

lncRNAs

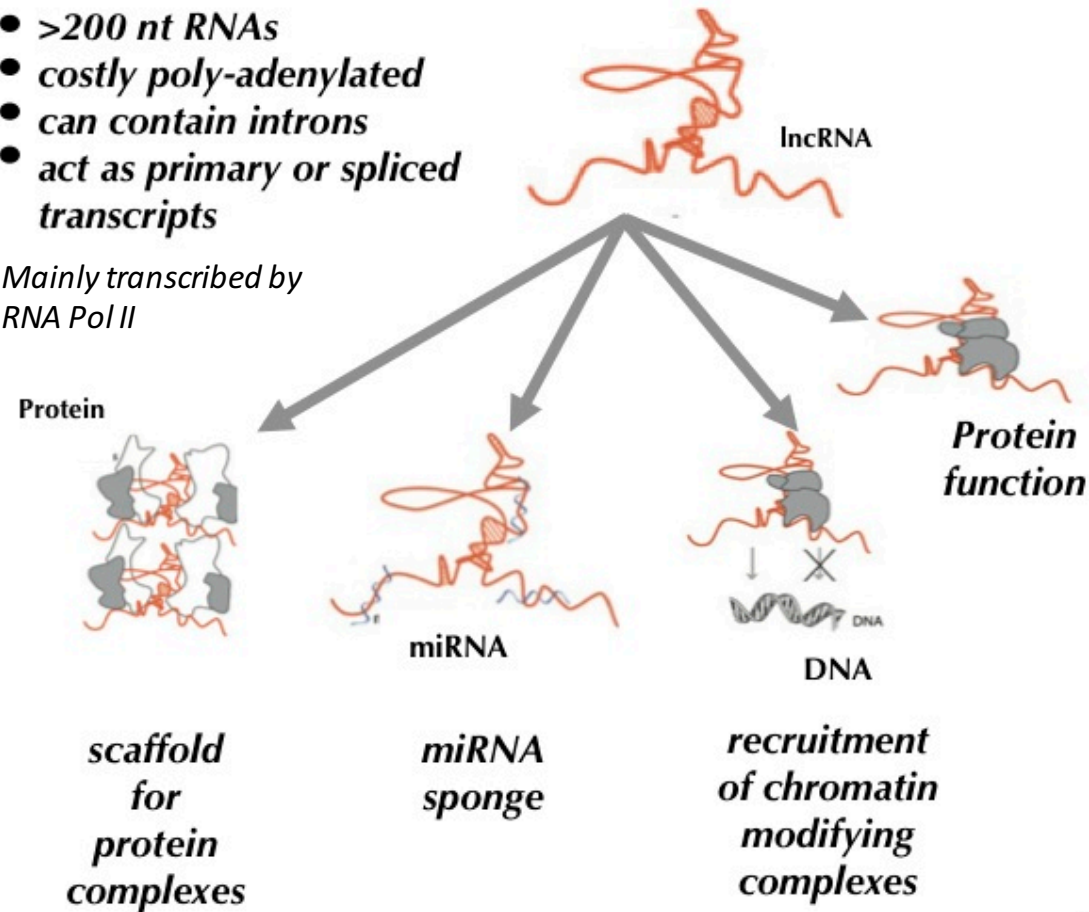
long, non-coding RNAs

Characteristics of lncRNAs

Long, non-coding RNAs (lncRNAs)

- *>200 nt RNAs*
- *costly poly-adenylated*
- *can contain introns*
- *act as primary or spliced transcripts*

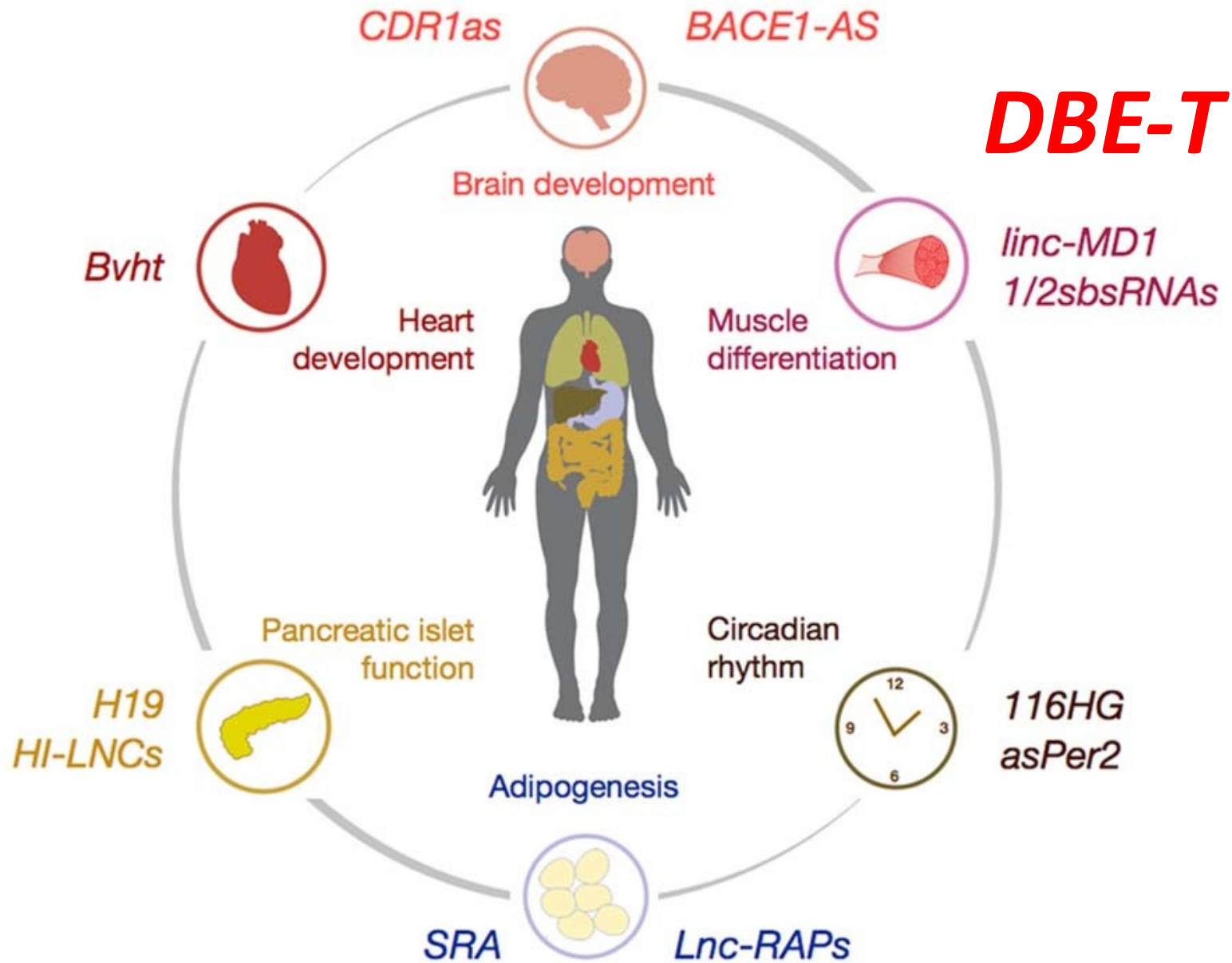
*Mainly transcribed by
RNA Pol II*



**Protein
function**

Nuclear
and/or
cytoplasmic
localization

lncRNAs in physiology



Resource

The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression

Thomas Derrien,^{1,11} Rory Johnson,^{1,11} Giovanni Bussotti,¹ Andrea Tanzer,¹ Sarah Djebali,¹ Hagen Tilgner,¹ Gregory Guernec,² David Martin,¹ Angelika Merkel,¹ David G. Knowles,¹ Julien Lagarde,¹ Lavanya Veeravalli,³ Xiaoan Ruan,³ Yijun Ruan,³ Timo Lassmann,⁴ Piero Carninci,⁴ James B. Brown,⁵ Leonard Lipovich,⁶ Jose M. Gonzalez,⁷ Mark Thomas,⁷ Carrie A. Davis,⁸ Ramin Shiekhattar,⁹ Thomas R. Gingeras,⁸ Tim J. Hubbard,⁷ Cedric Notredame,¹ Jennifer Harrow,⁷ and Roderic Guigó^{1,10,12}

¹Bioinformatics and Genomics, Centre for Genomic Regulation (CRG) and UPF, 08003 Barcelona, Catalonia, Spain; ²INRA, UR1012 SCRIBE, IFR140, GenOuest, 35000 Rennes, France; ³Genome Institute of Singapore, Agency for Science, Technology and Research, Genome 138672, Singapore; ⁴Riken Omics Science Center, Riken Yokohama Institute, Yokohama, Kanagawa 351-0198, Japan; ⁵Department of Statistics, University of California, Berkeley, California 94720, USA; ⁶Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan 48201, USA; ⁷Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, United Kingdom; ⁸Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; ⁹The Wistar Institute, Philadelphia, Pennsylvania 19104, USA; ¹⁰Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, 08002 Barcelona, Catalonia, Spain

COMPARING mRNAs - lncRNAs

lncRNAs can be

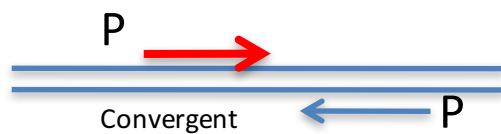
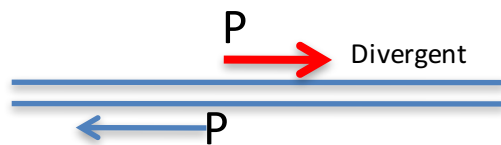
1. Intergenic (lincRNA):

Do not intersect with protein coding gene

2. Genic:

Intersect a protein coding gene

- Exonic
- Intronic
- overlapping



Closest protein coding gene

lncRNA

Intergenic lncRNA

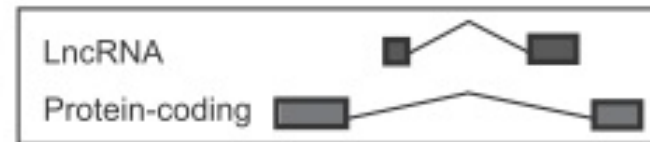


Genic lncRNA

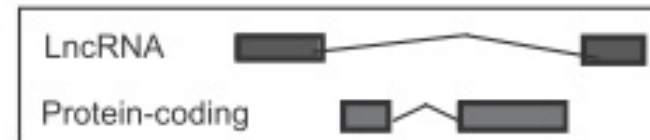
Exonic



Intronic



Overlapping



Gencode lncRNAs transcripts (14,880)									
Intergenic (9,518)					Genic (5,362)				
Same Strand	Convergent	Divergent	Exonic (2,411)		Intronic (2,784)		Overlapping (167)		
4,165	1,937	3,416	S	AS	S	AS	S	AS	
			NA	2,411	563	2,221	52	115	

P: promoter

COMPARING mRNAs - lncRNAs

A. lncRNAs do not have coding potential:

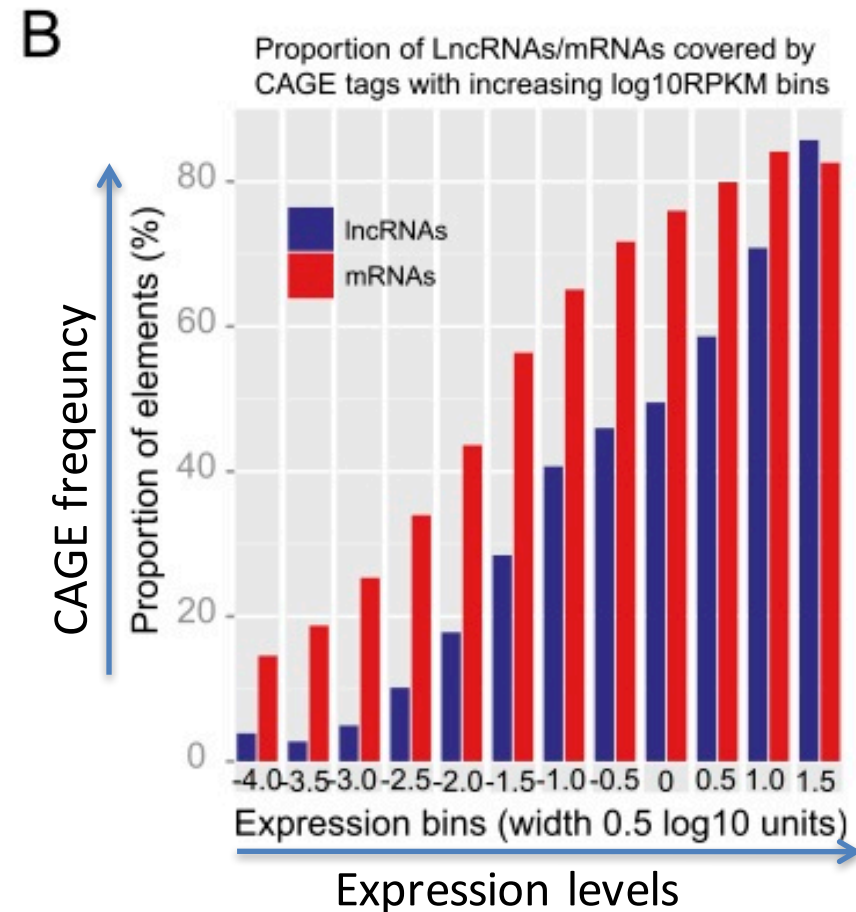
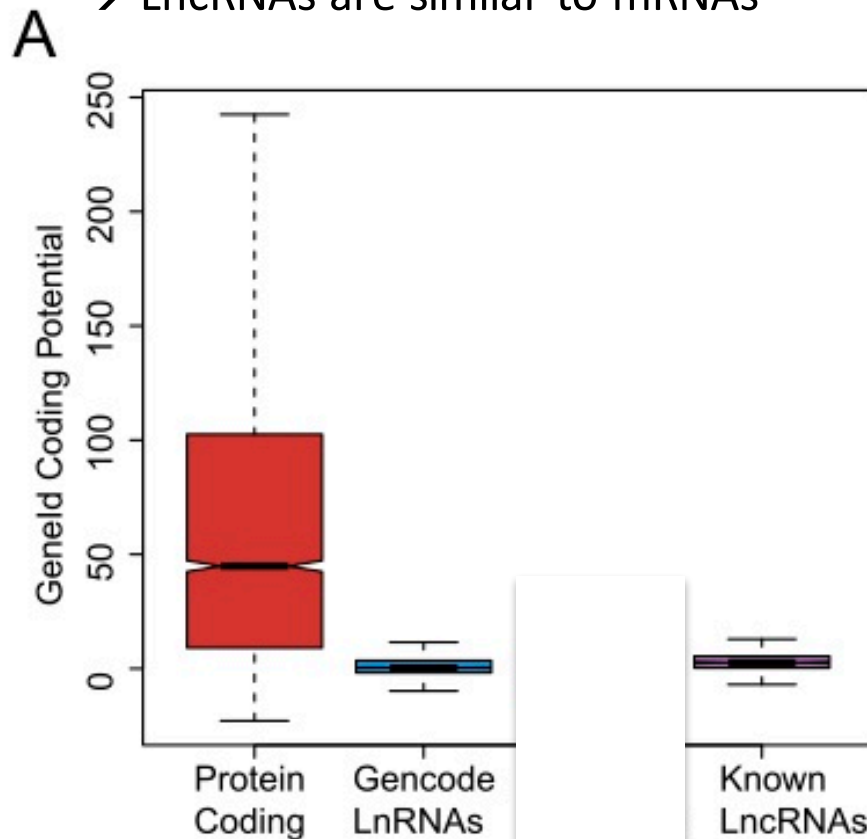
→ Longest possible ORF was searched in mRNA/lncRNAs

B. CAGE tags can be found in lncRNAs → defined transcriptional start site

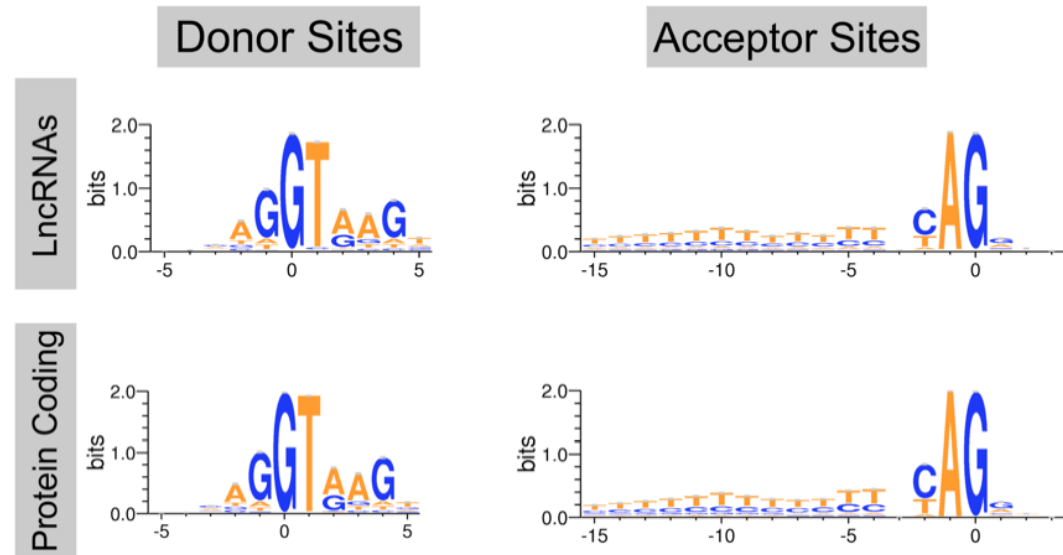
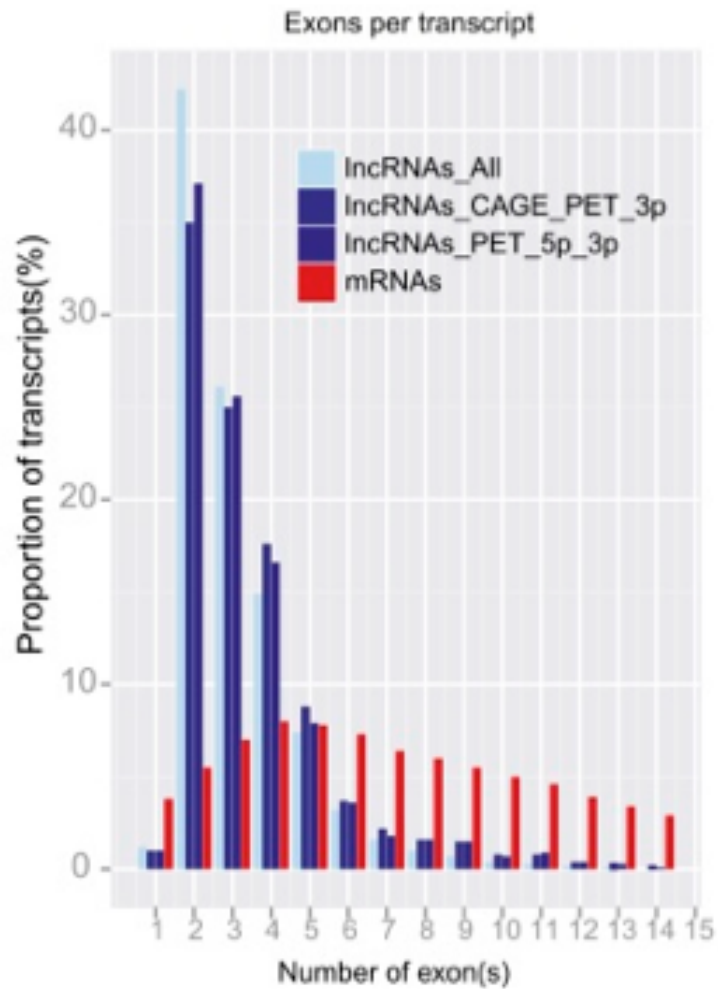
CAGE tag frequency increases with increased lncRNA expression levels.

mRNAs are characterized by more CAGE tags

→ lncRNAs are similar to mRNAs



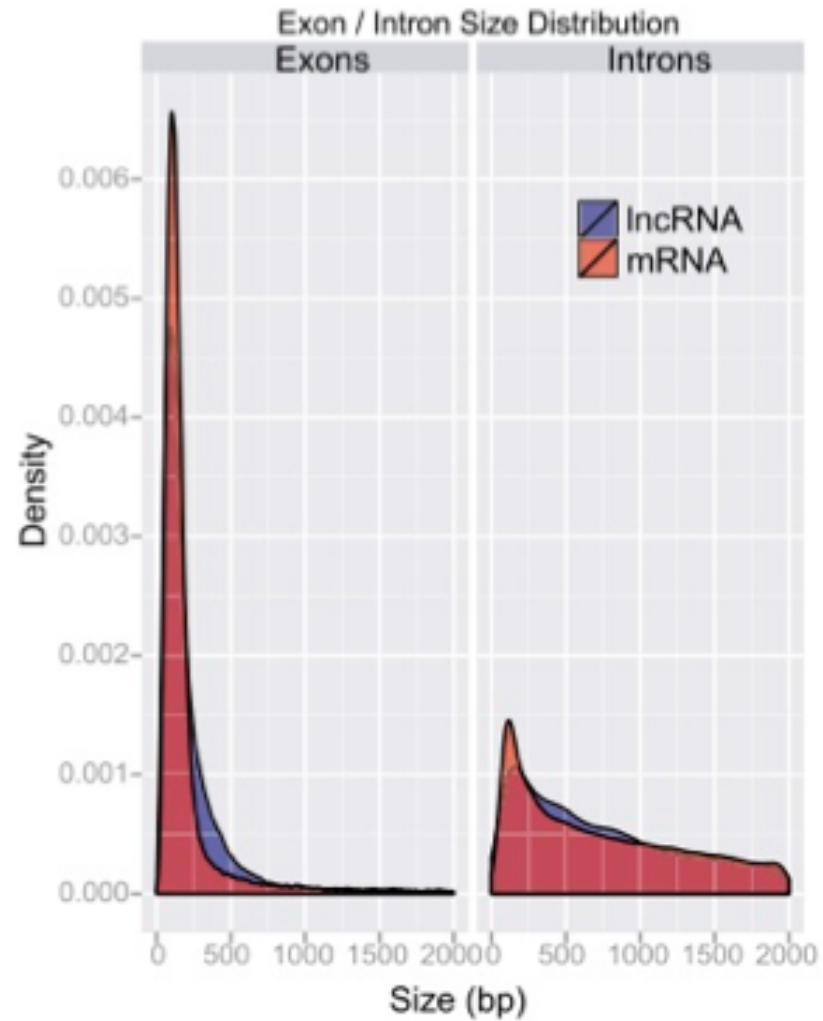
COMPARING mRNAs - lncRNAs



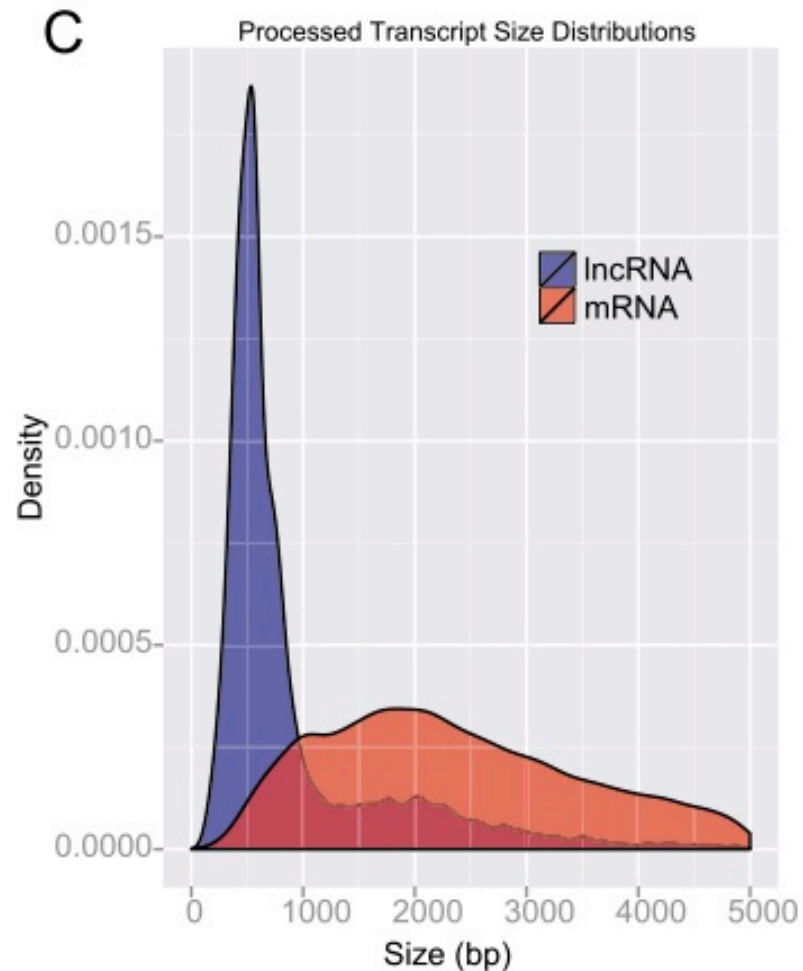
SPLICING: lncRNAs are spliced, relevant
Splice-site prerequisites at splice donor/acceptor are conserved
BUT: lncRNAs contain fewer INTRONS!! most lncRNA have only 1 intron!!!; mRNAs 3-6

COMPARING mRNAs - lncRNAs

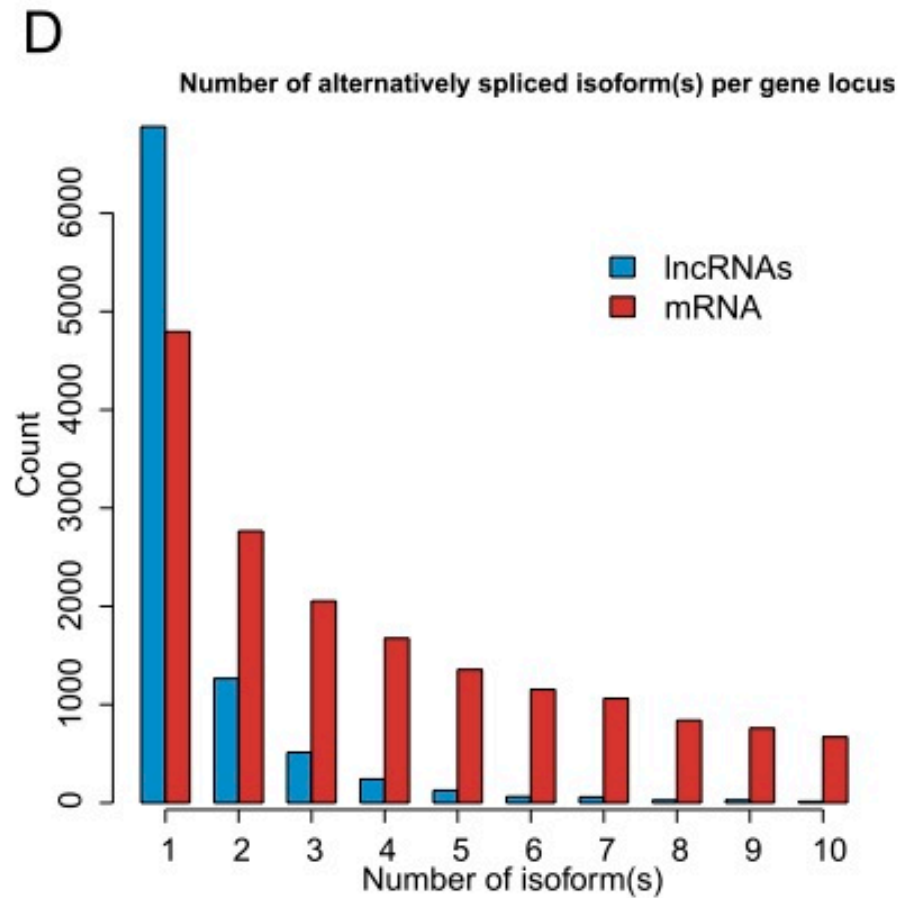
Introns/Exons from lncRNAs are slightly longer



COMPARING mRNAs - lncRNAs



lncRNAs are on average much shorter:
ca. 500nt
mRNAs are longer and have wider size
distribution



lncRNAs are uniform → little alternative
splicing
mRNAs: large variety of alternative splicing

COMPARING mRNAs - lncRNAs

EXONS:

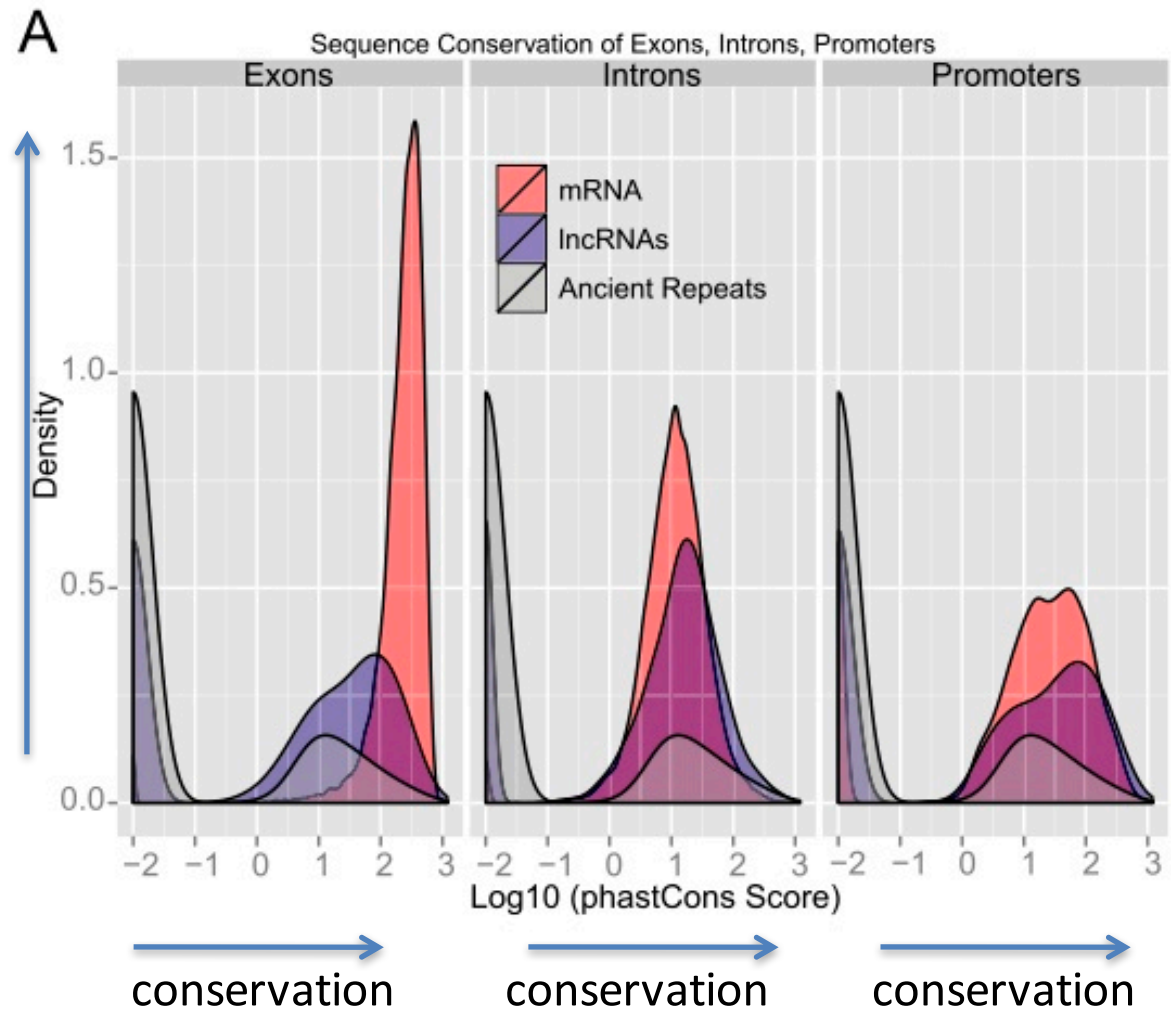
mRNA: high conservation

lncRNA: reduced conservation

But: conservation is higher than mRNA intron conservation

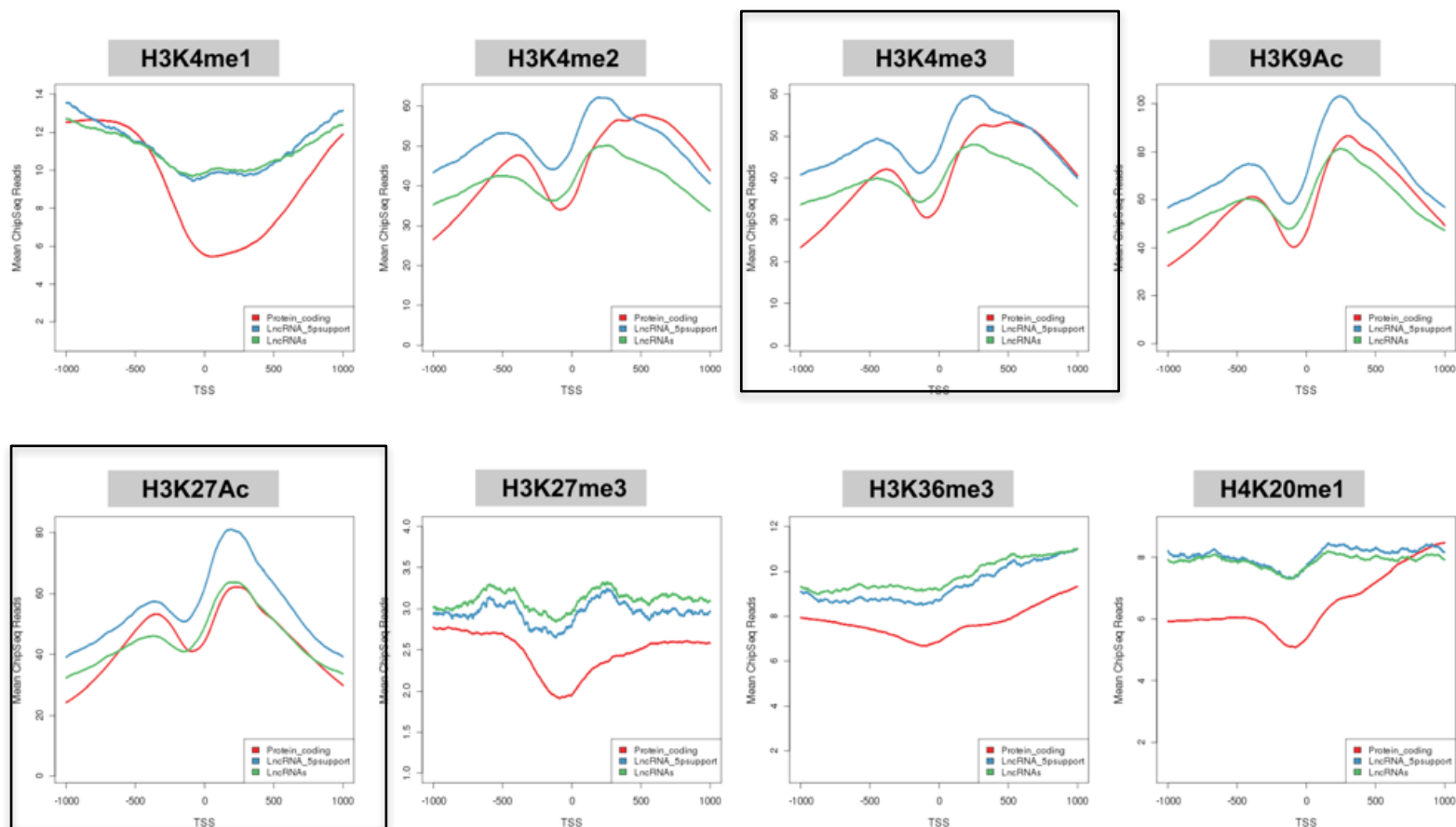
INTRONS:

mRNA: higher conservation than lncRNAs



COMPARING mRNAs - lncRNAs

Chromatin signatures at transcriptional start sites are conserved between mRNAs
And lncRNAs: MOST lncRNAs ARE TRANSCRIBED BY RNA Pol II



Chromatin signatures around TSS of protein-coding and lncRNA transcripts expressed in the same cell lines where the signatures were monitored by ChIP-seq. Shown on the y-axis is the average density of reads covering the TSS of various gene sets, with position plotted on the x-axis (bp relative to positive strand TSS). Protein coding genes are plotted in red, Gencode v7 lncRNAs in green, and lncRNAs with 5' experimental support (n=2,793) in blue. N.B. A more extensive analysis of histone modifications in multiple cell types is available at http://big.crg.cat/bioinformatics_and_genomics/lncrna_data.

K562

Not shown:
RNA Pol II peaks at transcription
start site of most lncRNAs

COMPARING mRNAs - lncRNAs

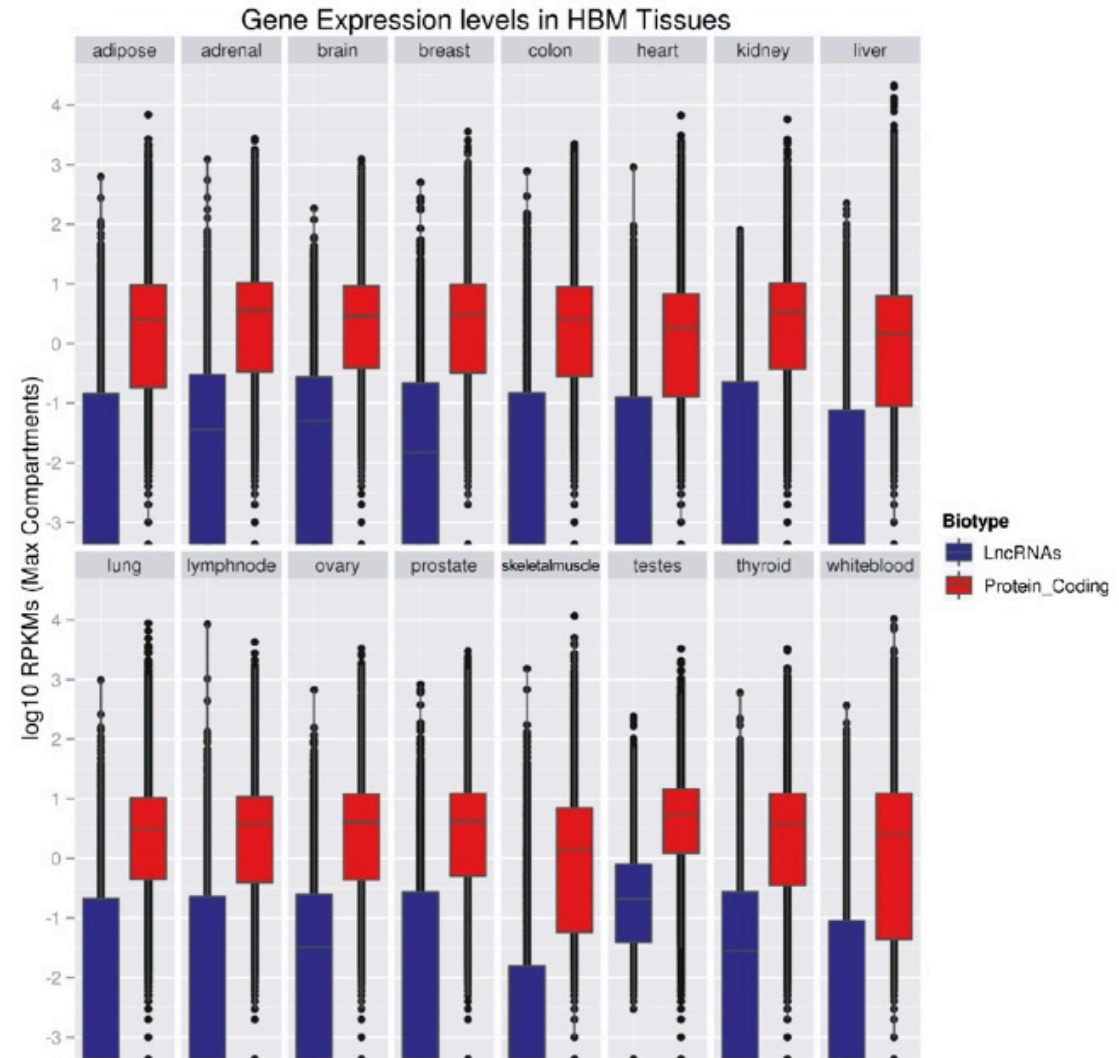
Absolute expression of lncRNAs is much lower than mRNA expression

Most lncRNAs localize to the Nucleus and chromatin

Many lncRNAs (30%) are PRIMATE SPECIFIC

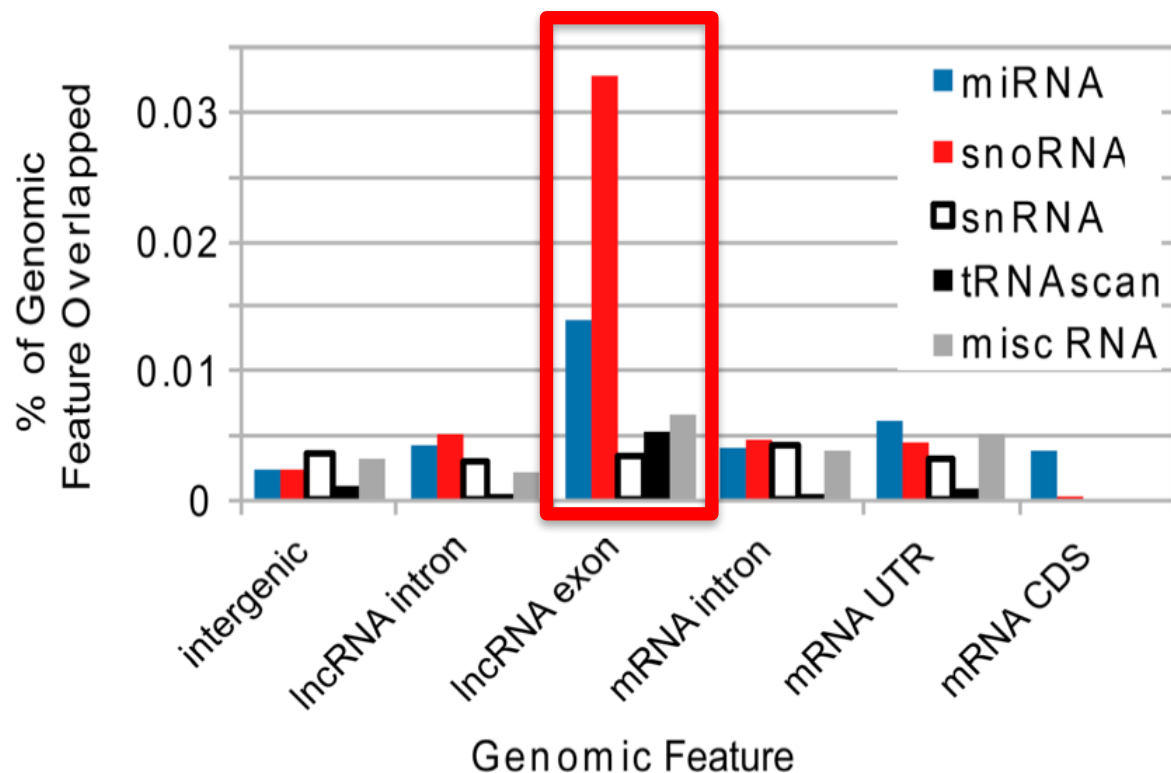
→ Evolutionary advantage

A



B

COMPARING mRNAs - lncRNAs



lncRNA exons
 Have a remarkable
 Potential to encode
 functional small RNAs
 (sRNAs)
**HIGHER THAN EXONS OF
 mRNAs!!!!**

snoRNA: Small nucleolar RNAs (snoRNAs) are a class of small RNA molecules that primarily guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs.

snRNA: Small nuclear ribonucleic acid (snRNA), also commonly referred to as U-RNA, is a class of small RNA molecules that are found within the splicing speckles and Cajal bodies of the cell nucleus in eukaryotic cells.

tRNA: translation

miscRNAs: MiscRNA is short for miscellaneous RNA, a general term for a series of miscellaneous small RNA. It serves a variety of functions, including some enzyme-like catalysis and processing RNA after it is formed. Besides, some of these small RNAs may serve as switches.

Figure S7: lncRNAs as a source of small RNAs. Shown is the proportion of nucleotides in exons/introns/(UTRS) from protein-coding genes and lncRNAs that overlap different classes of annotated small RNAs. Note that all values refer to cases where the small RNA is on the same strand as the indicated lncRNA/mRNA.

COMPARING mRNAs - lncRNAs

lncRNAs represent a big class of functional elements that

- controlled gene expression
- are processed
- lack protein coding potential
- defined localization
- frequently encode sRNAs
- low overall conservation
- 35% of lncRNAs are primate specific
- expression is rather low – but controlled!
- mostly transcribed by RNA Pol II

FUNCTION: FOR THE VAST MAJORITY OF lncRNAs THE BIOLOGICAL FUNCTION IS UNKNOWN!!!!

Examples of lncRNAs

DBE-T AND GENETIC DISEASE

THE ROLE OF THE lncRNA DBE-T IN FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY (FSHD)

FSHD: <http://www.omim.org/entry/158900>

Facioscapulohumeral muscular dystrophy-1 (FSHD1) is associated with contraction of the D4Z4 macrosatellite repeat in the subtelomeric region of chromosome 4q35.

The genetics of FSHD is, however, complex and detection of a D4Z4-reduced allele may not be sufficient for diagnosis.

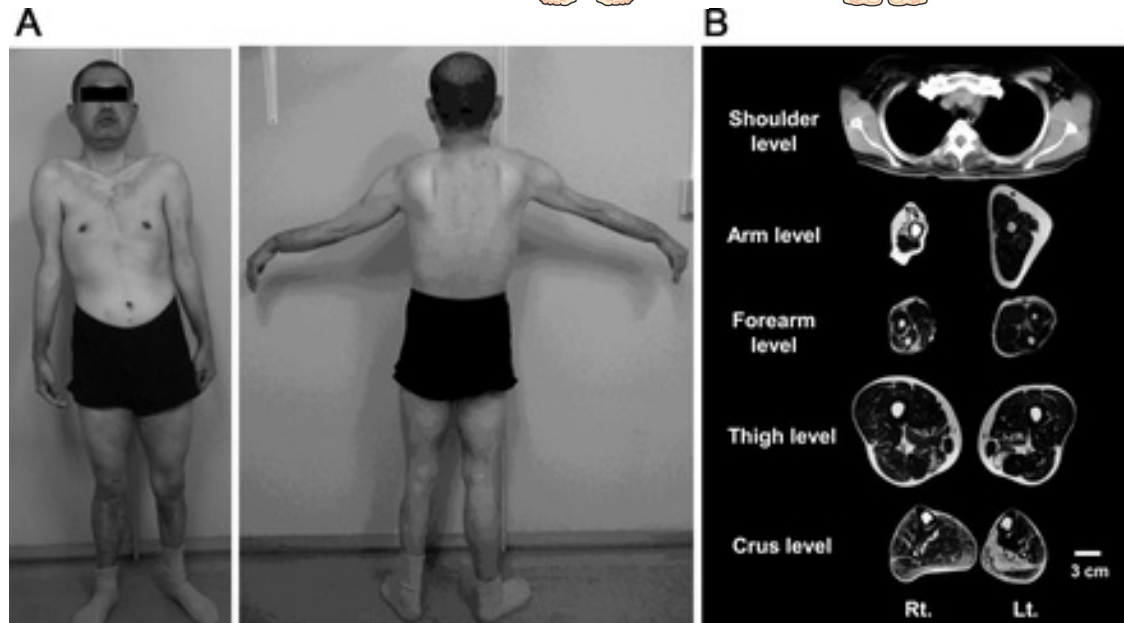
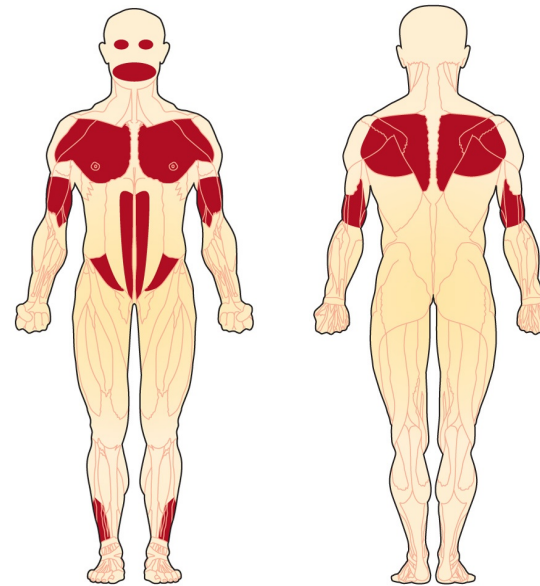
Facioscapulohumeral muscular dystrophy (FSHD) is a genetic muscle disorder in which the muscles of the face, shoulder blades and upper arms are among the most affected.

The long name comes from *facies*, the Latin word and medical term for face; *scapula*, the Latin word and anatomical term for shoulder blade; and *humerus*, the Latin word for upper arm and the anatomical term for the bone that goes from the shoulder to the elbow.

The term muscular dystrophy means progressive muscle degeneration, with increasing weakness and atrophy (loss of bulk) of muscles. In FSHD, weakness first and most seriously affects the face, shoulders and upper arms, but the disease usually also causes weakness in other muscles.

Aberrant expression of genes in vicinity to D4Z4 repeats, including DUX4, ANT1, FRG1, FRG2 in FSHD patients are thought to mediate the syndrome (have a “toxic” effect)

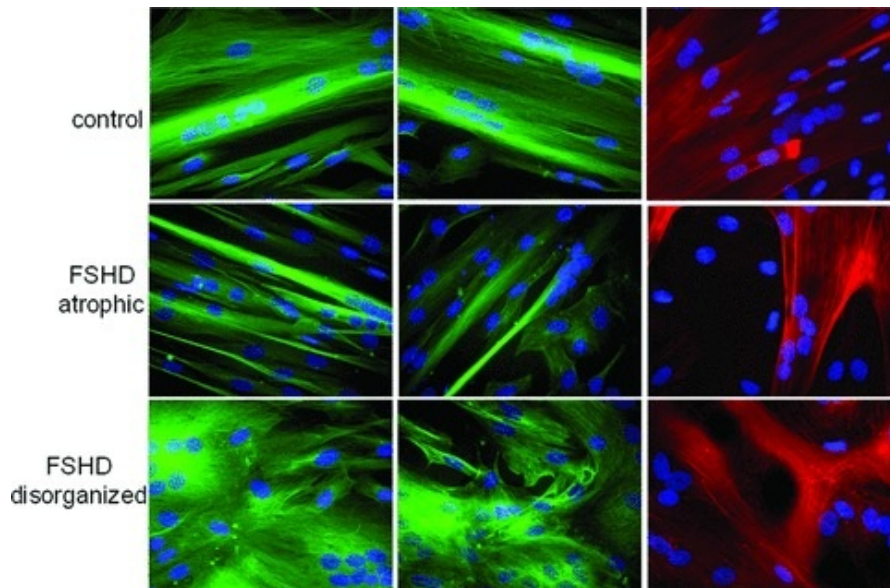
**Muscles affected
By wasting in
FSHD patients**



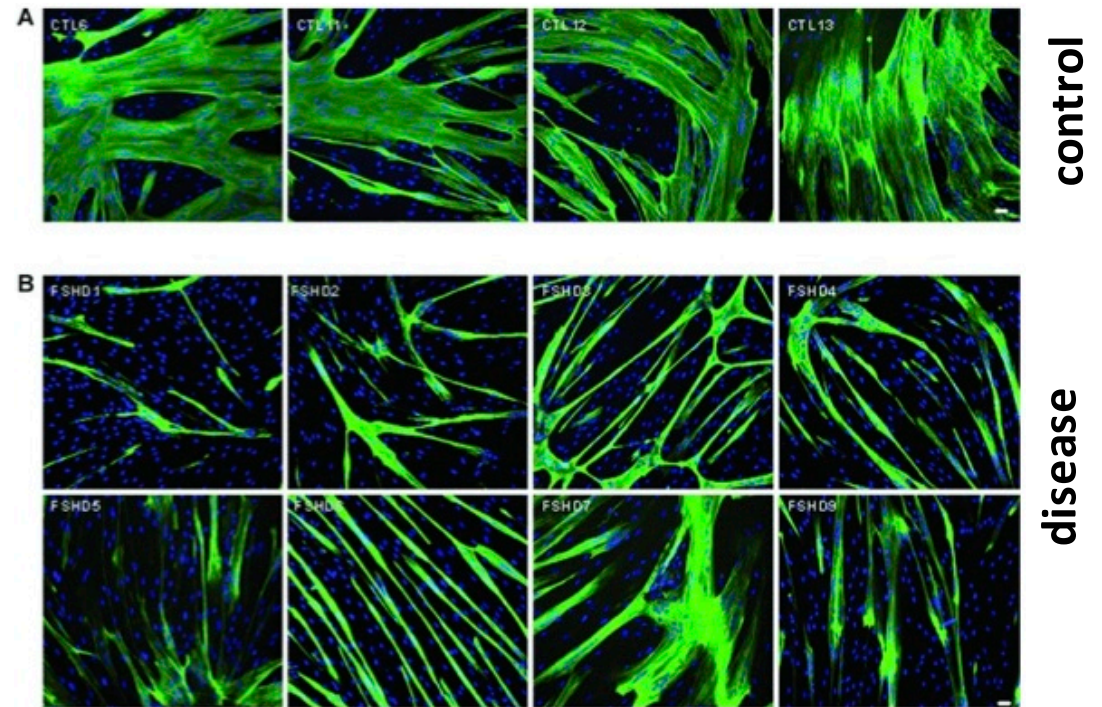
FSDH impairs muscle cell function

Moreover, in patients affected with FSHD, it is quite common to observe the co-existence of affected and apparently healthy muscles. Myoblasts, which were obtained from muscle typically affected in FSHD, manifested an increased susceptibility to oxidative stress during proliferation. Myotubes obtained from patient/healthy myoblasts show abnormal morphology and muscle marker expression

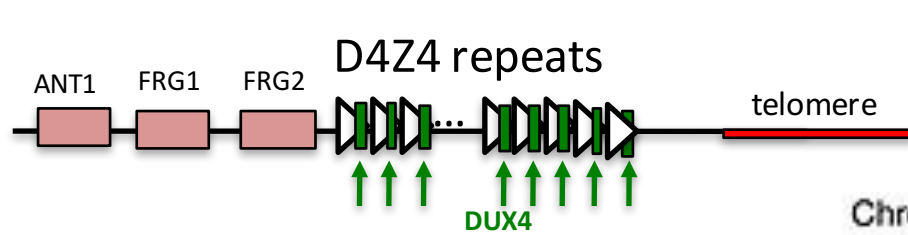
Anti **actin/tubulin** immunostaining of myotubes obtained by differentiating myoblasts cells isolated from healthy or FSHD patients



anti-**troponinT** immunostaining of myotubes obtained by differentiating myoblasts cells isolated from healthy or FSHD patients



D4Z4 repeats in Chr. 4q subtelomeres give rise to various transcripts



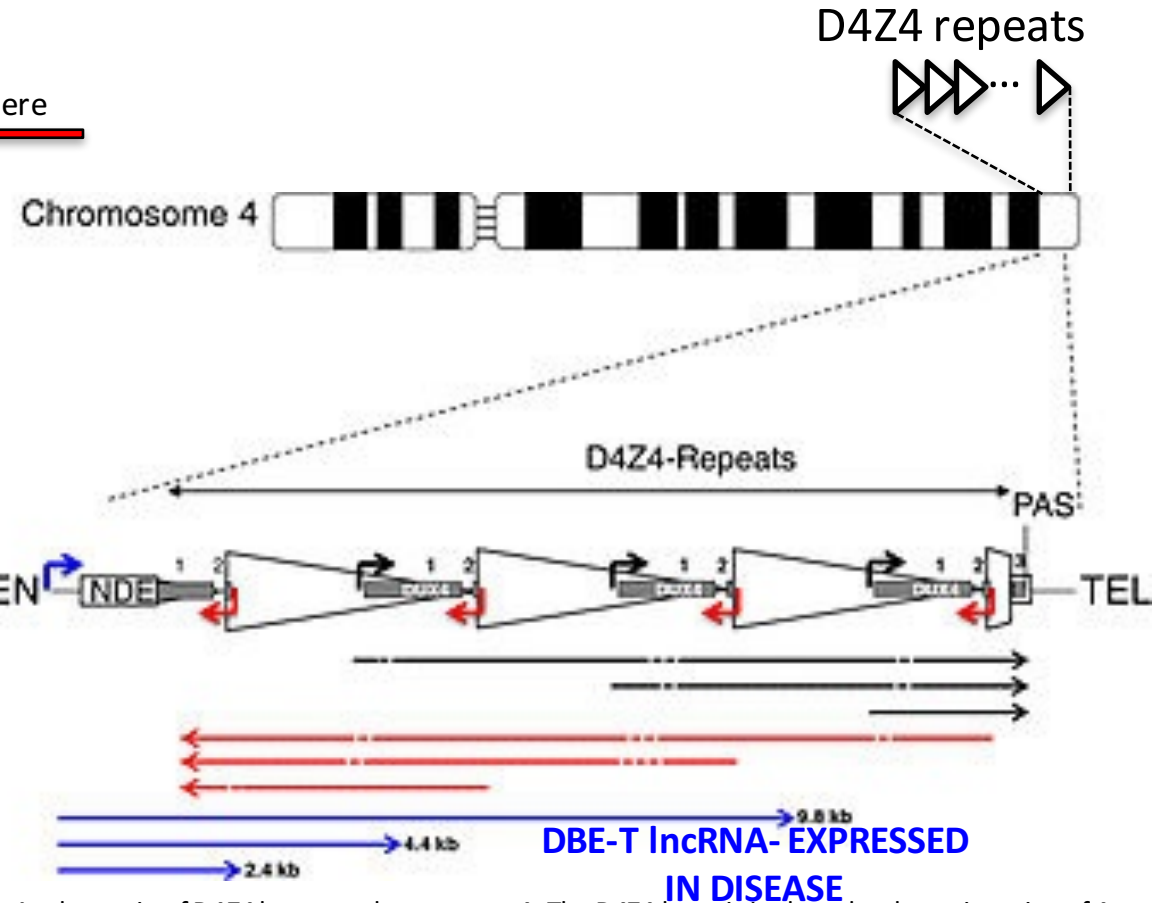
D4Z4 repeat: ca 3kb; extremely GC-rich

Healthy: 11-150 D4Z4 repeats:

ANT1, FRG1, FRG2, DUX4: *repressed*

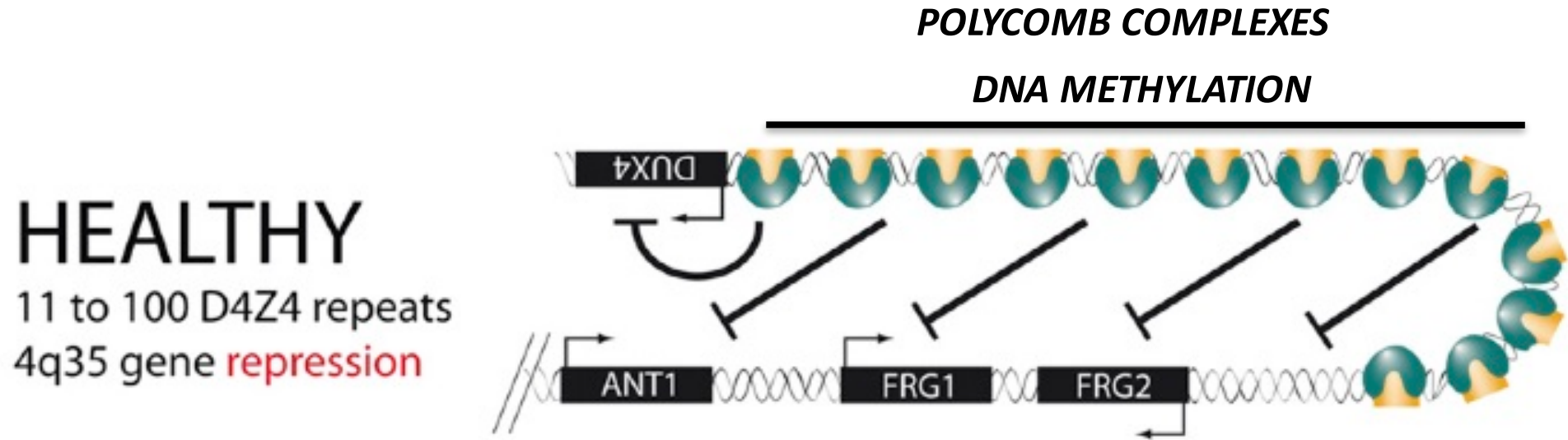
Diseased: 1-10 D4Z4 repeats:

ANT1, FRG1, FRG2, DUX4: *expressed*



A schematic of D4Z4 locus on chromosome 4: The D4Z4 locus is in the sub-telomeric region of 4q. The figure shows a three repeat D4Z4 array. CEN indicates the centromeric end and TEL indicates the telomeric end. The DUX4 gene is shown as a gray rectangle with exon 1 and exon 2 in each repeat and exon 3 in the pLAM region telomeric to the last partial repeat (numbered 1, 2, and 3). PAS indicates the polyadenylation site on the permissive 4qA allele that is not present on the non-permissive 4qB allele or on chromosome 10. The arrowed lines represent: **Blue, DBE-T transcripts (2.4, 4.4, and 9.8 kb) found in FSHD cells and reported to de-repress DUX4 expression;** Black and red, transcripts in the sense and antisense direction were detected in both FSHD and control cells and might originate from the mapped sense promoters (black) and anti-sense promoters (red) with dashed lines indicating areas that might be degraded or produce si-like small RNAs. NDE, non-deleted element identified as the transcription start site for the DBE-T transcripts.

HETEROCHROMATIN AT D4Z4 REPEATS SILENCES LOCAL GENE EXPRESSION



A Long ncRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy

Daphne S. Cebianca,¹ Valentina Casa,^{1,2} Beatrice Bodega,^{3,5} Alexandros Xynos,¹ Enrico Ginelli,³ Yujiro Tanaka,⁴
and Davide Gabellini^{1,*}

¹Dulbecco Telethon Institute at San Raffaele Scientific Institute, Division of Regenerative Medicine, Stem Cells, and Gene Therapy, 20132 Milan, Italy

²Università Vita-Salute San Raffaele, 20132 Milan, Italy

³Department of Biology and Genetics for Medical Sciences, University of Milan, 20133 Milan, Italy

⁴Genome Structure and Regulation, School of Biomedical Science and Biochemical Genetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

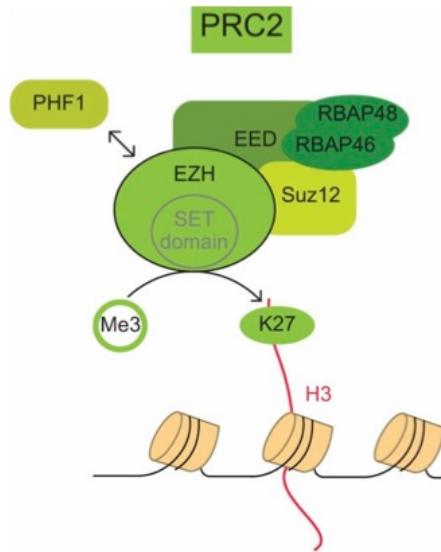
⁵Present address: Dulbecco Telethon Institute at Fondazione Santa Lucia, 00143 Rome, Italy

*Correspondence: gabellini.davide@hsr.it

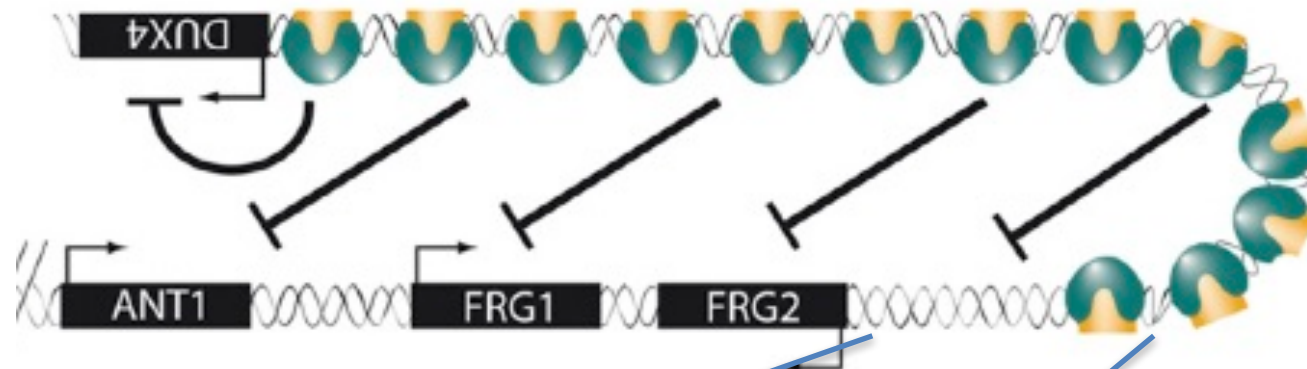
DOI 10.1016/j.cell.2012.03.035

IS SILENCING IMPAIRED IN FSDH PATIENTS??

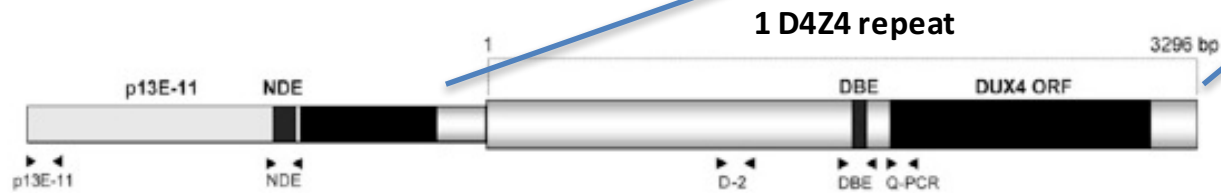
PRC2 → H3K27me3



POLYCOMB COMPLEX 2



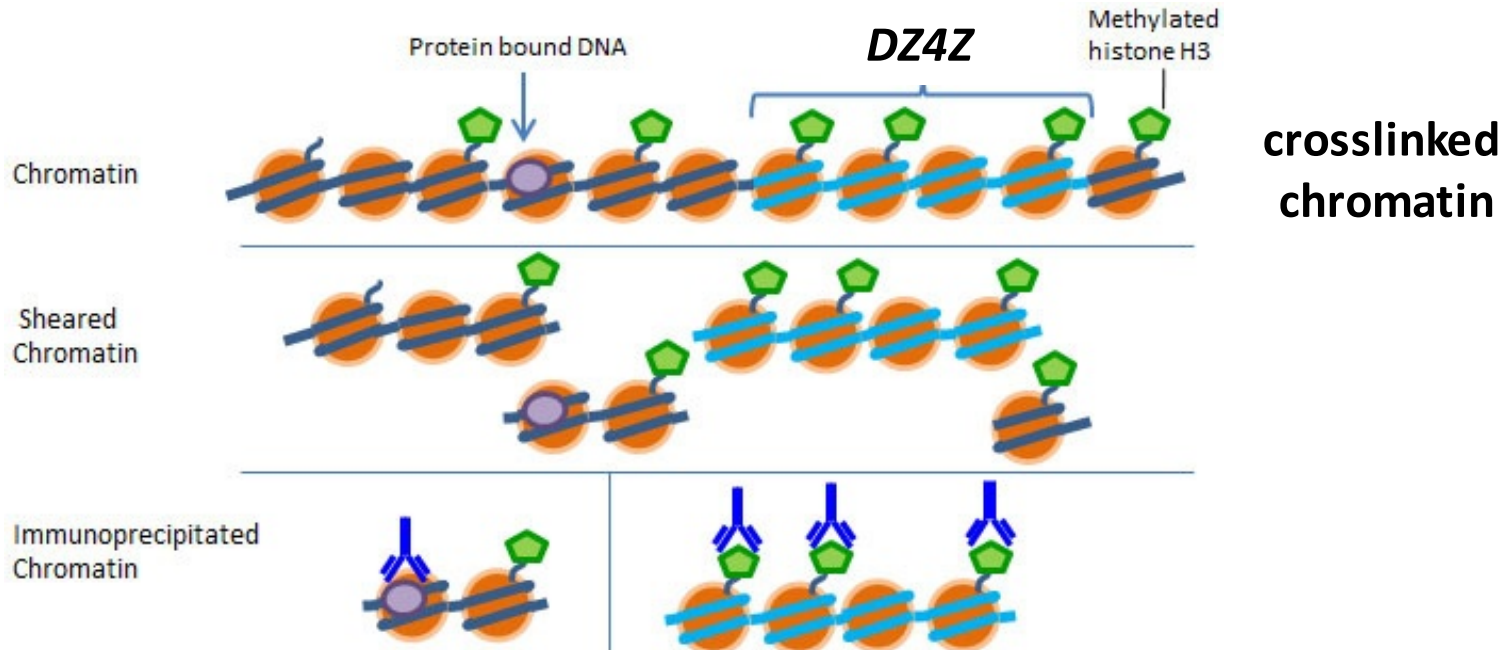
A



IS SILENCING IMPAIRED IN FSDH PATIENTS??

CHEMICALLY CROSSLINKED CHROMATIN ISOLATED FROM

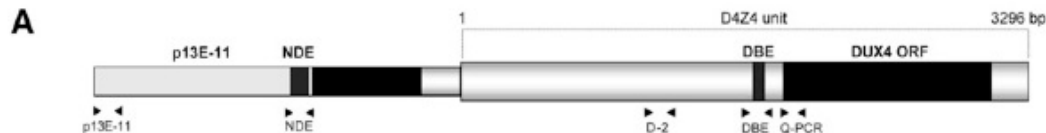
- a. Patient primary muscle cells
- b. Normal primary muscle cells



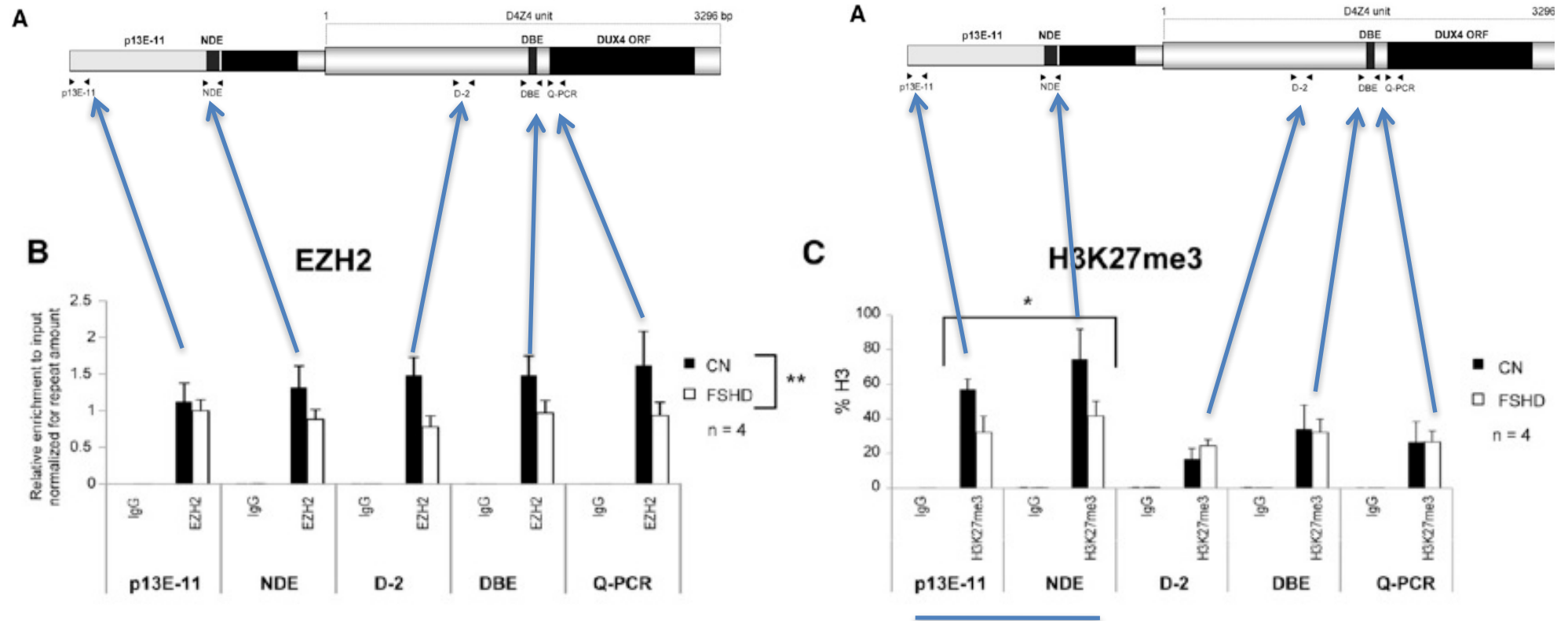
Precipitate-Ab-Chromatin complex with beads that bind heavy chain of antibody



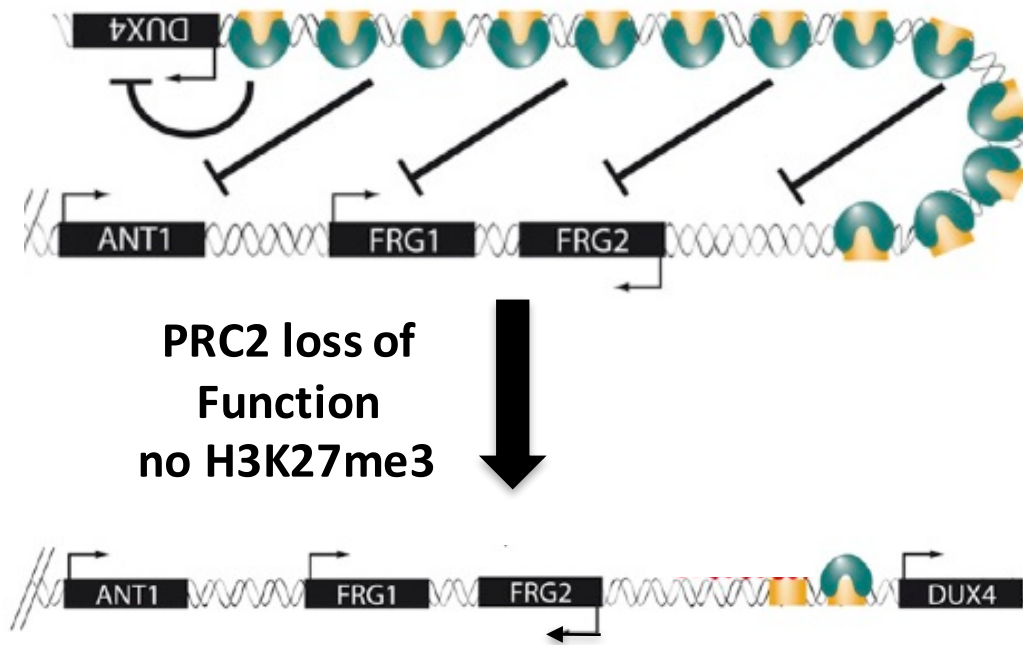
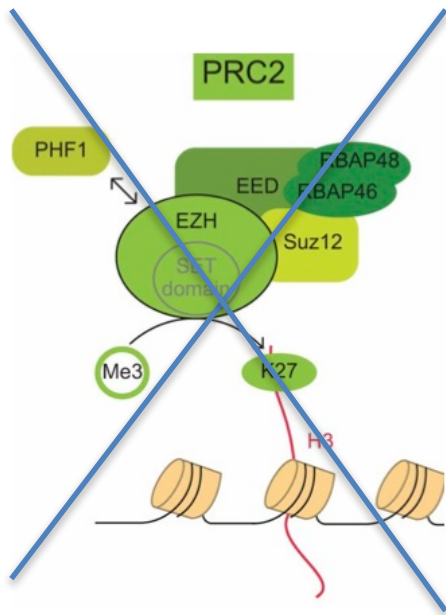
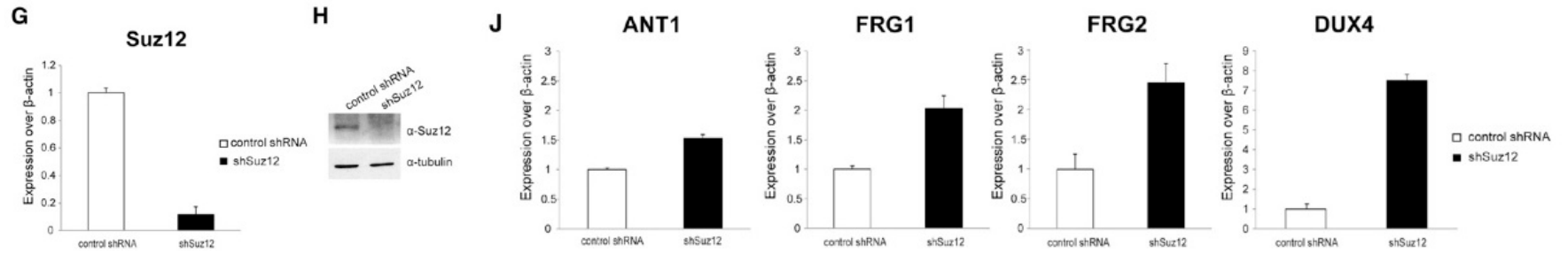
Make PCR with primers that amplify specific regions in D4Z4 repeats



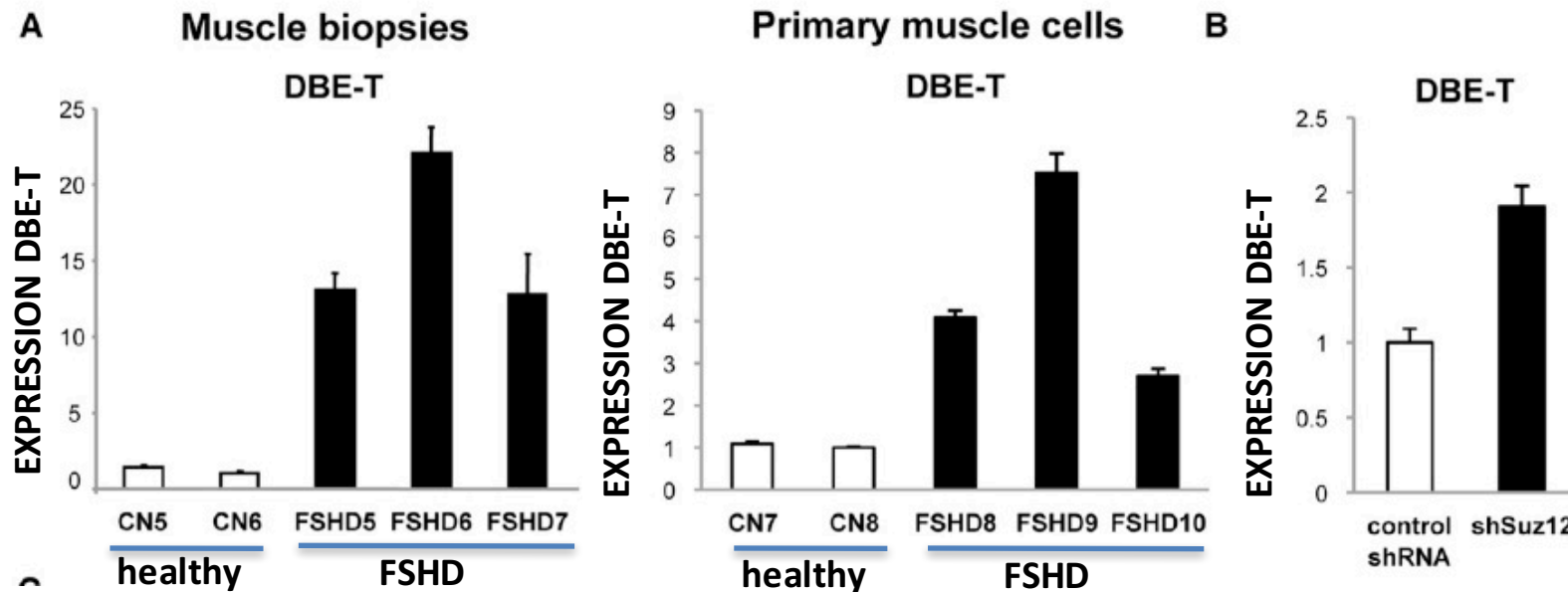
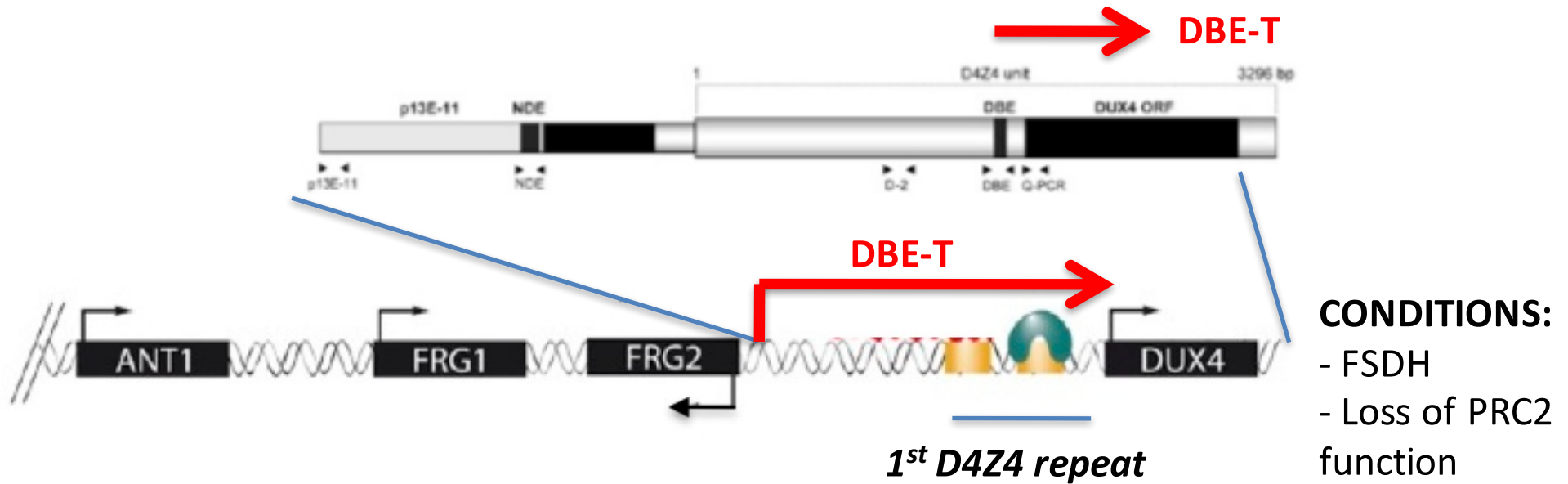
FSDH IS LINKED WITH LOSS OF PRC2 FUNCTION AT DZ4Z REPEATS



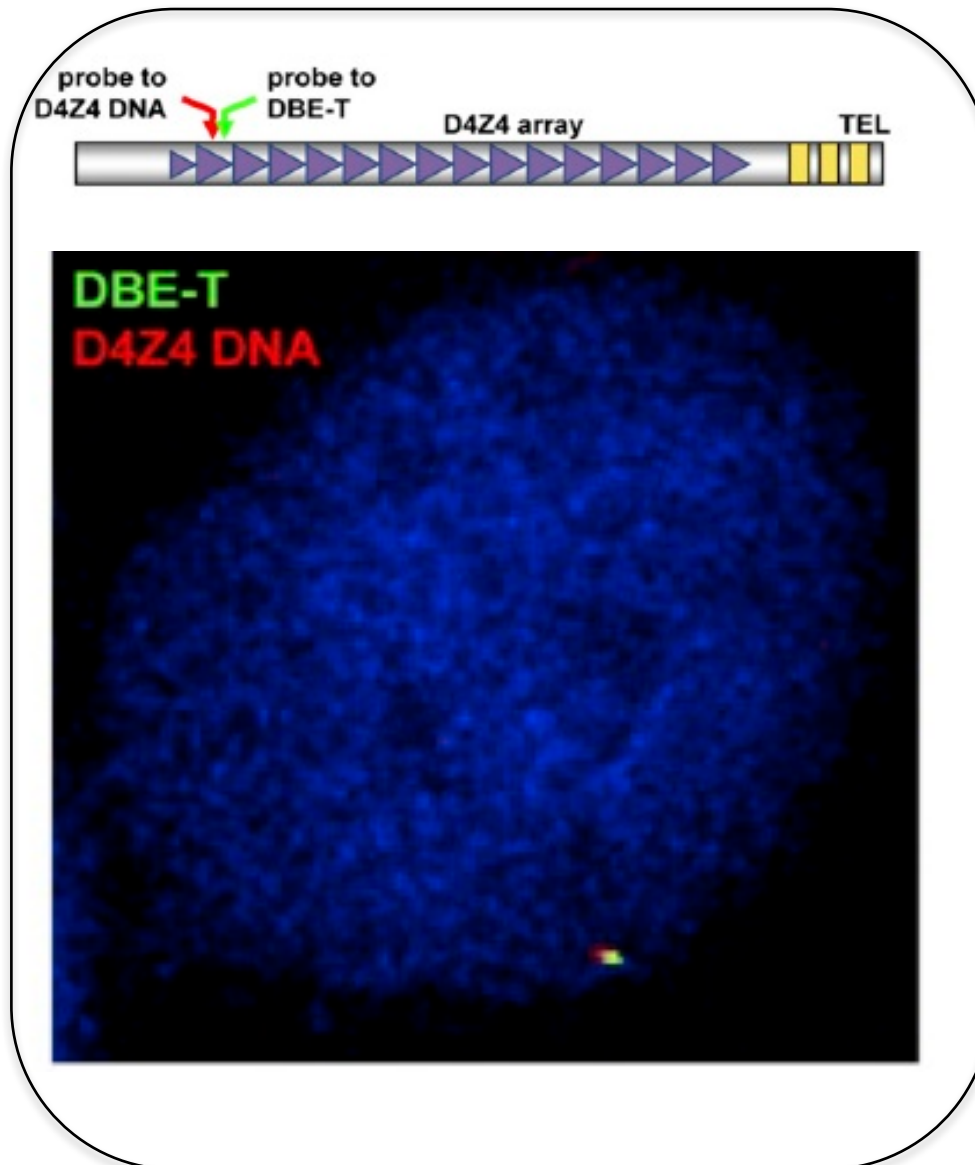
LOSS OF PRC2 FUNCTION INCREASES ANT1---DUX4 EXPRESSION



LOSS OF PRC2 FUNCTION CAUSES AN UPREGULATION OF A NOVEL lncRNA – DBE-T



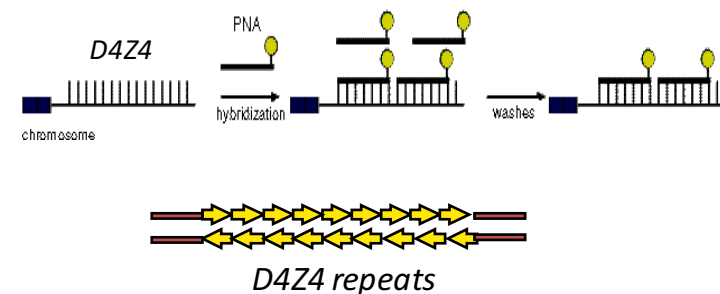
DBE-T IncRNA COLOCALIZES TO D4Z4 REPEATS



COMBINED DBE-T RNA-FISH AND D4Z4 DNA FISH

1. FIX CELLS
2. HYBRIDIZE A DBE-T PROBE (fluorescently labelled green)
2. CAPTURE IMAGE WITH
3. MICROSCOPE
4. WASH
5. DENATURE DNA (HEAT)
6. HYBRIDIZE D4Z4 PROBE (fluorescently labelled – red)
7. CAPTURE IMAGE
8. SUPERIMPOSE IMAGES

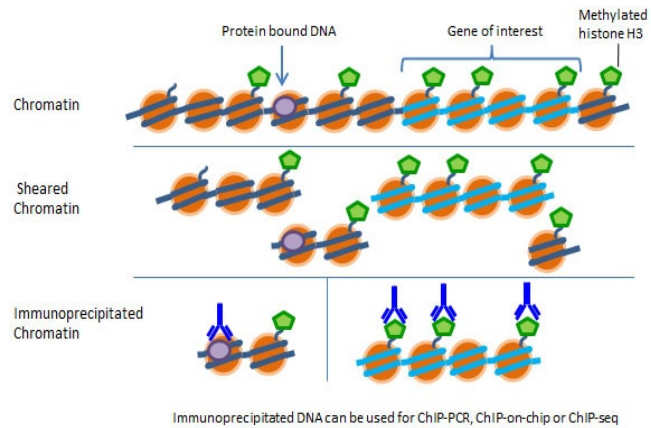
Fluorescence in situ hybridization



D4Z4 is a primate-specific repeat.

Model system: CHO hamster cell line, engineered with a single human chromosome 4 derived from a healthy subject (chr4/CHO).

The Trx group HMTase Ash1L localizes to DZ4Z REPEATS



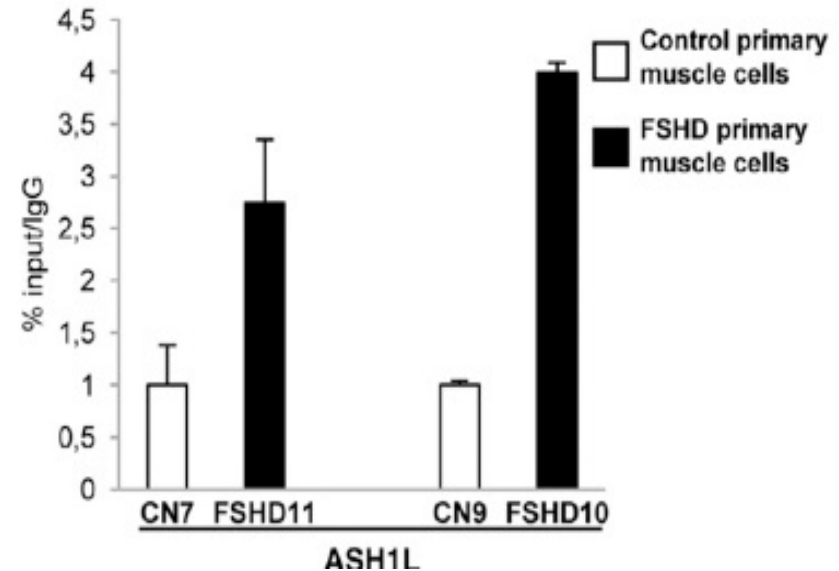
Ash1: Histone methyltransferase

→ H3K4me3

→ H3K36m2

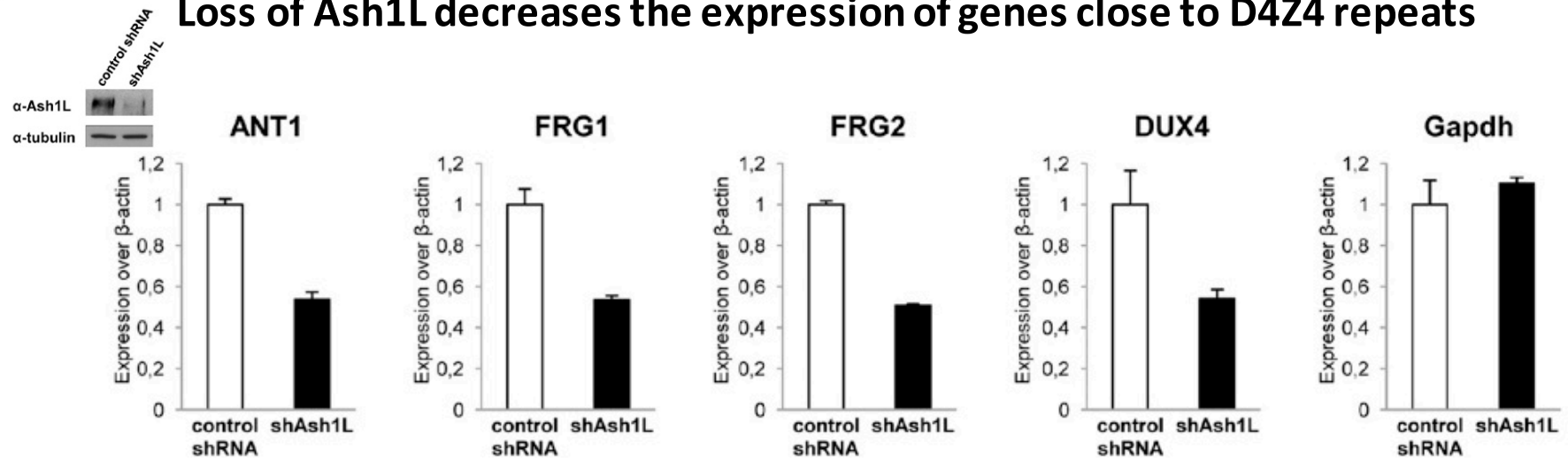
**ACTIVATES
TRANSCRIPTION**

B

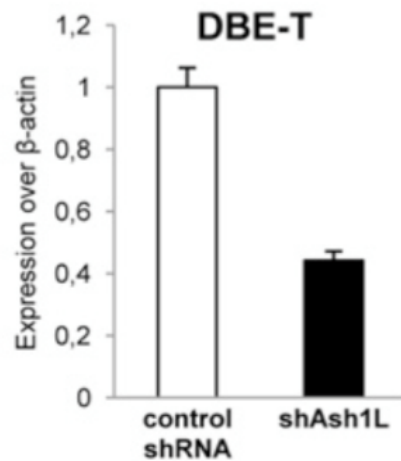


Loss of Ash1L increases ANT1/FRG1/GRF2/DUX4 and DBE-T expression

Loss of Ash1L decreases the expression of genes close to D4Z4 repeats

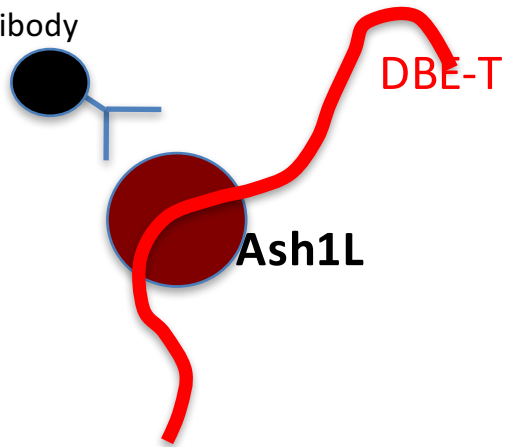


Ash1L shRNA experiments in Chr4/CHO cells



DBE-T INTERACTS WITH Ash1L

Bead+anti-Ash1L antibody

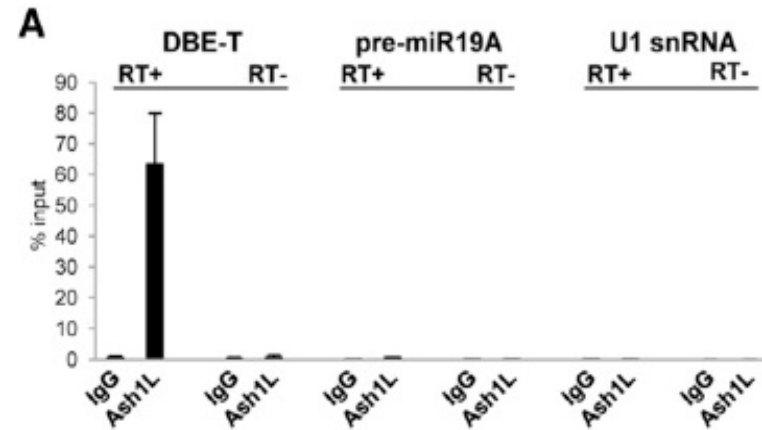


RNA-immunoprecipitation

↓
Elute bound RNA

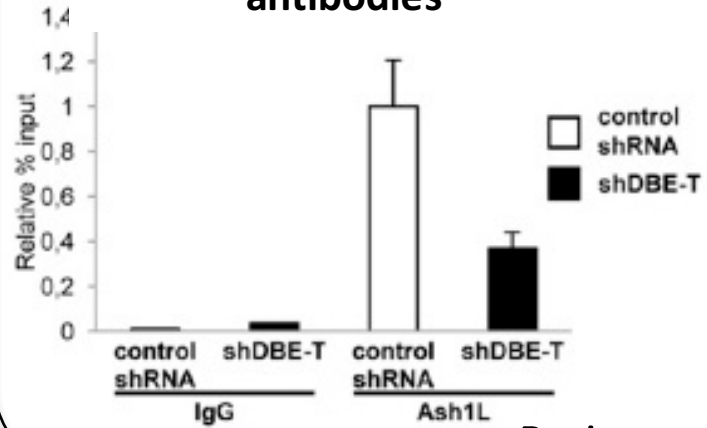
↓
Reverse transcription

↓
PCR specific for DBE-T and control RNAs (pre-miR19; U1 snRNA)

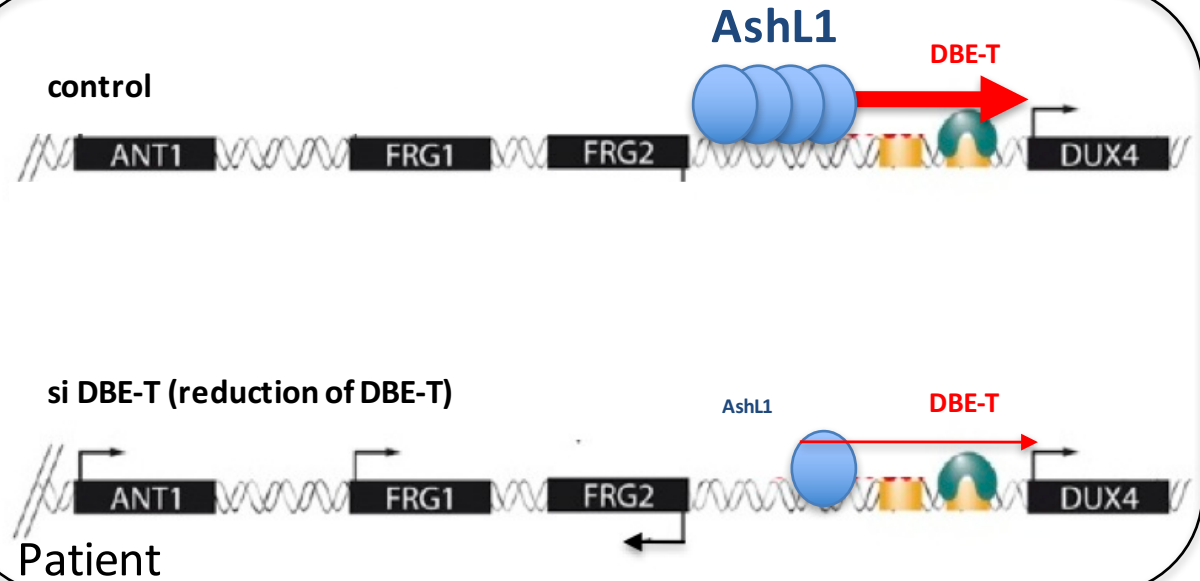


DBE-T BRINGS Ash1L TO D4Z4 REPEATS

ChIP D4Z4 using anti-Ash1L antibodies



Patient

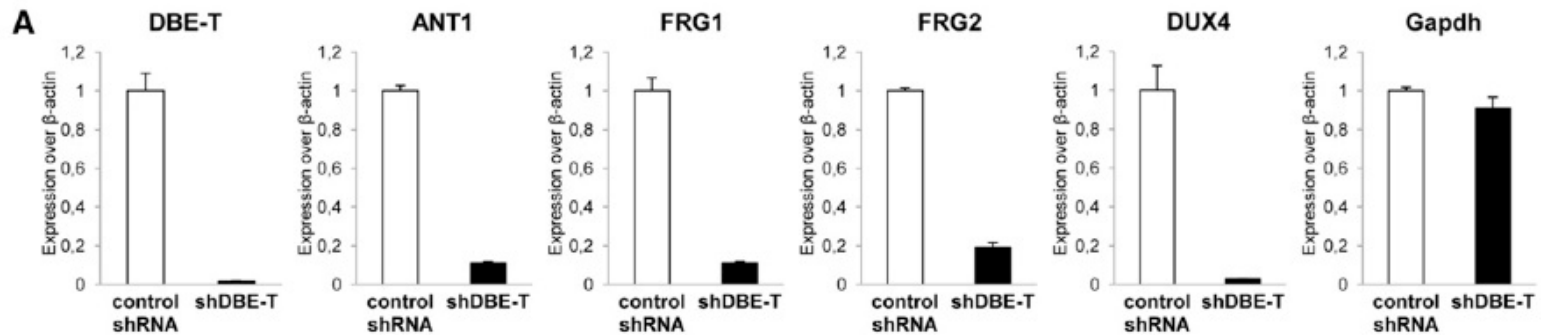


Patient

Loss of DBE-T results in decreased expression of D4Z4 neighboring genes

Gene expression RT-PCR

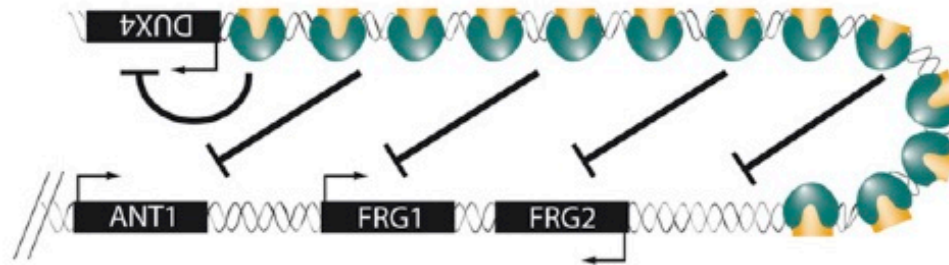
control and si DBE-T (reduction of DBE-T)



A lincRNA IS USED TO TRANSMIT D4Z4 REPEAT NUMBER INTO A DISEASE RELEVANT MECHANISM

HEALTHY

11 to 100 D4Z4 repeats
4q35 gene **repression**

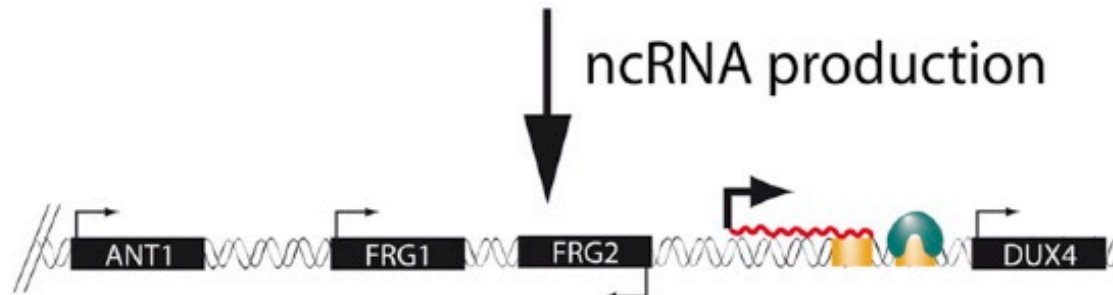


■ DBE

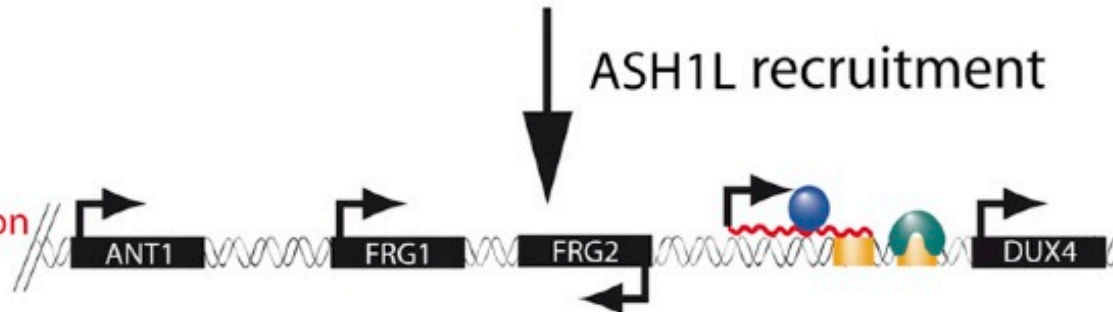
● PcG

● ASH1L

ncRNA production



ASH1L recruitment



FSHD

4q35 gene **de-repression**

