which is supported by enriched DHS peaks and binding of multiple transcription factors.

A subset of embryonic stem cell-specific enhancers are poised, characterized by presence of both activating and repressing marks (34); thus, we asked whether CNE reflects a poised enhancer in *HOTAIR*-unexpressed stem cells. We observed that H3K4me1, H3K27me3 and p300 signals are enriched at the CNE (Fig. 3B) in human H9-hESC (34) and mouse iPSC (33), supporting that *HOTAIR CNE* is a poised enhancer. However, the HOXD CNE lacks p300 and bimodal H3K27me3 peaks (Supplementary Fig. S4E). Poised enhancers have bimodal H3K27me3-modified nucleosomes (34), which is observed in *HOTAIR* CNE (H1-hESC and H9-hESC) (Fig. 3B; Supplementary Fig. S4F) but not in HOXD CNE.

We next analyzed all ENCODE transcription factors (TFs) (35) and observed CTBP2, YY1, SP1 and CHD1 are specifically enriched on the CNE exclusively in H1-hESC (Fig. 3C; Supplementary Table S4). YY1 has a known dual function in repressing or activating transcription depending upon the cofactors it recruits (36). The binding of YY1 on regulatory elements and their associated RNA species has been shown to play a role at enhancers (37). PRC2 components (SUZ12 and EZH2) are marked throughout the locus (Fig. 4C) that are not specific to stem cells (Supplementary Table S4). A recent study on mouse embryonic stem cells using promoter capture Hi-C data showed extensive promoter contacts for HoxC and HoxD clusters (38). Poised enhancers establish physical interactions with their target genes in ESCs in a PRC2-dependent manner (39), whether *HOTAIR* CNE might have a similar role in stem cells remains speculative. We conclude that in both human and mouse, the *HOTAIR* CNE possess characteristics features of a poised enhancer in stem cells.

HOTAIR CNE reflects an active enhancer in HOTAIR-expressed cells

Bidirectional transcription of enhancer RNAs (eRNAs) is a hallmark of active enhancers (40). We therefore analyzed the FANTOM5 (41) CAGE signals at the CNE and observed bidirectional transcription flanking the CNE (Fig. 4A; dashed rectangular box), supporting its role as an active enhancer. The CNE and the promoter region of *HOTAIR* primary and alternative transcripts are co-expressed (Fig. 4B; Supplementary Table S5), as exemplified during differentiation of myoblast to myotube (Supplementary Fig. S5A). The expression of *HOTAIR* promoter and CNE are positively correlated (Fig. 4C), which also holds true for two other alternative promoters except for the distal promoter (labelled as dp1) (Supplementary Fig. S5B). Negative correlation of distal promoter is mostly due to large number of samples where only distal promoter is expressed (Fig. 4B). The genomic region between bidirectional CAGE tags around the CNE is about the length of one nucleosome (Fig. 4D) and is conserved across vertebrates (Supplementary Fig. S2B). DHS, H3K4me1 and H3K27ac peaks are also enriched around the CNE (Fig. 4A) in myoblast and myotube (30), thus providing additional evidence as an active enhancer. The CNE is flanked by bimodal H3K4me1 peaks, a characteristic feature of active enhancers (Fig. 4E). Due to lack of

appropriate cell lines (embryonic hindlimbs, genital tubercle and piece of trunk corresponding to the sacro-caudal region) where *Hotair* is expressed (8-10), we did not detect CAGE tags in either *Hotair* CNE or promoter across mouse FANTOM5 samples (Supplementary Fig. S5C). However, we observed that H3K4me1 and H3K27ac are enriched in mouse embryonic (E10.5 days) hindlimbs (42) (Fig. 4F; Supplementary Fig. S5D).



Figure 4 The *HOTAIR* CNE reflects an active enhancer in *HOTAIR*-expressed muscle cells. (**A**) A genome browser view with FANTOM5 CAGE tags (combined tracks) and individual tracks on myoblast and myotube along with RNA-seq, H3K4me/me3, H3K27ac/me3, and DNase hypersensitive site (DHS) tracks from ENCODE. Horizontal bars across each histone and DHS tracks are annotated peaks. Forward and reverse strand CAGE tags are represented by blue and red. Bidirectional CAGE tags flanking the *HOTAIR* CNE is shown in dashed rectangular boxes. In both myoblast and myotube, the *HOTAIR* CNE is flanked by bidirectional CAGE tags and overlaps H3K4me1 and H3K27ac peaks. (**B**) Expression levels of the *HOTAIR* CNE and alternative promoters across FANTOM5 samples. (**C**) Expression level of the CNE is correlated with the *HOTAIR* promoter. (**D**) A zoomed view of CAGE tags (from myoblast and myotube differentiation time points) around the CNE shows bidirectional CAGE peaks that roughly represent the length of a nucleosome. (**E-F**) Bimodal H3K4me1 peaks flank the CNE in human myoblast (**e**) and embryonic 10.5days hindlimbs in mouse (**F**).

Thus, we have shown that *HOTAIR* CNE possesses characteristic features of an active enhancer in *HOTAIR*-expressed cells, in both human and mouse. Importantly, transcription of

the *HOTAIR* promoter is tightly linked to enhancer activity of the CNE, suggesting transcription of the CNE is a key function of the *HOTAIR* transcript. This notion is further supported by the evolution of CNE paralogs that reveal transcribed paralogous CNEs have retained sequence complementarity and thus have been actively selected during evolution (Fig. 2; Supplementary Fig. S3C-D). This observation suggests that transcription directly on the CNE might be a key contributor to the selection acting on the CNE.

HOTAIR CNE expression is positively correlated with *HOXC11* and negatively correlated with *HOXD11*

The lack of consensus on whether HOTAIR regulates the HoxD cluster posterior (HoxD11, HoxD12) genes in trans (5,9) or HoxC cluster (Hoxc10, Hoxc11) genes in cis (10) prompted us to ask how the expression levels of HOTAIR CNE is correlated with reported target genes in large samples of HOTAIR-expressed cells. We reasoned that HOTAIR CNE reflects an active enhancer and could potentially explain cis regulation, while the sequence complementarity between HOTAIR CNE and HOXD CNE (embedded within HOXD11) could potentially regulate HOXD cluster genes in trans. To this end, we analyzed 694 cell types from FANTOM5 (41) and plotted the expression levels of genes across four HOX clusters (Supplementary Fig. S6A). We next selected only the HOTAIR-expressed cells and observed positive correlation with HOXC posterior genes and a trend towards negative correlation with HOXD posterior genes (Fig. 5A). It is striking to note that among HOXC and HOXD cluster posterior genes, HOXC11 expression has the best positive correlation (R=0.64; p-value: 2.2E-13) in the HOXC cluster and HOXD11 coding (R=-0.25; p-value: 0.009) and noncoding transcript (R=-0.23; p-value: 0.017) has the best negative correlation, both of which are target genes regulated by HOTAIR in human (5) and mouse (9,10). The observed negative correlation on HOXD cluster genes will diminish if analyzed by combining HOTAIR-expressed and HOTAIR-unexpressed cells. To confirm whether the observed correlation is present in other datasets, we analyzed gene expression data of breast cancer patients from TCGA (43). We selected only those patients (N=605) where HOTAIR is expressed (see Materials and Methods) and observed a similar trend of positive correlation with HOTAIR and HOXC cluster posterior genes and a trend of weak negative correlation with HOTAIR and HOXD cluster posterior genes (Supplementary Fig. S6B). Similarly, HOXC11 has the highest positive correlation (R=0.207; p-value:2.63E-07) and HOXD11 has highest negative correlation (R=-0.107; p-value:0.008). The act of transcription by HOTAIR, transcribes the HOTAIR CNE (putative enhancer) and collectively regulates neighboring genes in cis. On the other hand, sequence complementarity of HOTAIR CNE with paralogous HOXD CNE embedded in the HOXD11 noncoding transcript might negate the expression of HOXD11. Our observations suggest that HOTAIR CNE might simultaneously regulate HOXC11 in cis and HOXD11 in trans.



Figure 5 Correlation of expression levels of *HOTAIR* CNE with HOXC and HOXD clusters posterior genes in *HOTAIR*-expressed cells across FANTOM5 samples. (**A**) Expression levels of *HOTAIR* CNE with HOXC and HOXD cluster posterior genes on individual cell types. X-axis represents expression level of *HOTAIR* CNE and y-axis represents the expression levels of HOXC and HOXD cluster genes. Expression level is measured as tags per million (TPM). Expression levels of *HOTAIR* CNE and HOXC cluster genes show a trend of positive correlation where *HOXC11* (highlighted in green) has the highest positive correlation. Expression levels of *HOTAIR* CNE and HOXD cluster genes showed a trend of weak negative correlation where *HOXD11* coding and noncoding (highlighted in green) is the most negatively correlated gene.

DISCUSSION

We have identified and characterized the *HOTAIR* CNE as the ancestral sequence, which also has a paralogous coy in the HoxD cluster across vertebrates. Lack of the CNE in teleosts HoxC cluster is unexplained given it is conserved in other fish (spotted gar and elephant shark) and is likely one of the consequences of teleosts lineage whole genome duplication. Conservation of *HOTAIR* CNE and flanking sequences even in early fish (elephant shark, spotted gar) and tetrapod (coelacanth) (Supplementary Fig. S2B) does suggest an ancestral form of *HOTAIR* across all vertebrates, but needs additional RNA-seq transcript evidence which is currently limited to only few species due to limited data. Importantly, the length of the *HOTAIR* CNE flanking sequences that is conserved across vertebrates (Supplementary Fig. S2B) roughly represents region of DHS peak and bidirectional CAGE tags.



Figure 6 Schematic representation of various features of the *HOTAIR* CNE and its paralog in HoxD cluster in human and mouse. (**A**) The *HOTAIR* CNE has a paralog in HoxD cluster. Transcribed paralogous CNEs exhibits sequence complementarity and this complementarity is conserved across vertebrates. Based on *HOTAIR* expression, cell types are divided into *HOTAIR*-expressed and *HOTAIR*-unexpressed groups, where HOTAIR-unexpressed cells are further divided into stem cells and terminally differentiated cells based on DHS signals. Chromatin states, transcription factor bindings and CAGE signals on homologous CNEs are shown schematically.

Many of the tested CNEs act as enhancers in reporter construct (29). In fact, the homologous Hoxd CNE drives expression in a proximal posterior part of developing forelimbs (44) but does not work in its own context in zebrafish, as it requires the mouse promoter. However,

deletion of Hoxd CNE in *vivo* revealed no phenotype (45), and it was speculated that the difference might be due to other phenotypes that were not detected or has a redundant copy of the sequence that have masked the effect. We have now identified a paralog of the Hoxd CNE in *Hotair* (Fig. 1). This raises an interesting question whether paralogous CNEs have redundant functions, such that deletion of one CNE might be compensated by the other CNE. Putting this into the context of deletion of *Hotair* locus (including CNE) *in vivo* (9,10), it remains unknown whether the effects of *Hotair* CNE deletion is compensated for, to a certain extent, by paralogous Hoxd CNE. Duplicated CNEs have retained only a single copy (28); thus, retention of both copies of duplicated CNE is rare and provides an interesting model to test redundancy/sub-functionalization of CNEs.

Mechanistically, *HOTAIR* was proposed to regulate HOXD cluster genes by recruiting PRC2 (5), but this model has now been questioned since PRC2 is dispensable from *HOTAIR*-mediated transcriptional repression (7). Our observation of sequence complementarity between *HOTAIR* CNE and HOXD CNE (Fig. 2; Supplementary Fig. S3A-B; Fig. 6), suggest sequence-specific binding. This notion is further supported by the observed simultaneous genetic substitutions of CNEs during evolution. Moreover, paralogous CNEs are generally co-expressed in the same cell lines (Fig. 6a), simultaneously maintaining open or closed chromatin (Fig. 6) and expression of the *HOTAIR* CNE and *HOXD11* (previously reported target gene of *HOTAIR* (5,9)) is negatively correlated which collectively provide strong arguments that these CNEs might interact.

HOTAIR has been the prototypic model of functional lncRNA and, thus, the presence of an embedded enhancer (characterization based on bidirectional CAGE tags, DHS and histone peaks) was unexpected (Fig. 6). Most importantly, transcriptional activity of the CNE is coupled to *HOTAIR* transcription, suggesting that a key function of the *HOTAIR* transcript is to provide active transcription for the CNE. Though the expression of the *HOTAIR* CNE is positively correlated with *HOXC11* (Fig. 5), it remains unknown whether the observed *cis* regulation of *Hotair* (10) is mediated through the CNE. However, we observed that *HOTAIR* expression is positively correlated with *HOXC11* and negatively correlated with *HOXD11* and suggest a unifying model of *HOTAIR* regulation through CNE. Our findings have opened an avenue for multiple testable hypotheses that should clarify ongoing controversies whether *HOTAIR* regulates HoxC or HoxD cluster genes *in cis* or *trans* (5,7-10) or can regulate both HoxC and HoxD cluster genes simultaneously. Our results should guide future experiments to resolve ongoing controversies. Future studies should focus on understanding how specific deletion of CNE(s) alters *HOTAIR* regulation in different cellular contexts.

MATERIALS AND METHODS

Genome assemblies and gene annotations

Analyses on human and mouse were done in hg19 and mm9 genome version respectively.

The genome assemblies of 37 species are listed in Supplementary Table S1. Gene models were downloaded from UCSC (46). The conserved noncoding elements (CNEs) were downloaded from ANCORA (19).

Roadmap Epigenome, ENCODE, mouse ENCODE and FANTOM5 data sets

Roadmap Epigenome data were downloaded from NIH Roadmap Epigenome browser (30). Annotated chromatin states were downloaded from 127 cell lines. Histone modifications (H3K4me1/3, H3K27ac/me3) and DNase I hypersensitive sites (DHSs) data were downloaded as mapped (tagAlign format) files. RNA-seq data for 57 cell lines were downloaded in the computed gene expression (RPKM) matrix (30). Only 29 cell lines that have all four (H3K4me1/3, H3K27ac/me3) histone modifications, DHS and RNA-seq were used for downstream analysis.

Histone modification and RNA-seq data from ENCODE were downloaded as mapped BAM files. ENCODE transcription factor ChIP-seq were downloaded as annotated peaks (35). Histone modifications (H3K4me1/3, H3K27ac/me3), DHS and RNA-seq data from mouse ENCODE were downloaded as mapped BAM files (32). Samples with replicates were merged into a single file. A threshold of 0.5 RPKM (reads per kilobase per million) was used as the cutoff expression to determine whether *HOTAIR* is expressed or not in the given RNA-seq samples.

Human and mouse CAGE-seq data were downloaded from FANTOM5 (41,47). Replicates were pooled into single file and resulting CAGE tags in each sample were quantified as tags per million (TPM) (21). CAGE tags with the highest expression level defined dominant transcription start site (TSS). CAGE based expression level was computed by summing all CAGE tags in defined the promoter region (300 bases upstream and downstream of TSS). To compare the expression correlation across four HOX clusters genes, we selected only 694 cell types only if any of HOX genes was expressed at the minimum of 5 TPM.

Data sets used from multiple studies

RNA-seq transcripts for multiple species were used from previous studies(21,22,27). Raw data during reprogramming of mouse embryonic fibroblast to iPSC were download from GEO (GSE90894) (33). Raw data for H3K27me3, H3K4me1, H3K27ac and p300 from H9-hESC cell lines were download from GEO (GSE24447) (34). Mouse embryonic (10.5 days) hind limb data were downloaded from GEO (GSE84793) (42). Raw fastq reads were mapped using bowtie2 (48). Only unique mapping reads were considered for downstream analysis. Breast cancer patients mRNA (Illumina Human v3 microarray) data (43) were downloaded from TCGA portal. Expression levels are measured in z-scores. We filtered samples where HOTAIR expression (<= 0.5) and were left with 605 patients.

Mapping of HOTAIR CNE across multiple species

The *HOTAIR* CNE sequences from both human and zebrafish was used as a query sequence. We used BLAST (blastall -p blastn -d -e 0.01 -m 8)(49) to find homologous sequences against 37 species (Supplementary Table S1). Even at the permissive e-value cutoff of 0.01, only two homologous sequences were identified.

Determining the orientation of transcripts overlapping CNEs

The *HOTAIR* transcript is annotated in multiple species (22,46), such as human, chimp, mouse, ferret and dog, and its orientation is antisense to *HoxC11* and *HoxC12* genes. Species lacking *HOTAIR* annotation, orientation of *HOTAIR* CNE was assigned antisense to annotated HoxC cluster genes. In multiple species, such as human, chimp, mouse, ferret, dog and chicken, HoxD CNE is embedded within the untranslated region of *HoxD11* noncoding transcript, which is an alternative splice variant of *HoxD11* coding gene. Thus, orientation of HoxD CNE was assigned similar to that of annotated *HoxD11* gene. Among teleosts fish, we analyzed RNA-seq transcripts in zebrafish(21,22) and tetraodon (27), and did not identify *HoxD11* noncoding transcript.

Software and tools

Multiple alignments were generated using ClustalW (50) and Jalview (51). Sequence logos were generated using WebLogo (52). Visualization of multiple data were done by uploading bigwig tracks on UCSC genome browser and images were downloaded. Bedtools (53), bash, perl and R scripts were used for data analysis.

FUNDING

C.N is the recipient of a postdoctoral fellowship from the Danish Medical Research Council (6110-00557A). The laboratory of J.B.A is supported by the Danish Medical Research Council (4183-00118A), Danish Cancer Society (R98-A6446) and Novo Nordisk Foundation (14040). P.M acknowledges funding from National Science Center (2014/14/E/NZ6/00162, Poland). F.M thanks funding support by the BBSRC and a Wellcome Trust Investigator Award, UK.

ACKNOWLEDGEMENTS

We are very grateful to ENCODE, FANTOM5, Roadmap Epigenome consortia and other researchers for making data freely available. We thank Colm J. O'Rourke, Michal Lubas and Albin Sandelin for critical comments on the manuscript. C.N acknowledges Boris Lenhard, Vidar Martin Steen, Piotr Mydel and Jesper B Andersen for supporting him.

AUTHOR'S CONTRIBUTION

C.N conceived the story, analyzed data and interpreted results. Y.H, S.P, P.M, B.L, F.M and J.B.A contributed to data analysis and critical discussions. All authors contributed to writing the manuscript.

CONFLICT OF INTEREST

Authors declare no conflict of interest and no competing financial interest.

SUPPLEMENTARY TABLES LEGENDS

Supplementary Table S1. The genomic coordinates of the CNE in HoxC and HoxD cluster across species. Lack of homolog is denoted by "-". The orientation of the *HOTAIR* CNE is antisense to the annotated *HoxC11* gene and the orientation of the transcript overlapping HoxD CNE is similar to that of annotated *HOXD11*. Number of mismatches between paralogous CNEs is shown. Evidence of Ensembl or RNA-seq annotated host transcript-encoding CNEs is denoted by "Y" and if there is no evidence, as in case of teleosts it is denoted by "N". No evidence of transcripts due to lack of data is denoted by "NA".

Supplementary Table S2. Summary of annotation of HoxC cluster genes in aves and teleosts. Only few HoxC genes are annotated in aves and are located in different contigs. Except chicken, other aves do not have annotated *HoxC11* and *HoxC12* genes. In teleosts, HoxC cluster genes are well annotated and located in the same cluster.

Supplementary Table S3. List of cell lines from Roadmap Epigenome samples that are classified into three groups based on *HOTAIR* expression from RNA-seq.

Supplementary Table S4. Enrichment of transcription factors bound on the *HOTAIR* CNE and promoter.

Supplementary Table S5. Expression level of the *HOTAIR* CNE and promoter across FANTOM5 samples.

REFERENCES

- 1. Hon, C.C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J. *et al.* (2017) An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, **543**, 199-204.
- 2. Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S. *et al.* (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet*, **47**, 199-208.
- 3. Guttman, M. and Rinn, J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339-346.
- 4. Mercer, T.R. and Mattick, J.S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol*, **20**, 300-307.
- 5. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311-1323.
- 6. Davidovich, C., Zheng, L., Goodrich, K.J. and Cech, T.R. (2013) Promiscuous RNA binding by Polycomb repressive complex 2. *Nat Struct Mol Biol*, **20**, 1250-1257.
- Portoso, M., Ragazzini, R., Brencic, Z., Moiani, A., Michaud, A., Vassilev, I., Wassef, M., Servant, N., Sargueil, B. and Margueron, R. (2017) PRC2 is dispensable for HOTAIR-mediated transcriptional repression. *EMBO J*, **36**, 981-994.

- 8. Schorderet, P. and Duboule, D. (2011) Structural and functional differences in the long non-coding RNA hotair in mouse and human. *PLoS Genet*, **7**, e1002071.
- 9. Li, L., Liu, B., Wapinski, O.L., Tsai, M.C., Qu, K., Zhang, J., Carlson, J.C., Lin, M., Fang, F., Gupta, R.A. *et al.* (2013) Targeted disruption of Hotair leads to homeotic transformation and gene derepression. *Cell Rep*, **5**, 3-12.
- 10. Amandio, A.R., Necsulea, A., Joye, E., Mascrez, B. and Duboule, D. (2016) Hotair Is Dispensible for Mouse Development. *PLoS Genet*, **12**, e1006232.
- 11. Li, L., Helms, J.A. and Chang, H.Y. (2016) Comment on "Hotair Is Dispensable for Mouse Development". *PLoS Genet*, **12**, e1006406.
- Selleri, L., Bartolomei, M.S., Bickmore, W.A., He, L., Stubbs, L., Reik, W. and Barsh, G.S. (2016) A Hox-Embedded Long Noncoding RNA: Is It All Hot Air? *PLoS Genet*, **12**, e1006485.
- 13. Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q. *et al.* (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell*, **143**, 46-58.
- 14. Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W. *et al.* (2014) Considerations when investigating IncRNA function in vivo. *Elife*, **3**, e03058.
- 15. Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O., Hughes, J.R., Hardison, R.C. *et al.* (2016) Unlinking an IncRNA from Its Associated cis Element. *Mol Cell*, **62**, 104-110.
- 16. Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M. and Lander, E.S. (2016) Local regulation of gene expression by IncRNA promoters, transcription and splicing. *Nature*, **539**, 452-455.
- 17. Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H. *et al.* (2015) Opposing Roles for the IncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. *Cell Stem Cell*, **16**, 504-516.
- 18. Lee, A.P., Koh, E.G., Tay, A., Brenner, S. and Venkatesh, B. (2006) Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters. *Proc Natl Acad Sci U S A*, **103**, 6994-6999.
- 19. Engstrom, P.G., Fredman, D. and Lenhard, B. (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biol*, **9**, R34.
- 20. He, S., Liu, S. and Zhu, H. (2011) The sequence, structure and evolutionary features of HOTAIR in mammals. *BMC Evol Biol*, **11**, 102.
- 21. Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A.M., Sheng, Y., Abdelhamid, R.F., Anand, S. *et al.* (2013) Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Res*, **23**, 1938-1950.
- 22. Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P. and Ulitsky, I. (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*, **11**, 1110-1122.
- 23. Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M. *et al.* (2014) Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, **507**, 381-385.
- 24. Holland, P.W., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994) Gene duplications and the origins of vertebrate development. *Dev Suppl*, 125-133.
- 25. Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R. and Duboule, D. (2013) Duplications of hox gene clusters and the emergence of vertebrates. *Dev Biol*, **378**, 194-199.
- 26. Peng, J.C., Shen, J. and Ran, Z.H. (2013) Transcribed ultraconserved region in human cancers. *RNA Biol*, **10**, 1771-1777.
- 27. Basu, S., Hadzhiev, Y., Petrosino, G., Nepal, C., Gehrig, J., Armant, O., Ferg, M., Strahle, U., Sanges, R. and Muller, F. (2016) The Tetraodon nigroviridis reference transcriptome: developmental transition, length retention and microsynteny of long non-coding RNAs in a compact vertebrate genome. *Sci Rep*, **6**, 33210.
- 28. Harmston, N., Baresic, A. and Lenhard, B. (2013) The mystery of extreme non-coding conservation. *Philos Trans R Soc Lond B Biol Sci*, **368**, 20130021.

- 29. Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D. *et al.* (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, **444**, 499-502.
- 30. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317-330.
- 31. Pundhir, S., Bagger, F.O., Lauridsen, F.B., Rapin, N. and Porse, B.T. (2016) Peakvalley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic Acids Res*, **44**, 4037-4051.
- 32. Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**, 355-364.
- 33. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J. and Plath, K. (2017) Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*, **168**, 442-459 e420.
- 34. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A. and Wysocka, J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279-283.
- 35. Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R. *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*, **489**, 91-100.
- 36. Deng, Z., Cao, P., Wan, M.M. and Sui, G. (2010) Yin Yang 1: a multifaceted protein beyond a transcription factor. *Transcription*, **1**, 81-84.
- 37. Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A. and Young, R.A. (2015) Transcription factor trapping by RNA in gene regulatory elements. *Science*, **350**, 978-981.
- Schoenfelder, S., Sugar, R., Dimond, A., Javierre, B.M., Armstrong, H., Mifsud, B., Dimitrova, E., Matheson, L., Tavares-Cadete, F., Furlan-Magaril, M. *et al.* (2015) Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet*, **47**, 1179-1186.
- 39. Cruz-Molina, S., Respuela, P., Tebartz, C., Kolovos, P., Nikolic, M., Fueyo, R., van ljcken, W.F.J., Grosveld, F., Frommolt, P., Bazzi, H. *et al.* (2017) PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*, **20**, 689-705 e689.
- 40. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455-461.
- 41. Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462-470.
- 42. Andrey, G., Schopflin, R., Jerkovic, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M. *et al.* (2017) Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res*, **27**, 223-233.
- 43. Pereira, B., Chin, S.F., Rueda, O.M., Vollan, H.K., Provenzano, E., Bardwell, H.A., Pugh, M., Jones, L., Russell, R., Sammut, S.J. *et al.* (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*, **7**, 11479.
- 44. Beckers, J., Gerard, M. and Duboule, D. (1996) Transgenic analysis of a potential Hoxd-11 limb regulatory element present in tetrapods and fish. *Dev Biol*, **180**, 543-553.
- 45. Beckers, J. and Duboule, D. (1998) Genetic analysis of a conserved sequence in the HoxD complex: regulatory redundancy or limitations of the transgenic approach? *Dev Dyn*, **213**, 1-11.
- 46. Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res*, **44**, D717-725.
- 47. Arner, E., Daub, C.O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., Lennartsson, A., Ronnerblad, M., Hrydziuszko, O., Vitezic, M. *et al.* (2015)

Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, **347**, 1010-1014.

- 48. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.
- 49. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.
- 50. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. and Thompson, J.D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*, **31**, 3497-3500.
- 51. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.
- 52. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188-1190.
- 53. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.



Supplementary Figure S1. A genome browser view of *HOTAIR* in human and mouse to show sequence conservation. (A) Conserved noncoding elements (CNEs) track from ANCORA browser is shown as horizontal black bars on top. CNE that is conserved across all mammals and vertebrates is highlighted. CNE is located eight nucleotides away from splice site. CNE is mapped across all vertebrates, but flanking sequences (corresponding *HOTAIR* exons) are mapped in other fish and tetrapod (excluding teleost). (B) Structure of mouse *Hotair* transcript is different than in human. Homolog of CNE is also located eight nucleotides away from splice site of *Hotair* transcript. CNE is conserved across mammals and vertebrates.





Supplementary Figure S2. Alignment of CNEs and flanking sequences across HoxC and HoxD clusters within and across species. (**A**) Schematic representation on the origin of HoxC and HoxD cluster from the ancestral HoxC/D cluster after second round of whole genome duplication (2R-WGD). Alignment of flanking sequences around paralogous CNEs, in human and elephant shark, show little conservation despite both sequences are duplicated from the same ancestral sequences. (**B**) Schematic representation of alignment HoxC and HoxD CNE and flanking sequences across vertebrates. Alignment of 70 nucleotides flanking the *HOTAIR* CNE (across 22 vertebrates (excluding aves and teleosts)) and HOXD CNE (across 32 vertebrates) show conservation of flanking sequences across vertebrates. Species are aligned in the same order as in Supplementary Figure S1A.

bioRxiv preprint doi: https://doi.org/10.1101/250621; this version posted January 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CCBY NC-ND 4.0 International license HoxC cluster В HoxD11 noncoding transcript HoxC11 HoxD12 HoxD11 HoxC12 Hotair chr15 | 102940000 102955000 chr2 74676000 74687000 chr27 | 1306000 1289000 chr36 | 19908500 19917000 1 ∎++-Ĩ Н GL897095 | 2201500 2193500 GL897098 1110000 1095000 H chr7 15821000 15810000 chrUN JH376221 chrUN AADN03020675 7900 4600 12000 7000 | Ĥ chr9|1976000 1967000 chr23 36117000 36129000 chr2 13445500 13439500 chr9 4201000 4214000 D Chimp Painted turtle JH584844-1101648-1125928 chr12:35275000-35295000 inferred orientation of hotair Hotair ŀ Hoxc11 Hoxc12 hoxc11 hoxc12 genomic DNA genomic DNA CTAACTTGACCTTCACTTTGTCATGGCGCCCCC CTAACTTGACCTTCACTTTGTAATGGCGCCCCC 5 3 CTAACTTGACCTTCACTTTGTCATGGCGCCCCC 3' CCCCGCGGTAATGTTTCACTTCCAGTTCAATC 5' 3' GATTGAACTGGAATTGAAACAGTCCCGCGGGG 5' GGGGCGCCATAACAAAGTTAAGGTCAAGTTAG 3' 5' GGGGCGCCCTGACAAAGTTAAGGTCAAGTTAG GGGGCGCCATAACAAAGTTAAGGTCAAGTTAG genomic DNA genomic DNA chr2B:180576000-180585000 JH584712:2164000-2178000 inferred orientation inferred orientation of hoxd11 isoform hoxd12 Hoxd11 Hoxd12 hoxd11 of Hoxd11 isoform 3' CTAACTTGACCTTCACTTTGTCATGGCGCCCC | | | | | | | * | | | | | *

Supplementary Figure S3. Evidence of *HOTAIR* and *HoxD11* noncoding transcript across multiple species and sequence complementarity of paralogous CNEs. (A) The *HOTAIR* CNE is represented by rectangular blue bar. A homolog of the *HOTAIR* CNE is undetected in chicken. Intergenic region between *hoxc11* and *hoxc12* remains unassembled as *hoxc11* and *hoxc12* genes are assembled in different contigs. Zebrafish *hoxc11* and *hoxc12* is assembled but lacks the CNE. (B) The HoxD CNE is represented by rectangular red bar. *HoxD11* noncoding transcript is detected across tetrapods but not in teleosts. (C-D) Schematic representation to show how the orientation of missing *Hotair* and *HoxD11* noncoding transcripts were inferred. *HOTAIR* is antisense to *HoxC11* genes across species, and thus the same convention was used to infer the orientation of *Hotair* CNE. HoxD11 noncoding transcript is an alternative splice variant of *HoxD11* across multiple species; thus, the same convention was used. Expected transcripts are represented by dashed rectangular boxes and lines and arrows indicate transcript orientation. The CNE sequences are zoomed in on to show the genomic DNA. Aligning sequences in 5' to 3' orientation show paralogous CNEs are complementarity. (E) Sequence logos of *HOTAIR* CNE and HoxD CNE show paralogous CNEs exhibit sequence complementarity in transcribed orientation.

5'

GATTGAACTGGAATTGAAACAGTcCCGCGGGG

С

E



Supplementary Figure S4. Chromatin environment of the *HOTAIR* CNE. (**A**) Twenty-nine cell lines were separated into *HOTAIR*-expressed and *HOTAIR*-unexpressed cell lines using a threshold of 0.5 RPKM (reads per kilobase per million mapped reads). *HOTAIR*-unexpressed cell lines were further separated into stem cells (N=1) and terminally differentiated cells (N=18). (**B**) Heatmaps show the normalized read counts in 250 nucleotides flanking *HOTAIR* CNE across 29 cell lines from Roadmap Epigenome Consortium. Cell line IDs are the same as provided by Roadmap Epigenome Consortium. (**C**) DHS signals around the HOXD CNE across three groups (*HOTAIR*-expressed cells, *HOTAIR*-unexpressed stem cells and *HOTAIR*-unexpressed cells) show similar trends as that of the *HOTAIR* CNE. Y-axis represents normalized read counts in reads per million (RPM). (**D**) Open chromatin state of mouse *Hotair* CNE is dynamically acquired during reprogramming of mouse embryonic fibroblast to iPSC. (**E**) Human HOXD CNE is not a poised enhancer as it lacks p300, H3K4me1 and bimodal H3K27me3 peaks in H9-hESC cell line. (**F**) Pattern of H3K27me3 marks around *HOTAIR* CNE across three groups.



Supplementary Figure S5. Transcriptional dynamics of *HOTAIR* promoters. (A) Expression levels of *HOTAIR* promoter, alternative promoters and CNE across nine different stages during differentiation from myoblast to myotube. (B) Expression correlation of *HOTAIR* promoter, alternative promoters and CNE across FANTOM5 samples are positively correlated except for distal promoter (dp1). (C) No significant CAGE tags are identified on mouse *Hotair* across FANTOM5 samples as it lacks tissues (embryonic hindlimbs, genital tubercle and a piece of trunk corresponding to sacro-caudal region) where Hotair is expressed. Orthologs of promoter region of HOTAIR are aligned to mouse. (D) *Hotair* CNE is flanked by bidirectional H3K27ac on mouse hindlimbs (E10.5 days).





Supplementary Figure S6. Co-expression analysis of *HOTAIR* with four HOXC clusters genes across FANTOM5 samples. (A) Heatmap shows the expression level of *HOTAIR* and four HOX cluster genes. Cell types are sorted based on *HOTAIR* CNE and *HOTAIR* distal promoter expression. Expression levels are scaled between 0-1 across each column, where the most highly expressed gene is assigned 1. (B) Expression levels of *HOTAIR* with HOXC and HOXD cluster posterior genes on individual breast cancer patients show a trend of positive correlation with HOXC cluster genes and a trend of weak negative correlation with HOXD cluster genes.

List of supplementary tables

Table Supplemental S1	Coordinates of CNE in HoxC and HoxD cluster across vertebrates.
Table Supplemental S2	Summary of annotation of hoxc cluster genes in aves and telesot.
Table Supplemental S3	Roadmap Epigenome cell lines classified into three groups based on HOTAIR expression.
Table Supplemental S4	Enrichment of transcription factors (TFs) bound on HOTAIR CNE and promoter.
Table Supplemental S5	Expression levels of HOTAIR CNE, promoter and distal promoter across FANTOM5 samples

Table Supplemental S1. Coordinates of CNE in HoxC and HoxD cluster across vertebrates.

Sequence used for homology search

hg19_chr12:54359708-54359739_HOXC_CTAACTTGACCTTCACTTTGTCATGGCGCCCCC hg19_chr2:176969157-176969189_HOXD_CTAACTTGACCTTAACTTTGTCAGGGCGCCCCC

	Coordinates of	CNE homologs	Count	Host Transcr	ipt Evidence
	HoxC cluster	HoxD cluster			
	Chr:Start-End(Strand)	Chr:Start-End(Strand)	Mismatches	HoxC	HoxD
Primates					
Human(hg19)	chr12:54359707-54359739(-)	chr2:176969157-176969189(+)	2	Y	Y
Chimp(panTro4)	chr12:35284661-35284693(+)	chr2B:180579781-180579813(+)	2	Y	Y
Gorilla(gorGor3)	chr12:52012160-52012192(-)	chr2B:64029540-64029572(+)	3	Y	Y
Orangutan(ponAbe2)	chr12:53740632-53740664(-)	chr2b:66260351-66260383(+)	3	Y	Y
Euarchontoglires					
Squirrel(speTri2)	JH393322:11187590-11187622(-)	JH393319:5591308-5591340(+)	3		
Mouse(mm10)	chr15:102947454-102947486(-)	chr2:74679612-74679644(+)	4	Y	Y
Rat(rn5)	chr7:142349588-142349620(-)	chr3:68068756-68068788(+)	4		Y
GuineaPig(chiLan1)	scaffold_9:45845103-45845135(+)	scaffold_3:14206941-14206973(-)	4		
Rabbit(oryCun2)	chr4:37750144-37750176(-)	chr7:115664606-115664638(+)	3		
Laurasiatheria					
Cow(bosTau8)	chr5:26225443-26225475(+)	chr2:20844410-20844442(-)	2		
Dog(canFam3)	chr27:1296692-1296724(+)	chr36:19912089-19912121(+)	3	Y	Y
HedgeHog(eriEur3)	JH835374:2711697-2711729(+)	JH835538:1267467-1267499(+)	4		
Afrotheria					
Elephant(loxAfr3)	scaffold_2:40399191-40399223(-)	scaffold_3:51280894-51280926(-)	3	NA	NA
Tenrec(echTel2)	JH980427:730939-730971(-)	JH980302:1668675-1668707(+)	4	NA	NA
Manatee(triMan1)	JH594633:22473672-22473704(+)	JH594643:8289255-8289287(-)	3	NA	NA
Mammals					
Armadilo(dasNov3)	JH569064:163083-163115(+)	JH575655:3643342-3643374(+)	0		
Platypus(ornAna1)	Contig13135:14368-14400(+)	Contig13670:14779-14811(+)	3		
Aves	3	0			
MediumGroundFinch(geoFor1)		JH739888:20155410-20155442(+)	-	NA	
ZebraFinch(taeGut2)		chr7:17035450-17035482(+)	-	NA	
Budgegerigar(melUnd1)		JH556613:11953860-11953892(-)	-	NA	
Chicken(galGal4)		chr7:15818298-15818330(-)	-	NA	Y
Turkev(melGal1)		chr7:16893541-16893573(-)	-	NA	
Sarcoptervgii					
AmericanAlligator(allMis1)	JH739658:224031-224063(-)	JH736379:590770-590802(-)	3		
PaintedTurtle(chrPic2)	JH584844:1113401-1113433(+)	JH584712:2171341-2171373(+)	3		
Frog(xenTro3)	GL172862:354043-354075(-)	GL172799:638809-638841(-)	7		Y
Coelacanth(latCha1)	JH126829:1709983-1710015(+)	JH127184:433817-433849(-)	4	Y	Ŷ
Holesti Fish					-
Spottedgar (lepOcu1)	LG12:15577478-15577509(-)	LG4·15815626-15815646(-)	3		
Teleost Fish	20.2		°,		
Tetraodon(tetNig2)		chr2:13442915-13442947(-)	-	Ν	Ν
Fuqu(fr3)		chr1.15039680-15039712(-)	-	N	N
Medaka(orvl.at2)		chr21.24604991-24605023(+)	-	N	N
AtlanticCod(gadMor1)		HE566826:374754-374786(-)	-	N	N
Zebrafish(danRer7)		chr9:1970782-1970814(-)	_	N	N
Cartilaginous Fish					i N
ElephantShark(calMal1)	KI636119:205590-205622(-)	KI635868-5106636-5106668(+)	7		
Jawless Vertebrates			,		
Lamprev(petMar2)			_	_	_
Deuterostome					
C Intestinalis(ci2)					
Insect			-	-	-
D melanogaster(dm6)			_		
Nematode			-	-	-
C elegans(ce6)			_		
0.009013(000)			-	-	-

Table Supplemental S2. Summary of annotation of hoxc cluster genes in aves and telesot.

	hox	a	hox	b	hox	с	ho	xd	Anno	otation	Same	Anno	otation	Same
	Genes	Chr	Genes	Chr	Genes	Chr	Genes	Chr	hoxc11	hoxc12	Chr	hoxd11	hoxd12	Chr
Aves														
ZebraFinch	9	1	5	1	2	2	5	2	NO	NO	NA	YES	YES	YES
Chicken	8	1	9	1	5	5	7	1	YES	YES	Diff	YES	YES	YES
Turkey	4	1	4	2	2	1	5	1	NO	NO	NA	YES	YES	YES
Teleost														
Tetraodon	13	2	9	2	11	2	6	1	YES	YES	YES	YES	YES	YES
Fugu	9	3	7	3	6	1	4	1	NO	YES	NA	NO	YES	NA
Atlantic Cod	12	5	9	3	10	1	6	1	YES	YES	YES	YES	YES	YES
Zebrafish	12	2	15	2	15	2	7	1	YES	YES	YES	YES	YES	YES

Table Supplemental S3: Roadmap Epigenome cell lines classified into three groups based on HOTAIR expression.

HOTAIR-expre	essed differentiated cells	
E055	Epithelial	Foreskin Fibroblast Primary Cells skin01
E056	Epithelial	Foreskin Fibroblast Primary Cells skin01
E117	CRVX.HELAS3.CNCR	HeLa-S3 Cervical Carcinoma Cell Line
E120	MUS.HSMM	HSMM Skeletal Muscle Myoblasts Cells
E057	Epithelial	Foreskin Keratinocyte Primary Cells skin02
E127	BONE.OSTEO	Osteoblast Primary Cells
E006	ES-deriv	H1 Derived Mesenchymal Stem Cells
E005	ES-deriv	H1 BMP4 Derived Trophoblast Cultured Cells

HOTAIR-unexpressed stem cells ESC

E003

H1-hESC cells

HOTAIR-unexpressed differentiated cells

E004	ES-deriv	H1 BMP4 Derived Mesendoderm Cultured Cells
E007	ES-deriv	H1 Derived Neuronal Progenitor Cultured Cells
E028	Epithelial	Breast variant Human Mammary Epithelial Cells (vHMEC)
E114	LNG.A549.ETOH002.CNCR	A549 EtOH 0.02pct Lung Carcinoma Cell Line
E119	BRST.HMEC	HMEC Mammary Epithelial Primary Cells
E100	Digestive	Gastric
E059	SKIN.PEN.FRSK.MEL.01	Foreskin Melanocyte Primary Cells skin01
E085	Digestive	Fetal Intestine Small
E123	BLD.K562.CNCR	K562 Leukemia Cells
E097	Ovary	Ovary
E084	Digestive	Fetal Intestine Large
E109	Digestive	Small Intestine
E050	HSC & B-cell	Primary hematopoietic stem cells G-CSF-mobilized Female
E082	Brain	Fetal Brain Female
E094	GI.STMC.GAST	Gastric
E098	Pancreas	Pancreas
E116	BLD.GM12878	GM12878 Lymphoblastoid Cells
E118	LIV.HEPG2.CNCR	HepG2 Hepatocellular Carcinoma Cell Line
E122	VAS.HUVEC	HUVEC Umbilical Vein Endothelial Primary Cells
E128	LNG.NHLF	NHLF Lung Fibroblast Primary Cells

Table Supplemental S4: Enrichment of transcription factors (TFs) bound on HOTAIR CNE and promoter.

chr12 54350907 54360239 HotairCNE 1000 - chr12 54359807 CHD1 261 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359800 TCF12 178 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359910 S4360141 CHD1 261 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359803 SRF 631 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359803 SRF 631 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359957 S4360239 HotairCNE 1000 - chr12 54369205 SU212 262 H1-hESC chr12 54360239 HotairCNE 1000 - chr12 54360205 SU212 24 H1-hESC chr12 54360239 HotairCNE 1000				FeakScore	16	LIIU_IFFEak	start_rreak	CIII	Stranu		region	enu	Start	CIII
chr12 54359507 54360239 HotairCNE 1000 - chr12 54360230 SUI2 224 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54360205 SUI2 224 H1-hESC chr12 54360235 HotairCNE 1000 - chr12 54360205 SUI2 244		H1-hESC	H1-hESC	261	CHD1	54360441	54359825	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54369297 54360239 HotairCNE 1000 - chr12 54359507 54360239 HotairCNE 1000 - chr12 54359957 54360239 HotairCNE 1000 - chr12 5435957 54360239 HotairCNE 1000 - chr12 5435957 54360239 HotairCNE 1000 - chr12 5435957 54360239 HotairCNE 1000 - chr12 54360230 SVI21 224 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 5436210 SVI21 1000 GM12878,H1-hESC,Hep62,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362		H1-hESC	H1-hESC	178	TCF12	54359800	54359540	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54359507 54360239 HotairCNE 1000 - chr12 54360230 SRF 631 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54360320 SU212 24 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54360320 SU212 24 H1-hESC chr12 54362255 5436239 HotairCNE 1000 - chr12 5436520 SU212 24 H1-hESC chr12 54362255 54362956 Hota		H1-hESC	H1-hESC	319	EGR1	54360194	54359918	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54359507 54360239 HotairCNE 1000 - chr12 54359673 54359803 SRF 631 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359955 S4360245 ZNF143 158 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54359307 S4360239 H0tairCNE 1000 - chr12 54359307 S4360239 HotairCNE 1000 - chr12 54359207 S4360239 HotairCNE 1000 - chr12 54360230 SU212 242 H1-hESC chr12 54360239 HotairCNE 1000 - chr12 5436020 SU212 244 H1-hESC chr12 54360235 54360239 HotairCNE 1000 - chr12 5436020 SU212 244 H1-hESC chr12 54360235 HotairPromoter 1000 - chr12 54362600 POLR2A 204 PANC-1,SK-N-MC chr12 54362255 54362956		H1-hESC	H1-hESC	240	SP1	54359844	54359520	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54369290 54360239 HotairCNE 1000 - chr12 54359307 54360239 HotairCNE 1000 - chr12 54360230 S4360245 SVE12 224 H1-hESC chr12 54360239 HotairPromoter 1000 - chr12 5436260 POLR2A 204 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436220 5436316 TCF/12 426 PANC-1 chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436216 EH21 1000 GM1287		H1-hESC	H1-hESC	631	SRF	54359803	54359673	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54359507 54360239 HotairCNE 1000 - chr12 54359308 54360436 CTP2 262 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 5436020 SU212 224 H1-hESC chr12 54359507 54360239 HotairCNE 1000 - chr12 54365366 EZH2 1000 GM12878,H1-hESC,Hep62,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPomoter 1000 - chr12 54365260 POLR2A 240 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPomoter 1000 - chr12 54365210 54365366 EZH2 1000 GM12878,H1-hESC,Hep62,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPomoter 1000 - chr12 54365210 54365316 TCF7L2 426 PANC-1 chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362337 GAT33 234 T-47D chr12 54362956 HotairPromo		H1-hESC	H1-hESC	158	ZNF143	54360245	54359955	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54359507 54360239 HotairCNE 1000 - chr12 54360279 54360230 SUZ12 224 H1-hESC,NT2-D1 chr12 54359507 54360239 HotairCNE 1000 - chr12 54360200 SUZ12 224 H1-hESC,NT2-D1 chr12 54360239 HotairCNE 1000 - chr12 54360200 SUZ12 224 H1-hESC,NT2-D1 chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362250 5436230 SUZ12 244 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436250 543635116 TCF/L2 426 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436237 S4363316 TCF/L2 426 PANC-1 chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436237 GATA 234 T-47D chr12 54362956 HotairPromoter 1000		H1-hESC	H1-hESC	262	CTBP2	54360436	54359398	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54359507 54360239 HotairCNE 1000 - chr12 54359205 54365366 EZH2 1000 GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362600 POLR2A 204 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436250 5436116 TCF12 426 PANC-1,SK-N-MC chr12 54362255 54362956 HotairPromoter 1000 - chr12 5436205 54365366 EZH2 1000 GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362051 GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362051 GAT13 234 T-47D chr12 54362956 HotairPromoter 1000 - chr12 54362537 SUZ12 167 H1-hESC,NT2-D1 chr12 176968457 176969389		H1-hESC,NT2-D1	H1-hESC,N	224	SUZ12	54360920	54360179	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chrl2 54362255 54362956 HotairPromoter 1000 - chrl2 54362600 POLR2A 204 PANC-1,SK-N-MC chrl2 54362255 54362956 HotairPromoter 1000 - chrl2 54362205 5436210 S4362600 POLR2A 204 PANC-1,SK-N-MC chrl2 54362255 54362956 HotairPromoter 1000 - chrl2 54362305 54365366 EZH2 1000 GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF chrl2 54362255 54362956 HotairPromoter 1000 - chrl2 54362337 GATA3 234 T-47D chrl2 54362955 54362956 HotairPromoter 1000 - chrl2 5436237 SUZ12 167 H1-hESC,NT2-D1 chrl2 176968857 176969839 HoxD11CNE 1000 + chrl2 176969091 176969421 CTCF 526 A549,Fibrobl,Gliobla,HEX293,SK-N-SH_RA		GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF	GM12878,	1000	EZH2	54365366	54359205	chr12	-	1000	HotairCNE	54360239	54359507	chr12
chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362600 POLR2A 204 PANC-1,SK-N-MC chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362250 54362956 POLR2A 204 PANC-1,SK-N-MC chr12 54362255 54362956 HotaiPromoter 1000 - chr12 5436255 5436116 TCF7L2 426 PANC-1 chr12 54362255 54362956 HotaiPromoter 1000 - chr12 5436237 GATA3 234 T-47D chr12 54362255 54362956 HotaiPromoter 1000 - chr12 5436237 SUZ12 167 H1-hESC,NT2-D1 chr12 176968657 176969389 HoxD11CNE 1000 + chr2 176969091 176969421 CTCF 526 A549,Fibrobl,Gilobla,HEX293,SK-N-SH_RA														
chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362250 54363116 TCF/12 426 PANC-1 chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362956 BitaiPromoter 1000 - chr12 54362976 EZH2 1000 GATA3 234 T-47D chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362237 GATA3 234 T-47D chr12 54362255 54362956 HotaiPromoter 1000 - chr12 54362237 SUZ12 167 H1-hESC,NT2-D1 chr2 176968857 176969839 HoxD11CNE 1000 + chr2 176969091 176969421 CTCF 526 A549,Fibrobl,Gilobla,HEX293,SK-N-SH_RA		PANC-1,SK-N-MC	PANC-1,SK	204	POLR2A	54362660	54362190	chr12	-	1000	HotairPromoter	54362956	54362255	chr12
chr12 54362255 54362956 HotairPromoter 1000 - chr12 5435236 EZH2 1000 GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362337 GATA3 234 T-47D chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362337 SU212 167 H1-hESC,NT2-D1 chr2 176968857 176969389 HoxD11CNE 1000 + chr2 176969091 176969421 CTCF 526 A549,Fibrobl,Gliobla,HEX293,SK-N-SH_RA		PANC-1	PANC-1	426	TCF7L2	54363116	54362520	chr12	-	1000	HotairPromoter	54362956	54362255	chr12
chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362337 GATA3 234 T-47D chr12 54362255 54362956 HotairPromoter 1000 - chr12 54362337 SUZ12 167 H1-hESC,NT2-D1 chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176969091 176969421 CTCF 526 A549,Fibrobl,Gliobla,HEK293,SK-N-SH_RA		GM12878,H1-hESC,HepG2,HUVEC,NH-A,NHLF	GM12878,	1000	EZH2	54365366	54359205	chr12	-	1000	HotairPromoter	54362956	54362255	chr12
chr12 54362255 54362956 HotairPromoter 1000 - chr12 54361277 54362537 SUZ12 167 H1-hESC,NT2-D1		T-47D	T-47D	234	GATA3	54362337	54362041	chr12	-	1000	HotairPromoter	54362956	54362255	chr12
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176969091 176969421 CTCF 526 A549,Fibrobl,Gliobla,HEK293,SK-N-SH_RA		H1-hESC,NT2-D1	H1-hESC,N	167	SUZ12	54362537	54361277	chr12	-	1000	HotairPromoter	54362956	54362255	chr12
		A549 Fibrobl Gliobla HFK293 SK-N-SH_RA	A549.Fibro	526	CTCF	176969421	176969091	chr2	+	1000	HoxD11CNF	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968367 176975129 F7H2 1000 H1-hFSC HenG2 HMEC HSMM HSMM tube HUVEC NH-A NHDF-Ad NHK N	HDF-Ad.NHFK.NHI F	H1-hESC.HepG2.HMEC.HSMM.HSMMtube.HUVEC.NH-A.NHDE	H1-hFSC.H	1000	F7H2	176975129	176968367	chr2	+	1000	HoxD11CNF	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968464 176968800 POLR2A 382 SK-N-MC		SK-N-MC	SK-N-MC	382	POLR2A	176968800	176968464	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968988 176969498 POLR2A 471 A549,SK-N-MC,Gliobla		A549,SK-N-MC,Gliobla	A549,SK-N	471	POLR2A	176969498	176968988	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968889 176969429 EP300 381 A549		A549	A549	381	EP300	176969429	176968889	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176969019 176969197 FOSL2 813 A549		A549	A549	813	FOSL2	176969197	176969019	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968813 176969389 TCF12 356 A549		A549	A549	356	TCF12	176969389	176968813	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176969043 176969353 SUZ12 96 H1-hESC		H1-hESC	H1-hESC	96	SUZ12	176969353	176969043	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968819 176969335 TCF7L2 185 HEK293		HEK293	HEK293	185	TCF7L2	176969335	176968819	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2
chr2 176968657 176969389 HoxD11CNE 1000 + chr2 176968726 176969412 KAP1 329 HEK293,U2OS		HEK293,U2OS	HEK293,U2	329	KAP1	176969412	176968726	chr2	+	1000	HoxD11CNE	176969389	176968657	chr2

inen 10. juuri 10. maa juuli juuri Tyykö 2. Maalailii 10. kilkiilii 20. kai Anto pasameen, mit juuri 10. maa juuli juuri Malli, 2011, 2014, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018,
Mar Albertales is republic data bahara Manie Symphy ng ummu ni jentich Childhild data data Mar Albertales is mender data bahara Manie Symphy
idae diferentiaten in providen dialiti barberra Mensie Joprophy idae diferentiaten in providen dialiti barberra Mensie Joprophy idae diferentiaten in menden dialiti barberra Mensie Joprophy
line differentiation of reproduct daplet control line differentiation of reproduct daplet barborne discussion (symmetry effects and second second second
inar alfanninin (n. panin, nafit an alfanninin (n. panin, taki bahana Manin Jamaja ina alfanninin (n. panin, taki bahana Manin Jamaja
hine differentiation to reportion dealer backware because by many
inizia da la constante de la constante de La constante de la constante de
no, marine proposition (discrimination) and the second sec
nan, ana N. Marine, Janni art, Maarin, Gale, Similari, Janay art, Marine, Gale, Similari, Janay
ining di Panenini ng Kapanang dagi kunani Kang di Panenini ng Kapanang dagi kunani Kang di Panenini ng Kapanang dagi kunang
n/ yanar yali jewili kiki Obusi BO 2006 2008 Najahalana yali jewili kuki Kiki Obusi BTI kili kiki Kan diferentakan iki menden dariki centri
Nadar yanisana yak ina 1962 McGalashi Katala 1960 samalamat yanar yak ina Pa 1991 2 Okustri Si 1983 Alekt Malashi yanar yak ina Pa 1991 2 Okustri Si 1983 Alekt
a material faith year part for material index materials and the internal part for AUX sets well as sensitive of sense part for PX APPE & Object 2 as about 2 as about 2 as ab
ndigeni (kan juk jenini jen ndigeni (kan juk jenini j
, nerve etc., maarin, jaardi, jaardi, 2004 il 2005, il 2006, il 2005 mit, Maarin, Sark
nnyn palentariana 11 parlinaria pal jine tellä sellä Silla Sekusifiki sille ikisiki 11 kompad jinen plako petikilaat
a yang dina ning yak yang meng yak berdini pilihi dikerakili nilihik alimi. Internali ketiy yake malanamata intertere danti.
ann, Salh anhymi, nam, aith, amhlinaí
Carriera de la constante de la
nyana jul jauna katalan katalan nya nakanan jul jauna katalah katalah
part yishiannai yur jali jirainta LObaidhia ilala ilali 1. ying perimeta jali jiraina jil Obaidhia ilalia Internet Judy jali, palameta jiriaina jila
samalama yanar yak bartalari. Obartalari alifa 1994 ahalama yak bia Abra da Obartala 1995, ashi Jahat Jartad yak bartalari, bir Jahashi Abra alifa 1995, ashi
ninger och 11 na men och sen Oberlählster a star men Oberlählster
andalaman yadi jiwa Marea Mili Oniya Mili a Mili a Mili a Jana Jana yangi diga sangi diga salaha salaha salaha
endepend plant, salls, adapted endepend plant, fails, adapted
n de fondergen fondelen de fondelen de fondelen en de fondelen de fondelen de fondelen en de fondelen en de fon En de fondelen de fondelen de fondelen de fondelen en de fondelen en de fondelen en de fondelen en de fondelen e
, pang pang disertah kelamat ang kelamat ing Pagman jadi Interpat (kela pada, pada majar pada na pada na pada
nie party services party for any party of a constant of a constant is party mark defended alles & and a constant marks party any distant defended of a constant of a constant of a
en hymal yn en ywla, adigana y Antani y adgegenia y nalanian, dagad annael nema y ad y hawlyd y a Chlor iddill ad Nai, ad AC Add y fenlargan y refamion, ddir
en hynd yn en yn hyndraeu y ferfan Cydraegoriu, yn hynnyr ferfan Mae'n yw y fereniau y defan a'r ferfan a'r fer Mae'n yw y ferfan y ferfan y ferfan a'r ferfan
lakai (serinana, sel (na Julai), (Milili sila delegata pastera antyrad part, pelo alignac, latas (adargario, industro) iki himo.
endyna (hen fak Hannis, idy 4 juni (hennes) dialan (hennes)
interprit (only, national and the state of t
e part particular part de la constantia de Constantia de la constantia de la constanti Constantia de la constantia de la constanti
enegas jagta Jahond Jahond Just Jawe (1964, a., 2016, jokanson (1964, 2016), 2016), 2016) and yord Jahon, Salik, Jama, Marrina
lenieryski (konty, nale, poslanovych, justanian, pisyki nal, posoba, pistana, posoba mal, parasena, pali jeszetta, pisie Tables a antik, aktika, aktika
LAND, Contemport, Johannan, Milor Malar, Jain, Jacomal Jacoma, Jacoma,
ling pines parter part, tras Welgell Debusit CE states, ability embyrnel parts, arlyses, derived, arlingeric, industries, fribelisteric and part of the states of the stat
nan gart pa prayan jaranan jai pa ka na ka na ka na k na ki jah Integri (hit, nih, na mangir, istaria, ispit
i Semeniar Johannia Jahr. A Semeniar Johannia Jahr. Nalai Jah
alaan Maaroone yuli joortheedh Olina siim siddi alafix 2 amaalaanna maadi alih aaniba alif ad bD fahr
4 Meanglei (Alb. Guidensenhams, och Jose Mellet Devaritien (2012) 2019 Abbierens, och Jose (MC 2010) 2012 2013 2013
ani, graphena jad, barikat Okinacola sido't sidot Ininysti, bally sele, minimuto jatantiar, dapi? Mart 100au Masenbrad
armente, cali fone fair d'O d'Anna Mille a d'Ali a state Rice par agriterical 2 calis transmit a del fonge mille
nganganak paté pangananganingki at Maringki kabupatén pagai partanangan Janahanan Mila daharaké at Kasaké pagai partanangan Janahanan Mila daharaké at Kasaké
Prost, party rail, bu, reports (n. 1671, 1071/1010, 201 2) pressures a party still, and the party and bill page (n. 1711) and the party of the pa
nant, yanan yakan da katanan (Sang Ding yaka da katanan katanan) Sang yaka yaka da katanan (Sang Ding yaka da katanan) Sang yaka yaka yaka yaka yaka yaka yaka yak
(a) a set of the part of th
paparana pal parinana pal juariaksi biraksi 20 kialasi kilik nana pal juarisi biraksi 0 kiaksi 2011 kilikk kiliki A palo pado pado
emany personany, and personalise debuilded and de
interpit (only pub, reclements (othertor, dapit) Meet Jammal 2 second meets which and hit and hit? Statistics
entymi pen jelo alpen, kriet jelongeri, isketer jihtilen nitymi pennen jelo anter jene jijti penj fanne stani
ning and general Developed and a static static Mill (Interpret) America, Mile 1 alsh and (Developed School and a
n, pala panti dina dalah dalah dalah m, basi panti dina dilan dalah - Jammarata Jamai path, manka pati, pad Jali Jalir
ny jedné jedník Odmáři GO 2001 I 2002 najše jemeckul A Starbani Jajskala Stala
 Jonanty participant, Jane programme, pre (1015), (Alternative, State - Communication, proceed painty, according painty, and (1017), Salar structures, the URI.
éperpir, sakadarana Interpirt katiy naka parlamentri jatantan dapit Anton dan baha parlamentri kati dalamentri dapit
) pasawarang baang pala panda pali pali kiti pilakinin.) pasawarang baang pila panda pali pali pilakini pilakini pilakini pali pilakini pilakini pali pilakini pilakin
 Commerciales, Canada, party paralitas, party party bally player Commerciales, Consoli, party paralitas, party party bally player Consolidad, Sales, Saleshinar, party
ngangag penamu 2 penamunang penang pala panahis pala pel July July I 2 penamunang penang pala panahis pala pel July July July Interiore
dilan yalan yalan yanan ke yangku 1 yana maranan yanan yalah yana ka yala yala yala yala yala yala yala
pland parameterial sense will be to be able about 2000 ables mail sensena pail (na 4001 040-020 ables ables plane sense ad beautiful (decast) ables ables
need on the second s
ng panan pang pang banakan pang pang pang pang pang pang pang pa
landning pelenseninens pel jine fillstaffille Obrail fill alleri andra anleinens pel jine Mers III. Obrail 100 alleri. 1019 el nemer tel line TTERNONA 1855 alleri 1865
ilius jumphali andynal prawne ynt bene name inne ninadeneninerse oet inw MCM/Obui 80% 80% 80%
in Palpenn, fall andyrnal, nem, selv, adamar, derinad, adamperis, industan, fölletilmin- idiger differentiand
e (polieta (se). necessa pel journi del Constanta all'Altanto L'Anno estato del los constantos (20), dividinte test
ner, same val beskriet Oberschie sinte all No (Norma) ander Janiela jul
mana juan, yangkara Malanama juai, kuraha Li daharaka adal Labha 1. Jawar juanan juai, kuryangana juryahi jilakalania, kur
nais pai panaini di Constant ad St. 2016. Anis mbyosi panana pai panin pana jin pany
a panananana panani piri panahi piri pal kilopin dina. maga jiri malan panana pal jiro papis palisi
nyte, beuat nyte, partunal into: Esthuliat Selb
nyhäisenä jait jos päällää periitä ari, ynyhens jait jos kaikikiheisiittä sissä sissä arida joik arav
Holda (all) Microsovier Holda (all) Terreto
ining lan, lan Marjuhannay Marjuh Johnshia, nyamina
interpisi (sala, pala, palamento, interior, dapit 1.69 (interpris (san pala, partiempopis (mianten, dapit Autor, jaonal (spila, talo
Julick Deer Just Deeth Sale
eta Junitera Calita Janoste
nte (Instance, Gale, Aportyn), erychwidiante, Jandarsta, programae, pre Janeira, Jailer Kirnin), denaard, parenter, pref (Jane, programae, pref 1997), jailweidinense, Janei) - Benaard, annese one of the seconderse one (1997), dielerchinen best
ne (belancy) also, by anyone (any beneficiant), by also many any granters (are, benefic, balanciant), programmer (are, benefic, balanciant), pro- anti-anterna (are) (are, anyonena (are, balanciantes), para (area (area (are
er, Senter, Santo, Santo, Santo Santo, Santo, Santo, Santo, Santo, Santo, Santo, Santo, Januar, Januar, Juli, Jan, Santonan, S., Kit, Mahdaha, Jani Santo, Januar, Jano, Jan, Santonan, S., Kit, Mahdaha, Jani Santo, Januar, Jano, Jan, Santonan, S., Kit, Mahdaha, Jani Santo, Januar, Jano, Jan, Santonan, S., Kit, Mahdaha, Jani Santon, Januar, Jano, Jan, Santonan, S., Kit, Mahdaha, Jani
ατό μπό τη μεγάλλη τη μαριτηρία τη τη μαρία τη