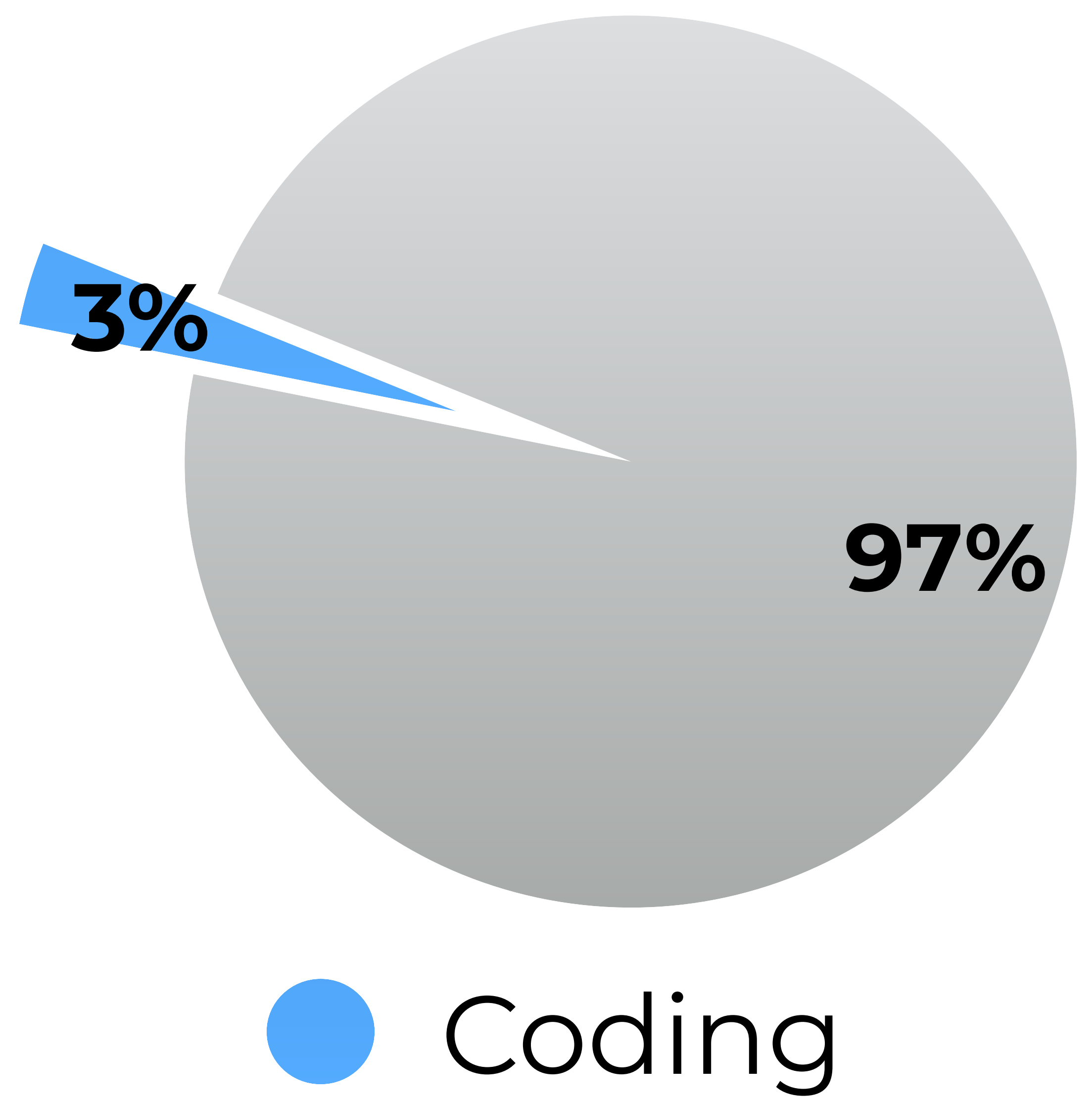
# High-throughput characterization of lncRNA regulation, evolution, and function

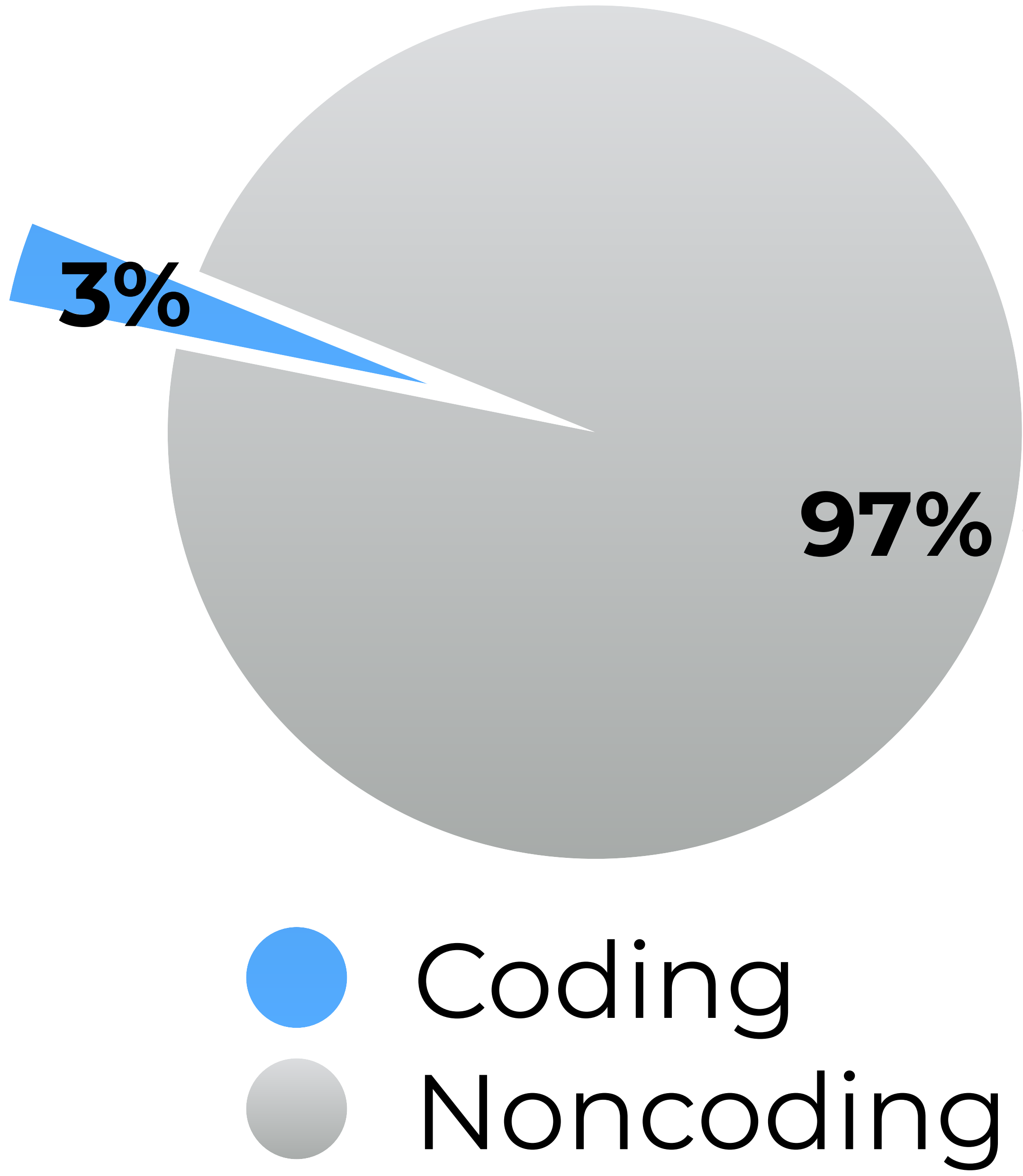
Kaia Mattioli, Ph.D.
Rinn (University of Colorado), Slack (Harvard Medical School), and Melé (BSC) Labs

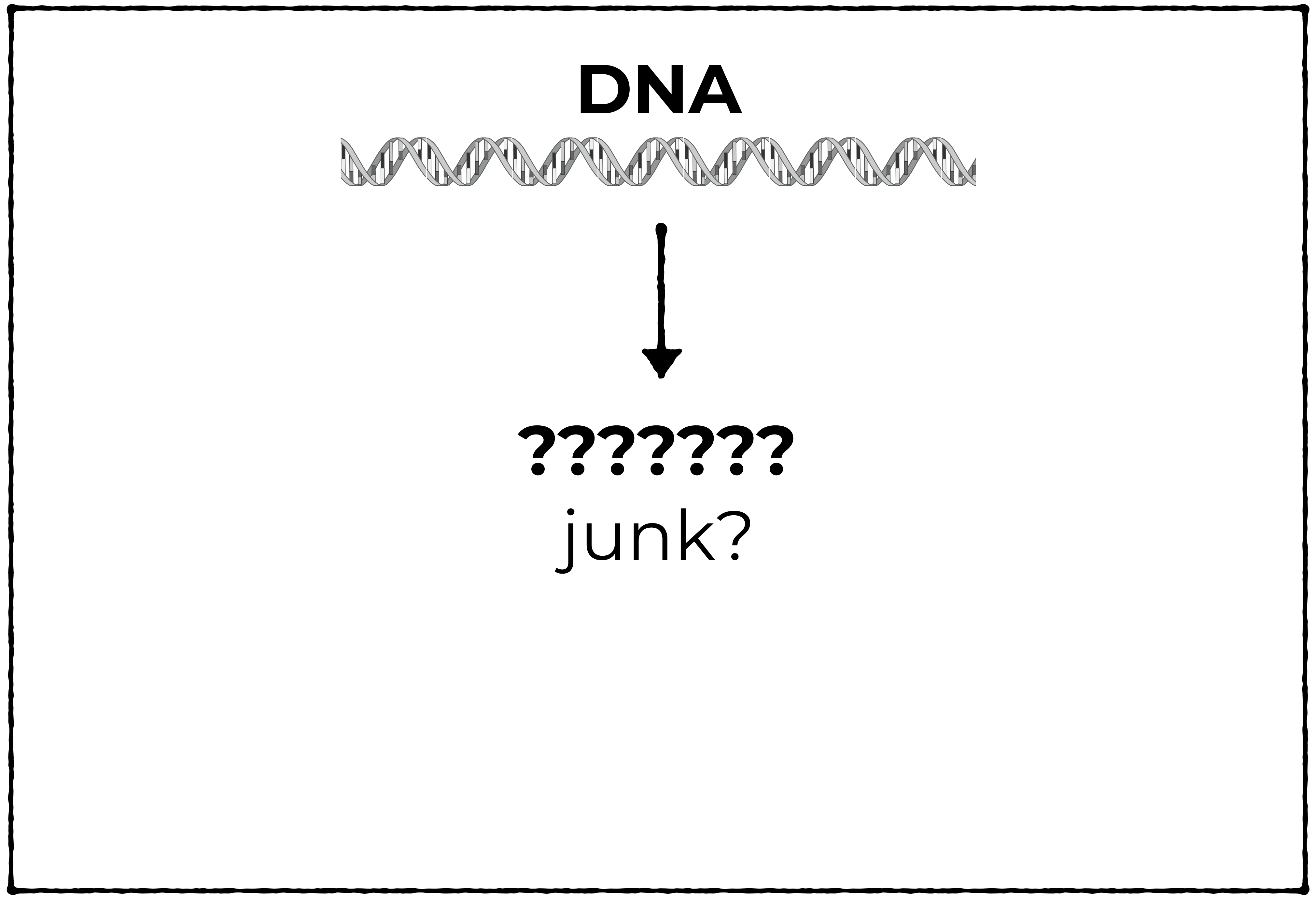
BSC Seminar
November 12th, 2019

# Only ~3% of the human genome is coding



# Only ~3% of the human genome is coding



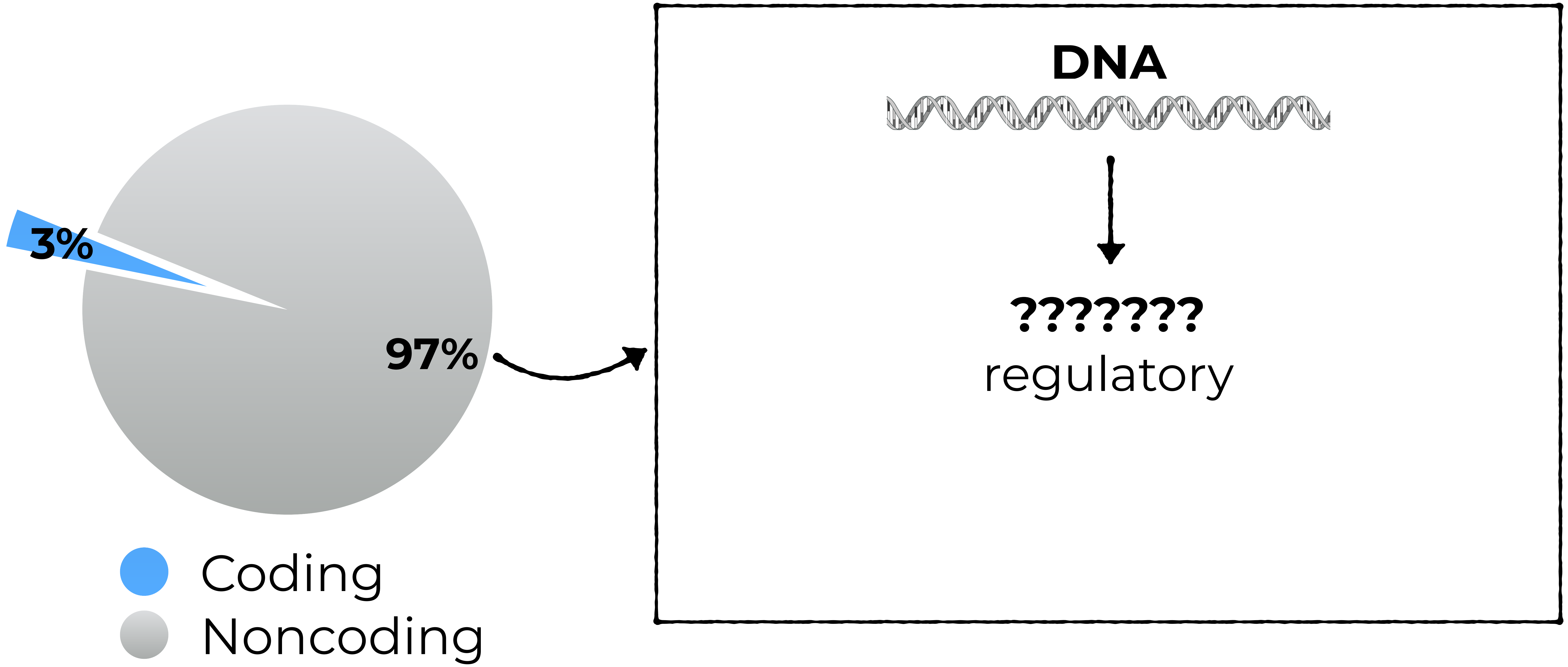


**DNA**

**???????**

junk?

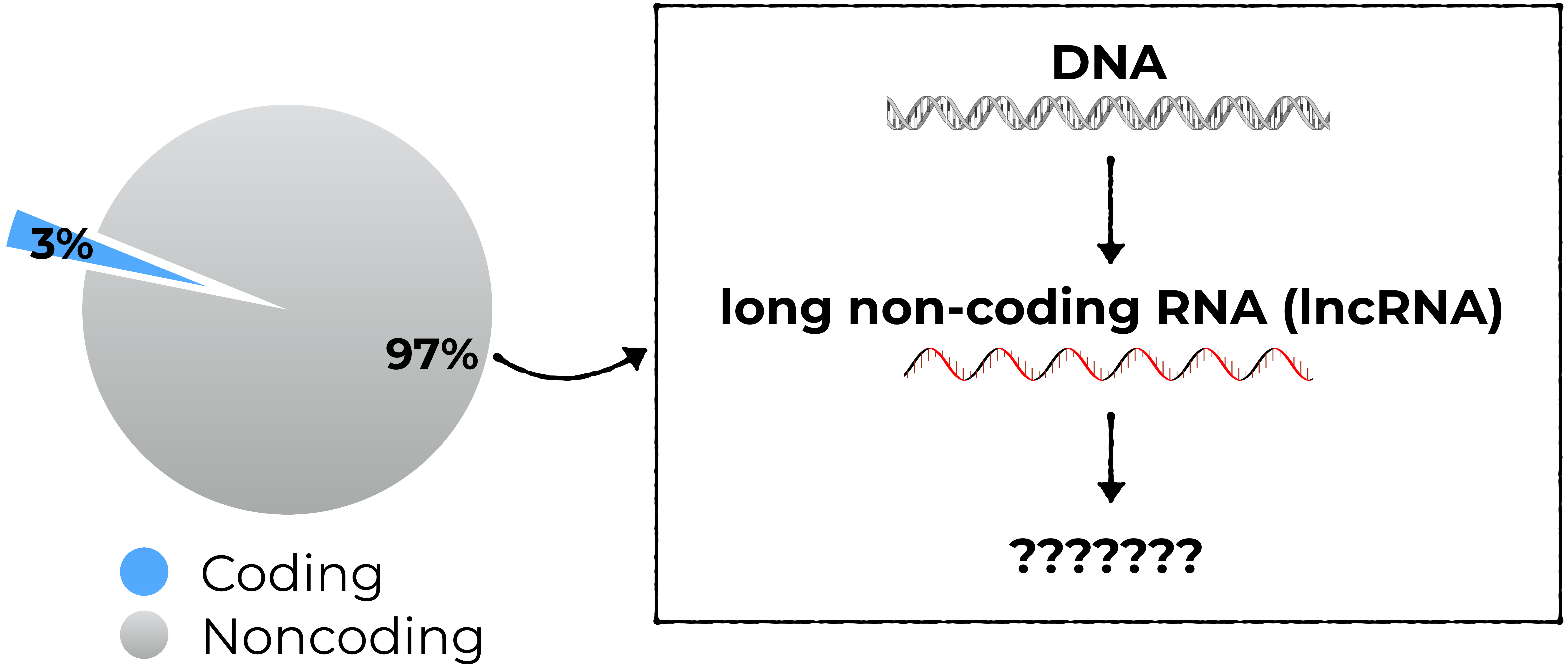
# Only ~3% of the human genome is coding



**DNA**

**???????**
regulatory

# Non-coding DNA is transcribed into RNA



**DNA**

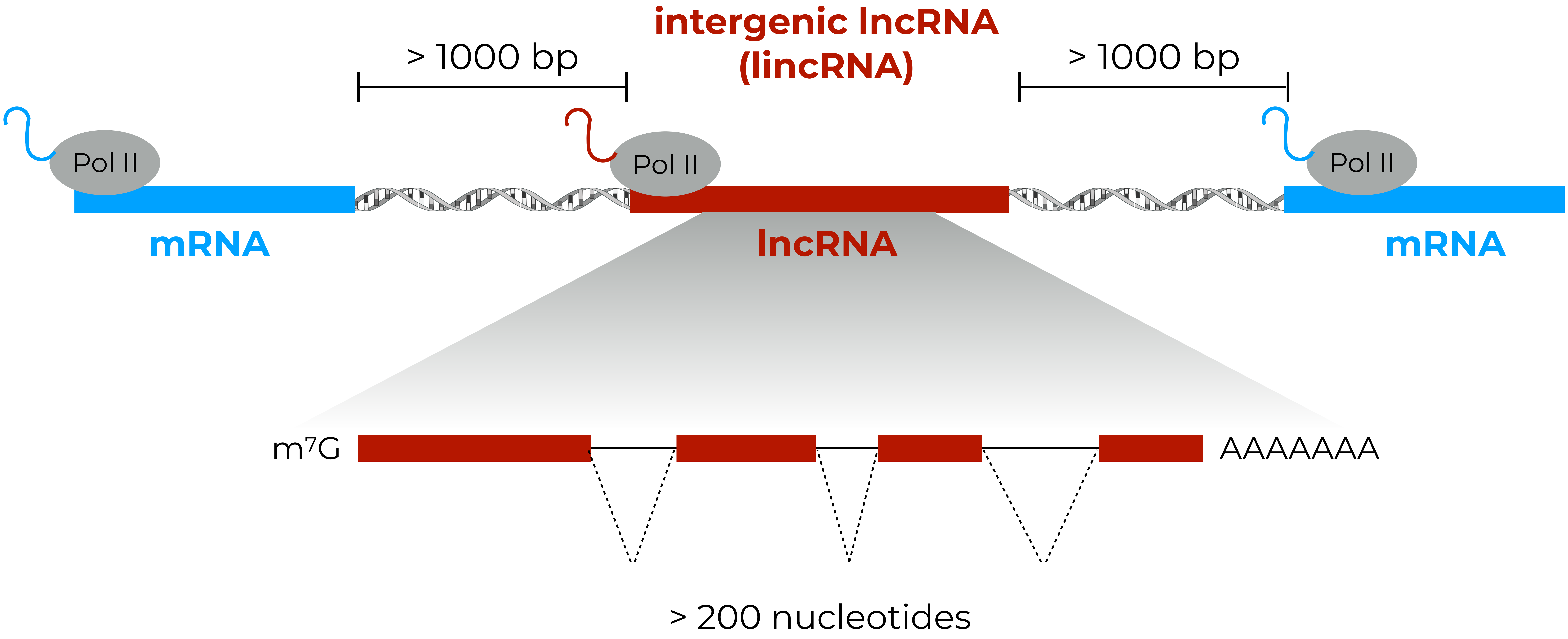
↓

**long non-coding RNA (lncRNA)**

↓

**???????**

# lncRNAs resemble mRNAs



**intergenic lncRNA (lincRNA)**

> 1000 bp

> 1000 bp

Pol II

Pol II

Pol II

**mRNA**

**lncRNA**

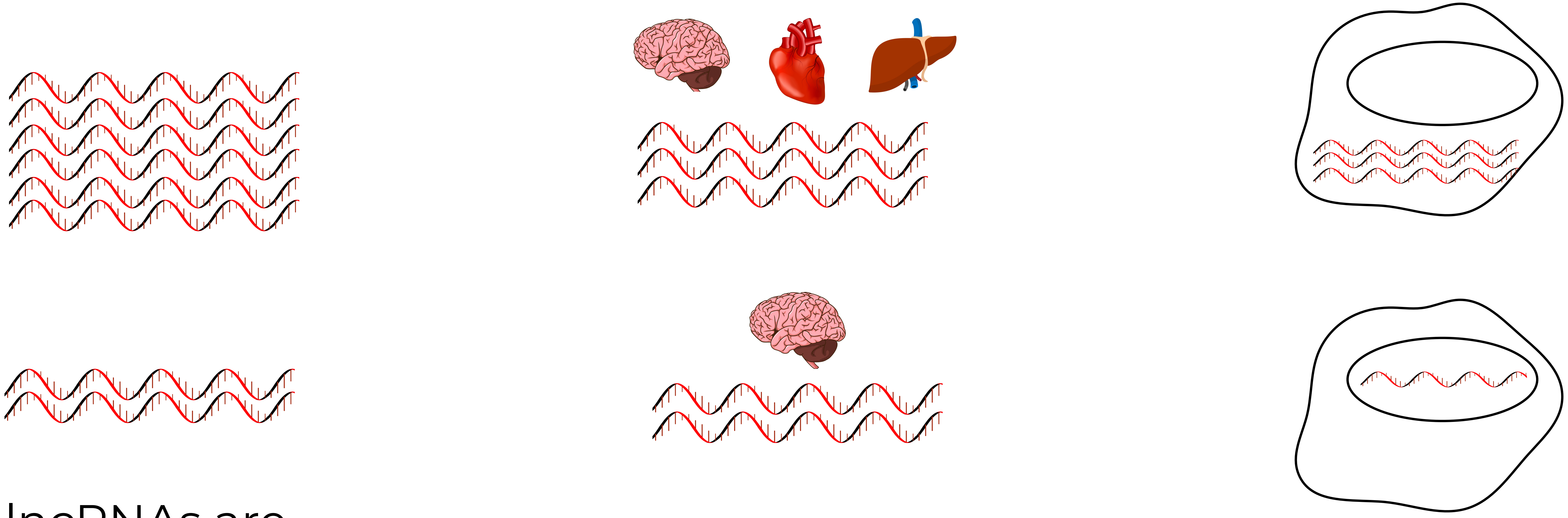
**mRNA**

$m^7G$

AAAAAAA

> 200 nucleotides

# lncRNAs differ from mRNAs in certain properties



**expression**

**specificity**

**localization**

**mRNAs**

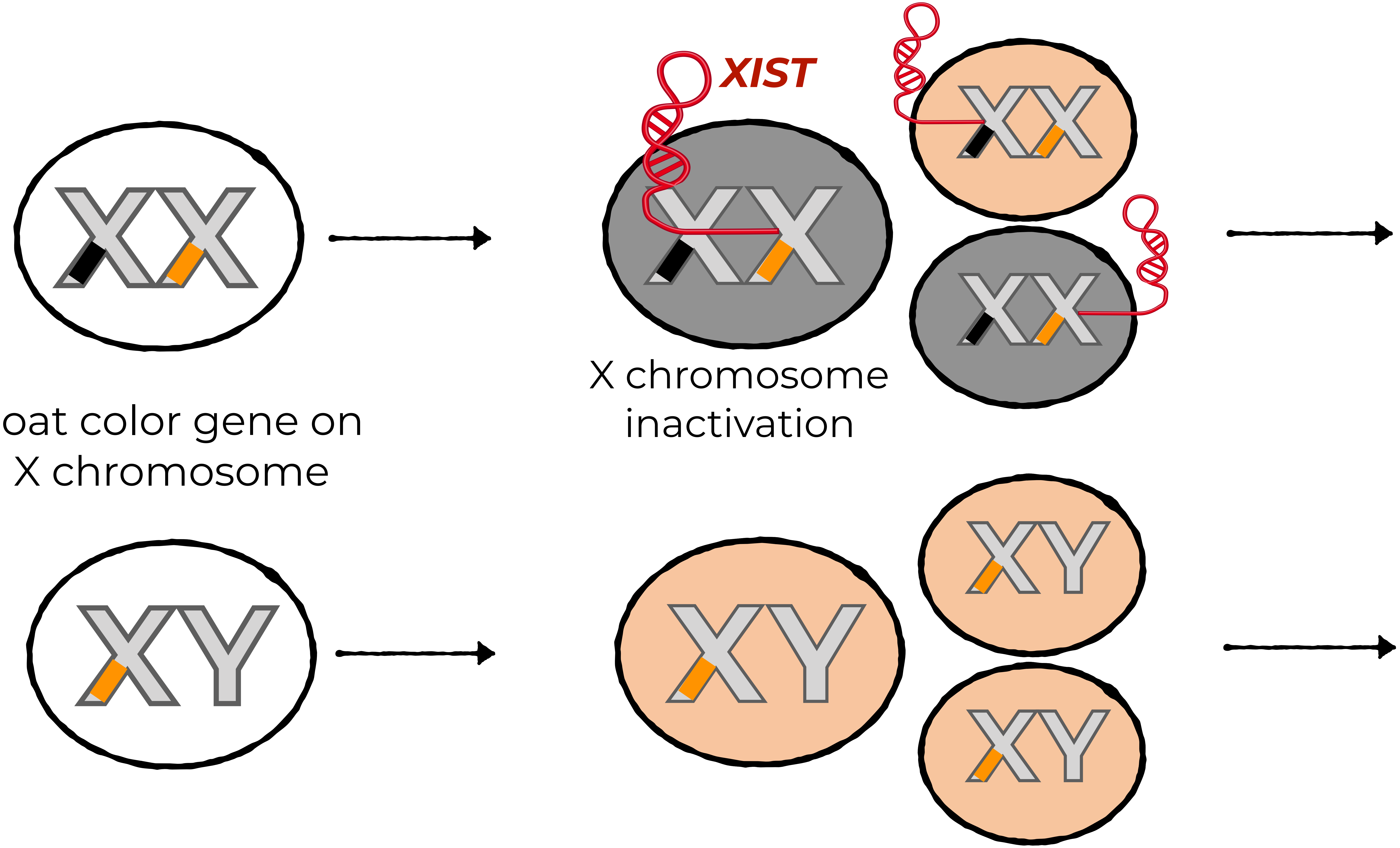
**lncRNAs**

lncRNAs are
lowly expressed

tissue-specific

& often retained
in nucleus

# A lncRNA is responsible for tortoiseshell cats



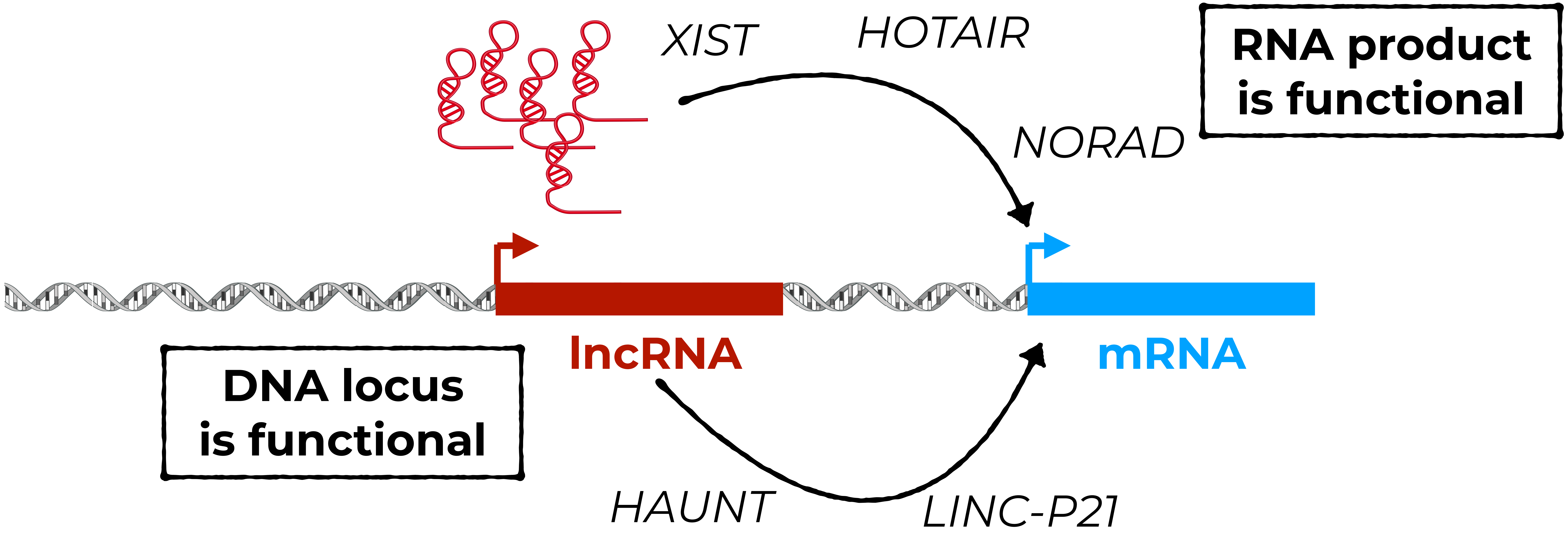
female cats

male cats

coat color gene on X chromosome

X chromosome inactivation

# lncRNAs can regulate gene expression via RNA- and DNA-based mechanisms



**RNA product is functional**

*XIST*

*HOTAIR*

*NORAD*

**lncRNA**

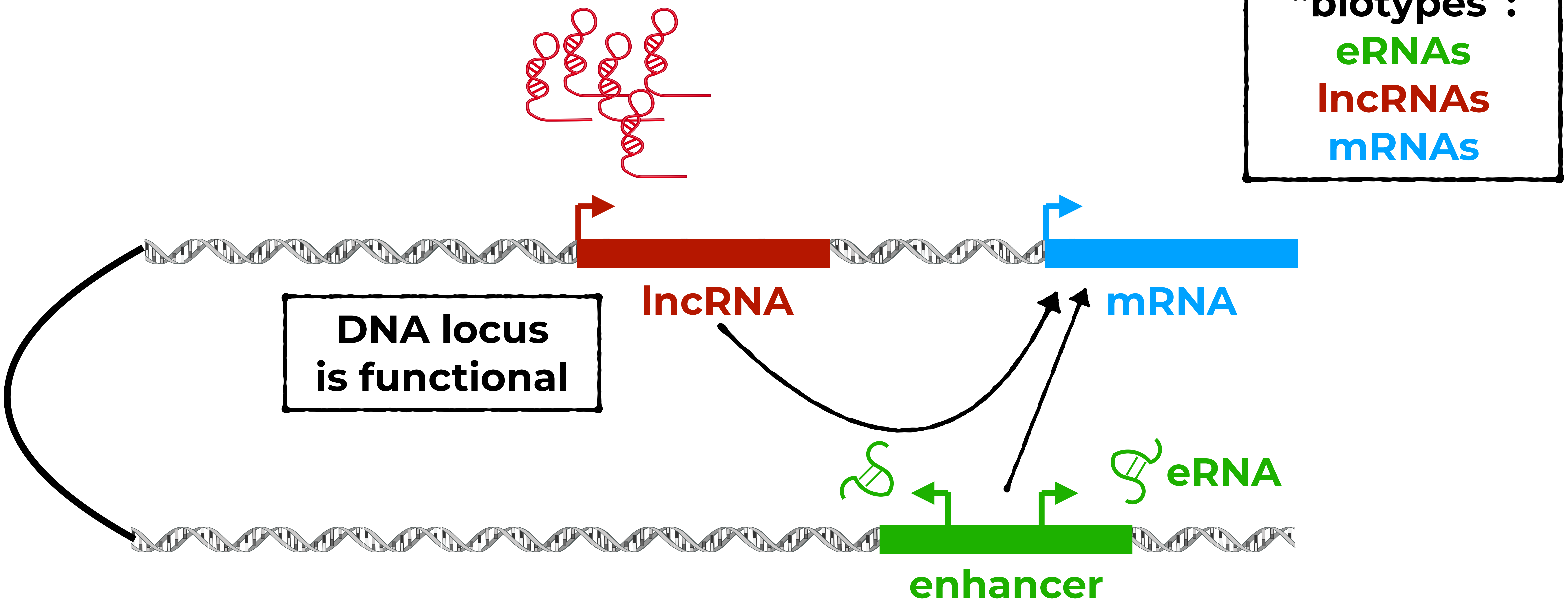
**mRNA**

**DNA locus is functional**

*HAUNT*

*LINC-P21*

# Enhancers also regulate gene expression via DNA-based mechanisms

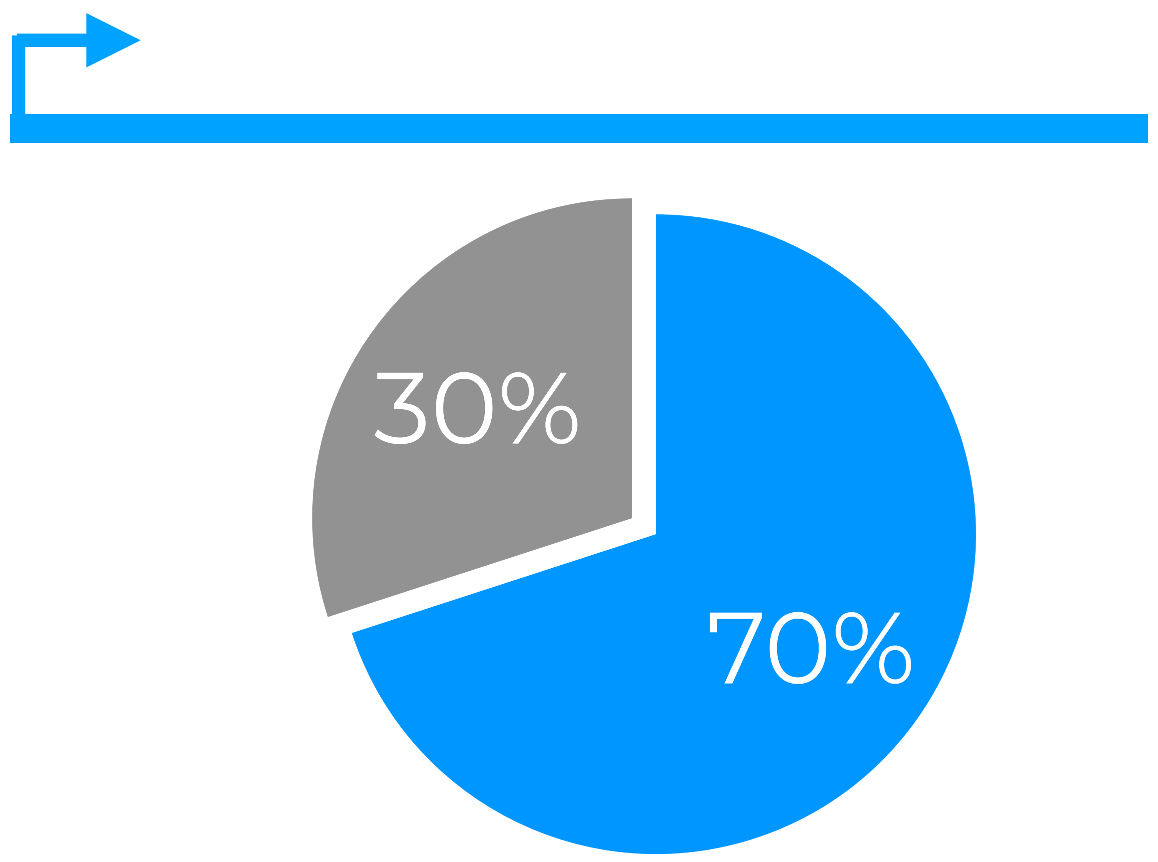


# Long non-coding RNAs are mostly uncharacterized

% with known function:

# in genome:

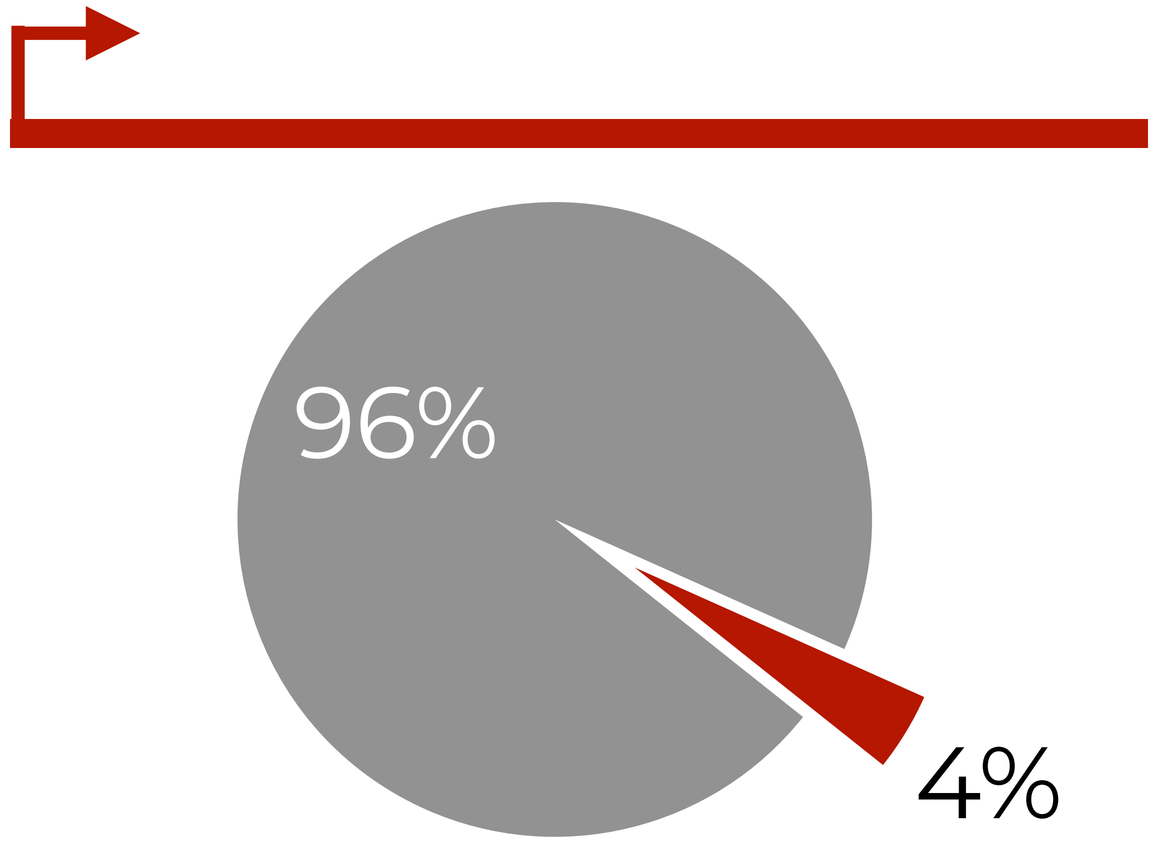
**mRNA**



* Ellens *et al.* NAR 2017

~20,000

**lncRNA**

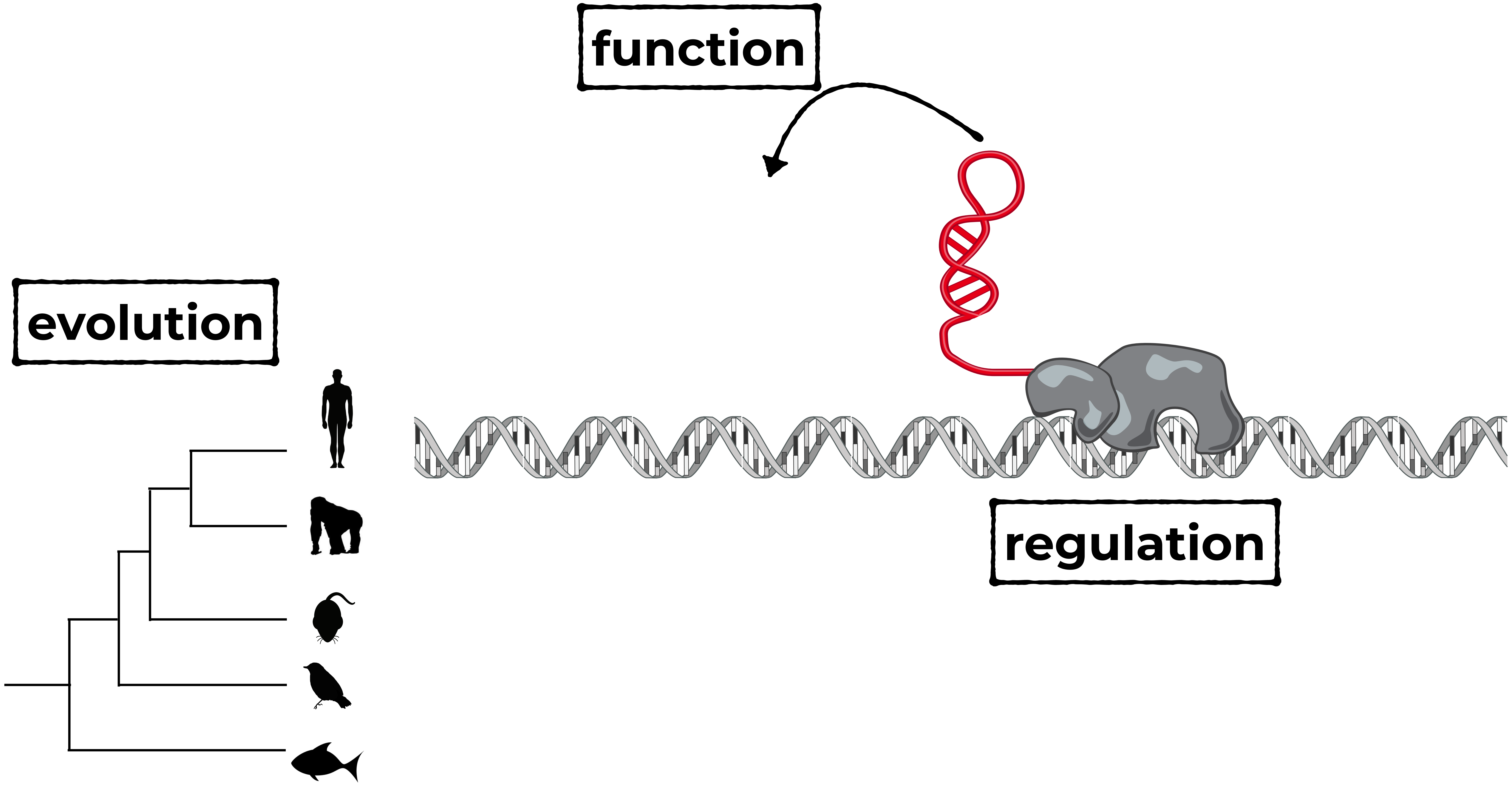


* Quek *et al.* NAR 2015

~15,000 to ~20,000

**tens of thousands of uncharacterized lncRNAs**

# Studying lncRNA biology *en masse*



**function**

**evolution**

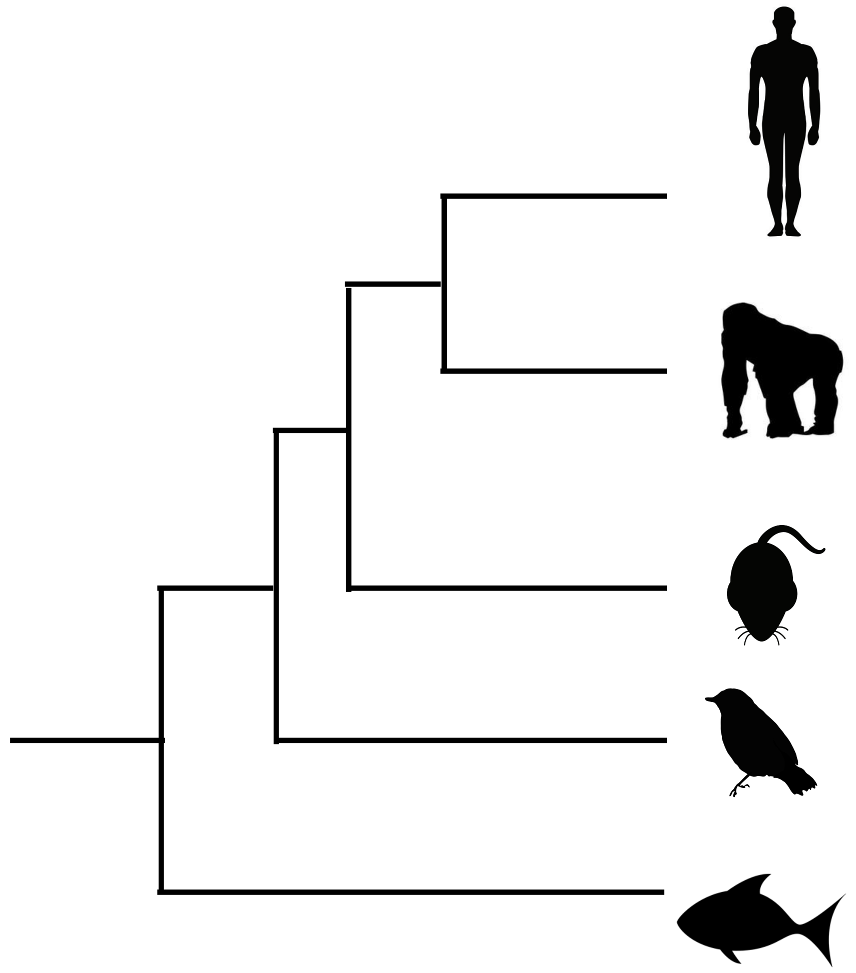
**regulation**

# Studying lncRNA biology *en masse*

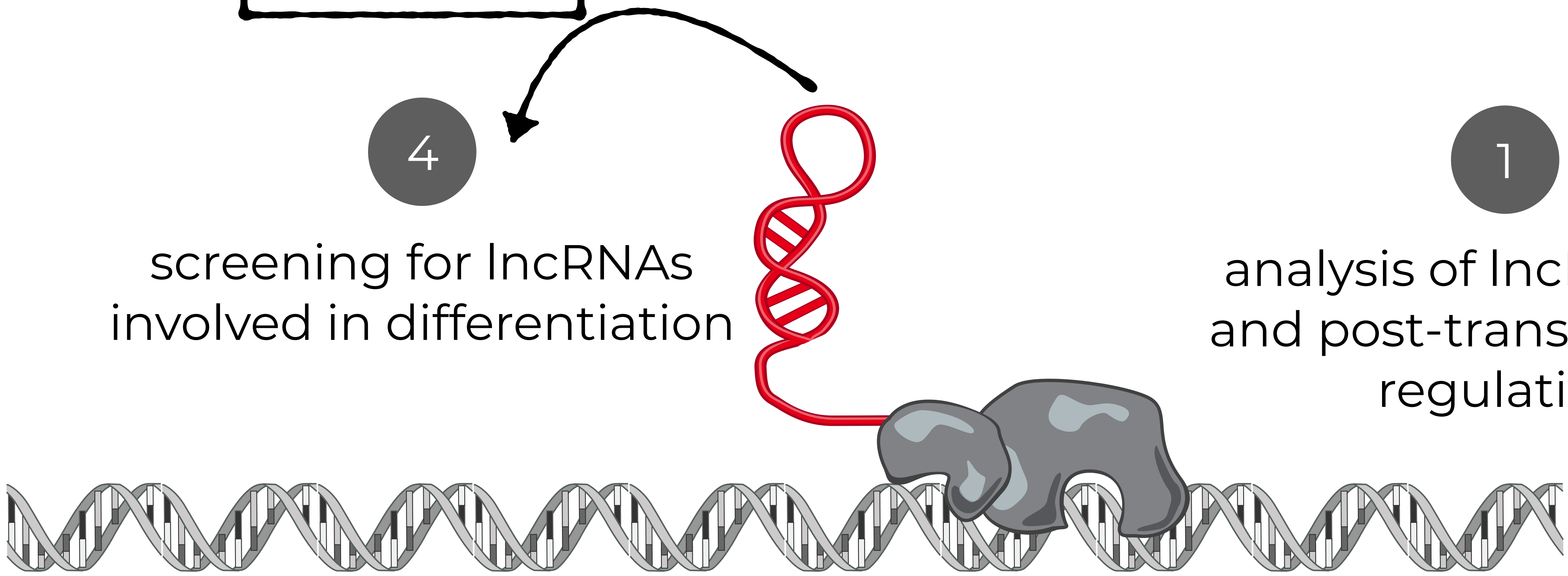
3

massively parallel
dissection
of lncRNA evolution

**evolution**



**function**



4

screening for lncRNAs
involved in differentiation

1

analysis of lncRNA pre-
and post-transcriptional
regulation

**regulation**

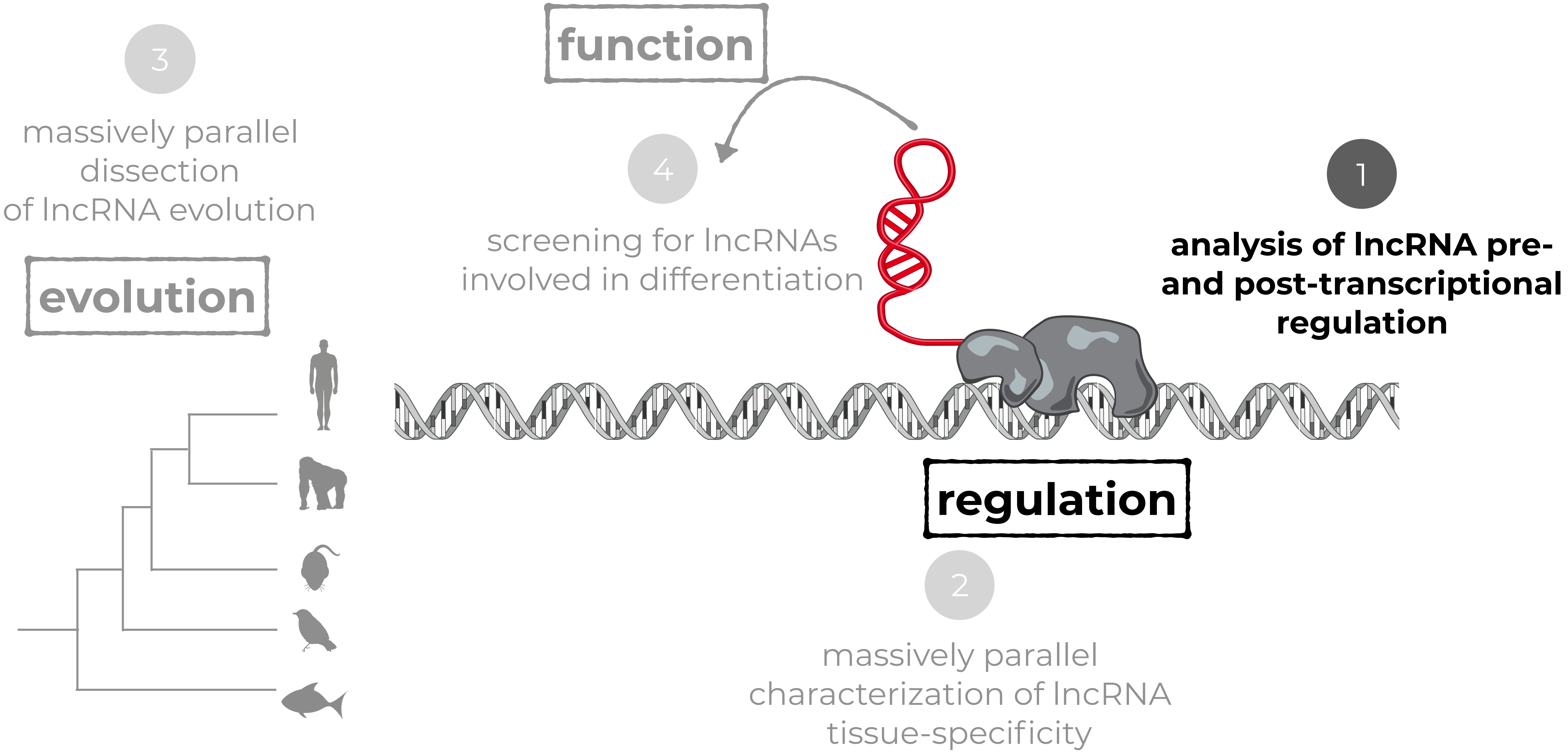
2

massively parallel
characterization of lncRNA
tissue-specificity

Marta Melé

Jeff Haswell

# Studying lncRNA biology *en masse*



3

massively parallel
dissection
of lncRNA evolution

**evolution**

**function**

4

screening for lncRNAs
involved in differentiation

1

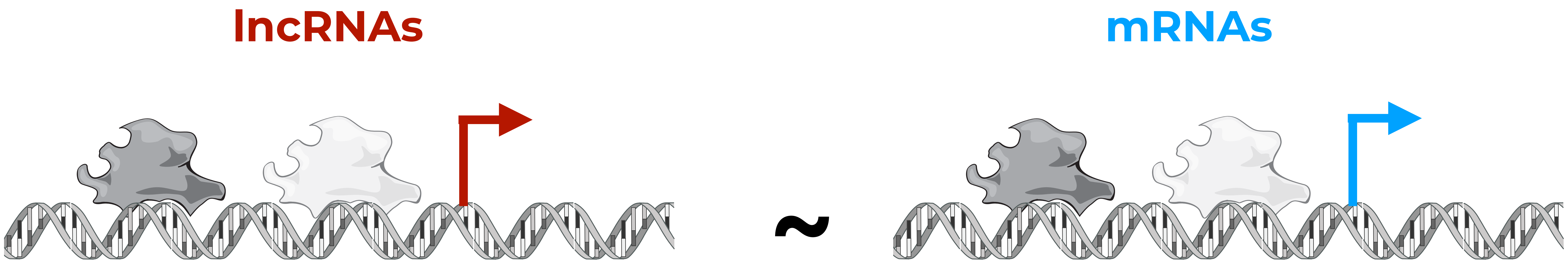
**analysis of lncRNA pre-
and post-transcriptional
regulation**

**regulation**

2

massively parallel
characterization of lncRNA
tissue-specificity

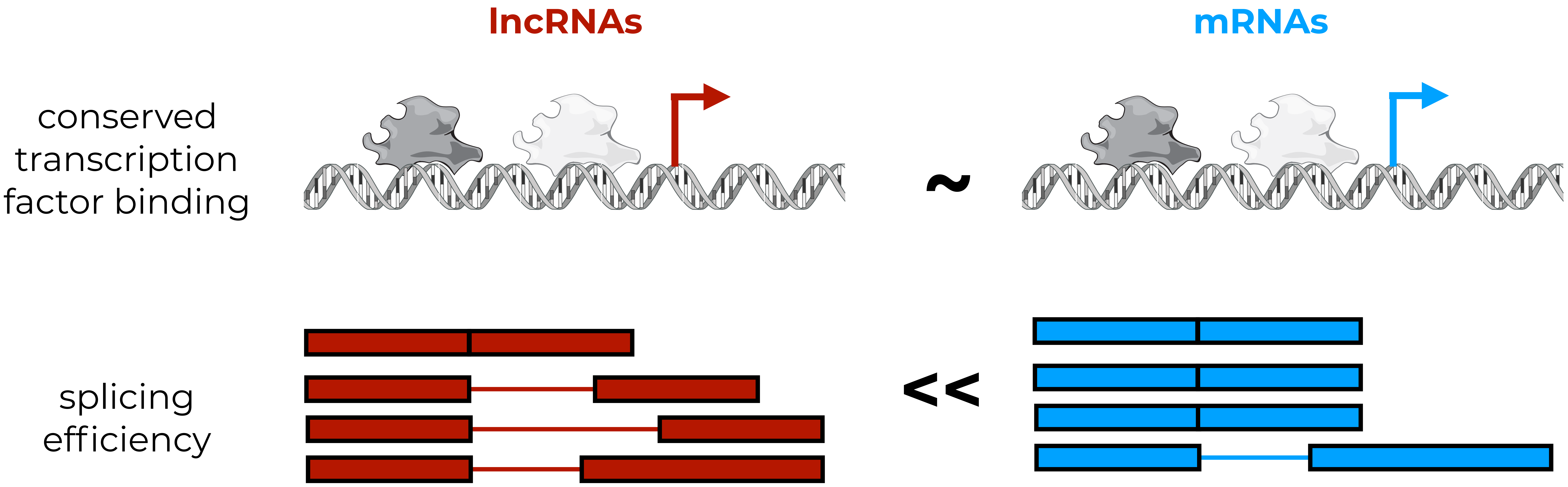
# LncRNAs have conserved transcription factor binding



conserved
transcription
factor binding

Melé, Mattioli *et al.* Genome Research 2017

# LncRNAs are inefficiently spliced

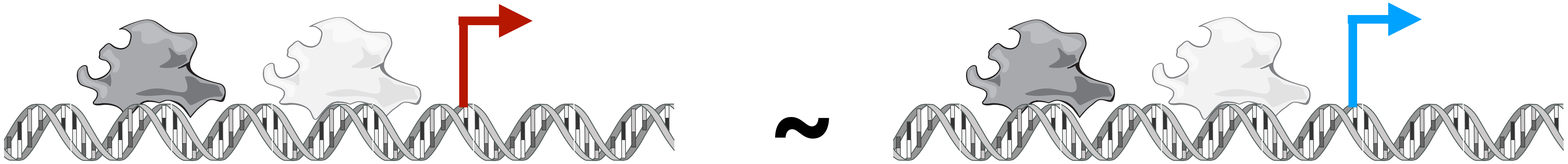


Melé, Mattioli *et al.* Genome Research 2017

# Efficient splicing is associated with functional lncRNAs

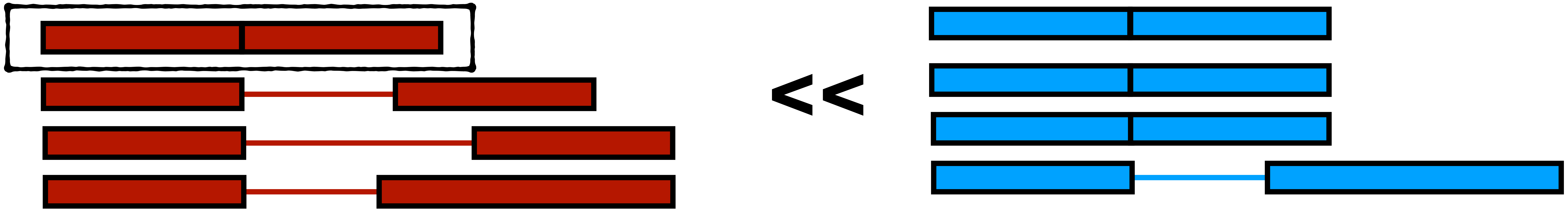
**lncRNAs**

**mRNAs**



conserved transcription factor binding

~



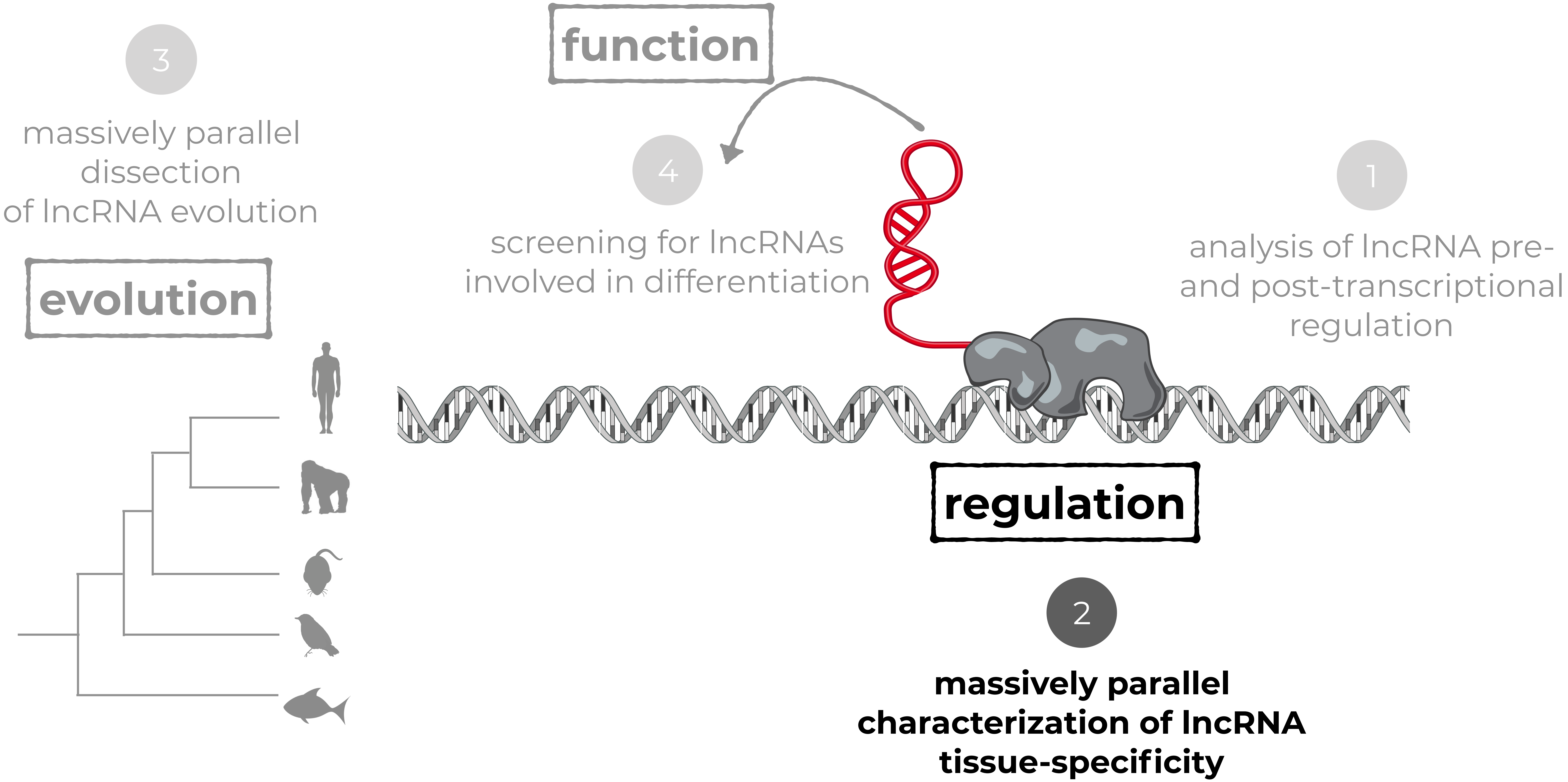
*XIST* & others

splicing efficiency

<<

Melé, Mattioli *et al.* Genome Research 2017

# Studying lncRNA biology *en masse*



3

massively parallel
dissection
of lncRNA evolution

**evolution**

**function**

4

screening for lncRNAs
involved in differentiation

1

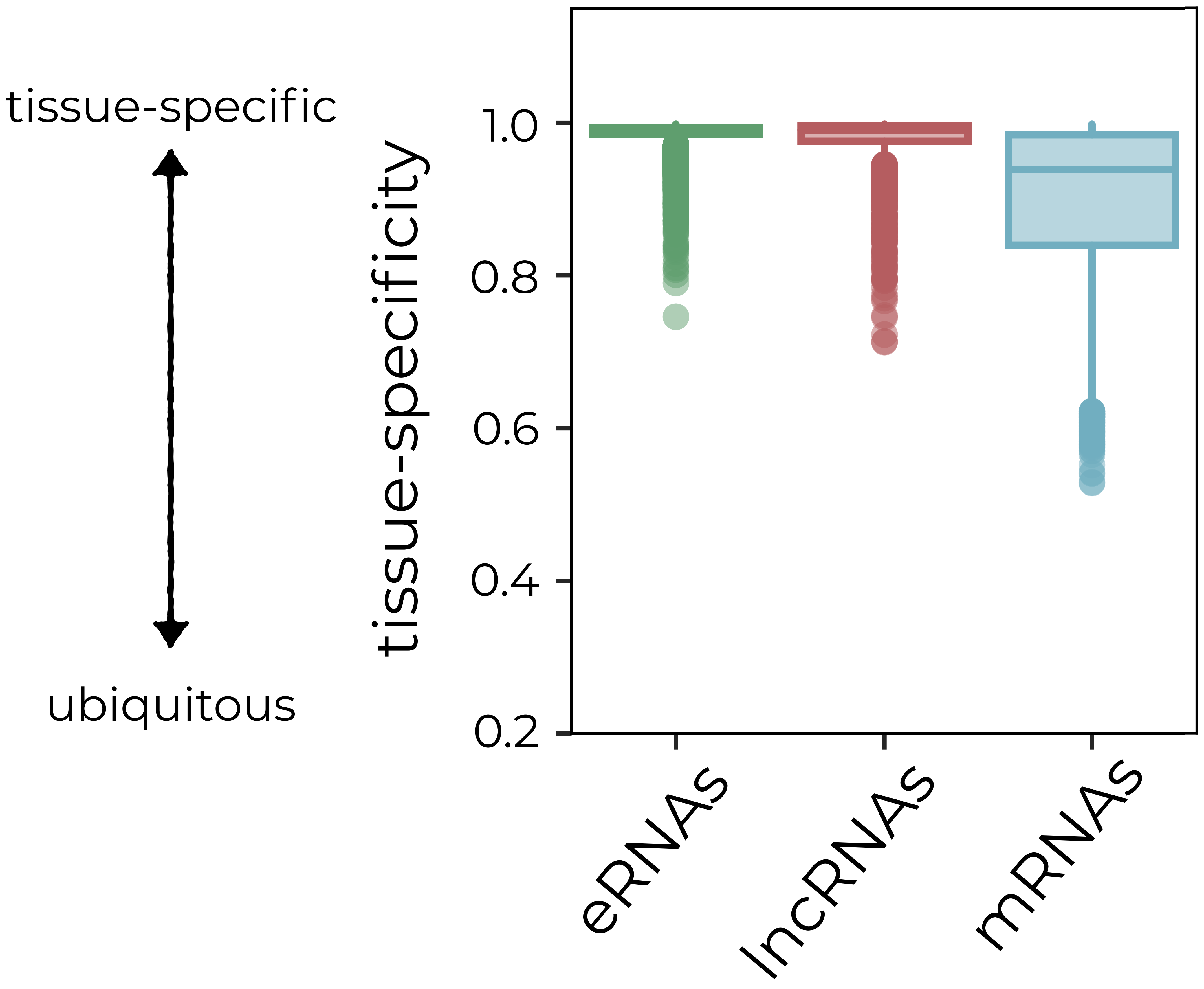
analysis of lncRNA pre-
and post-transcriptional
regulation

**regulation**

2

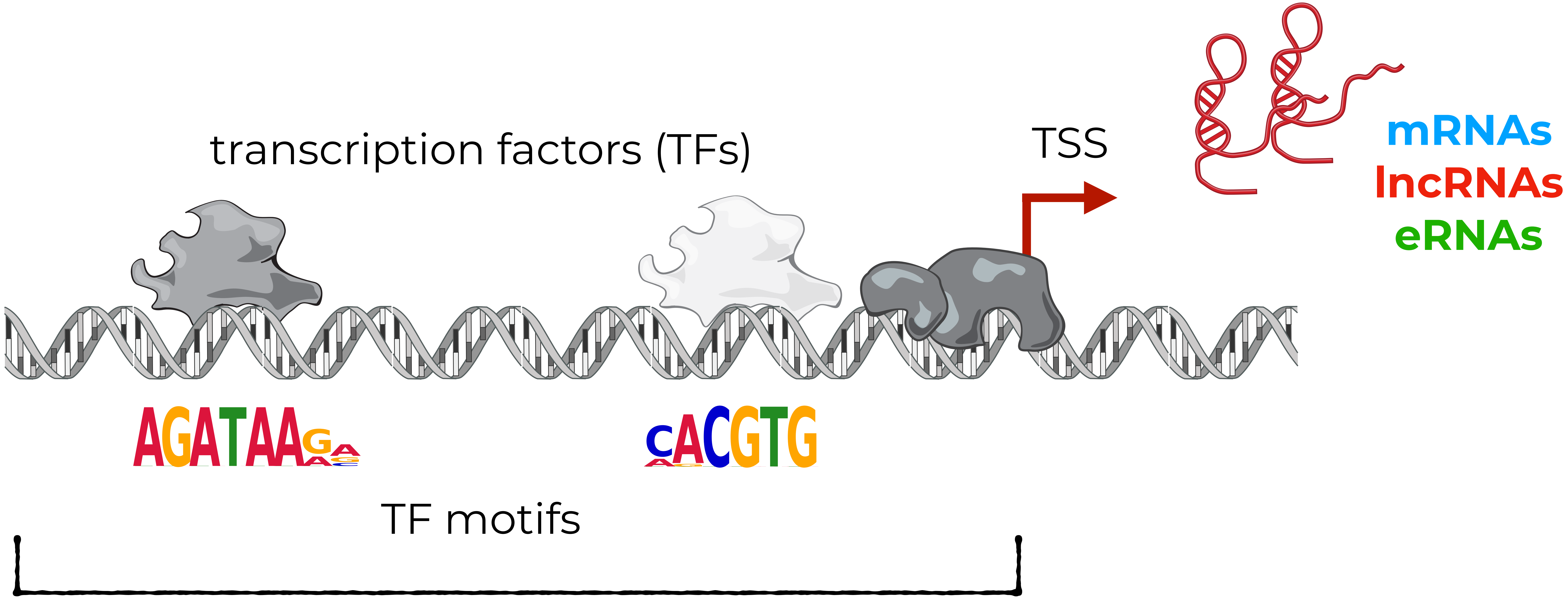
**massively parallel
characterization of lncRNA
tissue-specificity**

# LncRNAs are tissue-specific



Mattioli *et al.* Genome Research 2019

# Regulation of gene expression



transcription factors (TFs)

TSS

mRNAs

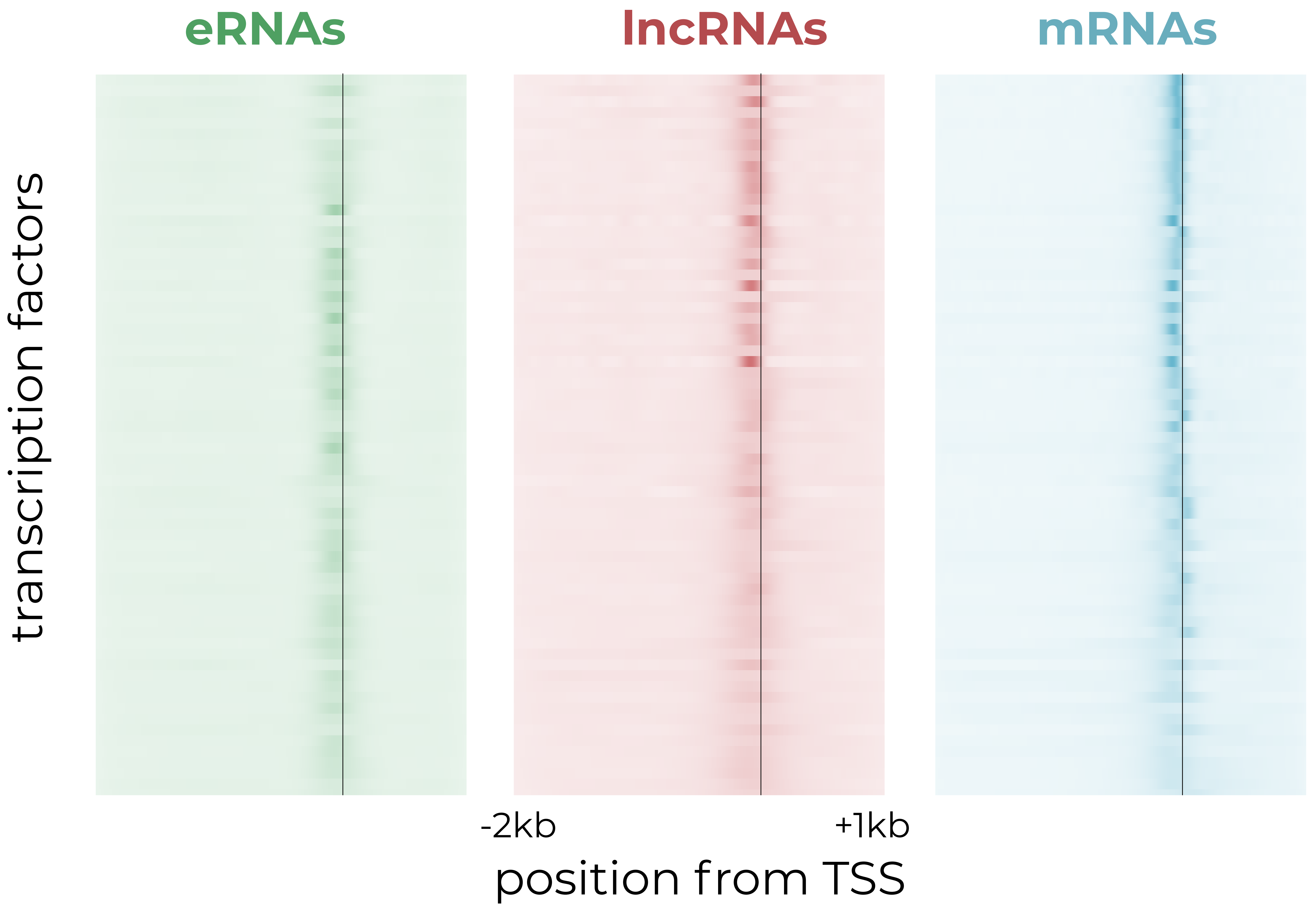
lncRNAs

eRNAs

TF motifs

**core promoter**

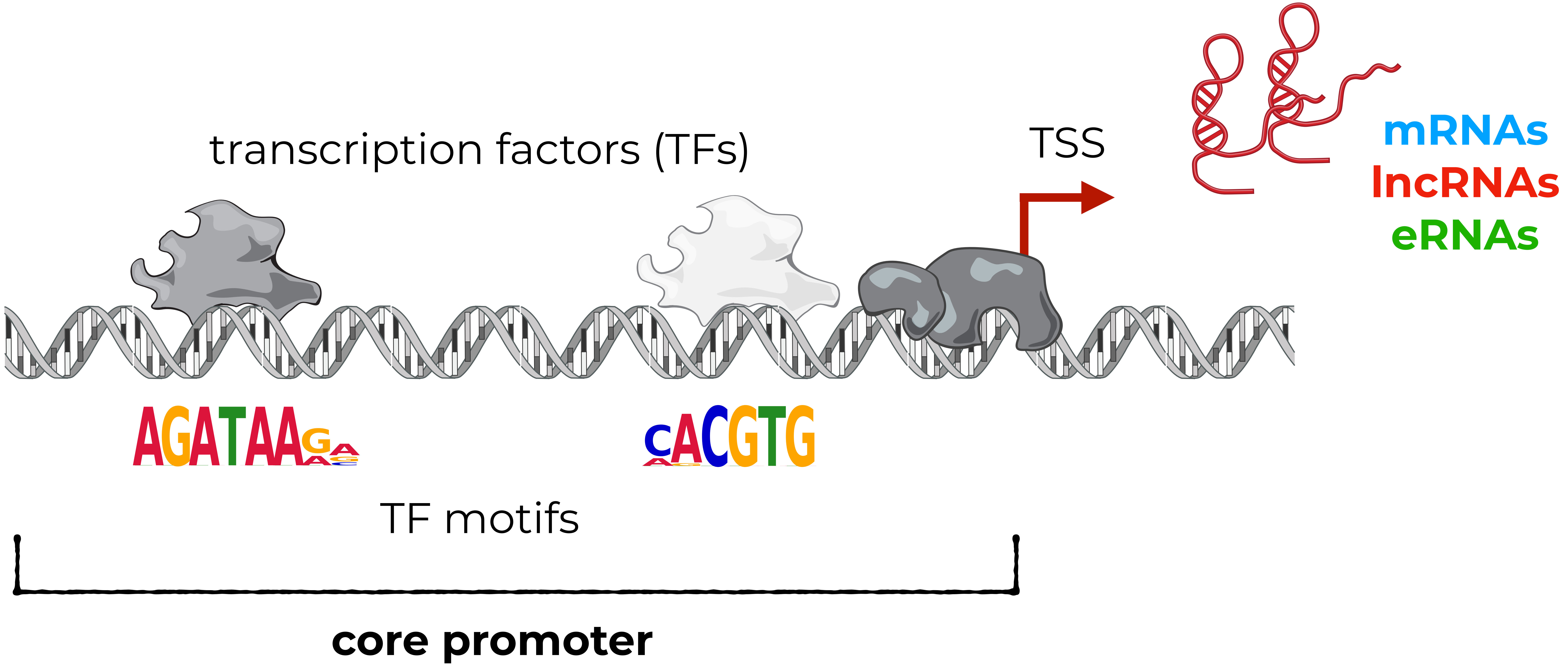
# TF binding is enriched in core promoter region



color =
ChIP-seq
peak intensity

Mattioli *et al.* Genome Research 2019

# Do sequence features in core promoters contribute to expression patterns?



transcription factors (TFs)

TSS

mRNAs

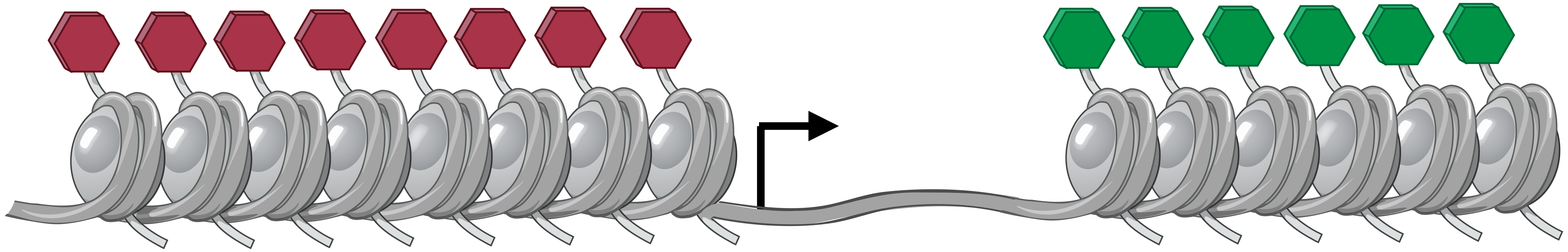
lncRNAs

eRNAs

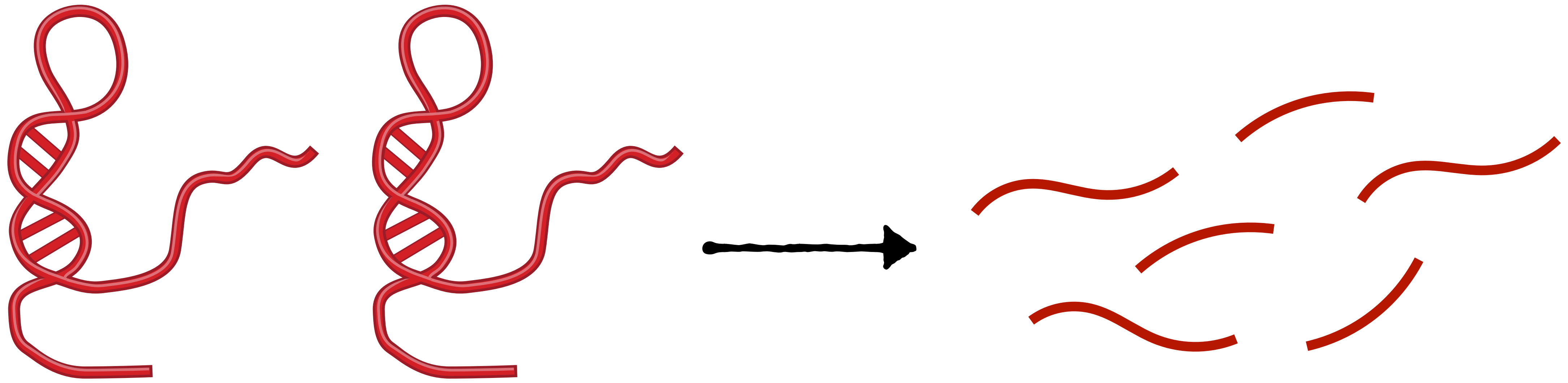
TF motifs

**core promoter**

# Examining the effect of DNA sequence on transcription is hard

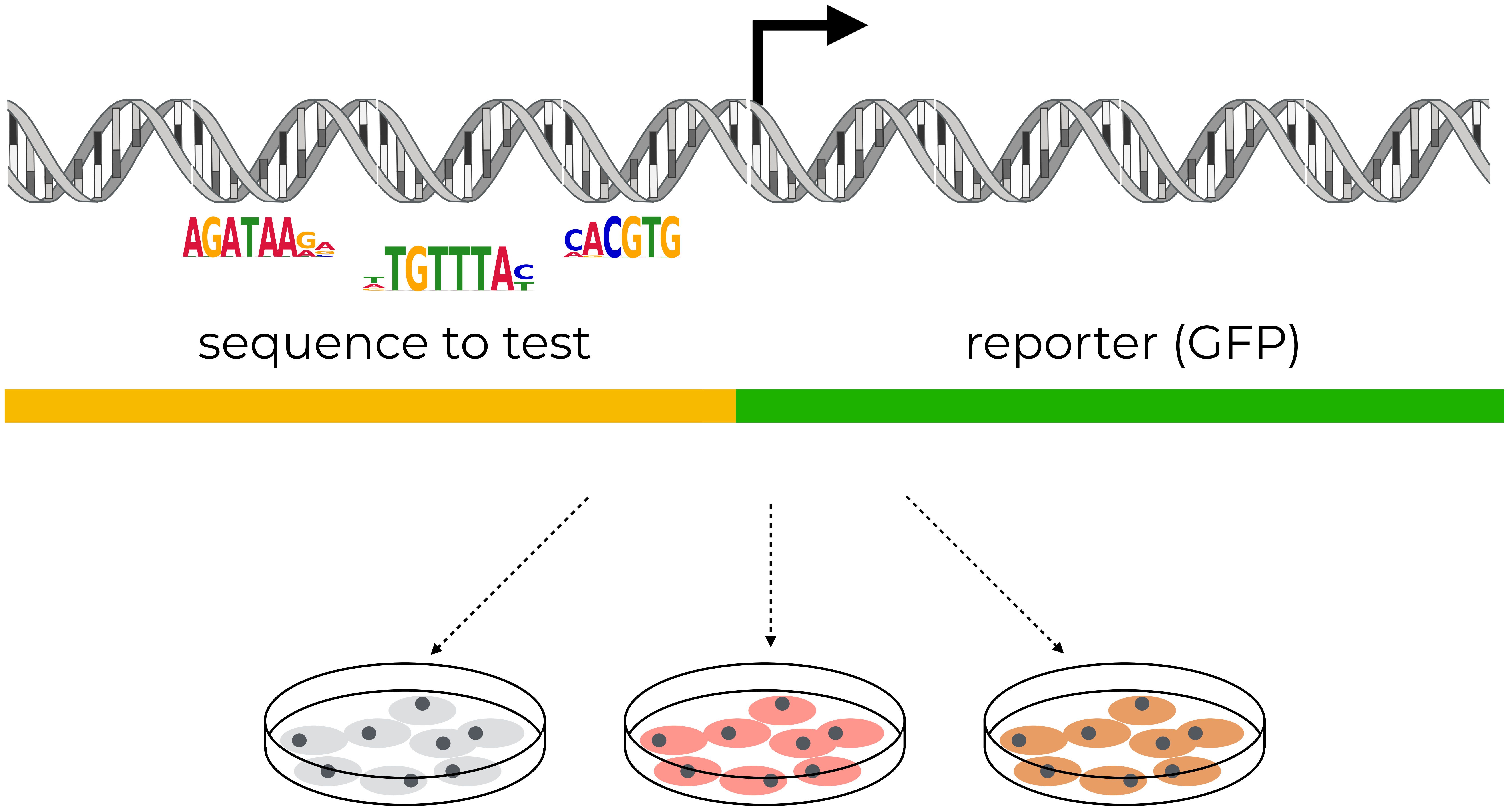


chromatin environment



post-transcriptional regulation

# Reporter assays



sequence to test

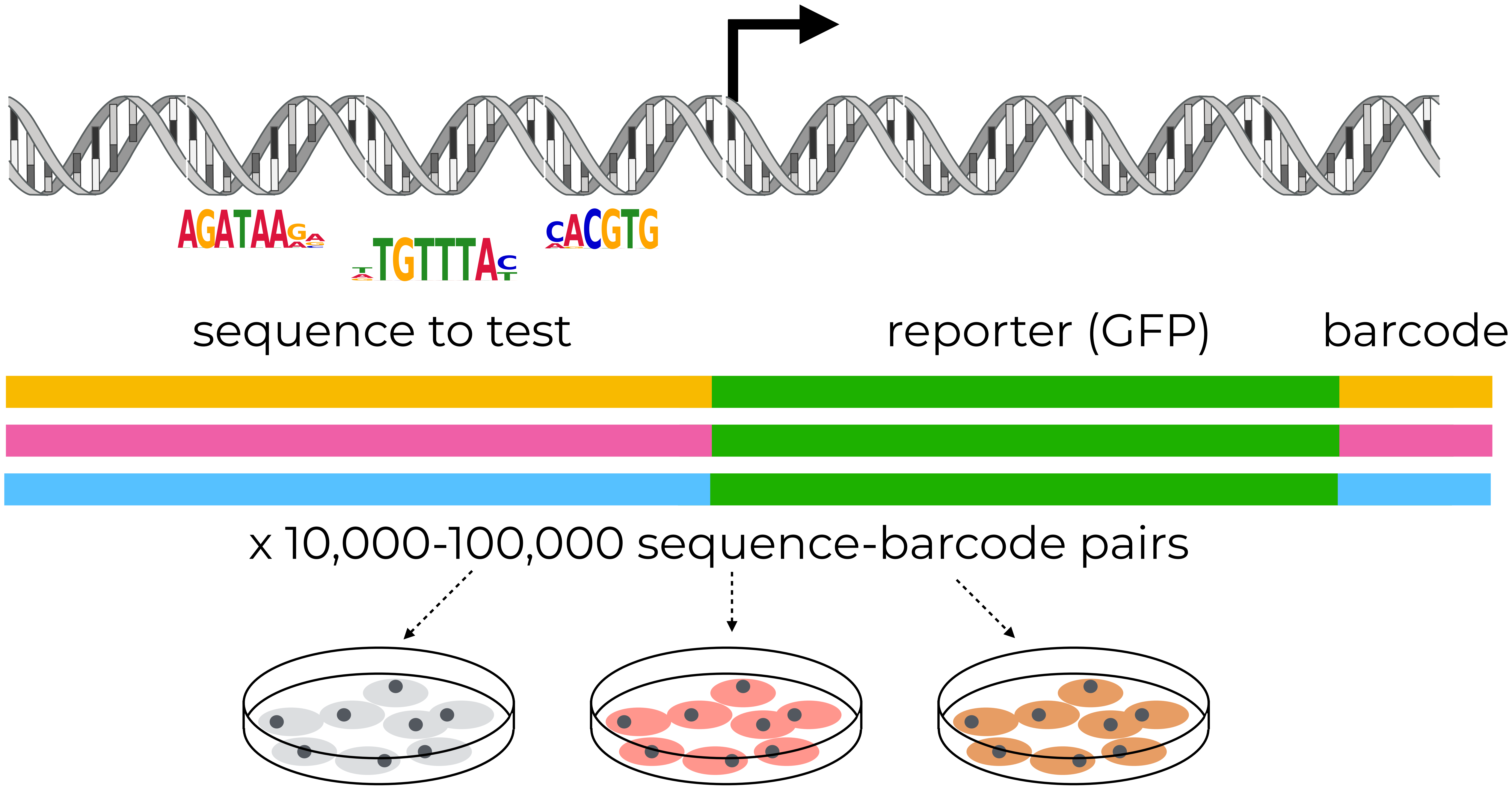
reporter (GFP)

cell type 1

cell type 2

cell type 3

# Massively parallel reporter assays (MPRA)



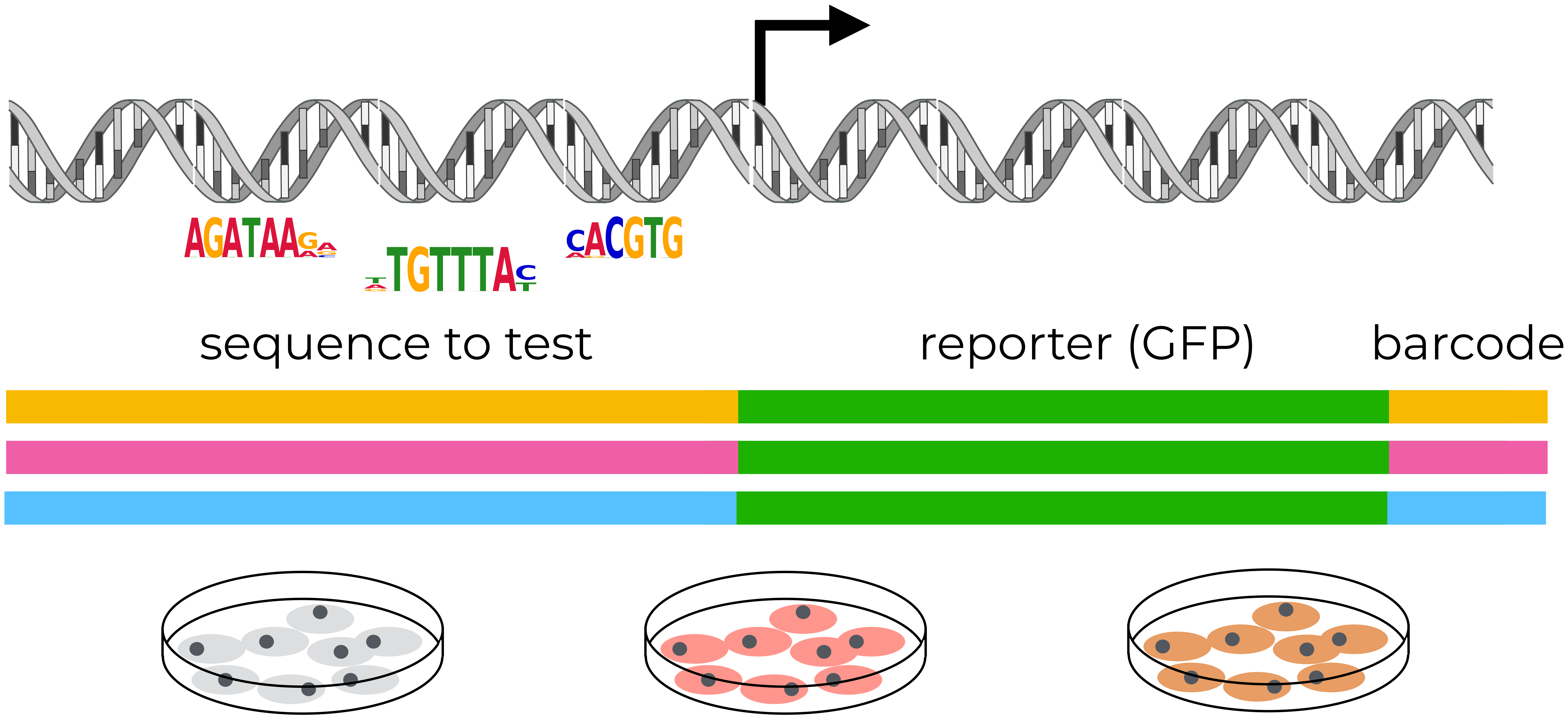
sequence to test

reporter (GFP)

barcode

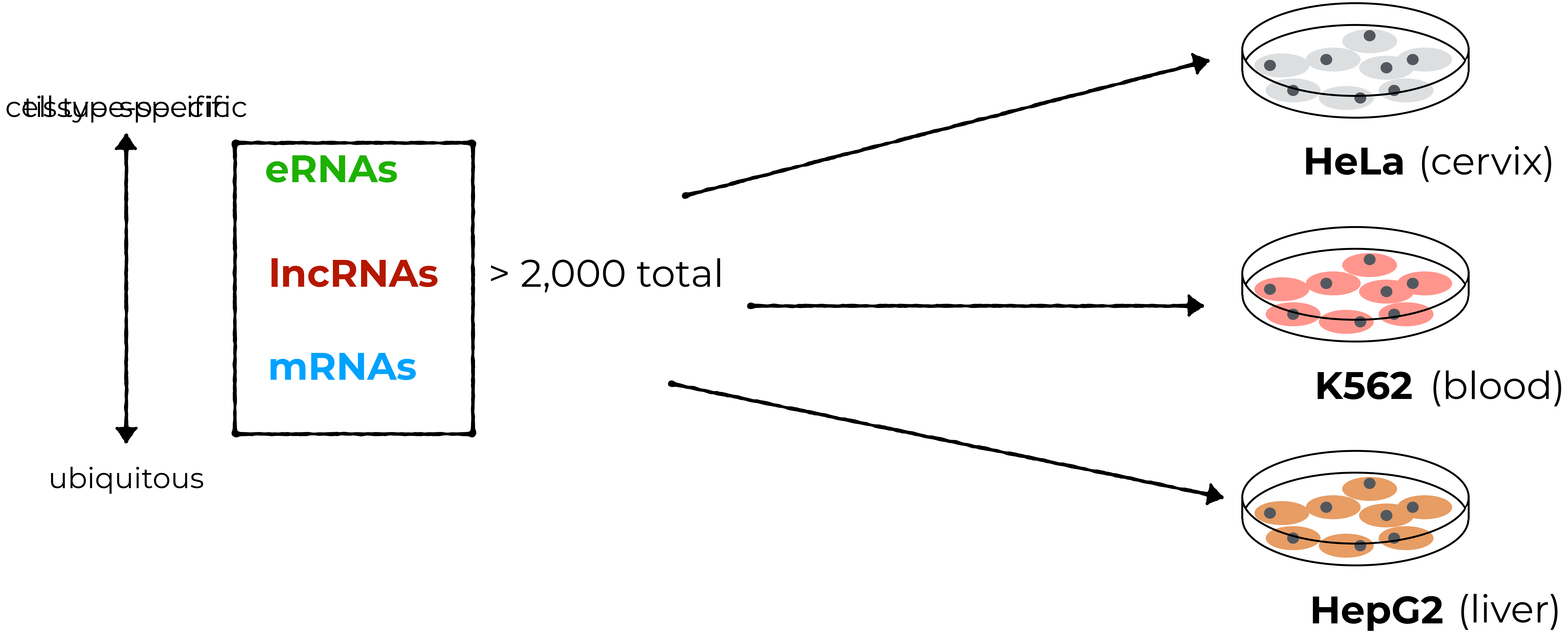
x 10,000-100,000 sequence-barcode pairs

# Massively parallel reporter assays (MPRA)



| barcode RNA-seq counts | | | |
|---|---|---|---|
| 🟨 | 100 | 105 | 120 |
| 🟪 | 2400 | 54 | 30 |
| 🟦 | 15 | 600 | 24 |

# MPRA on endogenous core promoters in 3 diverse cell lines



cell type-specific
tissue-specific

ubiquitous

eRNAs

lncRNAs

mRNAs

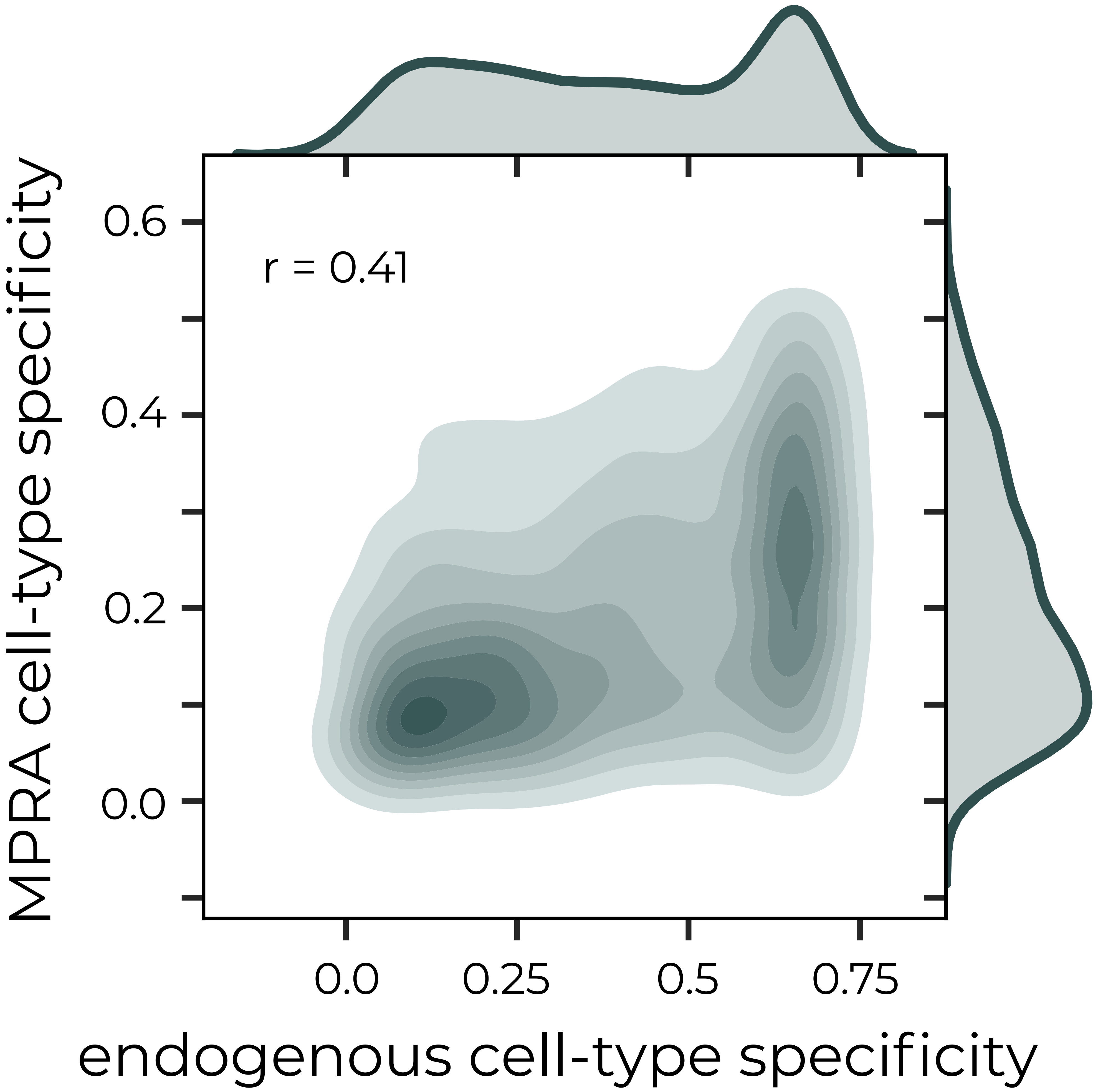
> 2,000 total

**HeLa** (cervix)

**K562** (blood)

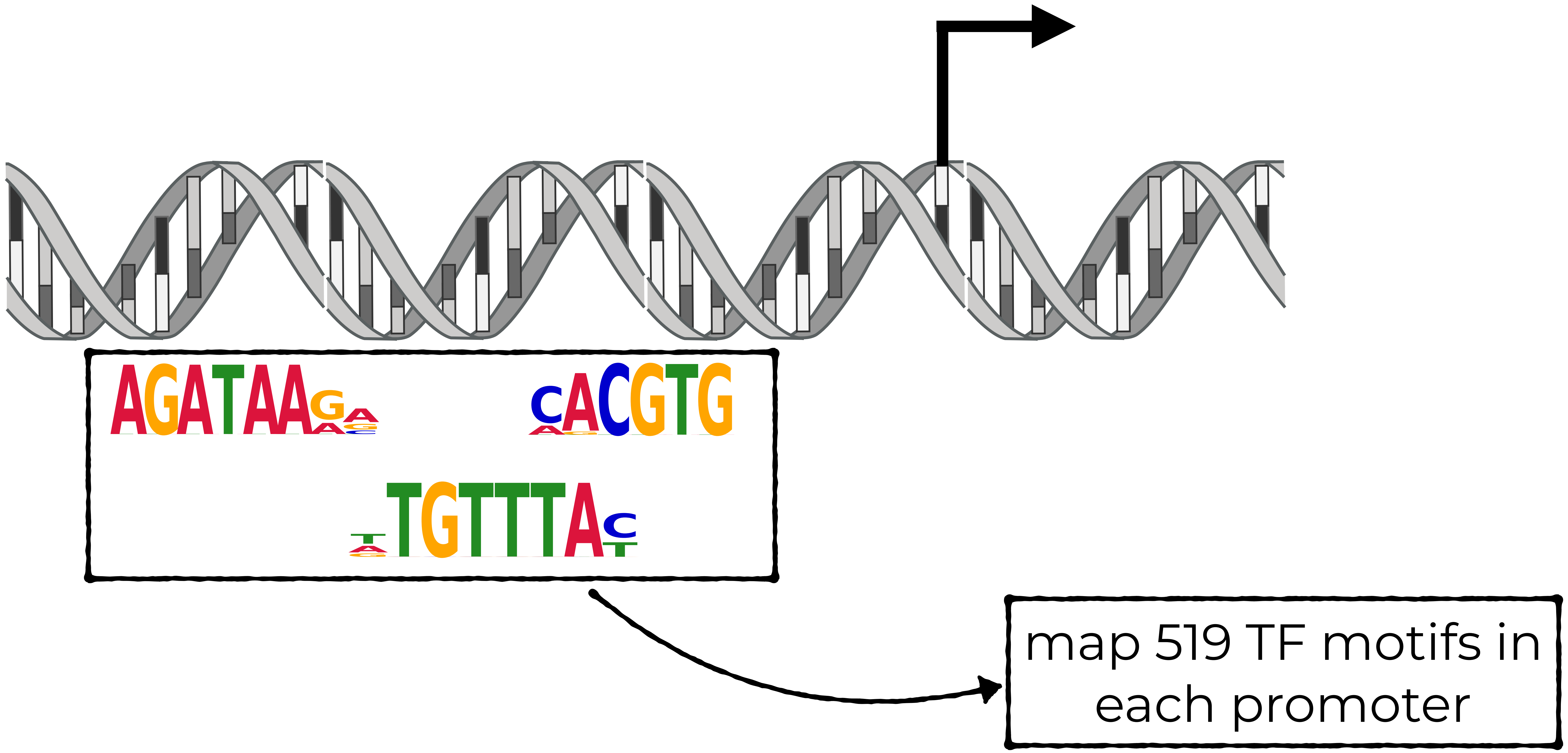
**HepG2** (liver)

# MPRA recapitulates endogenous expression and specificity



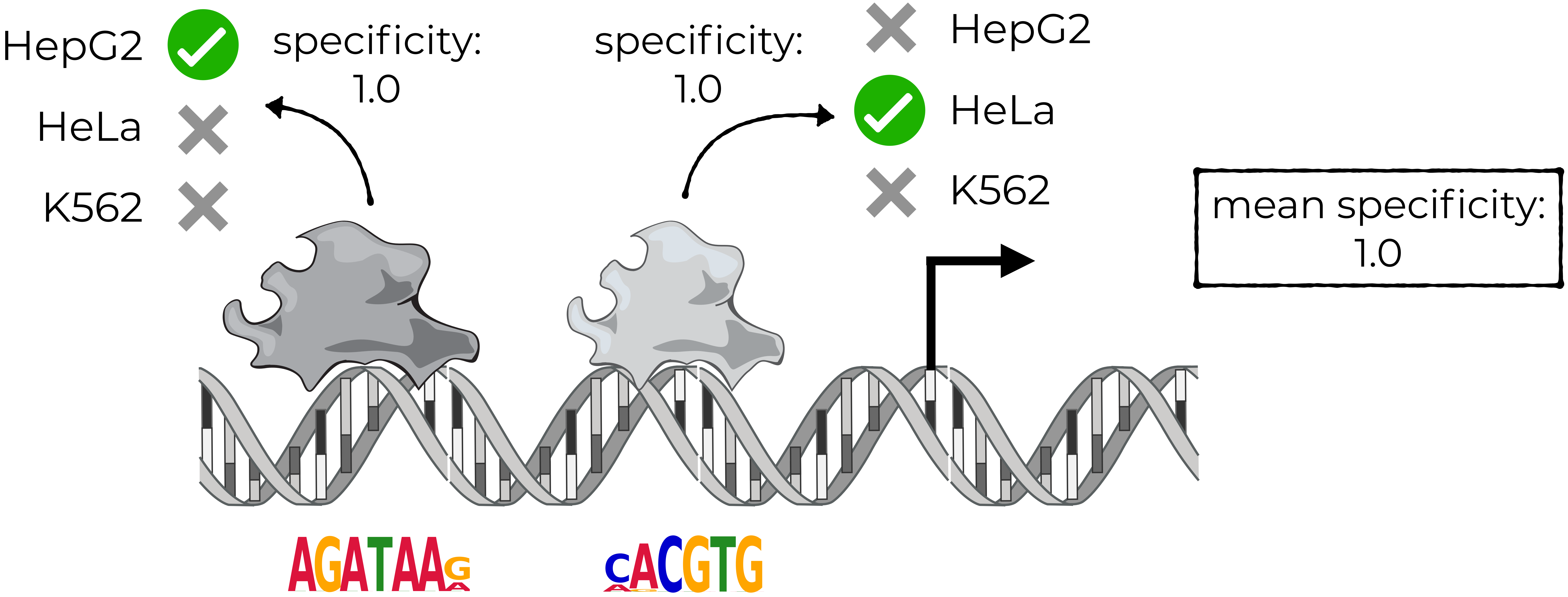
Mattioli *et al.* Genome Research 2019

# What sequence features in core promoters drive their specificity?



map 519 TF motifs in each promoter

# 1) cell-type specificities of the TF proteins themselves



HepG2 ✓
HeLa ✗
K562 ✗

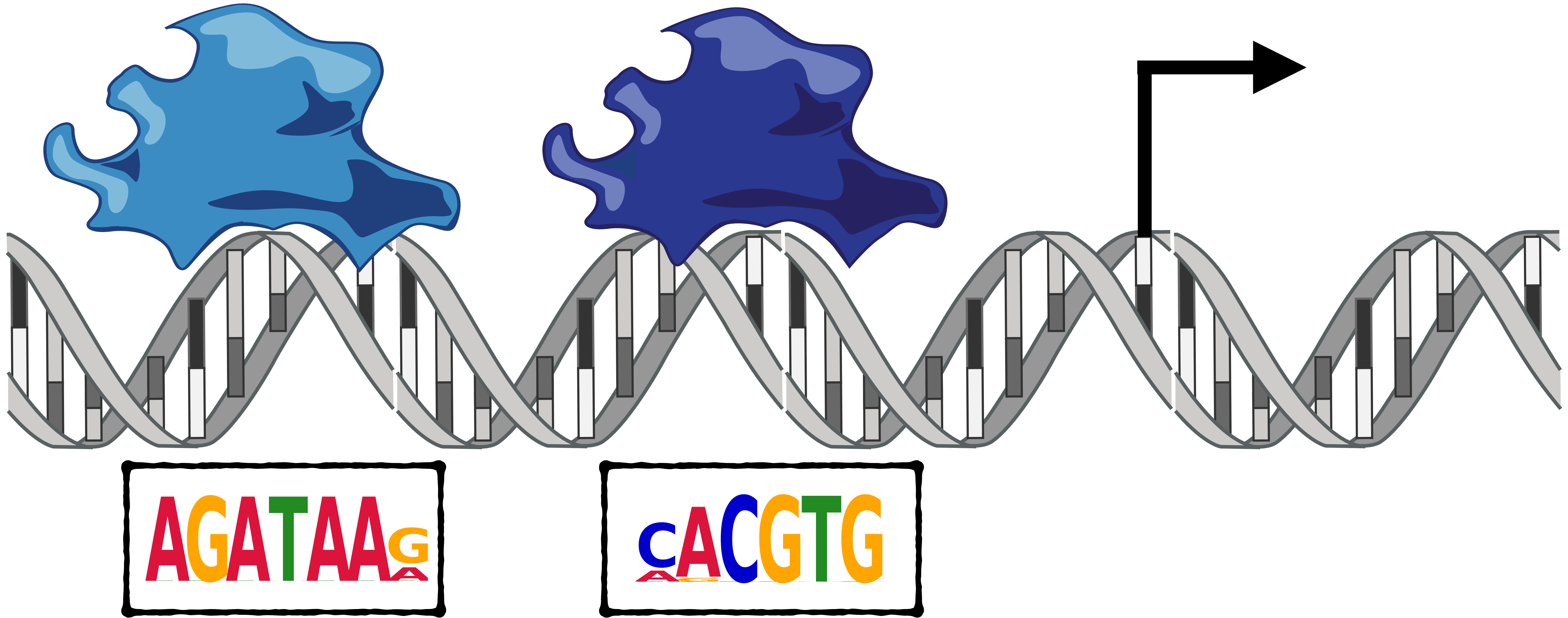
specificity:
1.0

specificity:
1.0

✗ HepG2
✓ HeLa
✗ K562

mean specificity:
1.0

# 2) number of basepairs covered by a TF motif

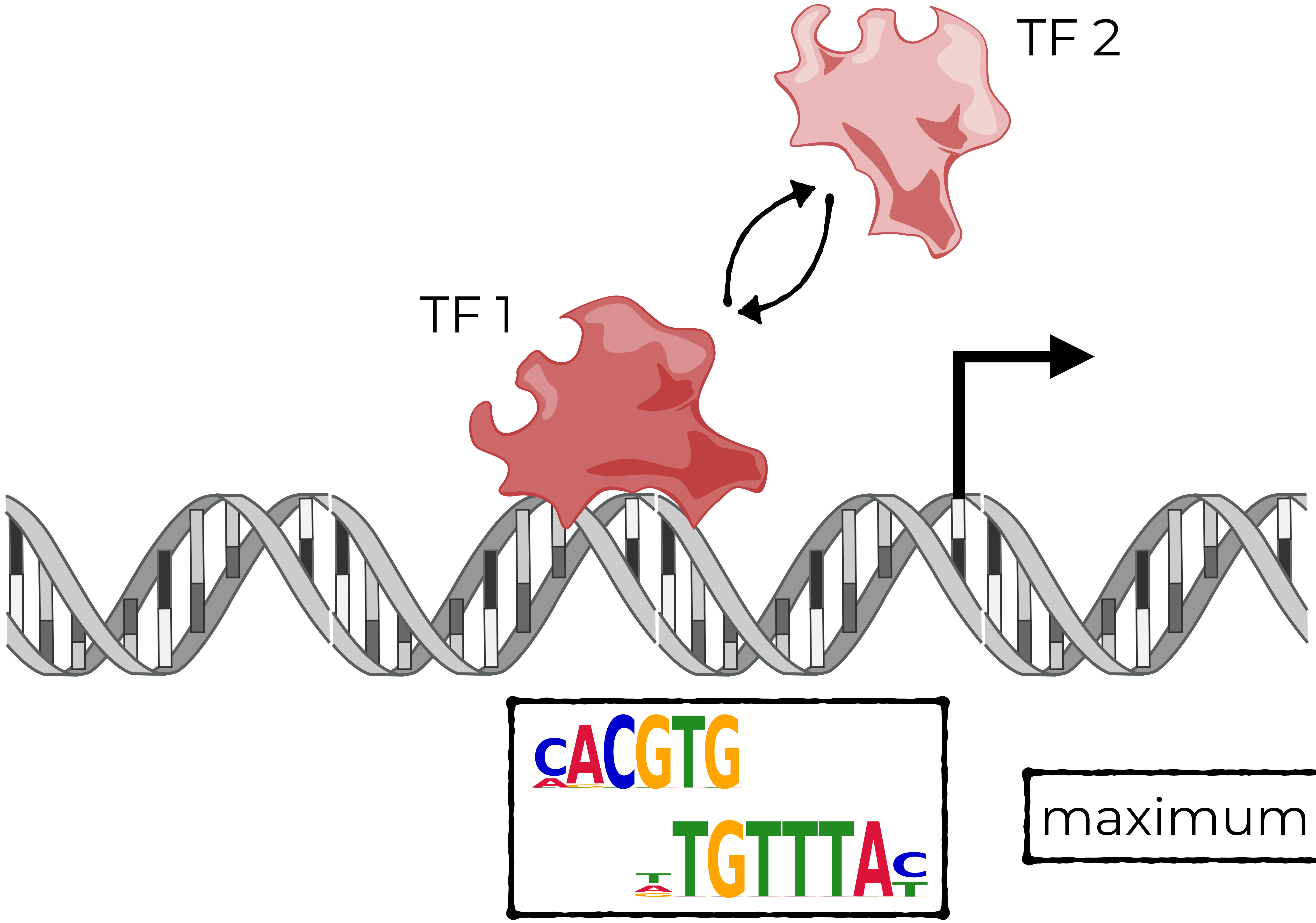


**7**

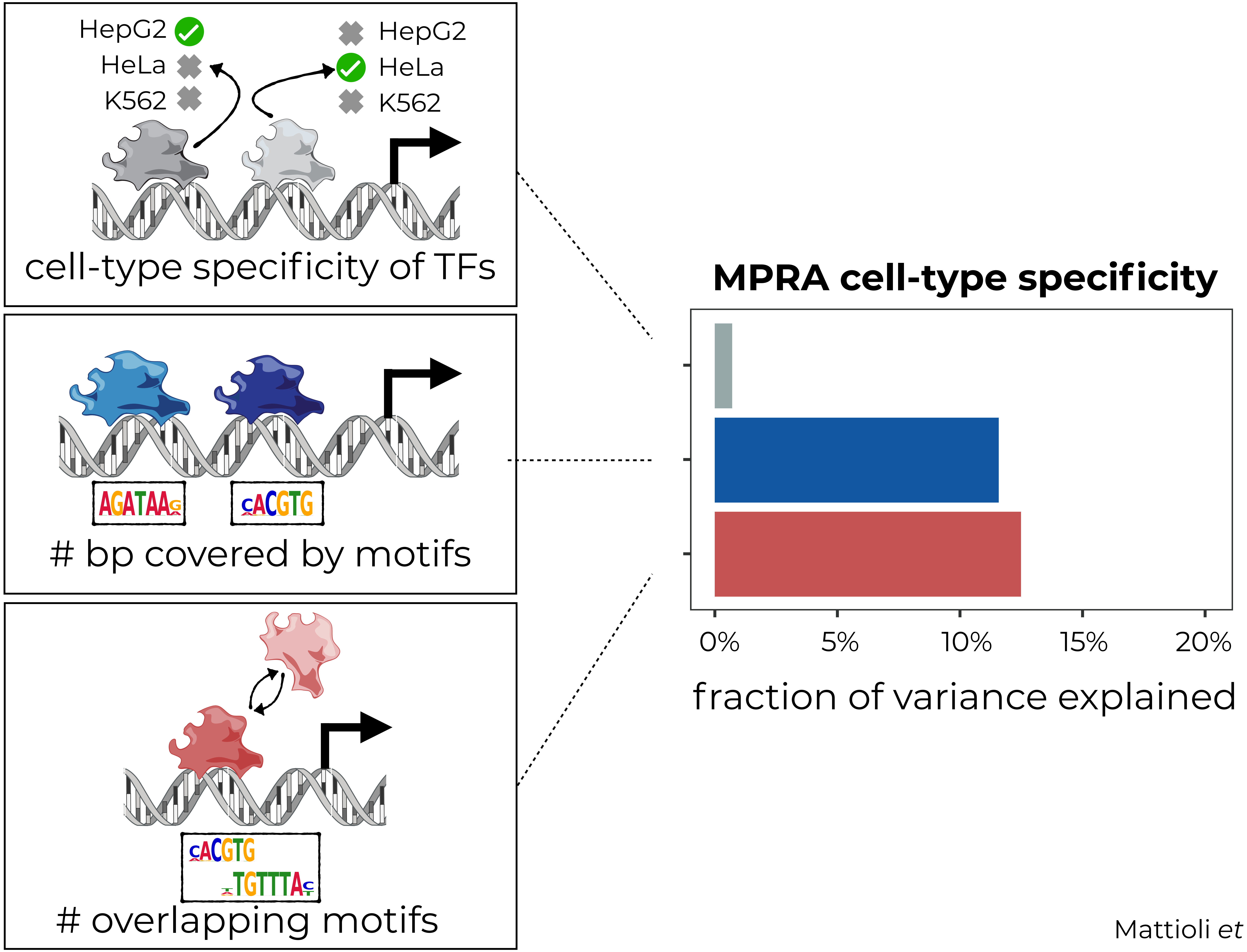
**6**

13 total bp covered by motif

# 3) number of overlapping motifs

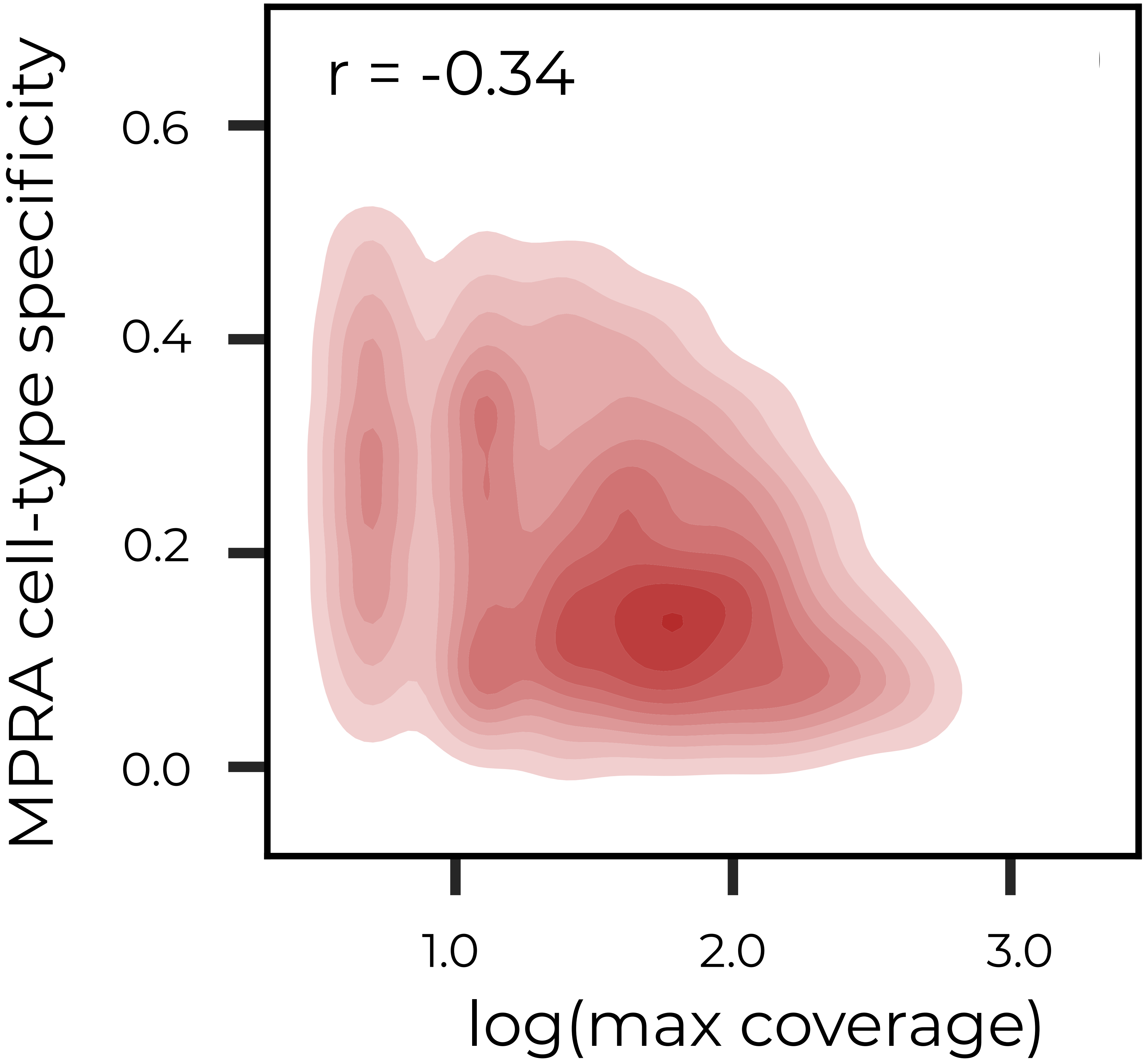


maximum coverage = 2



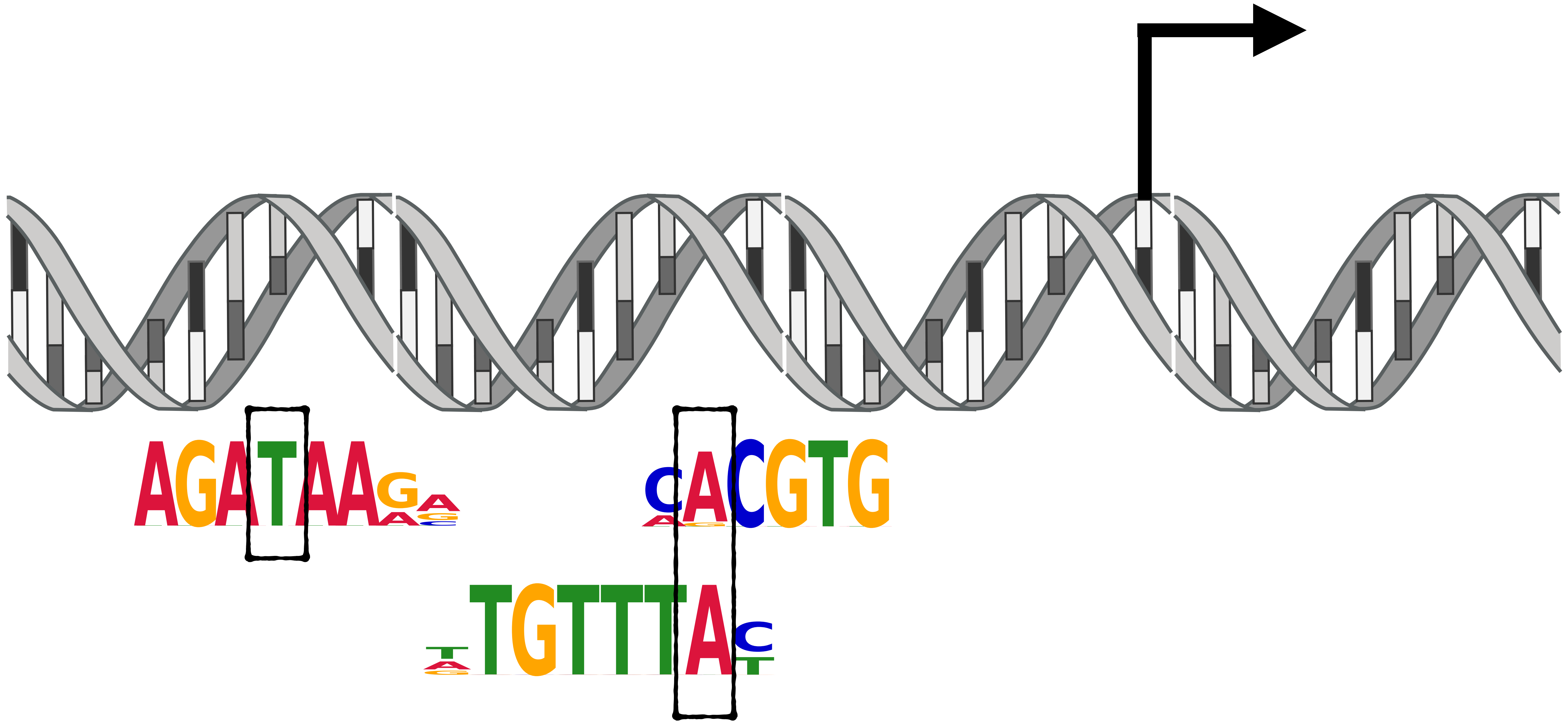
Mattioli *et al.* Genome Research 2019

# Overlapping motifs are associated with ubiquitous expression

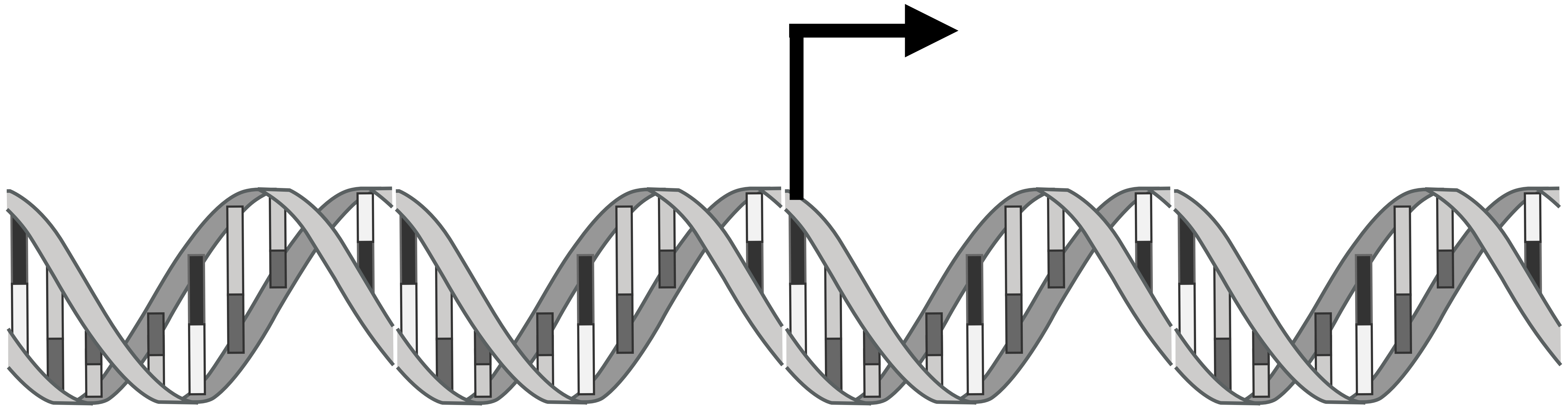


Mattioli *et al.* Genome Research 2019

# Does perturbing overlapping motifs result in strong effect sizes?



# Scanning mutagenesis MPRA

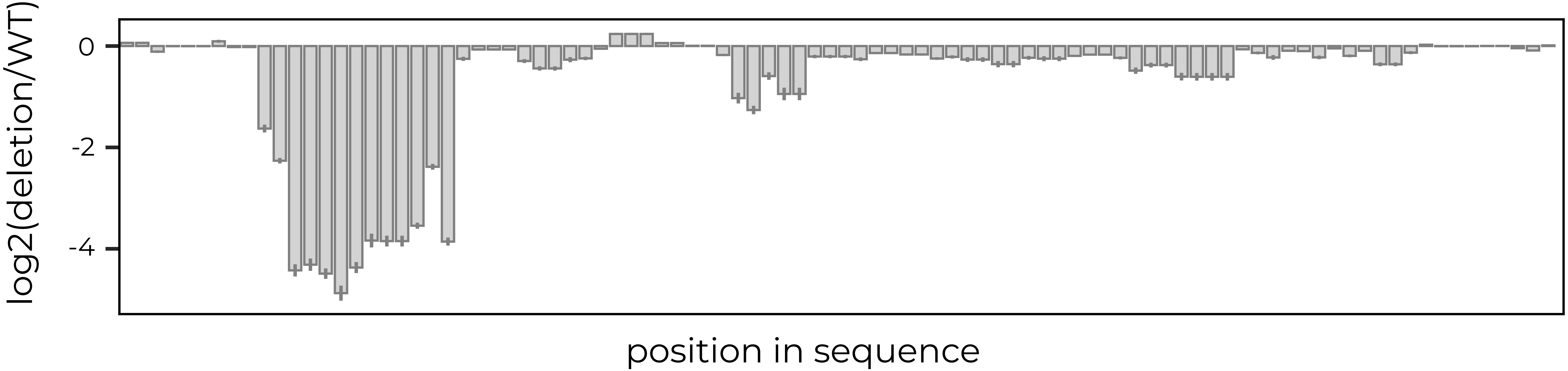


core promoter sequence CGGCCGCACCAAACCGCCGCGTACTATAAAAGTCGCTGC

scanning single nucleotide deletions

-GGCCGCACCAAACCGCCGCGTACTATAAAAGTCGCTGC
C-GCCGCACCAAACCGCCGCGTACTATAAAAGTCGCTGC
CG-CCGCACCAAACCGCCGCGTACTATAAAAGTCGCTGC
CGG-CGCACCAAACCGCCGCGTACTATAAAAGTCGCTGC

# Single-nucleotide deletions reveal TF binding sites

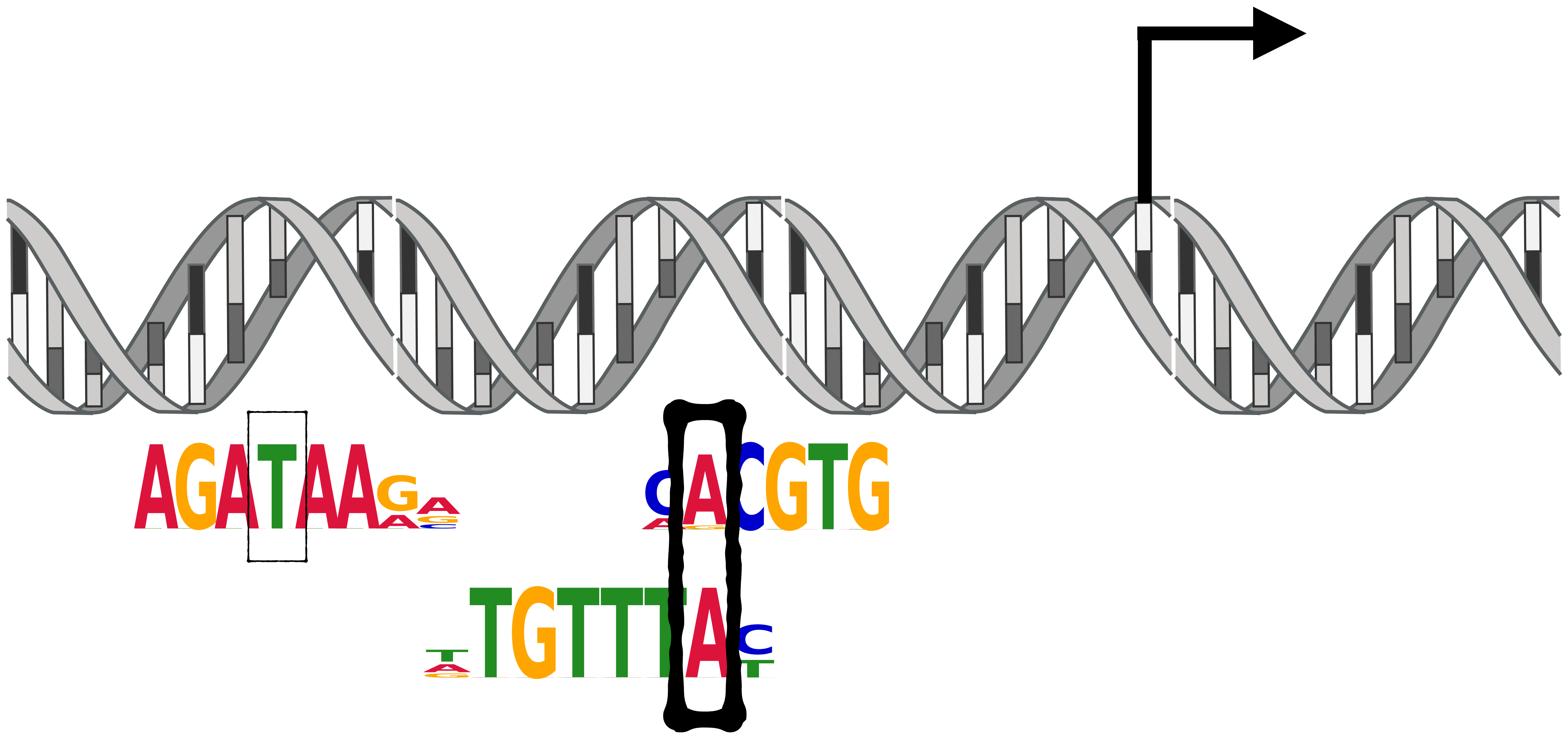


Mattioli *et al.* Genome Research 2019

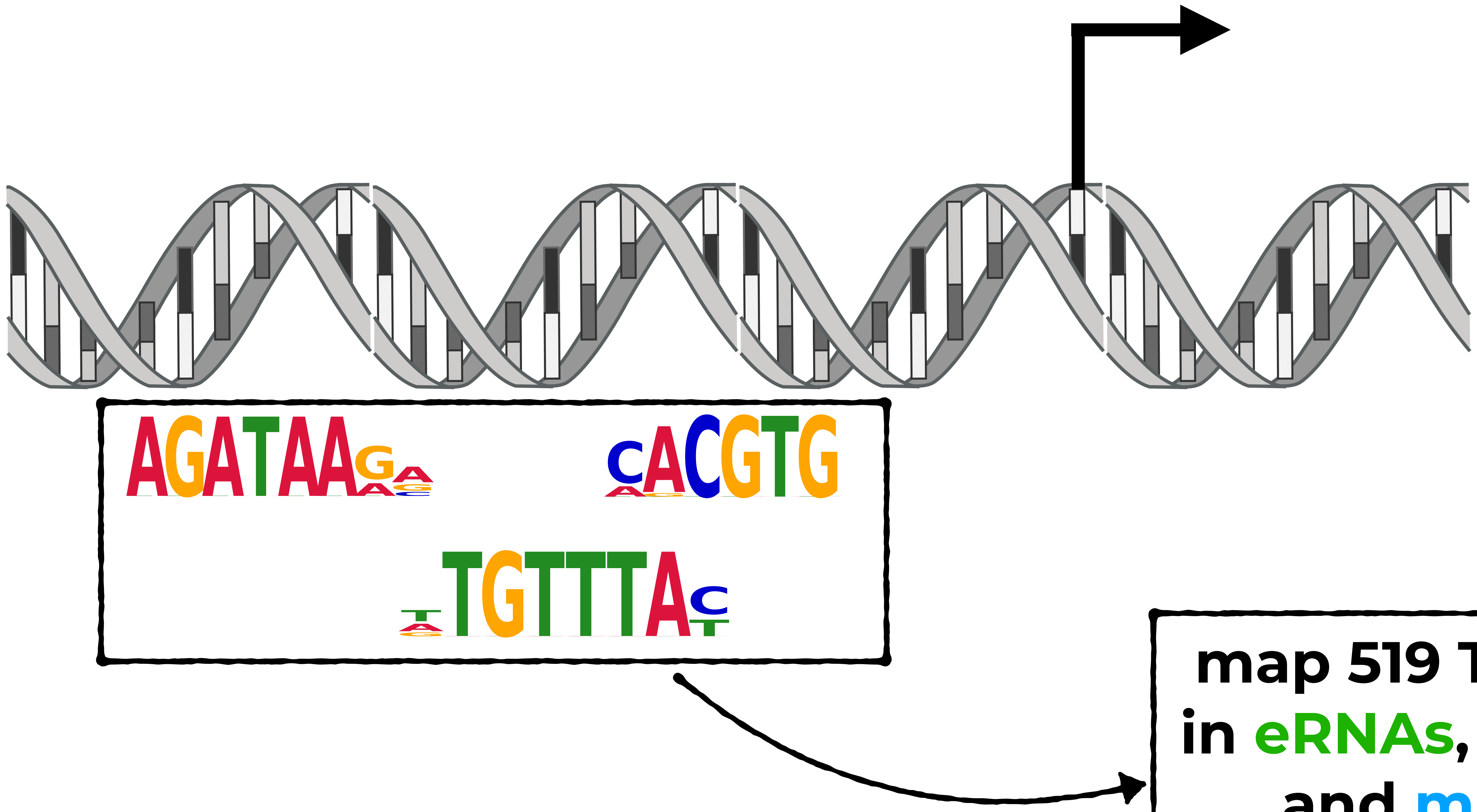
# Single-nucleotide deletions reveal TF binding sites

Mattioli *et al.* Genome Research 2019

# Perturbing overlapping motifs results in higher effect sizes

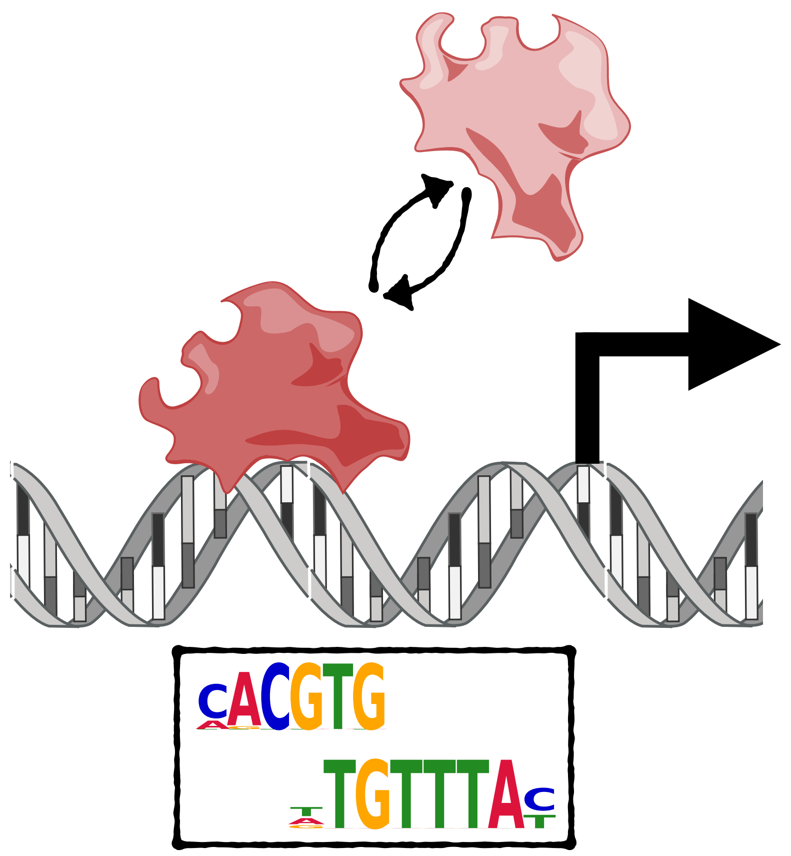


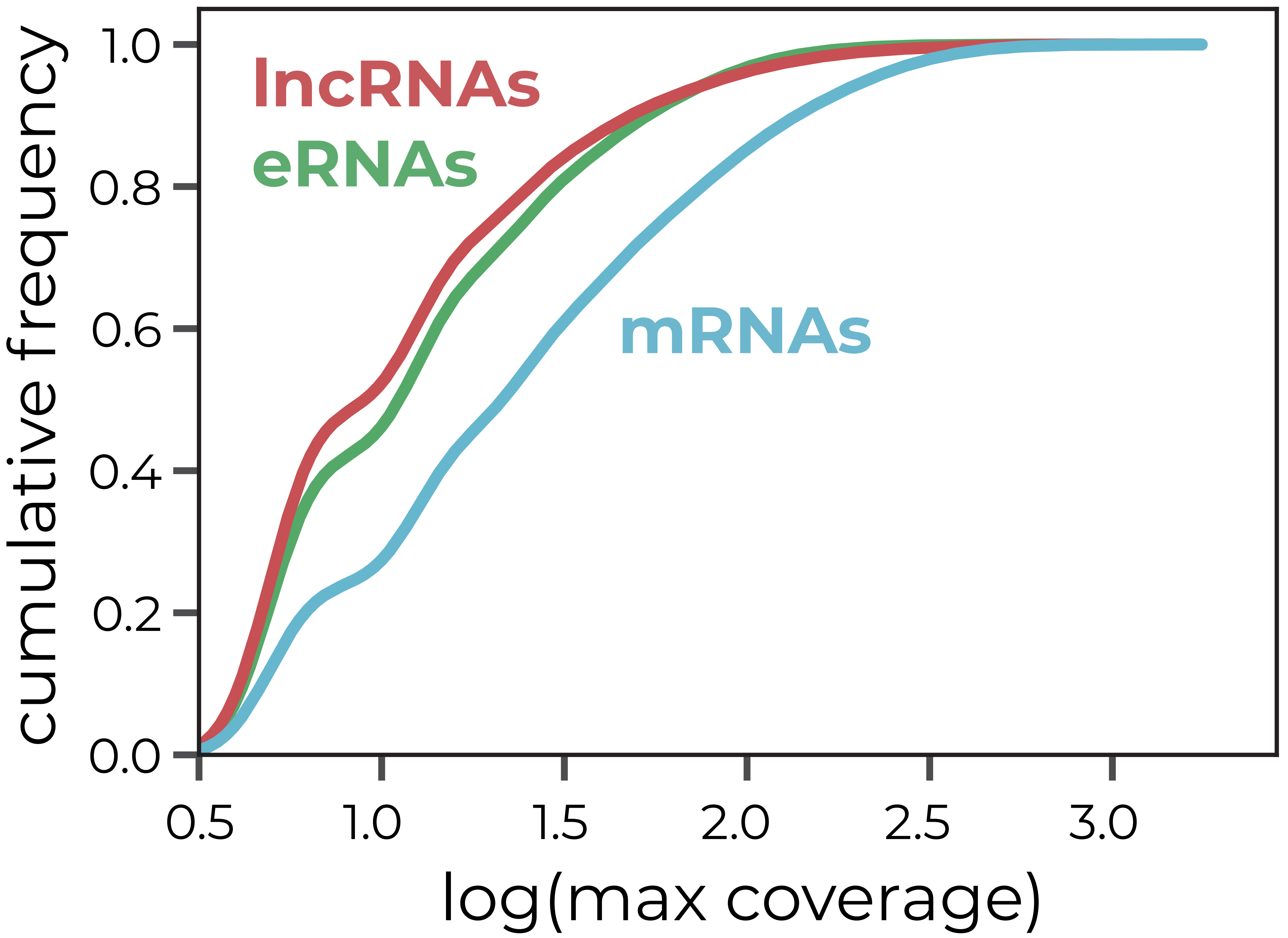
# What do the motif profiles of tissue-specific genes look like?



map 519 TF motifs
in eRNAs, lncRNAs,
and mRNAs,
genome-wide

# Genome-wide, mRNAs have more overlapping motifs than lncRNAs



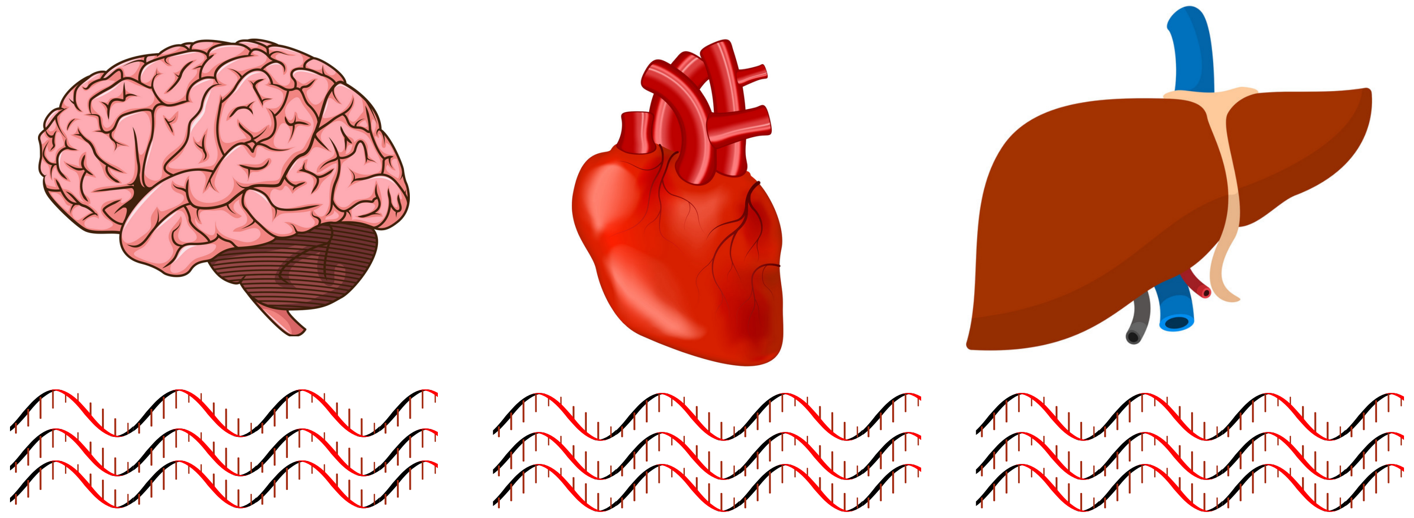


what if we
look within a
**single**
biotype?

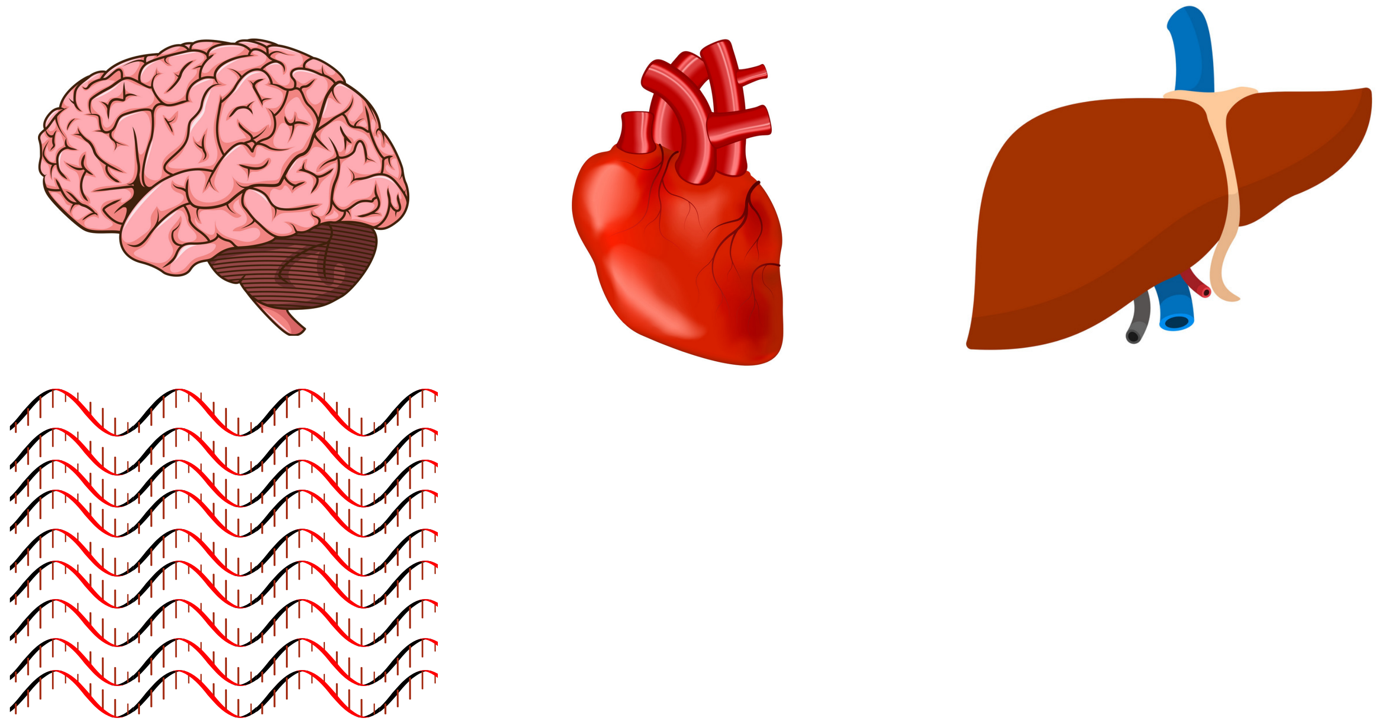
Mattioli *et al.* Genome Research 2019

# Defining expression profiles

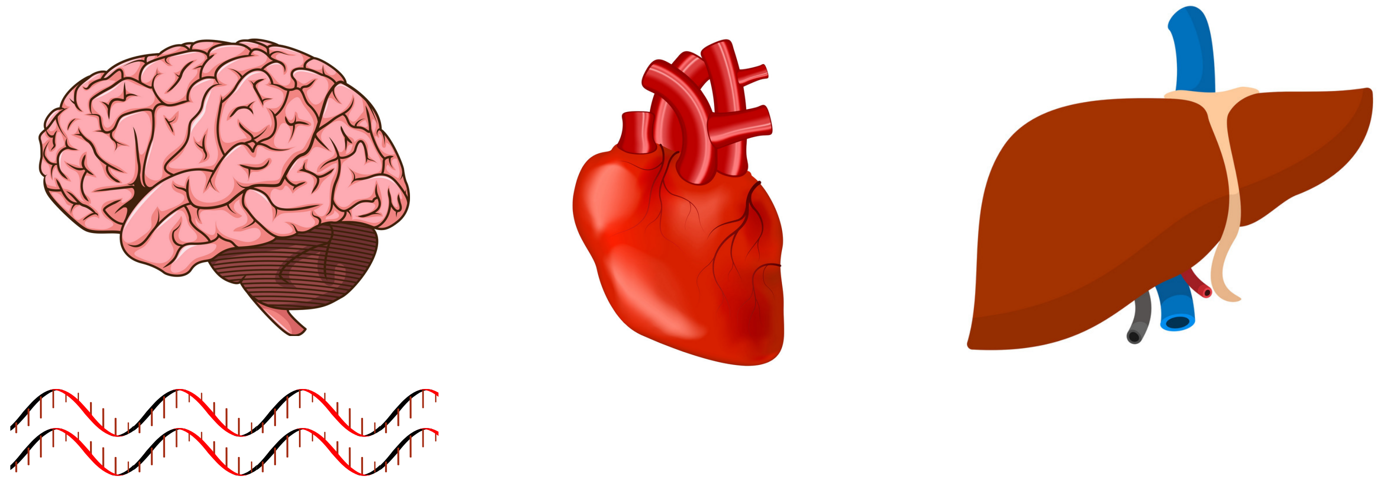
**ubiquitous**
expressed in >90% of samples



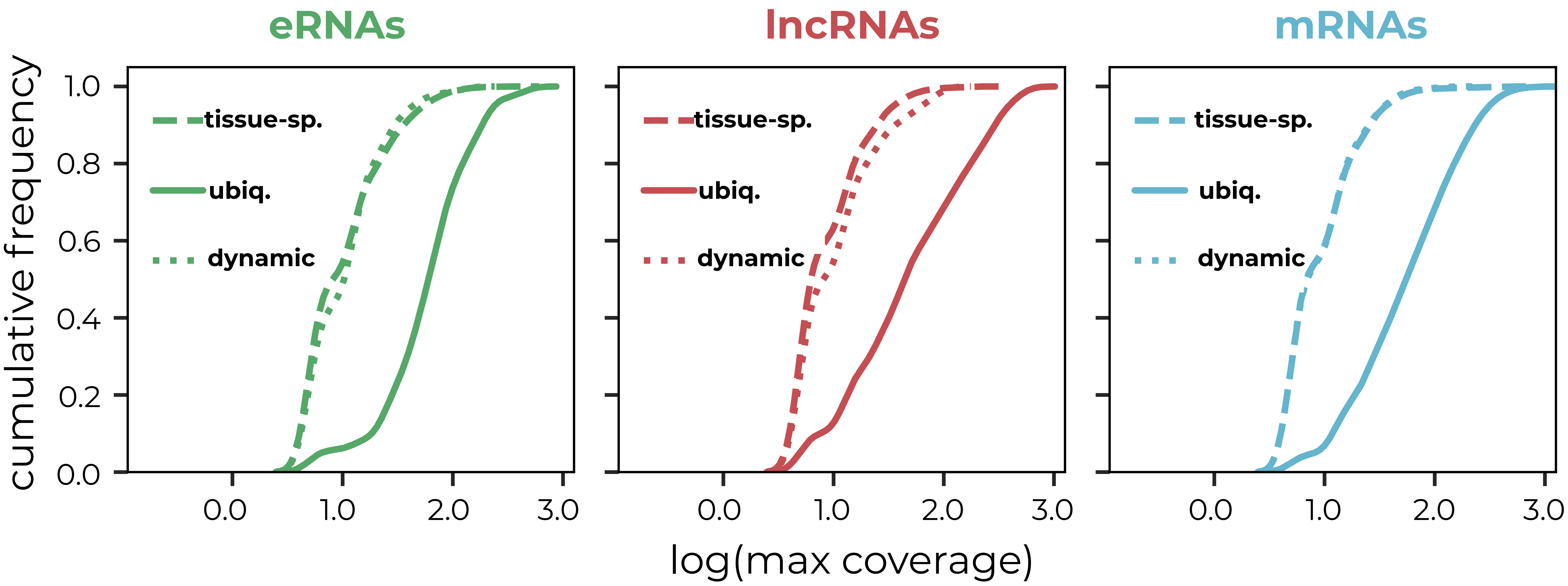
**dynamic**
expressed in <10% of samples,
with >=1 sample at >= 50 tpm



**tissue-specific**
expressed in <10% of samples

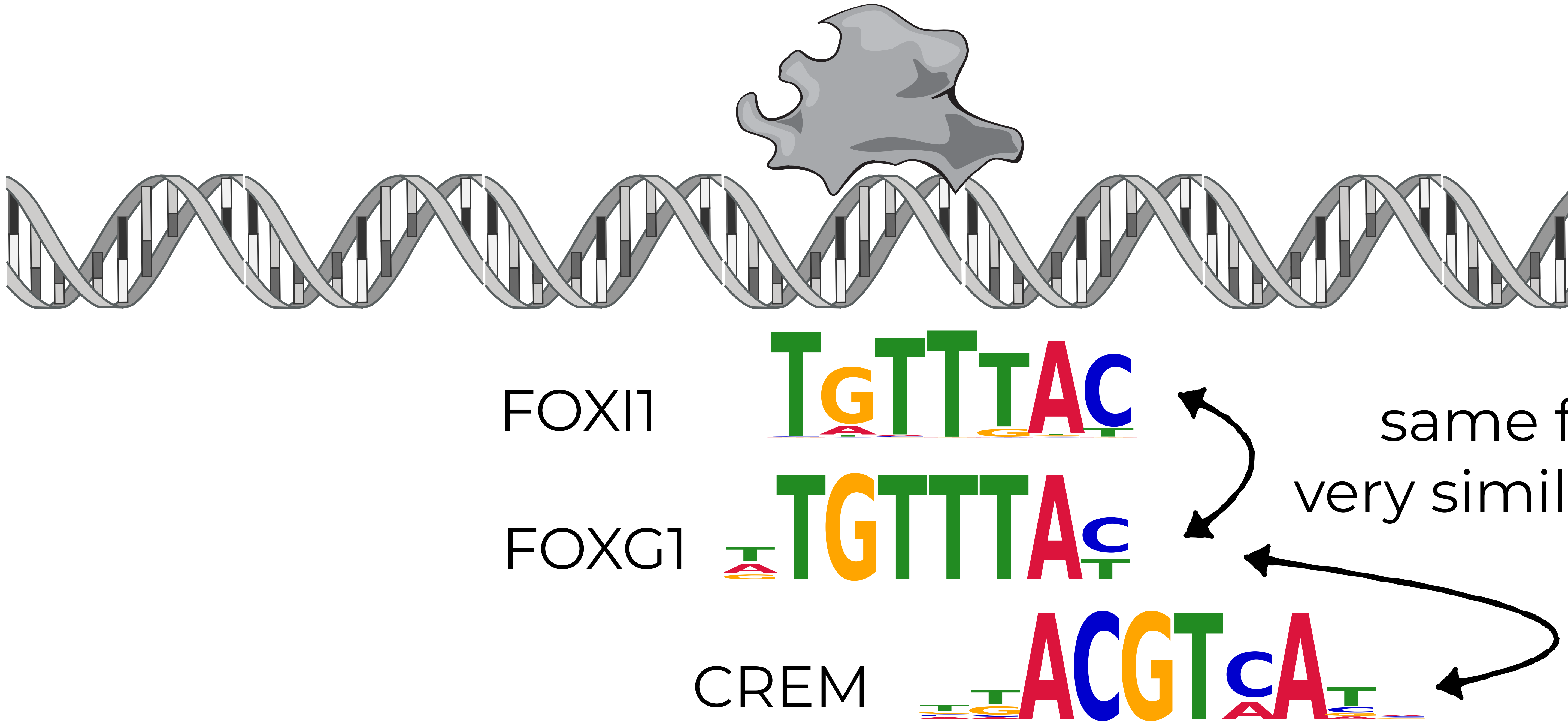


# Ubiquitously-expressed promoters within a biotype have more overlapping motifs



Mattioli *et al.* Genome Research 2019

# Are these overlapping motifs redundant?



FOXI1

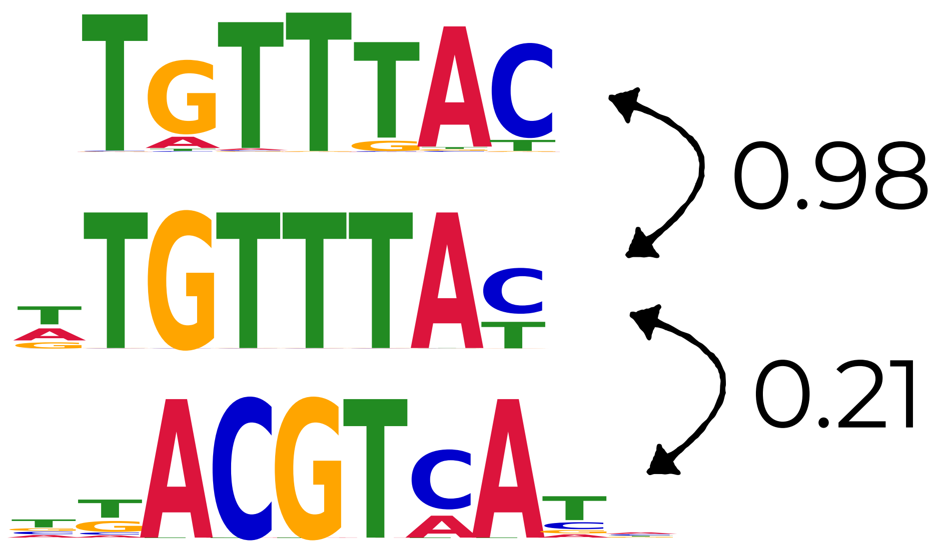
FOXG1

CREM

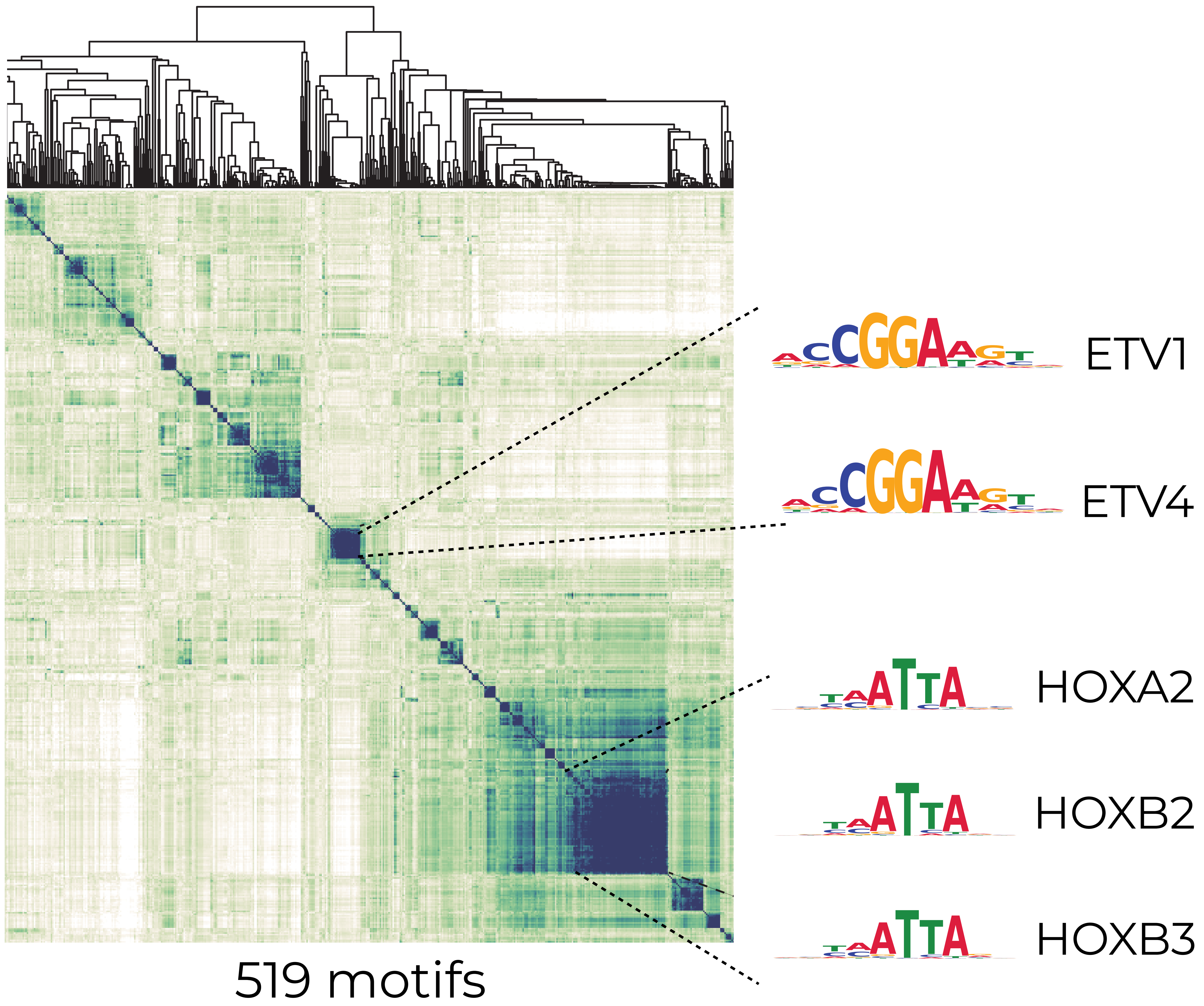
same family,
very similar motifs

different family,
different motifs

# Clustering similar motifs to control for redundancies

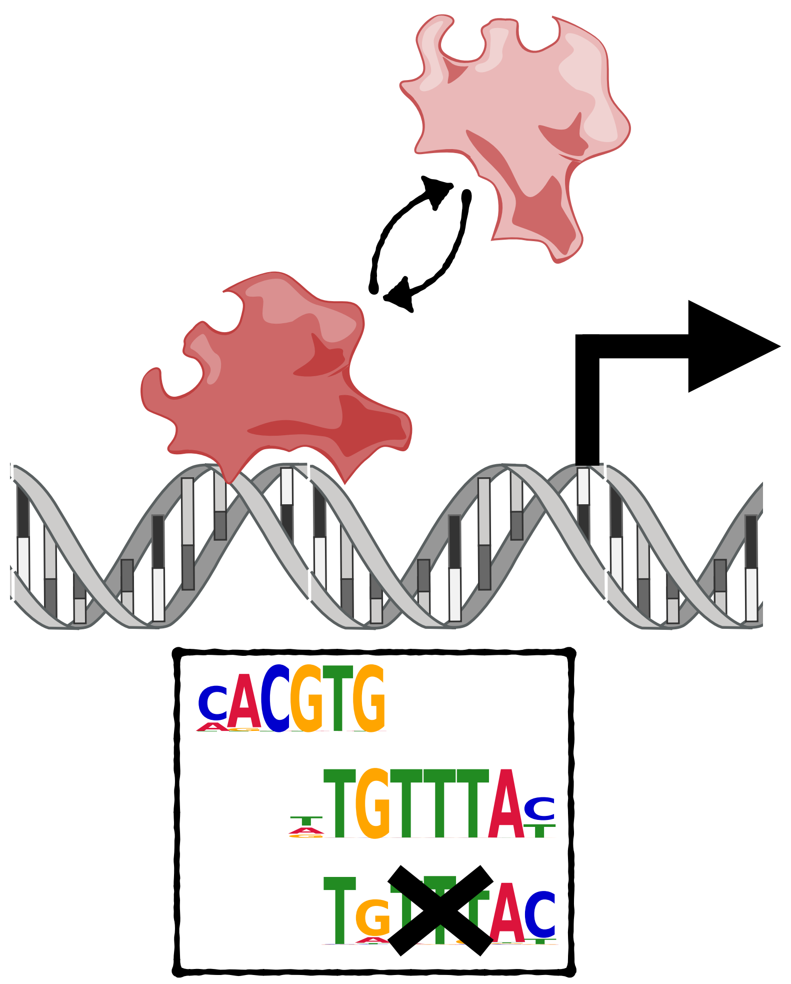


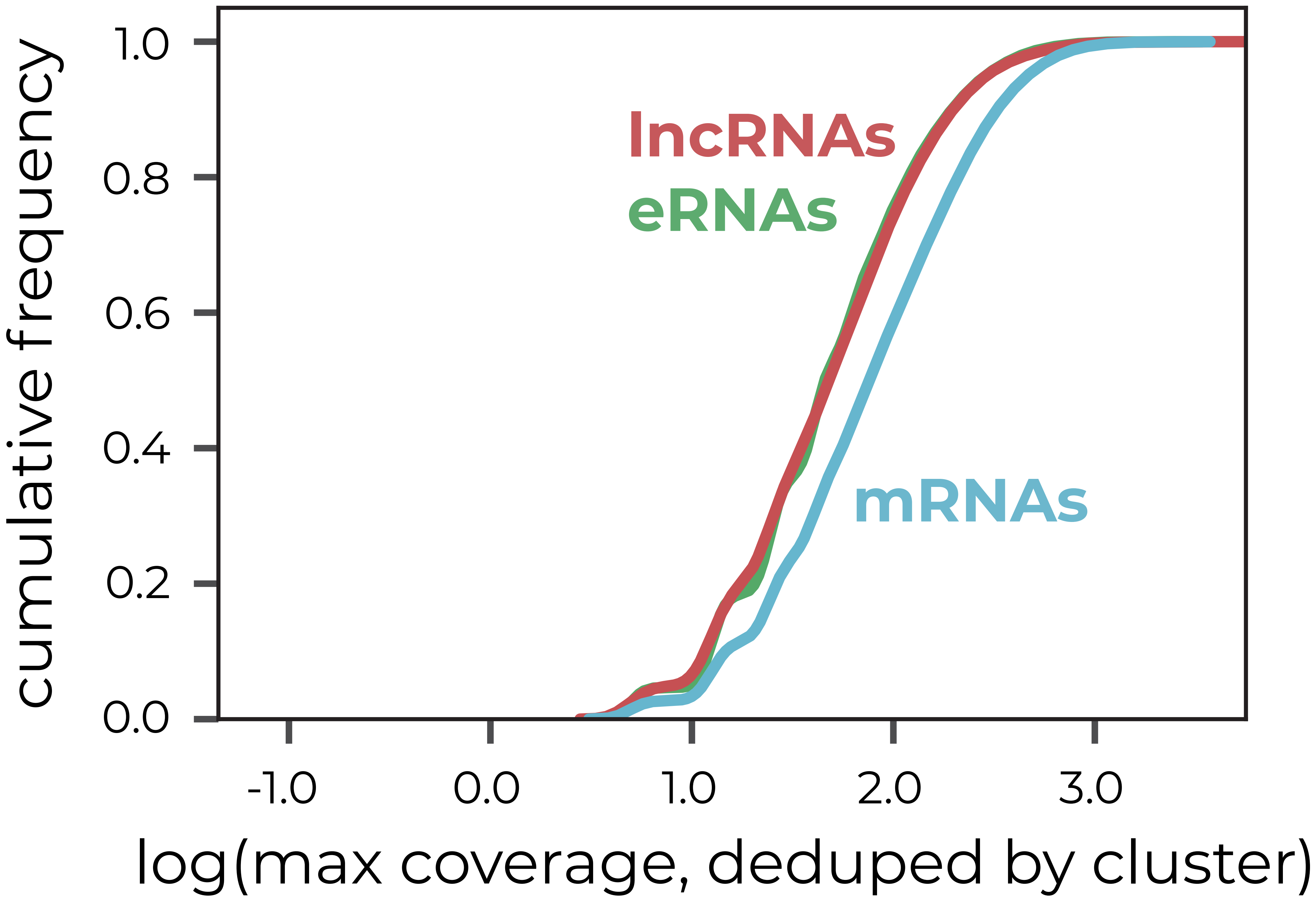
x 519 TF motifs



**220 motif clusters**

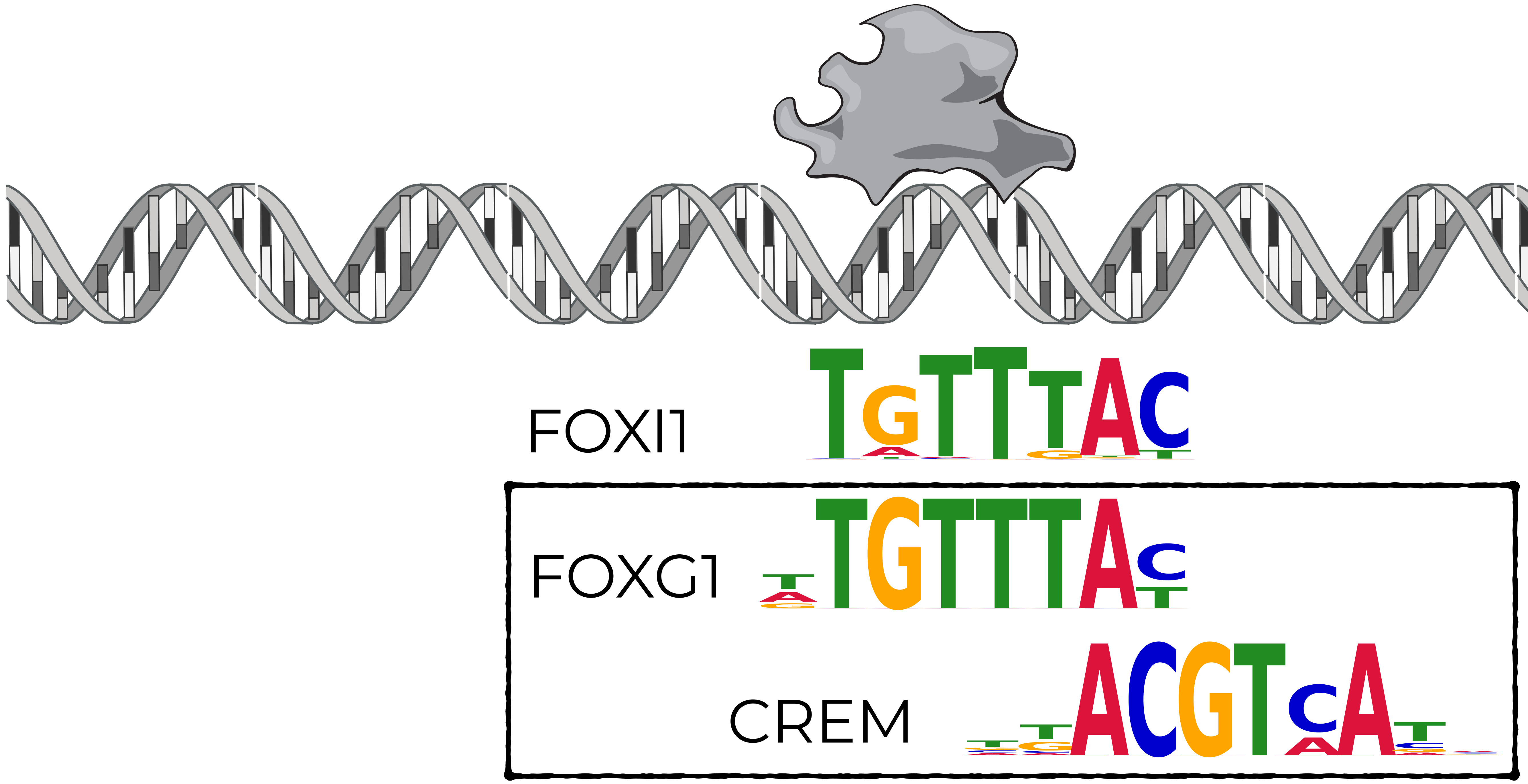
# Genome-wide motif observations hold after removing redundant motifs





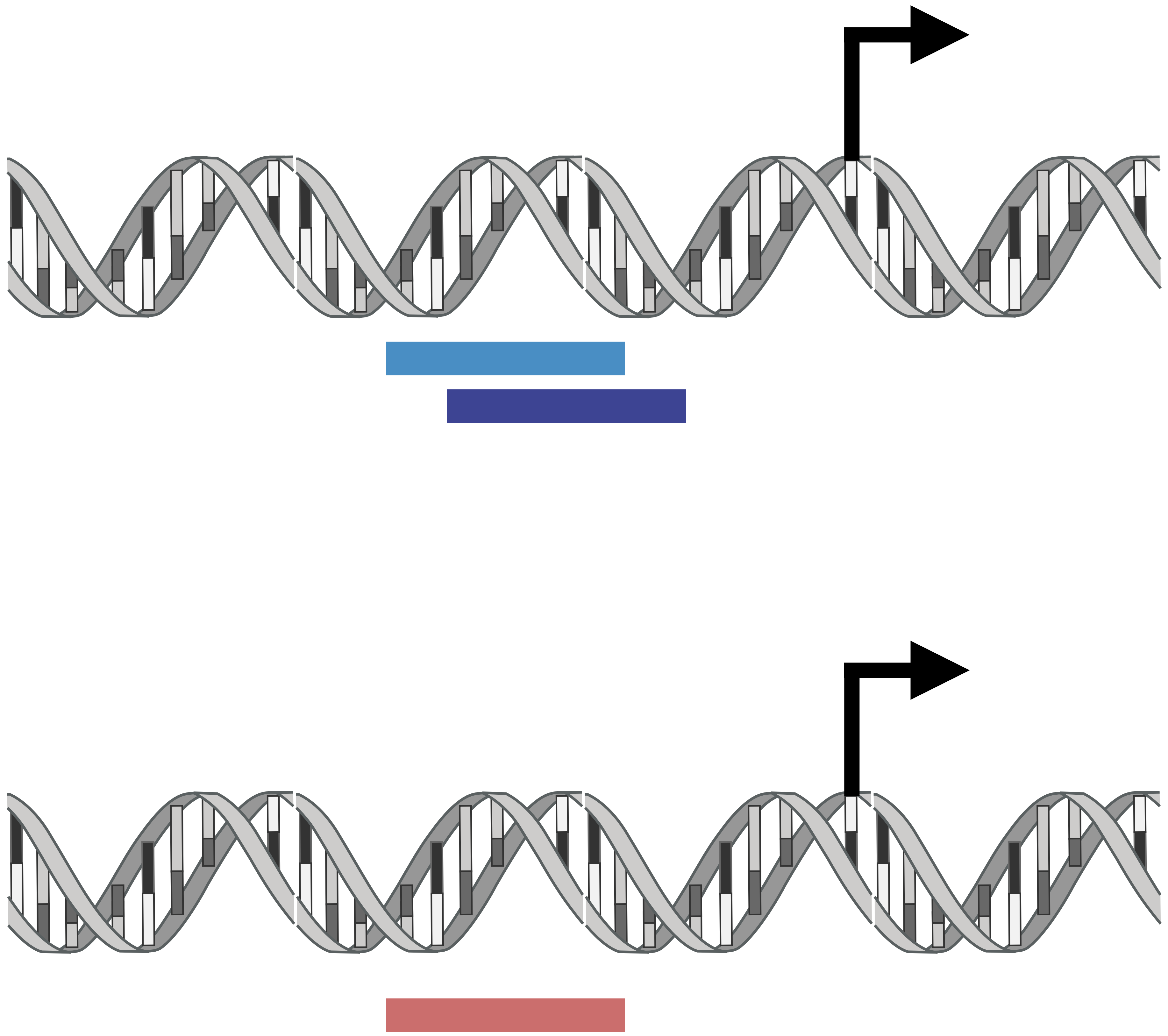
Mattioli *et al.* Genome Research 2019

# Are these overlapping motifs redundant?



non-redundant motifs dominate

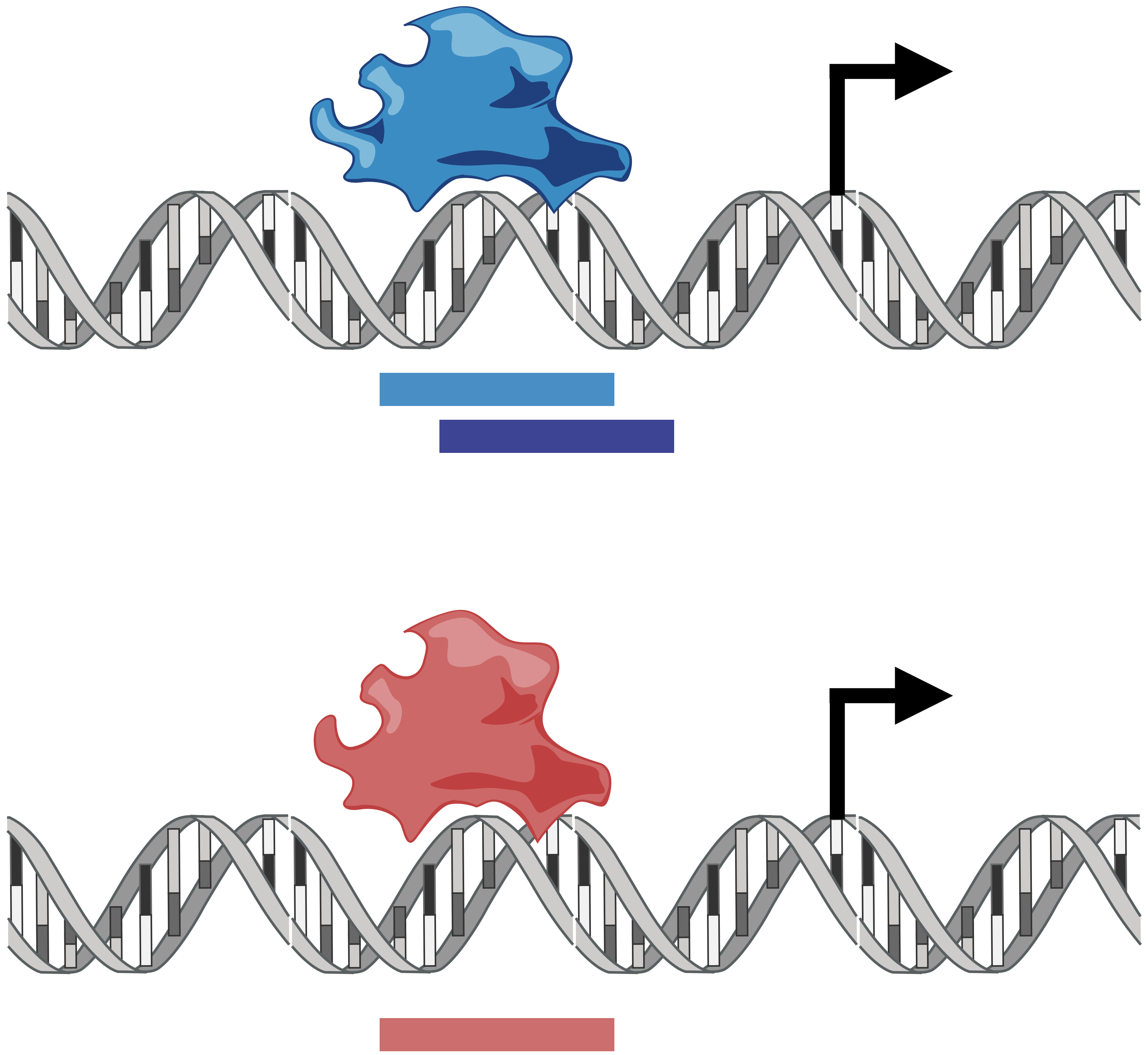
# Ubiquitously-expressed genes have diverse overlapping motifs



**mRNA**

**lncRNA**

# Ubiquitously-expressed genes have diverse overlapping motifs

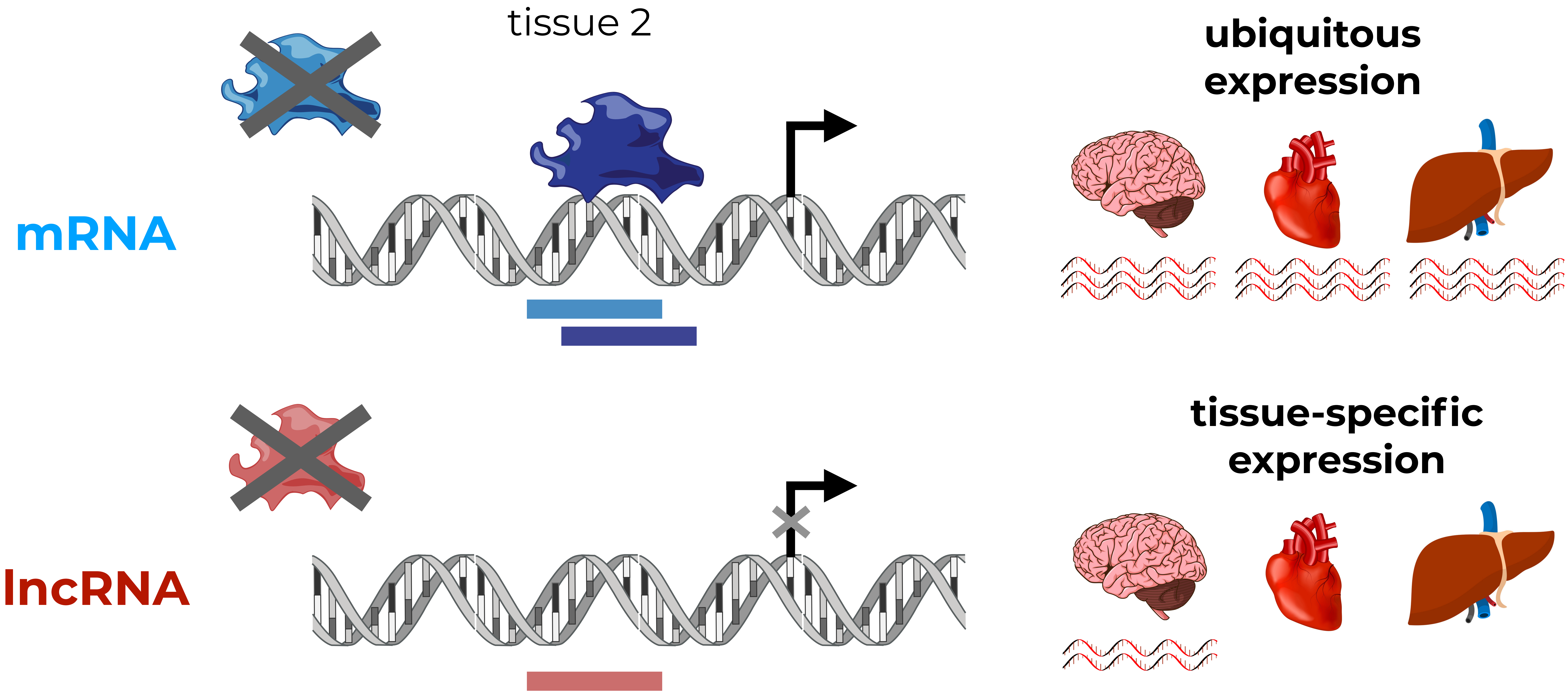


tissue 1

**mRNA**

**lncRNA**

# Ubiquitously-expressed genes have diverse overlapping motifs



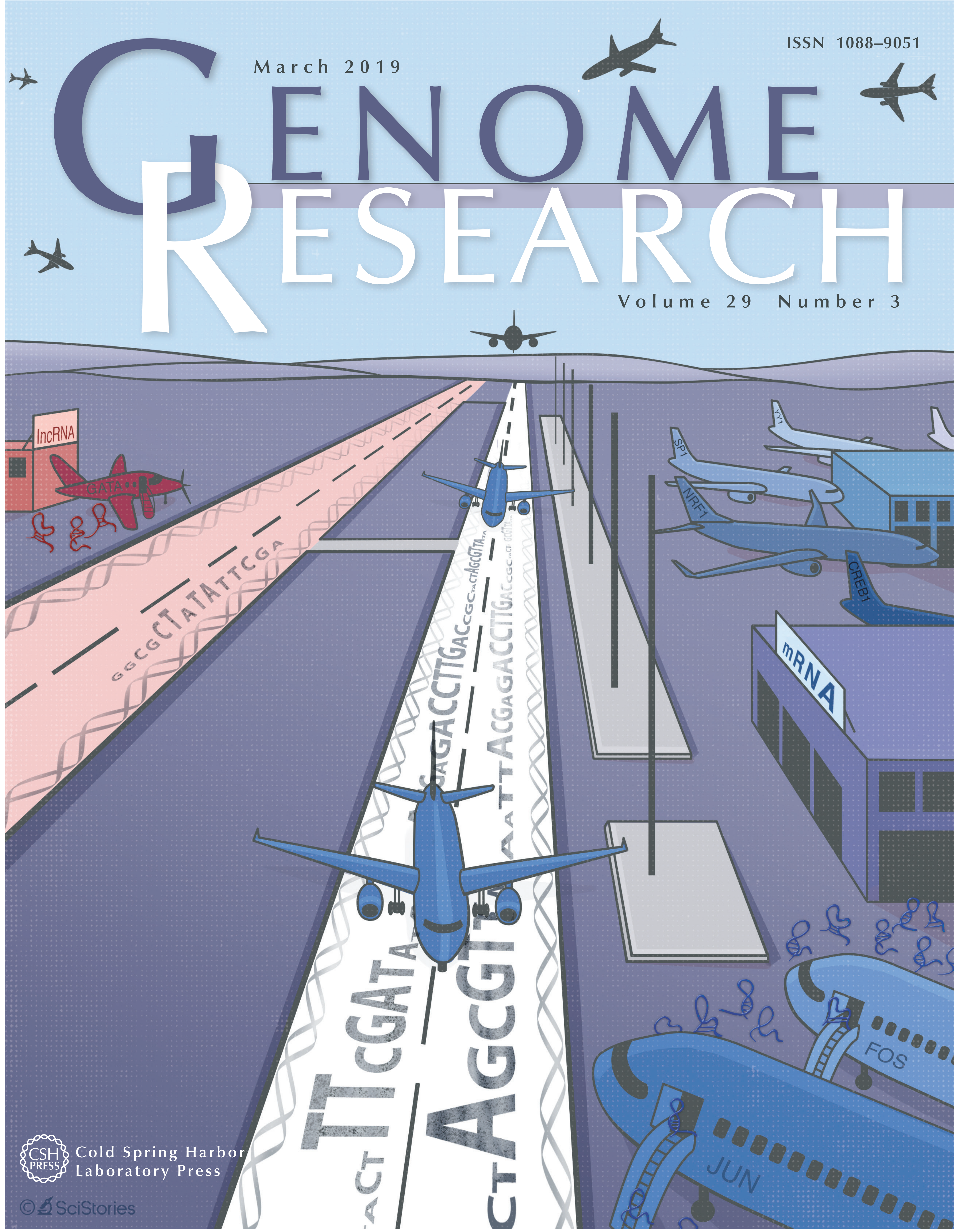
tissue 2

**mRNA**

**ubiquitous expression**

**lncRNA**

**tissue-specific expression**



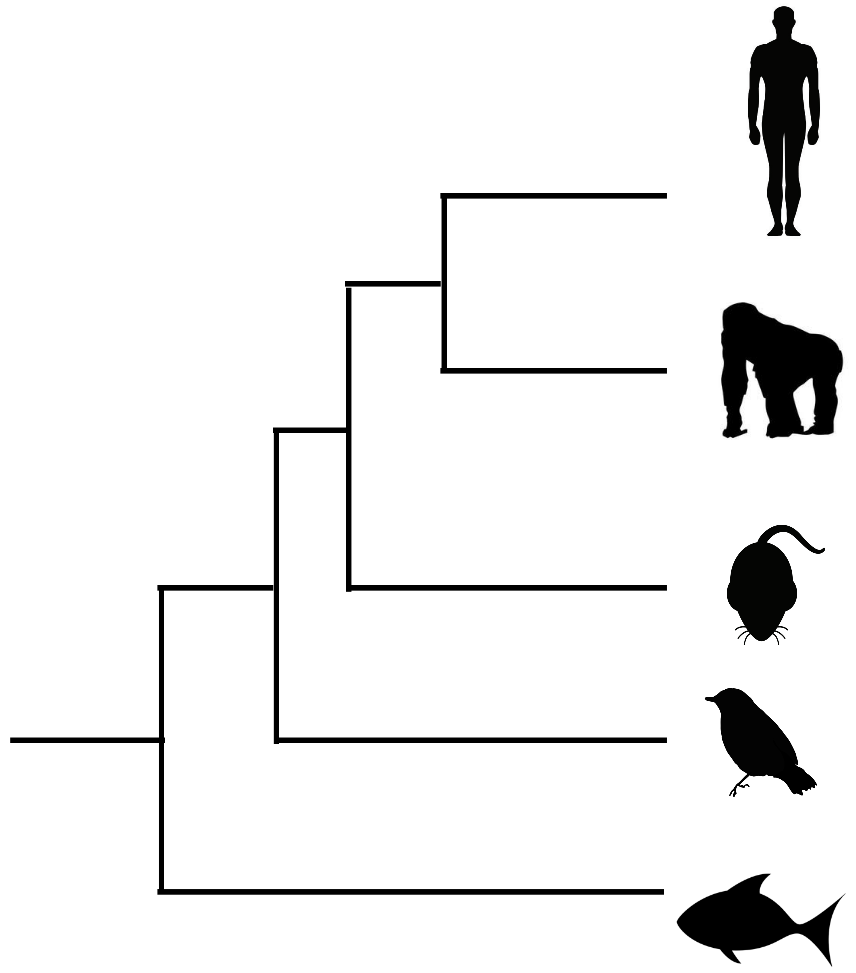
artwork by
Sigrid Knemeyer

# Studying lncRNA biology *en masse*

3

**massively parallel dissection of lncRNA evolution**

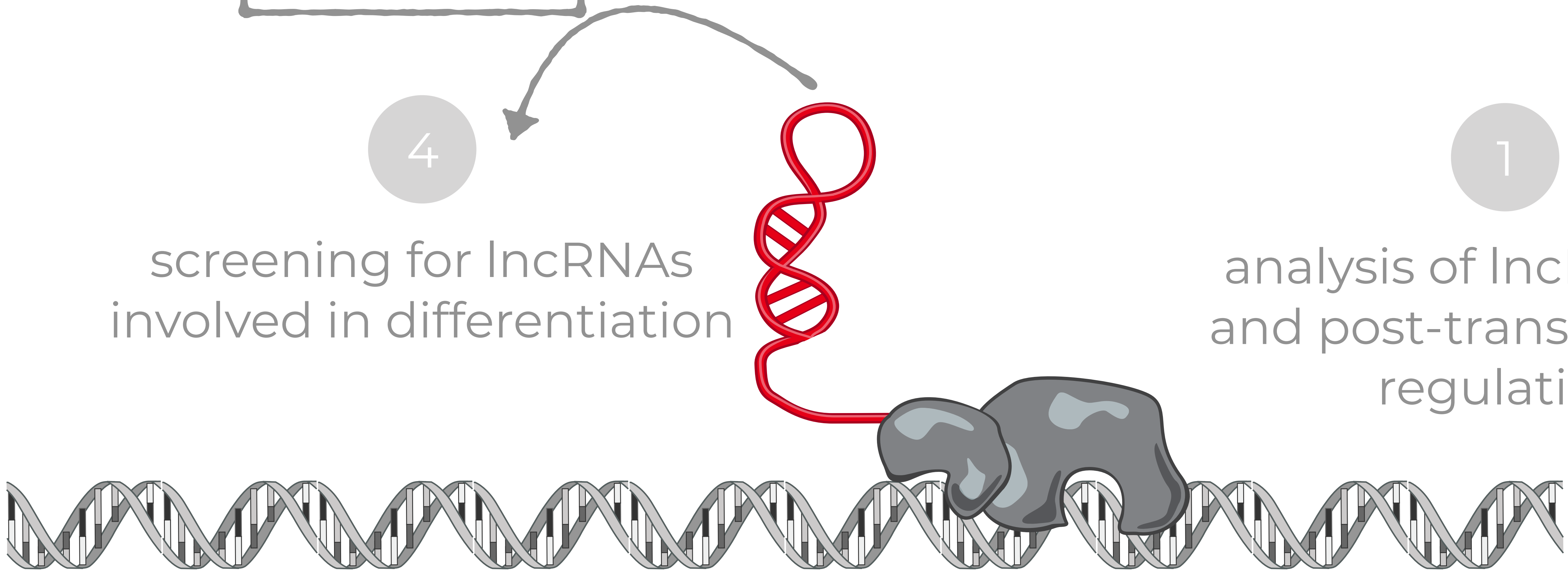
**evolution**



**function**

4

screening for lncRNAs involved in differentiation



1

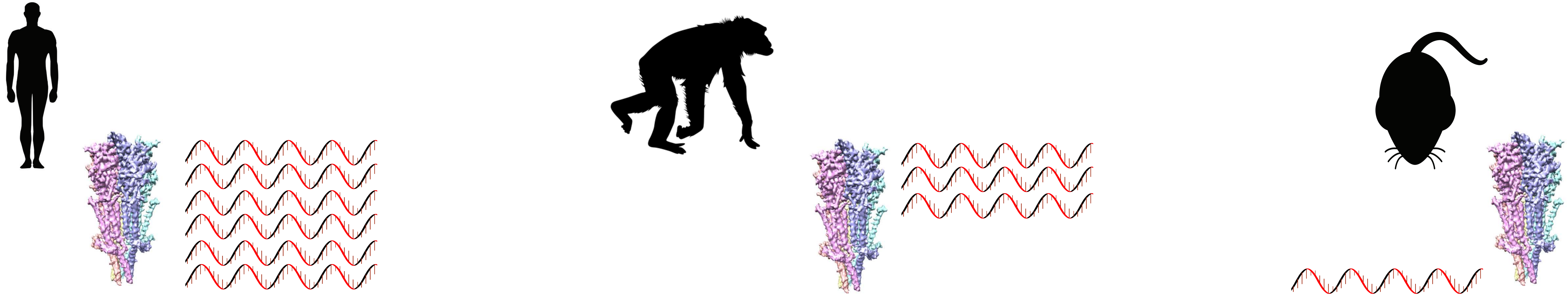
analysis of lncRNA pre- and post-transcriptional regulation

**regulation**

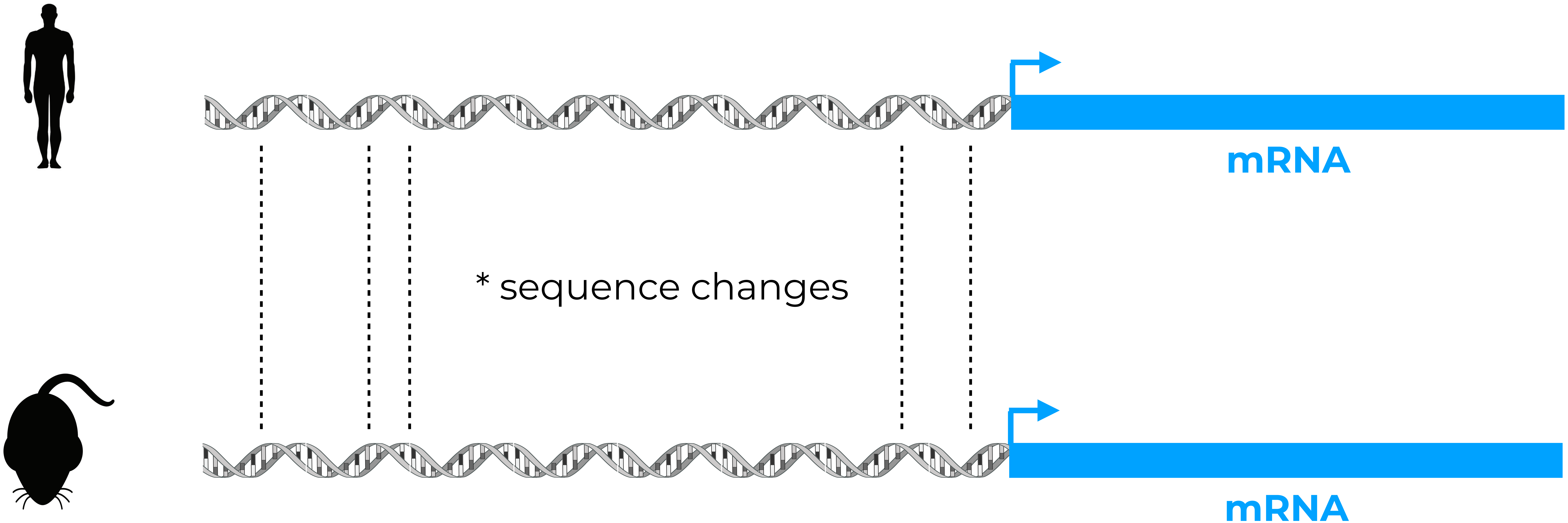
2

massively parallel characterization of lncRNA tissue-specificity

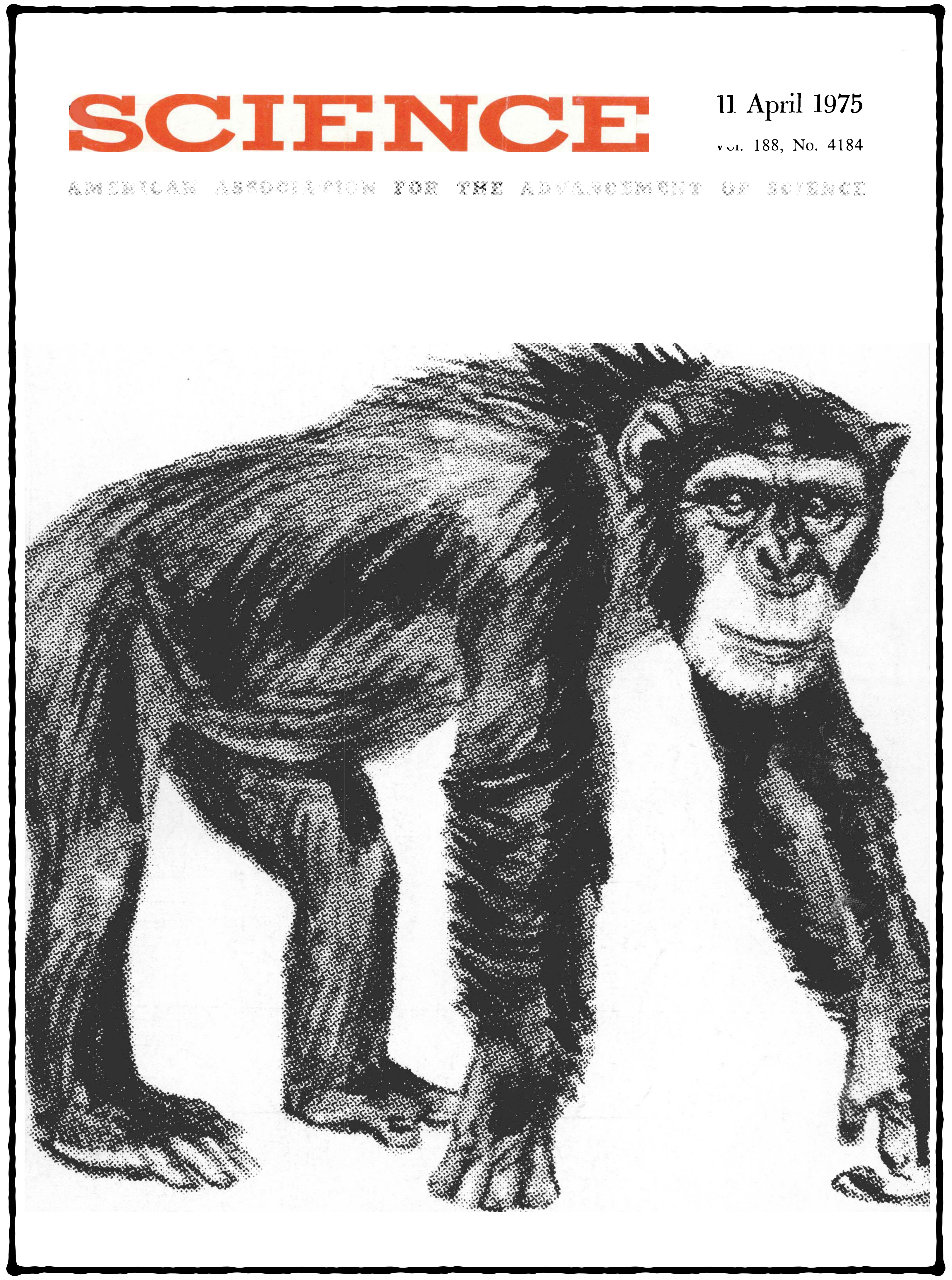
# Different species have different gene expression patterns



# Different species have different non-coding DNA



# Different species have different non-coding DNA



SCIENCE

11 April 1975

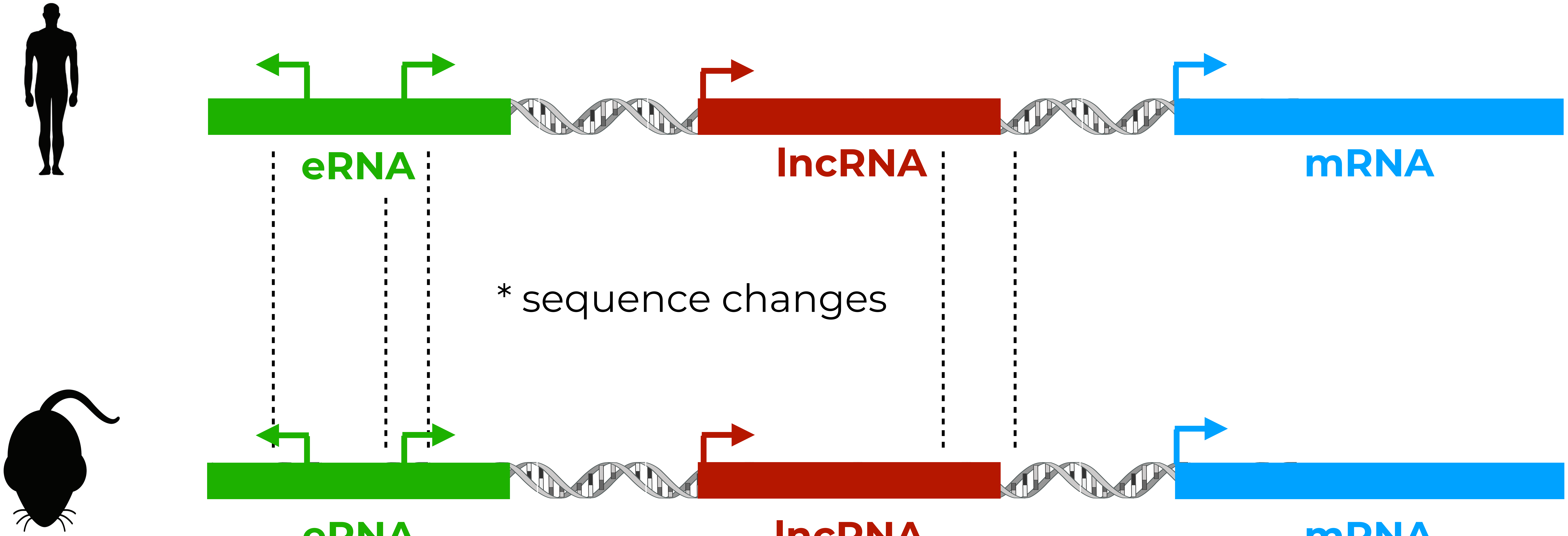
Vol. 188, No. 4184

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

**Evolution at Two Levels in Humans and Chimpanzees**

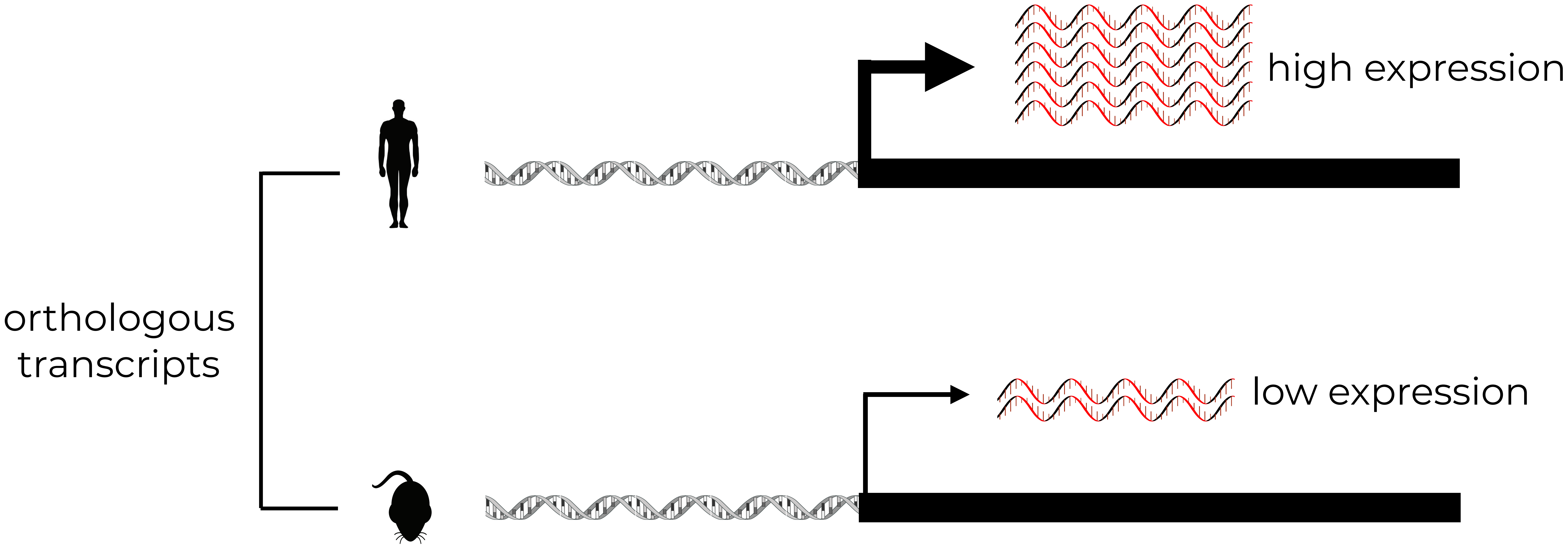
Mary-Claire King and A. C. Wilson

# Different species have different non-coding DNA

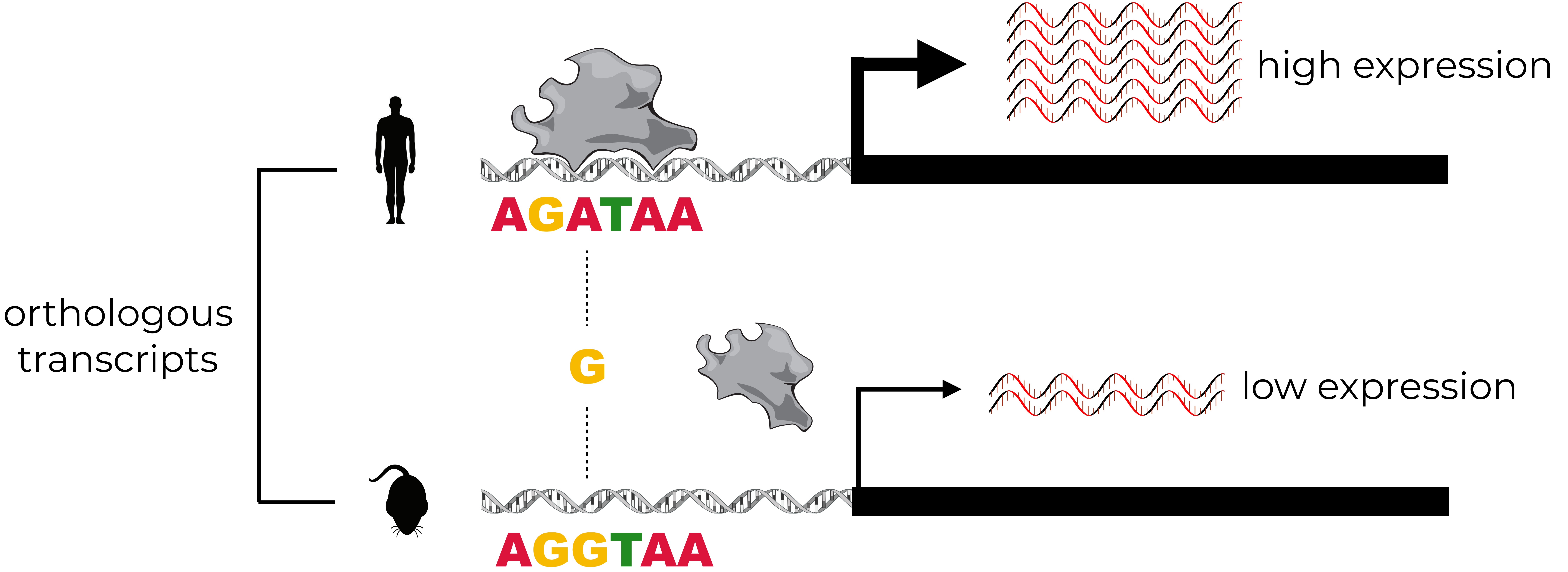


**how do sequence changes in each of these biotypes contribute to gene expression evolution?**

# Mechanisms of gene expression evolution



# *Cis* effects are proximal sequence changes to the transcripts they affect



high expression

orthologous transcripts

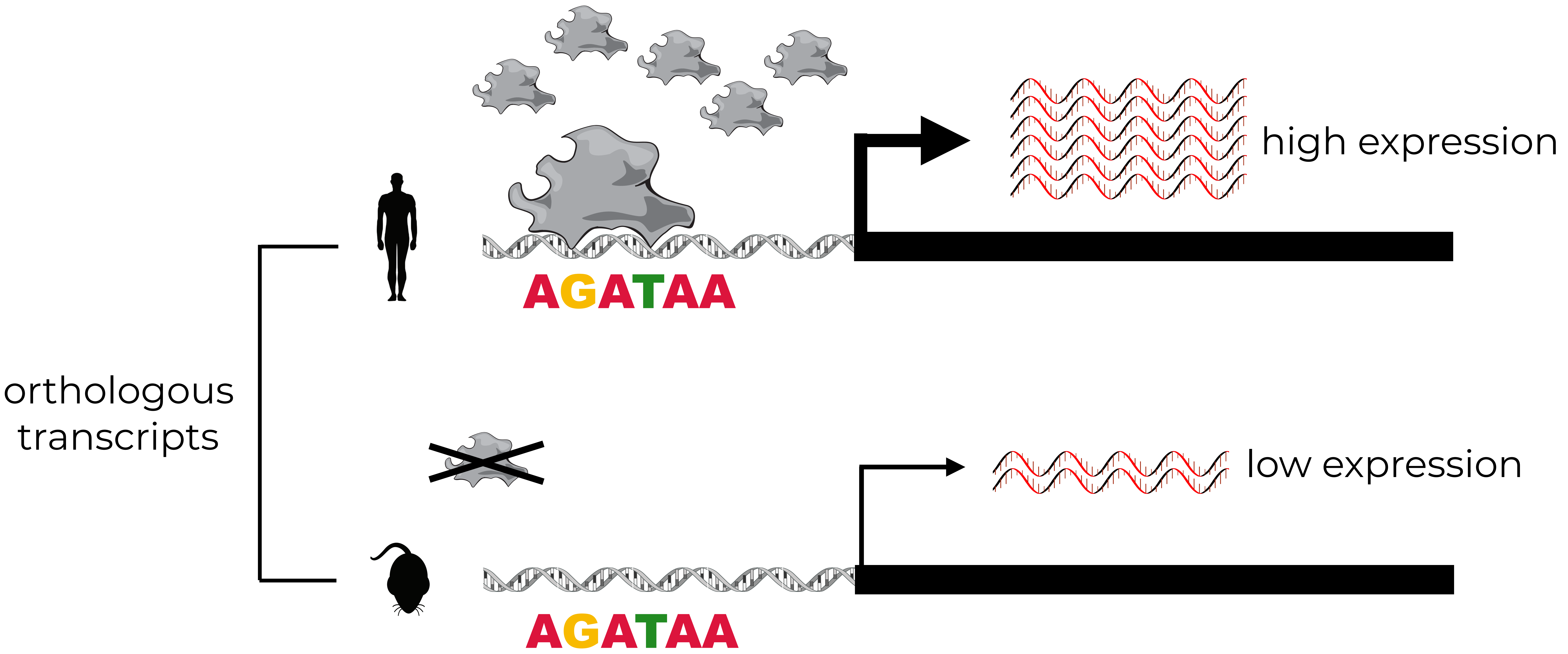
AGATAA

G

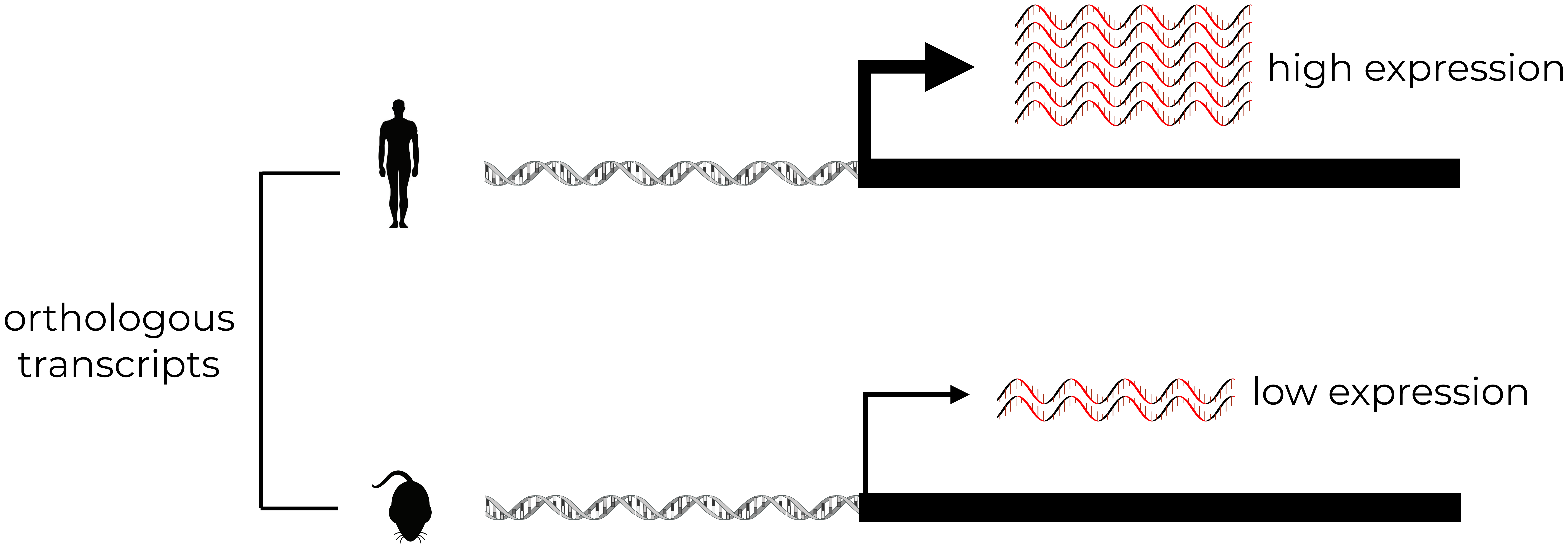
low expression

AGGTAA

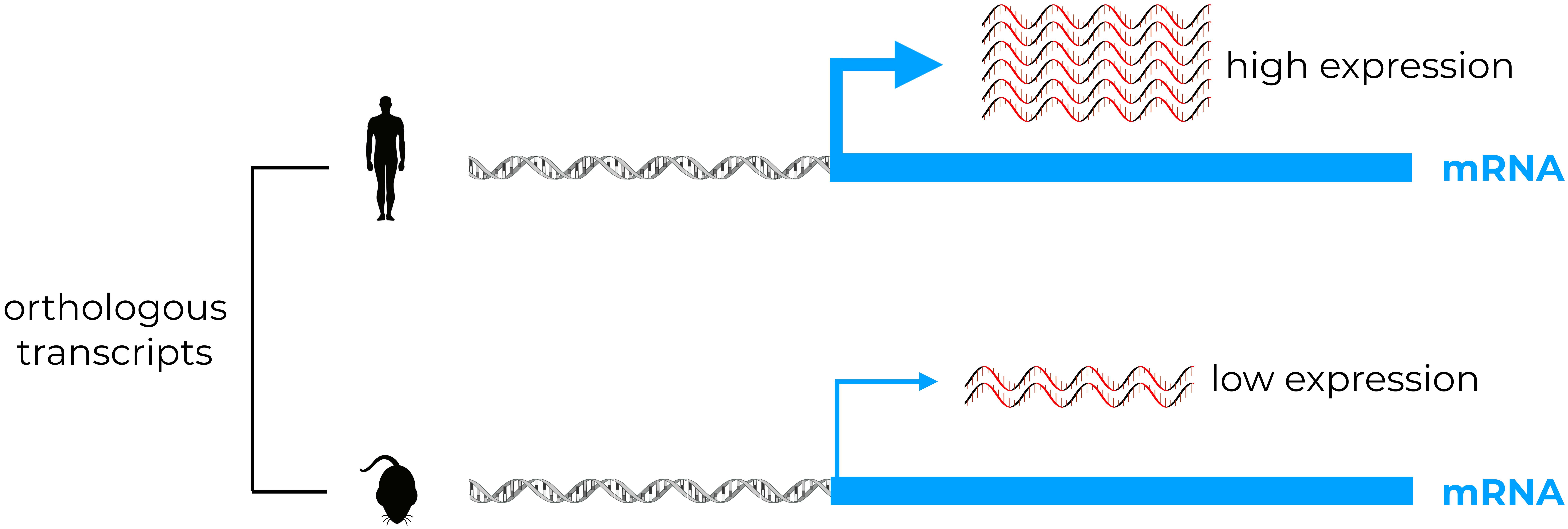
# *Trans* effects are other, distal changes that affect the cellular environment



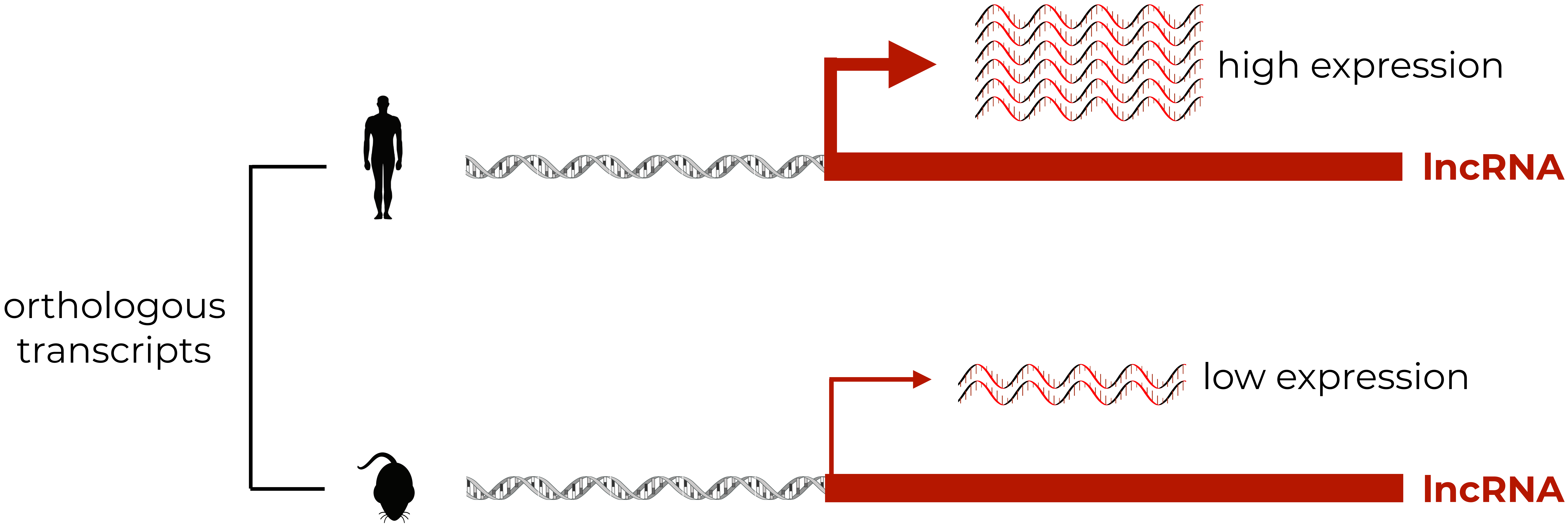
# How do *cis* and *trans* effects differ across biotypes?



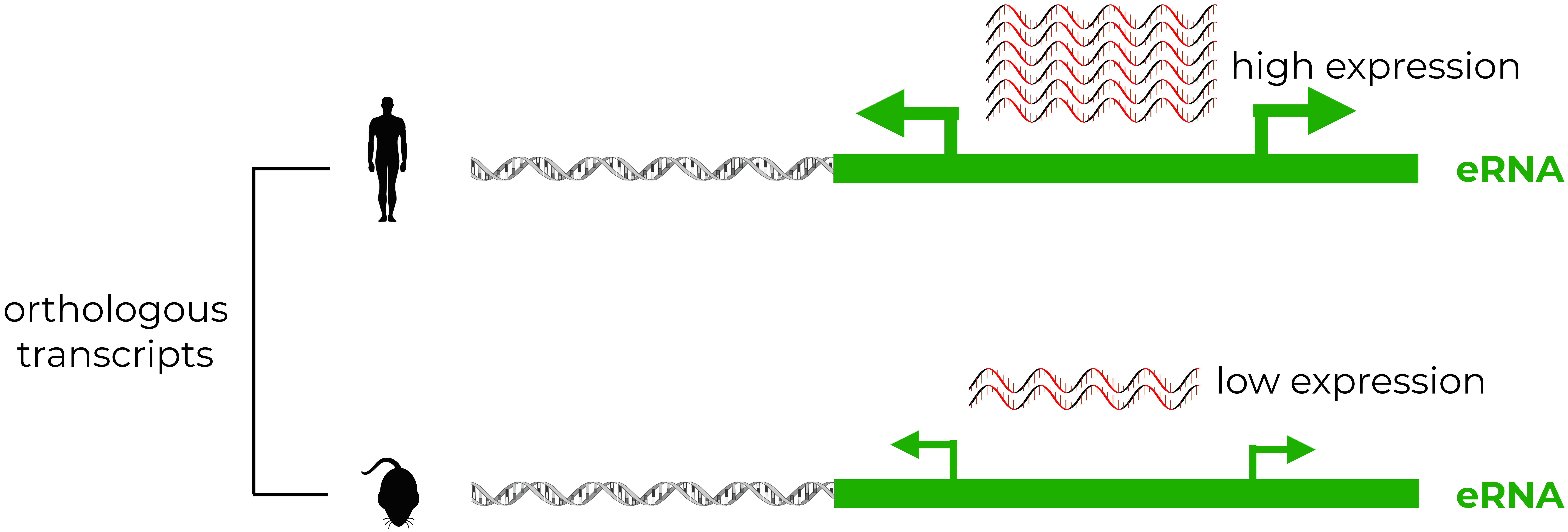
# How do *cis* and *trans* effects differ across biotypes?



# How do *cis* and *trans* effects differ across biotypes?



# How do *cis* and *trans* effects differ across biotypes?



# MPRA to *cis* and *trans* effects between human and mouse in embryonic stem cells

orthologous
core promoters

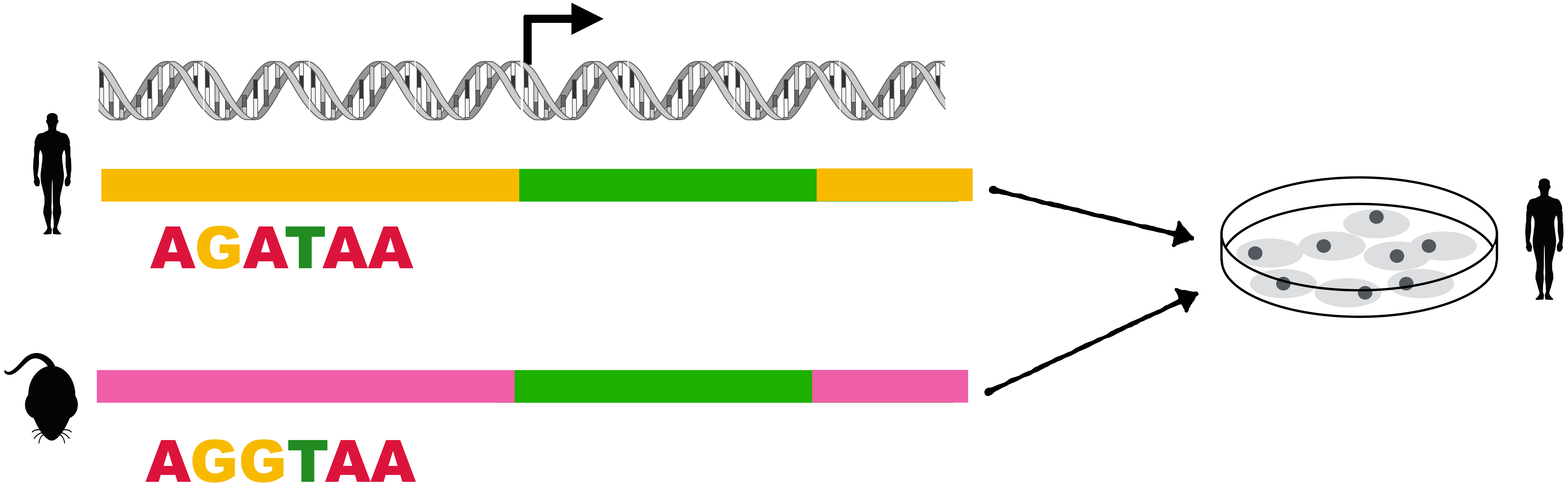
eRNAs
lncRNAs
mRNAs

> 3,000 total pairs

**human ESCs**

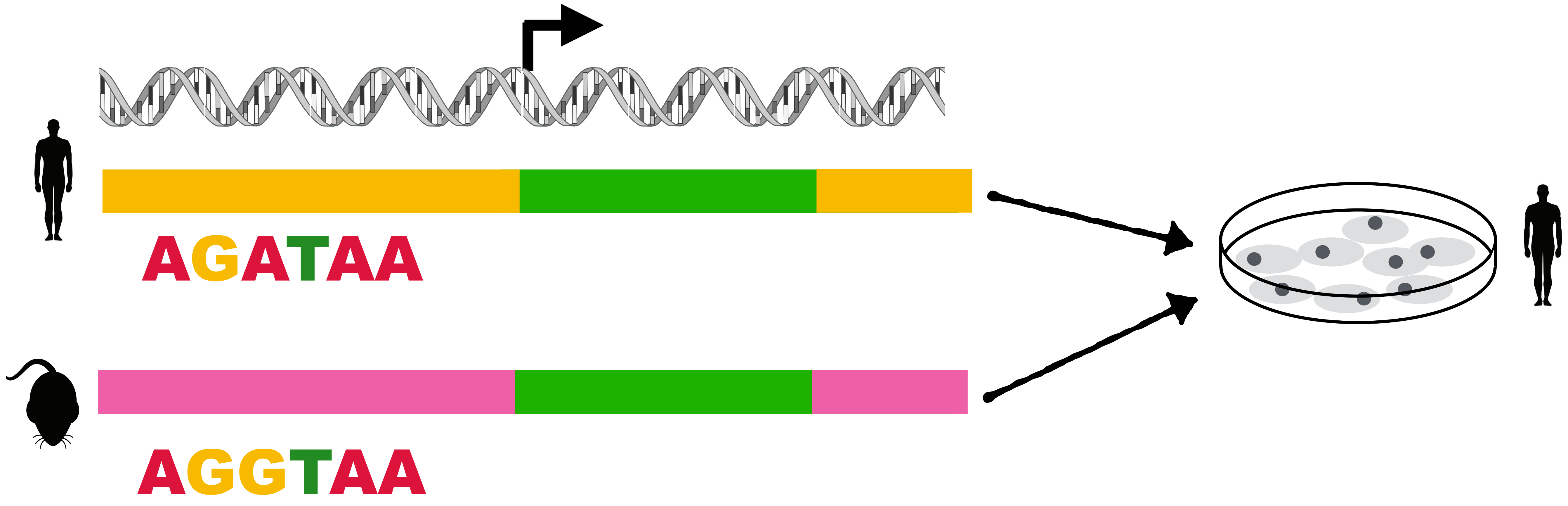
**mouse ESCs**

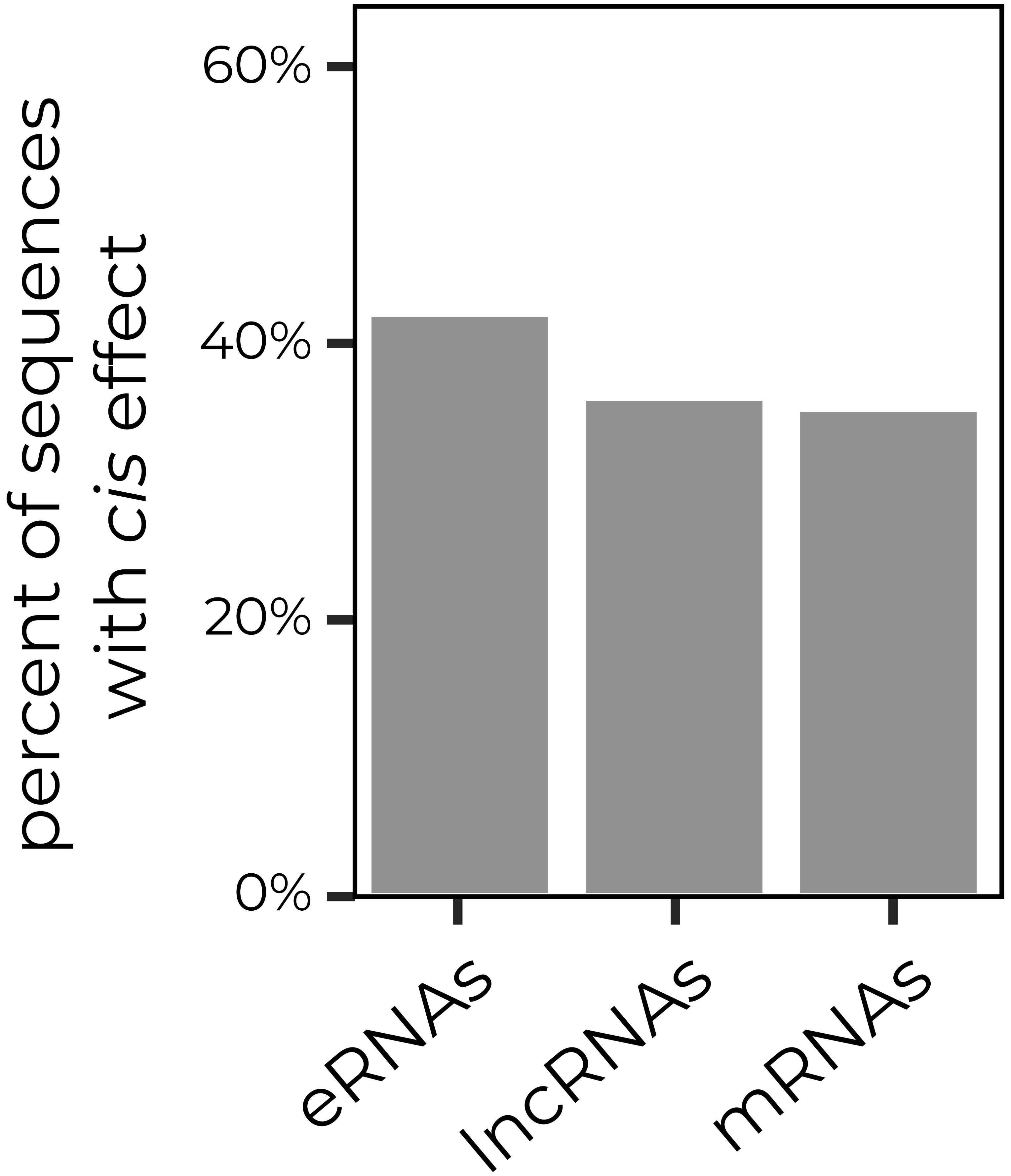
# Measuring *cis* effects in MPRA



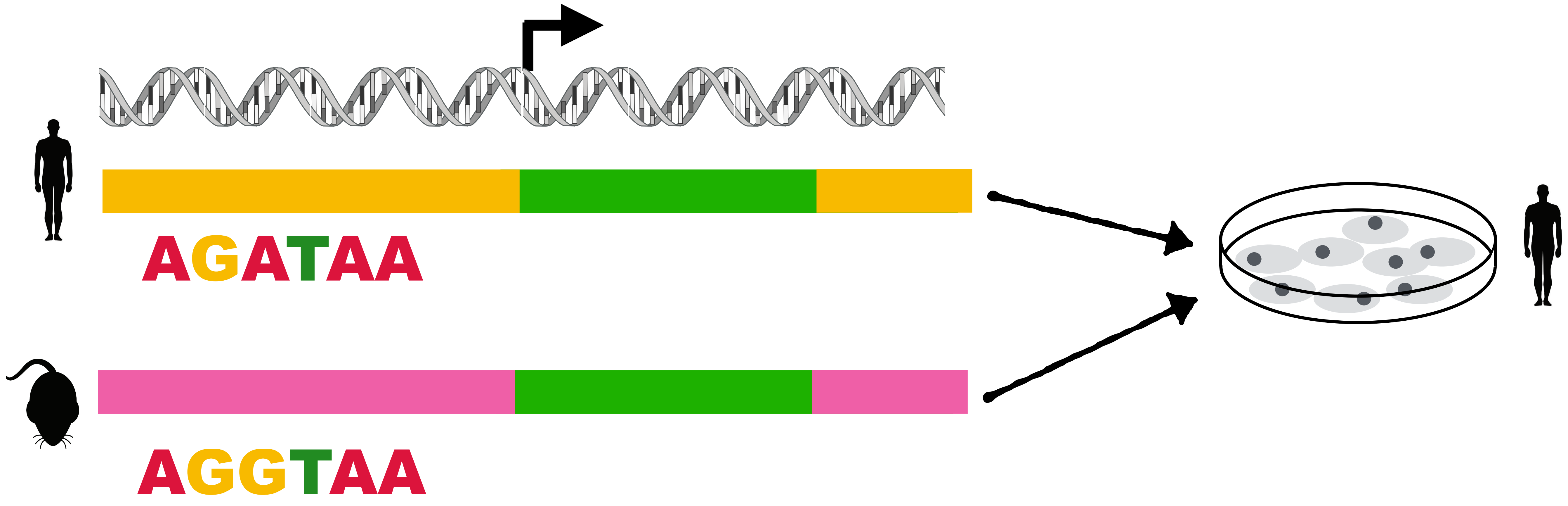
***cis* effects:** compare activities of core promoters from two species

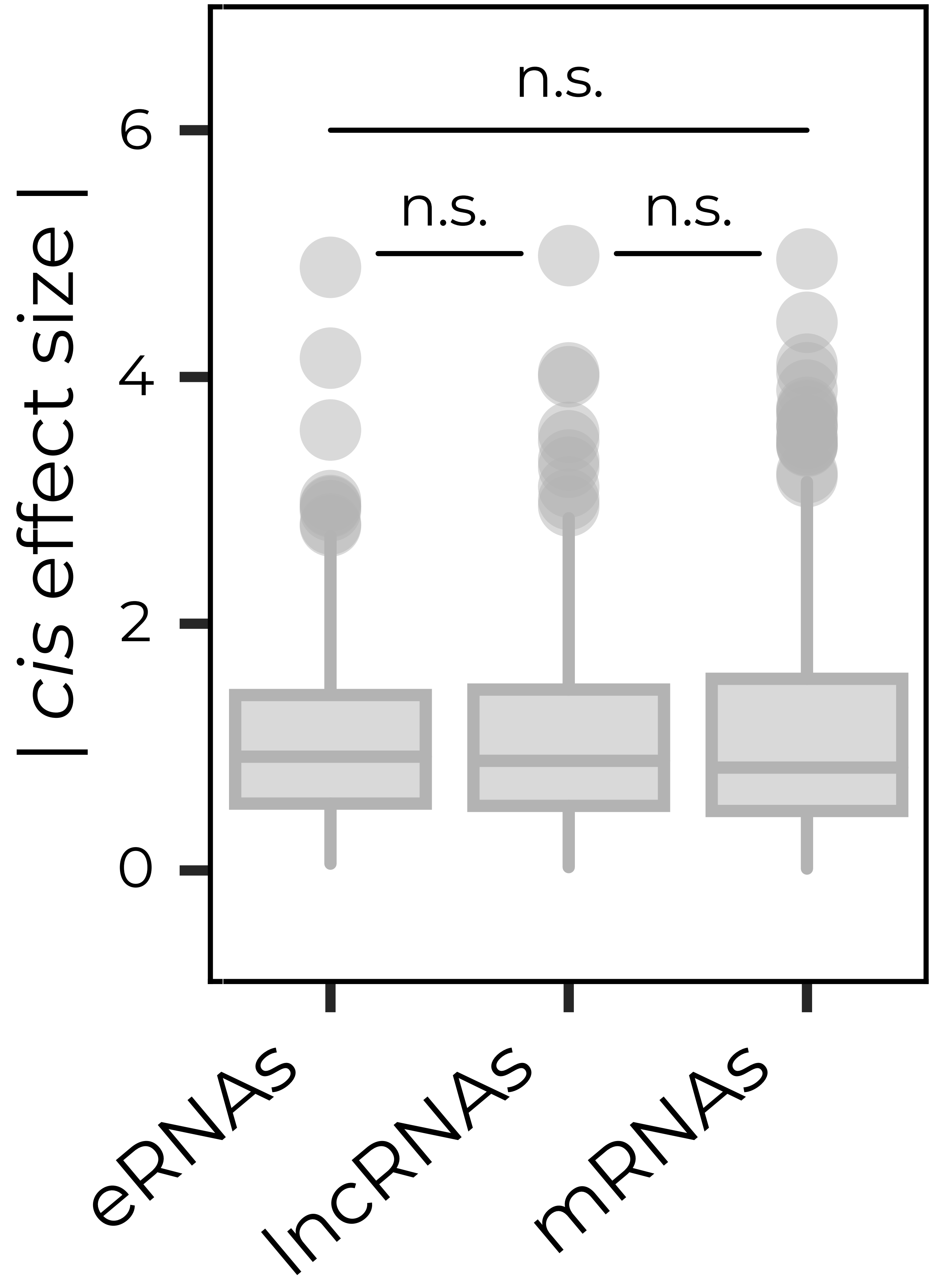
# *Cis* effects are common



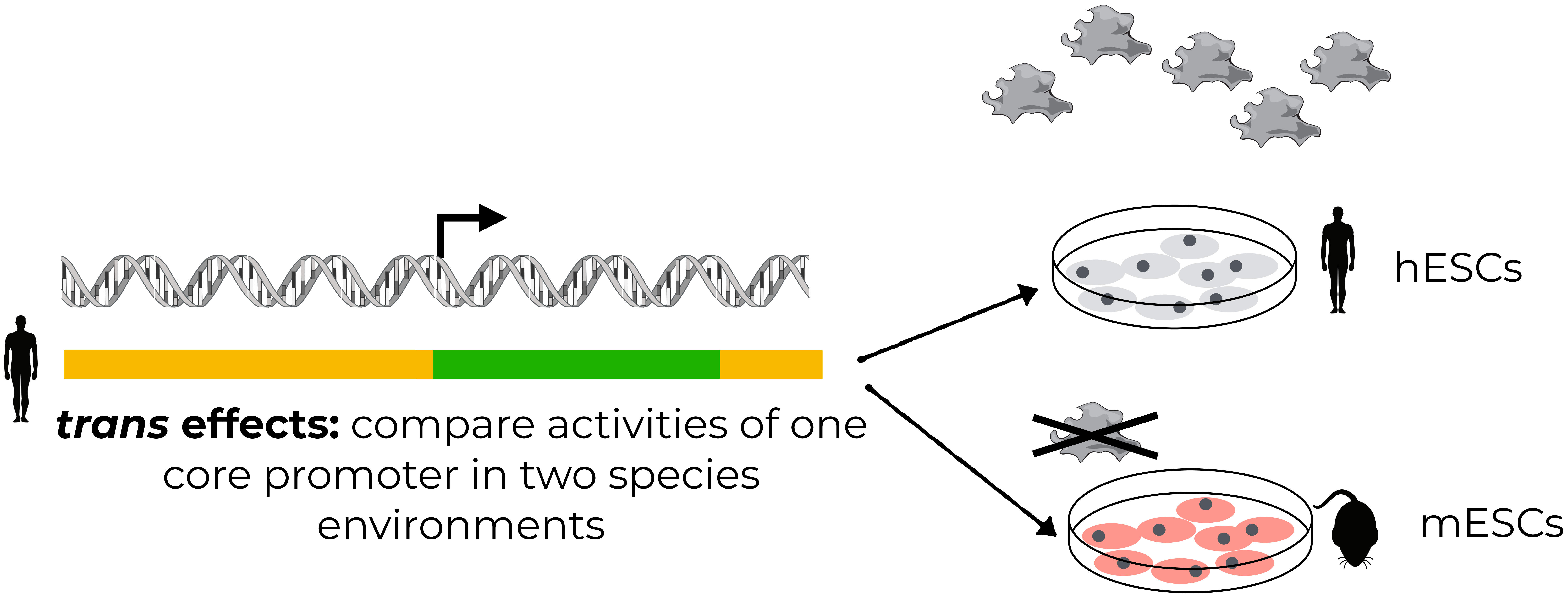


# *Cis* effects are uniform across biotypes



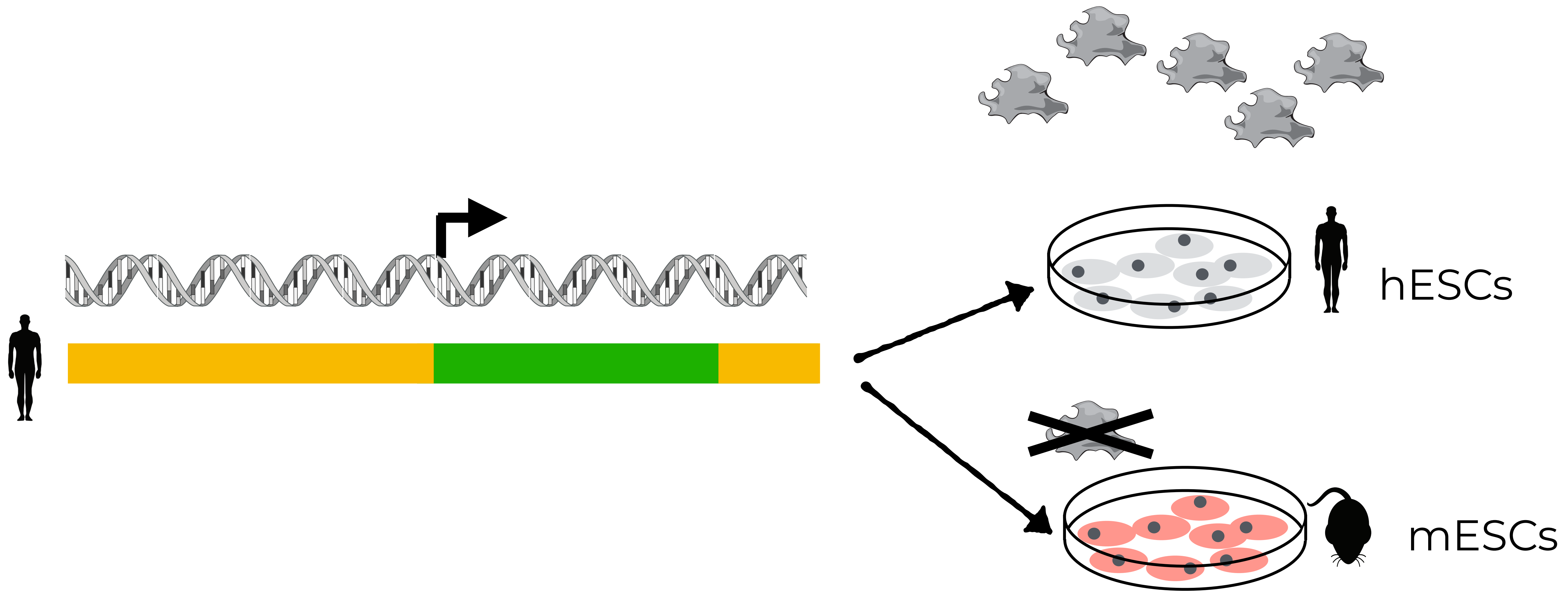


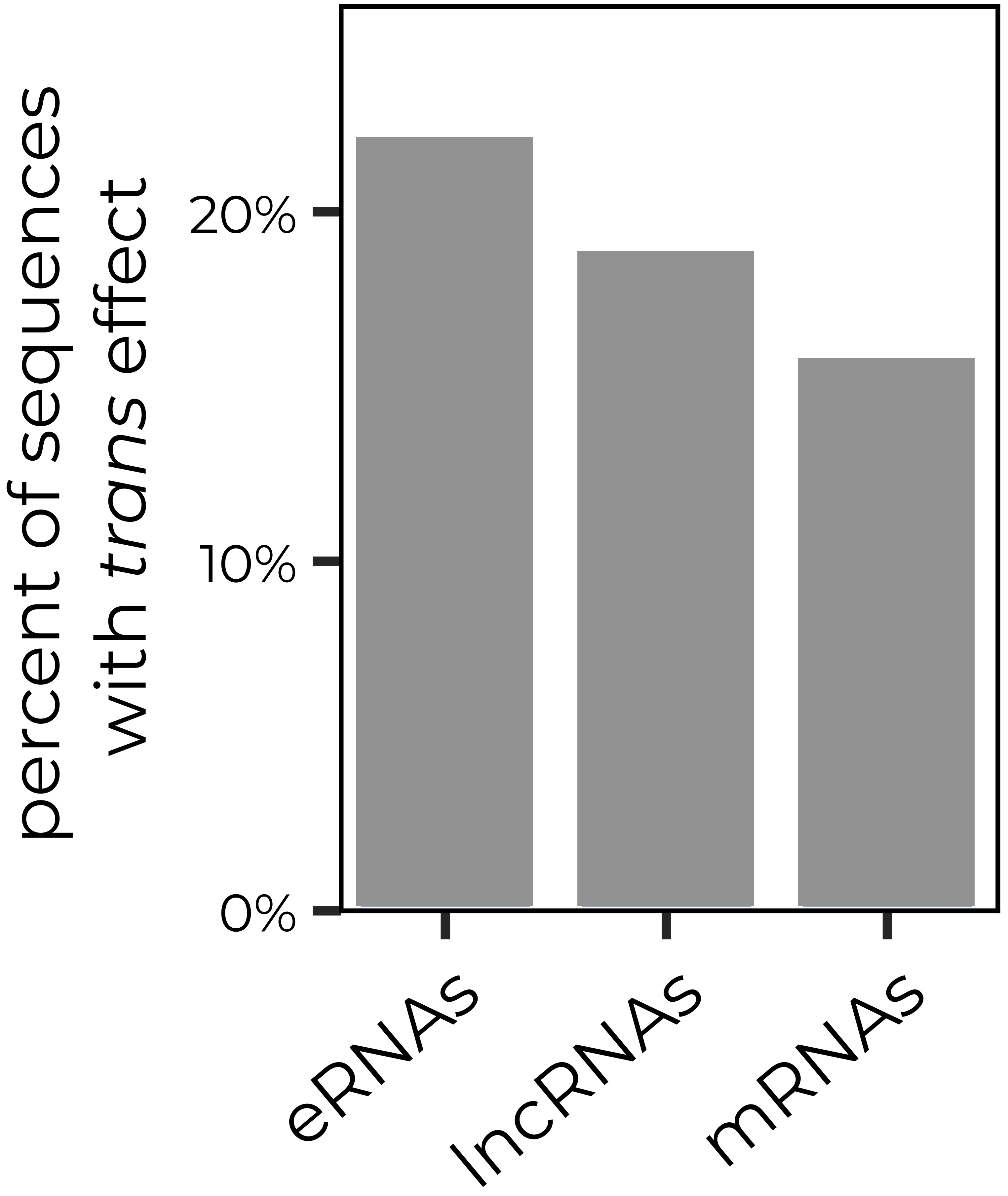
# Measuring *trans* effects in MPRA



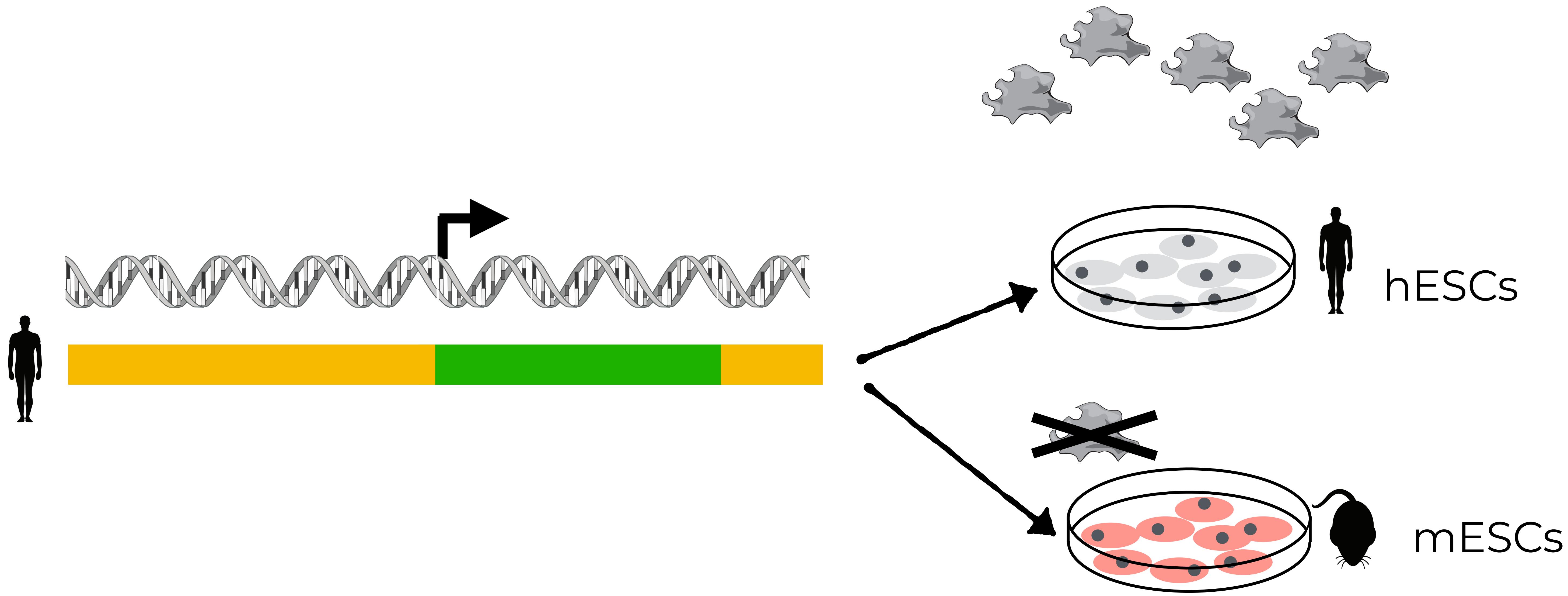
***trans* effects:** compare activities of one core promoter in two species environments

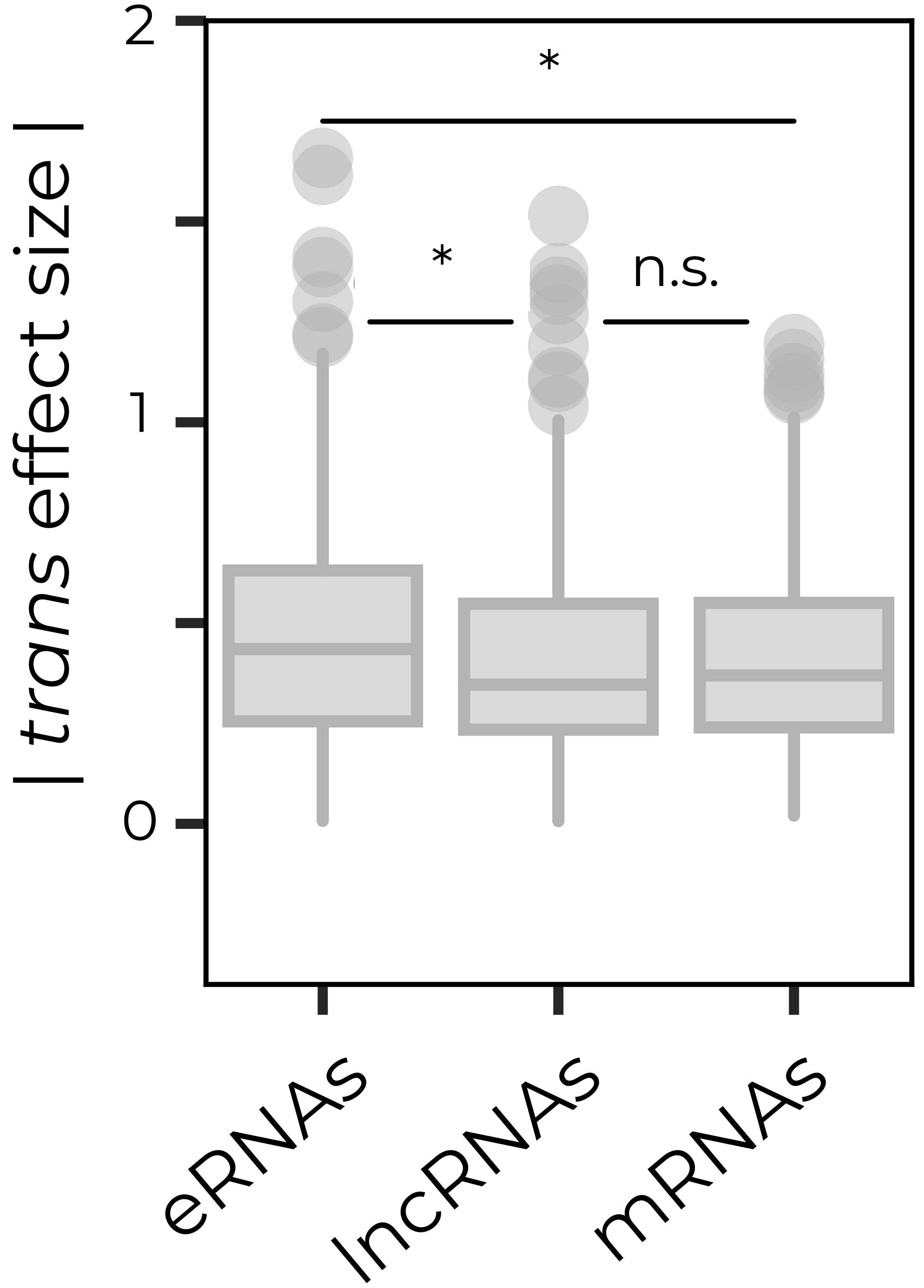
# *Trans* effects are rare



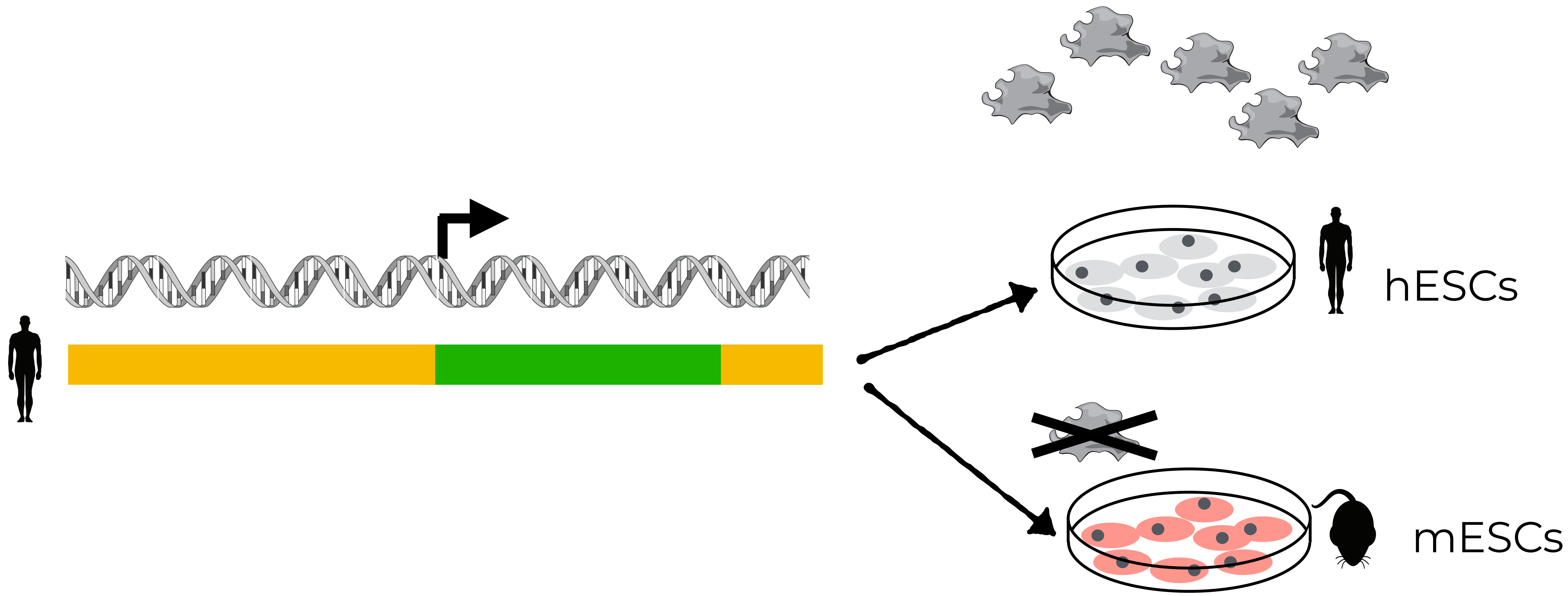


# *Trans* effects are highest in eRNAs





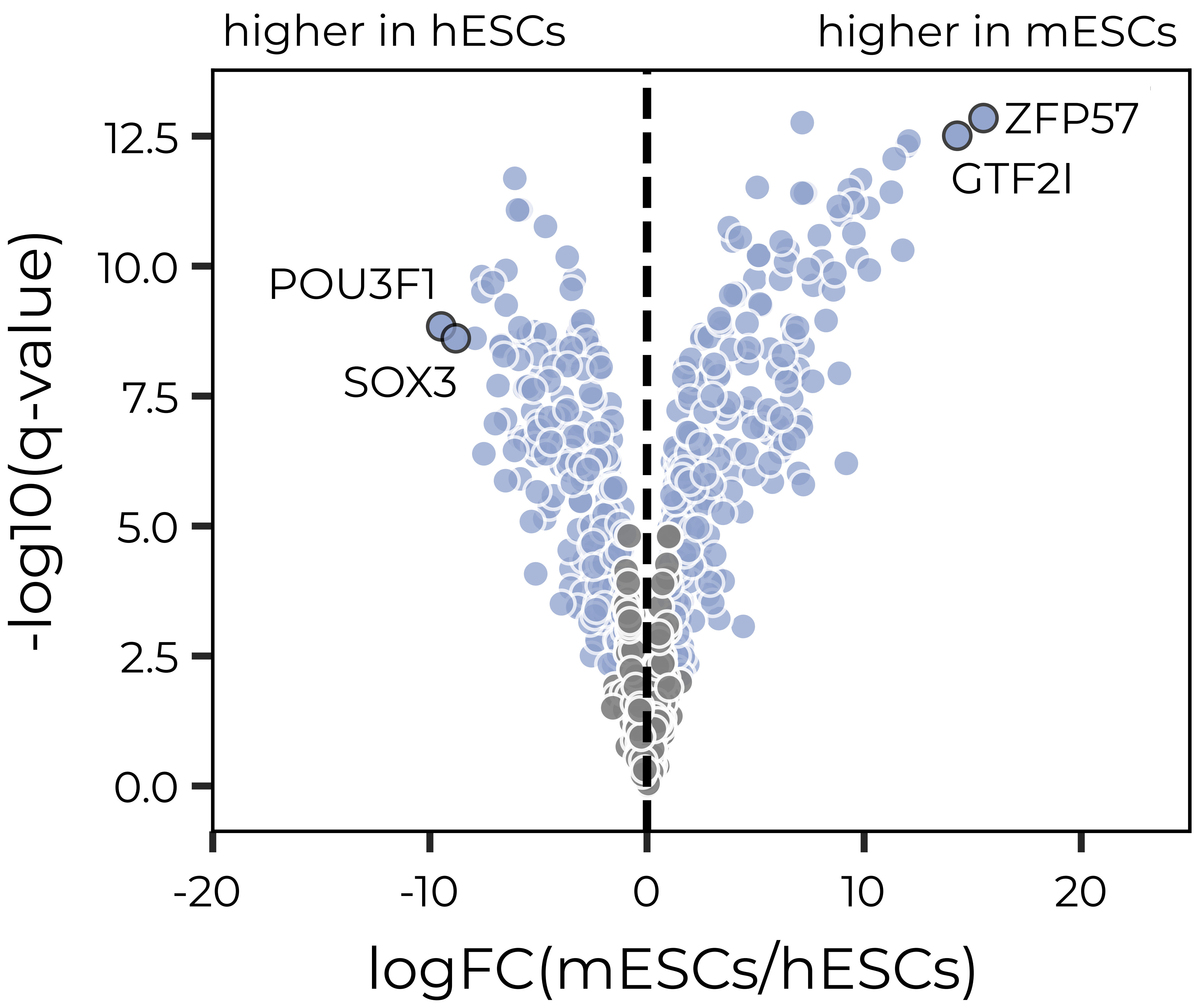
# Which TFs are driving *trans* effects?



perform **RNA-seq** of our hESCs and mESCs

# Many TFs are differentially expressed between hESCs and mESCs

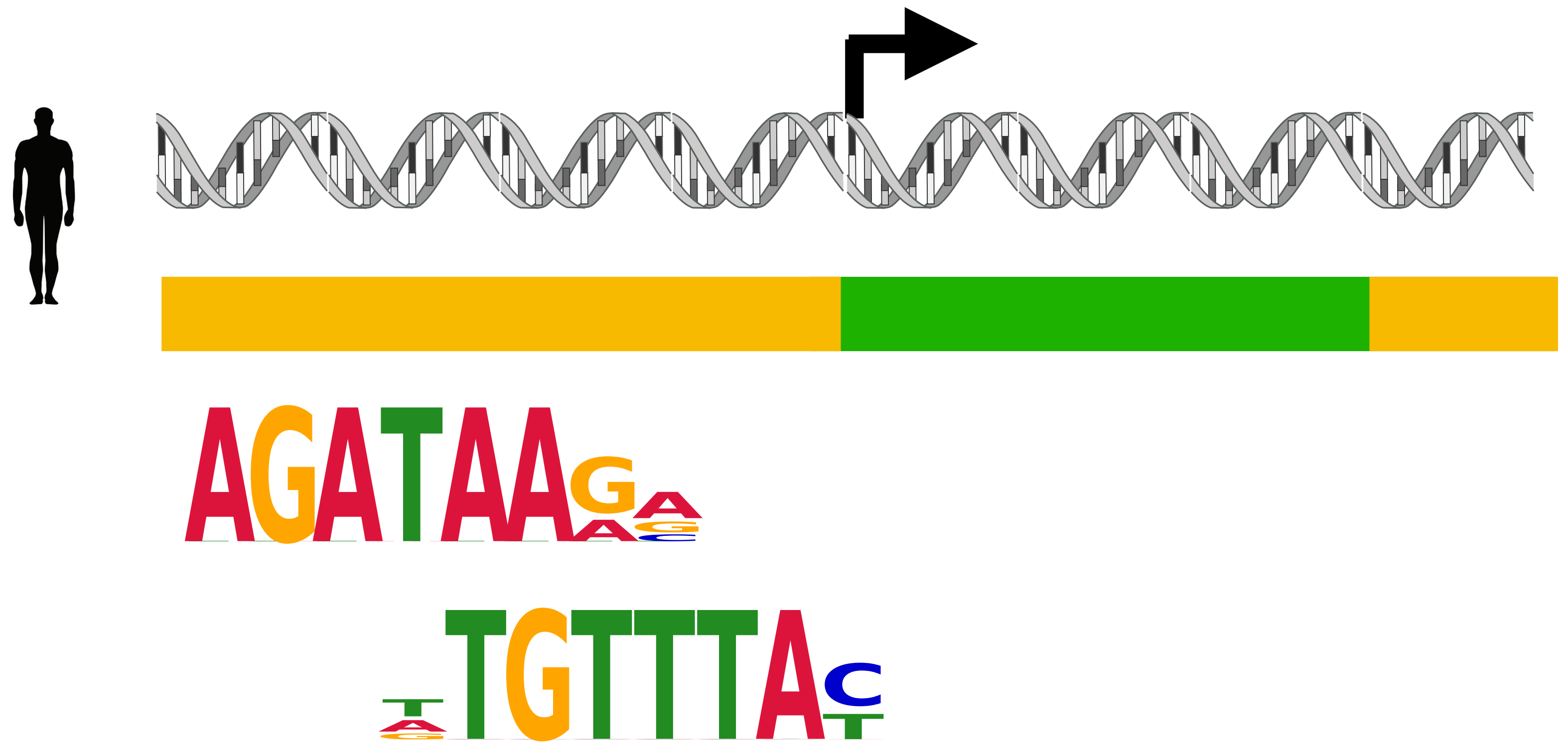
1,032 TFs that are one-to-one orthologs between human and mouse



428 differentially expressed (41%) (FDR < 0.01 and foldchange ≥ 2)

Winona Oliveros

# Finding the TFs that most likely drive *trans* effects

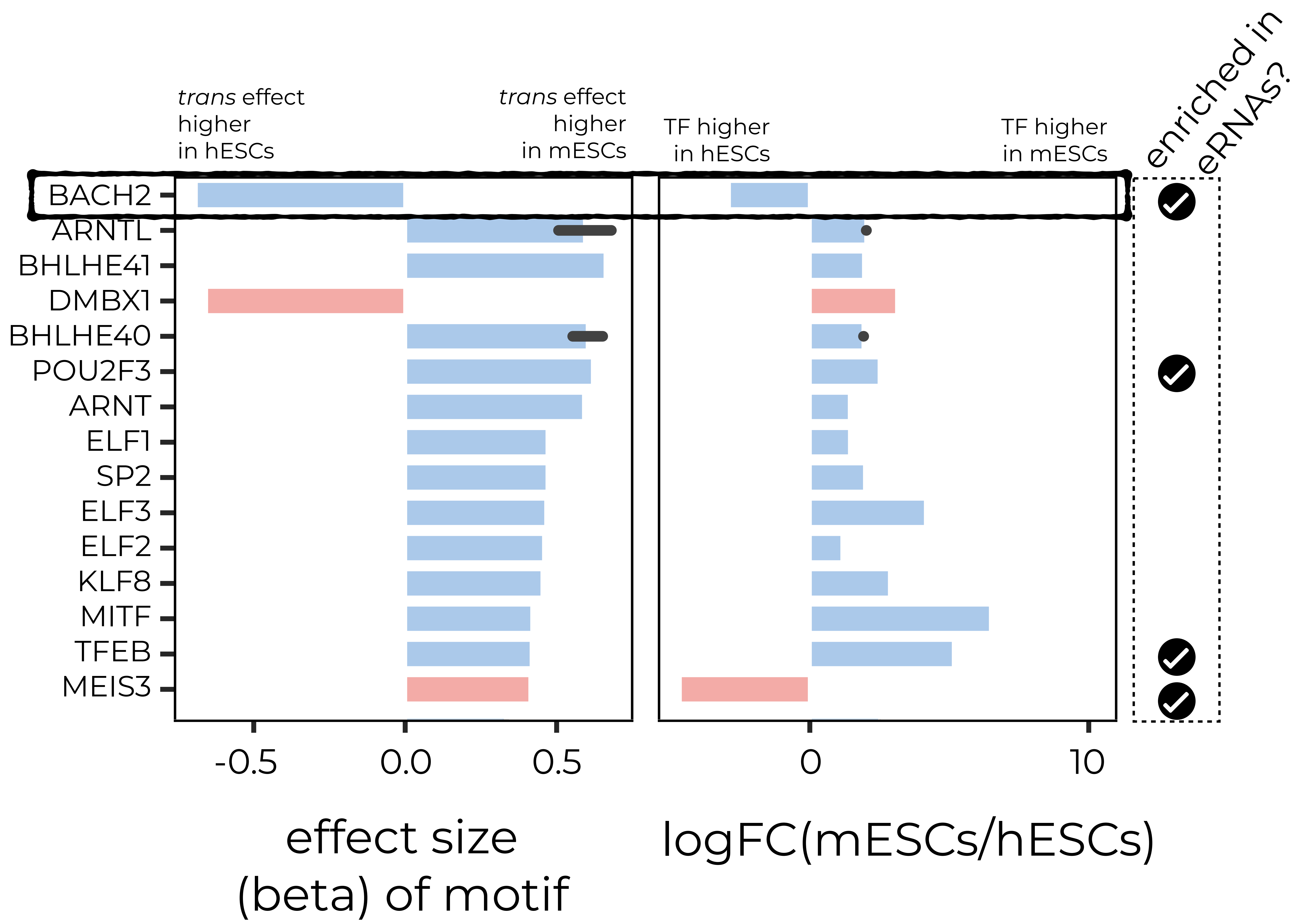


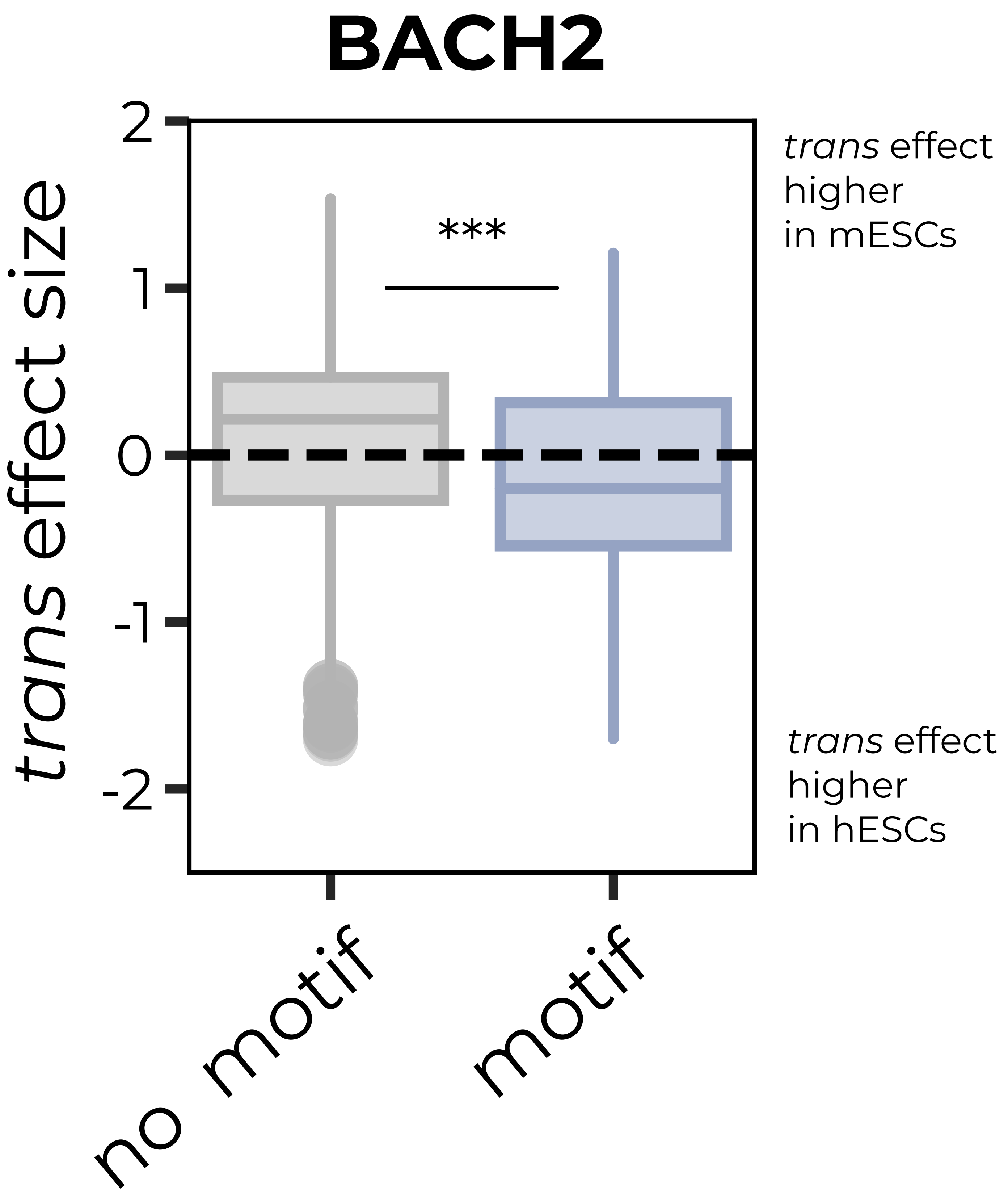
map TF motifs in each sequence

test whether presence of motif is associated with *trans* effects using linear regression

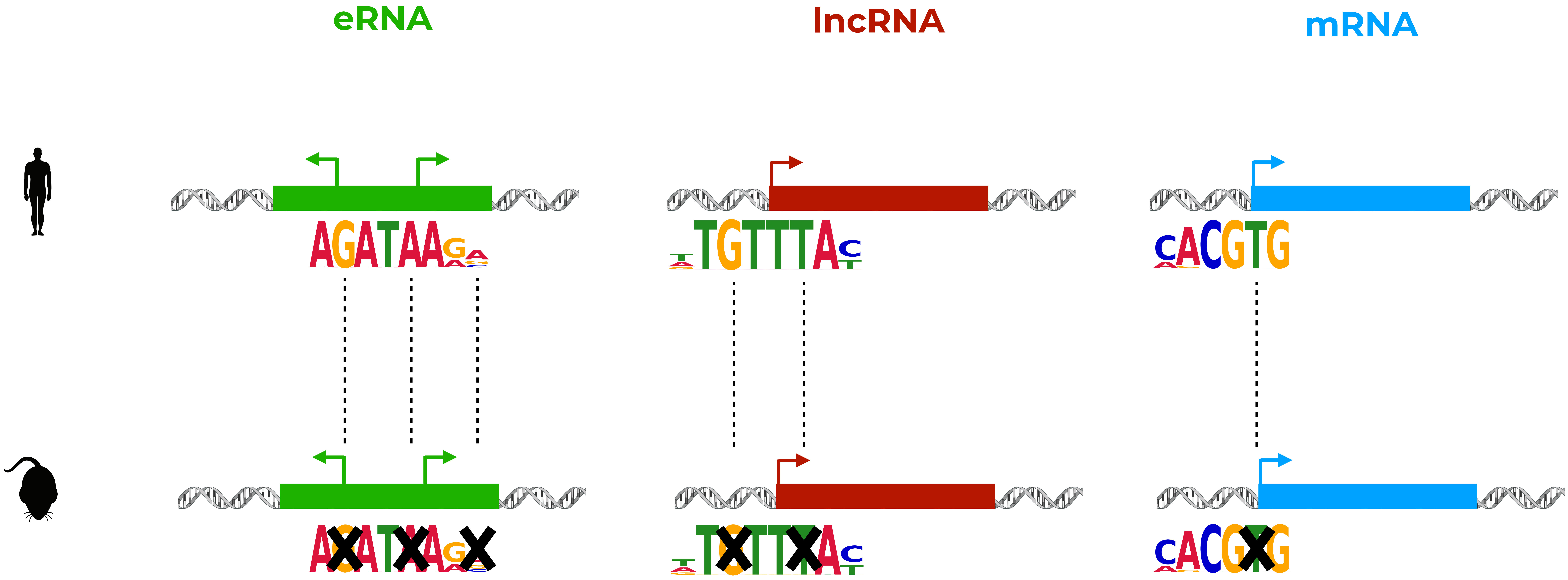
see if trans enrichment directionality matches RNA-seq directionality

# *Trans* effects are driven by a subset of TFs





# *Cis* effects are common across all biotypes

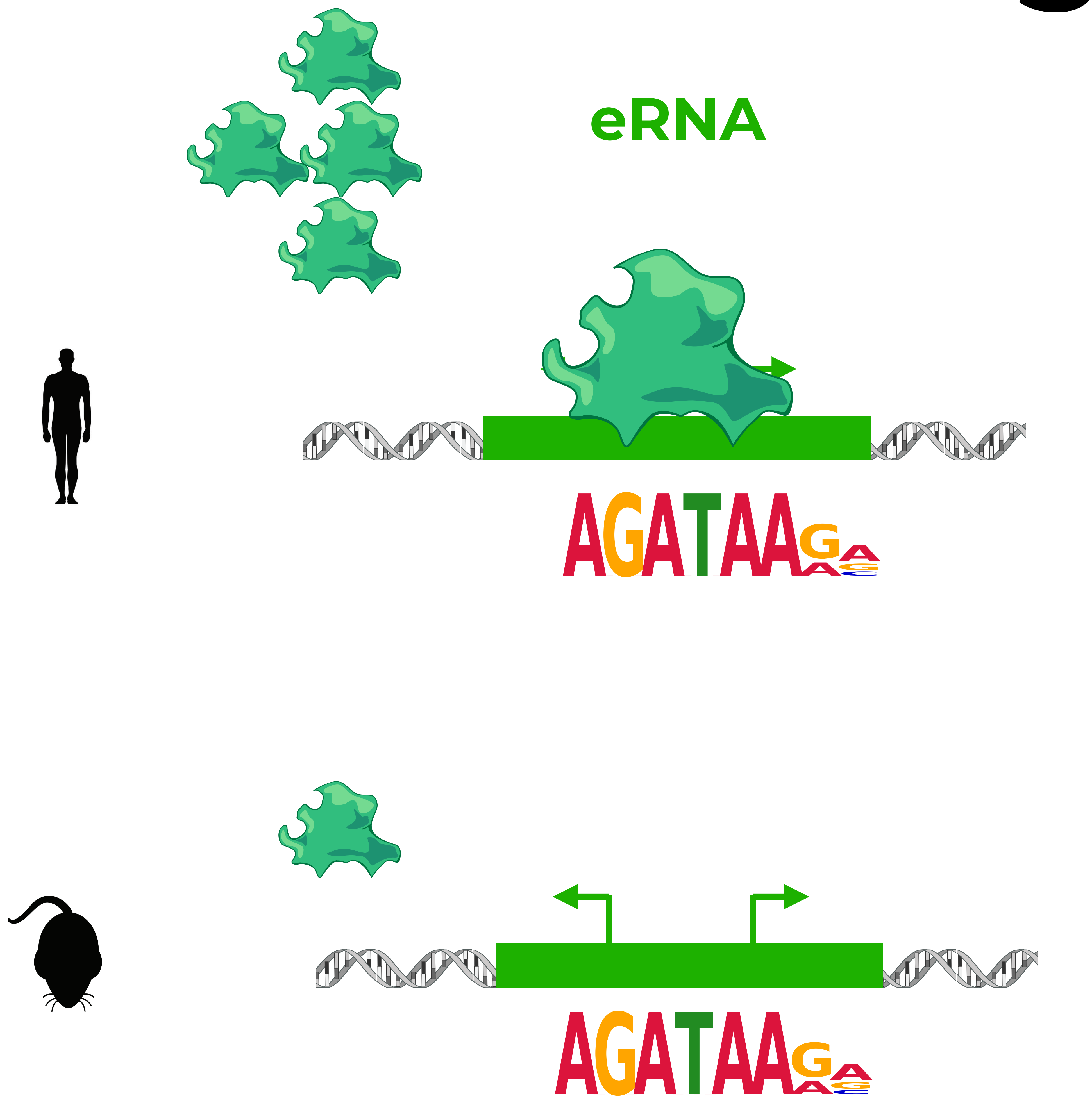


eRNA

lncRNA

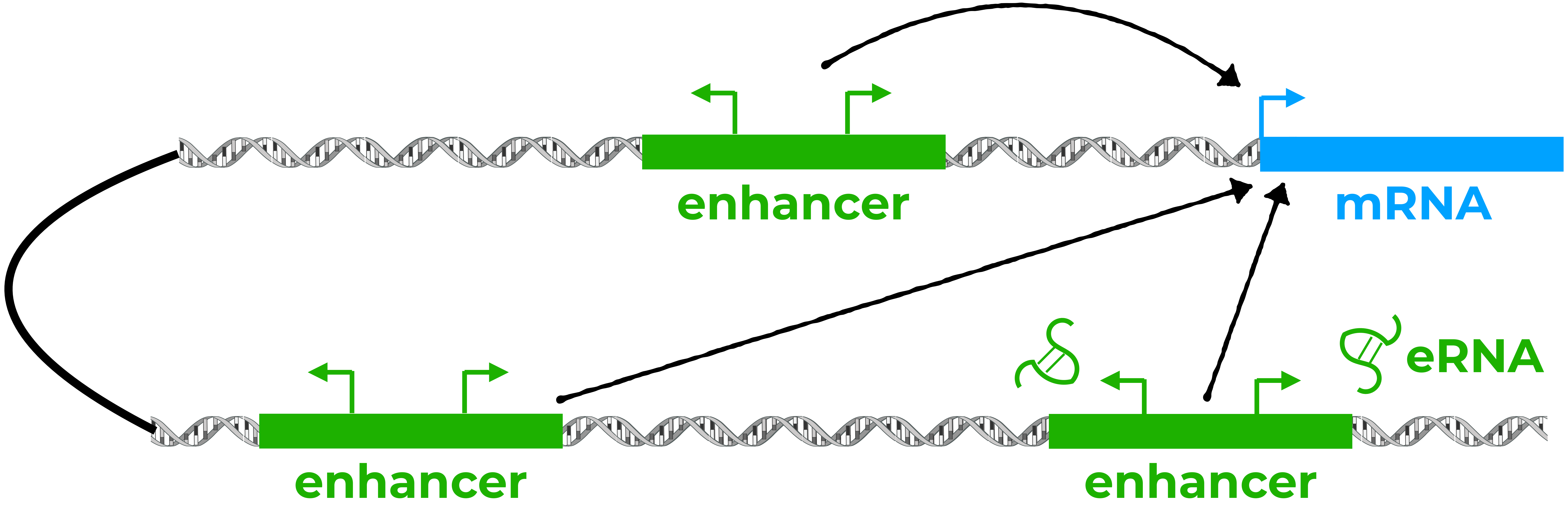
mRNA

# But *trans* effects are specifically high in eRNAs



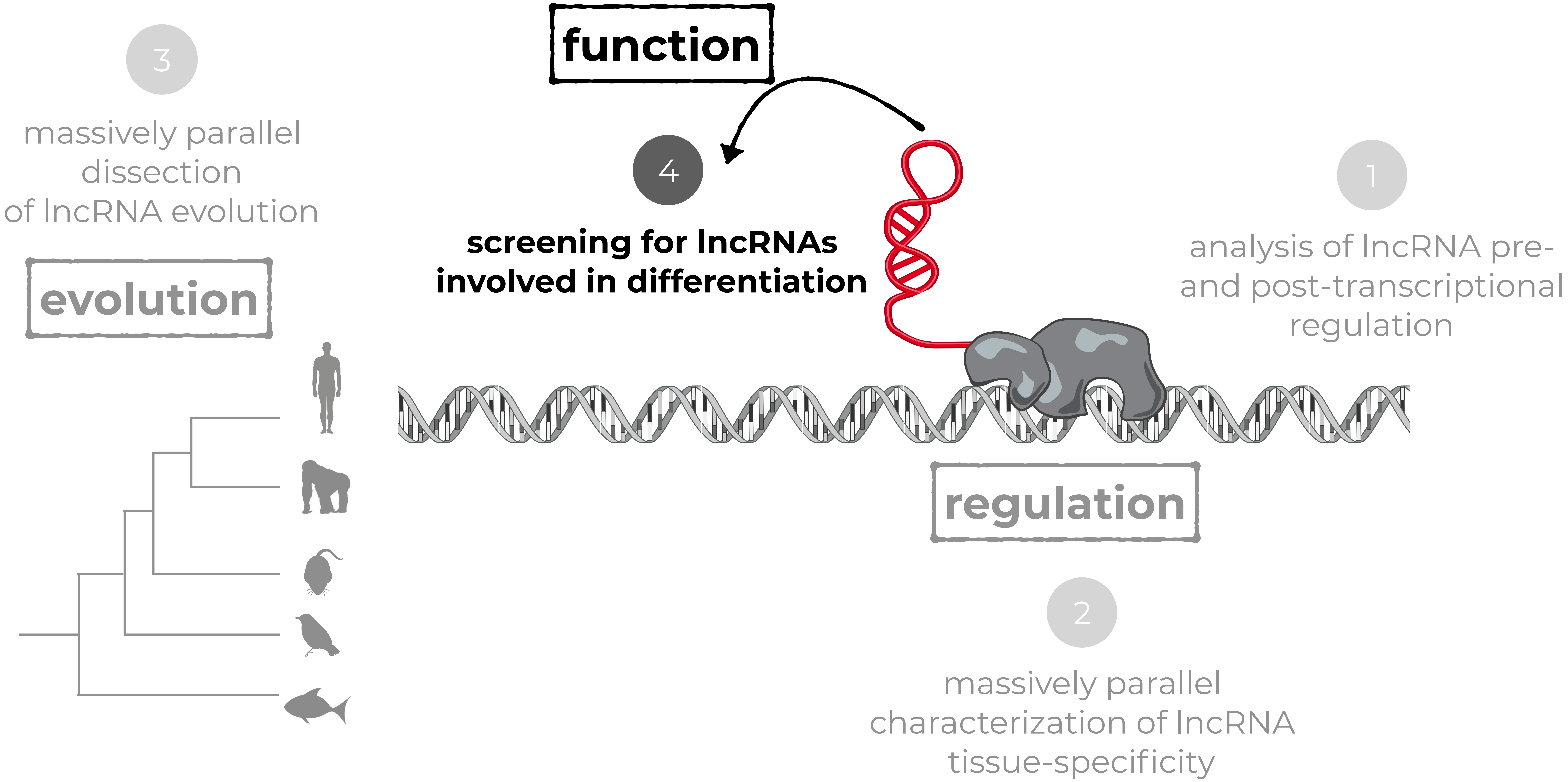
eRNAs are **uniquely** fertile targets for evolutionary change

# Does enhancer redundancy buffer *trans* effects?



**yes**! look out for our BioRxiv paper coming soon :)

# Studying lncRNA biology *en masse*



3

massively parallel
dissection
of lncRNA evolution

**evolution**

**function**

4

**screening for lncRNAs**
**involved in differentiation**

1

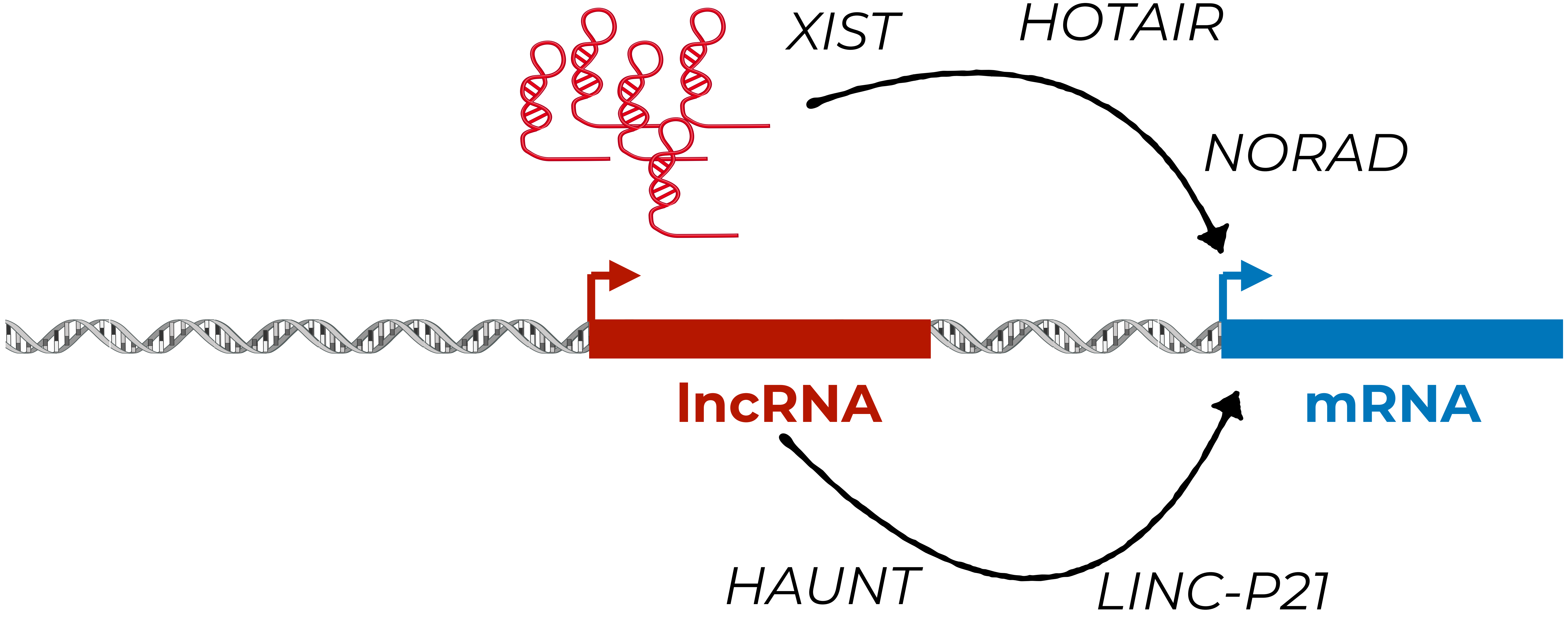
analysis of lncRNA pre-
and post-transcriptional
regulation

**regulation**

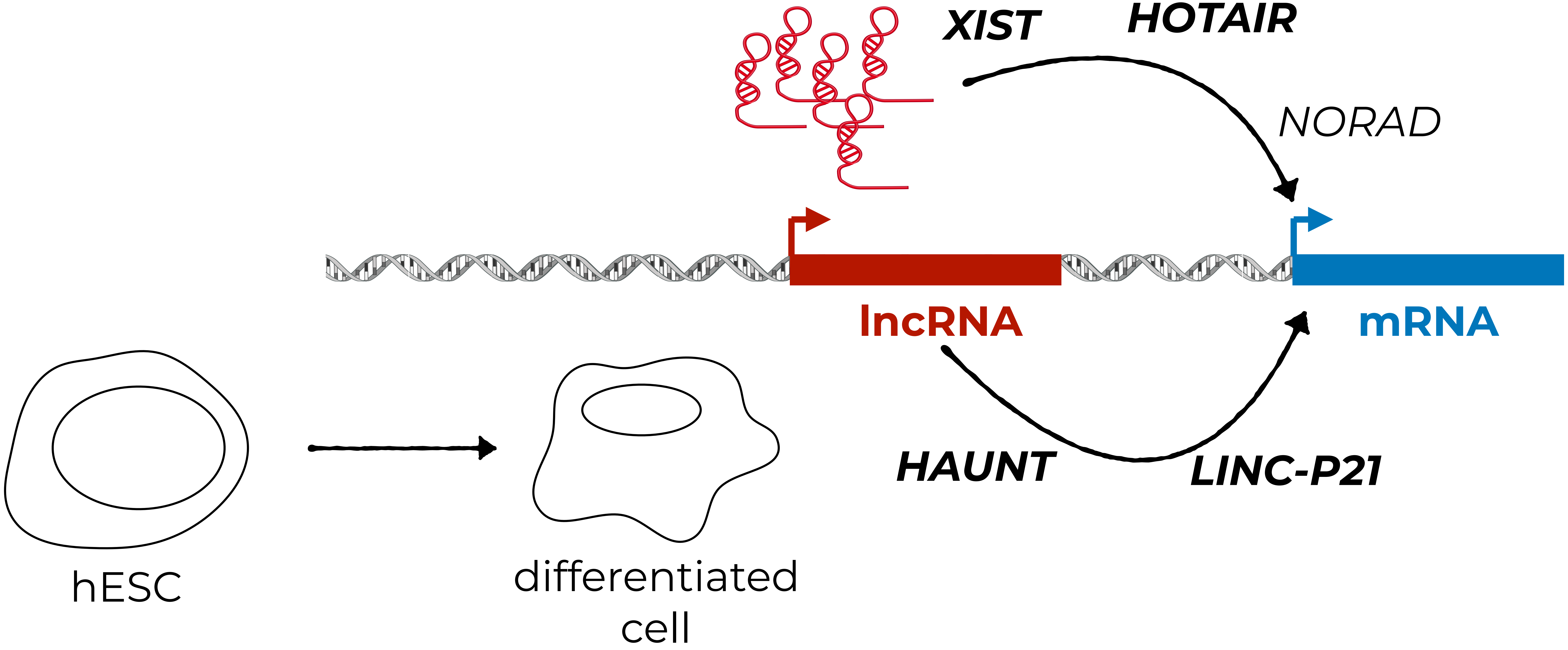
2

massively parallel
characterization of lncRNA
tissue-specificity

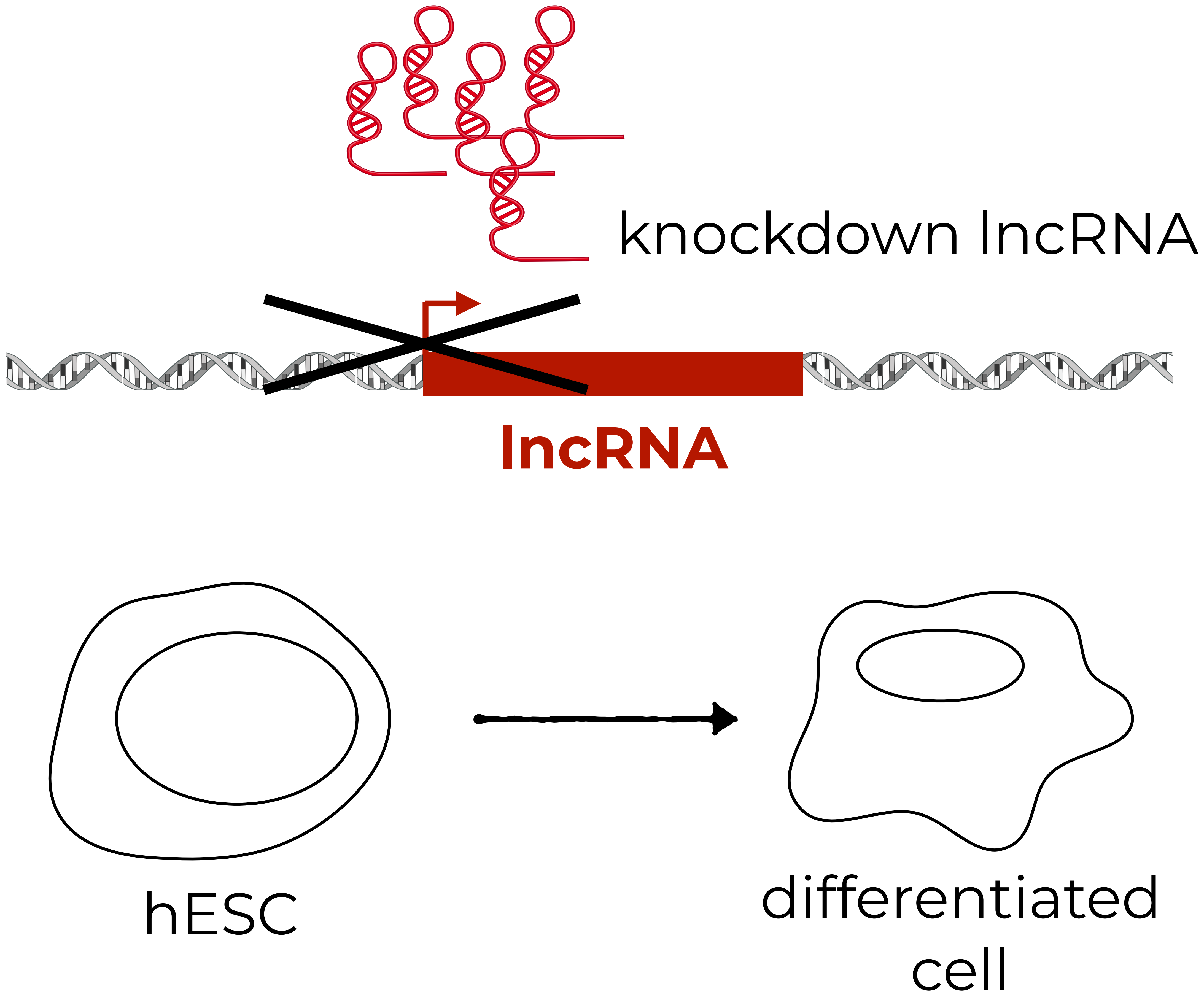
# Several known lncRNAs act in development



# Several known lncRNAs act in development



# Finding lncRNA loci that function in differentiation: individual knockdowns



knockdown lncRNA

**lncRNA**

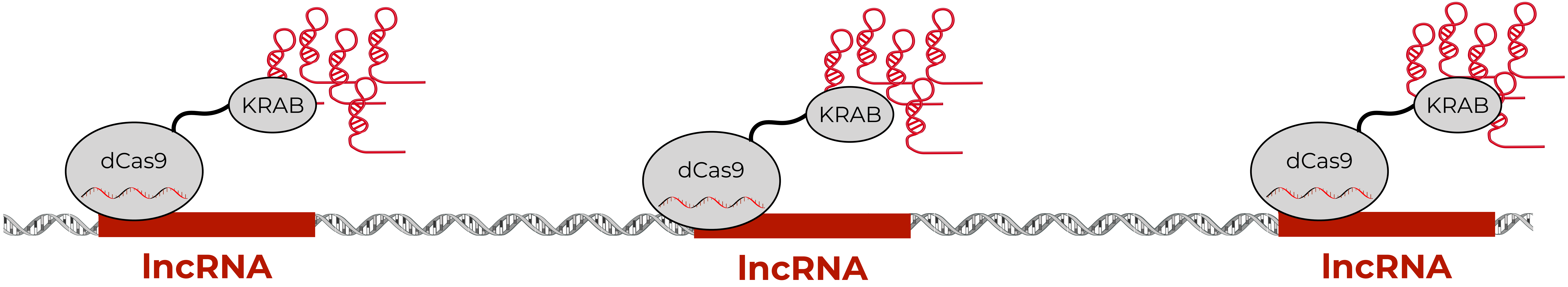
hESC

differentiated cell

**is differentiation affected?**

# Finding lncRNA loci that function in differentiation *en masse* using CRISPRi

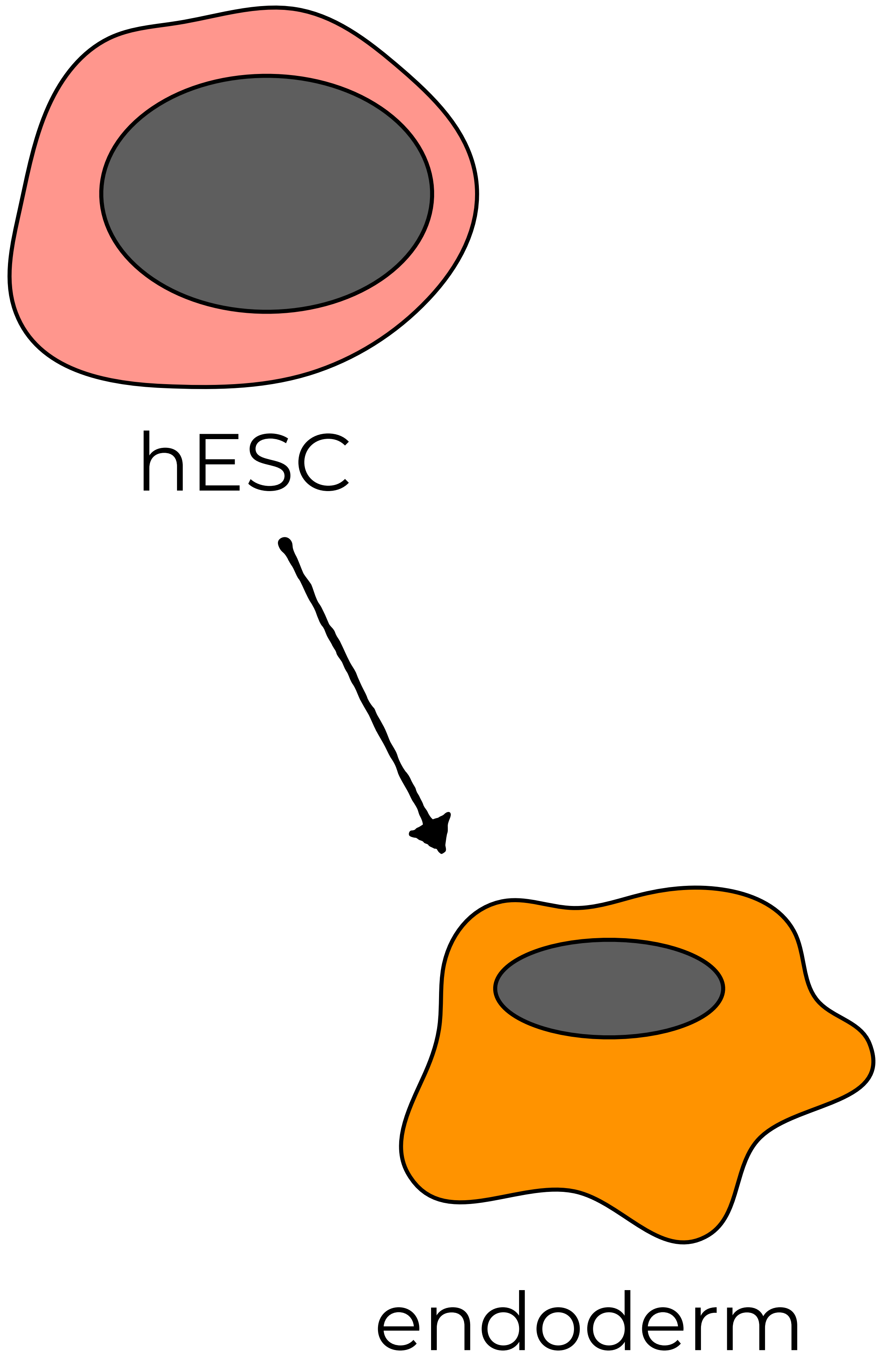
dCas9-KRAB silences genes



**design a library of sgRNAs to target lncRNAs possibly functioning in differentiation**

# Designing an sgRNA library to target lncRNAs that may act in differentiation

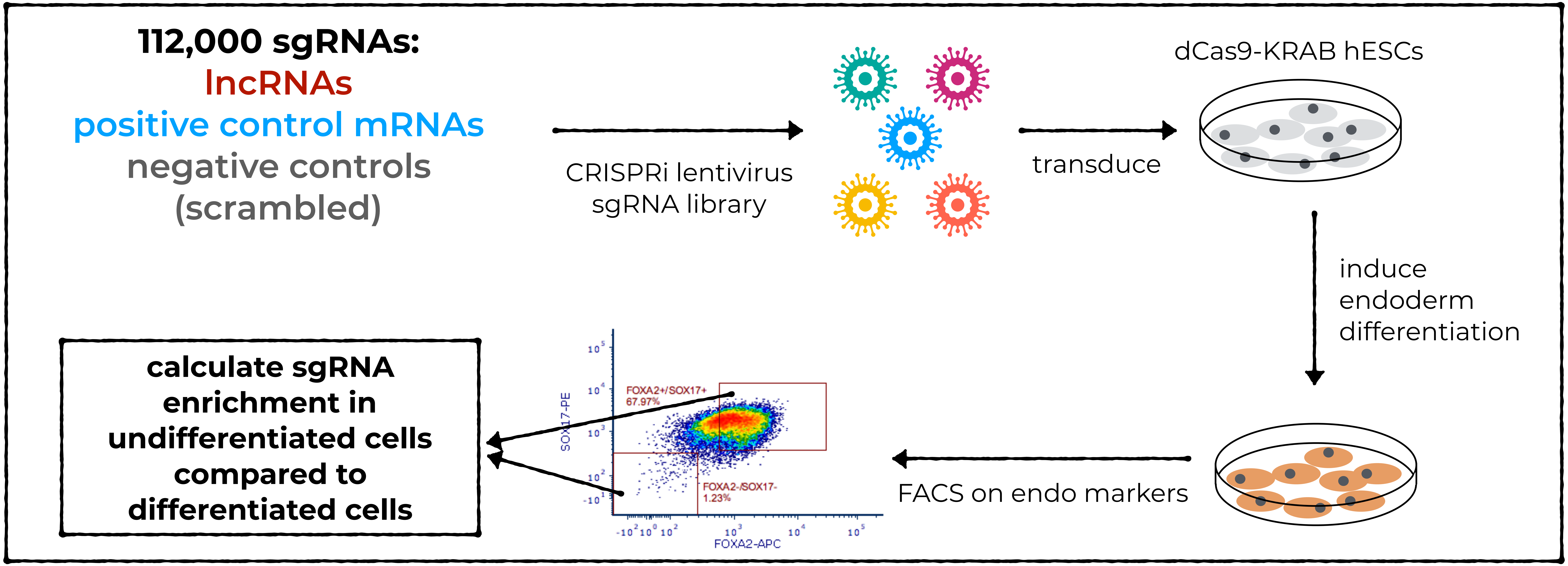
**RNA-sequencing**



target any lncRNA expressed in **hESCs** or **endoderm:** 10,800 transcripts

# CRISPRi screen for lncRNA loci functioning in endoderm differentiation

2 biological replicates



**112,000 sgRNAs:**
lncRNAs
positive control mRNAs
negative controls
(scrambled)

CRISPRi lentivirus
sgRNA library

transduce

dCas9-KRAB hESCs

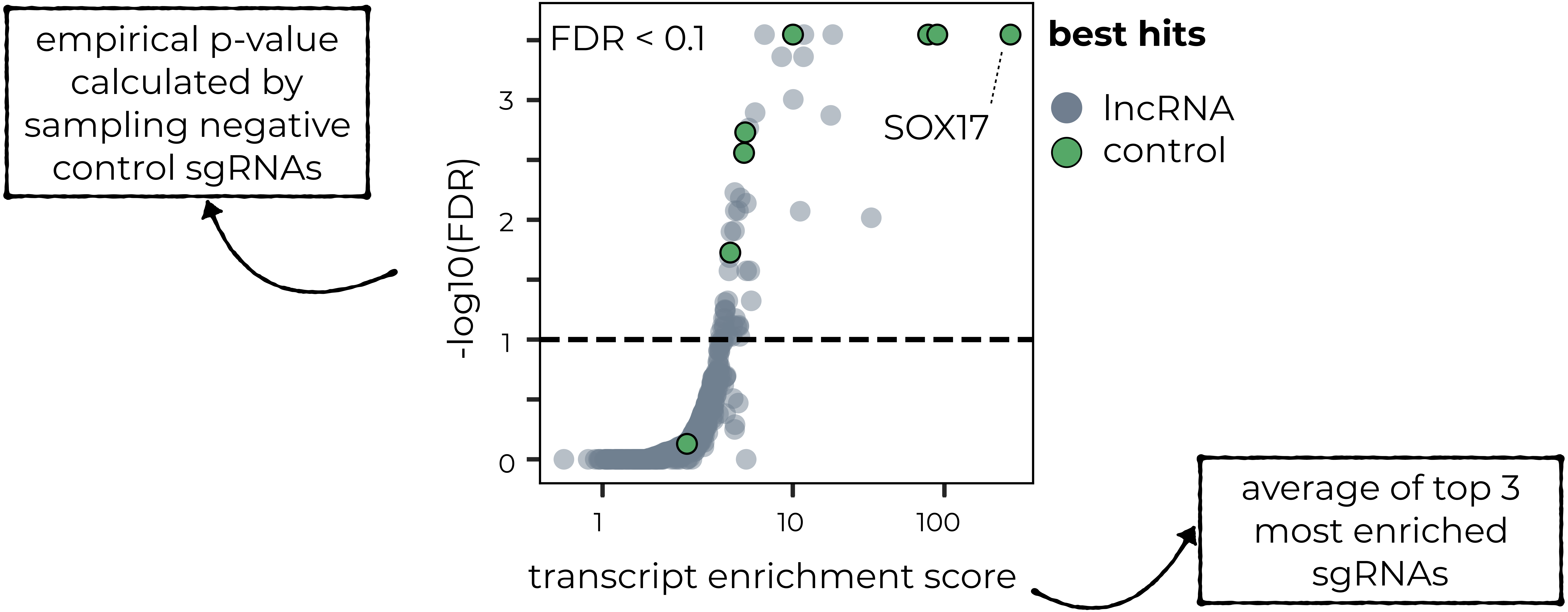
induce
endoderm
differentiation

FACS on endo markers

**calculate sgRNA enrichment in undifferentiated cells compared to differentiated cells**

collaboration with Jeff Haswell (Slack Lab)

# CRISPRi screen reveals lncRNA loci required for endoderm differentiation



empirical p-value calculated by sampling negative control sgRNAs

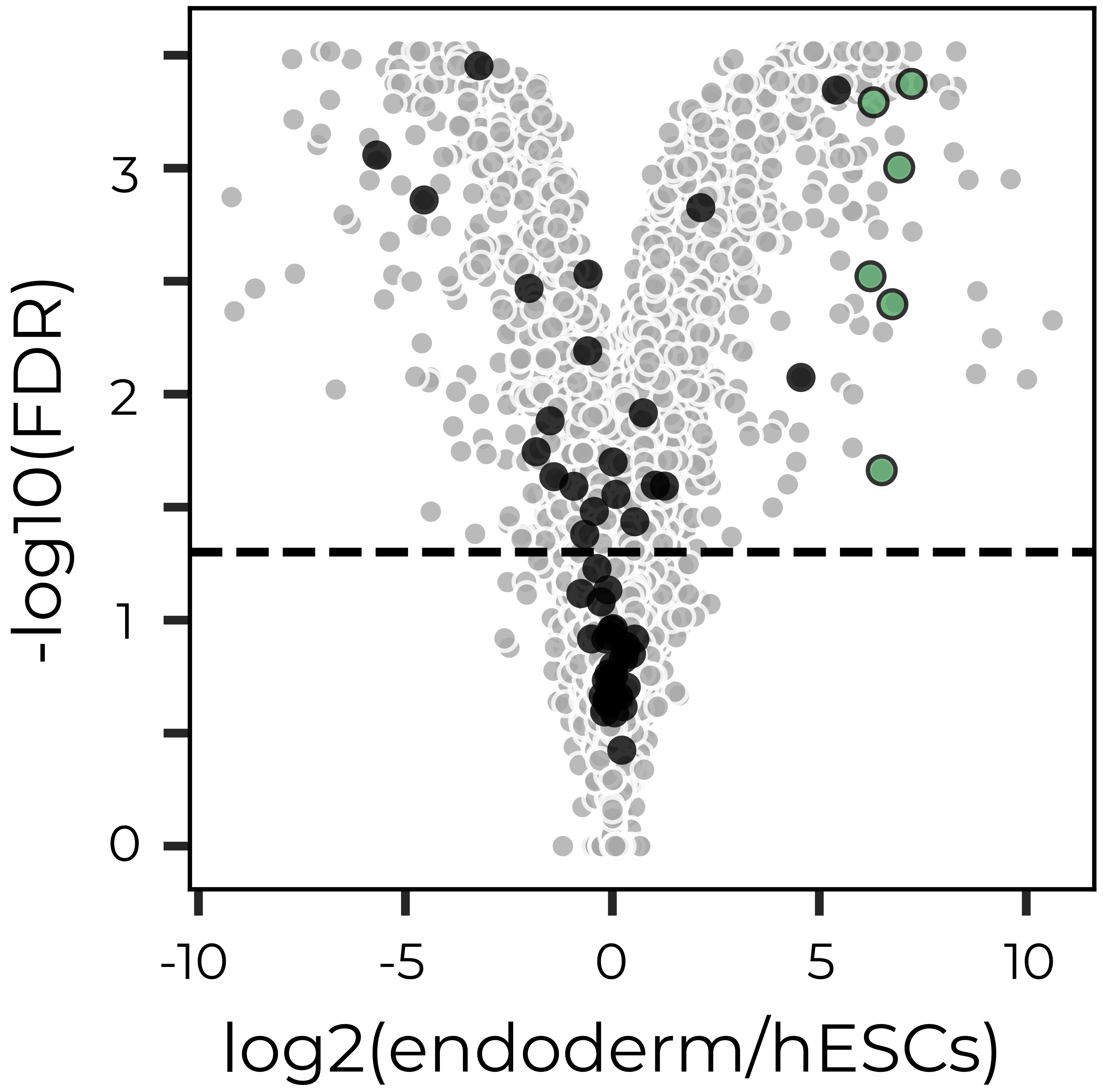
average of top 3 most enriched sgRNAs

# Comparing hits to RNA-seq data

● non-hit lncRNA

● hit lncRNA

🟢 hit control

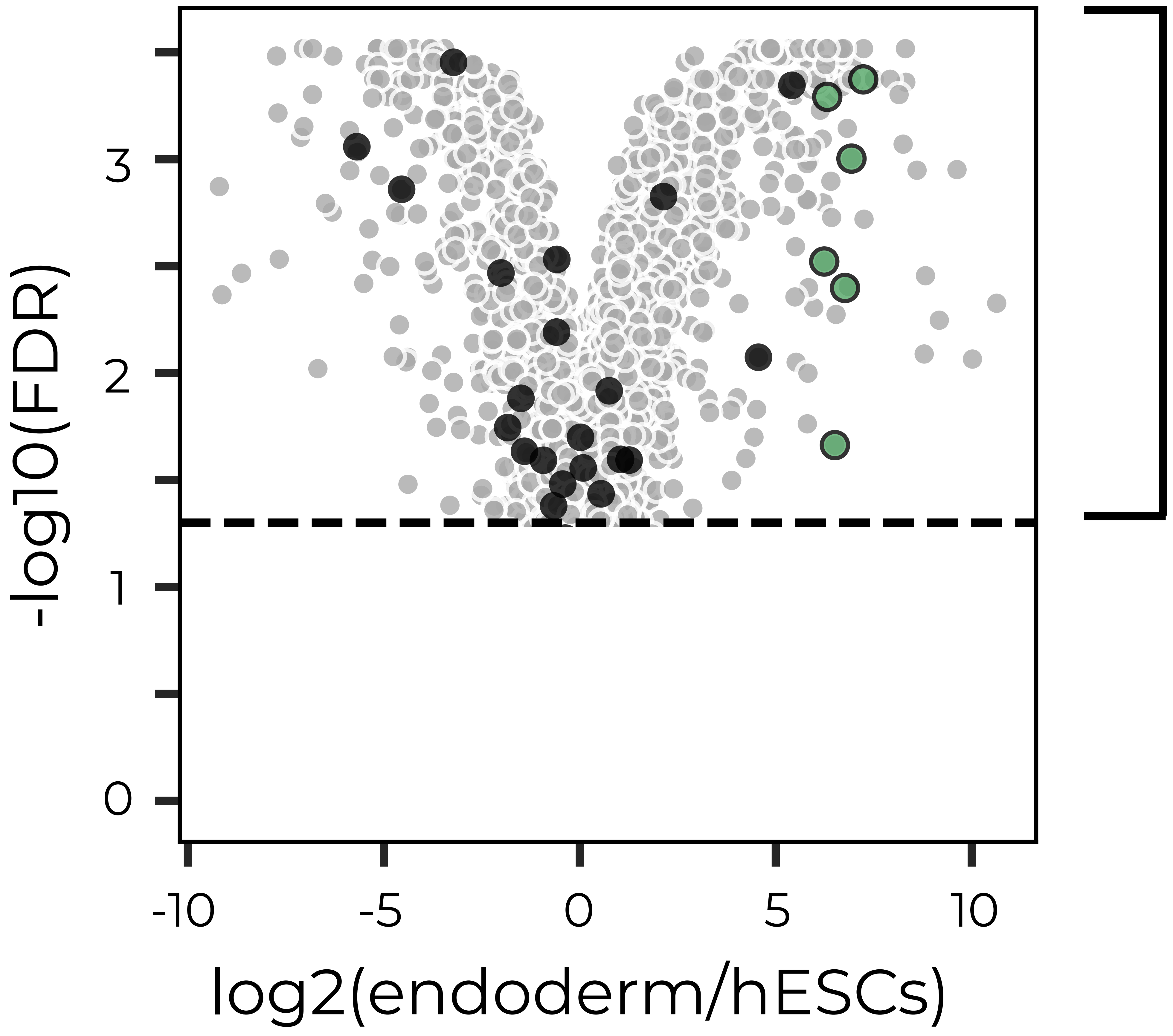


# Many differentially-expressed lncRNAs are not hits

● non-hit lncRNA

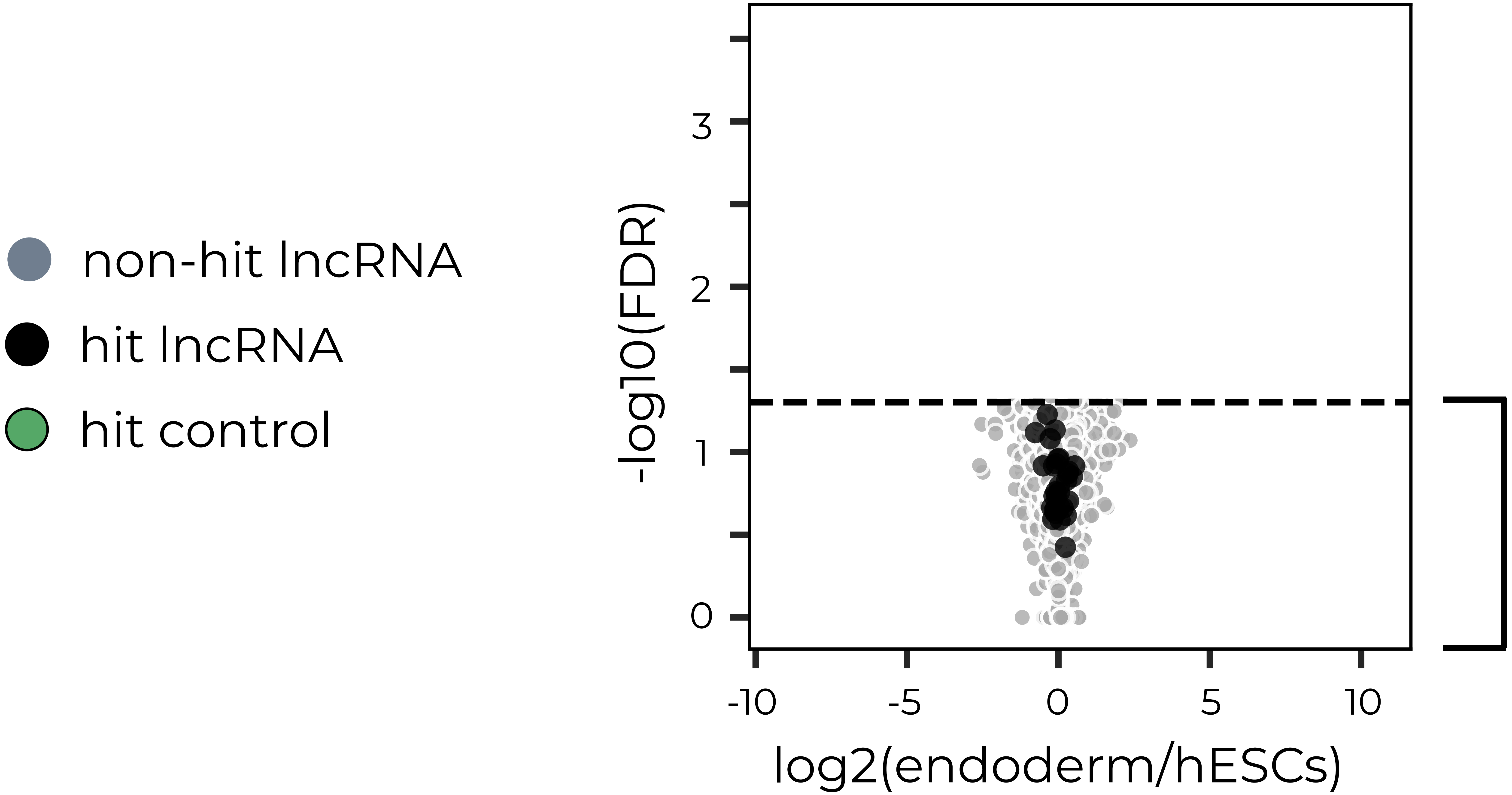
● hit lncRNA

🟢 hit control



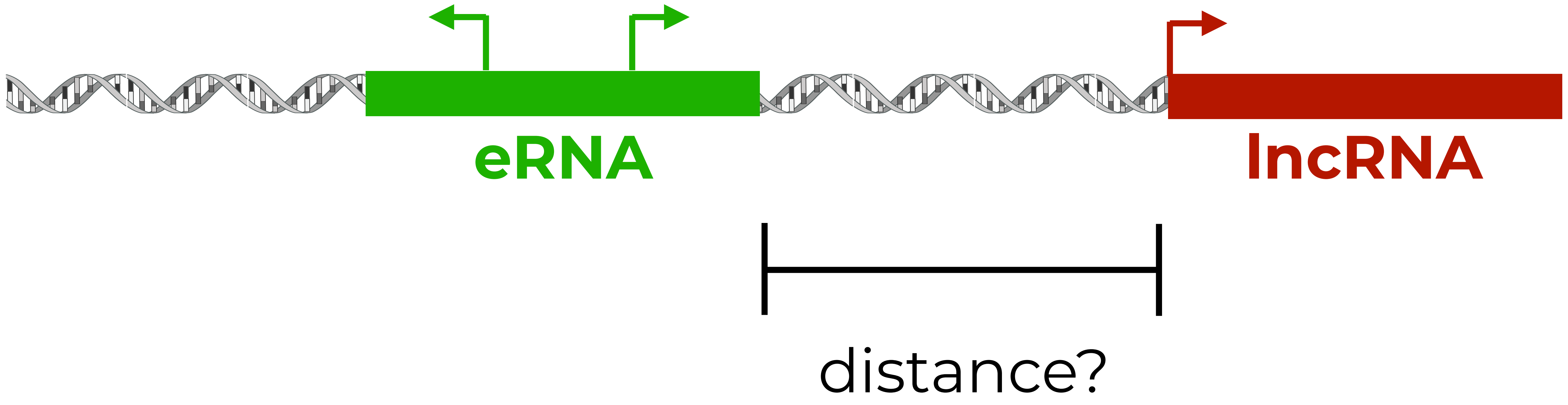
**phenotypic screens are important to identify functional loci**

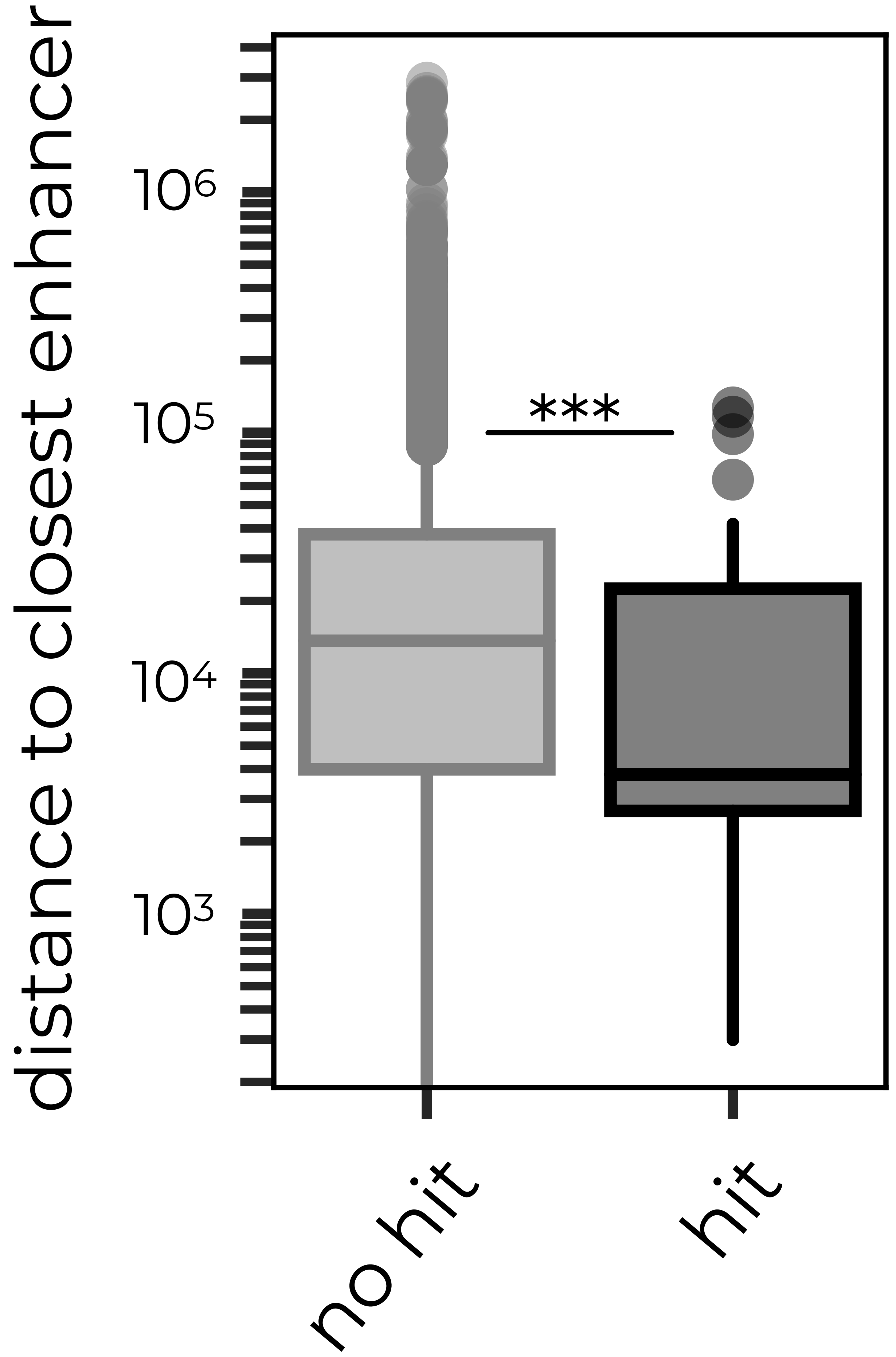
# Many lncRNA hits are not differentially expressed



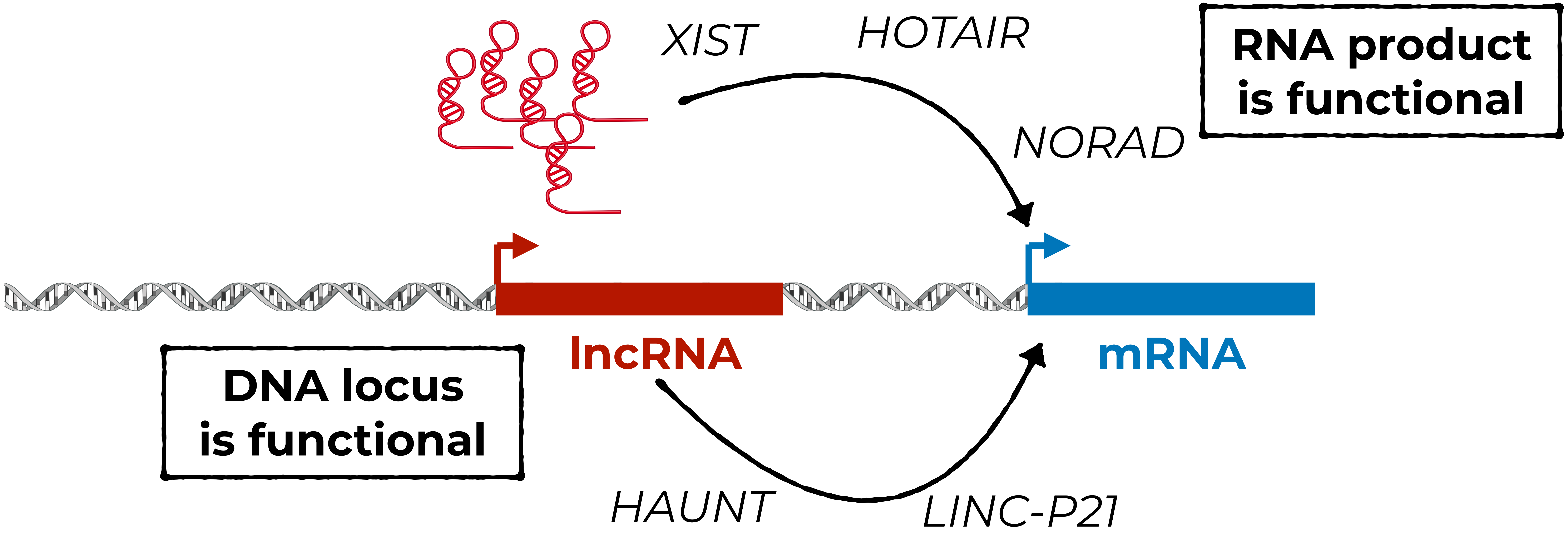
**RNA-based mechanisms unlikely?**

# lncRNA hits are close to eRNAs





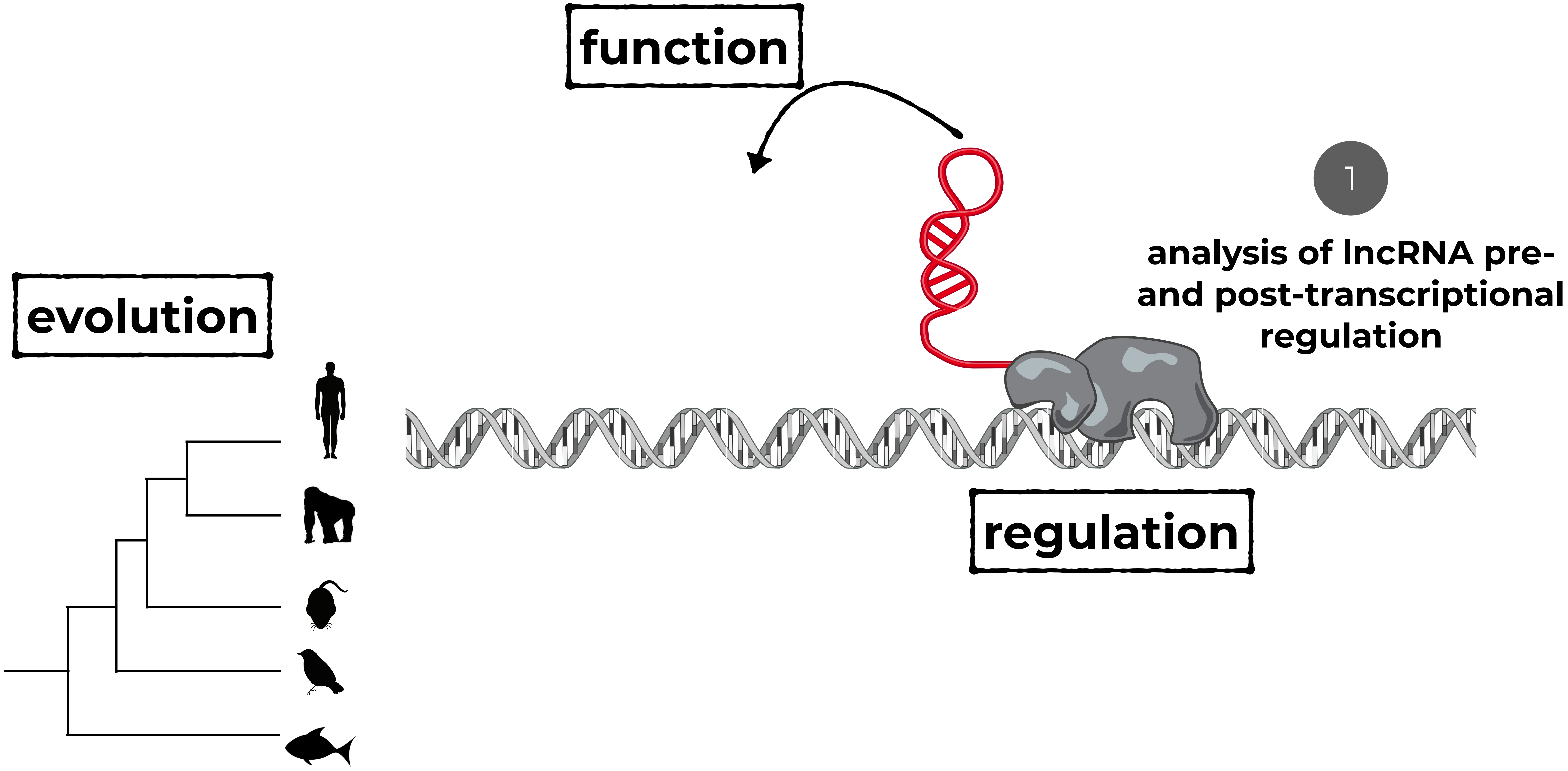
# Many DNA-based lncRNAs acting in endoderm differentiation?



**RNA product is functional**

**DNA locus is functional**

# Conclusions from studying lncRNA biology *en masse*



**function**

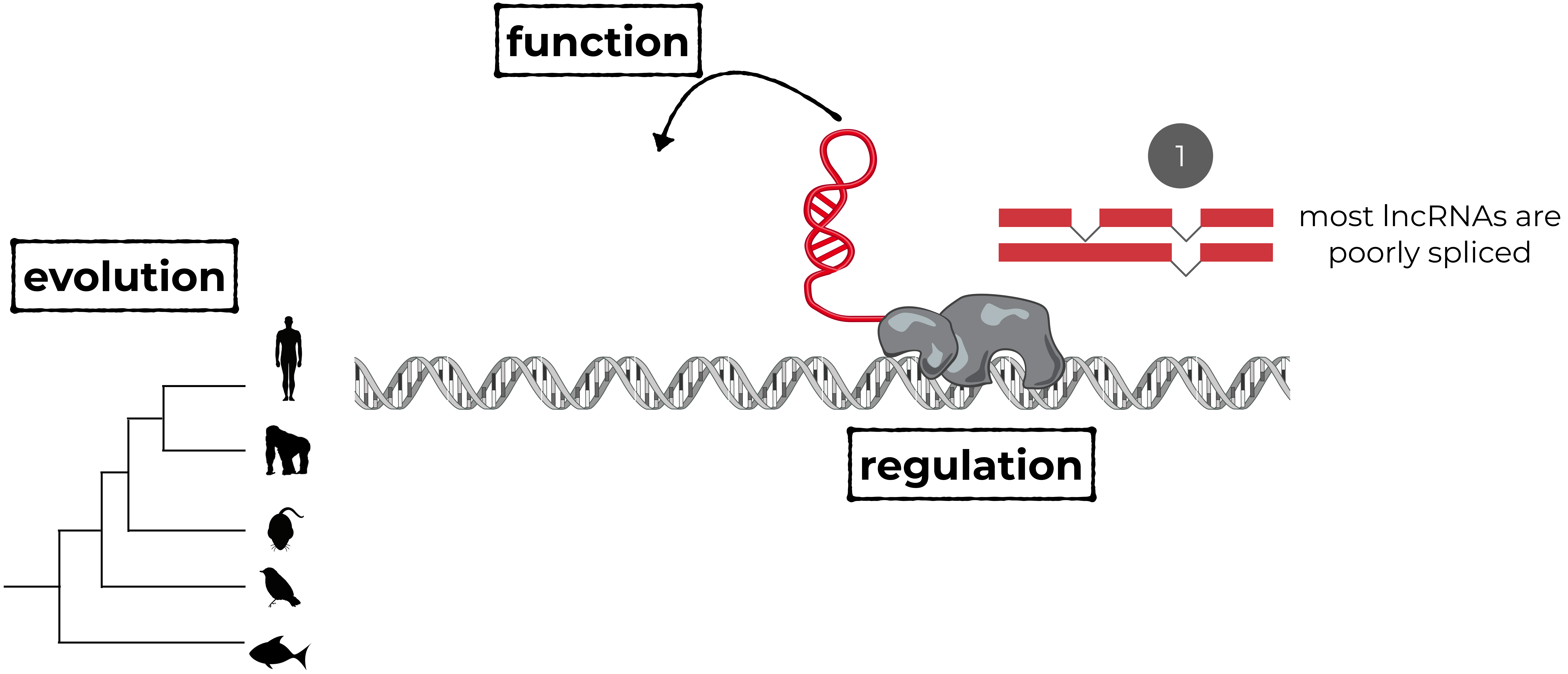
**evolution**

**regulation**

①

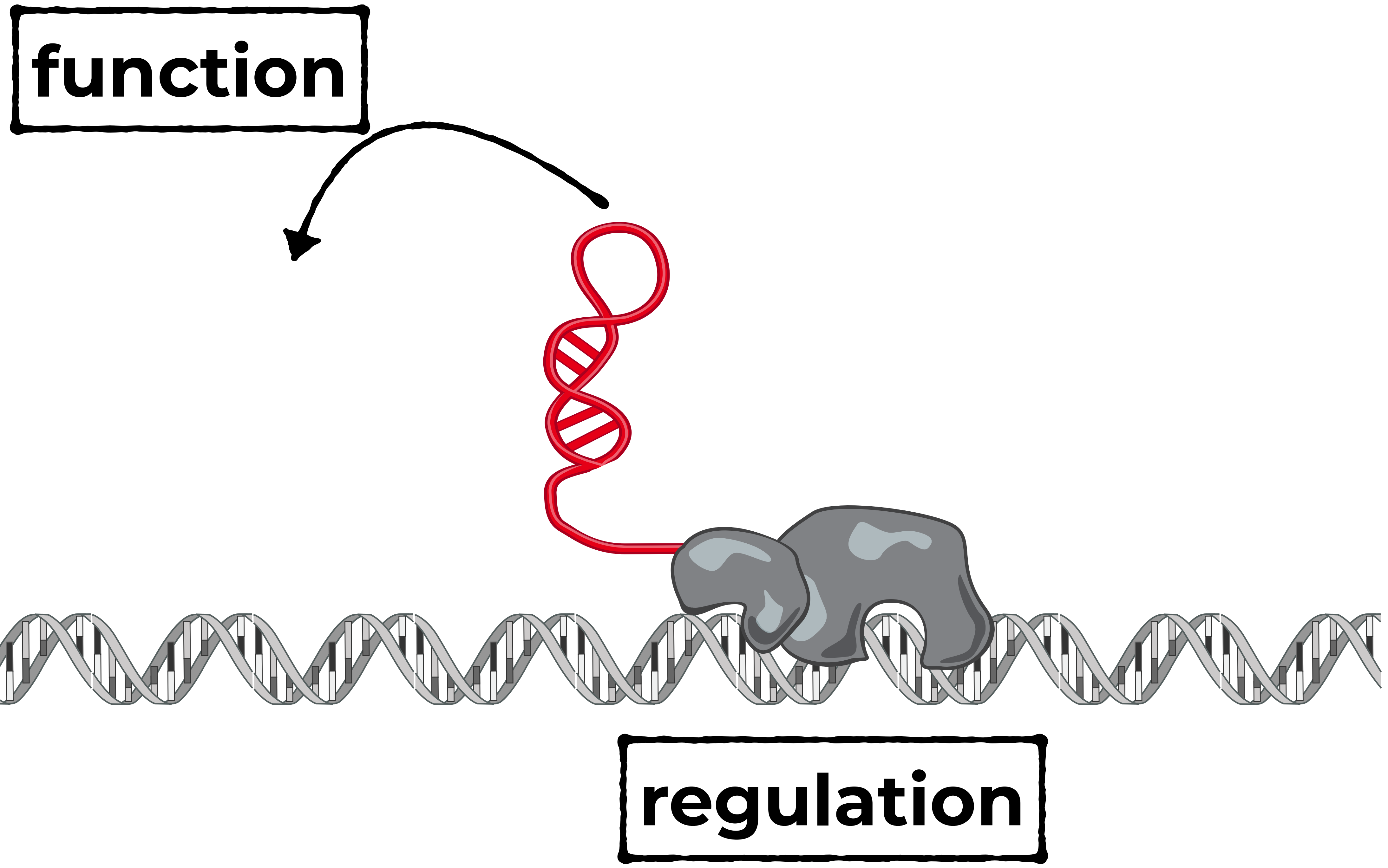
**analysis of lncRNA pre- and post-transcriptional regulation**

# Conclusions from studying lncRNA biology *en masse*



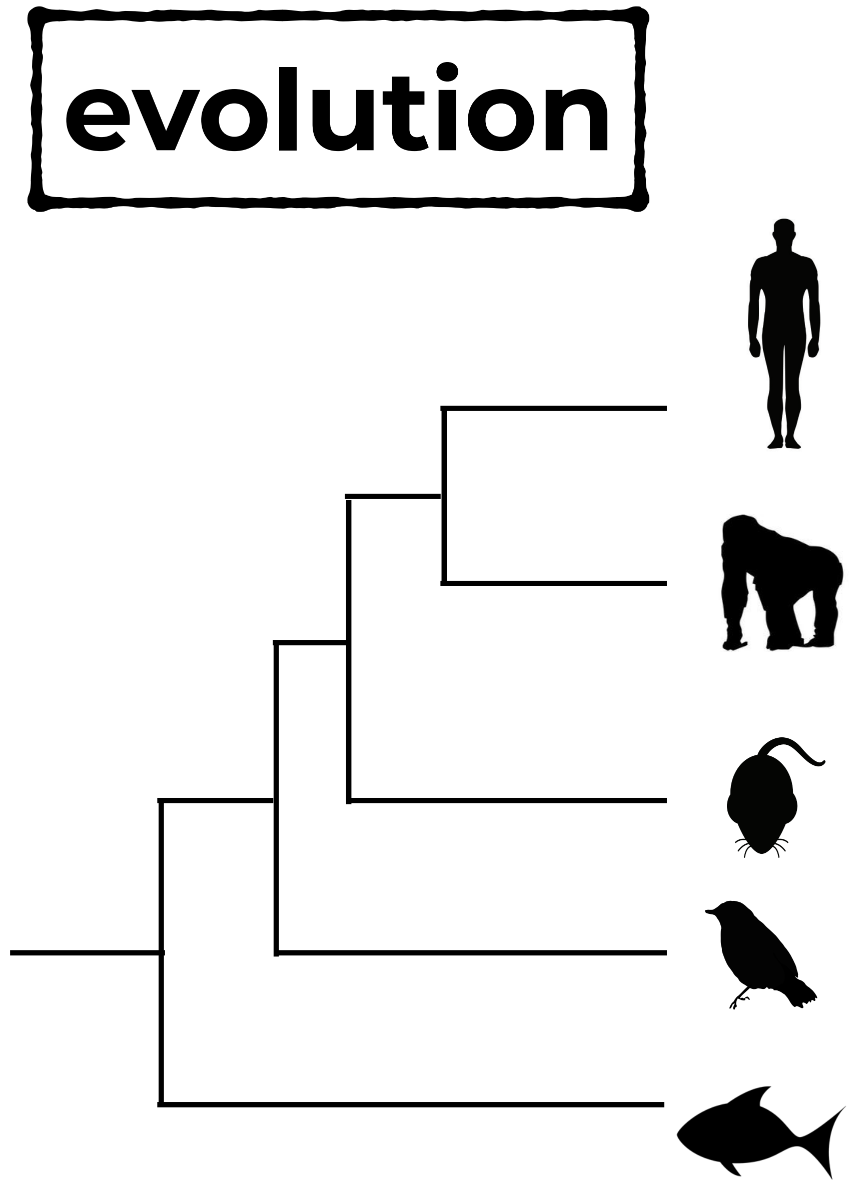
most lncRNAs are poorly spliced

# Conclusions from studying lncRNA biology *en masse*



**function**

**regulation**

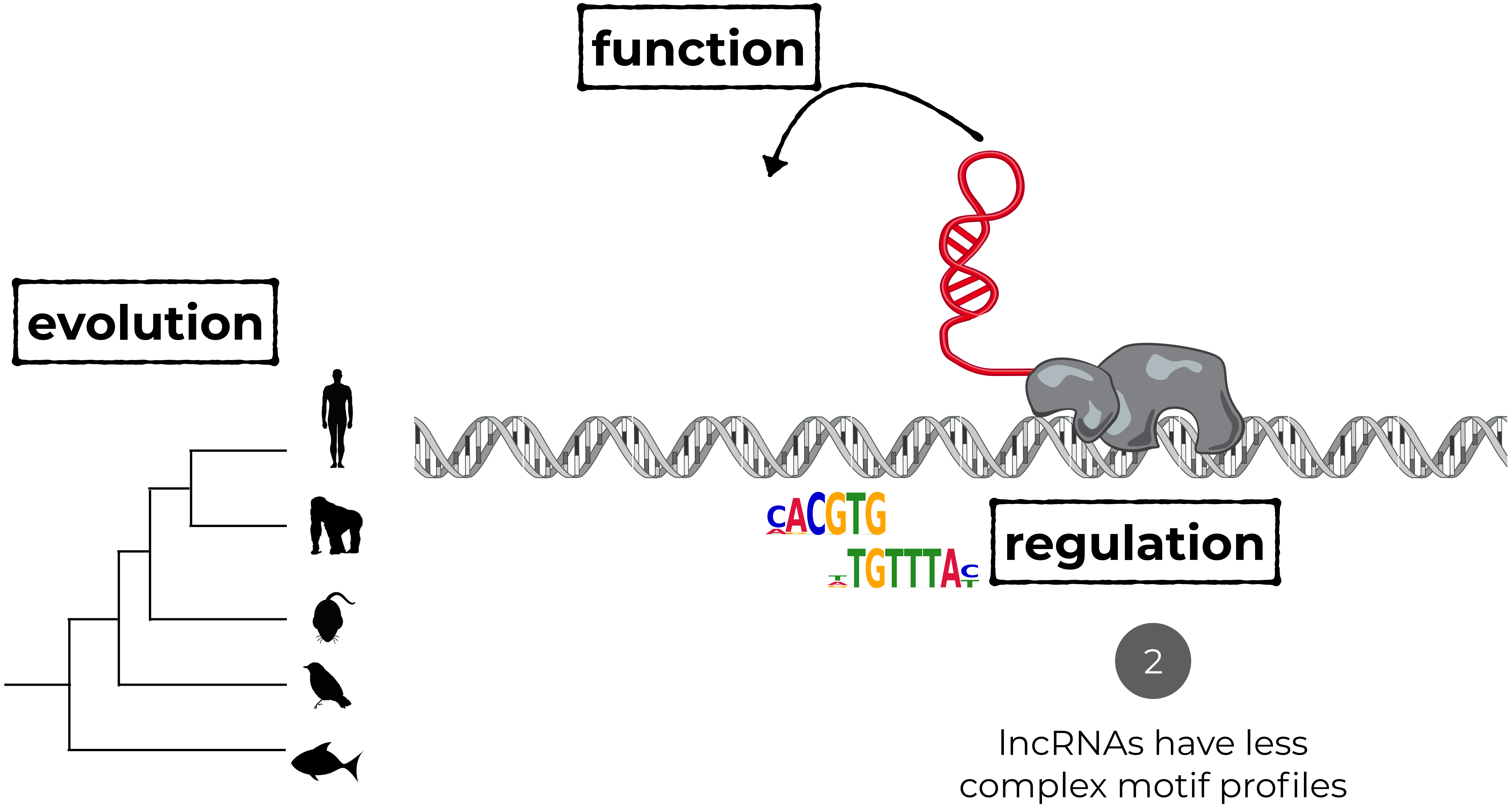


**evolution**

2

**massively parallel characterization of lncRNA tissue-specificity**

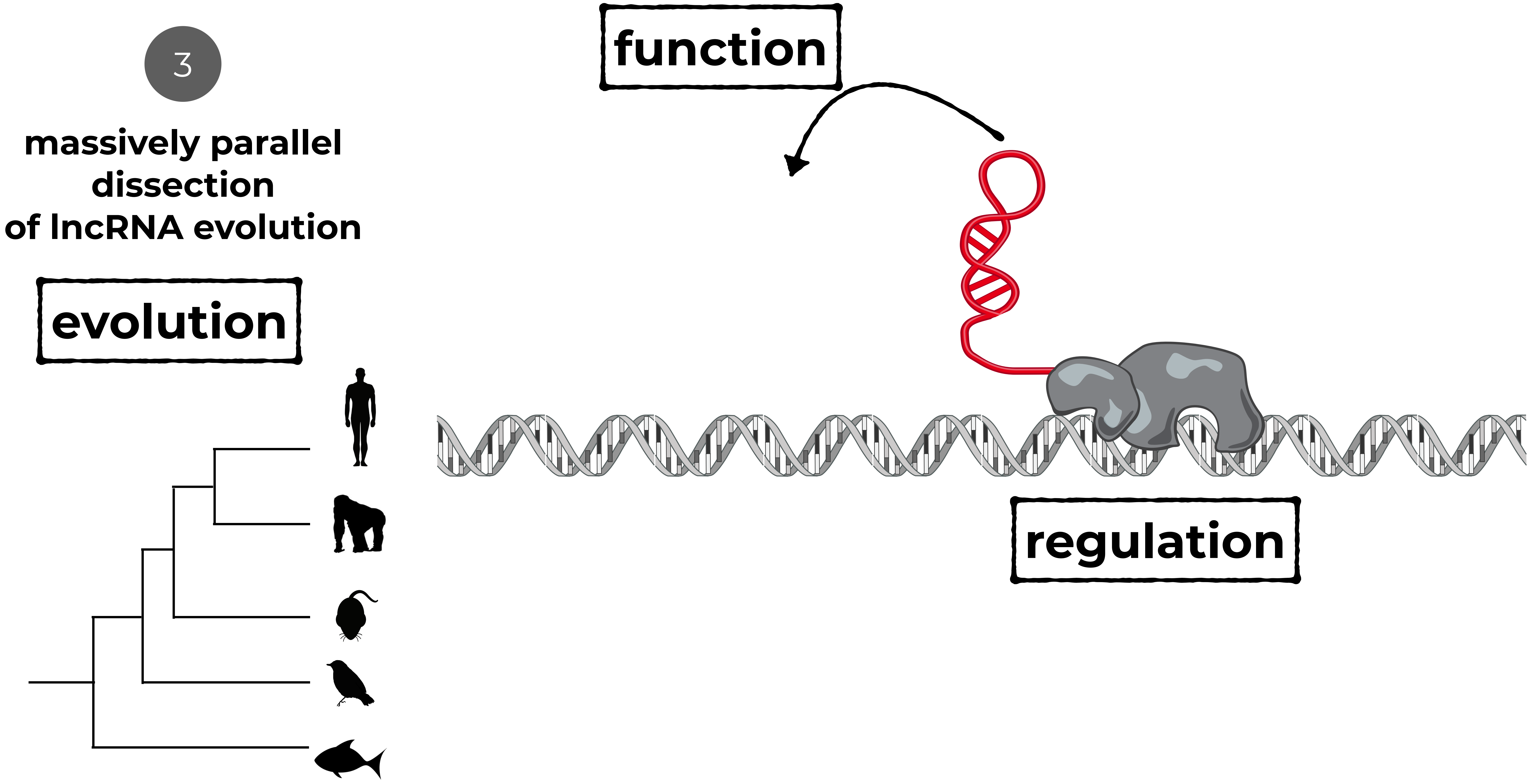
# Conclusions from studying lncRNA biology *en masse*



2

lncRNAs have less complex motif profiles

# Conclusions from studying lncRNA biology *en masse*



3

**massively parallel
dissection
of lncRNA evolution**

**evolution**

**function**

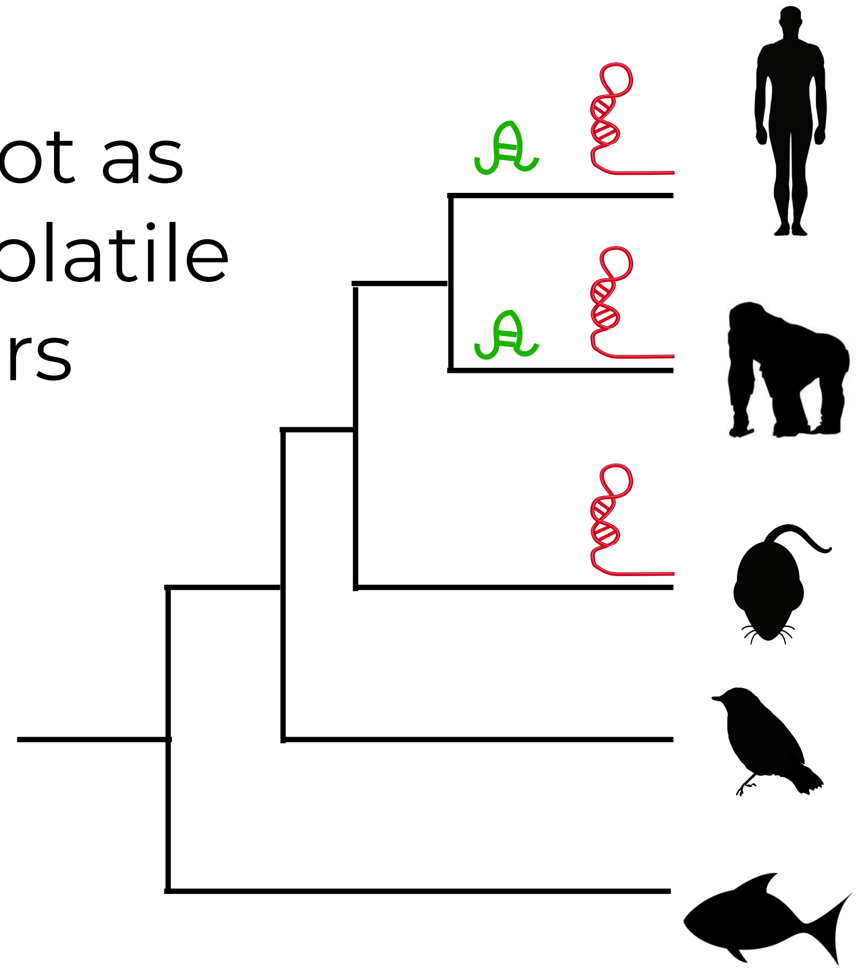
**regulation**

# Conclusions from studying lncRNA biology *en masse*

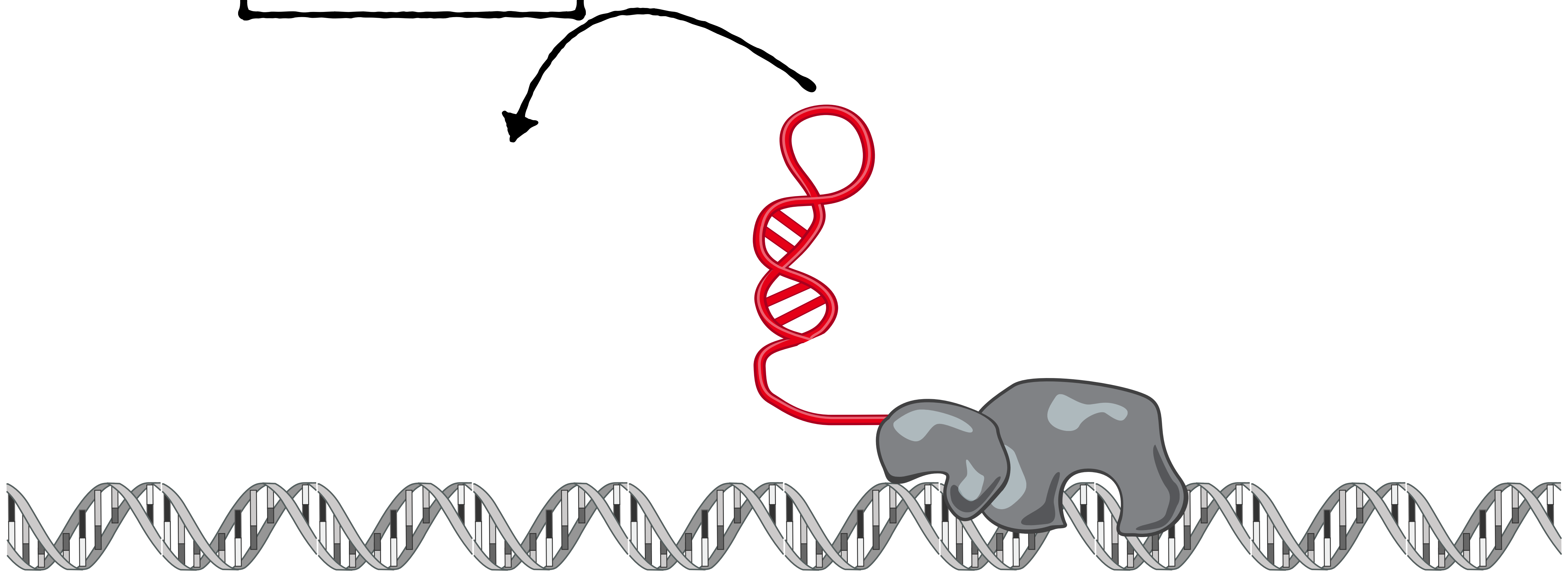
3

**evolution**

lncRNAs are not as evolutionarily volatile as enhancers

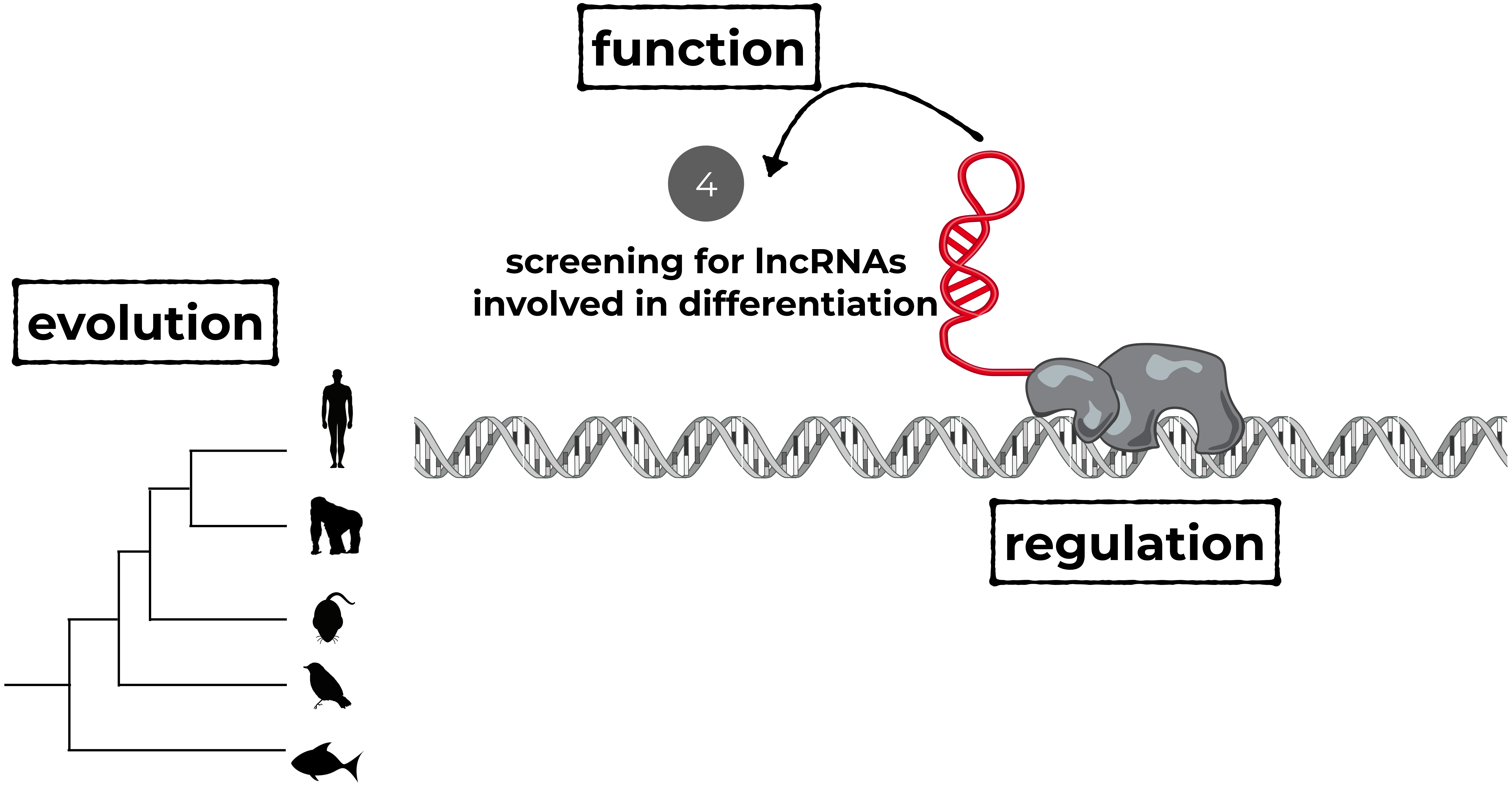


**function**



**regulation**

# Conclusions from studying lncRNA biology *en masse*



**function**

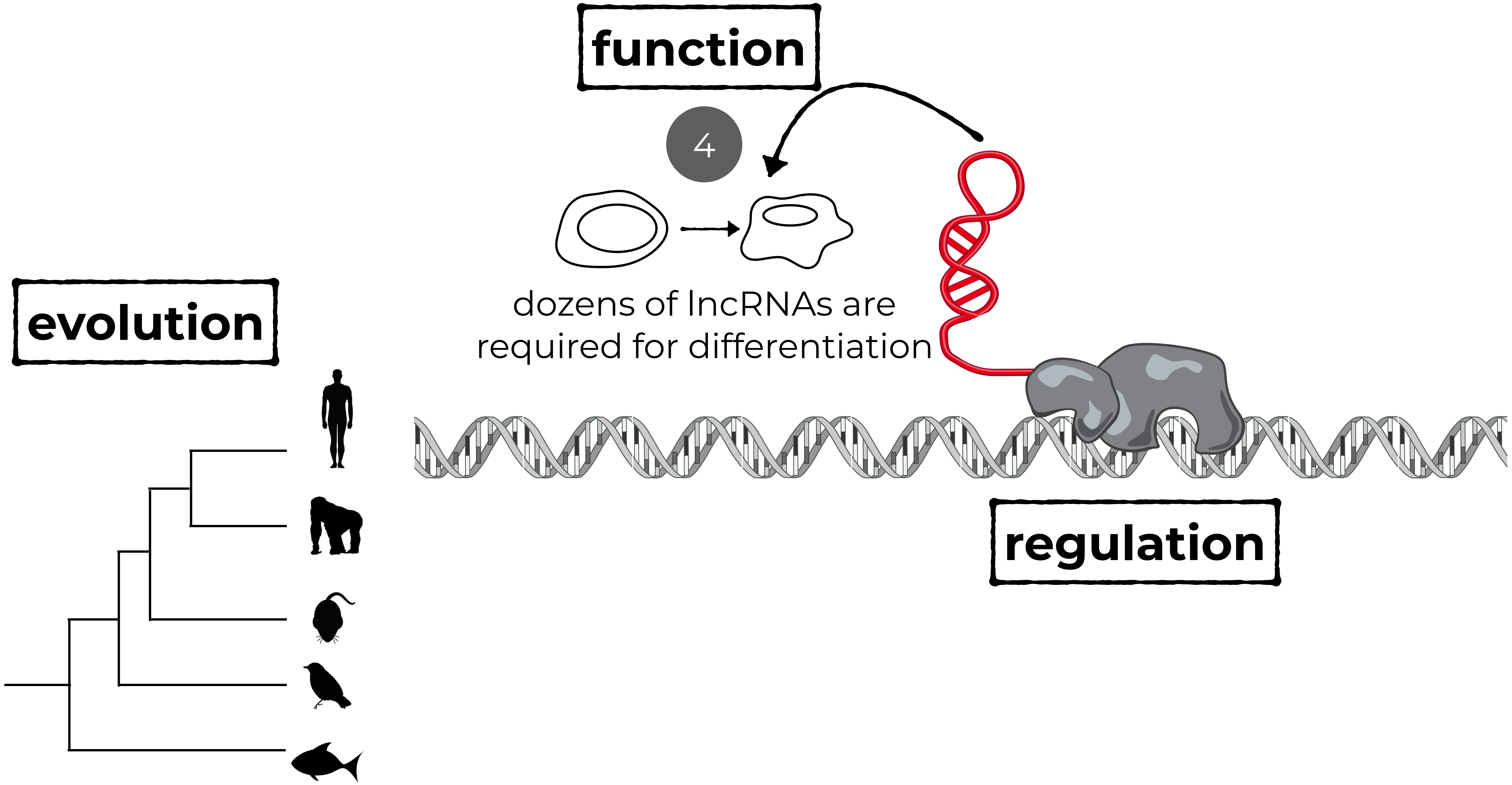
4

**screening for lncRNAs involved in differentiation**

**evolution**

**regulation**

# Conclusions from studying lncRNA biology *en masse*



**function**

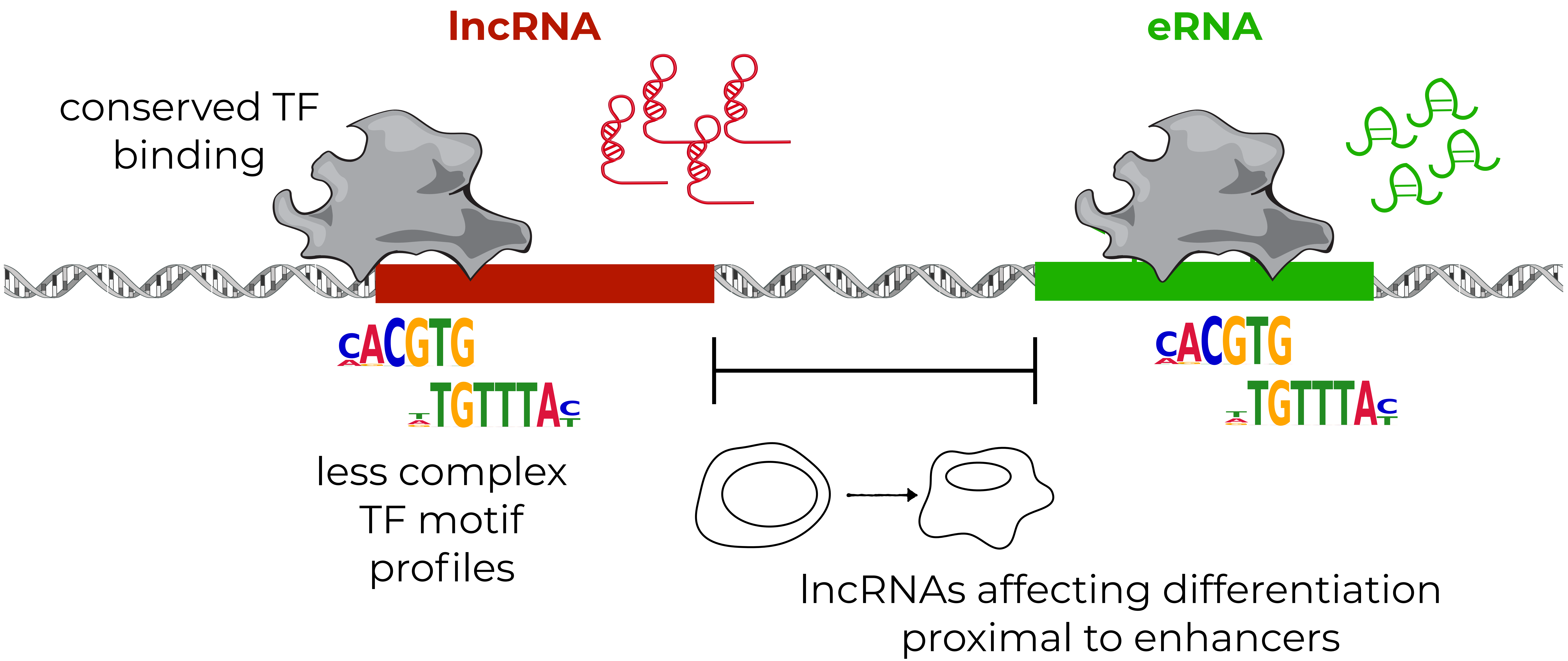
4

dozens of lncRNAs are
required for differentiation

**evolution**

**regulation**

# LncRNAs and eRNAs are similar



**lncRNA**

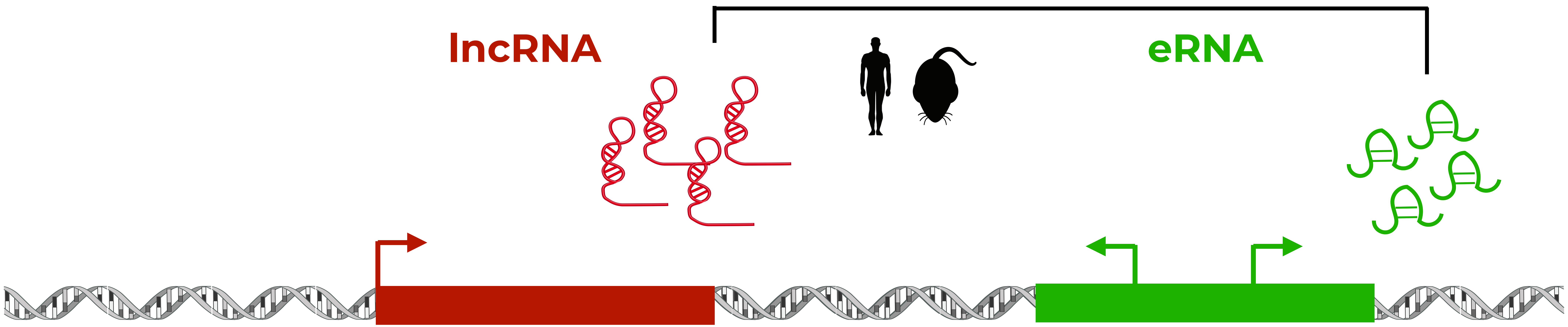
**eRNA**

conserved TF
binding

less complex
TF motif
profiles

lncRNAs affecting differentiation
proximal to enhancers

# … but also different



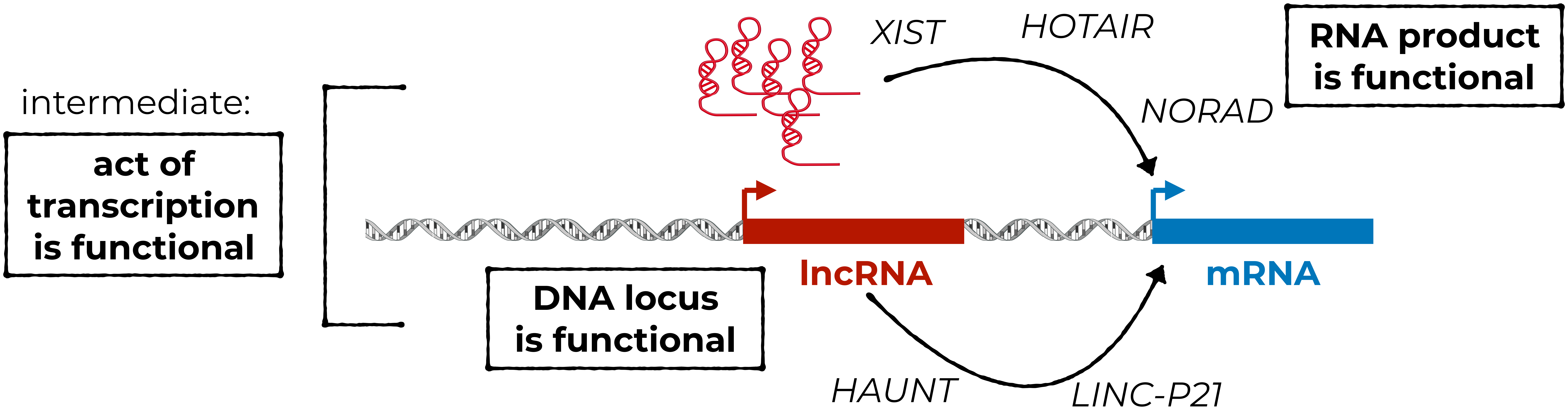
lncRNAs are less evolutionarily volatile

**lncRNA**

**eRNA**

lncRNAs are long,
stable RNA species

# LncRNAs: eRNAs, but where the act of stable transcription is necessary?

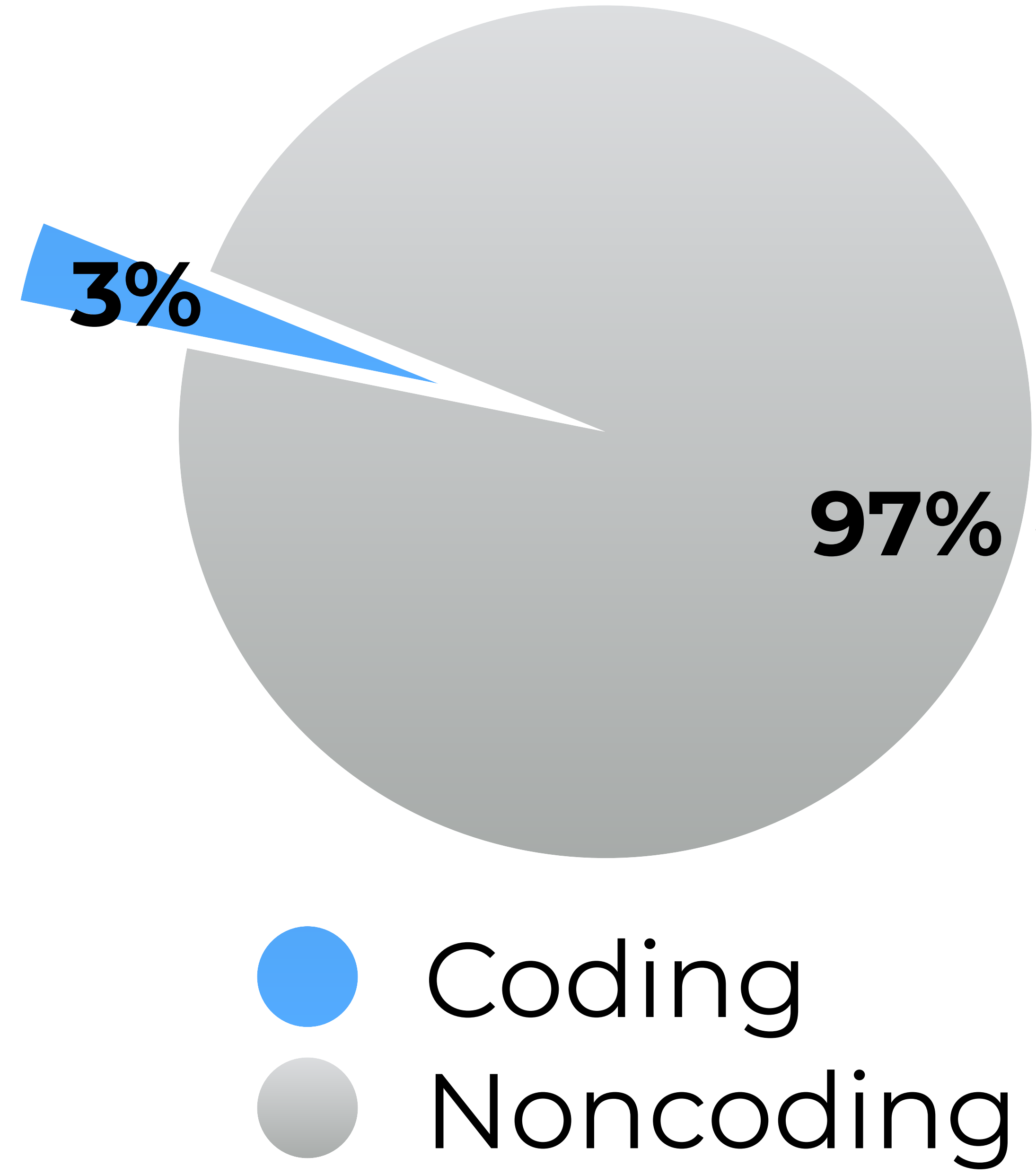

intermediate:

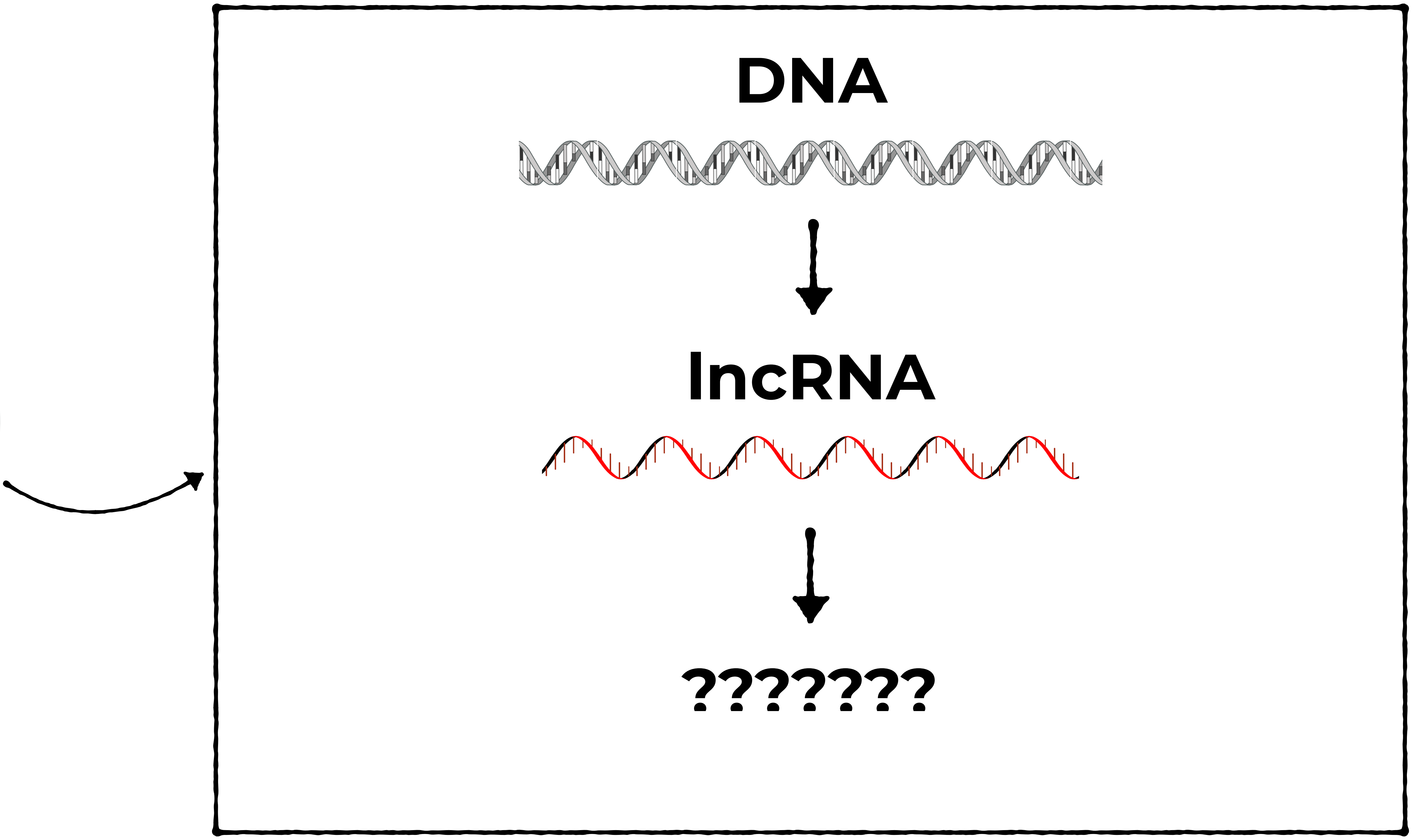
**act of transcription is functional**

**DNA locus is functional**

**RNA product is functional**

# Towards decoding the human genome



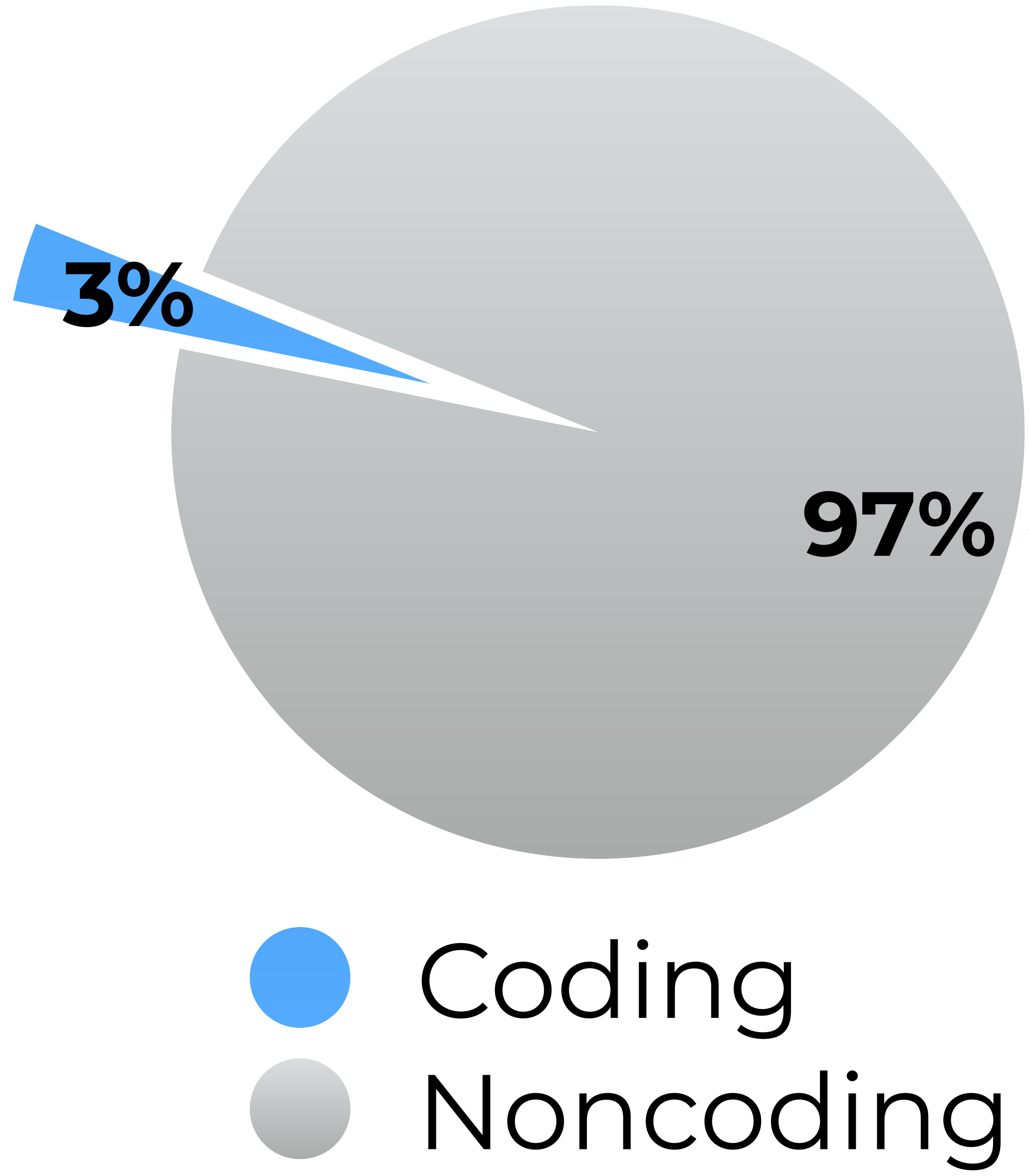


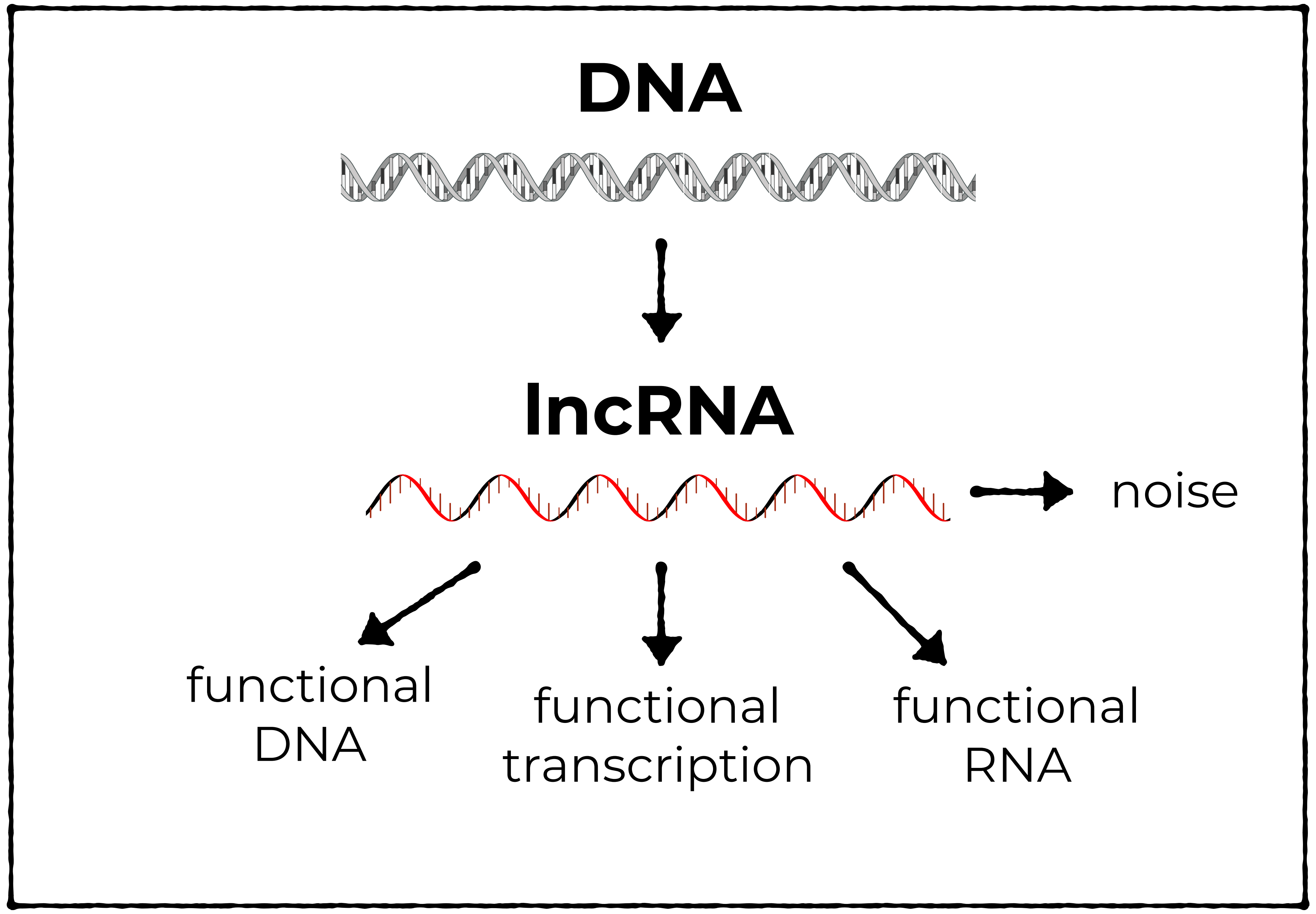
**DNA**

**lncRNA**

**???????**

# Towards decoding the human genome





**DNA**

**lncRNA**

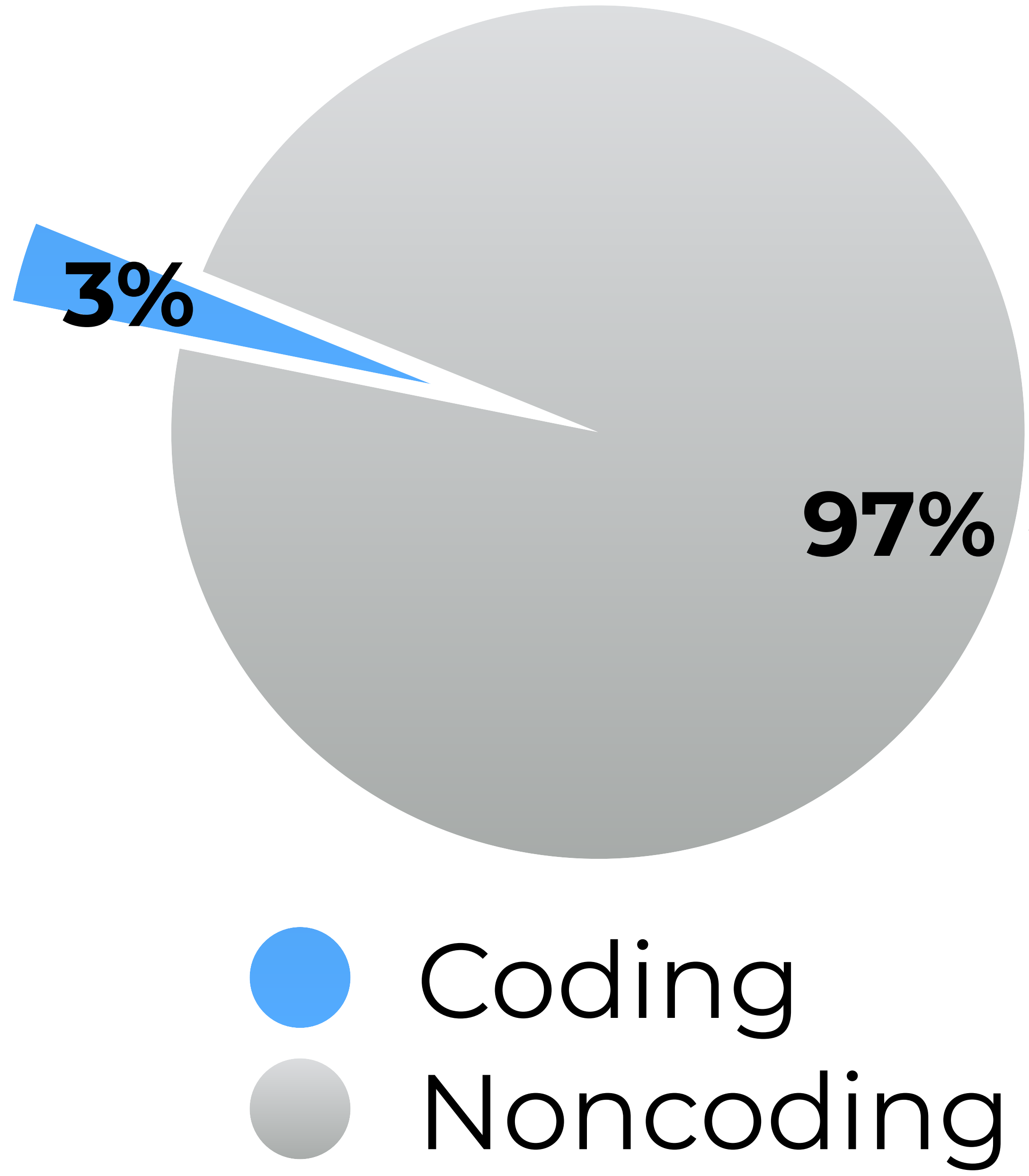
noise

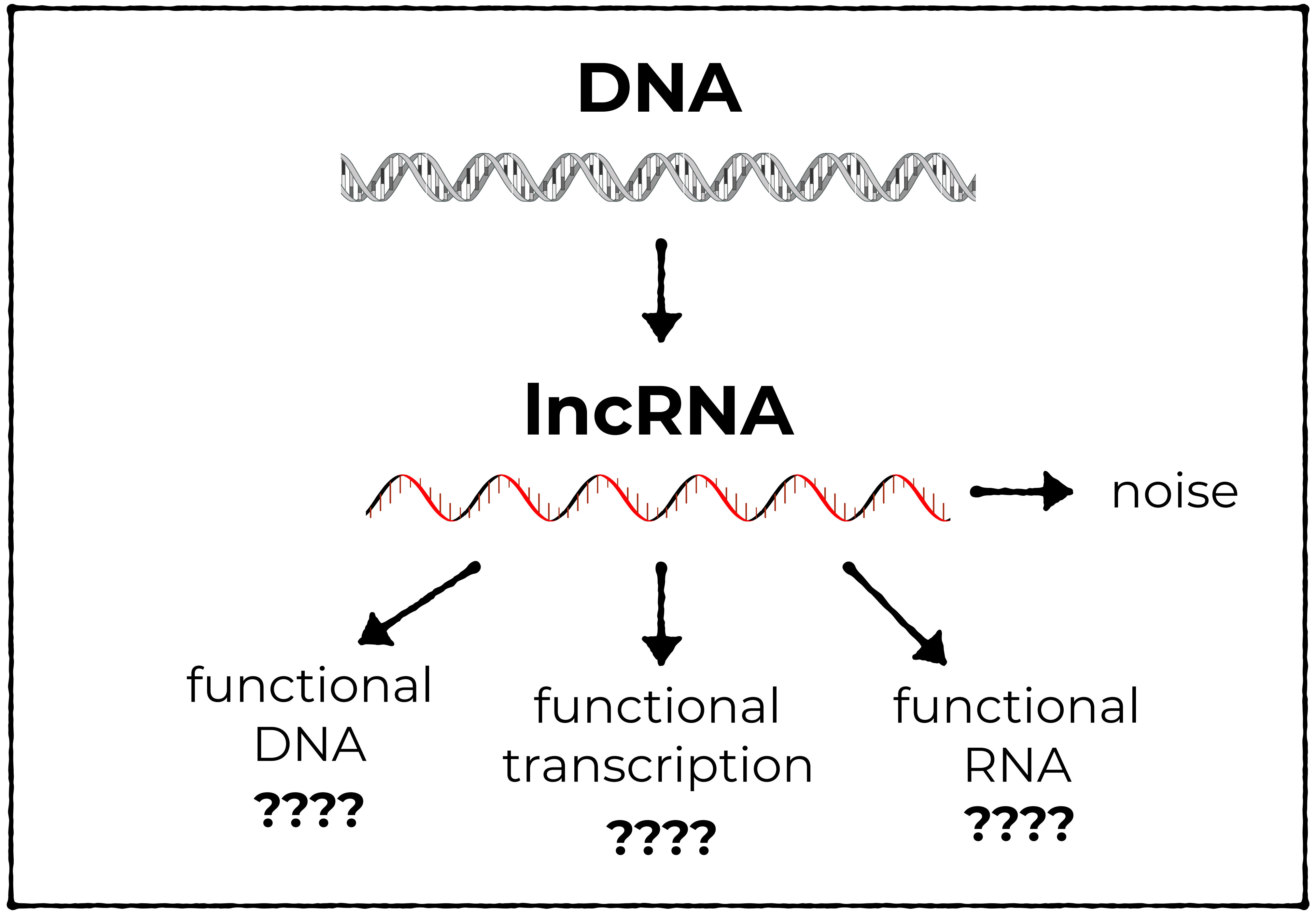
functional DNA

functional transcription

functional RNA

# Towards decoding the human genome





**DNA**

**lncRNA**

noise

functional DNA
????

functional transcription
????

functional RNA
????

# Acknowledgements

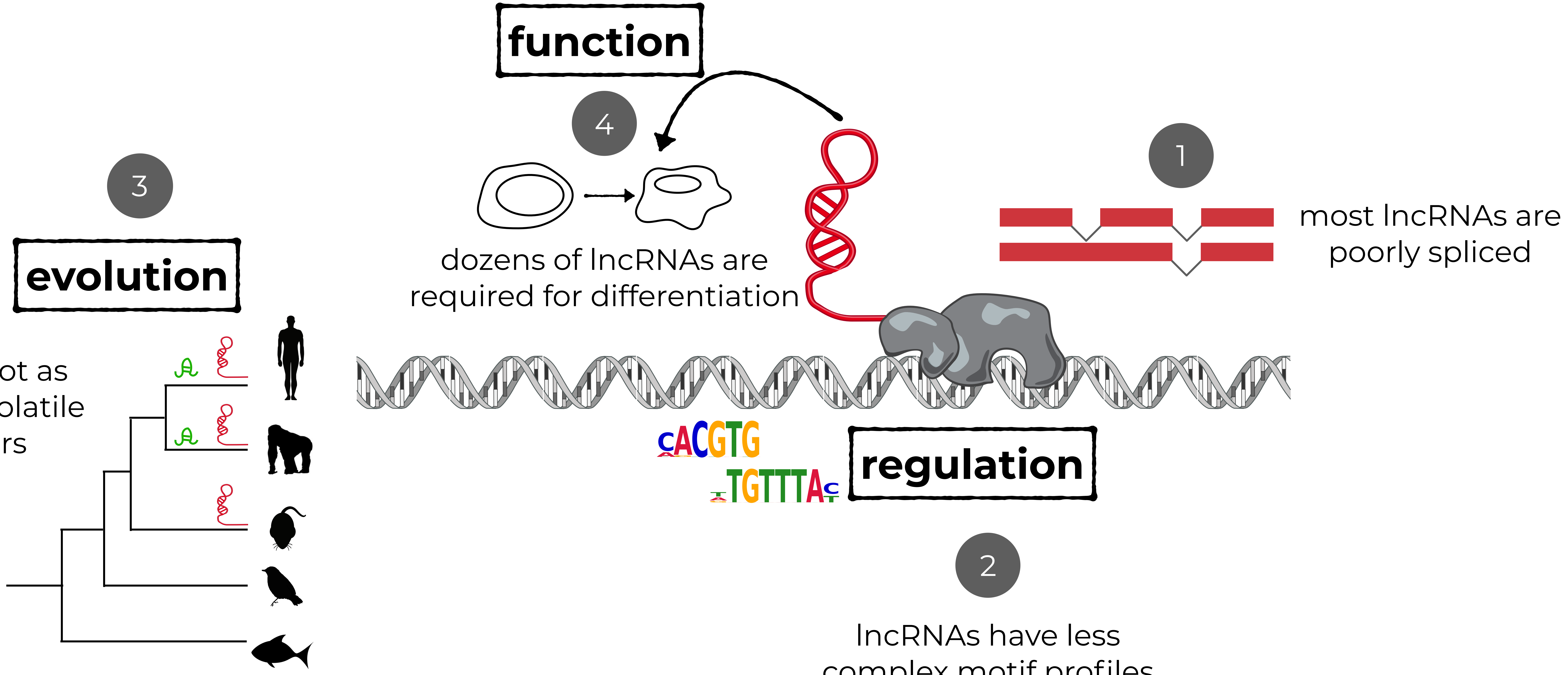
**John Rinn**
Chiara Gerhardinger
Philipp Maass
PJ Volders
Daniel Andergassen
Jordan Lewandowski
James Lee
Will Mallard
David Shechner

**Frank Slack**
Jeff Haswell

**DAC**
Martha Bulyk
Allon Klein
Shirley Liu

**Marta Melé**
Winona Oliveros

# Questions?



1

most lncRNAs are
poorly spliced

**regulation**

2

lncRNAs have less
complex motif profiles

3

**evolution**

lncRNAs are not as
evolutionarily volatile
as enhancers

**function**

4

dozens of lncRNAs are
required for differentiation