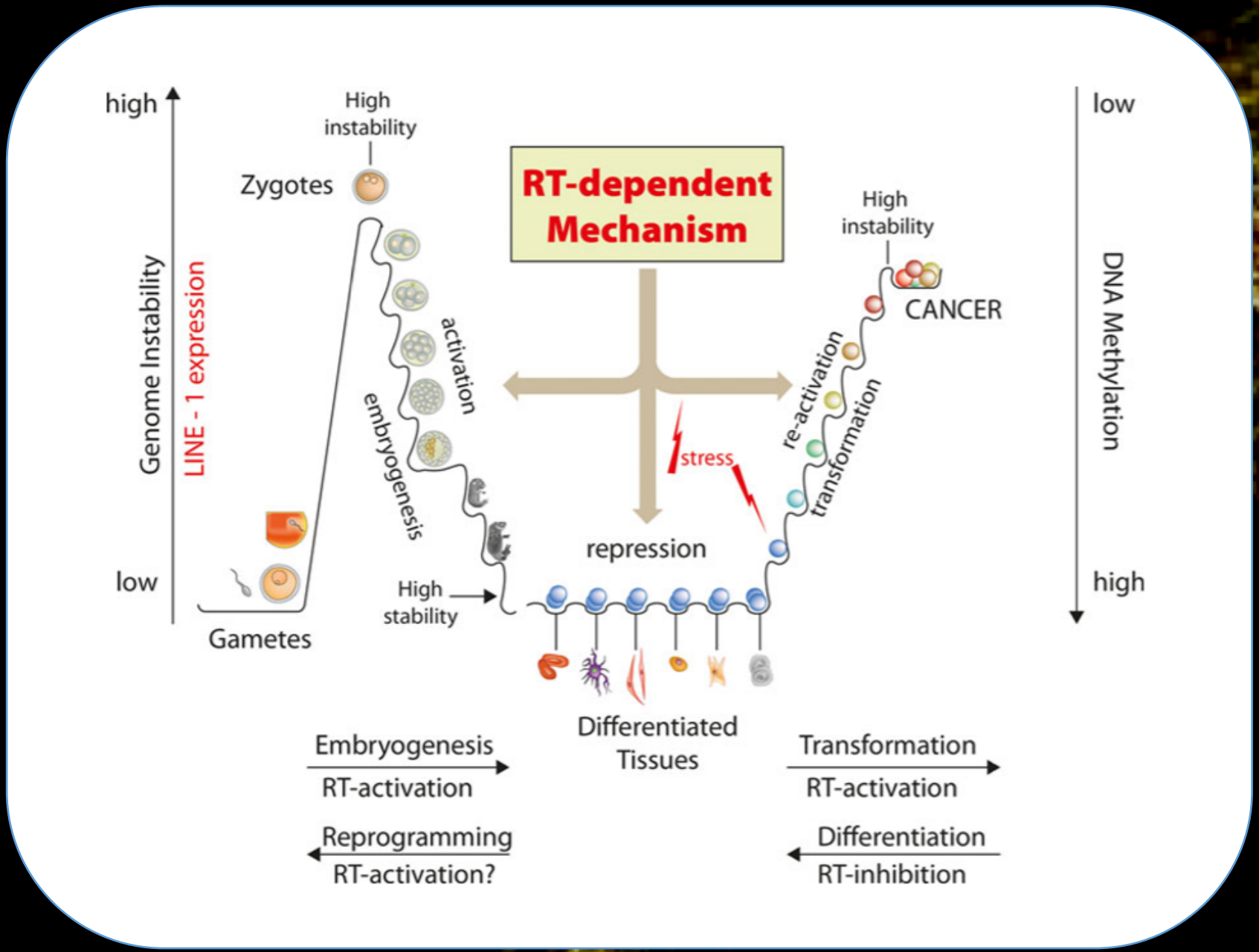
PSEUDOGENE lncRNAs EXAMPLE 1

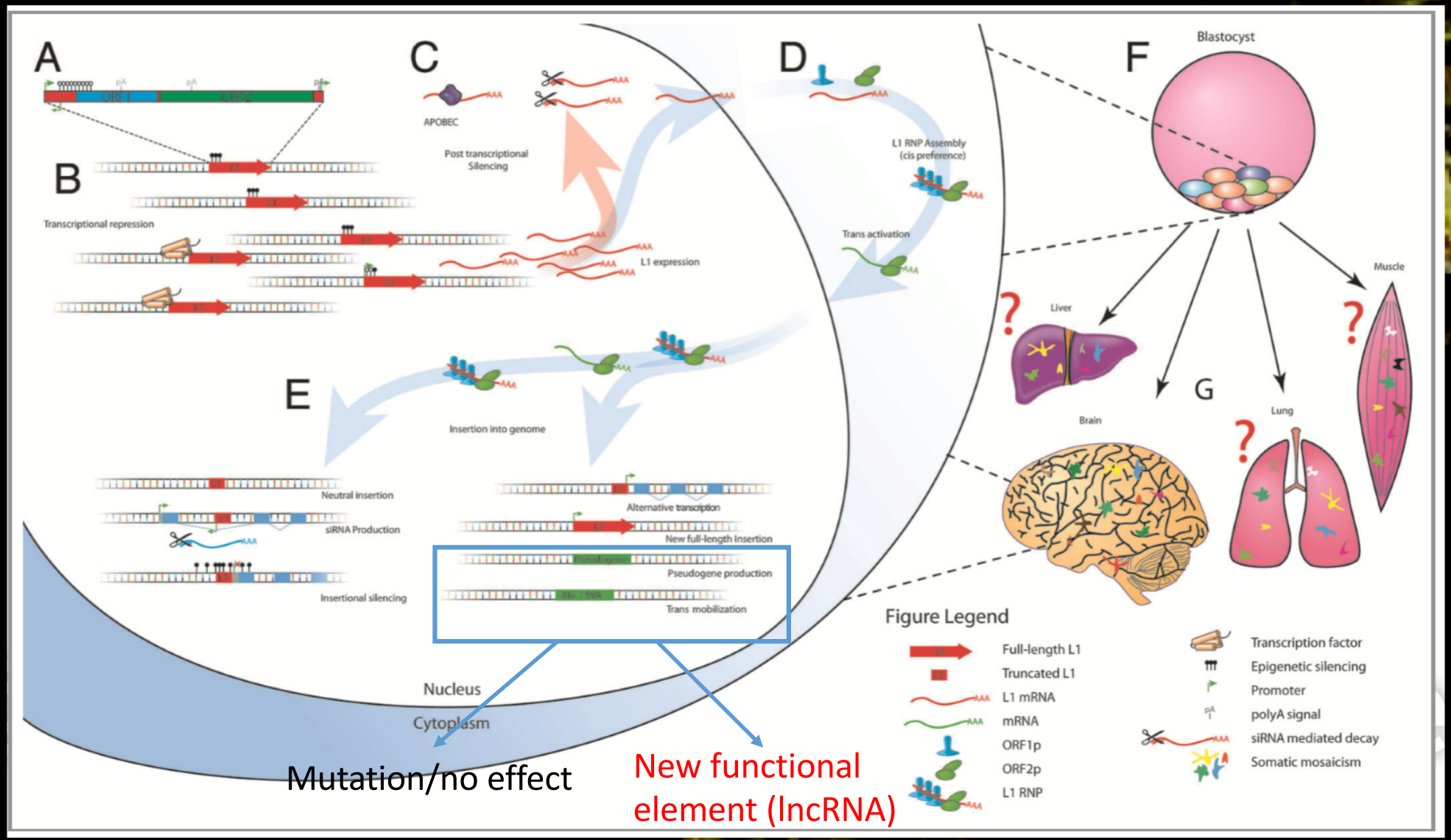
Pseudogene lncRNA that controls embryonic stem cell Self-renewal

A Oct4P4 pseudogene derived lncRNA silences the ancestral Oct4 gene in trans

Retrotransposon activity during development

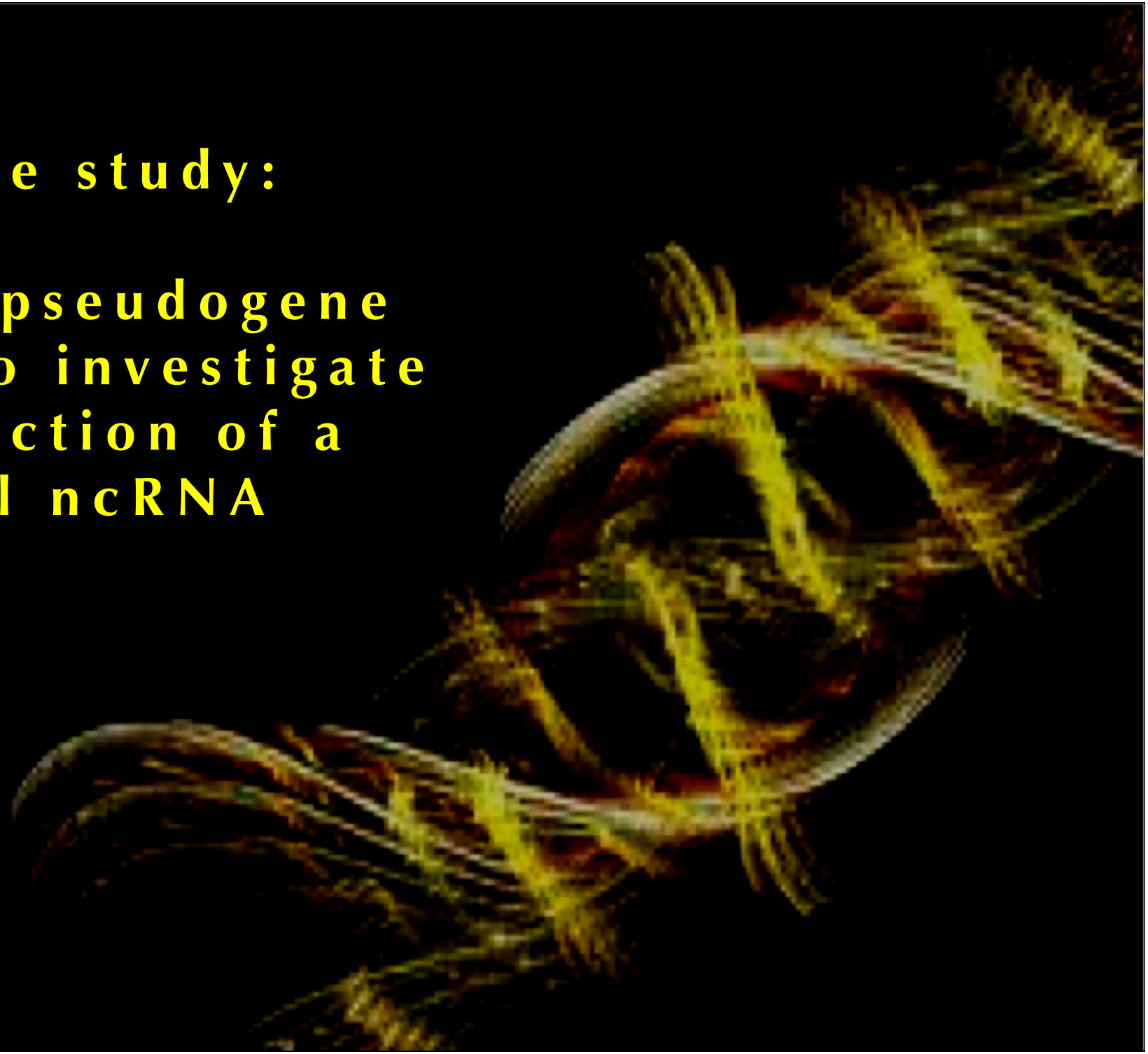


Retrotransposons can change genetic Context → mosaic somatic tissues



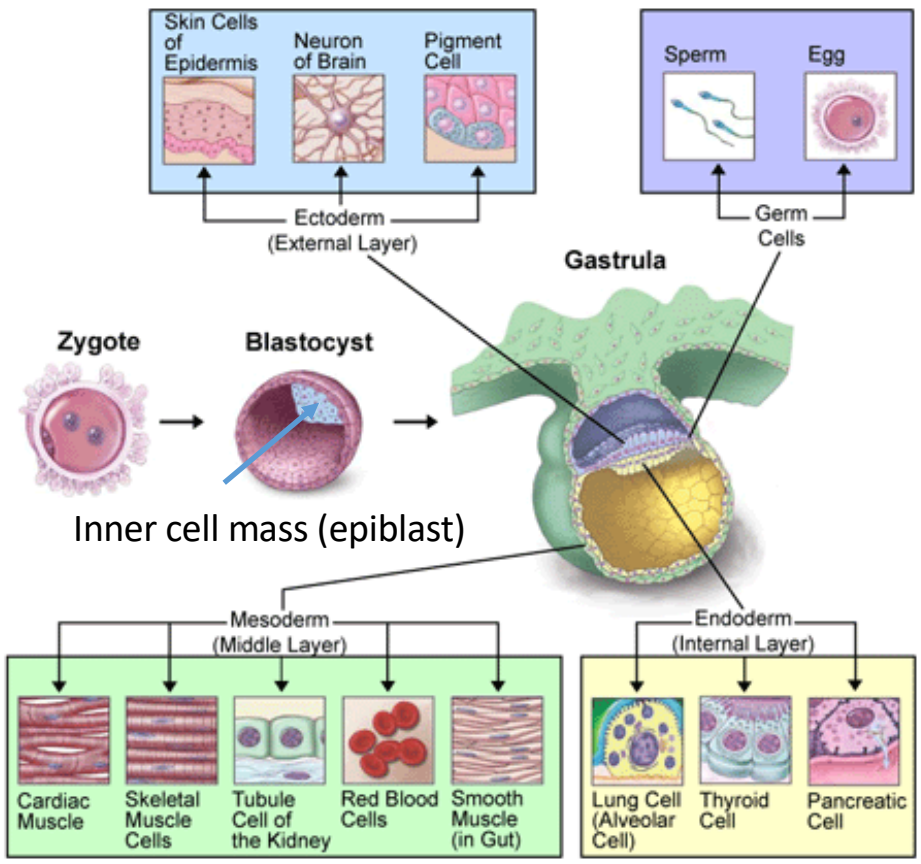
A case study:

Using a pseudogene lncRNA to investigate the function of a novel ncRNA

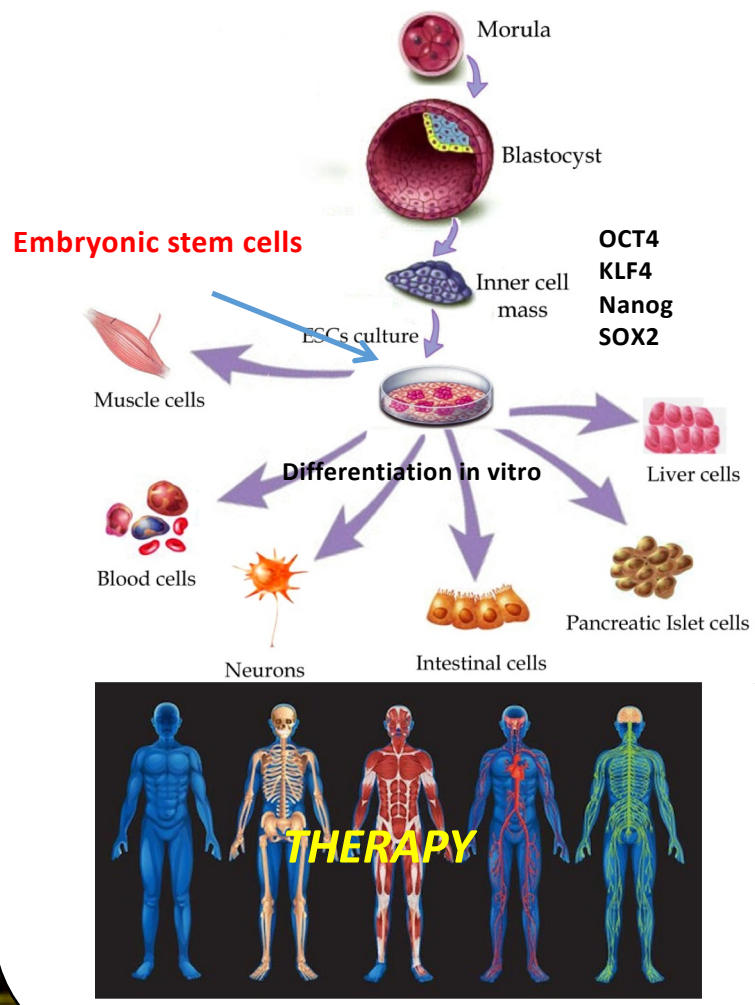


The inner cell mass of the blastocyst are the source of pluripotent embryonic stem cells

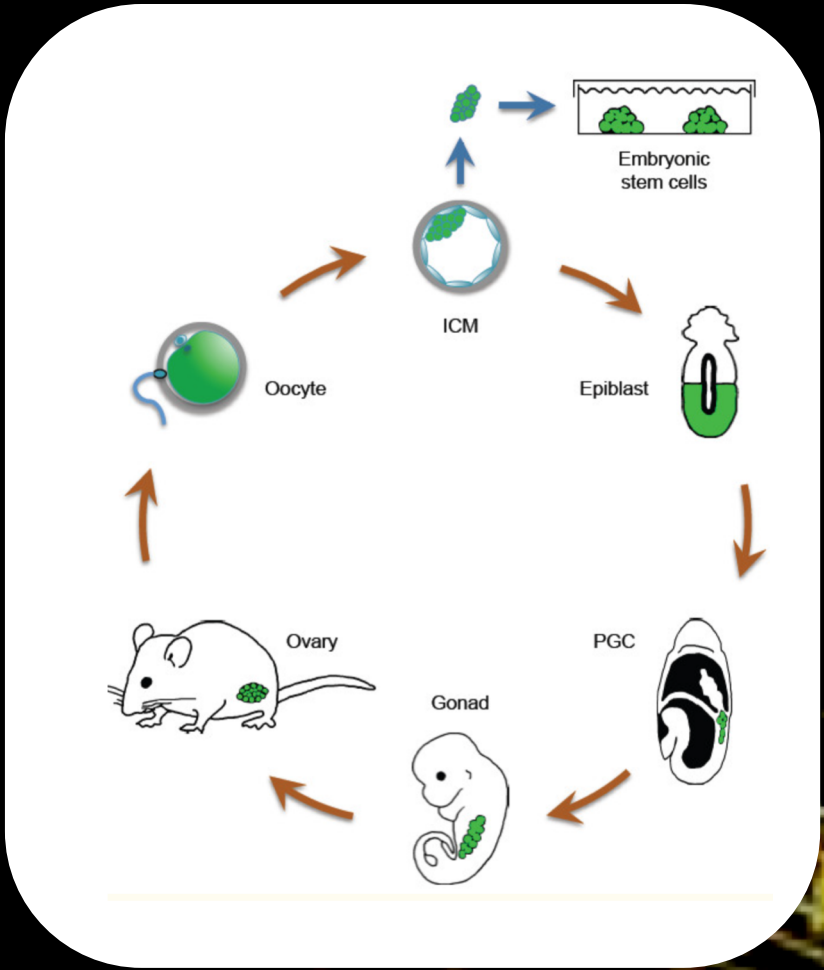
Blastocyst cells give rise to all organs and cell types



Blastocyst cells can be Isolated and cultivated



1. OCT4 EXPRESSION
Oct4 expression during the mouse life cycle



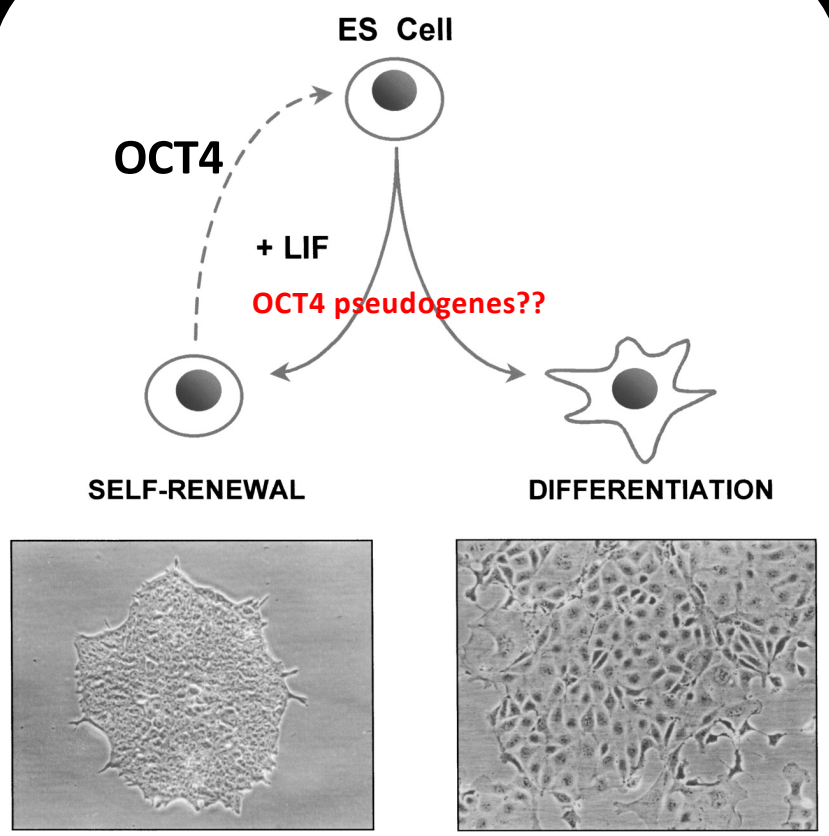
Cells and tissues expressing Oct4 are marked in green.

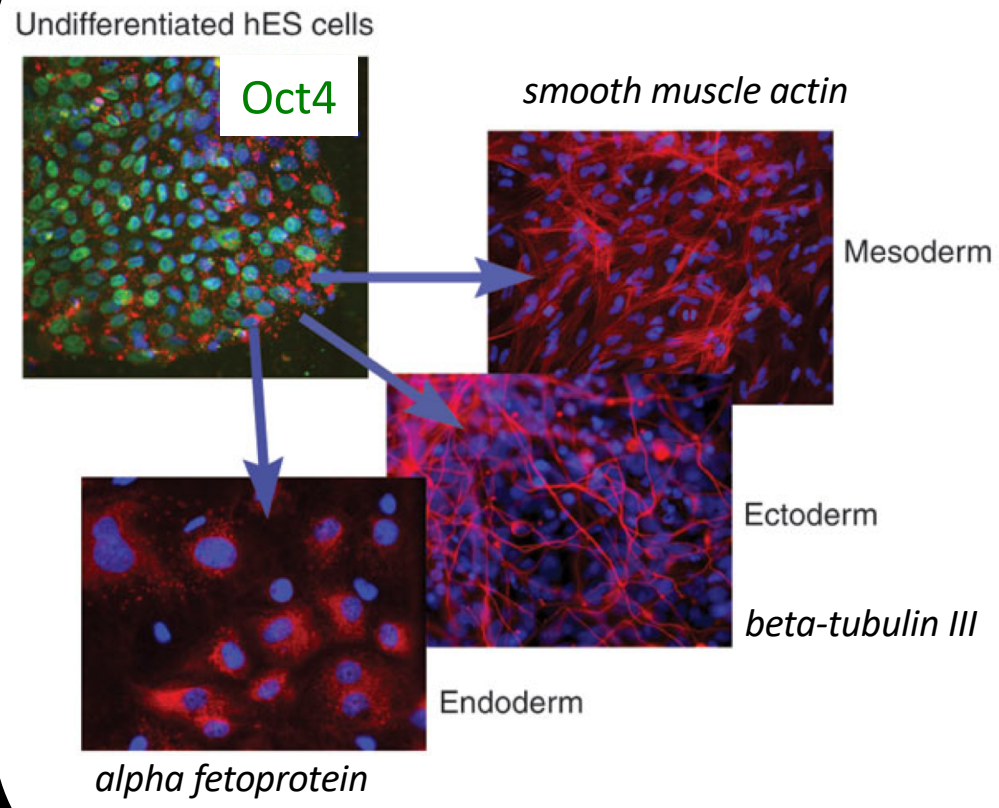
- Oct4 is expressed in mouse oocytes as a maternal transcript and protein.
- Zygotic Oct4 expression is activated prior to the 8-cell stage and is abundant and uniform in all cells of the embryo throughout the morula stage.
- However, as the outer cells of the embryo differentiate into the Trophectoderm, Oct4 expression is restricted to cells of the ICM in the blastocyst.
- After implantation, Oct4 expression is maintained in the epiblast of the blastocyst
- Oct4 expression becomes restricted to primordial germ cells (PGCs), which are first specified in the extraembryonic mesoderm at the base of the allantoic bud during gastrulation. PGCs give rise to gametes, which following fertilization will develop into a new organism of next generation.

Wu and Schöler: Role of Oct4 in the early embryo development Cell Regen. 2014

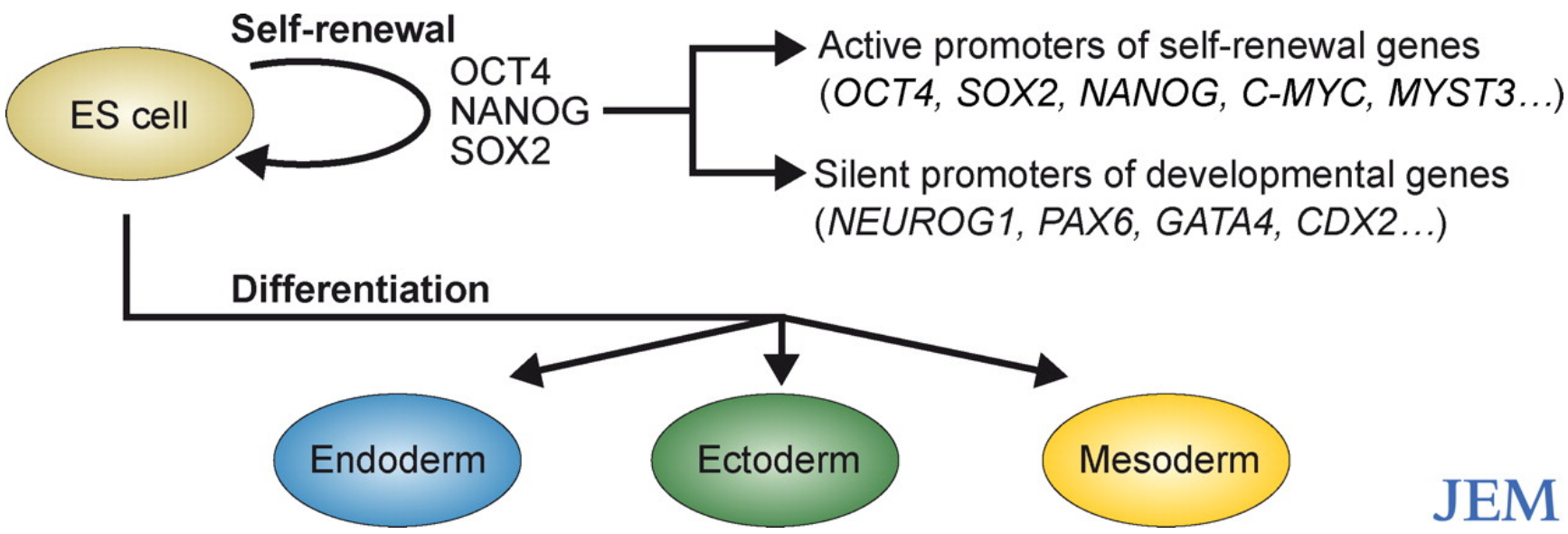
2. THE MODEL SYSTEM: ES CELL DIFFERENTIATION

OCT4 expressing ES cells have self-renewing and differentiation potential *in vitro*





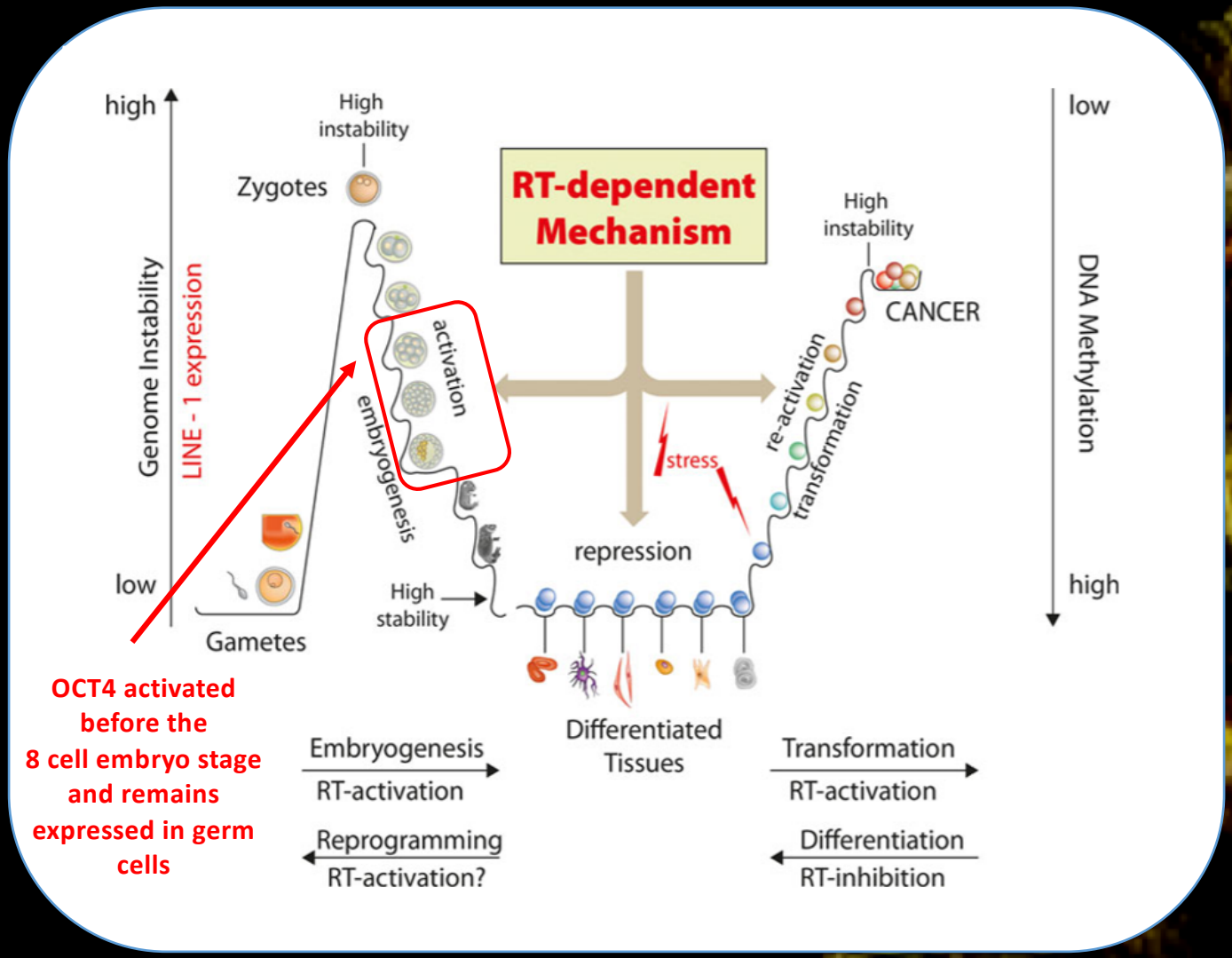
The self-renewal transcription factor Oct4 is essential for embryonic stem cell self-renewal



JEM

Nicolaj Strøyer Christophersen, and Kristian Helin J Exp Med 2010;207:2287-2295

Mouse and human contain several processed OCT4 pseudogenes



OCT4 activated before the 8 cell embryo stage and remains expressed in germ cells

Human:
1 ancestral OCT4
7 processed OCT4 pseudogenes

Mouse:
1 ancestral Oct4
5 processed OCT4 pseudogenes

3. Sequence analysis:
Are there other sequences in the mouse genome that are similar to Oct4?

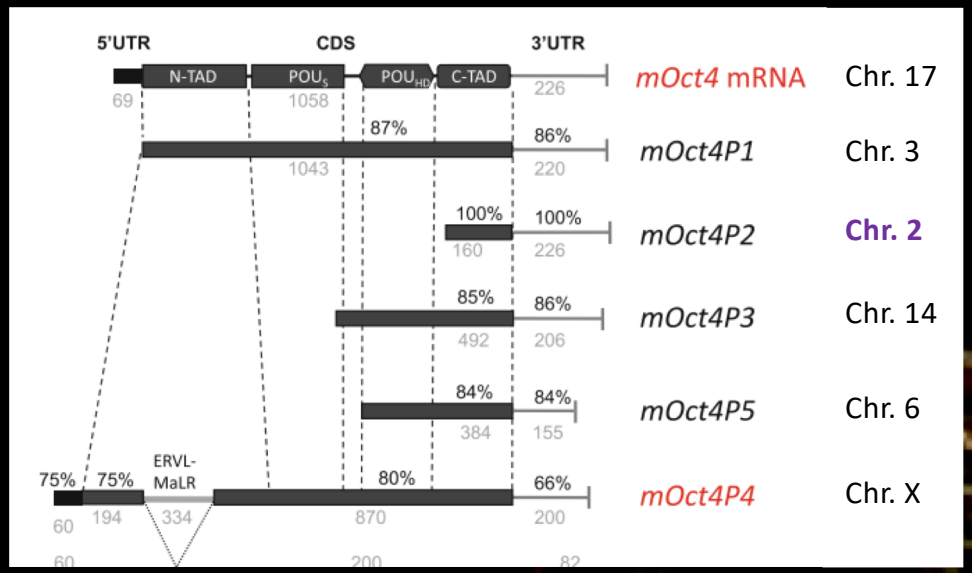
Use Oct4 cDNA and make BlastN

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download New Select columns Show 100

select all *6 sequences selected* GenBank Graphics Distance tree of results New MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Mus musculus strain C57BL/6J chromosome 3, GRCm39	Mus musculus	1177	2059	100%	0.0	87.15%	159745316	NC_000069.7
Mus musculus strain C57BL/6J chromosome 17, GRCm39	Mus musculus	710	1977	100%	0.0	100.00%	95294699	NC_000083.7
Mus musculus strain C57BL/6J chromosome X, GRCm39	Mus musculus	638	638	82%	2e-180	80.20%	169476592	NC_000086.8
Mus musculus strain C57BL/6J chromosome 14, GRCm39	Mus musculus	520	520	46%	8e-145	85.71%	125139656	NC_000080.7
Mus musculus strain C57BL/6J chromosome 6, GRCm39	Mus musculus	383	383	36%	1e-103	84.62%	149588044	NC_000072.7
Mus musculus strain C57BL/6J chromosome 1, GRCm39	Mus musculus	303	303	15%	9e-80	100.00%	195154279	NC_000067.7



Unspliced sequence on Chr. 2 with high similarity to Oct4

Mus musculus strain C57BL/6J chromosome 3, GRCm39

Sequence ID: NC_000069.7 Length: 159745316 Number of Matches: 2

Range 1: 129227322 to 129228364 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1177 bits(637)	0.0	936/1074(87%)	46/1074(4%)	Plus/Plus

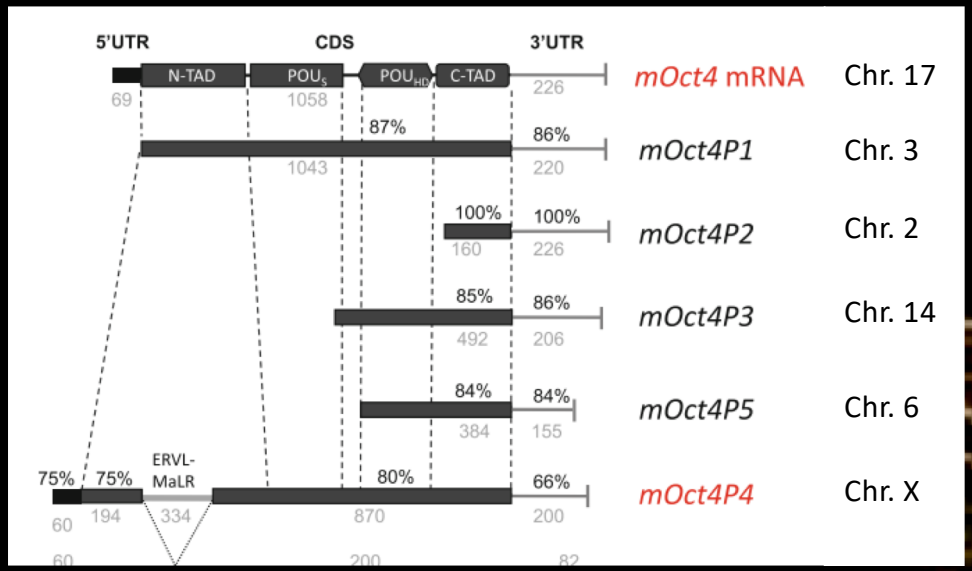
Features: 101192 bp at 5' side: glutamyl aminopeptidase
97965 bp at 3' side: elongation of very long chain fatty acids protein 6

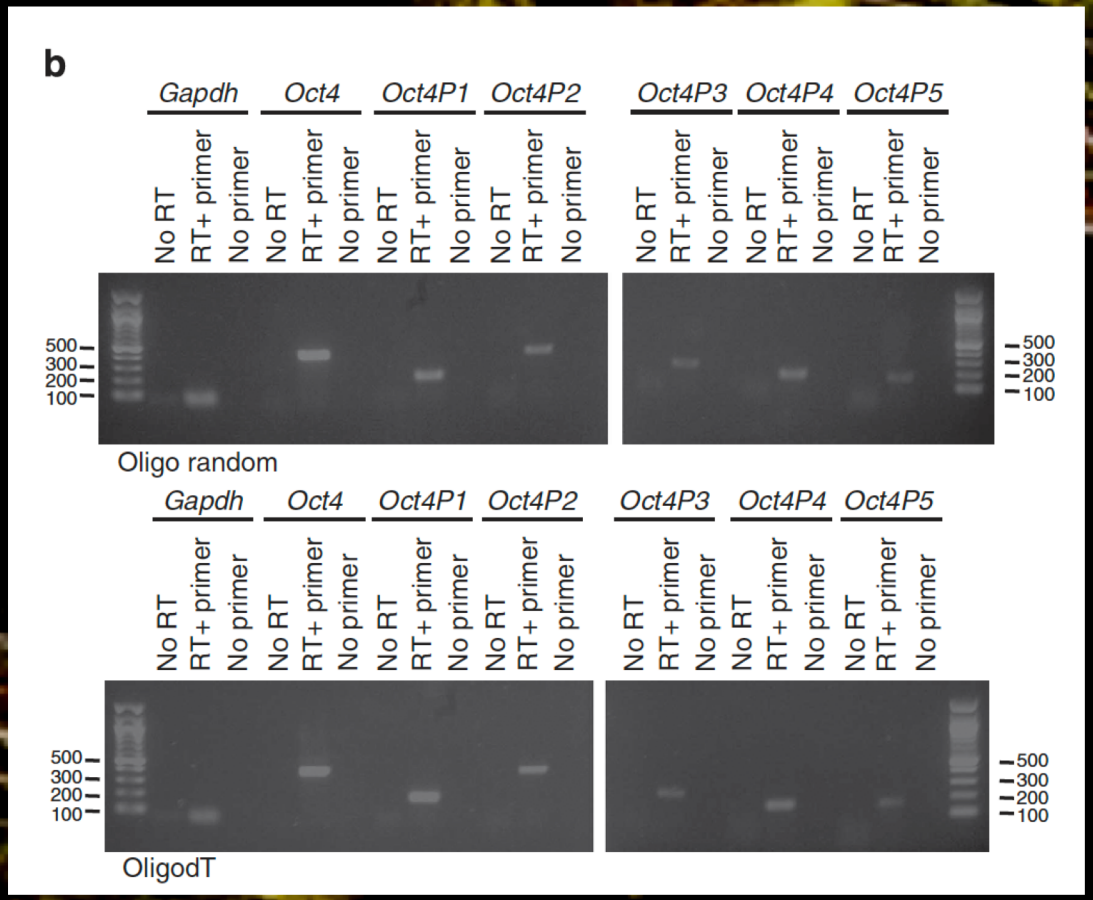
Oct4 Chr2 sequence

```
Query  1          ATGGCTGGACACCTGGCTTCAGACTTCGCCTTCTCACCCCCACCAGGTGGGGGTGATGGG  60
Sbjct  129227322  ATGGCTGGACACCTGGCTTCAGACTTCGCCTTCTCA-CCCCACCAGGTGGGGGCGATGGG  129227380
Query  61         TCAGCAGGGCTGGAGCCGGGCTGGGTGGATCCTCGAACCTGGCTAAGCTTCCAAGGGCCT  120
Sbjct  129227381  TCAGCAGGGCTGGAGCCGGGCTGGGTGGACCCTCGAACCTGGCTAAGCTTCCAAGGGCCT  129227440
Query  121        CCAGGTGGGCCTGGAATCGGACCAGGCTCAGAGGTATTGGGGATCTCCCCCATGTCCGCCC  180
Sbjct  129227441  CTGGGTGGGCCTGGAATAGGACCAGGCTTAGAGGTATT-GGGATCTCCCCATGTCCACCG  129227499
Query  181        GCATACGAGTTCTGCGGAGGGATGGCATACTGTGGACCTCAGGTTGGACTGGGCCTAGTC  240
Sbjct  129227500  GTATATGAGTTCTGCGGAGGGATGGCATACTGTGGACCTCAG------------------  129227541
Query  241        CCCCAAGTTGGCGTGGAGACTTTGCAGCCTGAGGGCCAGGCAGGAGCACGAGTGGAAAGC  300
Sbjct  129227542  ------GCTGGCGTGGAGACTTTGCAGCCTGAGAGCCAGGCAAGAGCAGGTGTGGAGAGC  129227595
Query  301        AACTCAGAGGGAACCTCCTCTGAGCCCTGTGCCGACCGCCCCAATGCCGTGAAGTTGGAG  360
Sbjct  129227596  AACTCAGAGGAAGCCTTCTCTGGGCCTTGTGCCGACCACCCCAGTGCCGTGAAGTTGGAG  129227655
Query  361        AAGGTGGAACCAACTCCCGAGGAGTCCCAGGACATGAAAGCCCTGCAGAAGGAGCTAGAA  420
Sbjct  129227656  ATGAGGGAACCACGTCCCGAGGGGTCCCAGGACATGAAAGTCCTGCAGAAGGAGCTAGAG  129227715
Query  421        CAGTTTGCCAAGCTGCTGAAGCAGAAGAGGATCACCTTGGGGTACACCCAGGCCGACGTG  480
Sbjct  129227716  AAGTTTGCCAAGCTGCTGAAGCAGAAGAGGATCACCTTGGGGTACACCCAGACCGACGTG  129227775
Query  481        GGGCTCACCCTGGGCGTTCTCTTTGGAAAGGTGTTCAGCCAGACCACCATCTGTCGCTTC  540
Sbjct  129227776  GGGCTCACCCTGGGCGTTCTCTTCAGAAAGGTG-TCAGCCAGACAACCATCTACTGCTTT  129227834
Query  541        GAGGCCTTGCAGCTCAGCCTTAAGAACATGTGTAAGCTGCGGCCCCTGCTGGAGAAGTGG  600
Sbjct  129227835  GAGGCCCGGCAGCTCAGCCTTATGACCATGCATAAGCTGCGGCCCCTGCTGGAGAAGTGG  129227894
Query  601        GTGGAGGAAGCCGACAACAATGAGAACCTTCAGGAGATATGCAAATCGGAGACCCTGGTG  660
Sbjct  129227895  GTAGAGGAAGCGGAAGACAATGAGAACCTTCAGGAGATATGCAAATCGGAGACCATGGTG  129227954
Query  661        CAGGCCCGGAAGAGAAAGCGAACTAGCATTGAGAACCGTGTGAGGTGGAGTCTGGAGACC  720
Sbjct  129227955  CAGGCCTGGAAGAAAAAGAGCGCGAGCATTGAGAACCATGTGAGGCAGAGCCTGGAGAAT  129228014
Query  721        ATGTTTCTGAAGTGCCCGAAGCCCTCCCTACAGCAGATCACTCACA-TCGC--CA--AT-  774
Sbjct  129228015  GAGTTTCTGAAGTGCCCAGAGCCCTCCCCGCAGCAGATCACTAGCATTAGCATCACCATG  129228074
Query  775        CAGCTTGGGCTAGAGAAGGATGTGGTTCGAGTATGGTTCTGTAACCGGCGCCAGAAGGGC  834
Sbjct  129228075  CAGCTTGGGATGGAGAAGGACATGGTTCGAGGATGGTTCTGTAACCGGCGCCAGAAGGGC  129228134
Query  835        AAAAGATCAAGTATTGAGTATTCCCAACGAGAAGAGTATGAGGCTA-CAGGG-ACACCTT  892
Sbjct  129228135  AAAAGATCAAGTATTGAGTATTCCCAACGAGAAAAGTATGAGGCTTTCCGGGGATGCCTT  129228194
Query  893        TCC---CA-ggg--ggggCTGTATCCTTTCCTCTGCCCCCAGGTCCCCACTTTGGCACCC  946
Sbjct  129228195  TCGGGGCGTGGGGTGGGGCTGTATCCTTTCCTCTGCCCCCAGG-CCCCACTTTGGCACCC  129228253
Query  947        CAGGCTATGGAAGCCCCCACTTCACCACACTCTACTCAGTCCCTTTTCCTGAGGGCGAGG  1006
Sbjct  129228254  CAAGCTATGGGAGCCCCCACTTCACCACACT--ACTCGATACCTG-TCCTGAGGGCGAGG  129228310
Query  1007       CCTTTCCCTCTGTTCCC-GTCACTGCTCTGGGCTCTCCCATGCATTCAAACTGA  1059
Sbjct  129228311  CCTTTCCCTCTGTCCCCTGTCATTGCTCTGGGCTCTCCCATGCATTCAAACTGA  129228364
```

4. EXPERIENTAL EVALATION OF lncRNA EXPRESSION IN MODEL CELLS

Self-renewing mESCs

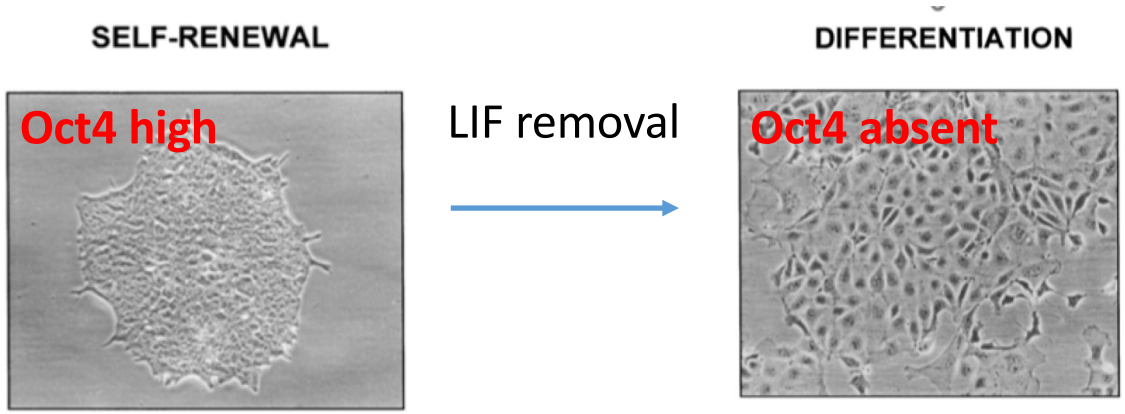




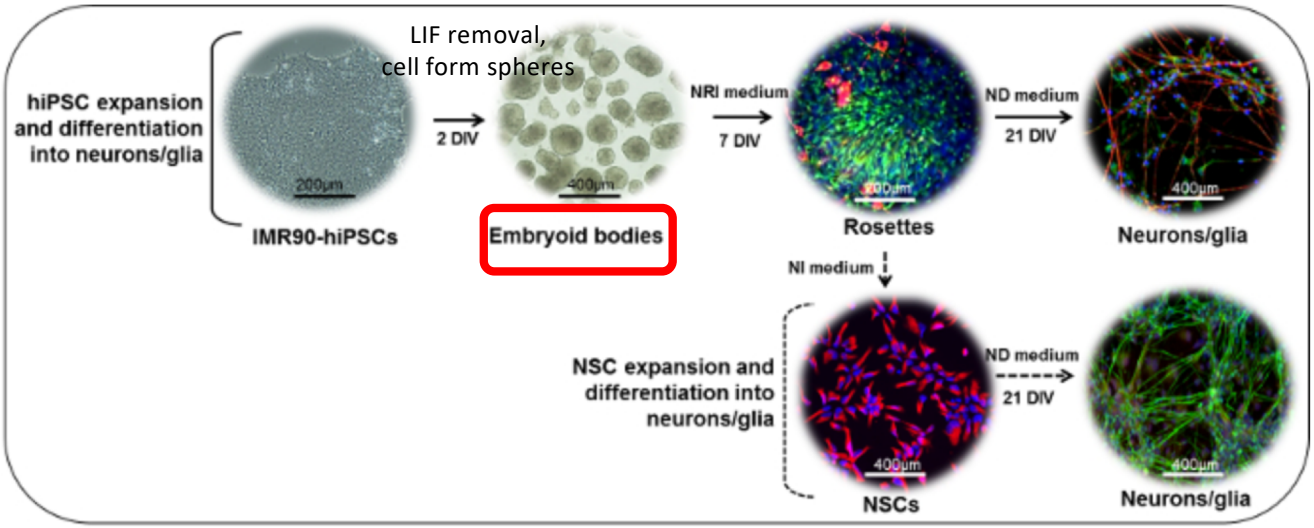
Ancestral OCT4 gave rise to 5 processed pseudogenes that are expressed in mESCs

4. EXPRESSION DURING IN VITRO DIFFERENTIATION

In vitro differentiation; 2D cell culture, attached on petri dishes

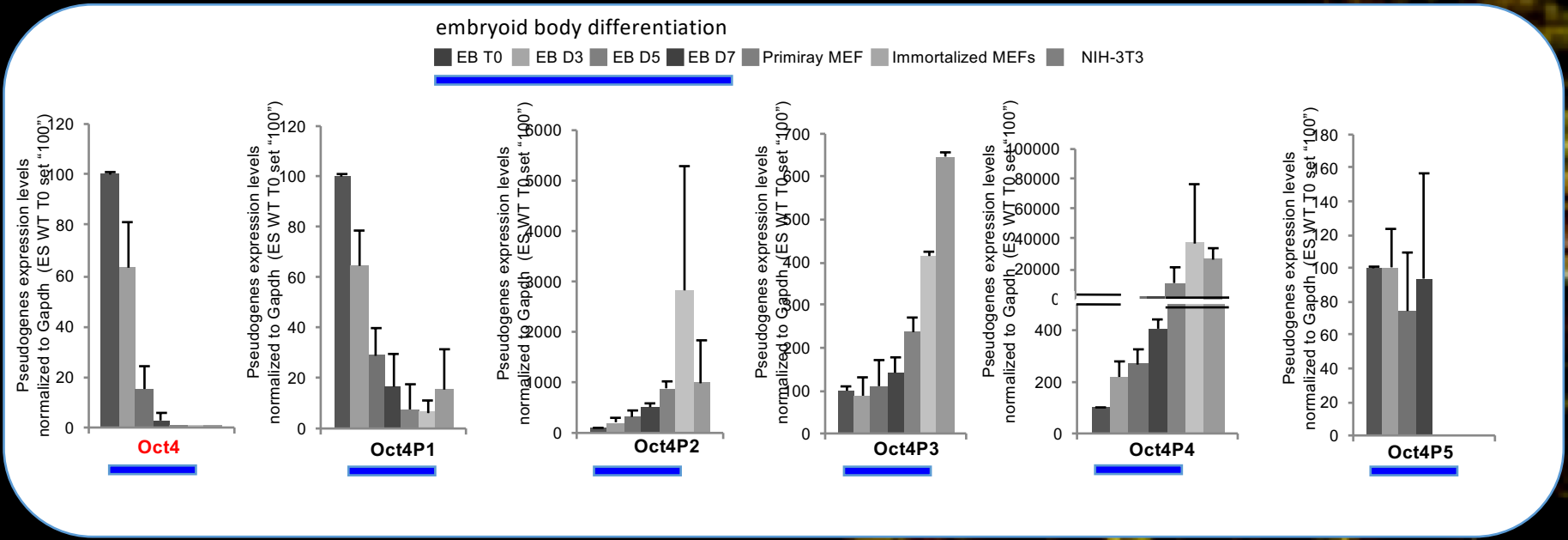


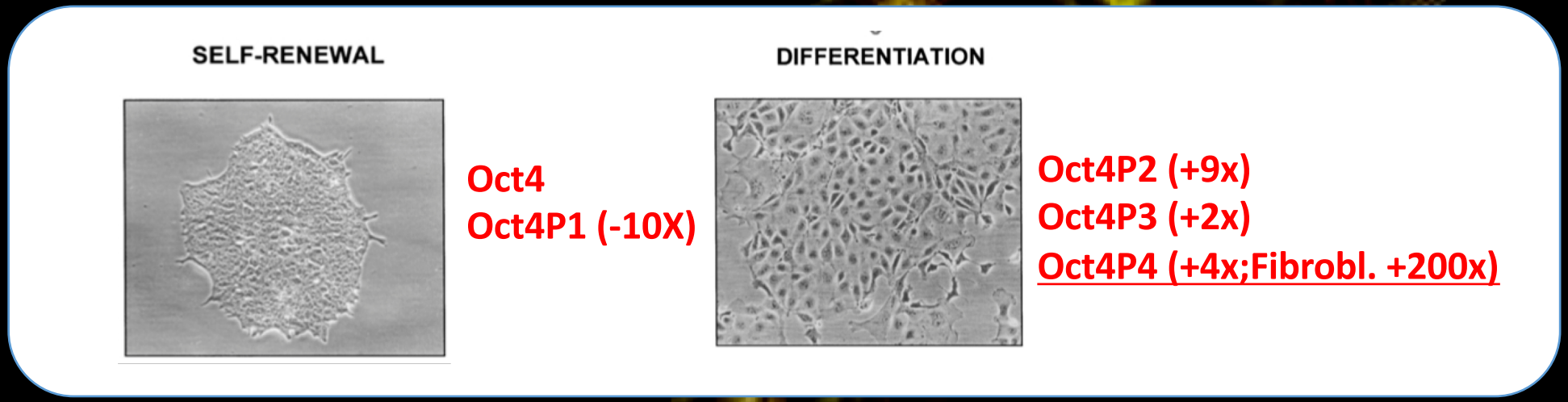
In vitro differentiation; 3D cell culture, embryoid body differentiation



EB: embyoid body differentiation (a model for ES differentiation)

Oct4 pseudogenes are tightly controlled during the differentiation of mESCs

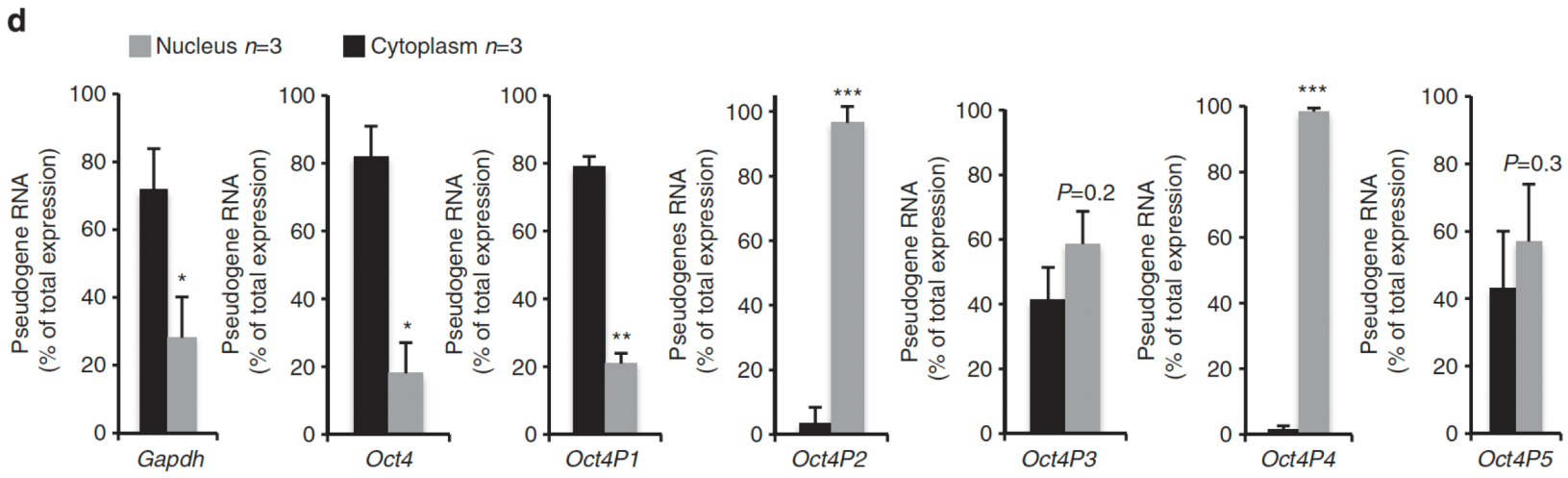


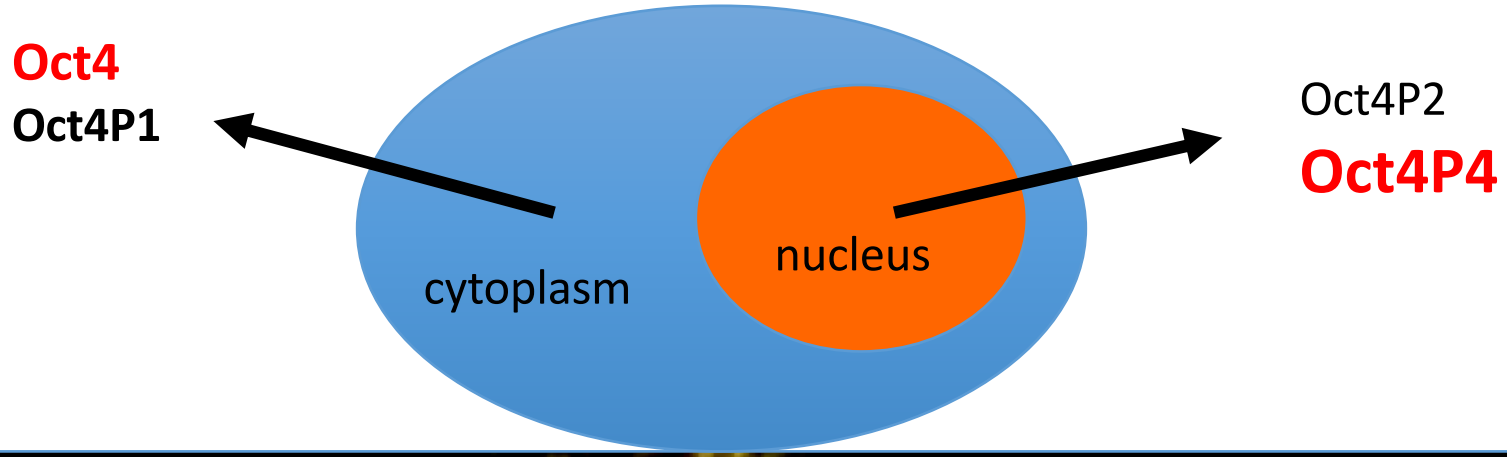


EB: embyoid body differentiation (a model for ES differentiation)

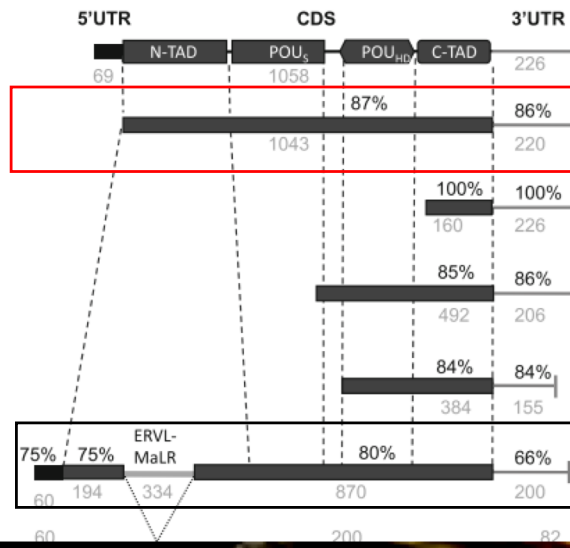
5. LOCALIZATION OF PSEUDOGENE lnCRNA

OCT4 pseudogenes are localized to nucleoplasm or cytoplasm



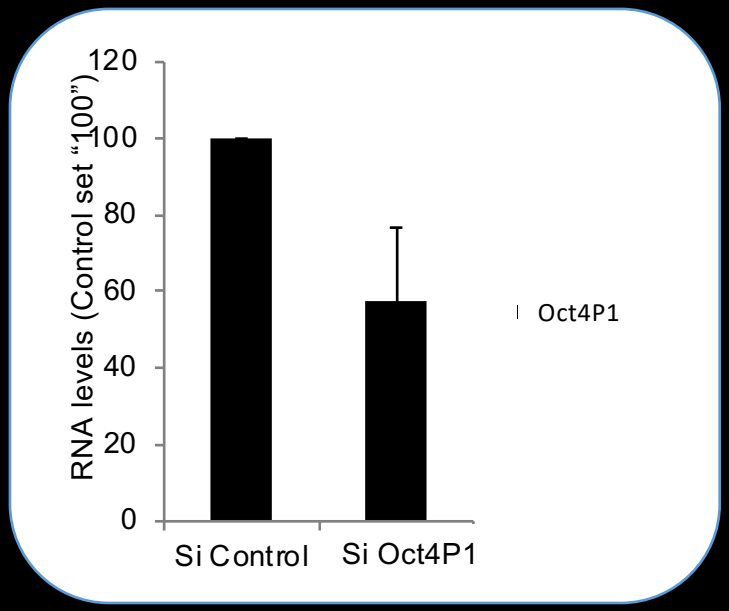


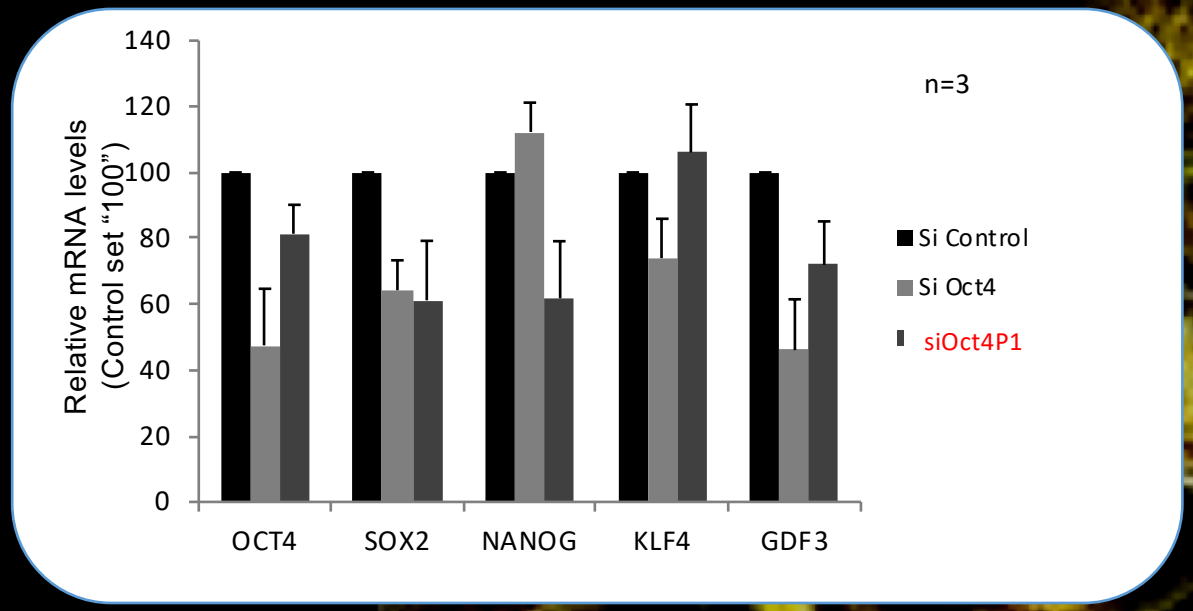
6. FUNCTIONAL ASSAY – lncRNA and self renewal

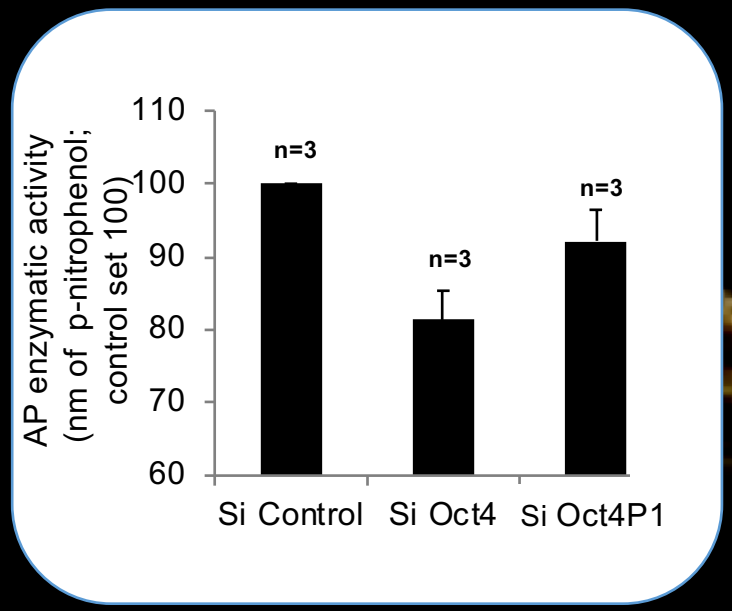


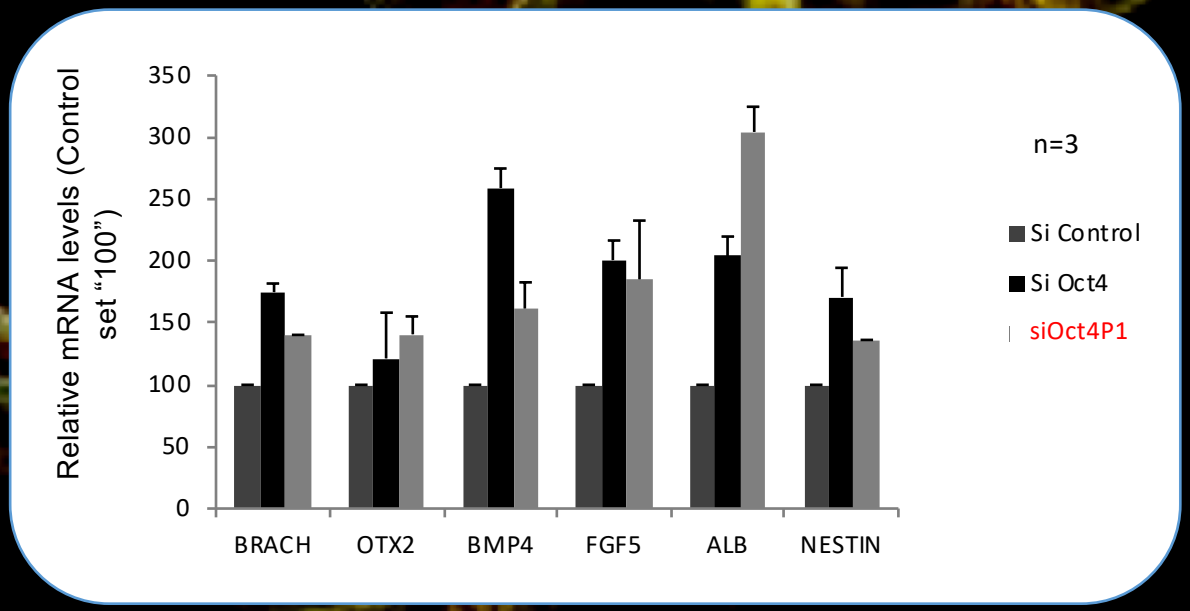
	RNA expression levels: self-renewing mESCs	RNA expression levels: differentiated mESCs	RNA localization
mOct4 mRNA	High	Low	Cytoplasm
mOct4P1	High	Low	Cytoplasm
mOct4P2	Low	High	Nucleus
mOct4P3	Low	Low	Non-specific
mOct4P5	Low	Low	Non-specific
mOct4P4	Low	High	Nucleus

6. FUNCTIONAL ASSAY – lncRNA and self renewal



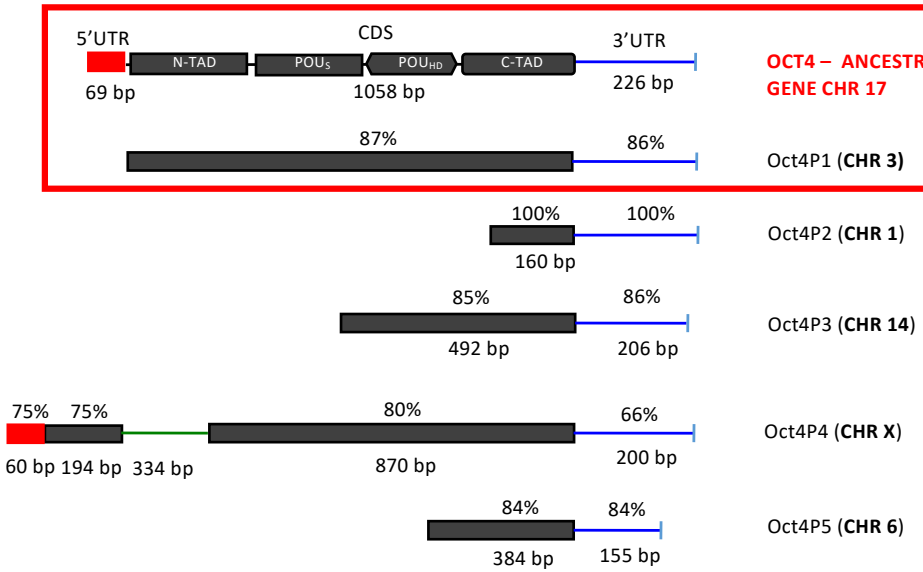


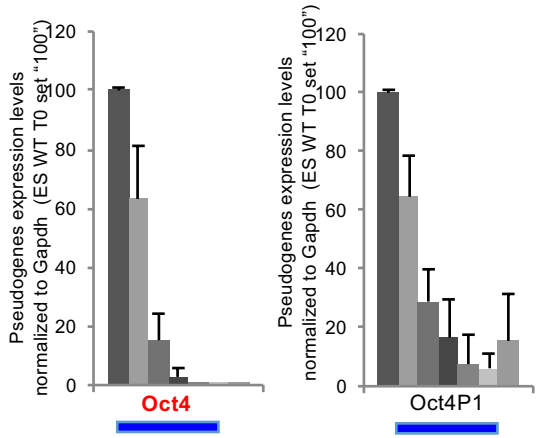




Cytoplasmatic OCT4P1 promotes mESC self-renewal

Ancestral OCT4 gave rise to 5 processed pseudogenes that are expressed in mESCs





siOct4: high in ESC

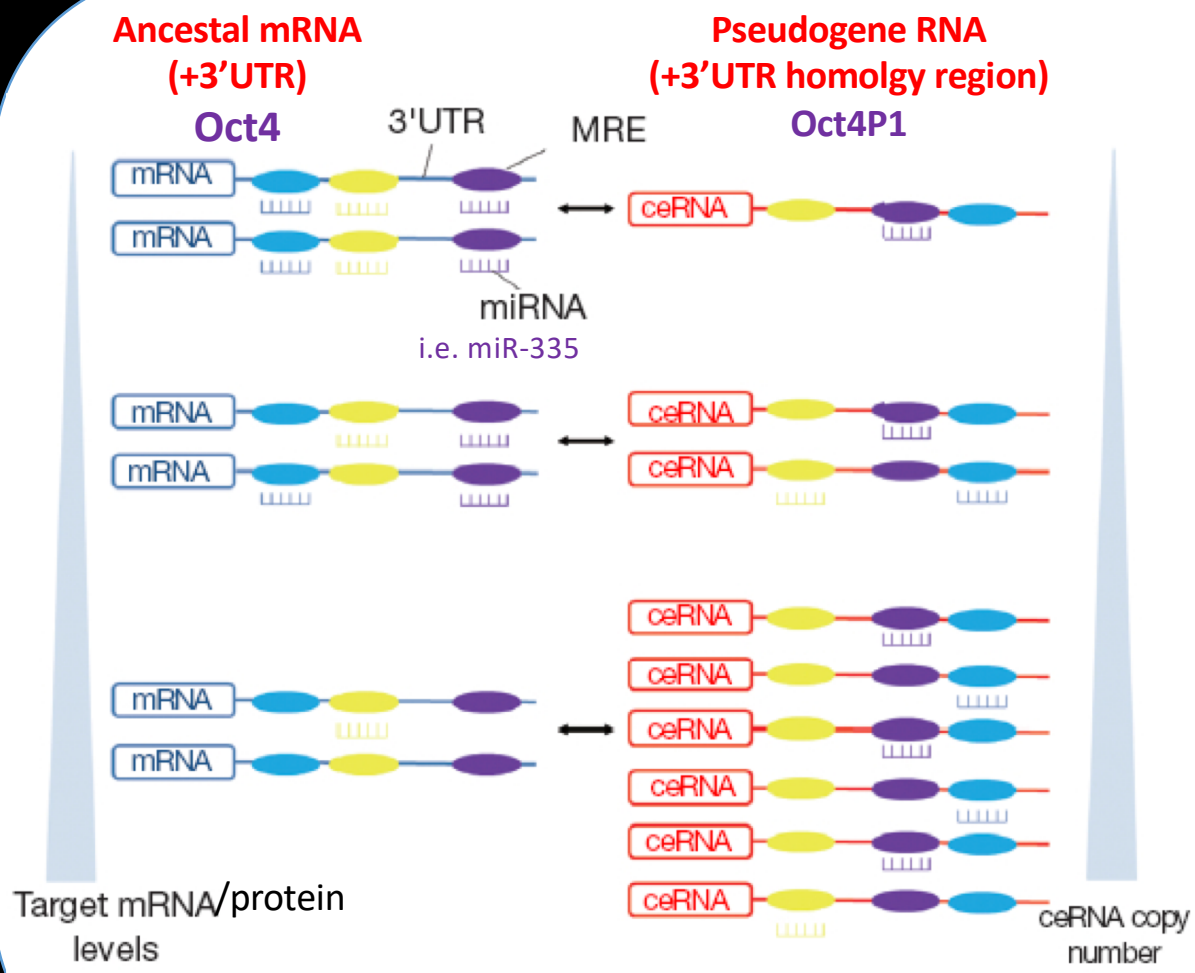
siOct4P1: high in ESC, low in differentiated cells

siOct4: differentiation

siOct4P1: differentiation

Oct4P1 needed to keep high Oct4 expression in self renewing ESC??

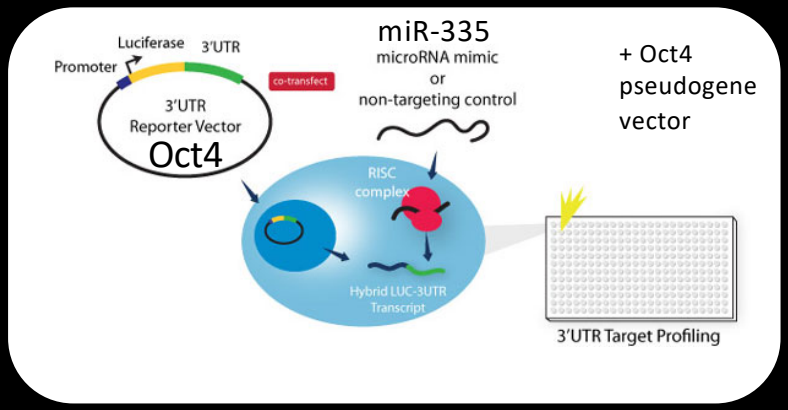
Pseudogene sponge miRNAs that target the ancestral gene

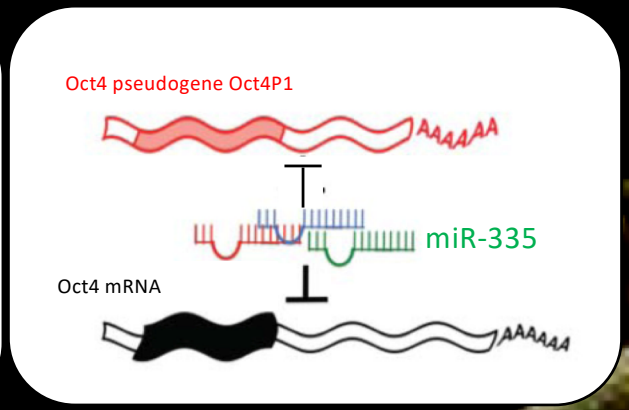


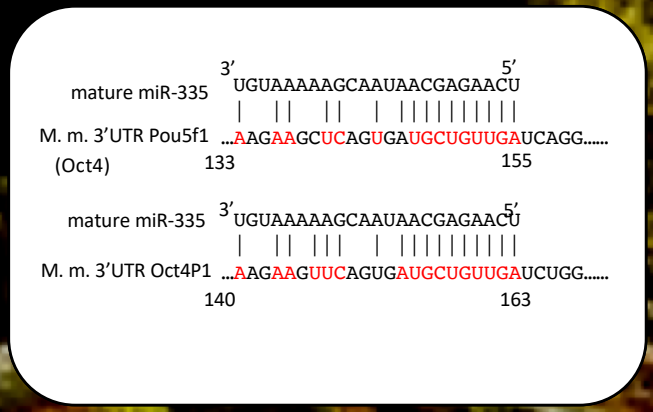
miRNA that targets Oct4: i.e. miR-335
MRE: miRNA response element

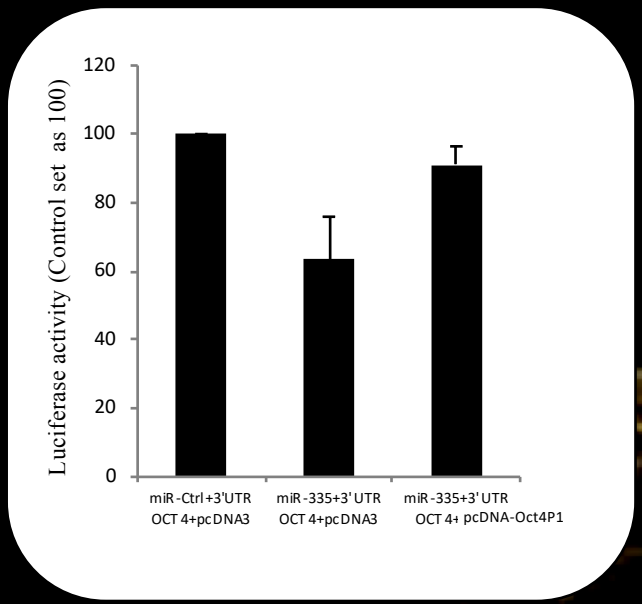
Evolution of pseudogenes to fine-tune the expression of ancestral genes

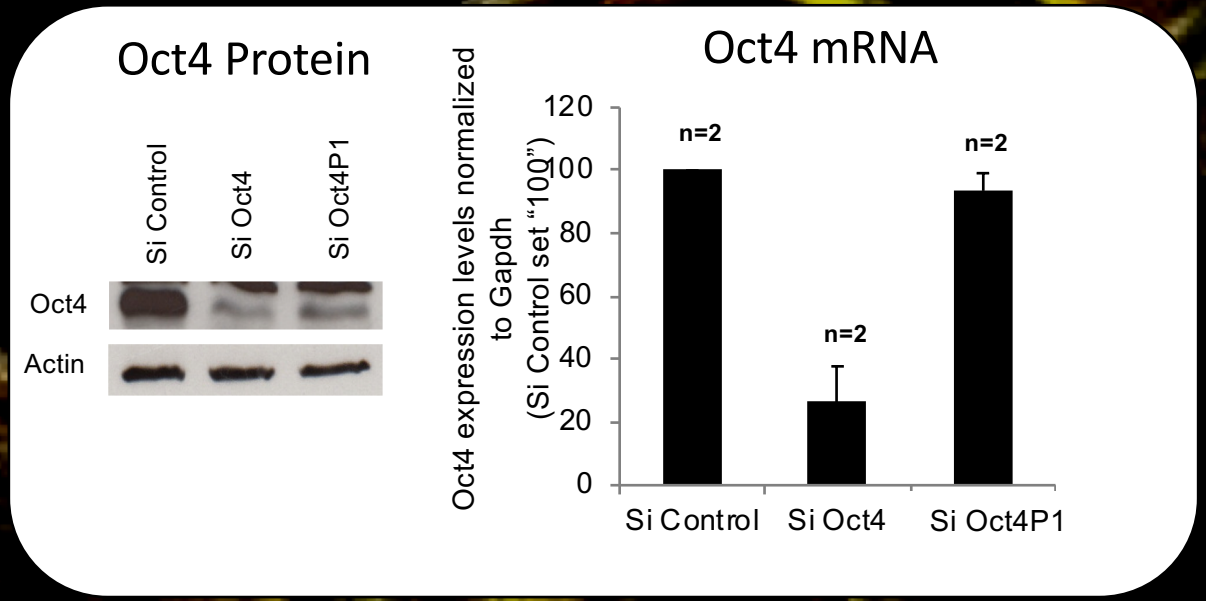
Cytoplasmatic OCT4P1 acts as Oct4/Rb1 ceRNA





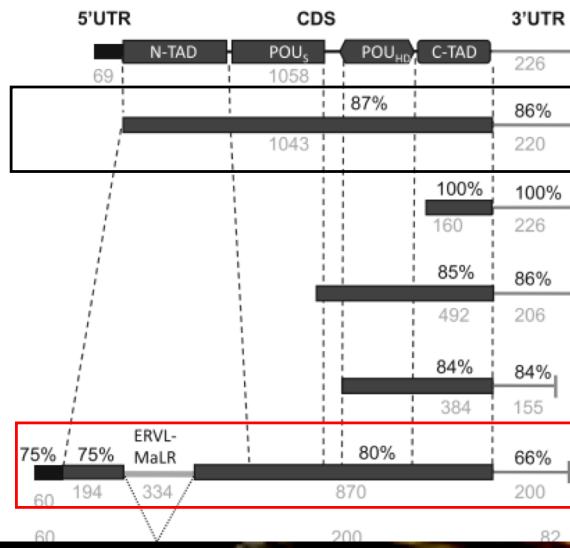






OCT4P1 promotes self-renewal by sponging miRNAs that target Oct4

6. FUNCTIONAL ASSAY – lncRNA and self renewal

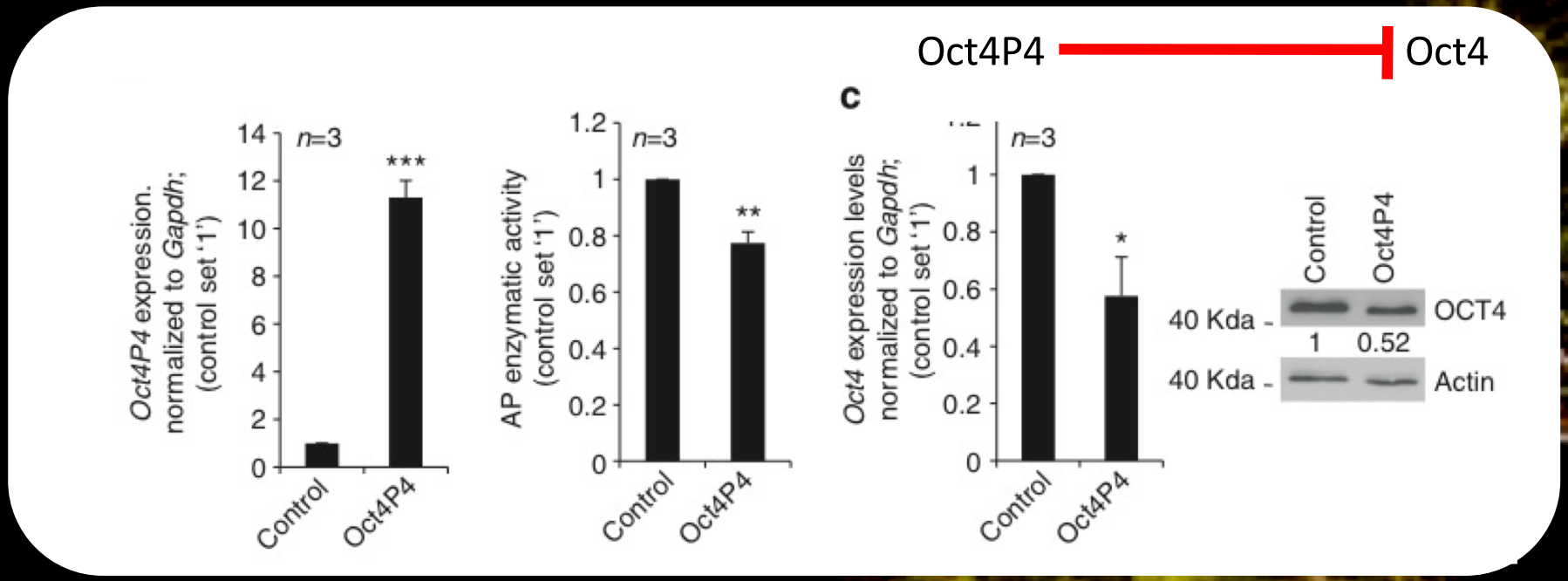


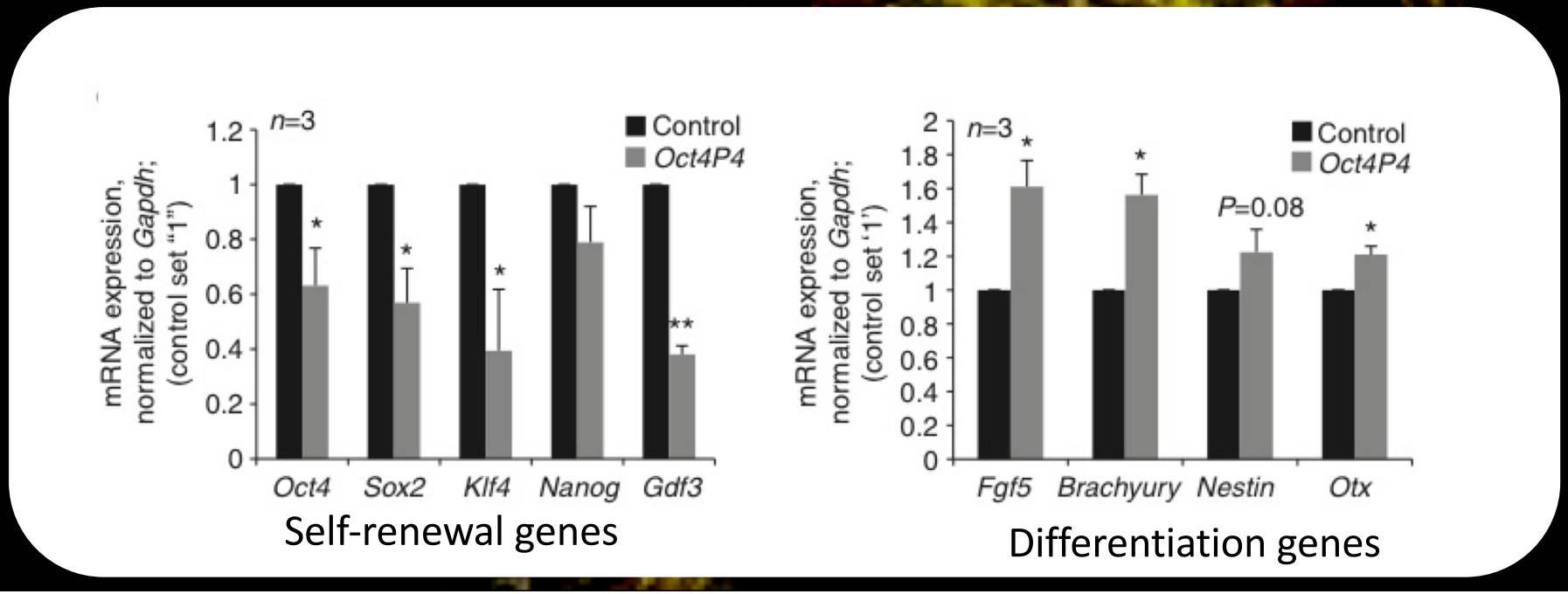
	RNA expression levels: self-renewing mESCs	RNA expression levels: differentiated mESCs	RNA localization
mOct4 mRNA	High	Low	Cytoplasm
mOct4P1	High	Low	Cytoplasm
mOct4P2	Low	High	Nucleus
mOct4P3	Low	Low	Non-specific
mOct4P5	Low	Low	Non-specific
mOct4P4	Low	High	Nucleus

Scarola et al. Nat Comms. 2015

Overexpression of OCT4P4 down regulates Oct4 to promote mESC differentiation

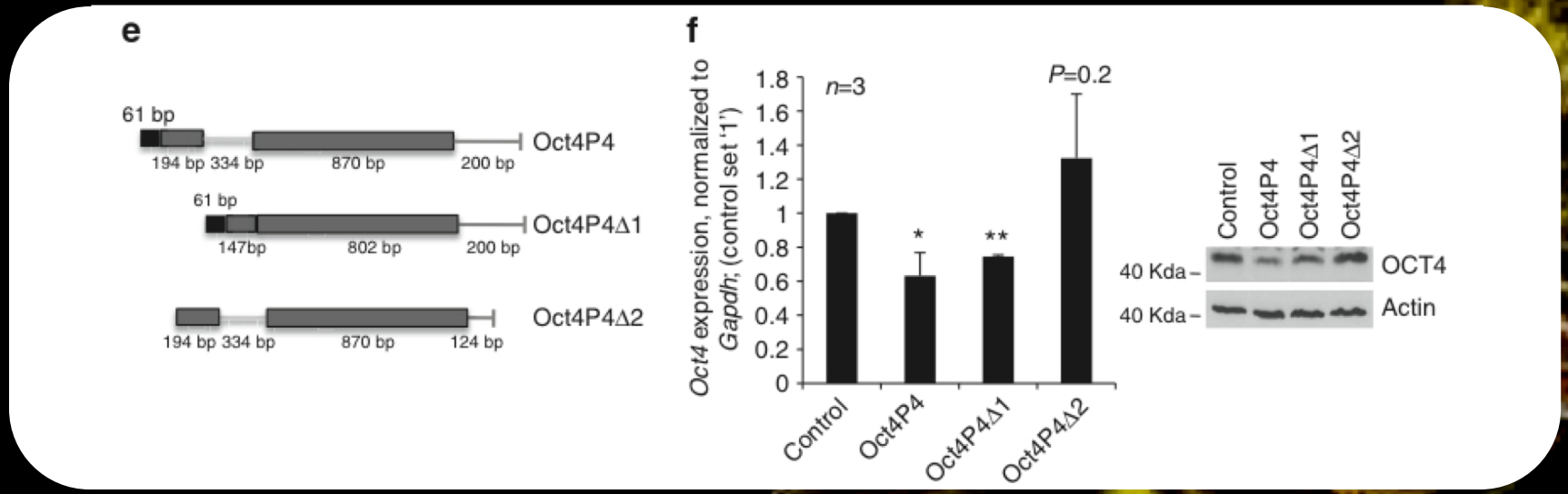
mESCs with stable *mOCT4P4* expression construct



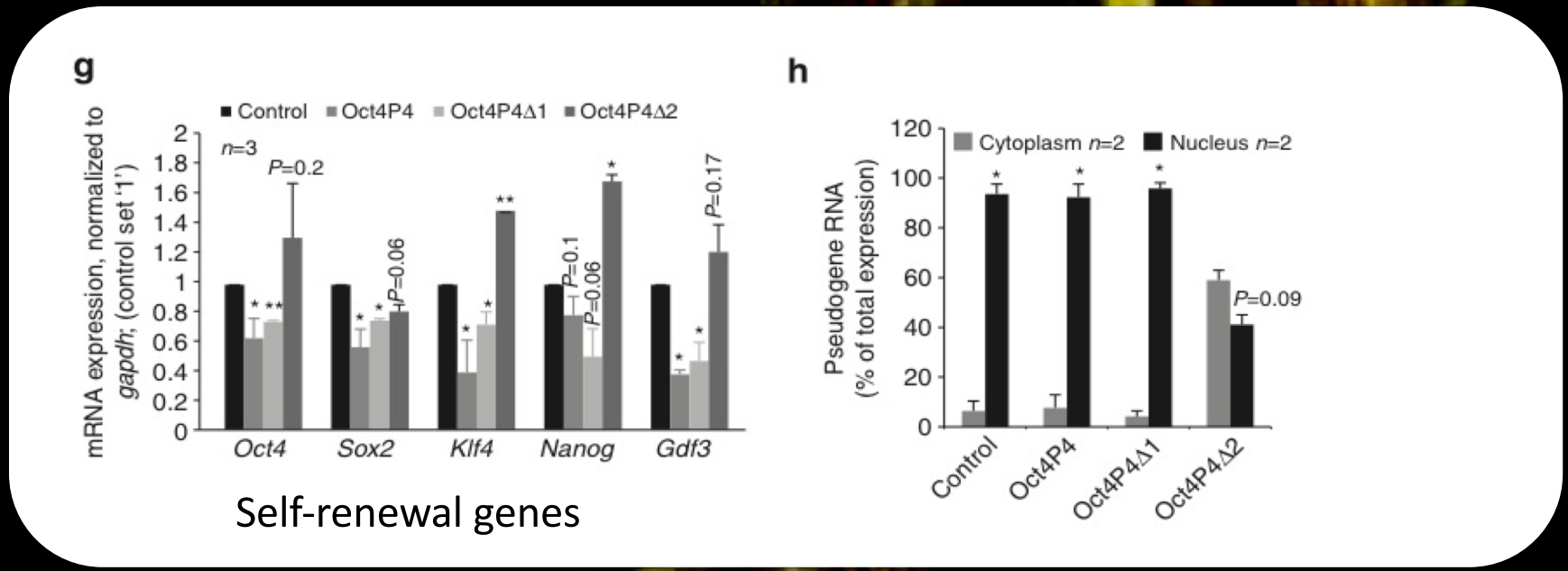


7. MOLECULAR ANALYSIS: WHAT ARE THE RELEVANT SEQUENCES??

Express Oct4P4 deletion constructs and follow parental gene expression



Express Oct4P4 deletion constructs and follow self-renewal programs



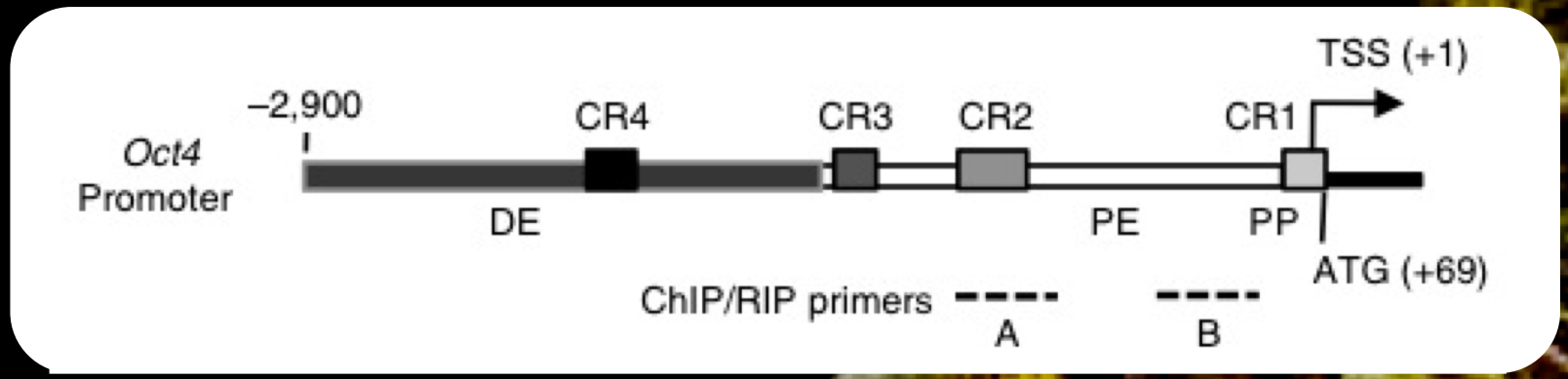
Self-renewal genes

5′ and 3′ UTR homology regions are required to repress self-renewal marker genes

8. MOLECULAR ANALYSIS: USE SIMPLIFIED MODEL TO GET INDICATION ON MOLECULAR MECHANISM??

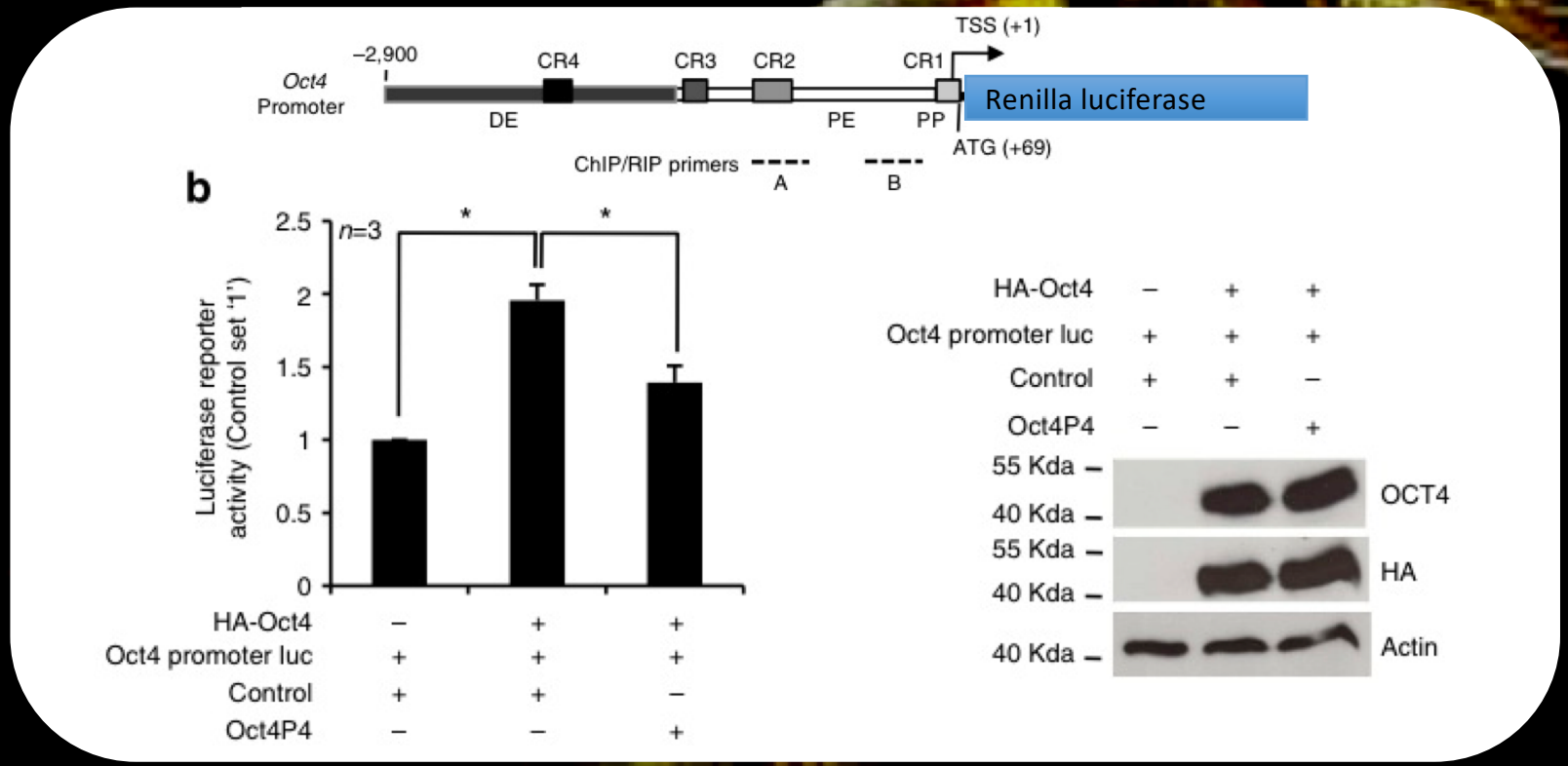
Oct4P4 is a nuclear RNA that suppresses Oct4 transcription

Hypothesis: transcriptional suppression of Oct4 promoter?



Use simplified models to support hypothesis:

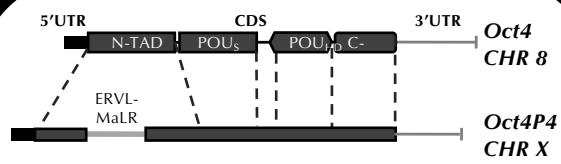
- Clone Oct4 promoter
- Clone luc reporter downstream
- Test wehter Oct4P4 lncRNA suppresses reporter

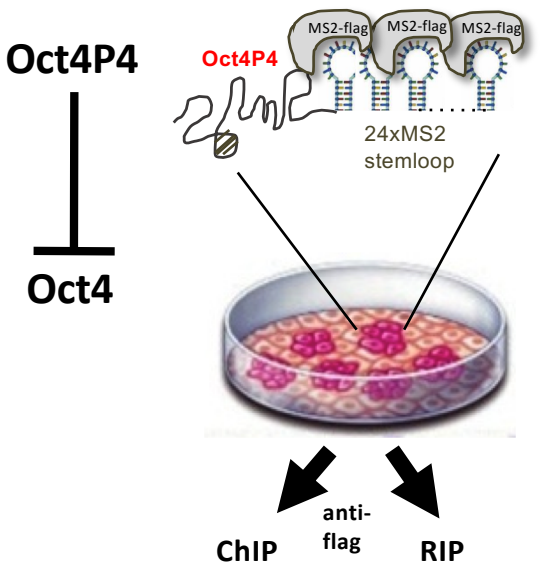


Oct4P4 <u>may</u> interfere with the ancestral Oct4 promoter

9. MOLECULAR ANALYSIS: DEMONSTRATE MEACHINISM AT ENDOGENOUSE Oct4 PROMOTER??

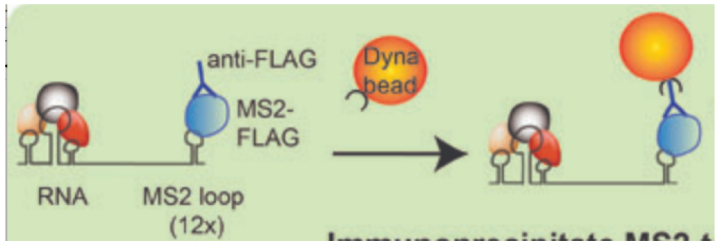
Generate model system to:
1. study Oct4P4 lncRNA localization
2. study chromatin at OCT4 promoter

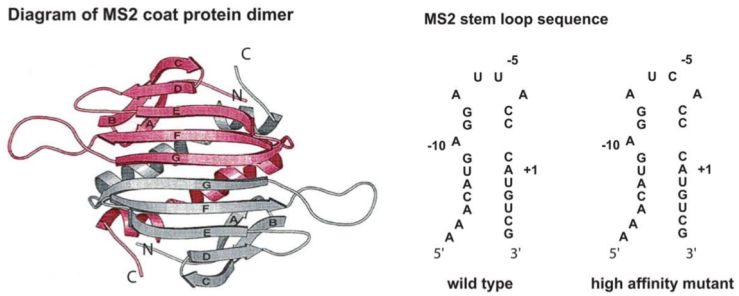




Step 1: stably transfect mESCs with MS2-flag

Step 2: stably transfect mESCs with an expression vector encoding mOct4P4 lncNRA fused to 24 MS loop stem loops

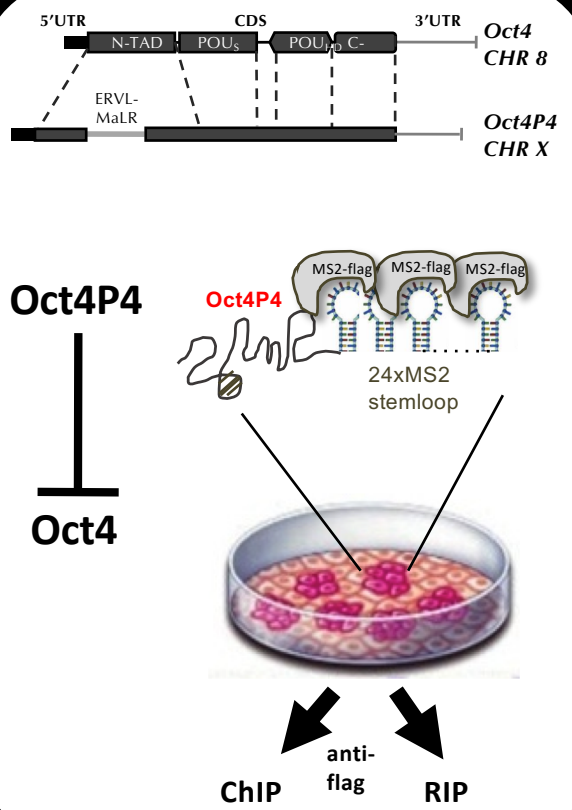


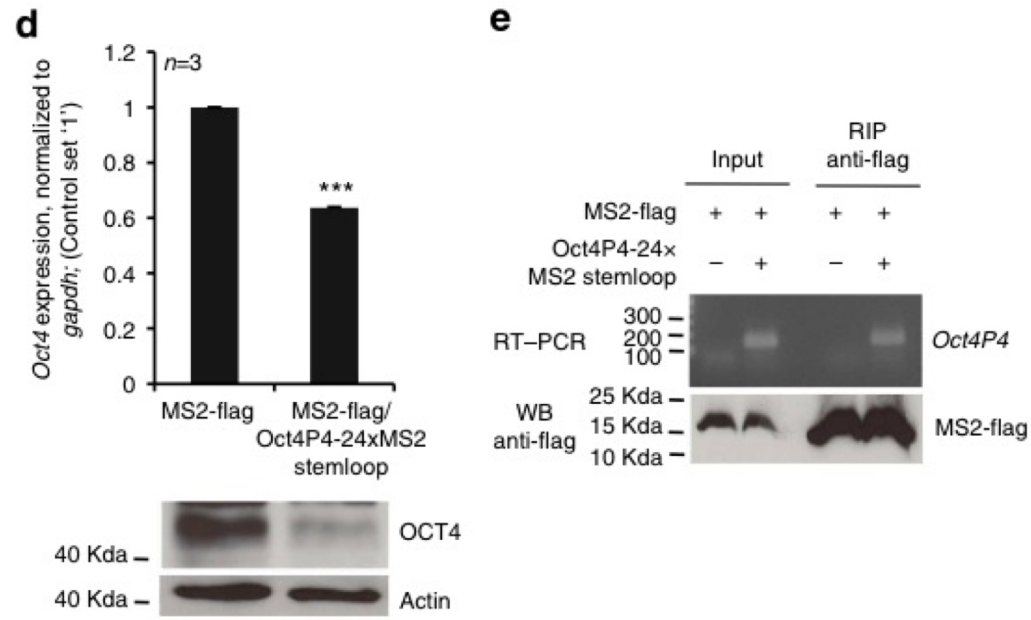


MS2 stem loop tagged Oct4P4 co-expressed with flag-MS2 in mESCs

9. MOLECULAR ANALYSIS: DEMONSTRATE MEACHINISM AT ENDOGENOUSE Oct4 PROMOTER??

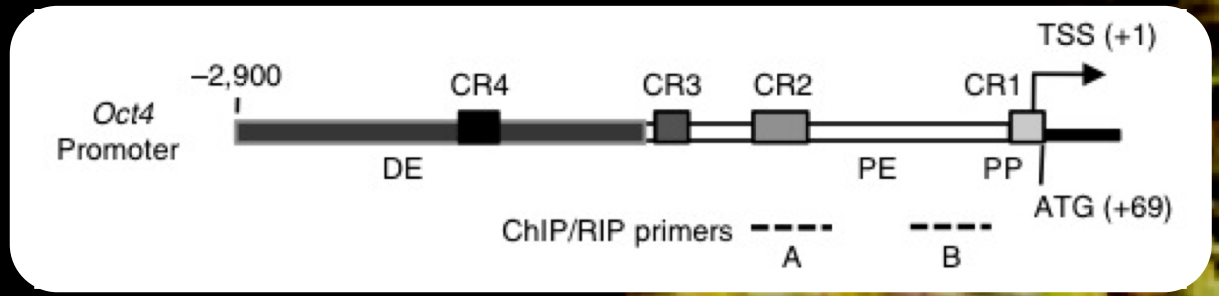
Generate model system to:
1. study Oct4P4 lncRNA localization
2. study chromatin at OCT4 promoter

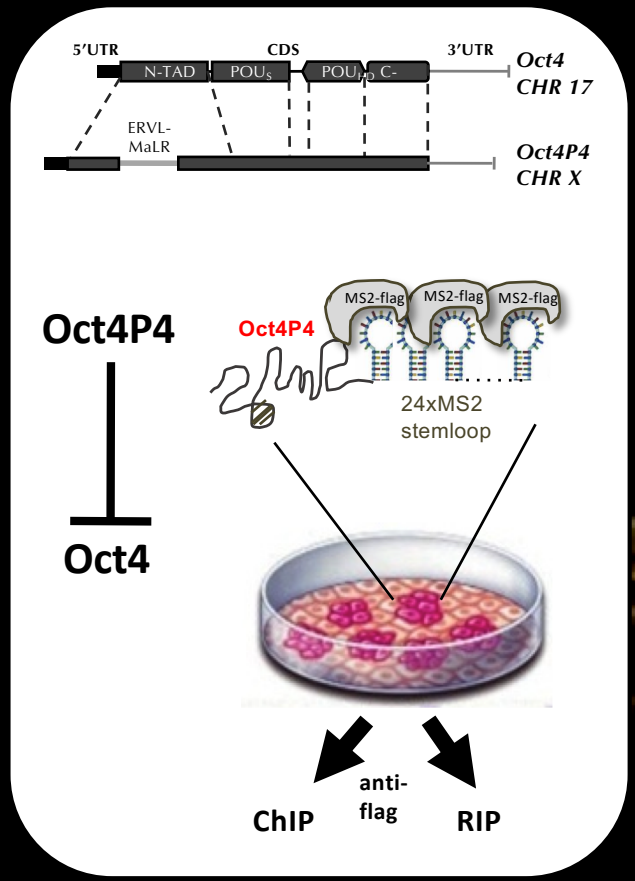




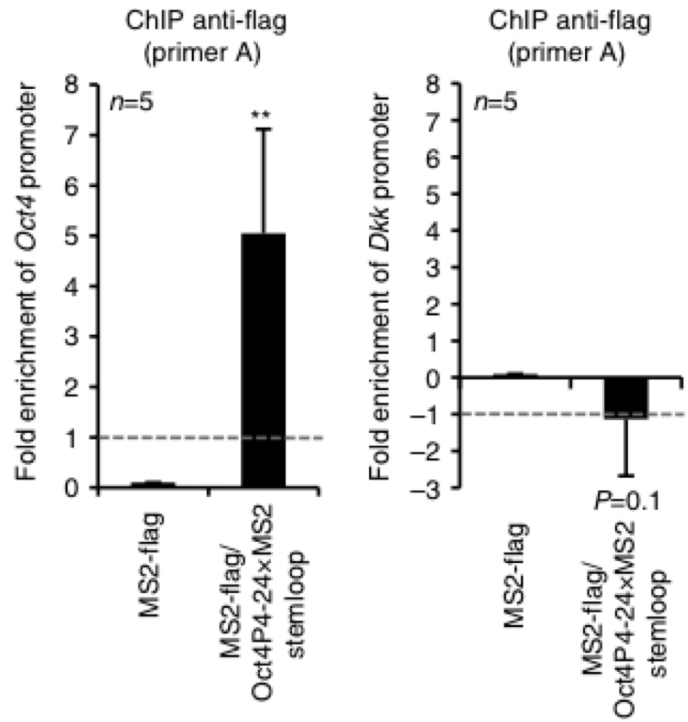
MS2 stem loop tagged Oct4P4 co-expressed with flag-MS2 in mESCs

A model system to study Oct4P4 lncRNA localization





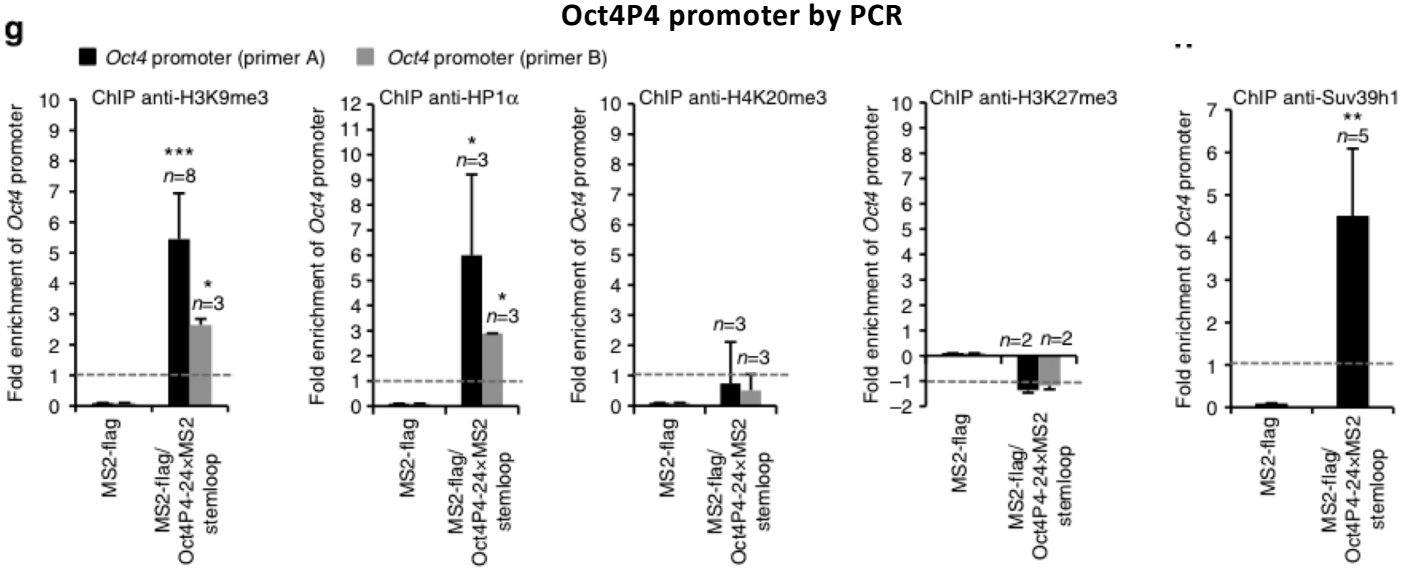
ChIP using anti-flag then use the immuno-precipitate to detect the Oct4P promoter by PCR

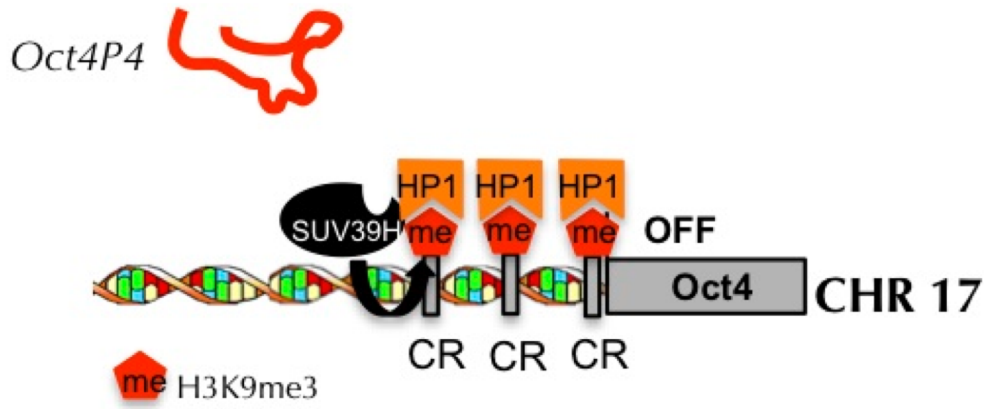


Oct4P4-MS2 lncRNA localizes to Oct4 promoter

Oct4P4-MS2 directs Suv39h1 to Oct4 promoter

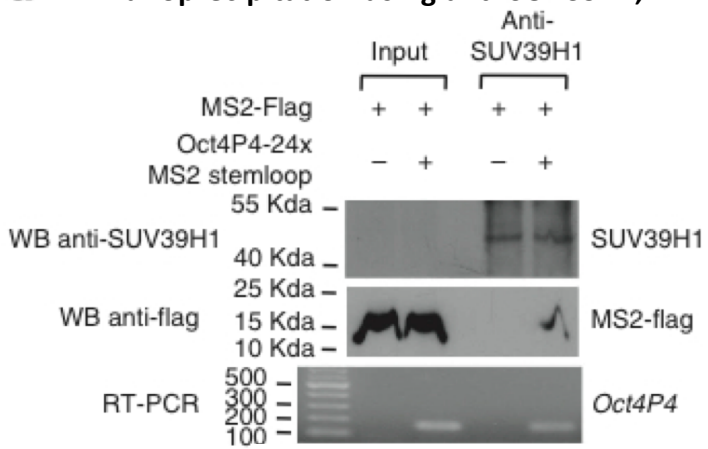
ChIP using specific antibodies then use the immuno-precipitate to detect the Oct4P4 promoter by PCR



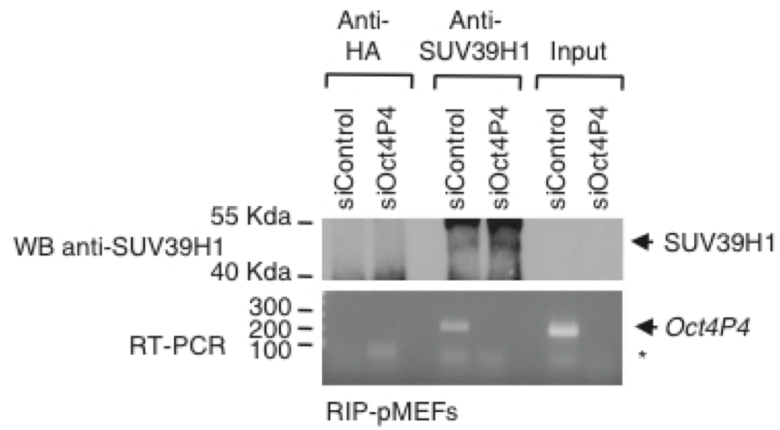


Oct4P4-MS2 directly interacts with Suv39h1

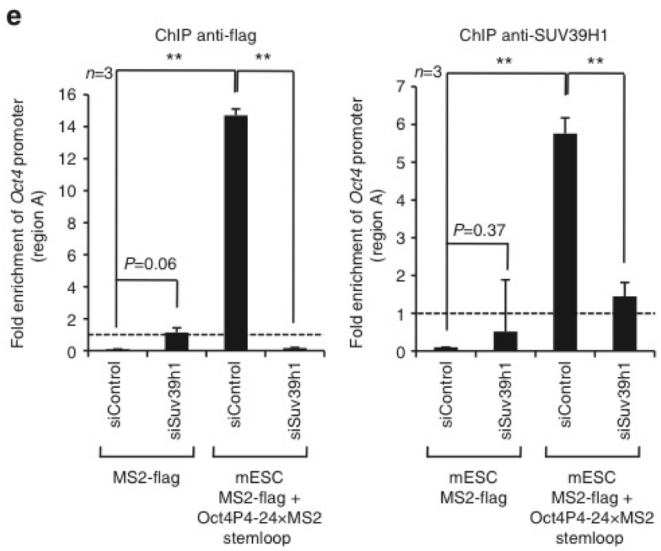
RNA immunoprecipitation using anti-SUV39h1;



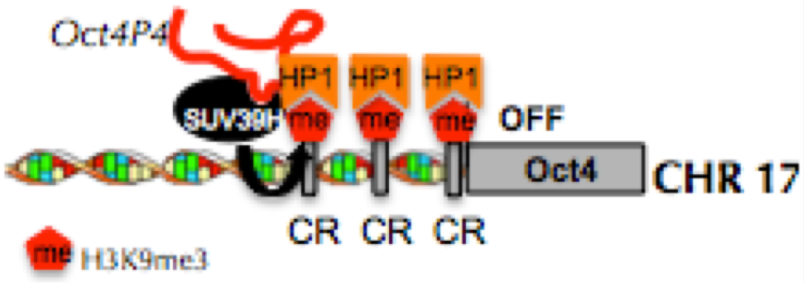
Western for MS2-flag



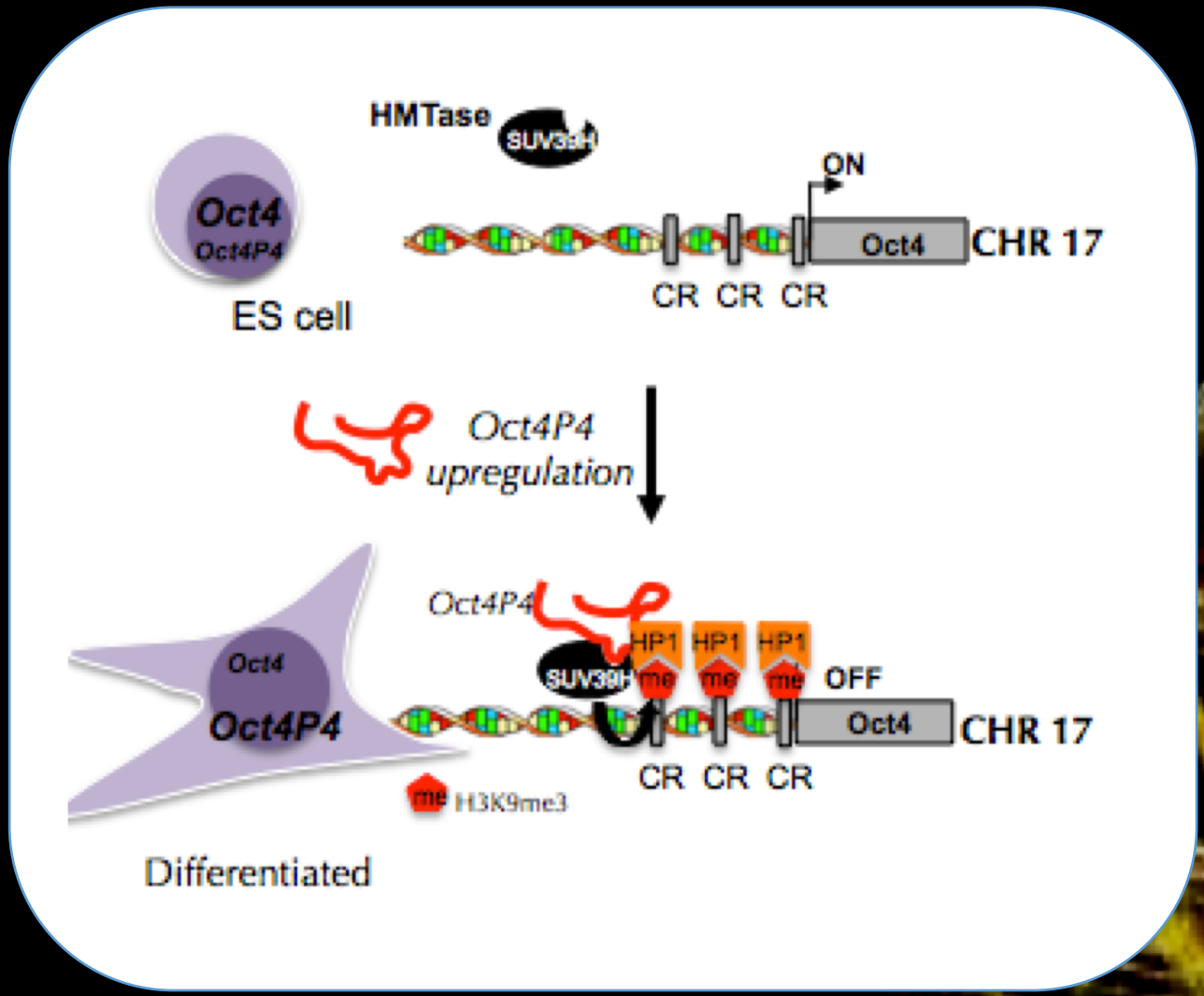
RT-PCR for Oct4P4 lncRNA



ChIP using specific antibodies then use the immuno-precipitate to detect the Oct4P4 promoter by PCR



Oct4P4-MS2 recruits Suv39h1
To direct silencing of the Oct4 promoter

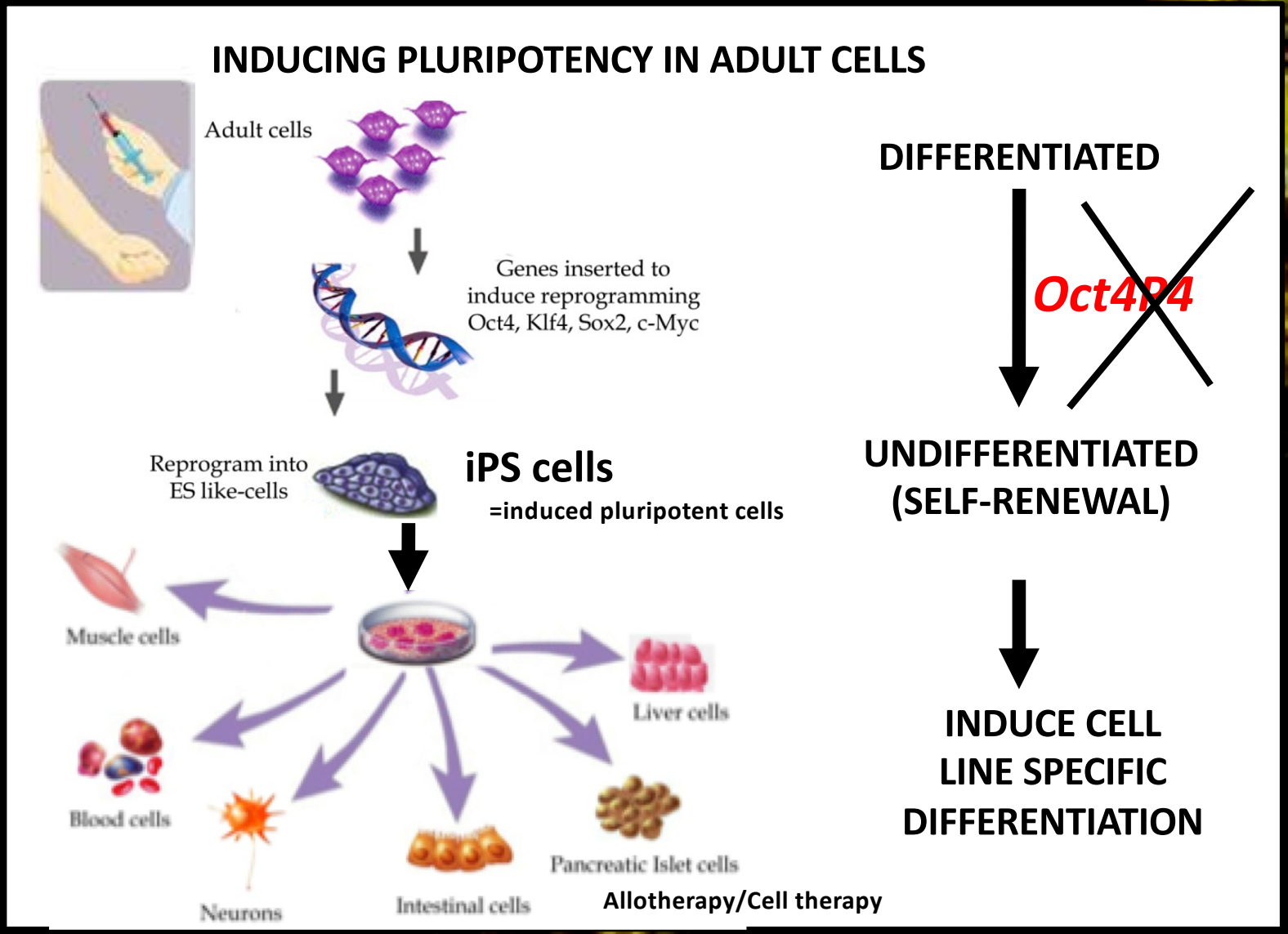


FUNCTIONAL RELEVANCE????

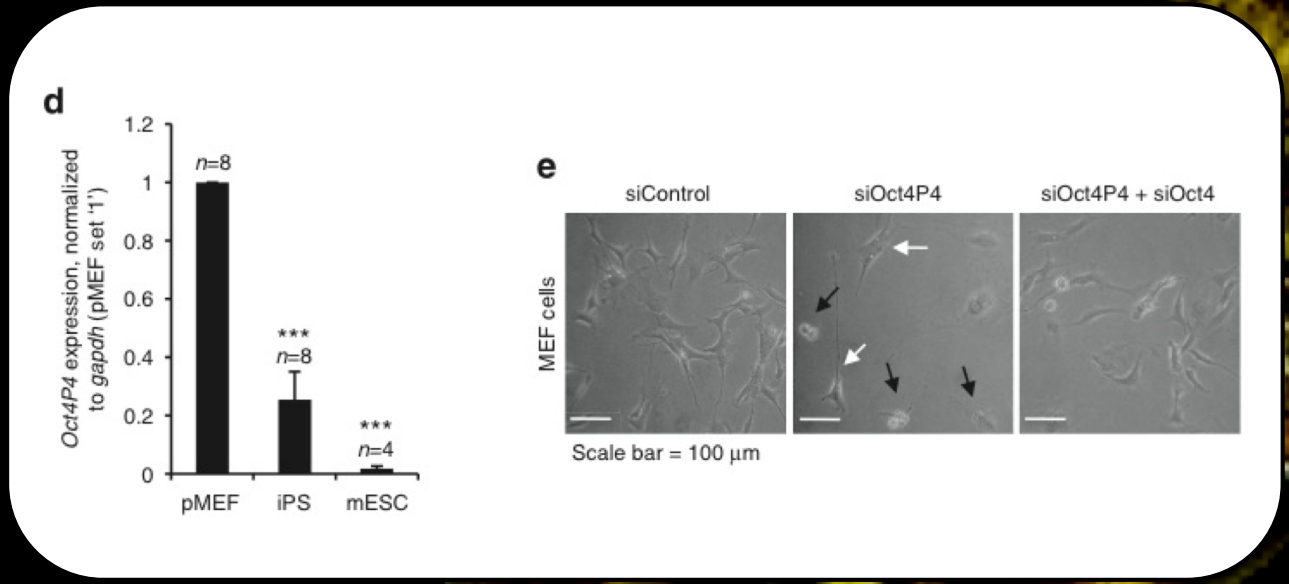
REVERSIBILITY???

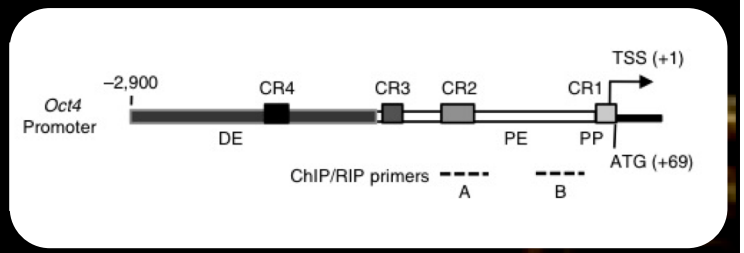
10. Functional relevance:

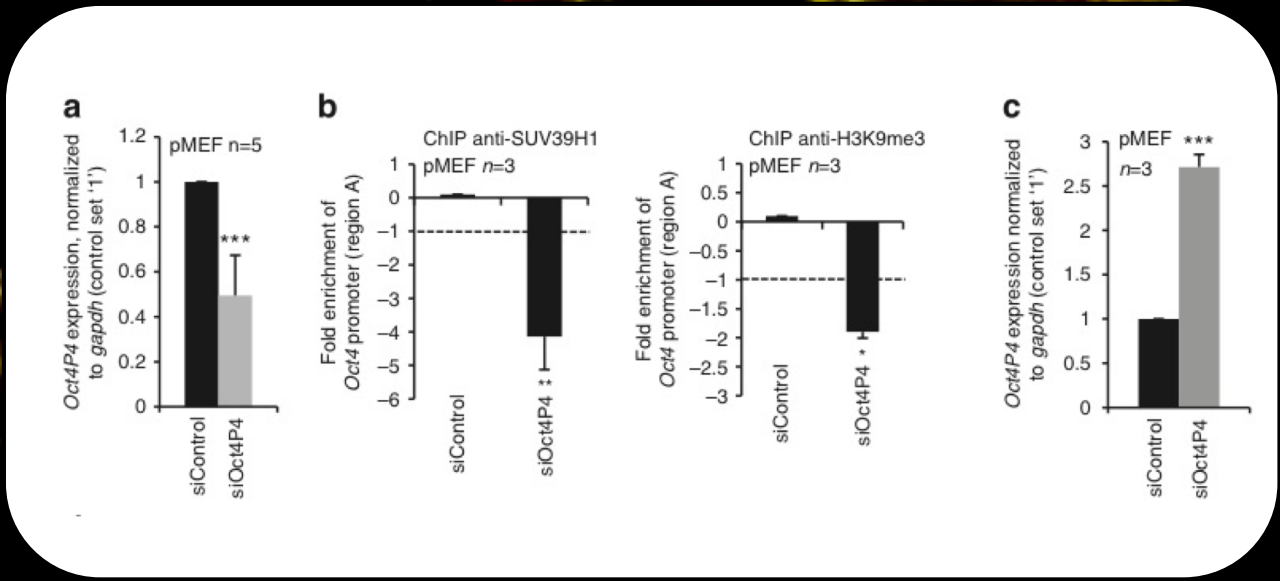
Oct4P4 depletion in pMEFs causes the re-acquisition of self-renewal features



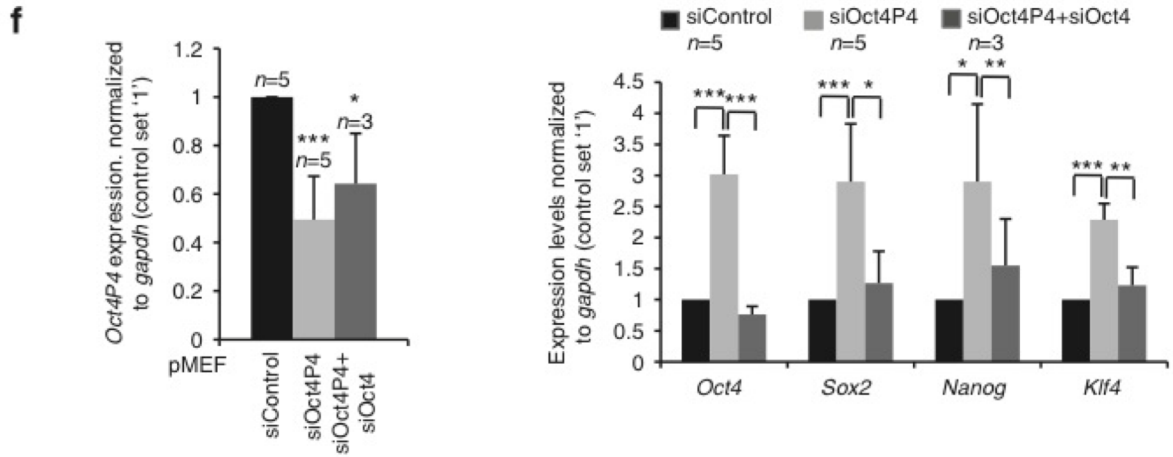
Oct4P4 depletion in pMEFs causes the re-acquisition of self-renewal features

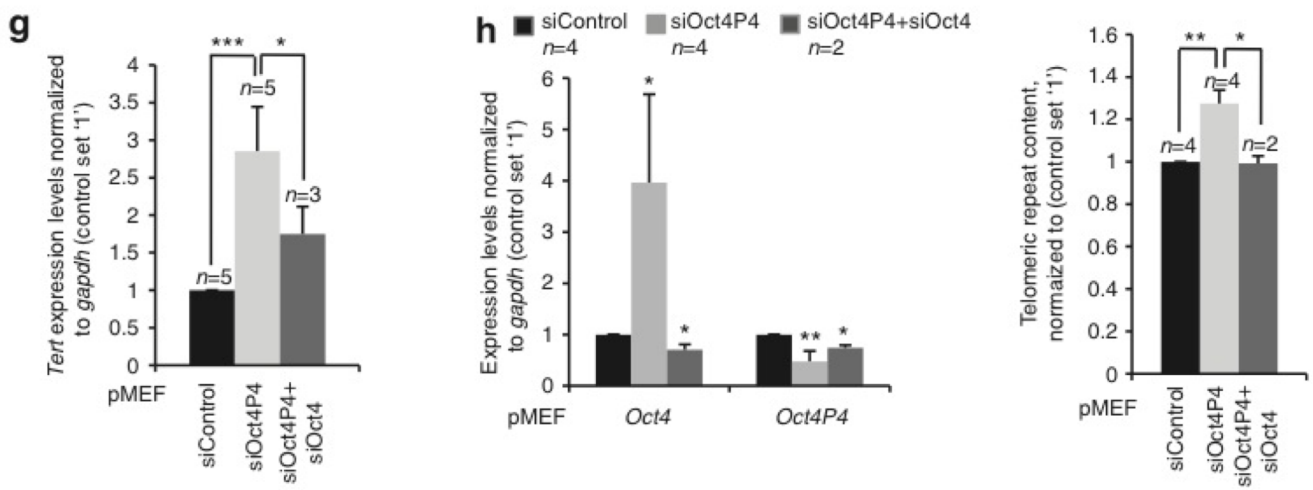




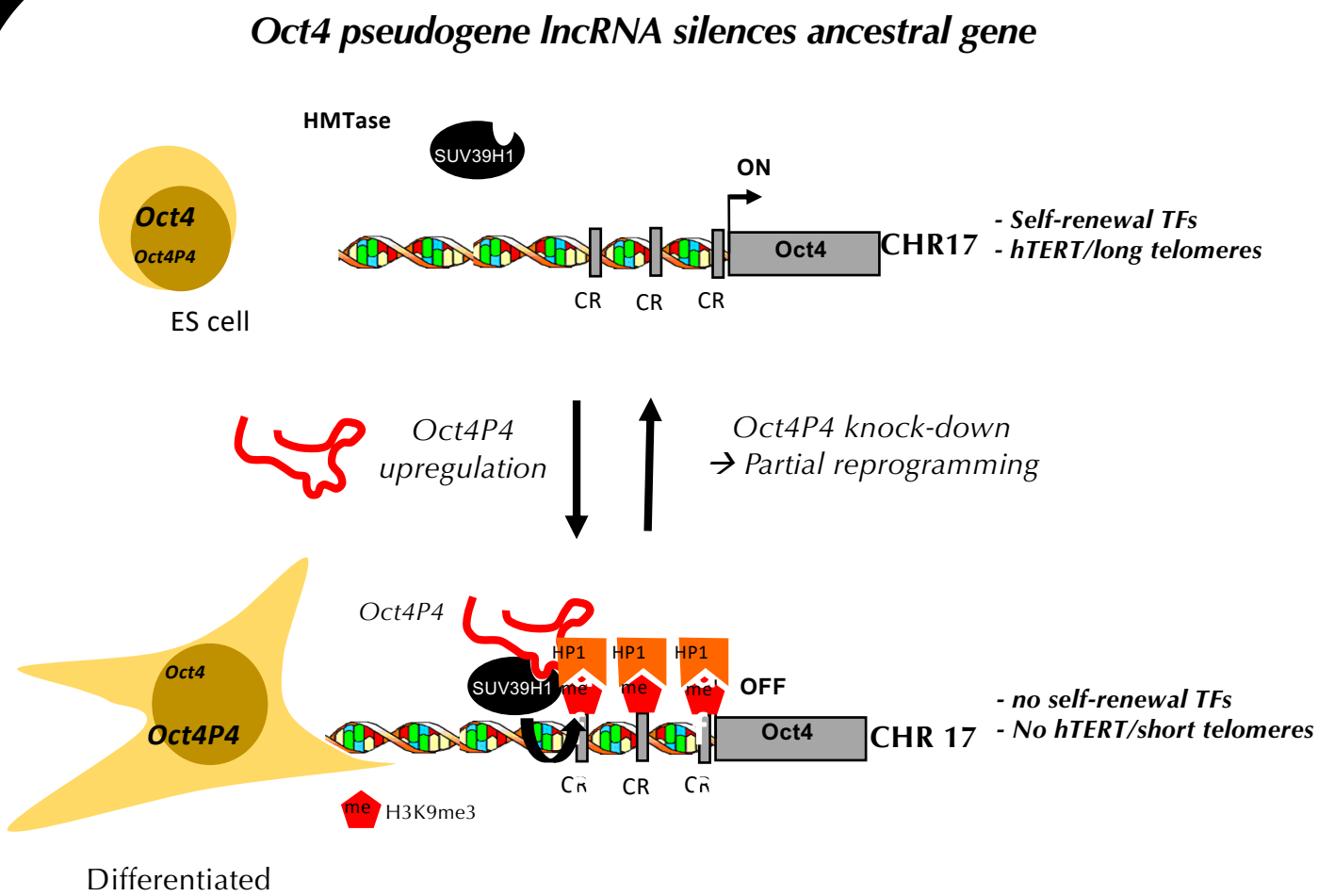


Oct4P4 depletion in pMEFs causes the re-acquisition of self-renewal features

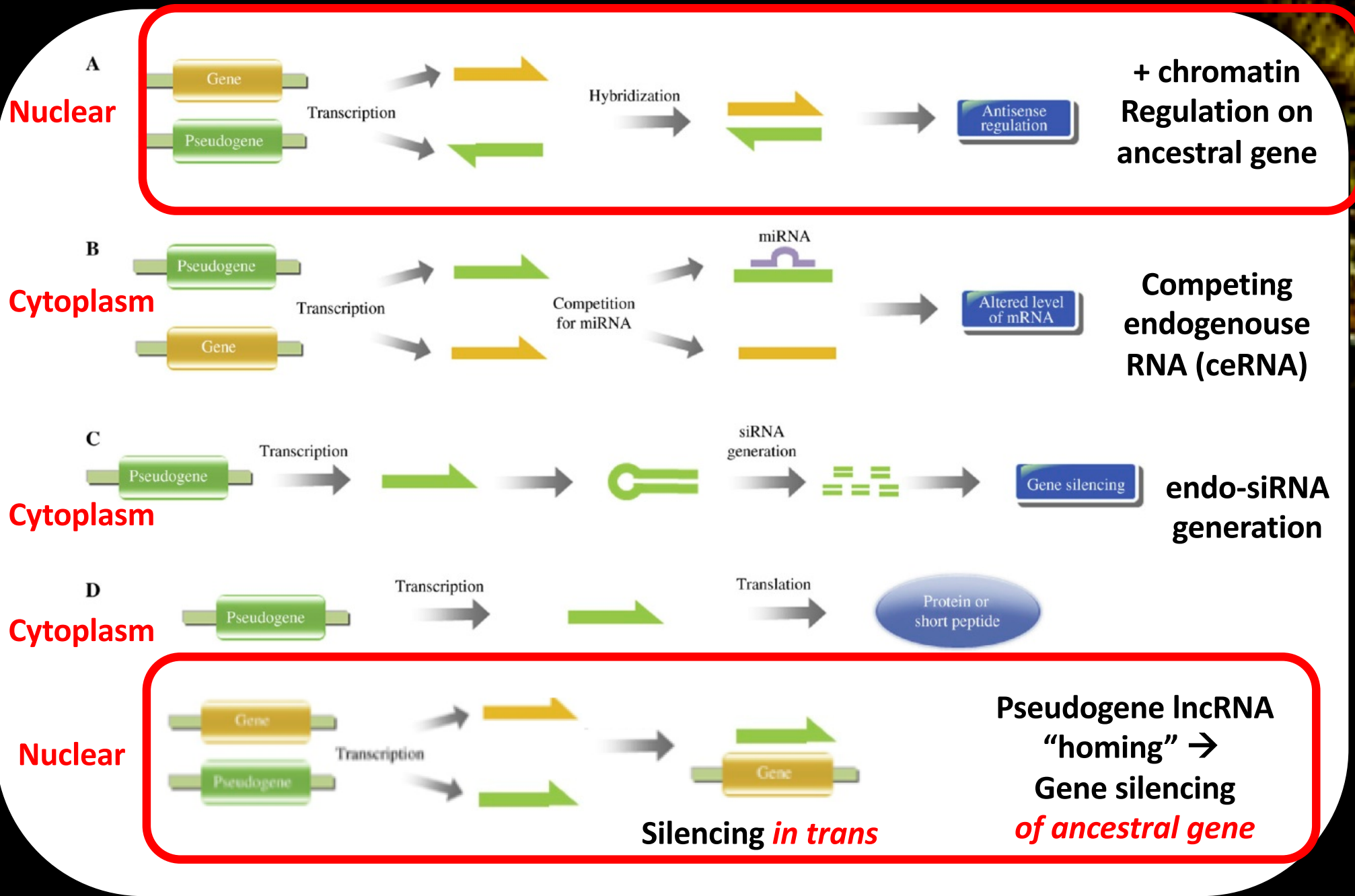




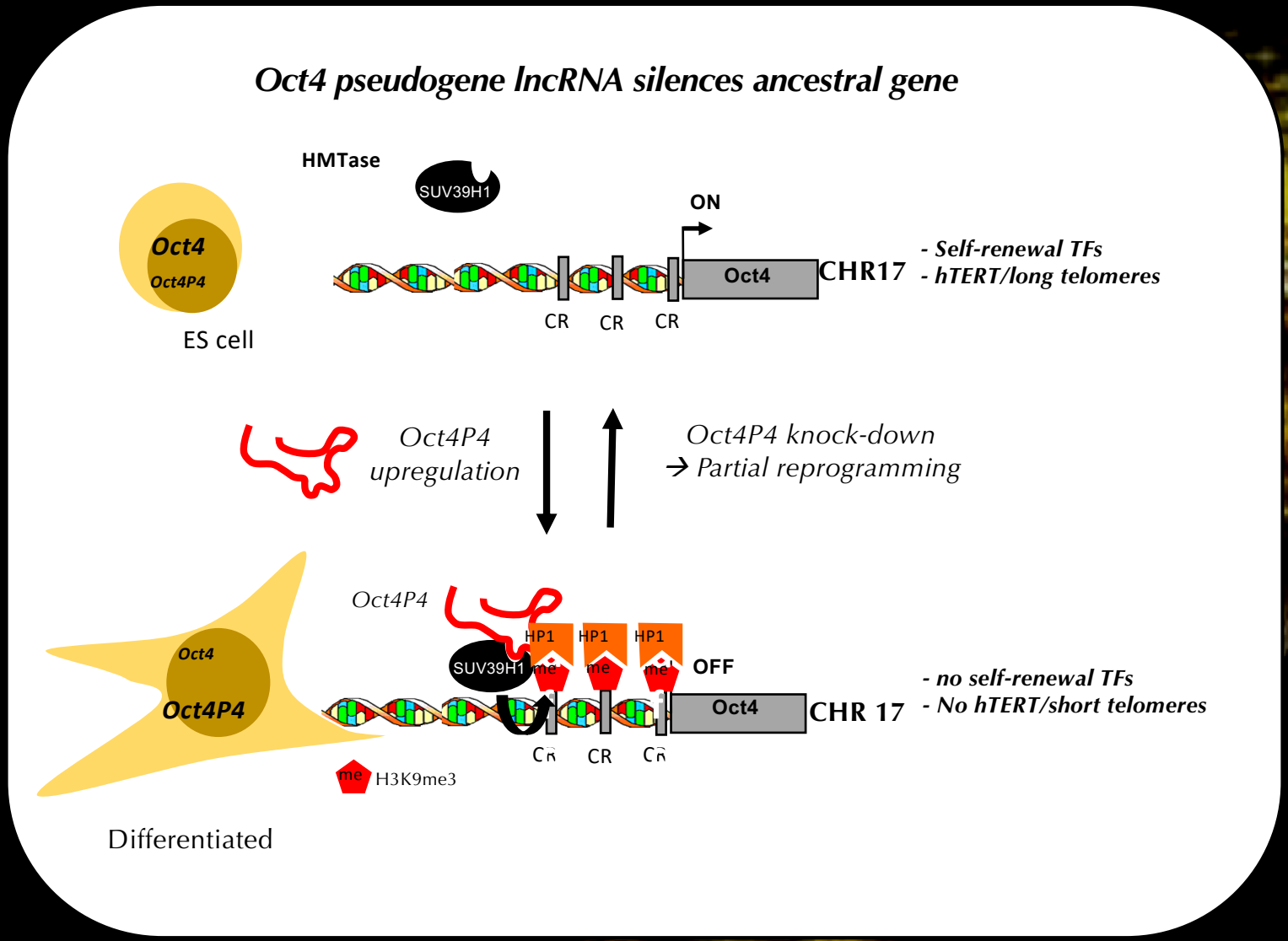
Pseudogenes control the epigenetic status of ancestral genes



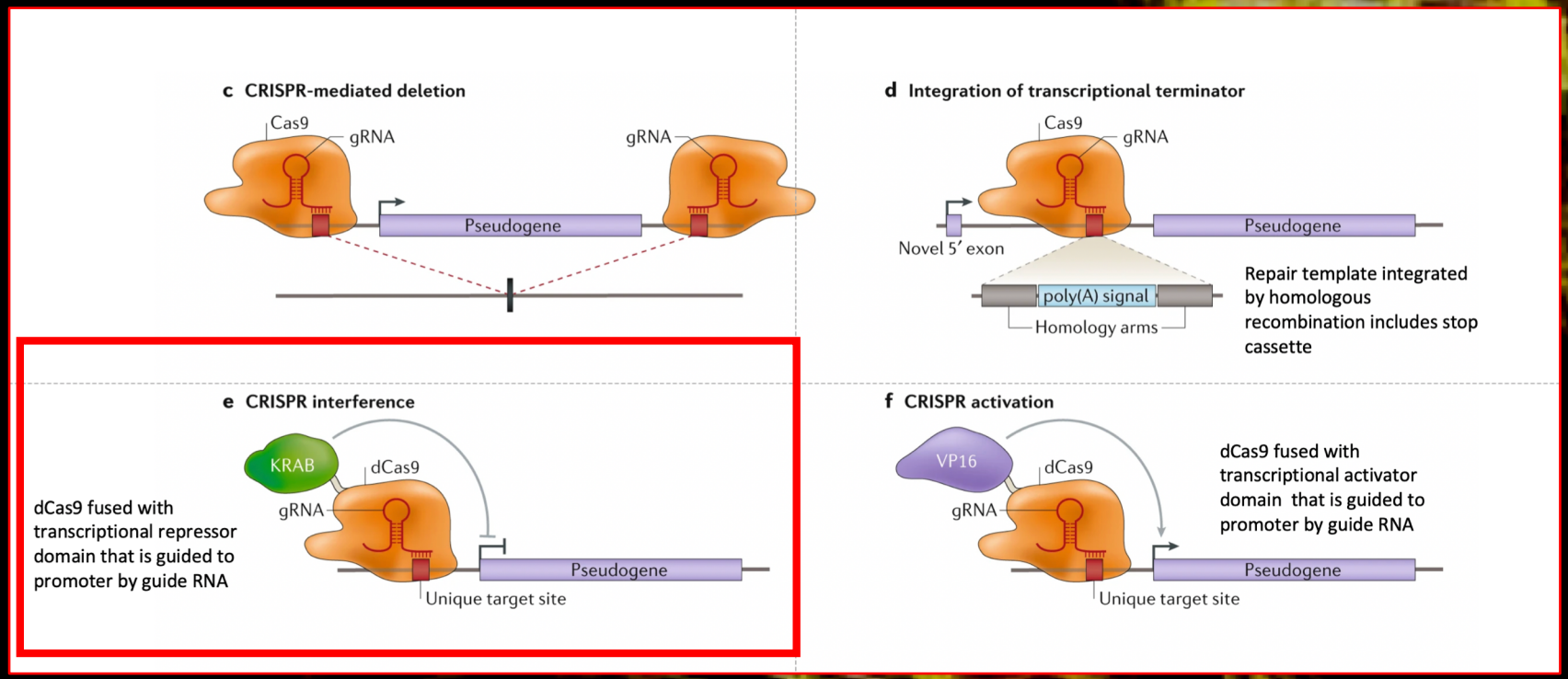
Pseudogenes are powerful regulators of gene expression



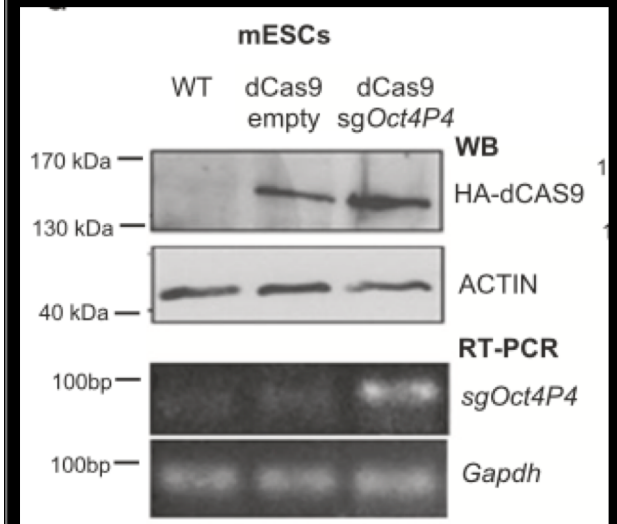
WHAT WOULD YOU DO NEXT TO UNDERSTAND OCT4P4 FUNCTION?



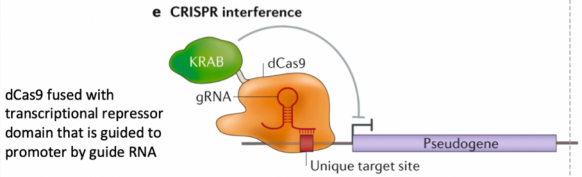
- GENETIC MODEL SYSTEM
- FUNCTIONAL RELEVANT REGIONS
- INTERACTING PROTEINS



GENETIC Oct4P4 loss of function model system

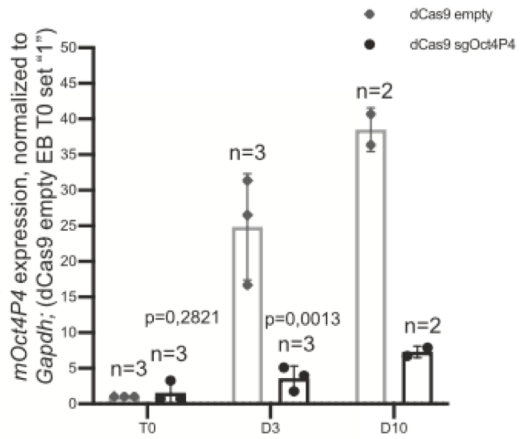


Stable expression of dCAS9 and sgOct4P4 promoter

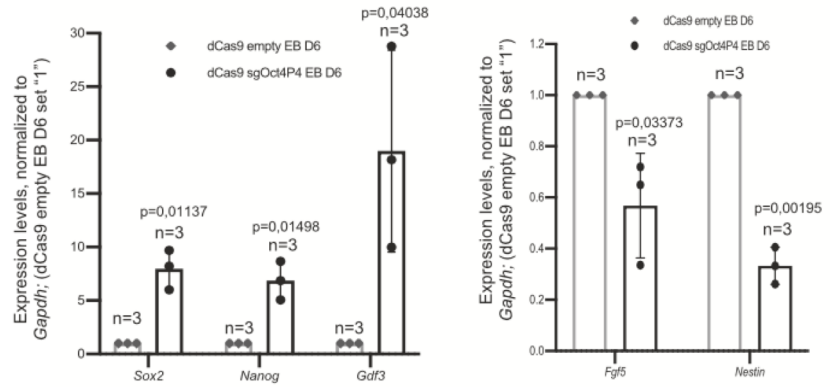


dCas9 fused with transcriptional repressor domain that is guided to promoter by guide RNA

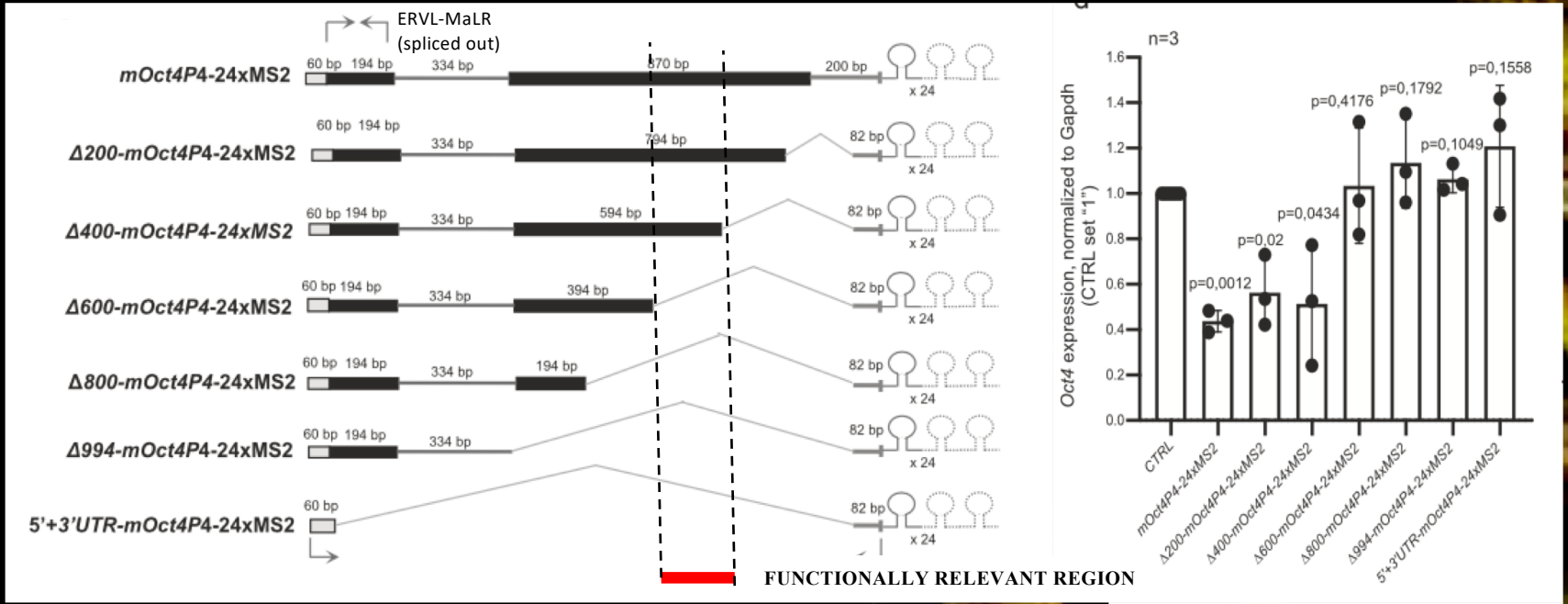
dCAS9 - sgOct4P4 suppresses Oct4P4 expression in embryoid body differentiation



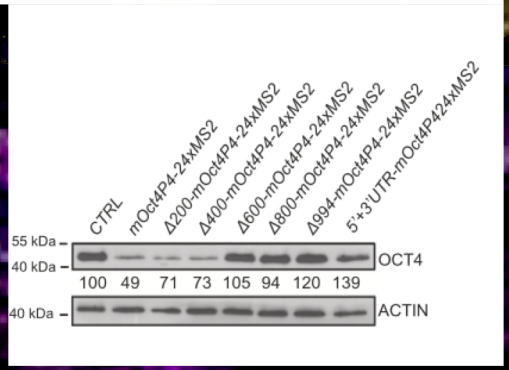
dCAS9 - sgOct4P4 releases aberrant self renewal transcription factor expression in embryoid bodies



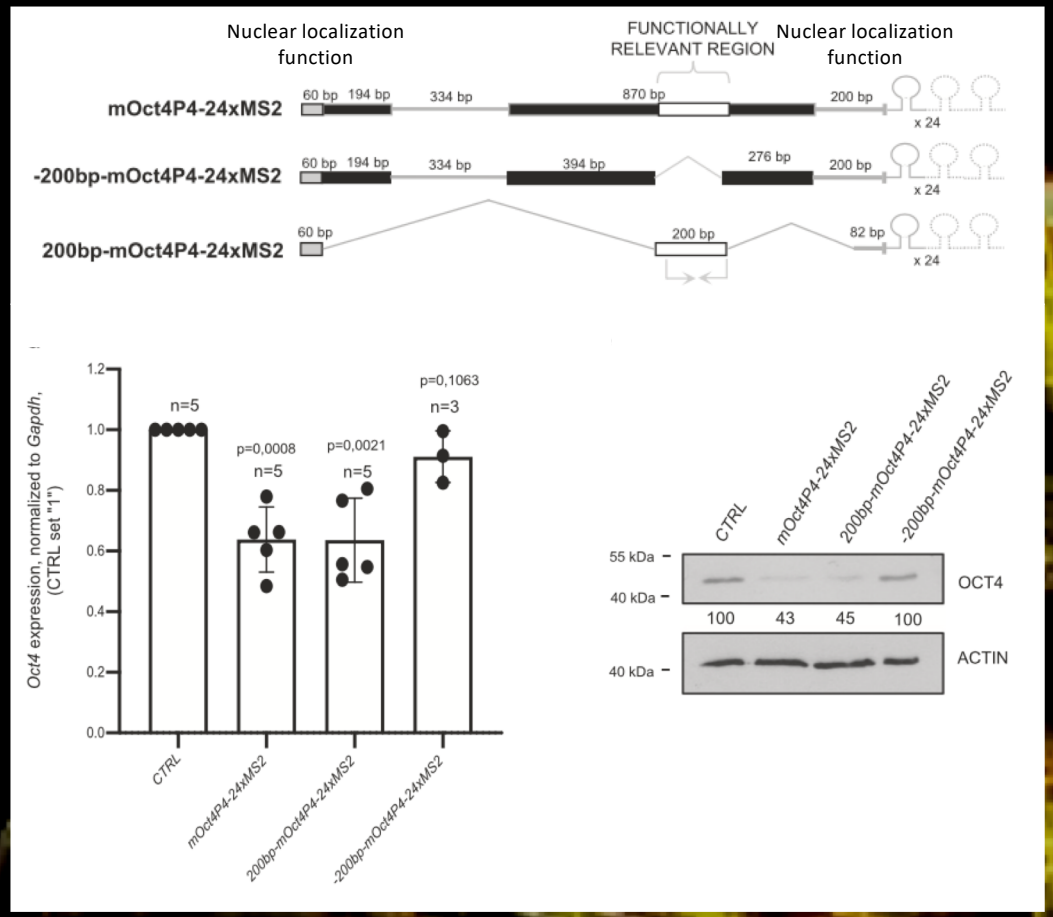
A detailed deletion analysis to identify functional relevant RNA sequences



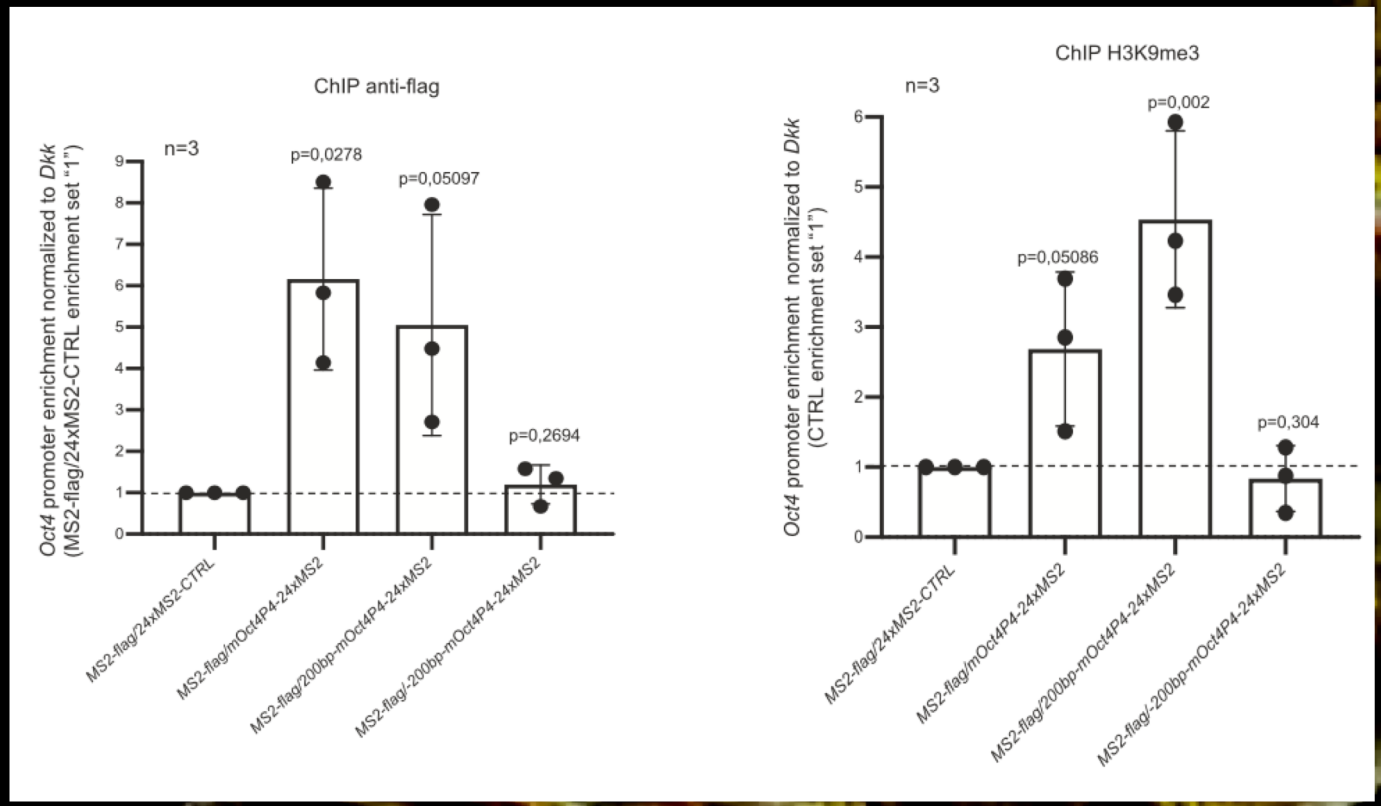
A 200 nucleotide element in mOct4P4 is functionally relevant for mOct4P4 function



A 200nt RNA sequence recapitulates mOct4P4 function

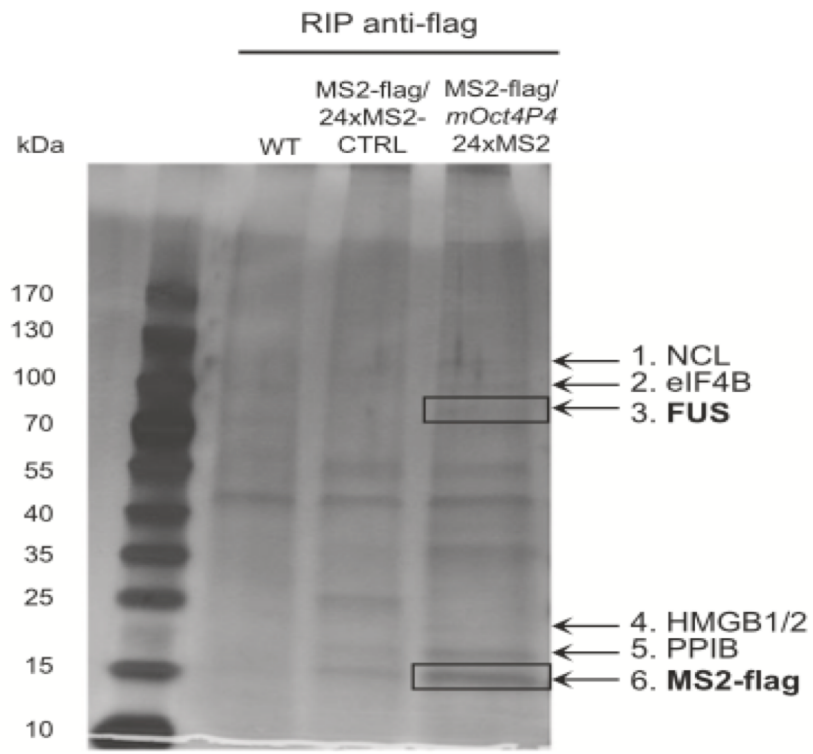


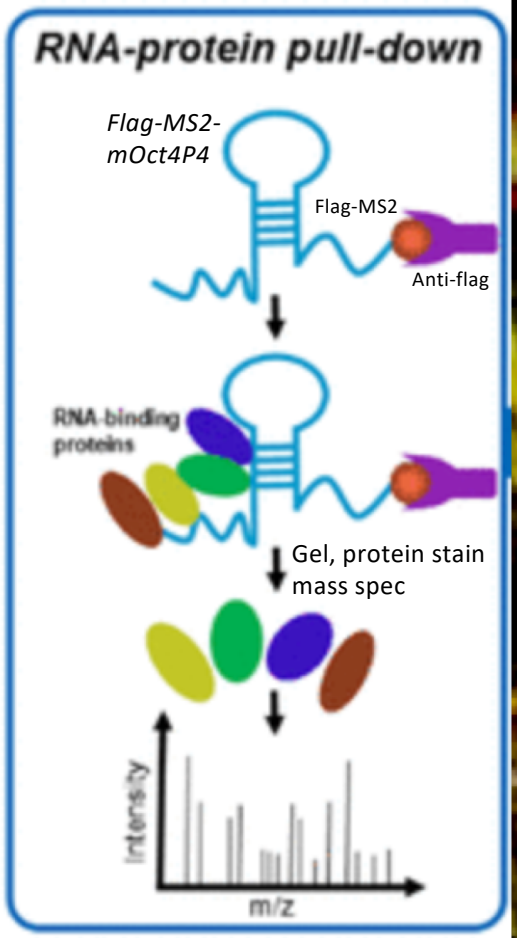
A 200nt RNA sequence recapitulates mOct4P4 function



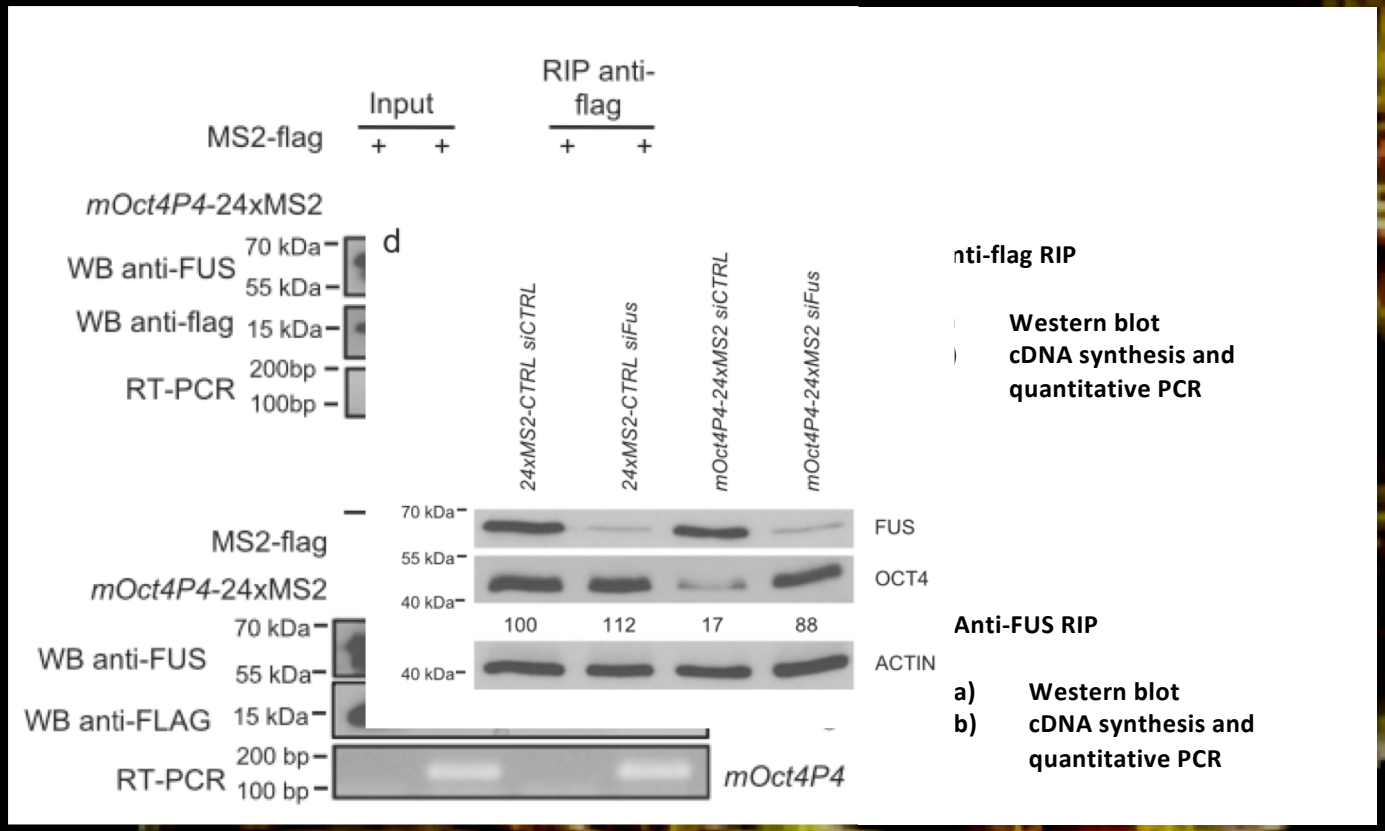
Are there proteins that bind Oct4P4 lncRNA to mediate silncing function?

RNA Immuoprecipitation (flag-MS2 tagged RNA)





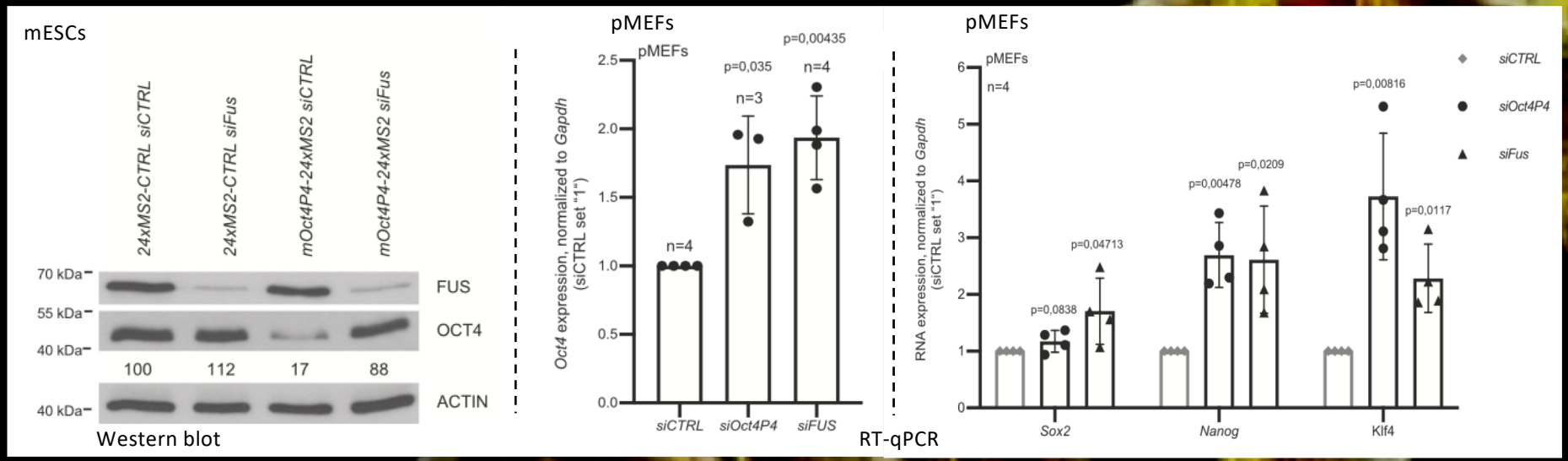
FUS binds the mOct4P4 lncRNA



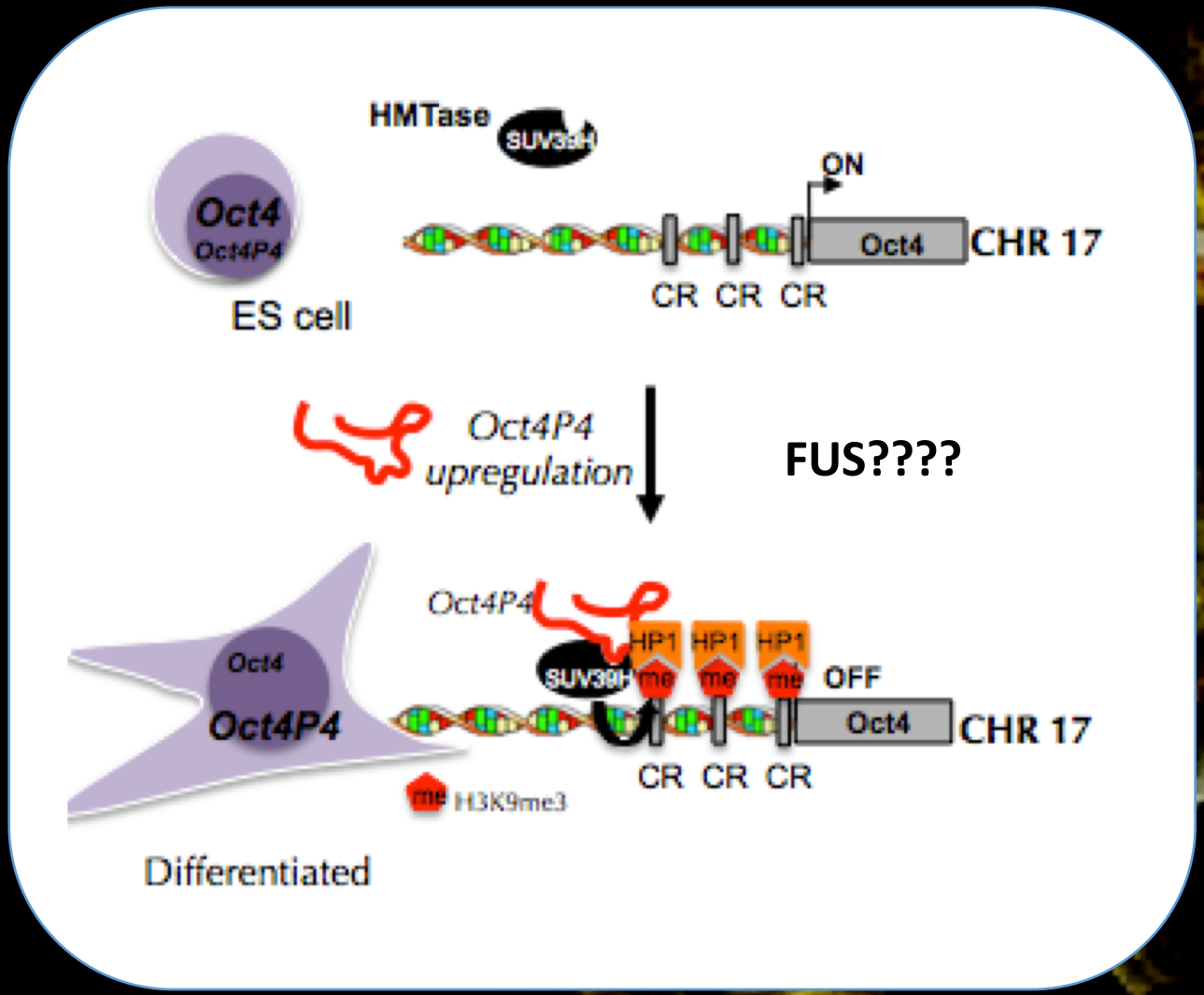
FUS is required for mOct4P4 lncRNA mediated silencing of Oct4

Question 1: Does FUS loss of function break mOct4P4 function in mESCs?

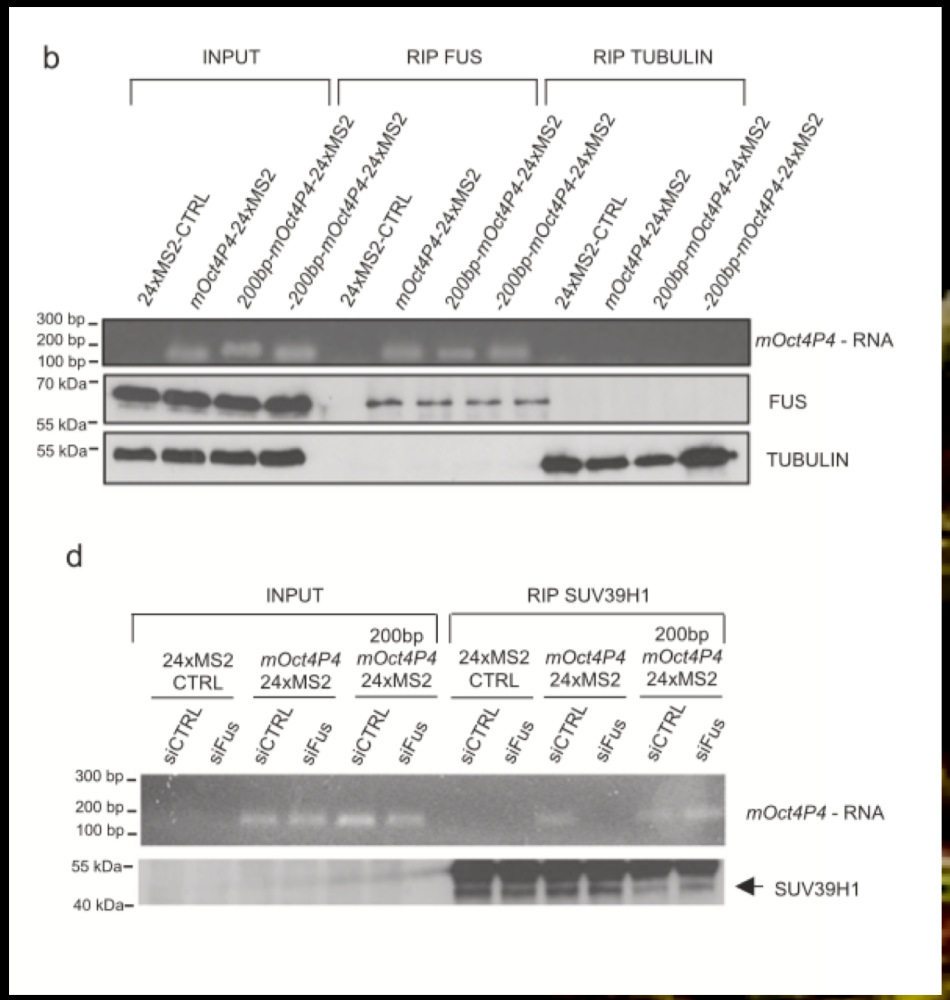
Question 2: Does loss of function of FUS break silencing of Oct4 in differentiated cells?



HOW DOES FUS RELATE TO SUV39H1???



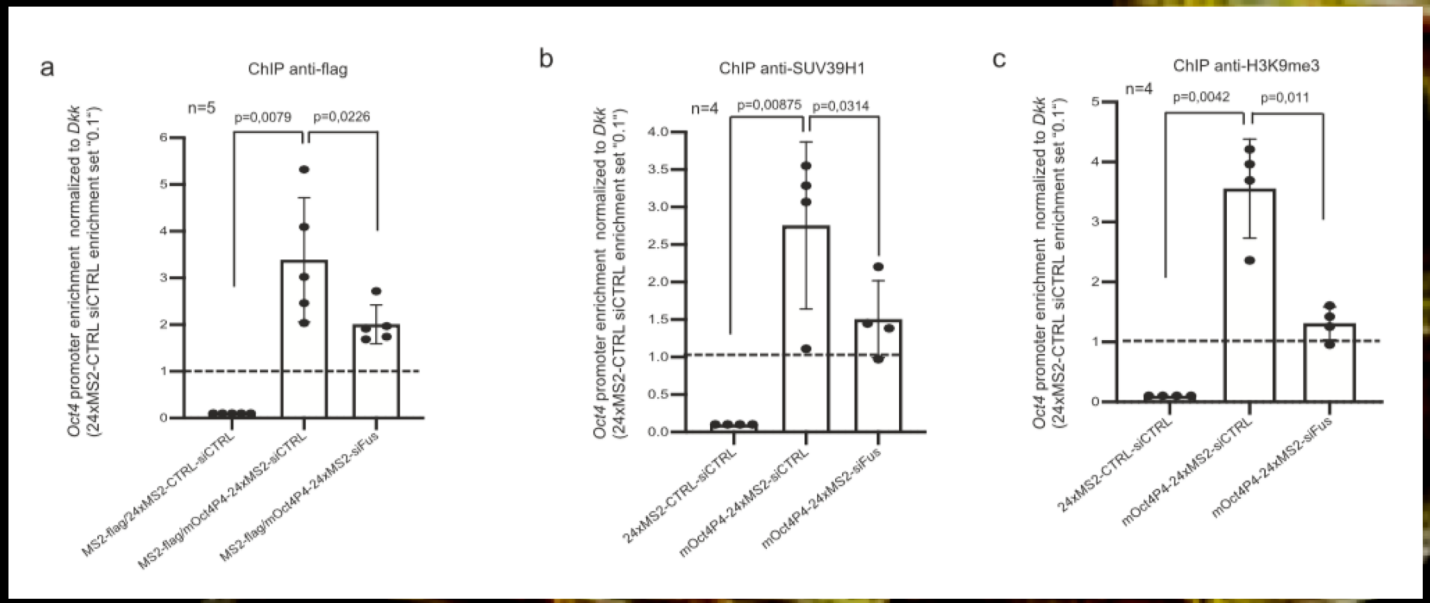
FUS binds to Oct4P4 to allow Suv39h1 recruitment



FUS binds mOCT4P4 with or without 200bp region

Suv39h1 binds mOCT4P4 only when FUS binds mOct4P4 lncRNA

FUS binds to Oct4P4 to allow Suv39h1 recruitment and targeting of the Oct4 locus



FUS binds to Oct4P4 to allow Suv39h1 recruitment and targeting of the Oct4 locus