Trieste, October 11, 2024

Genome regulation by long non-coding RNAs

Piero Carninci

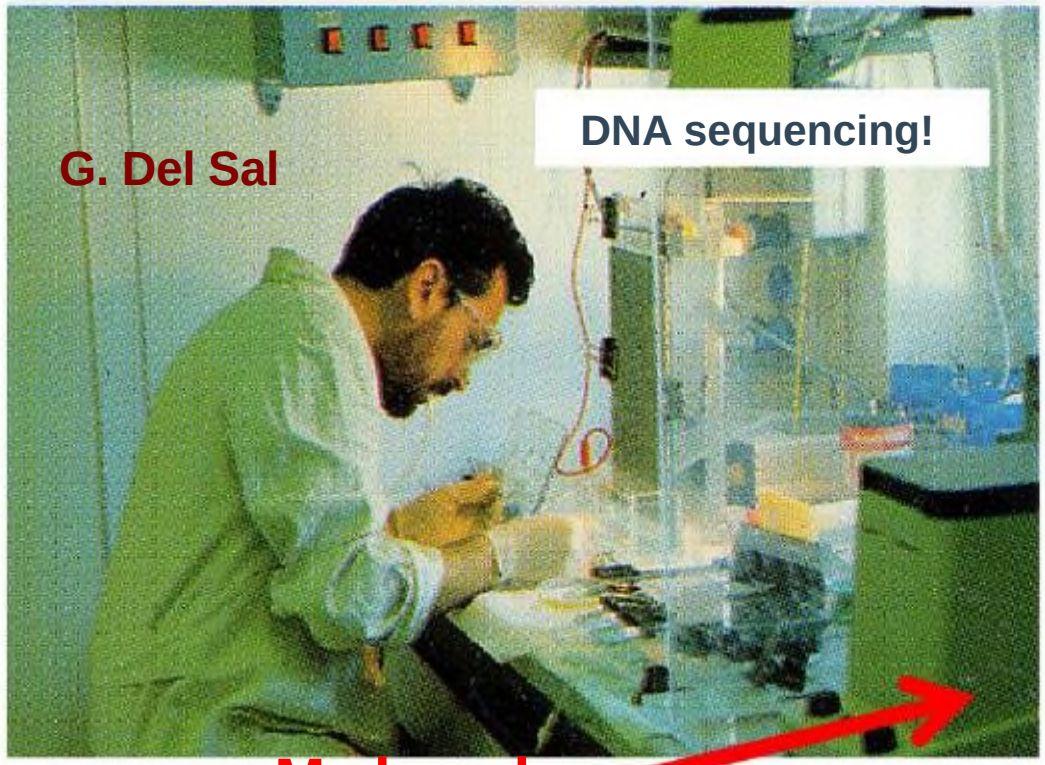
Human Technopole
Head of the Genomics Research Centre

RIKEN Center for Integrative Medical Sciences (IMS)
Team Leader

Emails: carninci@riken.jp, piero.carninci@fht.org
Twitter: @carninci

Trieste, Italy

ICGEB (AREA science Park)

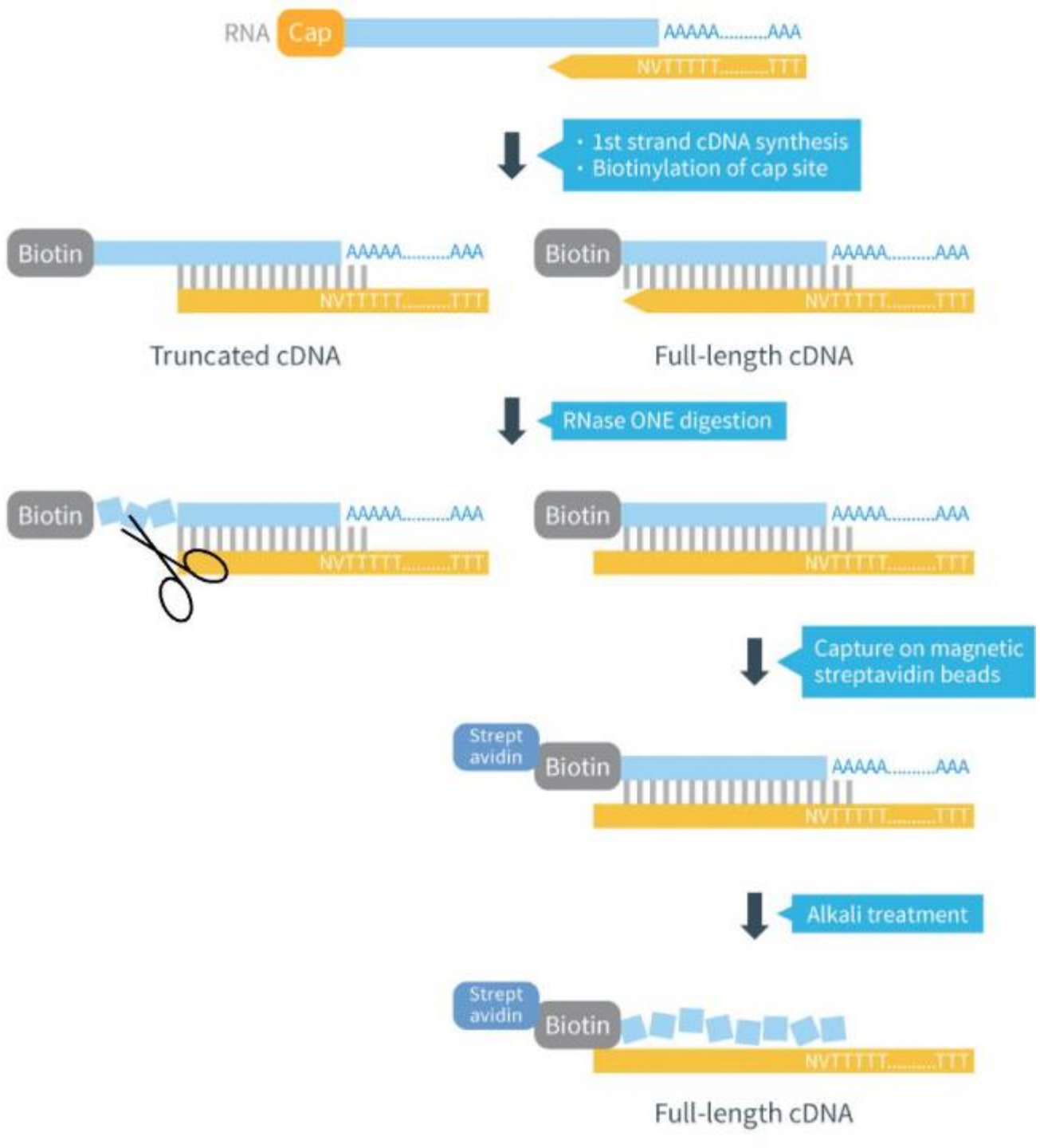


Lessons Learned

Technologies necessaries for revolutions

The Cap-trapper

- Cap chemical biotinylation
- cDNA synthesis
- Digest incomplete cDNAs
- Selecting only cDNAs that reach the cap site



Where to go?

1990-95 Discussion around the Genome Project

- All discussions on how to sequence the genome.
- How could possibly RIKEN contribute significantly in a such large project?

1995: started technology development, 1998 started data production for the:

“Mouse genome encyclopedia project”

Development of a series of technologies.

- Full length cDNA technologies
- A large scale sequencing system
  384 Capillary sequencer
  40000 clones/ day capacity of plasmid preparator

How to deal with lots of sequences??

- >103,000 cDNA sequences
- How to analyze so much data?
  - Call your friends
  - Make new friends
  - Bold proposal: invite many new friends to look at the data
- FANTOM Project:

  Functional ANnoTation Of Mammalian Genome


WHAT is FANTOM?
Work hard!
Think hard!
Discuss hard!

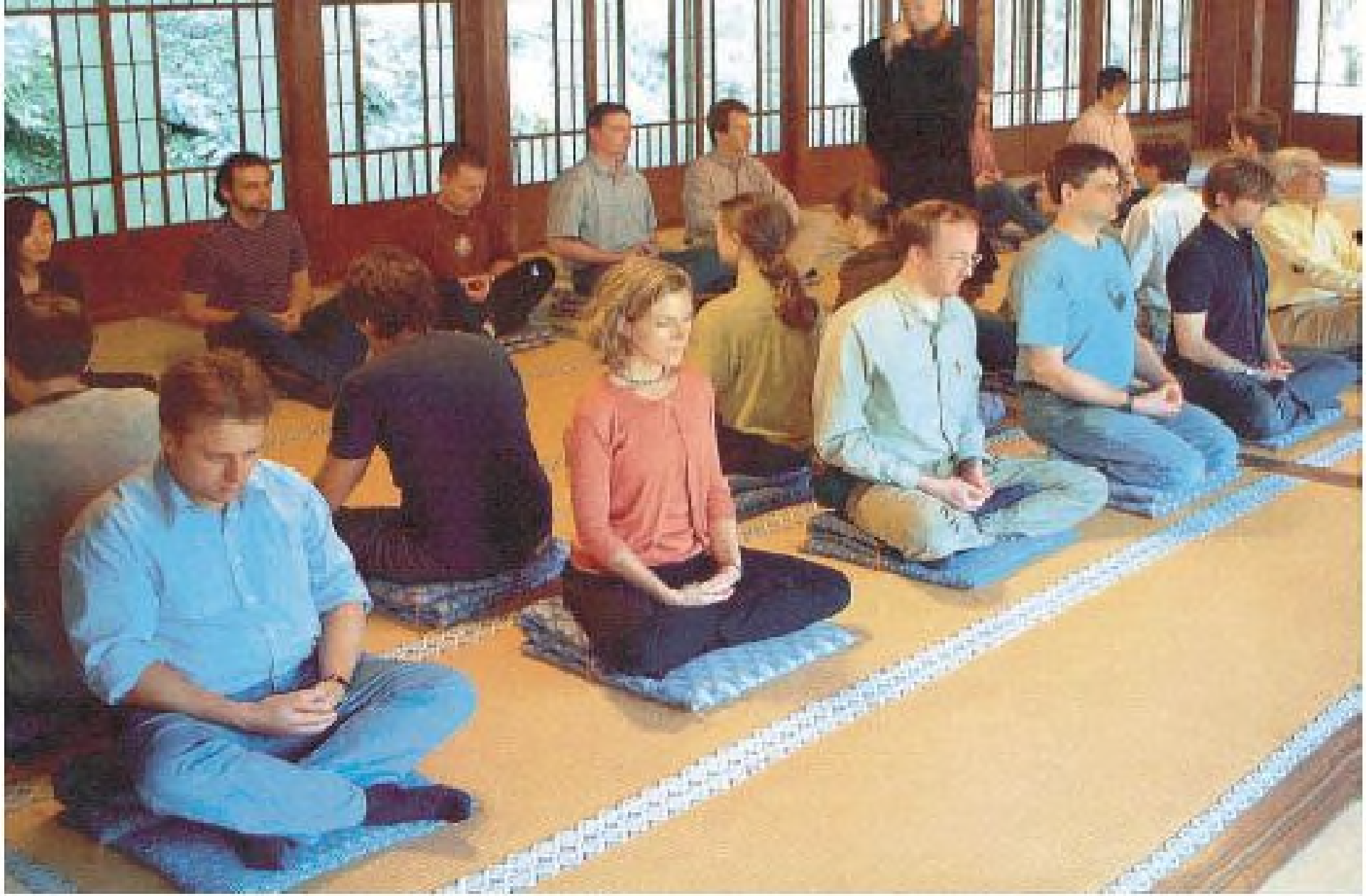
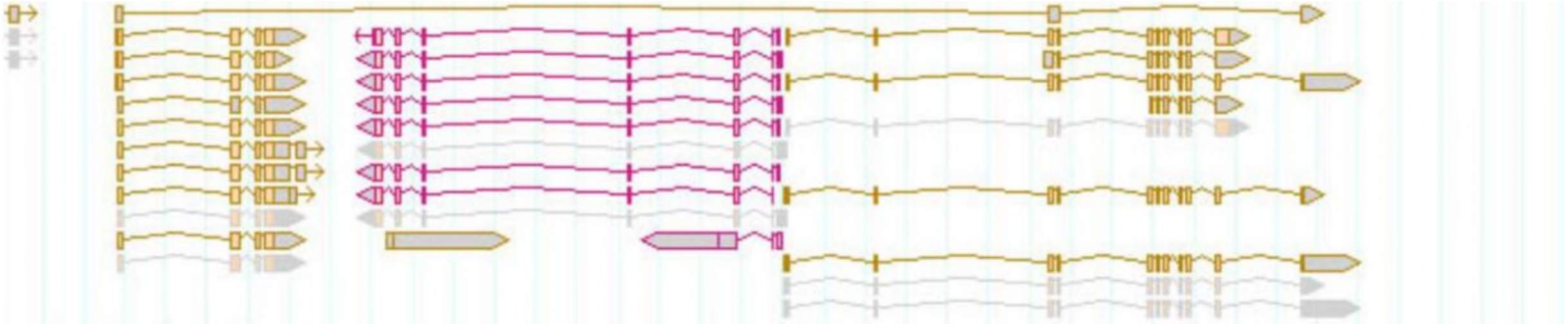


Figure I A shot from the Zen meditation ceremony held as an excursion during the FANTOM2 Cherry Blossom Meeting. The Zen meditation was a good break and p rovi ded the partici pants with novel inspi ration.

Unexpected discoveres

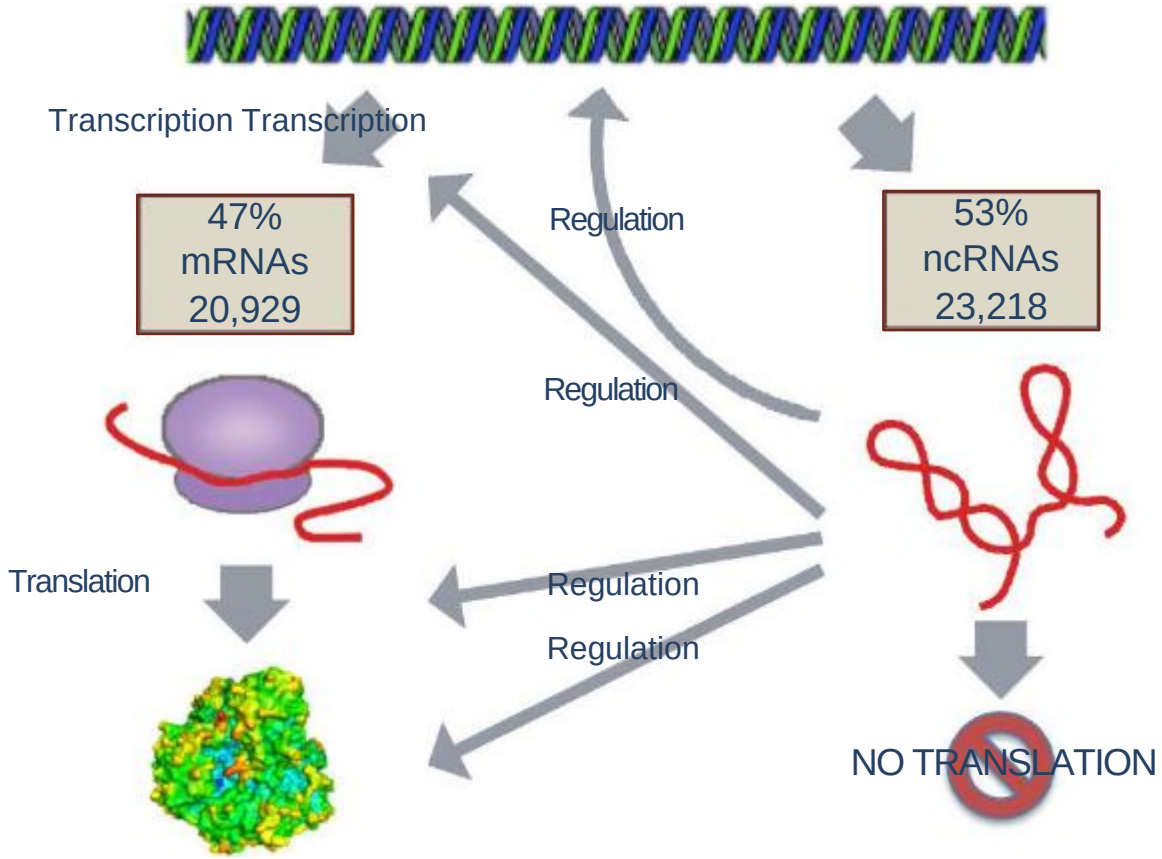
There is much more than what is in textbooks

Expectations ← ??? → Data



Discovery of the “RNA continent”

Unti! FANTOM2 in 2002, on!y ~ 100 ncRNAs genes were reported except tRNAs and rRNAs. More than ha!f of the genes missing from the gene maps ti!! FANTOM2, 3



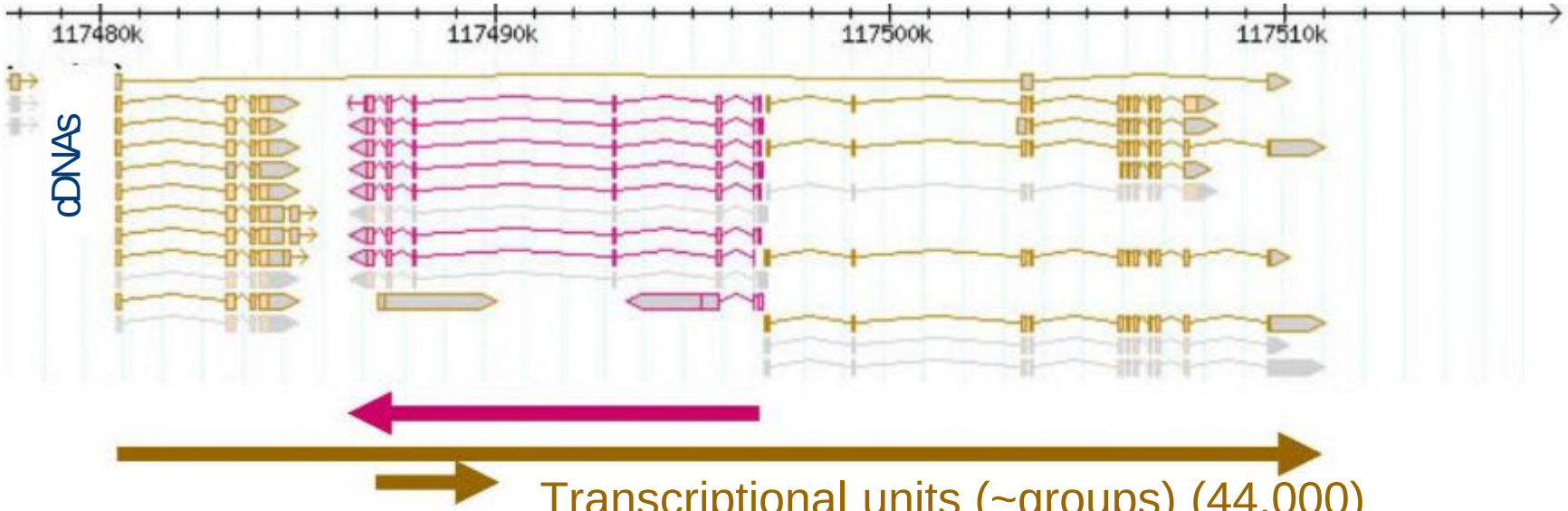
FANTOM3 revealed:

- 63% of the genome is transcribed into lncRNAs
- More than half of the transcripts is lncRNAs
- 73% of gene shows antisense transcription

Okazaki ***et a!. Nature*** (2002), Carninci ***et a!. Science*** (2005)

Gene fusion and merging forests

Retinol dehydrogenase 1 Retinol dehydrogenase 9



Transcriptional units (~groups) (44,000)

Gene forest (<19,000)

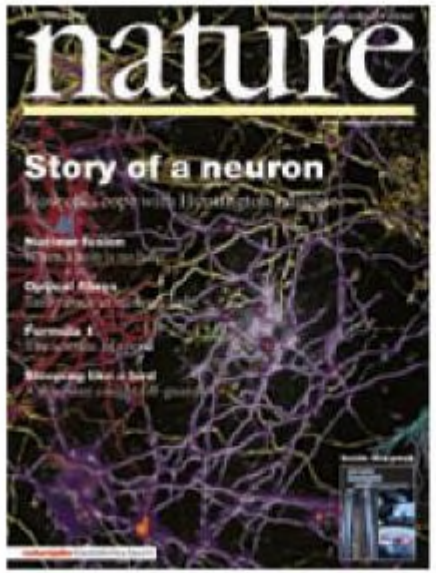
Gene desert

The new transcriptional landscape

lncRNAs Odyssey, 2002- 2012 and beyond

- Some attacks
  - The RIKEN cDNA libraries are full of “junk”.
    - -) Discovery of lncRNAs
  - The RIKEN library are full of cDNA in the wrong orientation. -)Discovery of antisense RNAs.
- Papers attacking the findings
- Papers supporting the findings
- Subsequent papers confirming the findings

Papers criticizing our findings



Jun Wang *et al*. *Nature* 431, (2004).

doi:10.1038/nature03016

1. coding 1 (probably protein)
2. coding 2 (marginal protein)
3. non-coding 1 (marginal RNA)
4. non-coding 2 (probably RNA)

Mouse transcriptome

Neutral evolution of 'non-coding' complementary DNAs

Arising from: Y. Okazaki *et al. Nature* **420,** 563–573 (2002)

Okazaki *et al.* have argued that as many as 15,815 of 33,409 non-redundant mouse complementary DNAs may represent functional RNA genes[1], on the basis of their findings that some of these cDNAs are confirmed by expressed sequence tagging and are found near CpG islands or polyadenylation signals[2] — although many are expressed at such low levels that they could not be detected by microarray analysis[3]. We show here that conservation of these 'non-coding' cDNAs in rats or humans is no better than in an evolutionarily neutral control. Our results indicate that they are either non-functional or, if they are functional, are specific to a given species.

We downloaded FANTOM release 2.0 cDNAs from the authors' website. Table 1 shows the data from the four categories defined by the authors, which we refer to as coding 1 (probably protein), coding 2 (marginal protein), non-coding 1 (marginal RNA), and non-coding 2 (probably RNA). Overall transcript sizes average about 2 kilobases (kb) in each category; most known RNA genes are much smaller than this — for example, the 587 mouse entries in the Rfam database[4] average 96 base pairs (bp) in length. Larger RNA genes do exist (such as *H19* and *Xist*) and many are stored in the Erdmann database[5]. Another striking difference between the given categories is the increase from 13.4% single-exon genes in coding 1 to 68.7% and 73.1% single-exon genes in non-coding 1 and non-coding 2, respectively.

As an evolutionarily neutral control, we use 'intergenic' sequences of 2 kb in length that are at least 5 kb distant from genes annotated by Ensembl, predicted by FgeneSH, or aligned to cDNAs. Transposons identified by RepeatMasker are excluded, as is the 5% of

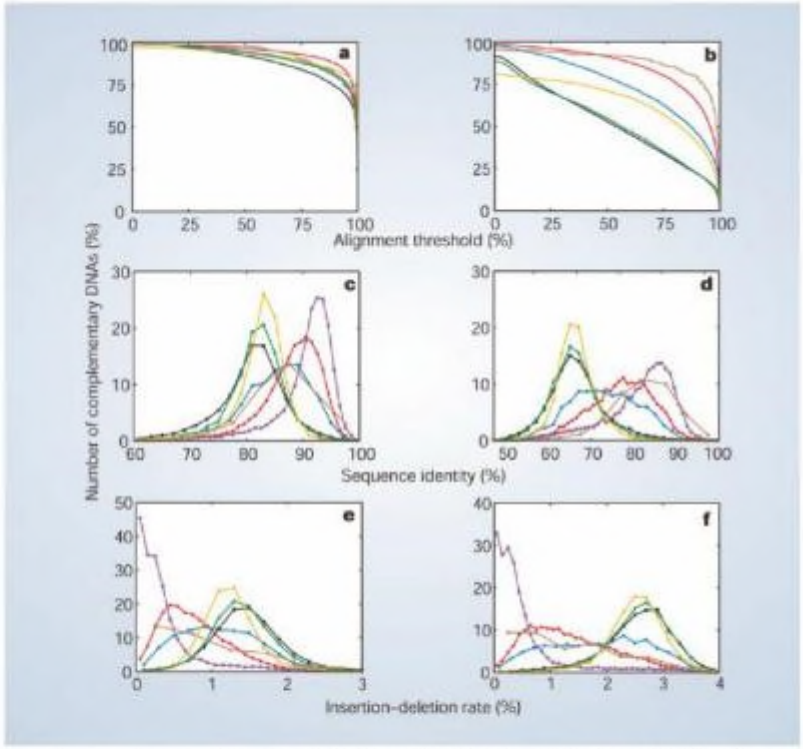
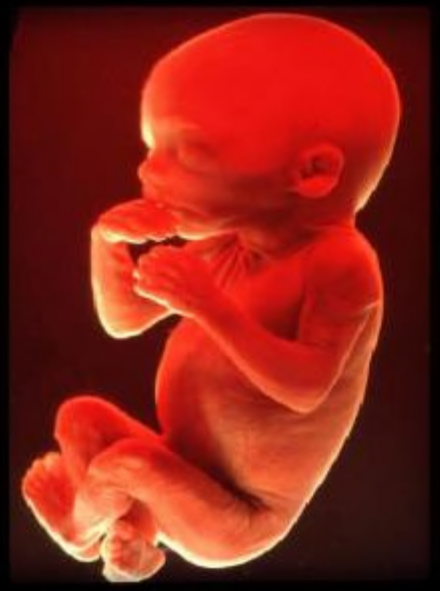


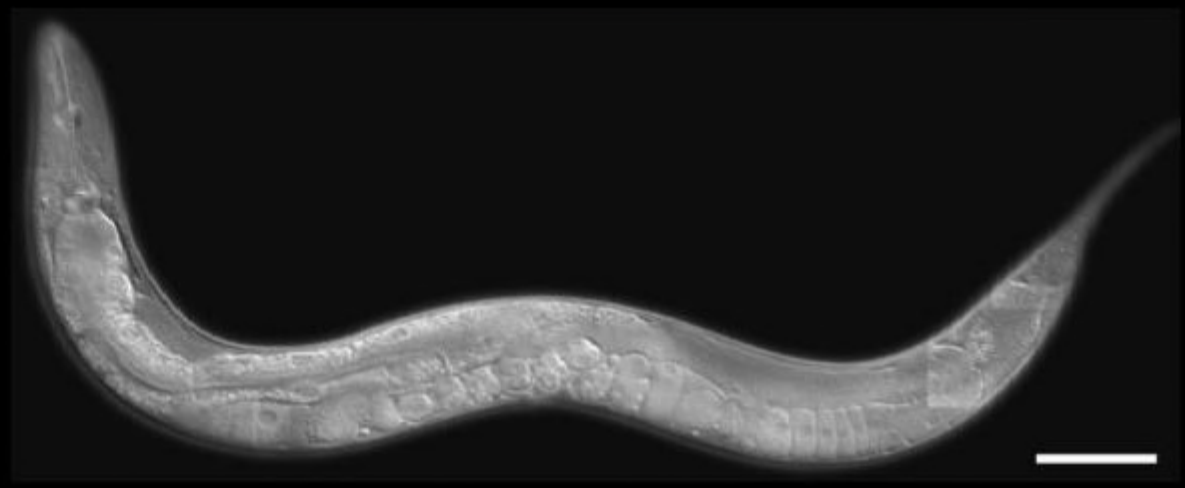
Figure 1 Comparisons between rat (left) and human (right) data. **a, b,** The number of good alignments. **c–f,** Distribution of sequence identities (**c,d**) and insertion–deletion rates (**e,f**) restricted to the good alignments. Each solid dot shows the centre of the bin over which signals were averaged. Red, coding 1; blue, coding 2; black, non-coding 1; green, non-coding 2; brown, ncRNAs; and yellow, intergenic. For panels **c** to **f**, a purple line is added for the CDS region of coding 1.

Table 1 **Other attributes of mouse complementary DNAs**

	FANTOM categories				Control data sets	
	Coding 1	*Coding 2*	*Non-coding 1*	*Non-coding 2*	*ncRNAs*	*Intergenic*
No. of cDNAs	14,317	3,277	11,526	4,280	321	3,450
No. in a single exon	13.4%	35.4%	68.7%	73.1%	90.7%	100%

The genetic basis of developmental complexity

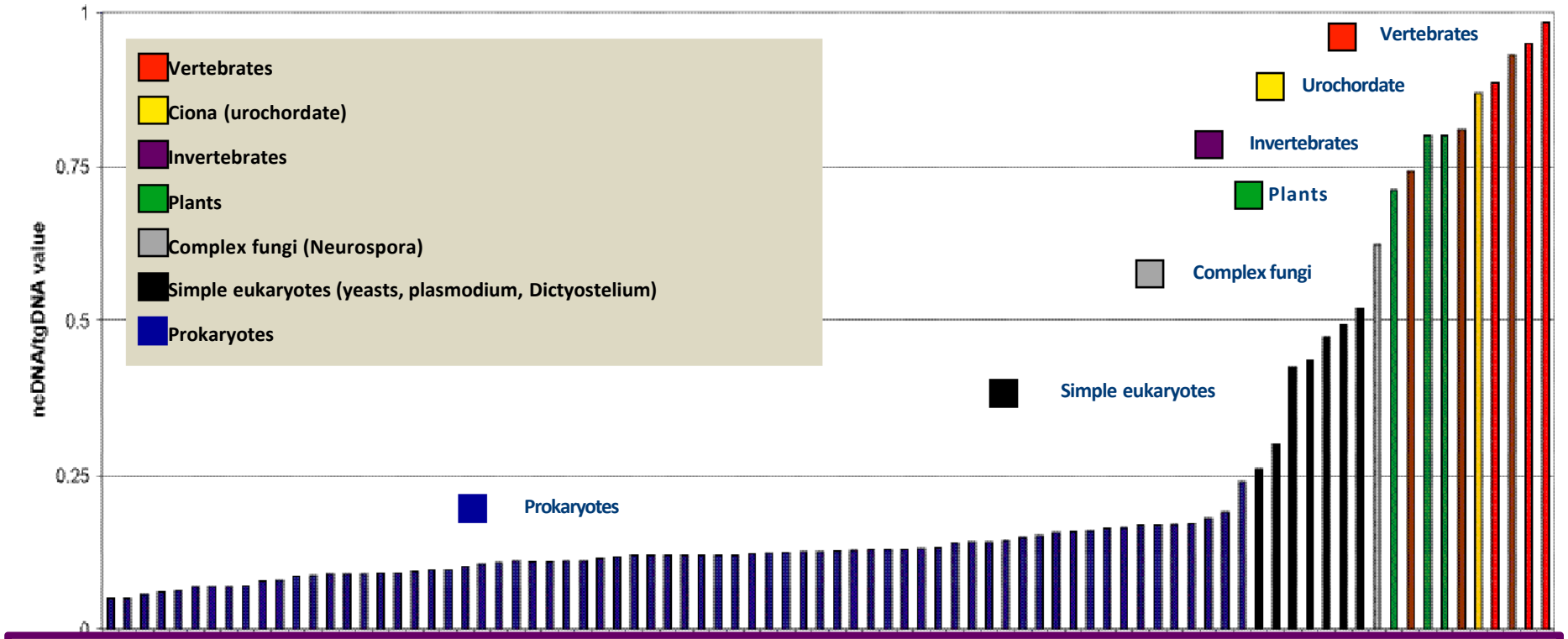




- Humans (and other vertebrates) have approximately the same number of protein-coding genes (~20,000) as *C. elegans*.
- Most of the proteins are orthologous and have similar functions from nematodes to humans, and many are common with yeast.
- Where is the information that programs our complexity?

Slide courtesy of John Mattick

The proportion of noncoding DNA broadly increases with developmental complexity



Irrespective of the extent of non-coding sequences, it is now evident that the vast majority of the genomes of all organisms is transcribed in a dynamic manner in different cells and tissues at different developmental stages.

J.S. Mattick *Nature Reviews Genetics* 5, 316-323 (2004)
R.J. Taft, M. Pheasant and J.S. Mattick, *Bioessays* 29, 288-299 (2007)

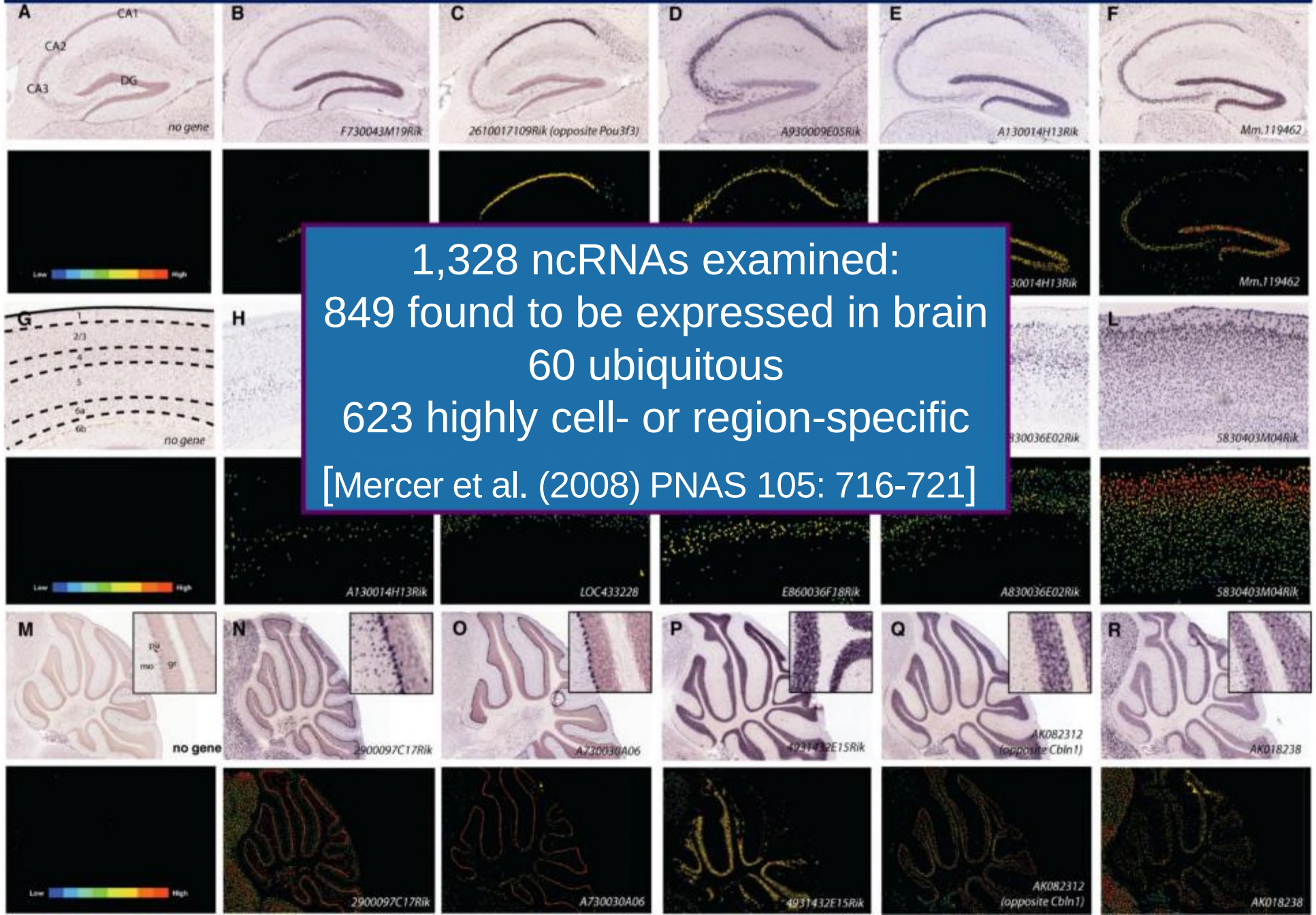
lncRNA seem to be essential

- **Theoretical basis: regulators of very complex operations.**

- **More validations:**
- **Are they expressed and where?**
- **Can we prove the function?**

- **Some beautiful and meaningful picture of long-noncoding RNA (lncRNAs) and their localization.**

Non-coding RNA expression in mouse brain



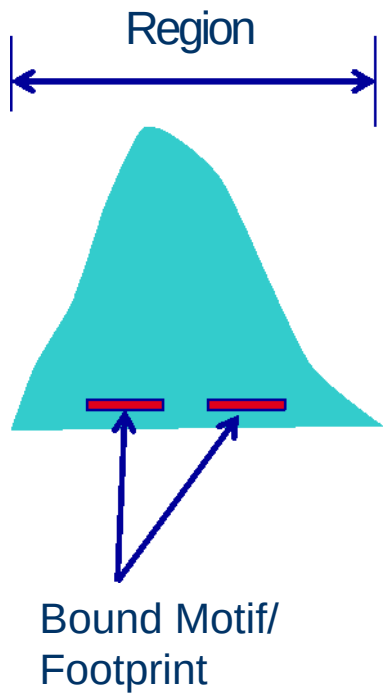
1,328 ncRNAs examined:
849 found to be expressed in brain
60 ubiquitous
623 highly cell- or region-specific

[Mercer et al. (2008) PNAS 105: 716-721]

✓ Subsequent papers confirming the findings

ENCODE

Element Type	Coverage	Cumulative Coverage
Exons	3%	3%
Chip-seq bound motifs	4.5%	5%
DNaseI Footprints	5.7%	9%
Chip-seq bound regions	8.1%	12%
DNaseI HS regions	15.2%	19.4%
Histone Modifications (*)	44%	49%
RNA	**62%**	**80%**
(* excluding broad marks)		



(Union over all experiments and cell types)

In 2012, the ENCODE confirms that genome is broadly transcribed

ENCODE Project Consortium, Nature 489, 57 (2012)

✓ Subsequent papers confirming lncRNAs

Large intervening noncoding RNAs (lincRNAs)

> “Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals”

Mitchell Guttman, Ido Amit, Manuel Garber, Courtney French, Michael F. Lin, David Feldser, Maite Huarte, Or Zuk, Bryce W. Carey, John P. Cassady, Moran N. Cabili, Rudolf Jaenisch, Tarjei S. Mikkelsen, Tyler Jacks, Nir Hacohen, Bradley E. Bernstein, Manolis Kellis, Aviv Regev, John L. Rinn & Eric S. Lander

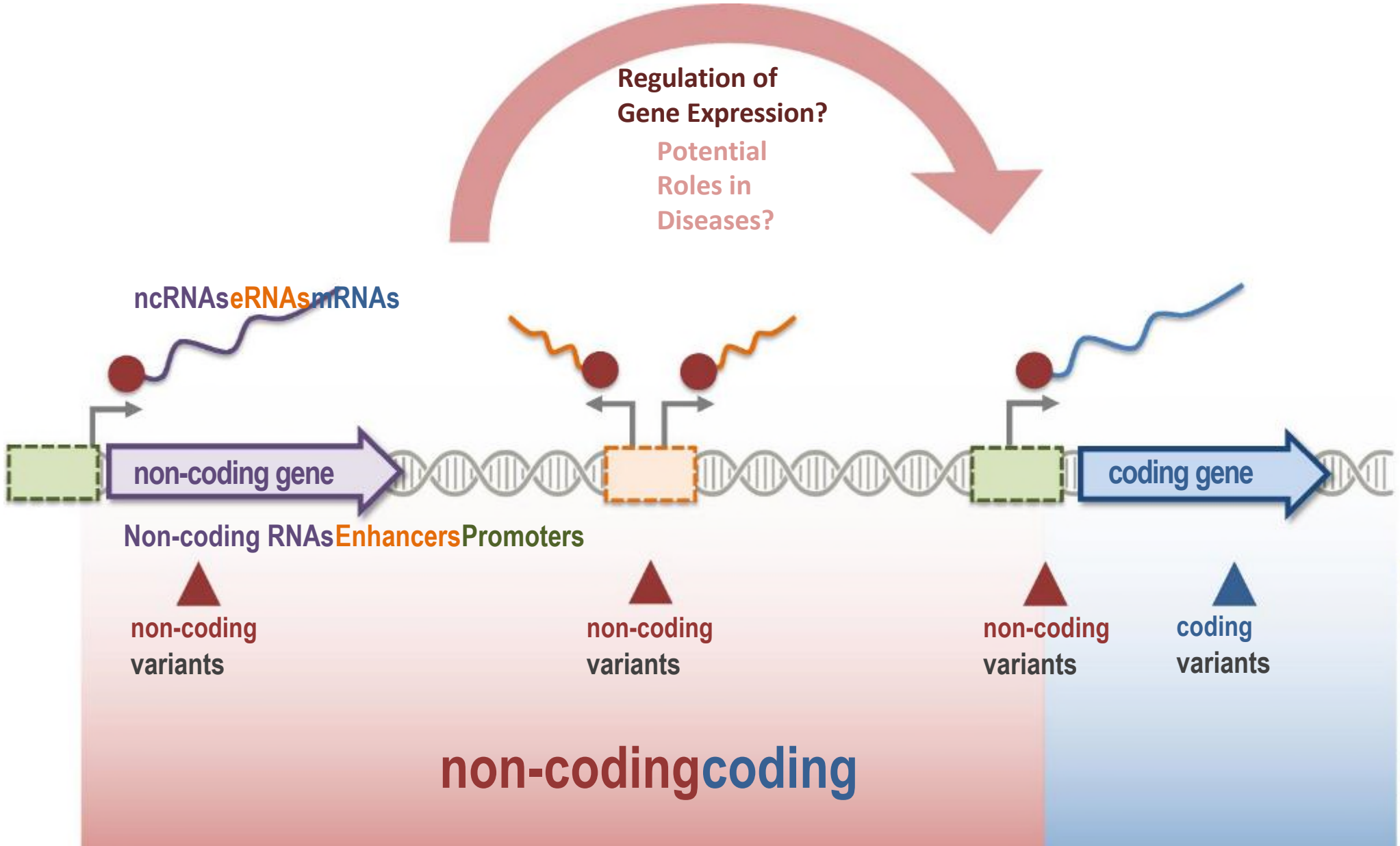
Nature 458, 223-227 (2009)

Second key technology: developed for identifying regulatory

elements

→ comprehensive mammalian regulatory elements

Mapping genome elements and their regulation

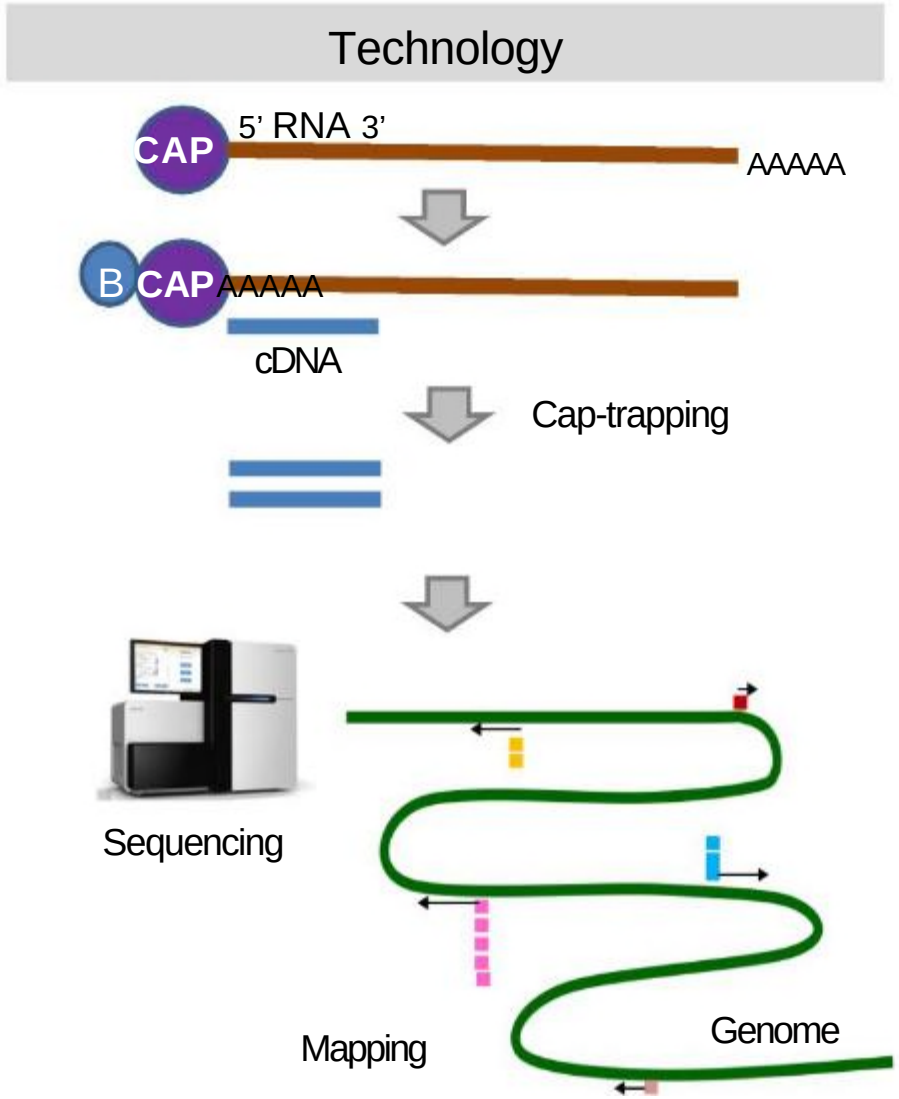


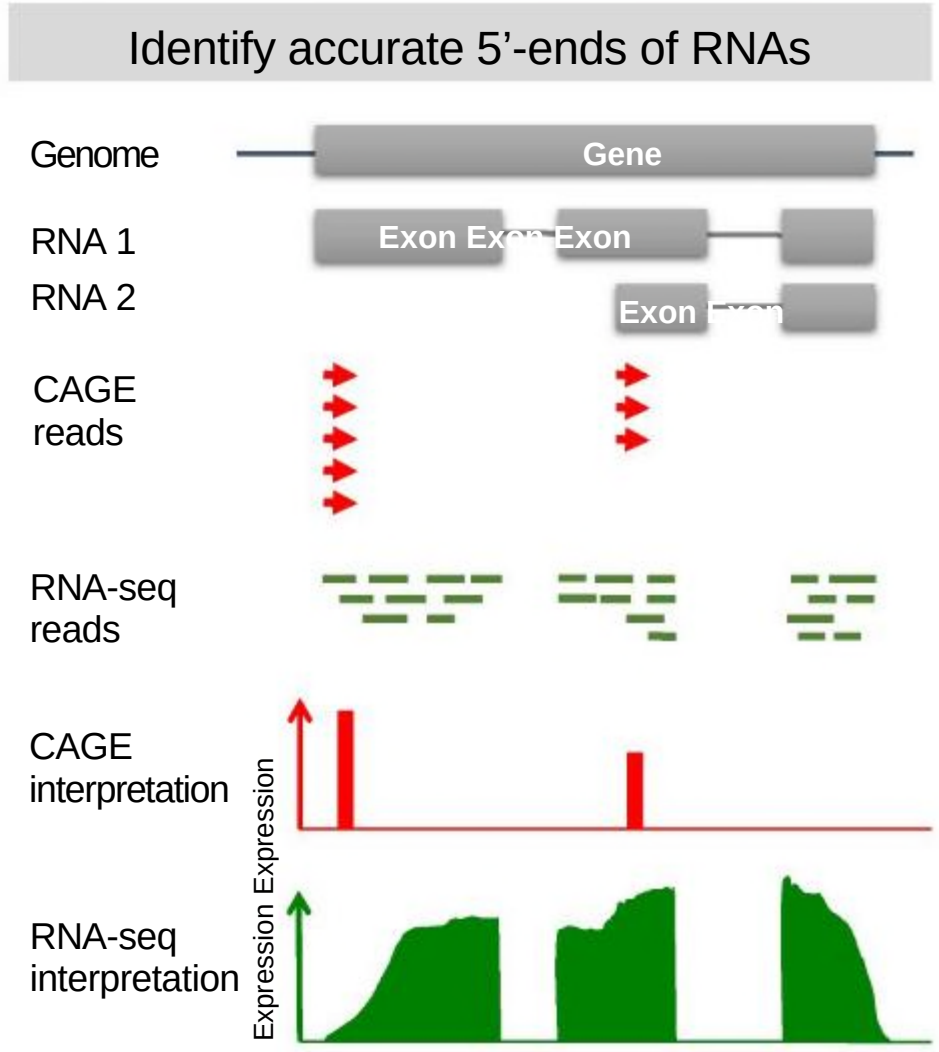
CAGE to broadly map promoters (and enhancers) → 5’ UTRs

- Precise identification of TSSs.
  - Quantitative analysis of TSS activity and promoter maps
- Genome-wide

CAGE

Cap Analysis of Gene Expression



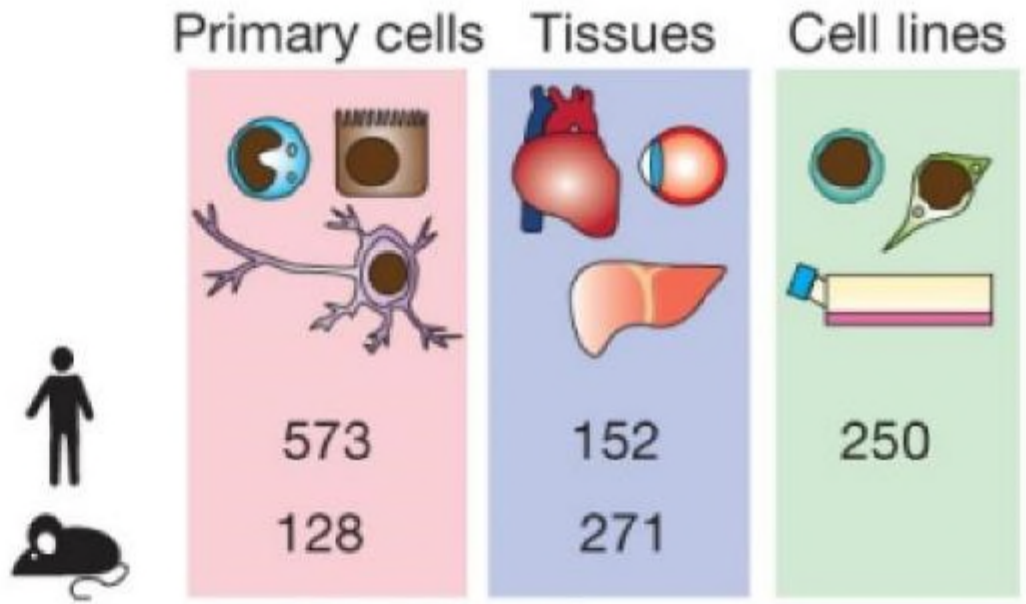


FANTOM5: *Regulatory elements in primary cell types & lncRNAs*

- 261 authors
- 120 international collaborators
- 19 countries

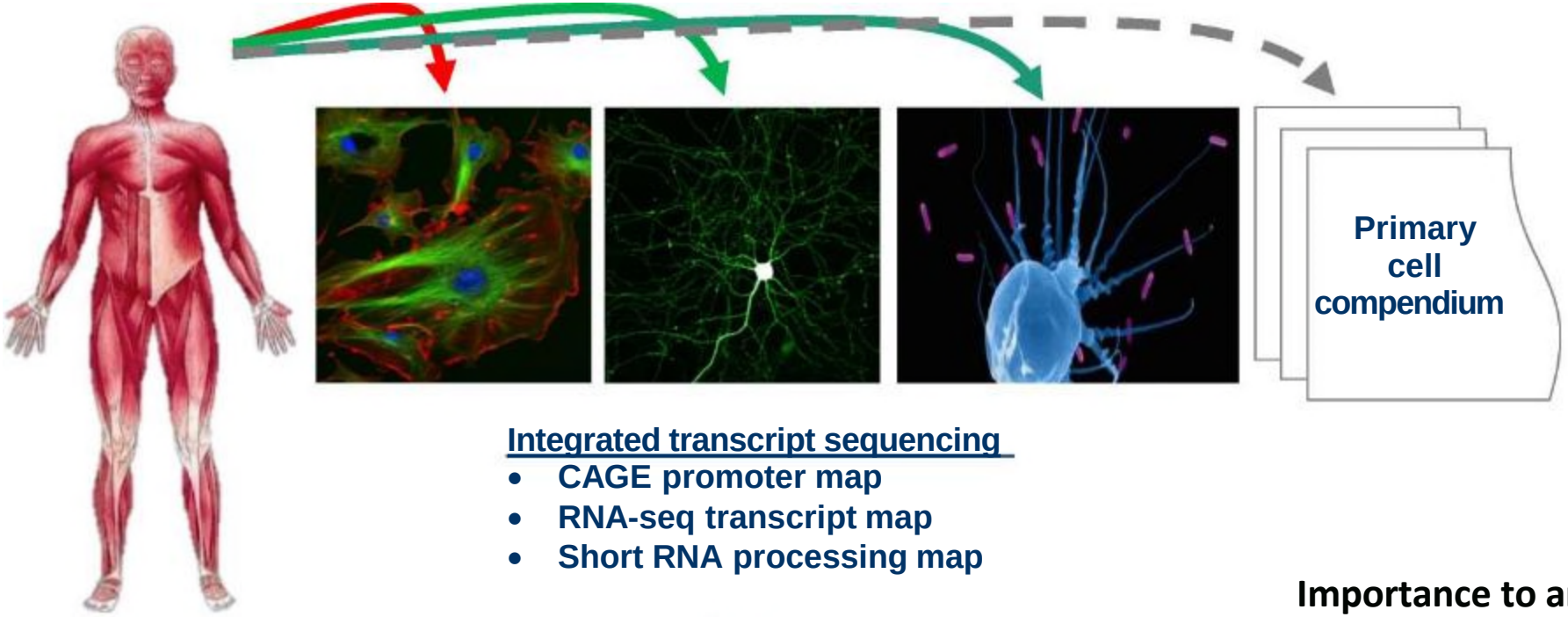
Samples: *Variety of cells*

~3,000 human and mouse libraries in total



Forrest *et al. Nature* 507, 462 (2014)

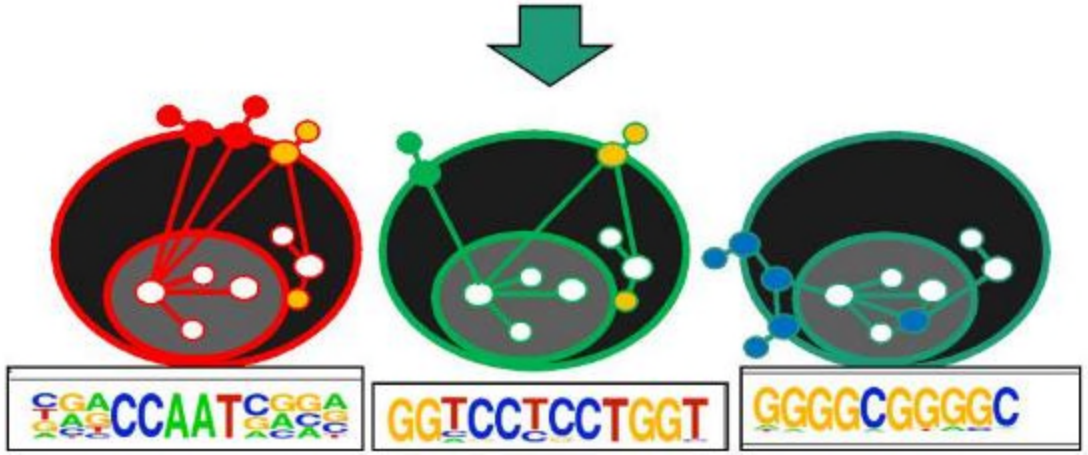
3000 CAGE libraries, many promoters, enhancers, networks



Integrated transcript sequencing

- **CAGE promoter map**
- **RNA-seq transcript map**
- **Short RNA processing map**

Importance to annotate TFBS and promoters



Cell specific network models **947 samples**

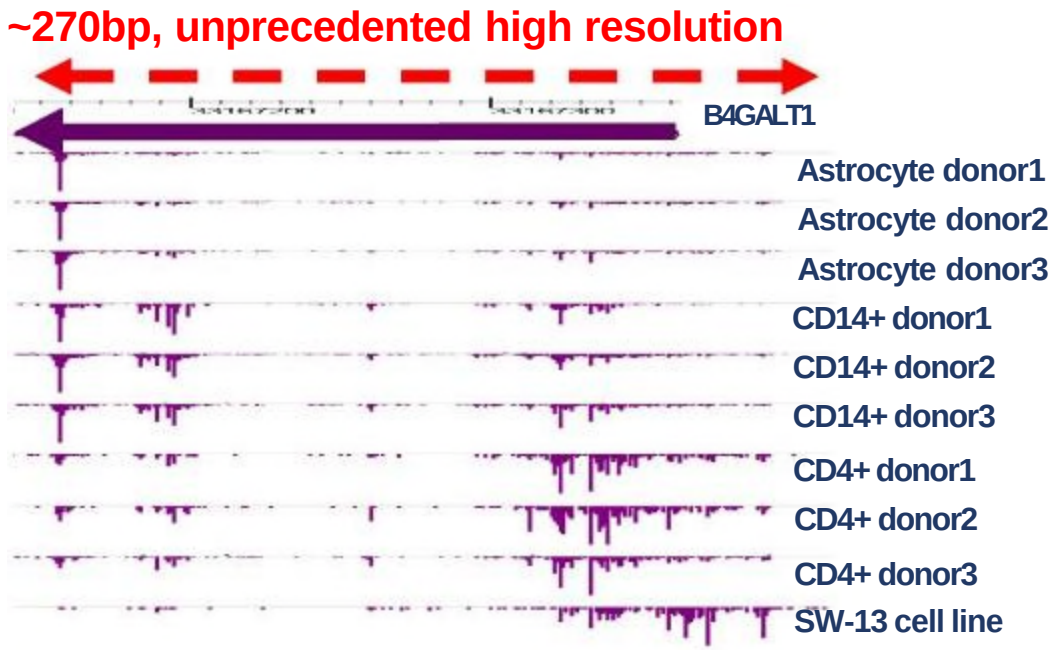
Key transcription factors, Key motifs Forrest *et al. Nature* 507, 462 (2014)

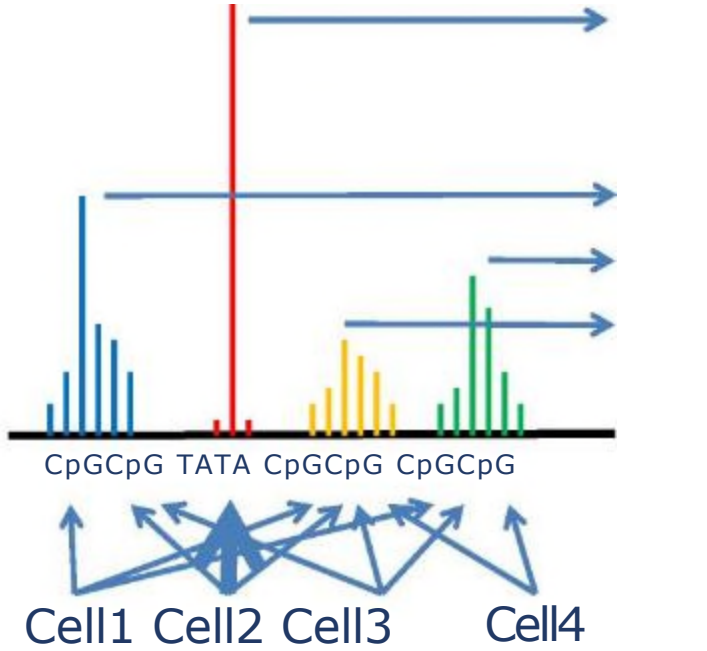
Expected and unexpected results

- Measure gene expression
- Infer networks globally
- Map promoters

- Surprises?

FANTOM5: *Promoter architectures differ in different cells*





TSS preferences:

- **B4GALT1 core promoter**
- **Primary Astrocytes**
- **CD14+ monocytes**
- **CD4+ T-cells**

- **Blue, yellow and green:**

Broadly used

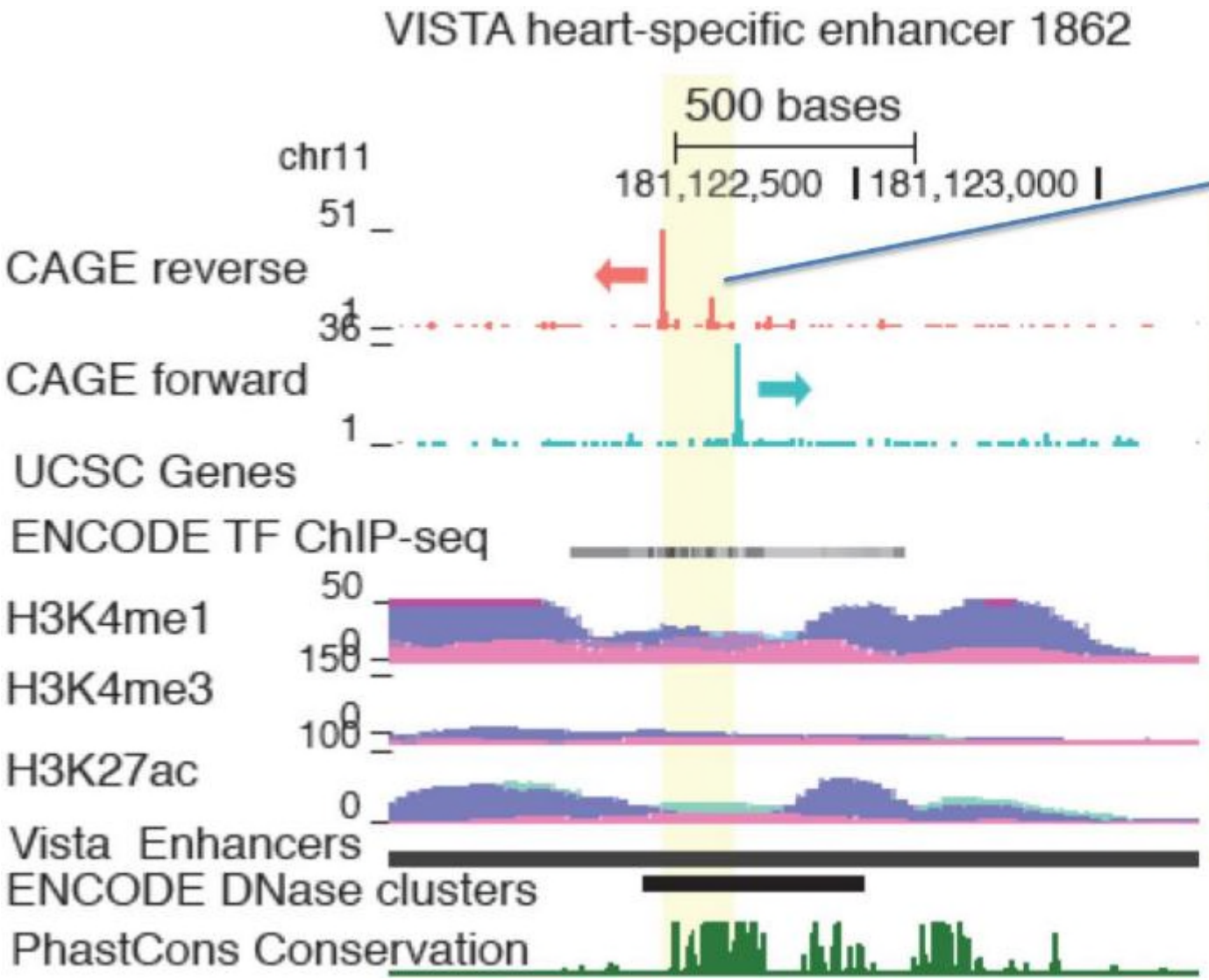
- **Red:**

Cell2 specific and highly expressed

223,428 in human and 162,264 in mouse of reference TSS

Forrest *et al. Nature* 507, 462 (2014)

CAGE instrumental to identify cell specific enhancers and eRNAs



Enhancers have bidirectional CAGE transcription.

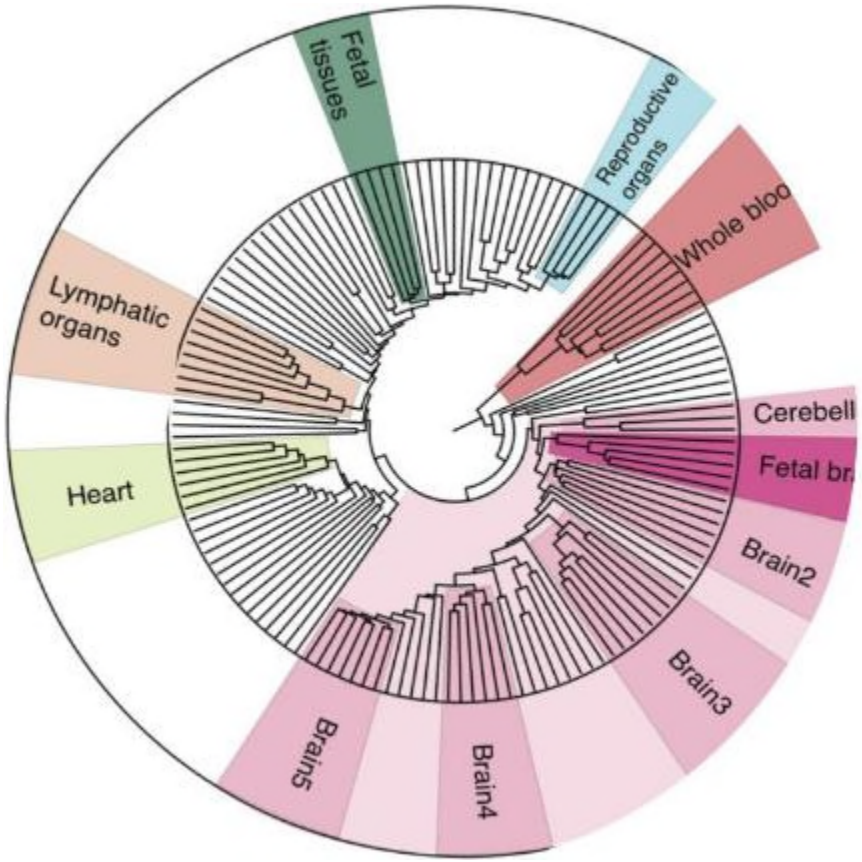
Bidirectional transcription identifies ~ nucleosome boundary.

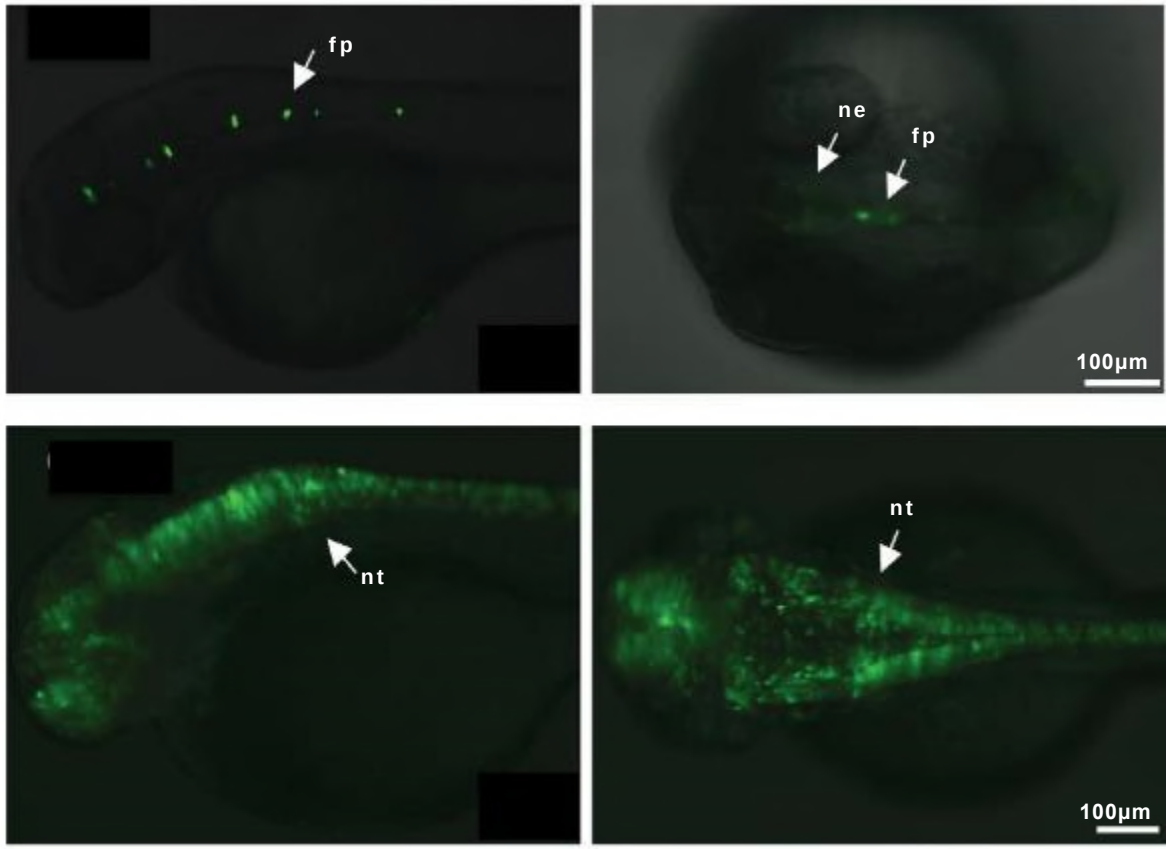
Based on this, we defined a rule to locate novel transcribed enhancers over the whole FANTOM collection.

eRNAs believed to be 200-300 nt lncRNAs

Andersson *et al. Nature* 507, 455 (2014)

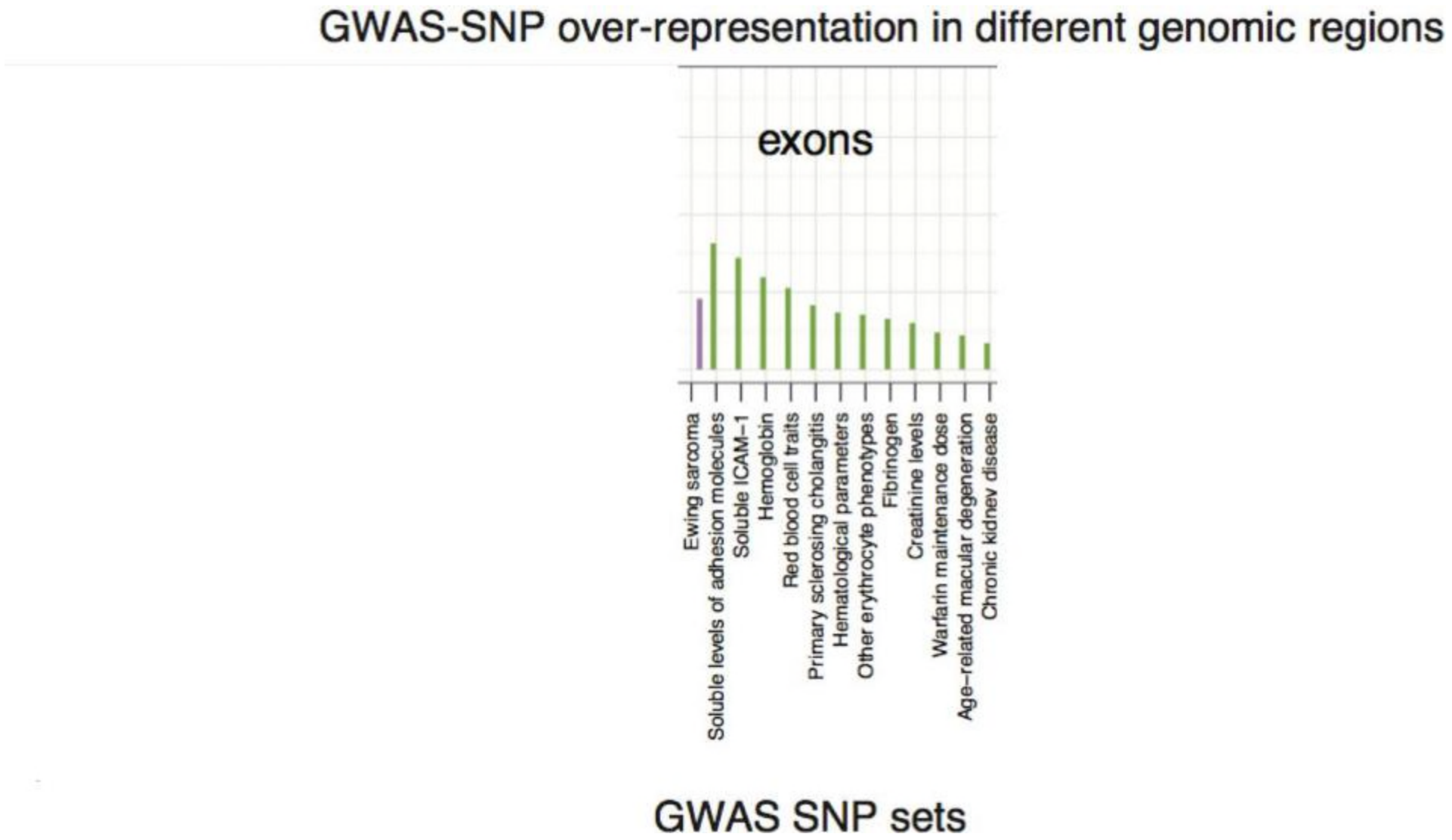
- identified 65,423 and 44,459 enhancers in human and mouse.
- 60% are over-represented in one cell/tissue group.
- Human GWAS SNPs map often on enhancers and promoters (less frequently on coding exons).
- Promoter and enhancer usage and QTL analysis. *Garieri et al. Nat Commun.* (2017)





Andersson *et al. Nature* 507, 455 (2014)

Disease-associated SNPs are enriched in enhancers

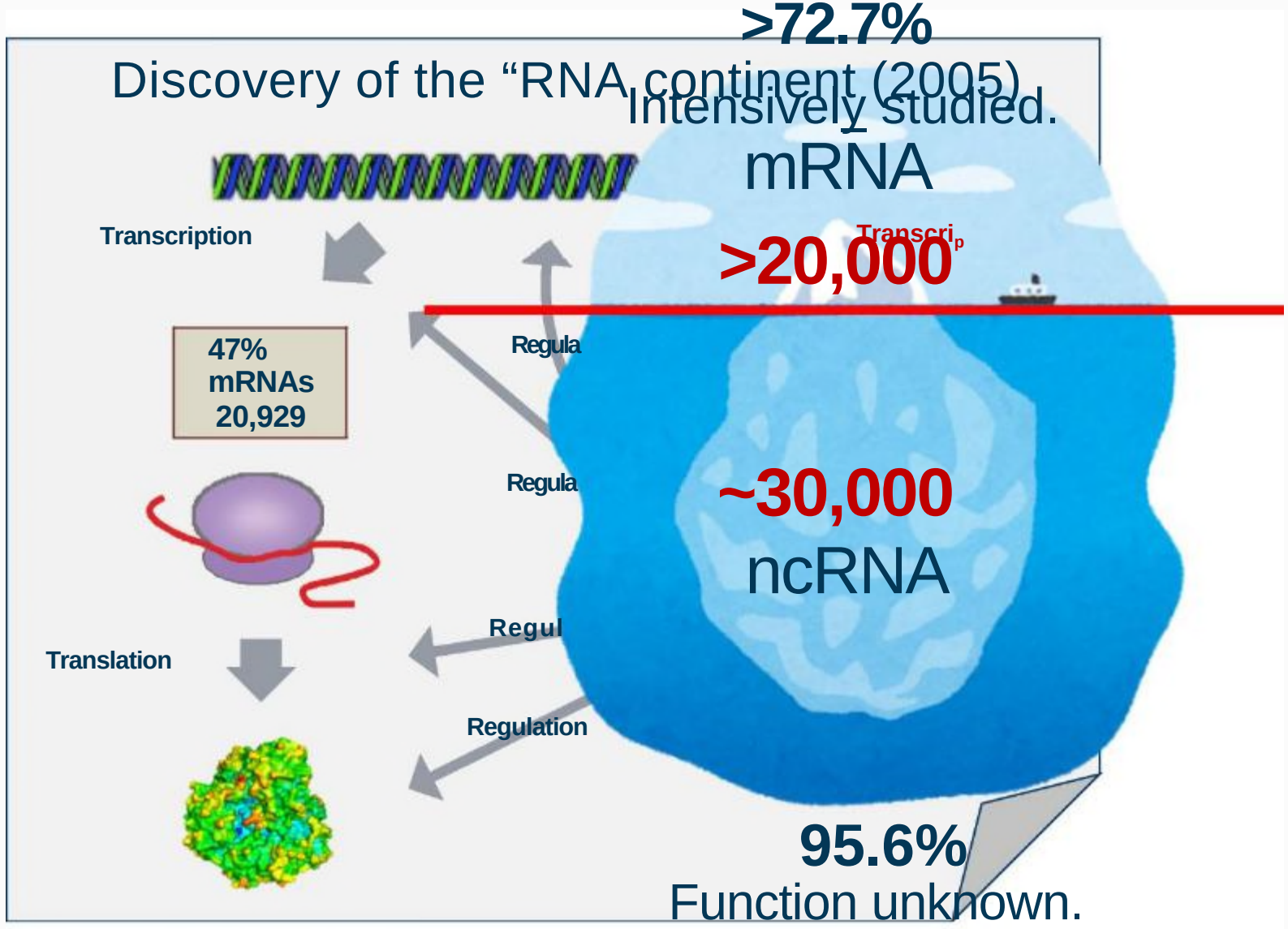


Mapping genome regions that regulate gene activity in diseases Andersson *et al. Nature* 507, 455 (2014)

Long “non-coding” RNAs

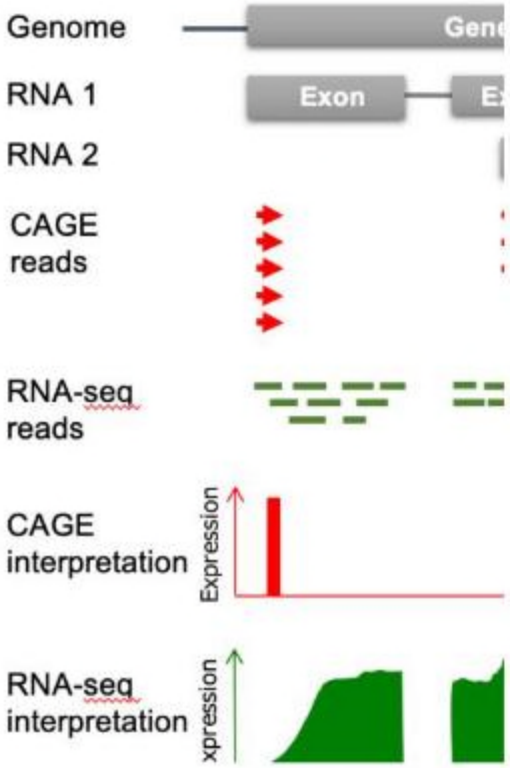
to be renamed as -
Regulatory RNA -
Structural RNAs
...

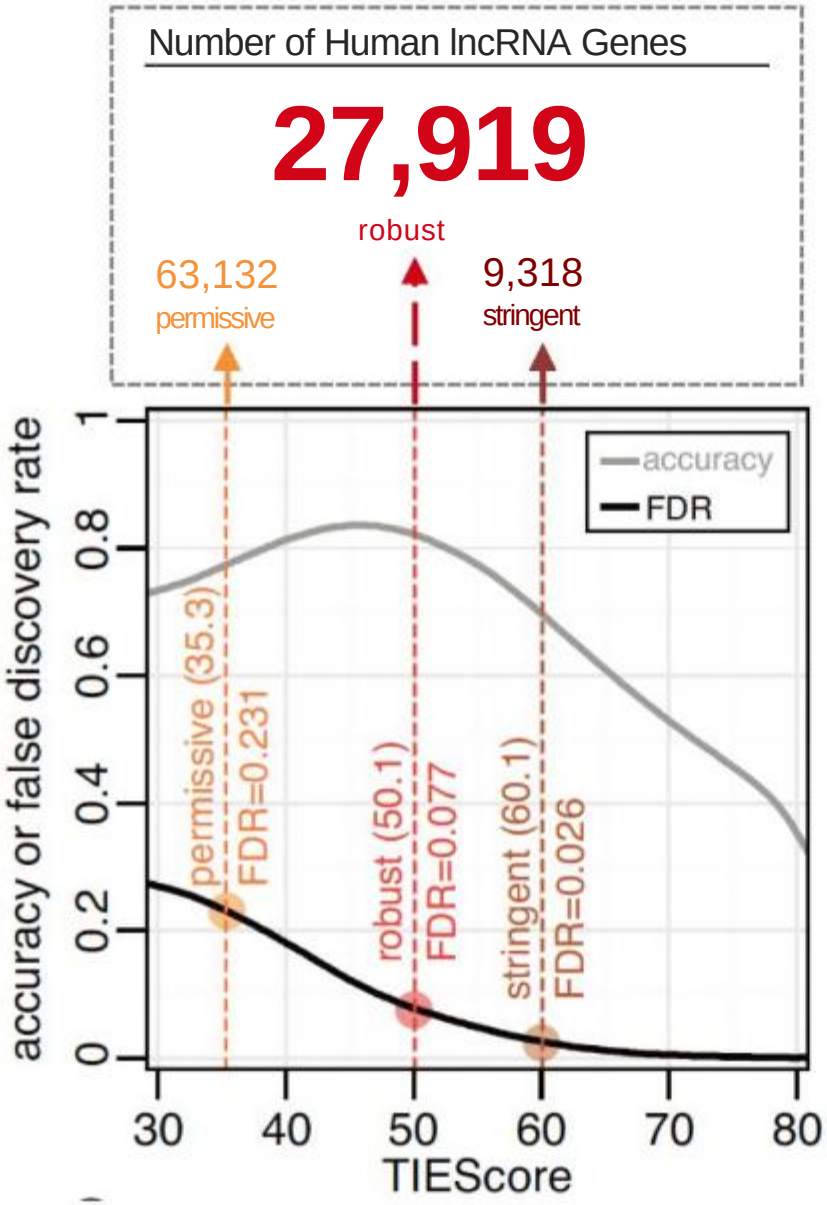
Discovery of “non-coding” RNA (FANTOM-3, ~2005)



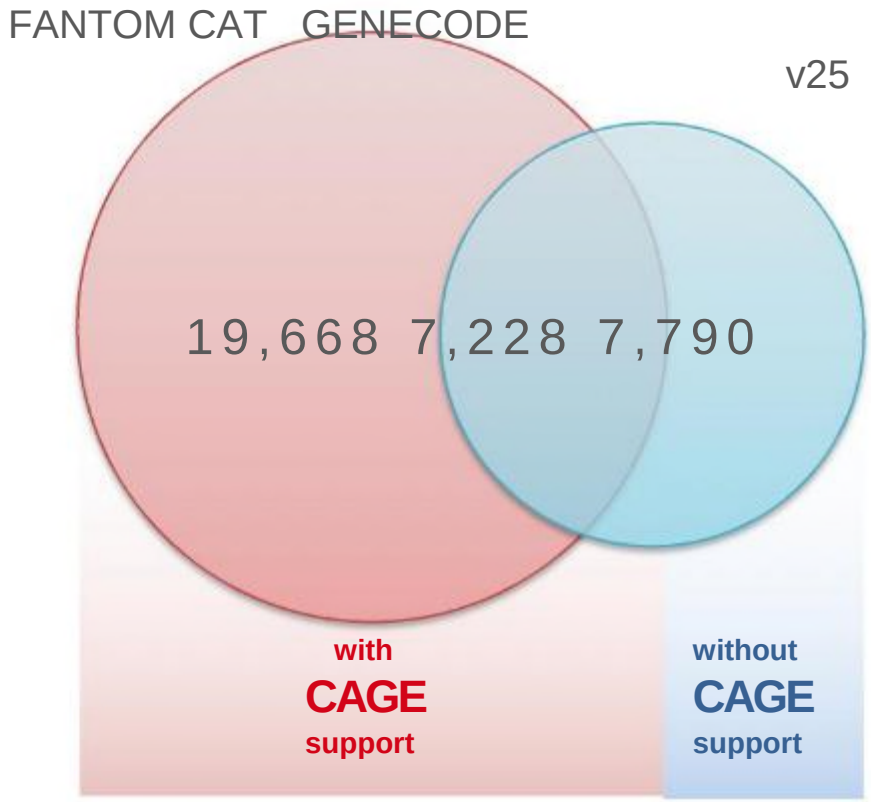
Carninci et al, Science, 2005

Human lncRNA: FANTOM CAGE Associated Transcripts





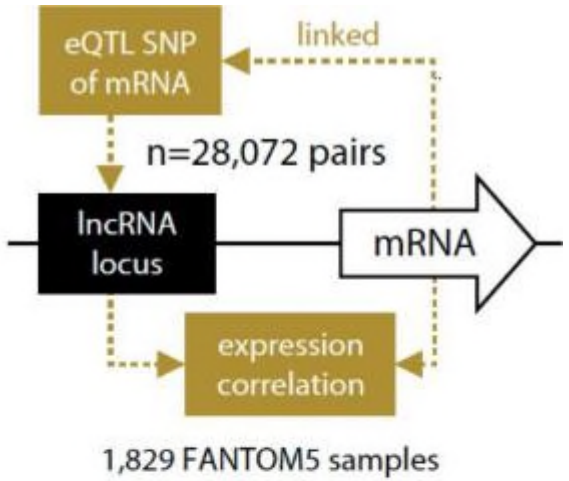
Versus GENCODEv25 lncRNA Catalogue (number of genes)

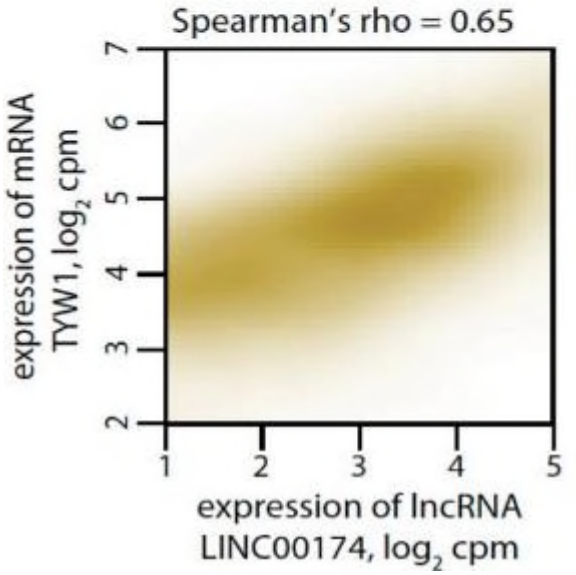


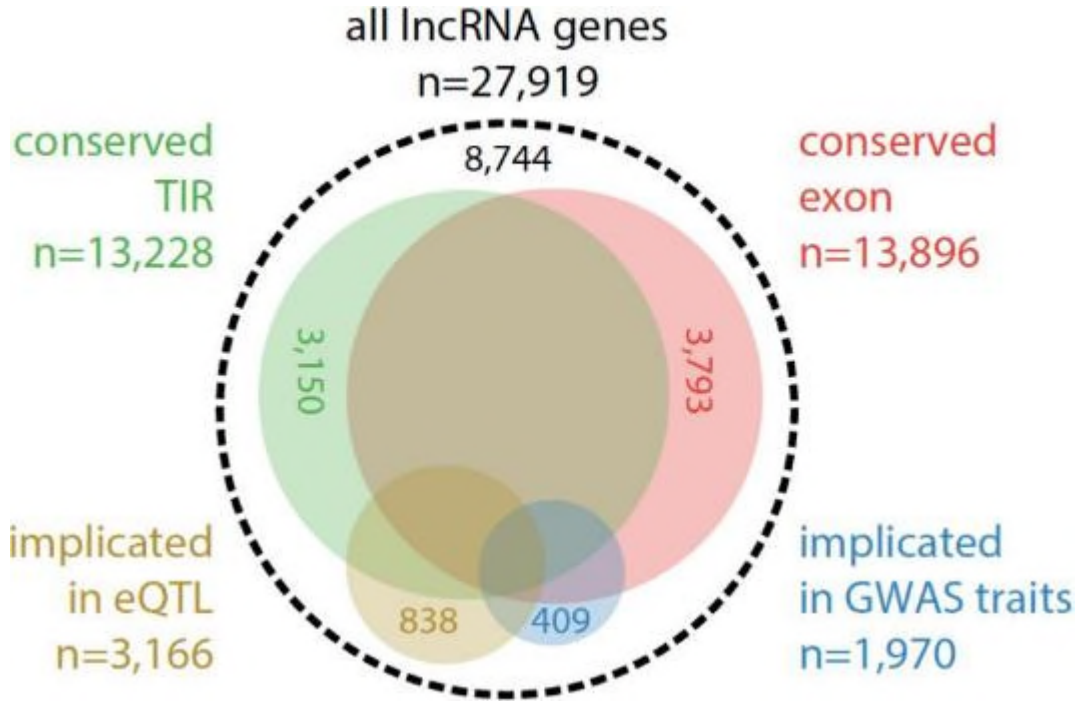
Hon *et al*., *Nature* 543, 199 (2017)

Many lncRNAs with potential function

Co-expression of lncRNA-mRNA pairs linked by eQTL







Identified 19,175 potentially functional lncRNAs in human

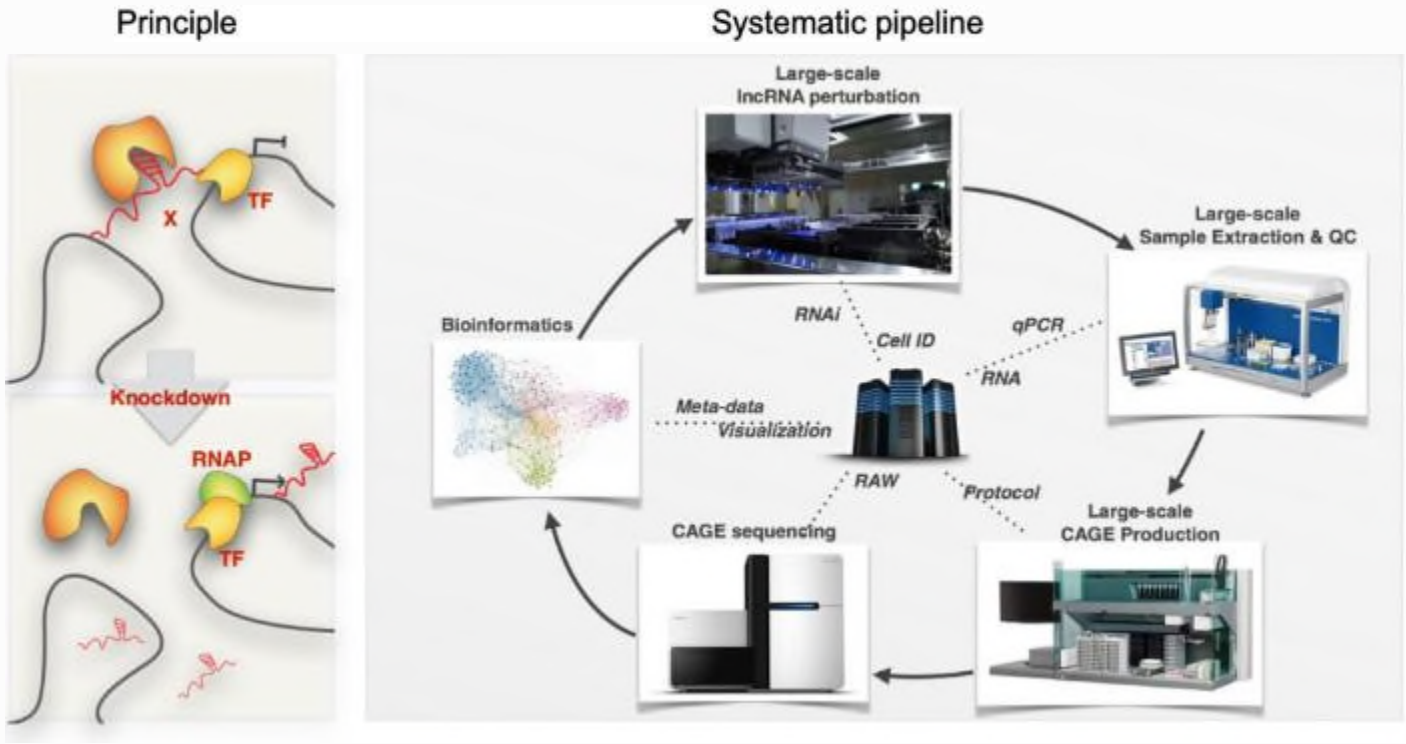
Hon *et al*., *Nature* 543, 199 (2017)

Function:

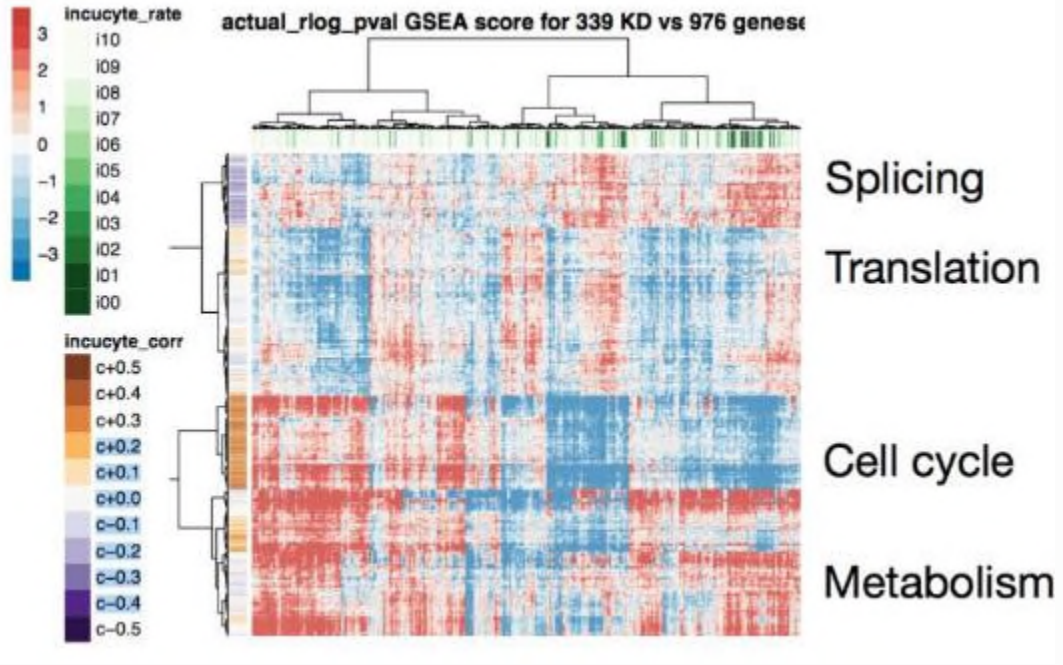
Do we need experimental evidence?

FANTOM6: *Functional lncRNA catalogue*

Measuring transcriptional phenotypes



Molecular phenotype induced by lncRNA KD



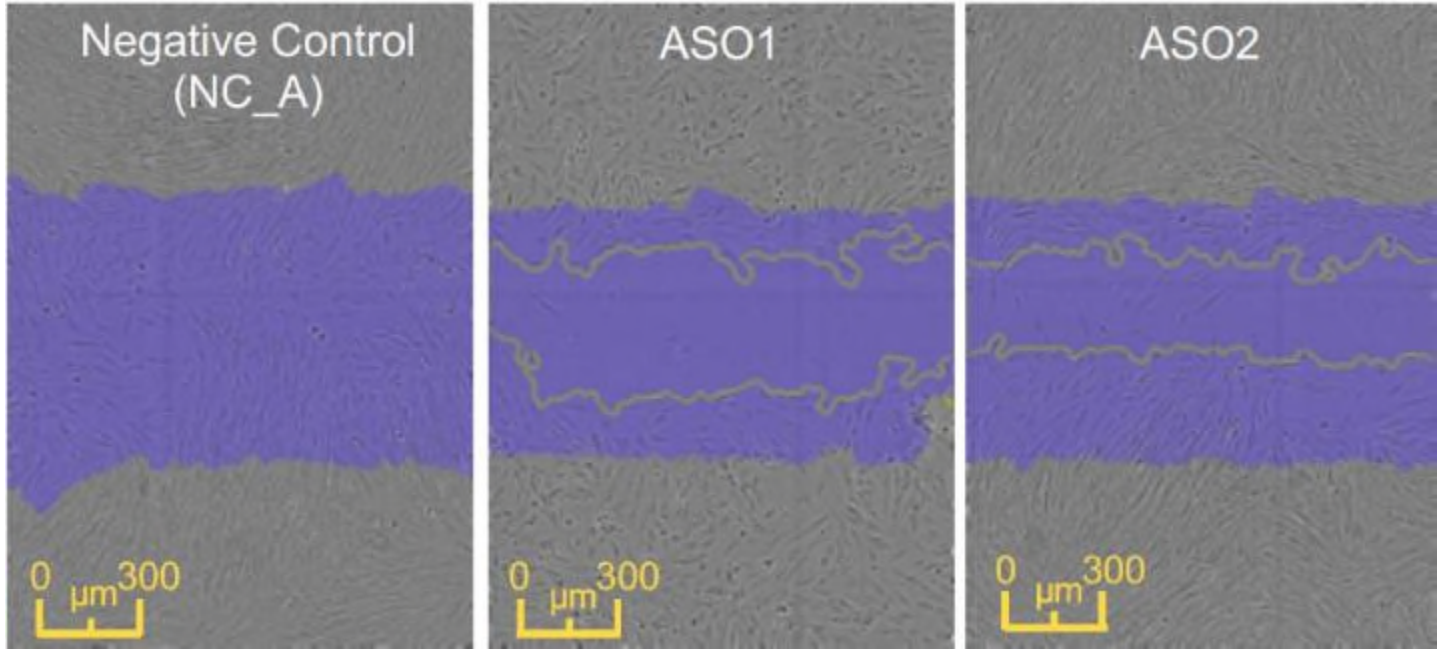
Characterization of the “new continent” of lncRNA:
- lncRNA functional
- Regulatory role among others
- Many genes with natural antisense RNAs

Alternatives:
Perturb-seq (@single cell level)

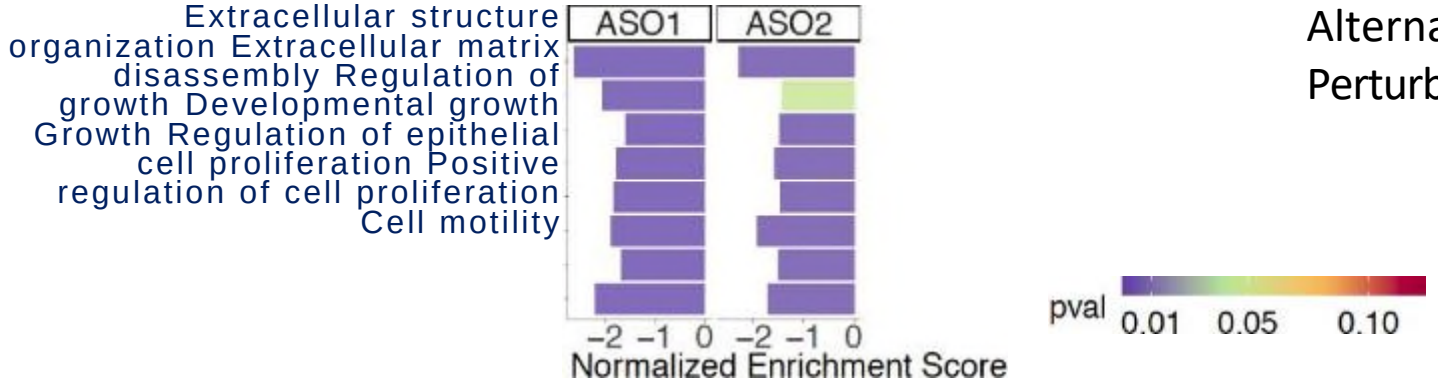
Ramilowski *et al. Genome Res.* 30, 1060 (2020) **Jay W. Shin, Michiel de Hoon, Jordan Ramilowski**

Example: KD of ZNF213-AS1 impacts cell growth and migration

Wound healing assay for ASOs targeting ZNF213-AS1
35-73% impairment of wound-closure



Selected enriched biological pathways

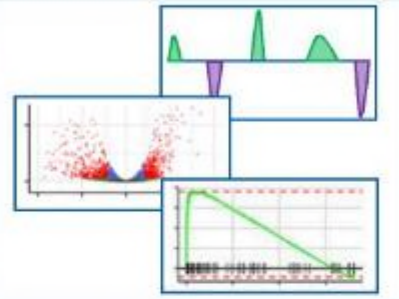


Alternatives:
Perturb-seq (@single cell level)

Annotation of lncRNAs: molecular and cellular phenotypes

Knockdown

Cell Phenotype CAGE Transcriptomic Profiling

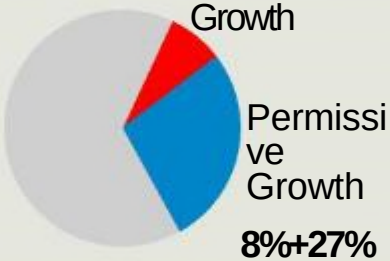


HDF

285 lncRNAs

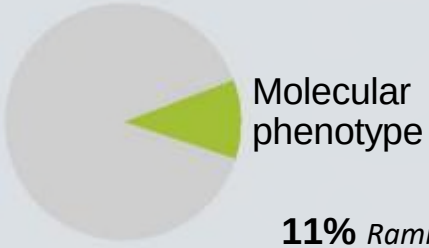
68%

194 lncRNAs



8%+27%

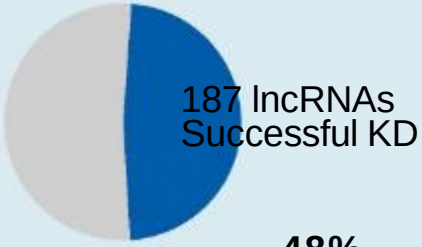
119 lncRNAs



11% *Ramilowki JA, Yip CW. Genome Research 2019*

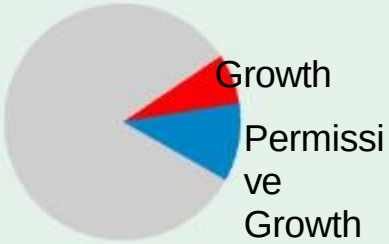
iPSC

390 lncRNAs



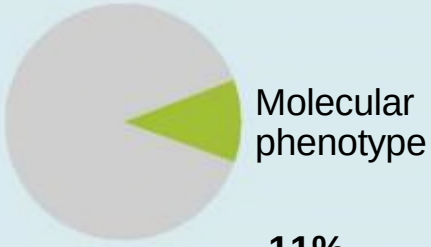
48%

187 lncRNAs

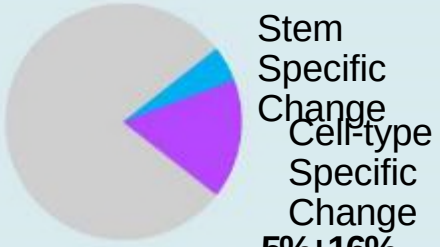


6%+12%

123 lncRNAs



11%



5%+16%

Note: lncRNAs are often functional in another cell type or assay

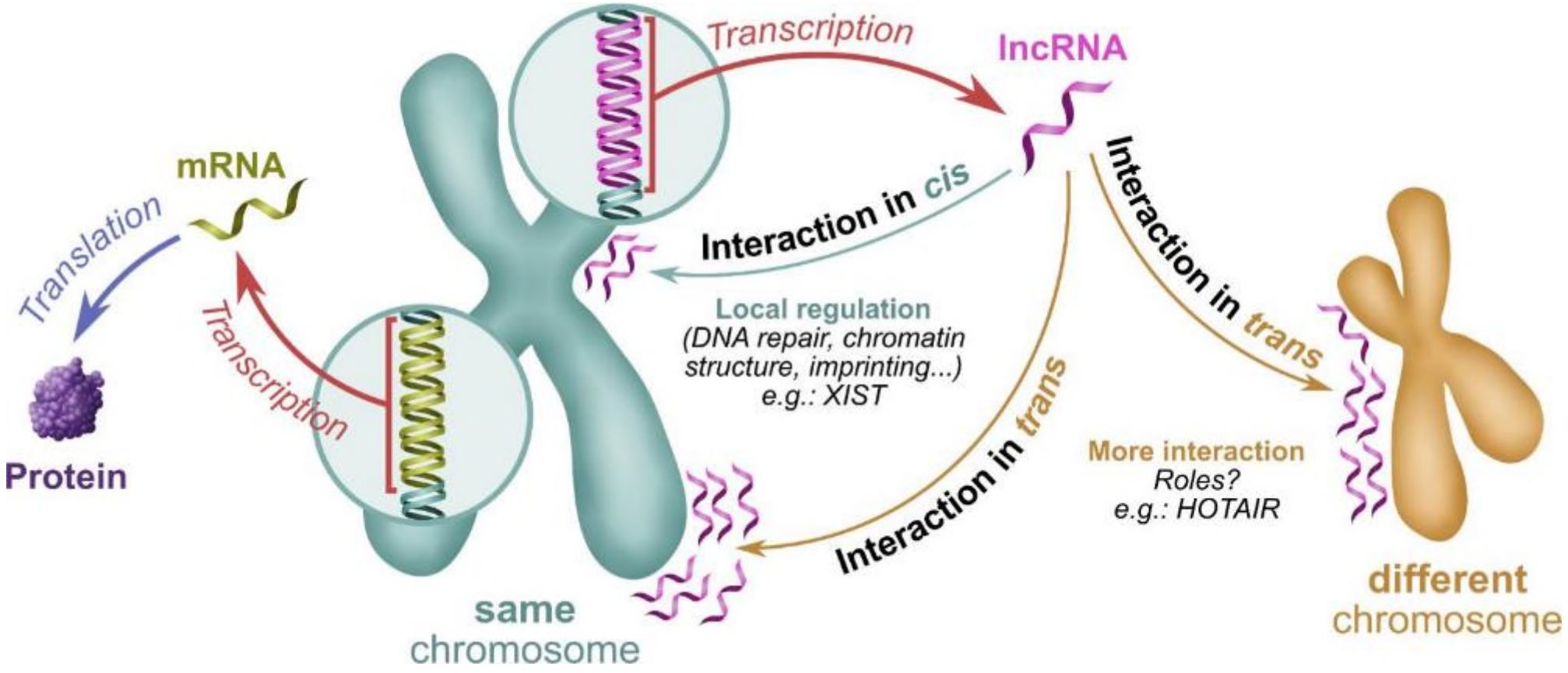
Chi Wai Yip *et al. Cell Reports* 41, 111893 (2022)

Chi Wai Yip

How to gather functional insights about lncRNAs efficiently?

Can we collect how and where lncRNAs interact, their “interactome”?

RNA-chromatin interactions affect epigenome and gene regulation



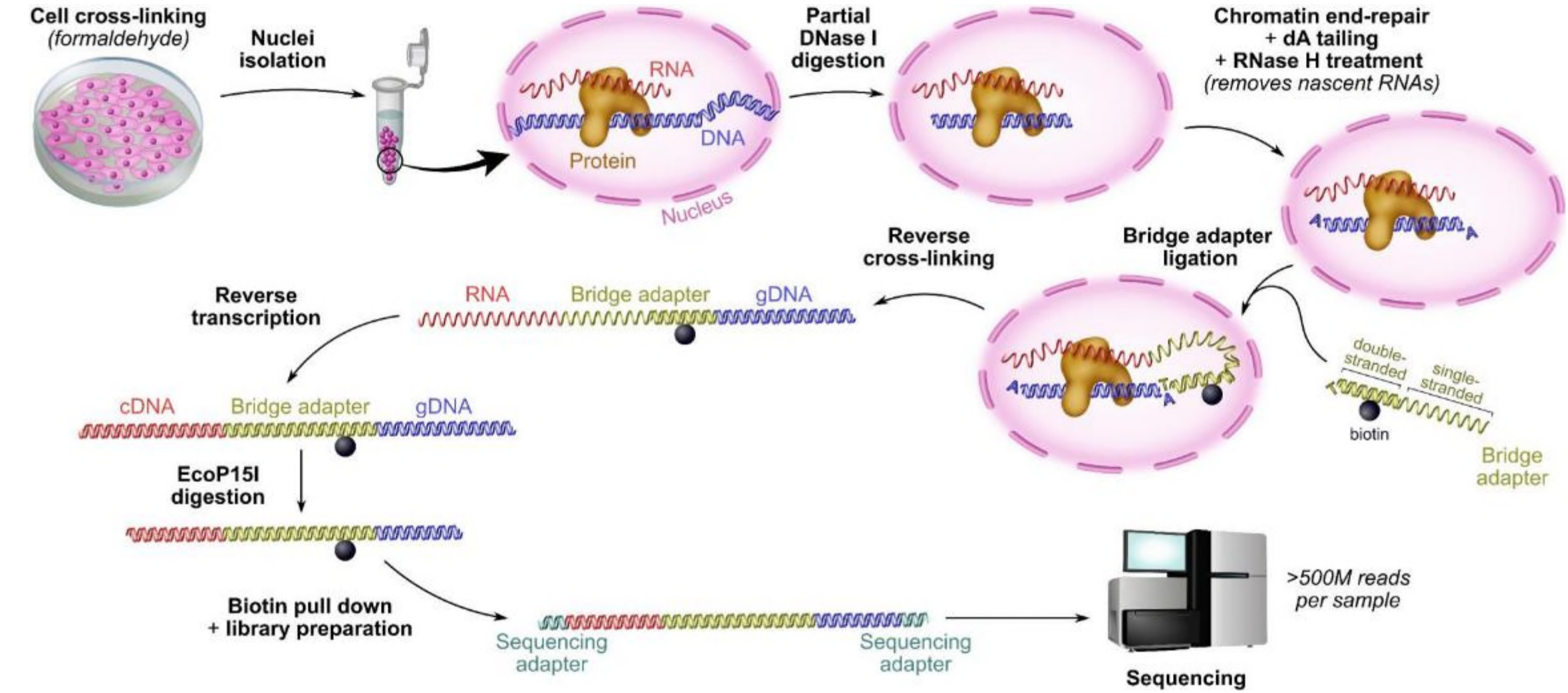
>37% of lncRNAs chromatin-bound

What is the role of RNA-chromatin interaction *in cis* and *in trans*?

- **Activate genes? Promoter? Enhancer?**
- **Repress genes? Insulator?**

More technologies

RADICL-seq captures RNA-DNA interactions in intact crosslinked nuclei

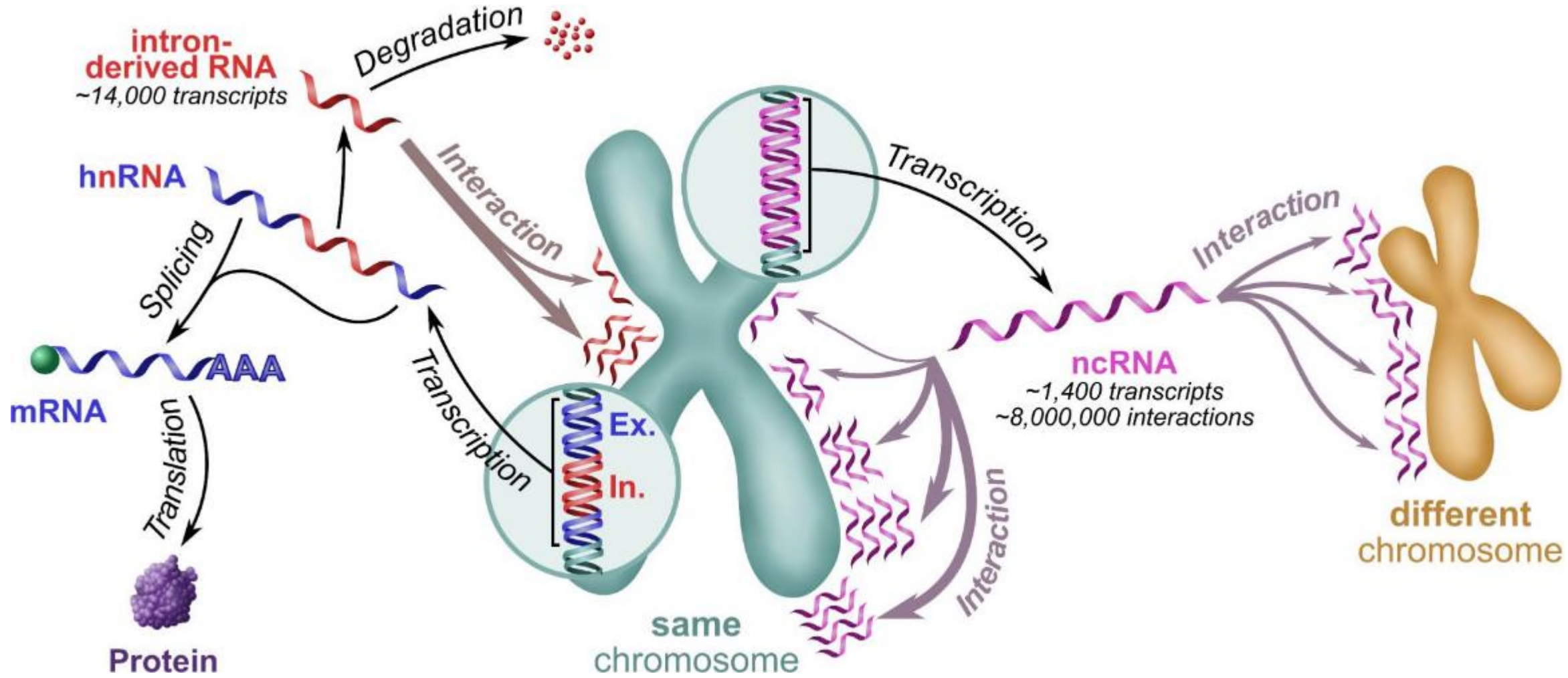


Other cell types that have been successfully tested with RADICL seq: miPSCs, mMEF, mOPC.
Typical seq. depth: 1 lane HiSeq2500/each replica.

Bonetti *et al. Nature Communications* 11, 1018 (2020)**Alessandro Bonetti**

Interactome analysis: *mostly introns in cis; lncRNAs interact more often in trans*

Bonetti *et al. Nature Communications* 11, 1018 (2020)



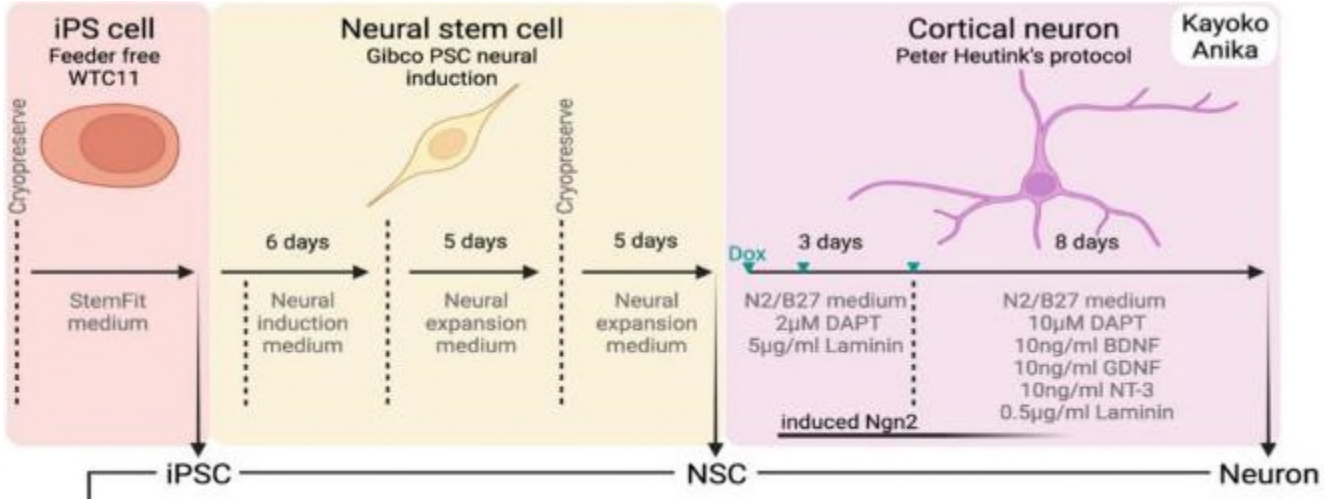
Enormous complexity from the first interactome in mouse ES cells

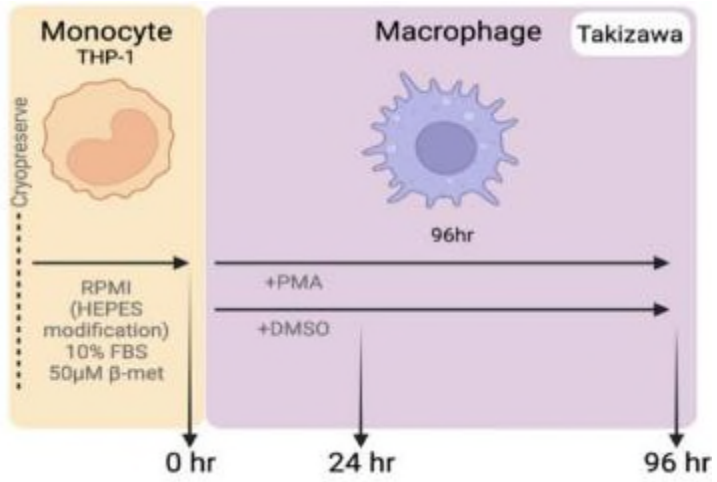
Exploring chromatin RNA interactions

- Biological systems
  - Static (map) + dynamic (to capture dynamic interactions and connect with known networks)
- Large scale Chromatin/RNA + complementary data

Cell models and FANTOM6 RADICL libraries (~15 cell types)

iPSC-Neuron series (2 reps each): iPS to neurons THP-1 series: THP-1 monocyte to macrophage





iPSC-mNGN2 / NSC from iPSC-mNGN2 / Neuron from iPSC-mNGN2 THP-1 (DMSO, PMA 24h, 96h)

Each rep 120"240 M(mapq10) Each rep 80"120 M(mapq10)

Pre F6 cell lines

iPSC (F6) (3 reps) Each rep "200 M (mapq10)

HDF (human dermal fibroblast) (2 reps) Each rep "130 M (mapq10)

ENCODE cell lines

MCF7 (breast cancer cell line) (2 reps) Each rep "80 M (mapq10)

MCF10A (mammary gland, non-tumorigenic epithelial) (2 reps)

K562 (chronic myeloid leukemia (CML) cell line) (2 reps)

HCT116 (colorectal carcinoma cell line) (2 reps)

Each rep "120 M(mapq10)

T cell CD4+

(naïve and active(Th1)) (Beatrice Bodega lab) (2 donors each)

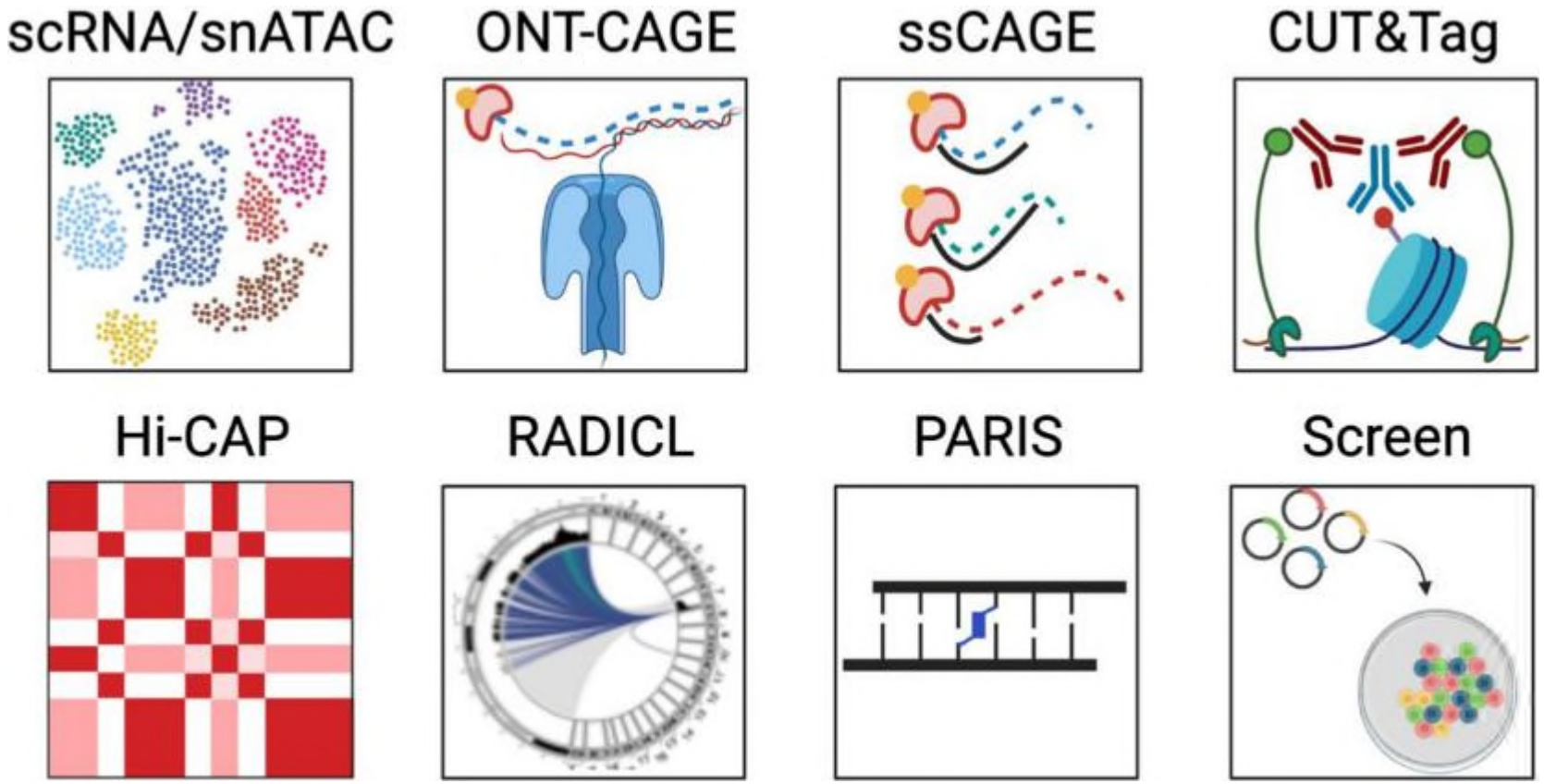
Each rep 22"33 M (mapq10)

AMI samples from patients

(acute myeloid leukemia) (Andreas Lennartsson lab) (2 donors)

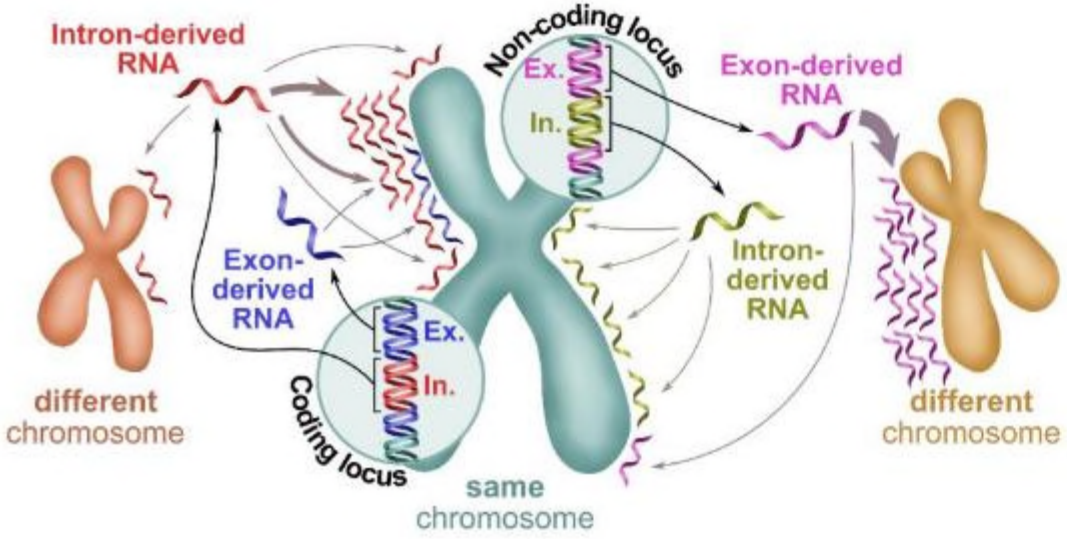
Each rep "15 M (mapq10)

FANTOM6-Interactome
Complementing the chromatin RNA interactome



Summary of patterns of RNA-chromatin interaction

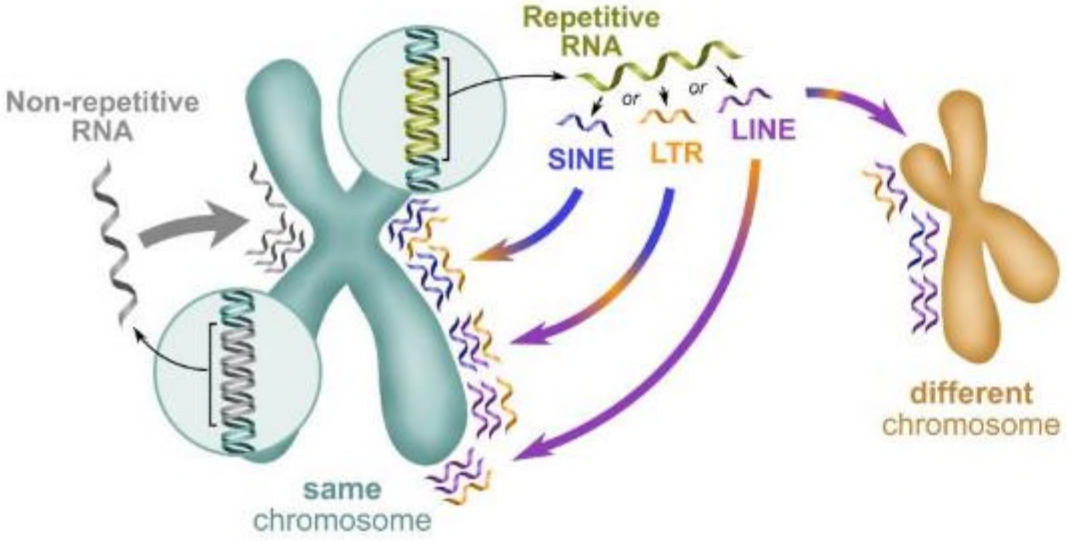
Coding *v.s.* non-coding locus-derived RNAs



Coding gene-derived RNAs: more often *in cis*

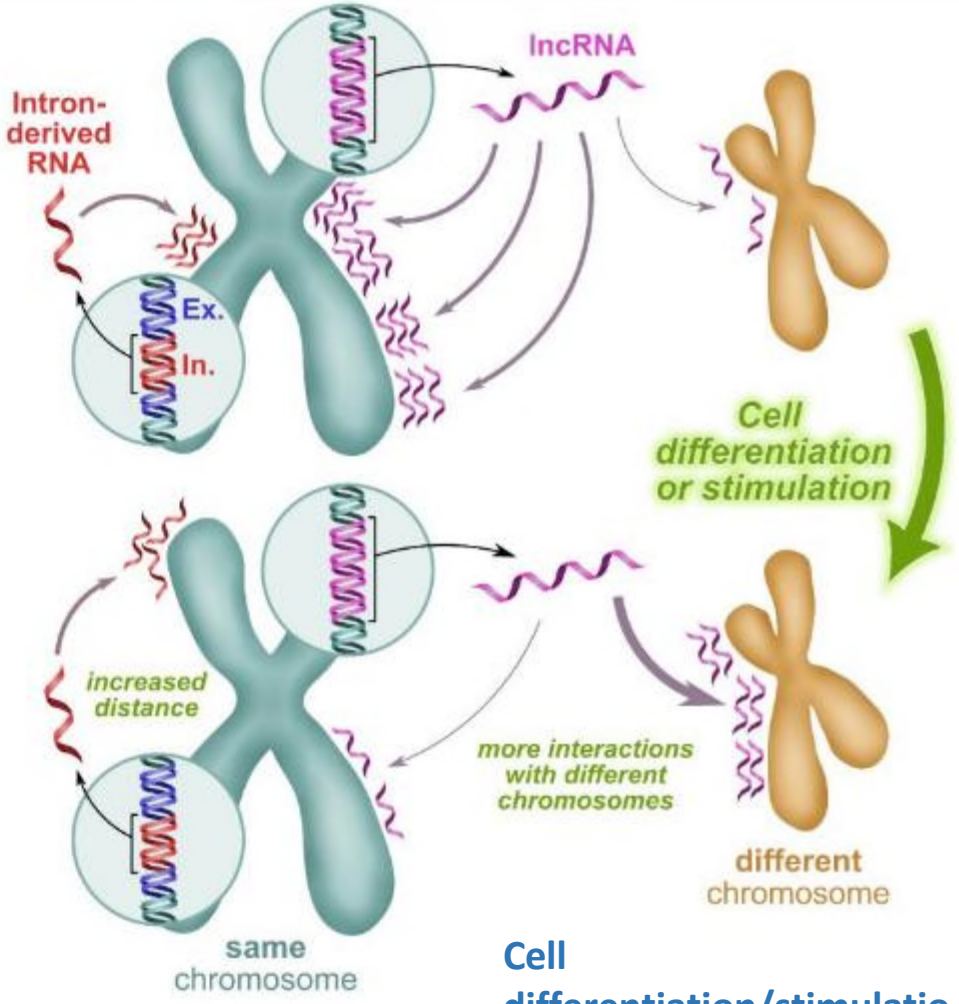
lncRNAs: more often *in trans*

Repetitive *v.s.* non-repetitive RNAs



Retrotransposon: longer range interactions, SINE more *in cis*, LINE more *in trans*

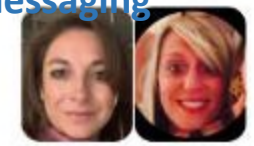
During cell differentiation/stimulation



Cell differentiation/stimulation ↳ "RNA messaging"

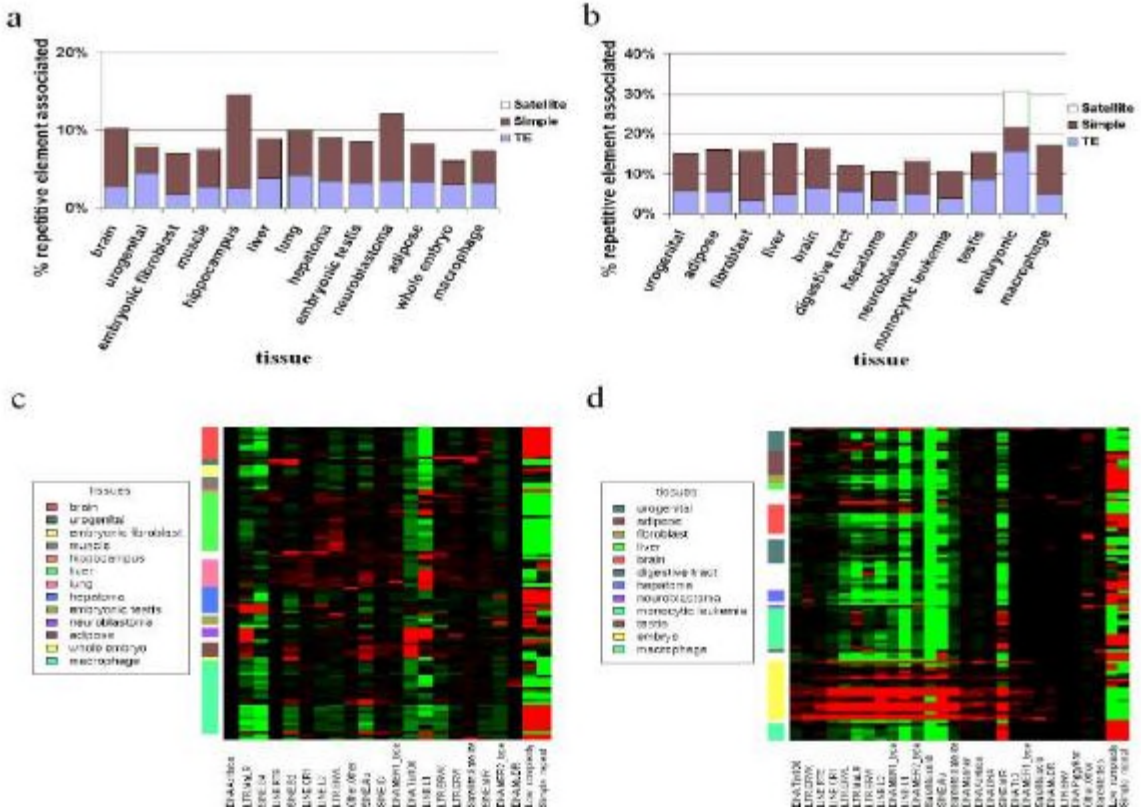
Laura Carpen, Aslihan Karabakack, Rodrigo Pracana Human Technopole + RIKEN IMS

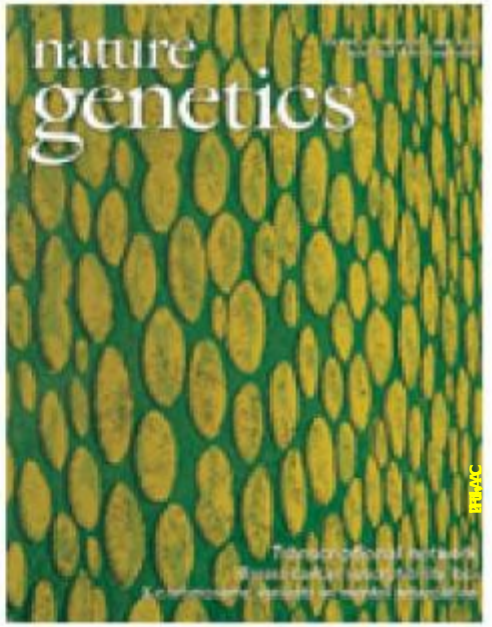
Beatrice Bodega, Valeria Ranzani University of Milan



Surprise!
Retrotransposon elements are broadly expressed in mammalian cells/tissues

- FANTOM4 -
- transcription is often initiated on retrotransposon elements
- The signal was actively removed from data with “repeat masks”
- Geoff: PhD students from computer sciences was fearless



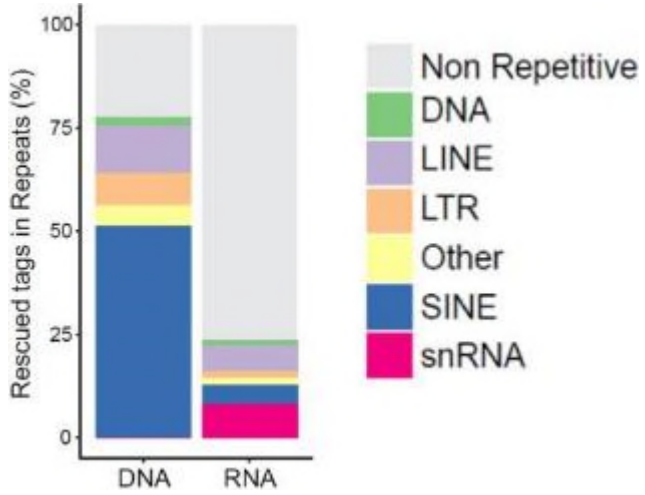


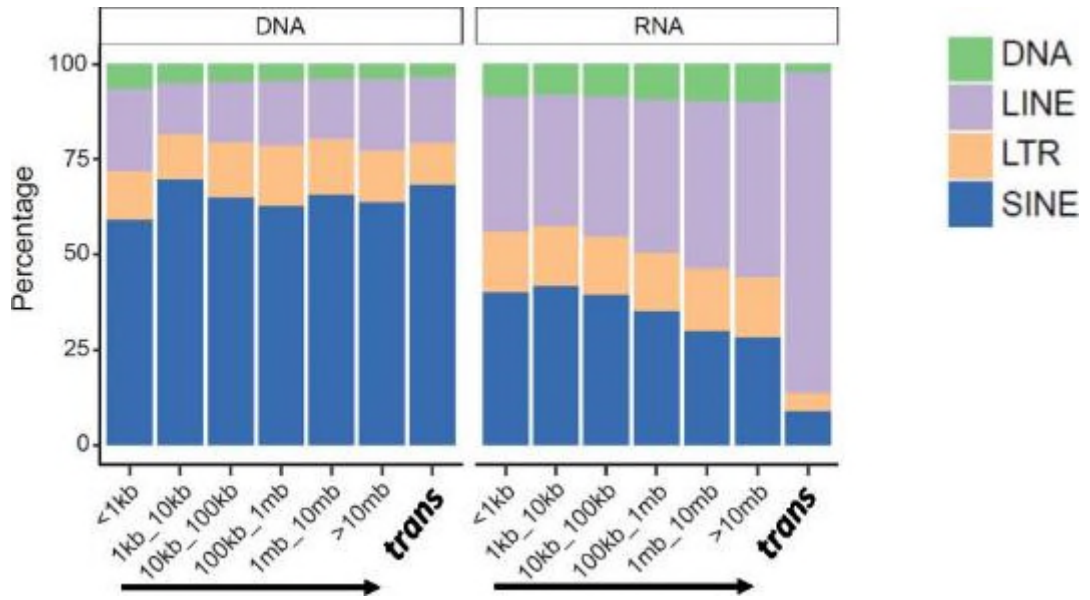
Nat Genet. 2009 May;41(5):563-71

The regulated retrotransposon transcriptome of mammalian cells

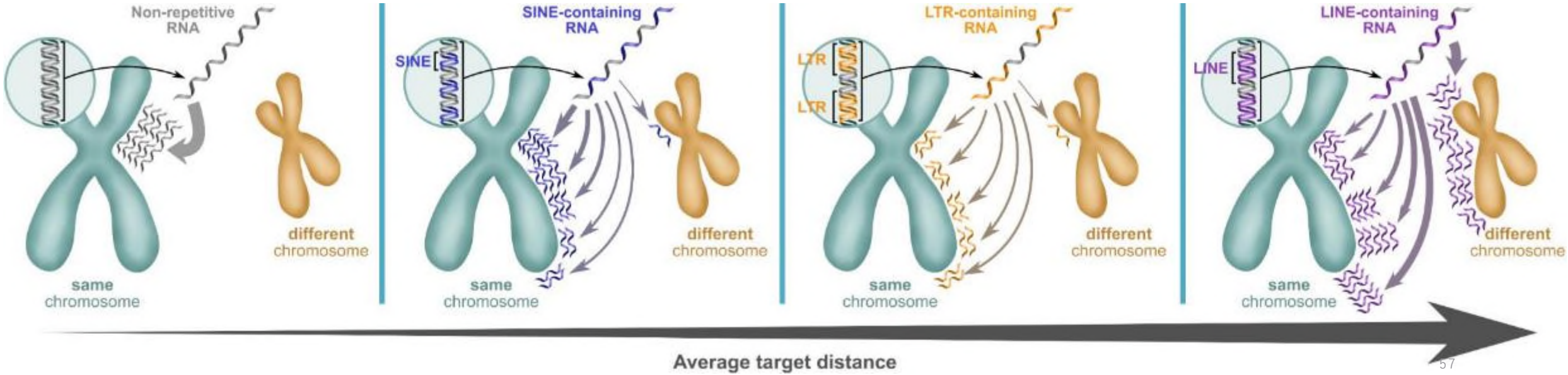
Geoffrey J Faulkner[1], Yasumasa Kimura[2], Carsten O Daub[2], Shivangi Wani[1], Charles Plessy[2], Katharine M Irvine[3], Kate Schroder[3], Nicole Cloonan[1], Anita L Steptoe[1], Timo Lassmann[2], Kazunori Waki[2], Nadine Hornig[4,5], Takahiro Arakawa[2], Hazuki Takahashi[2], Jun Kawai[2], Alistair R R Forrest[2,6], Harukazu Suzuki[2], Yoshihide Hayashizaki[2], David A Hume[7], Valerio Orlando[4,5], Sean M Grimmond[1] & Piero Carninci[2]

Transposable Elements in RNA contacting chromatin





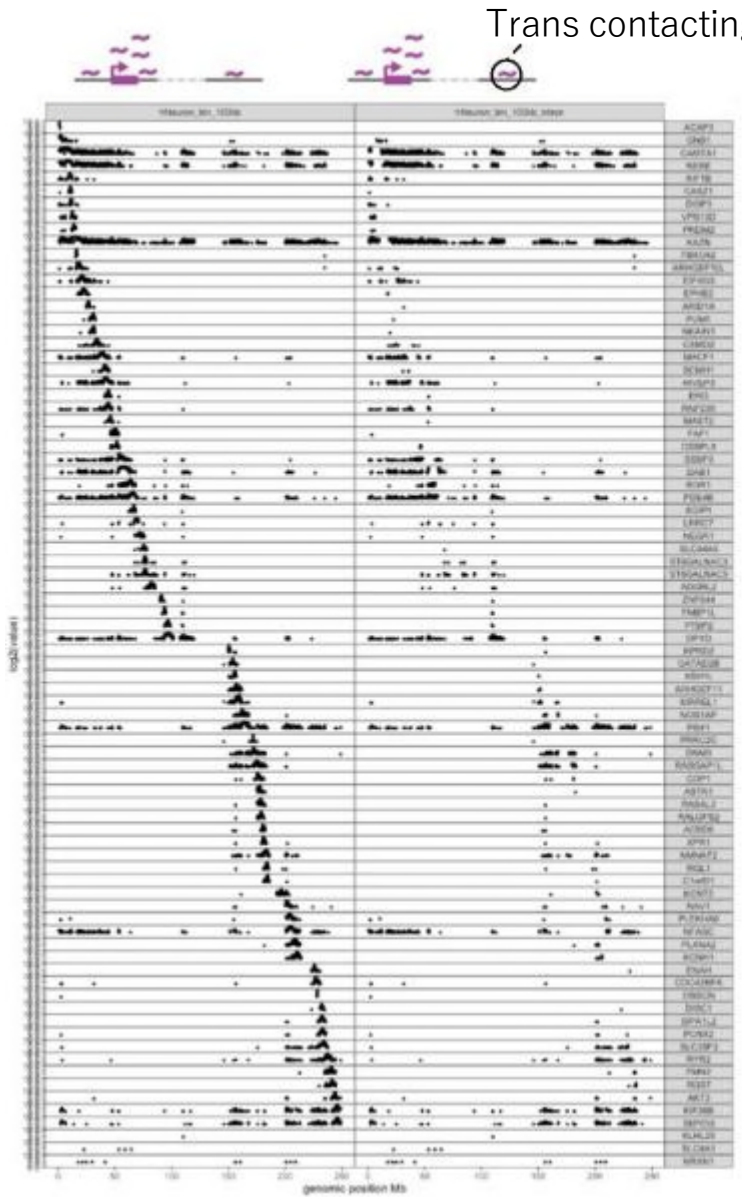
Distance in ***cis*** Distance in ***cis***

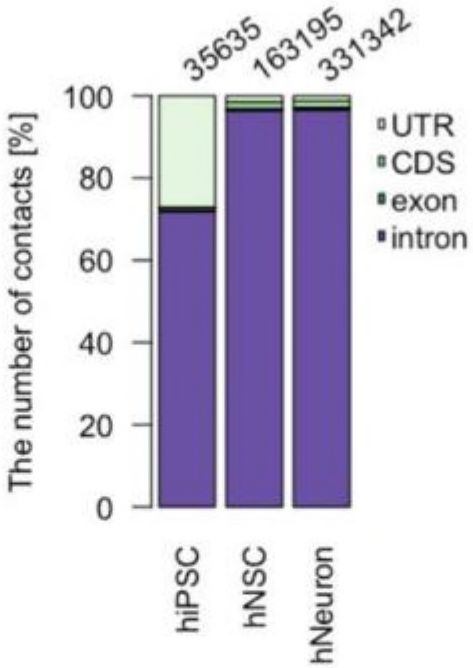


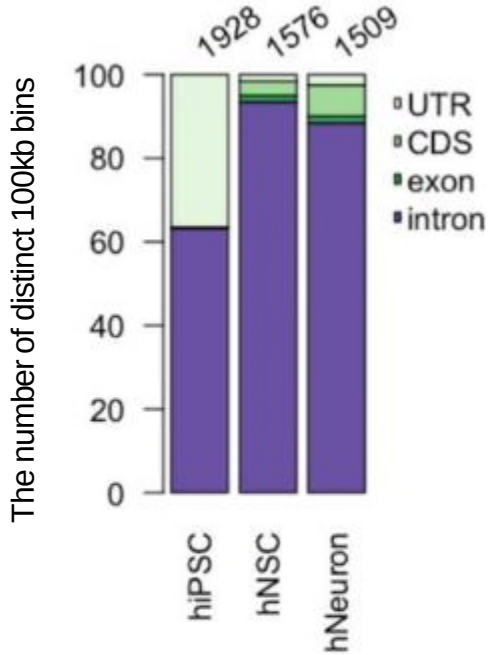
TIR: Trans-contacting Intronic RNA (another surprise: we know they are usually degraded)

- Transcribed from intronic region
- Making trans contacting (receiving locus > 5Mb apart from the source gene locus)

Trans-contacting Intronic RNAs (from intronic regions)

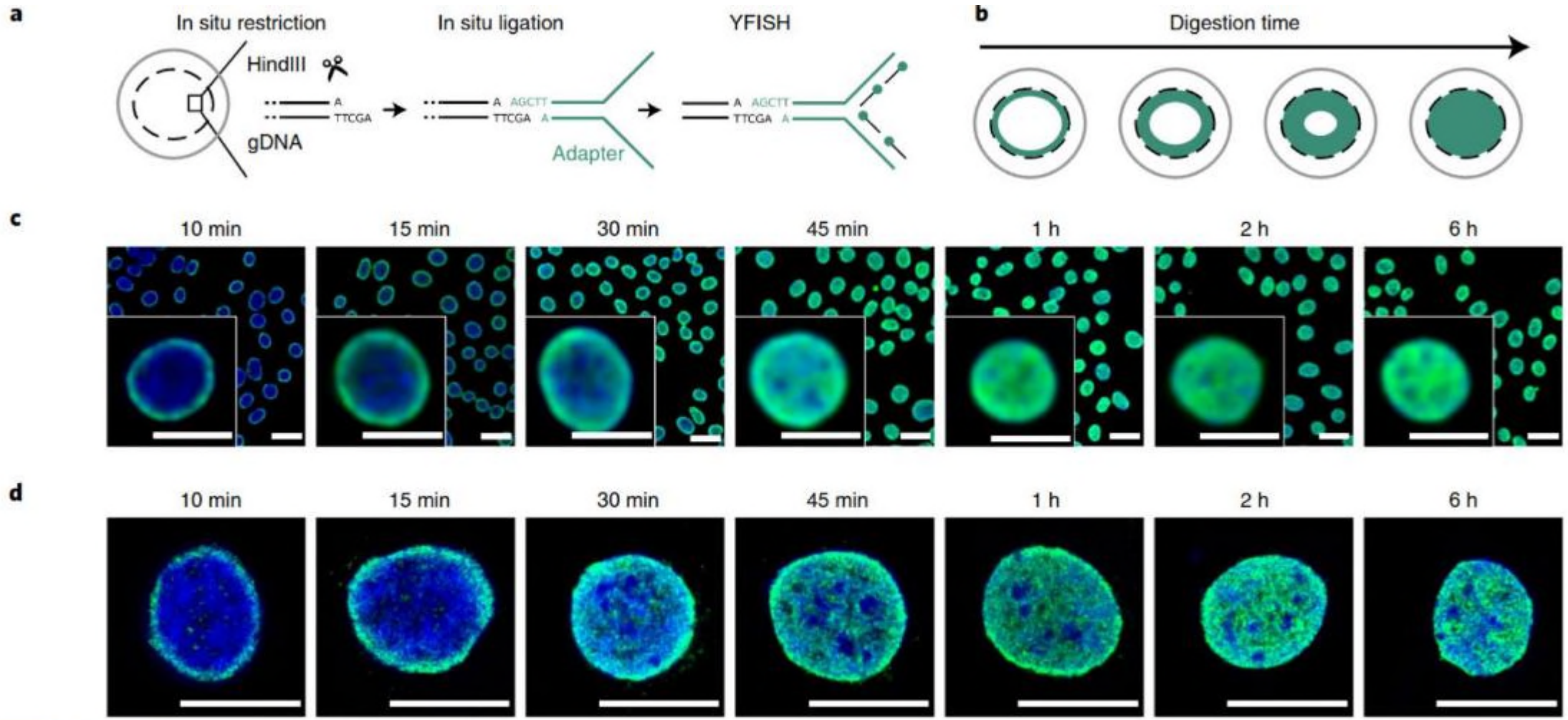






Wenjing Kang and Magda Bienko

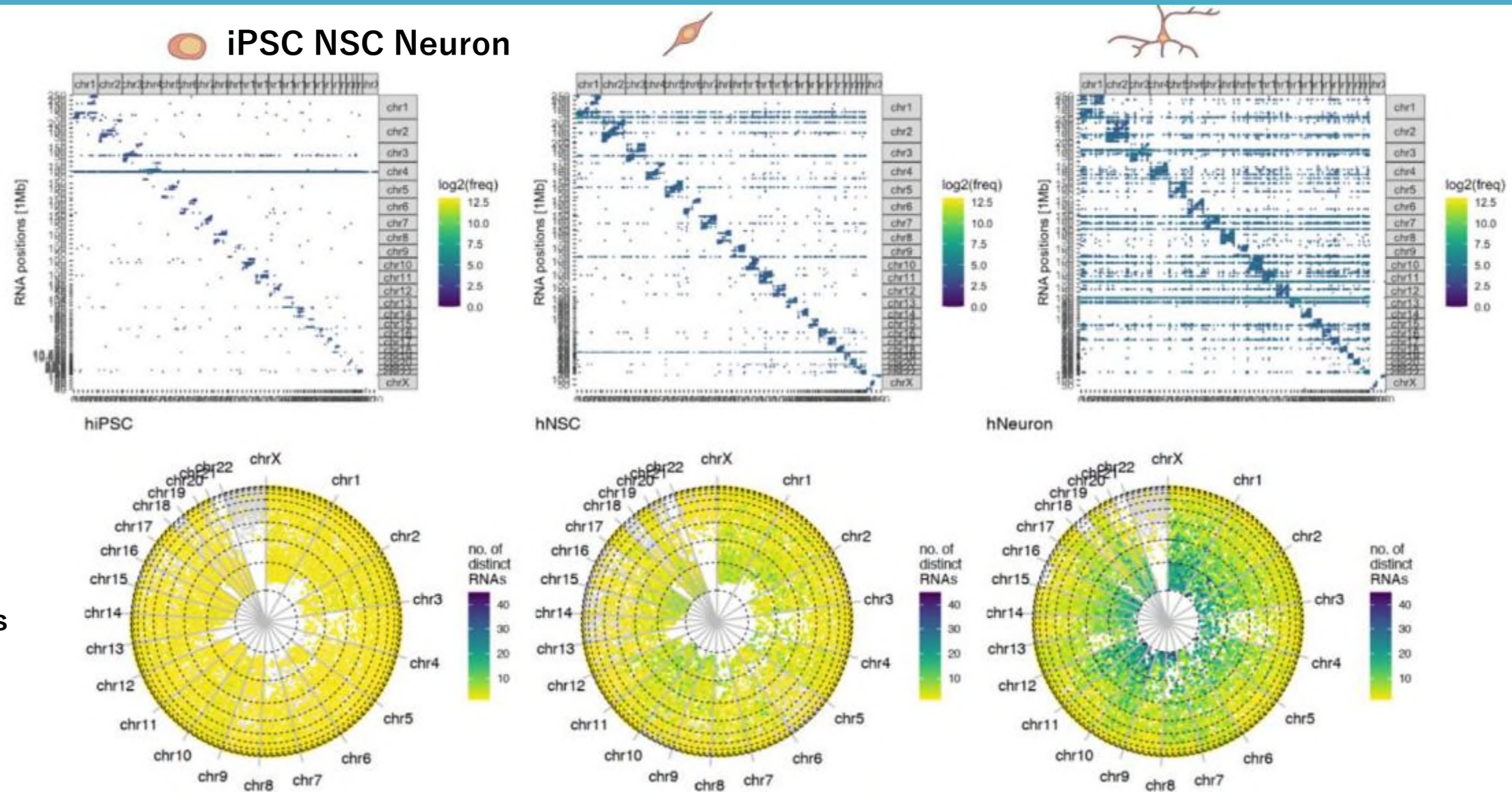
GPSeq (genomic loci positioning by sequencing)



Girelli, G. et al, Nature Biotech. 2020

Wening Kang and Magda Bienko

TIRs increasingly accumulate at nuclear center during differentiation



TIR from long neuron-specific genes, contacting short neuron-specific genes

MALAT1 and other TIR forming clouds are exclusive

RED: MALAT1 clusters

Other colors: various TIRs analyzed

Nuclei of neurons showing different RNA clouds territories

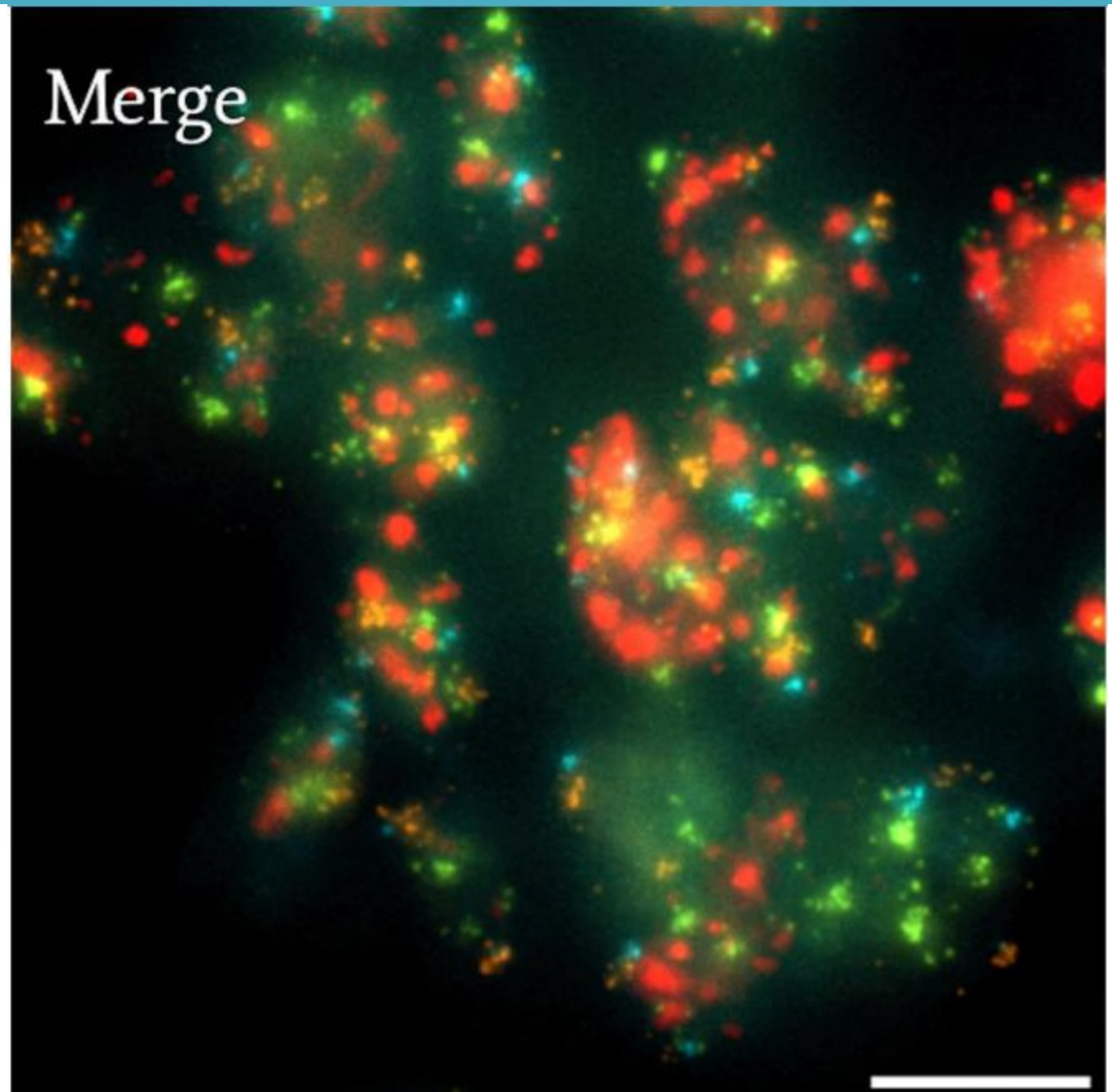
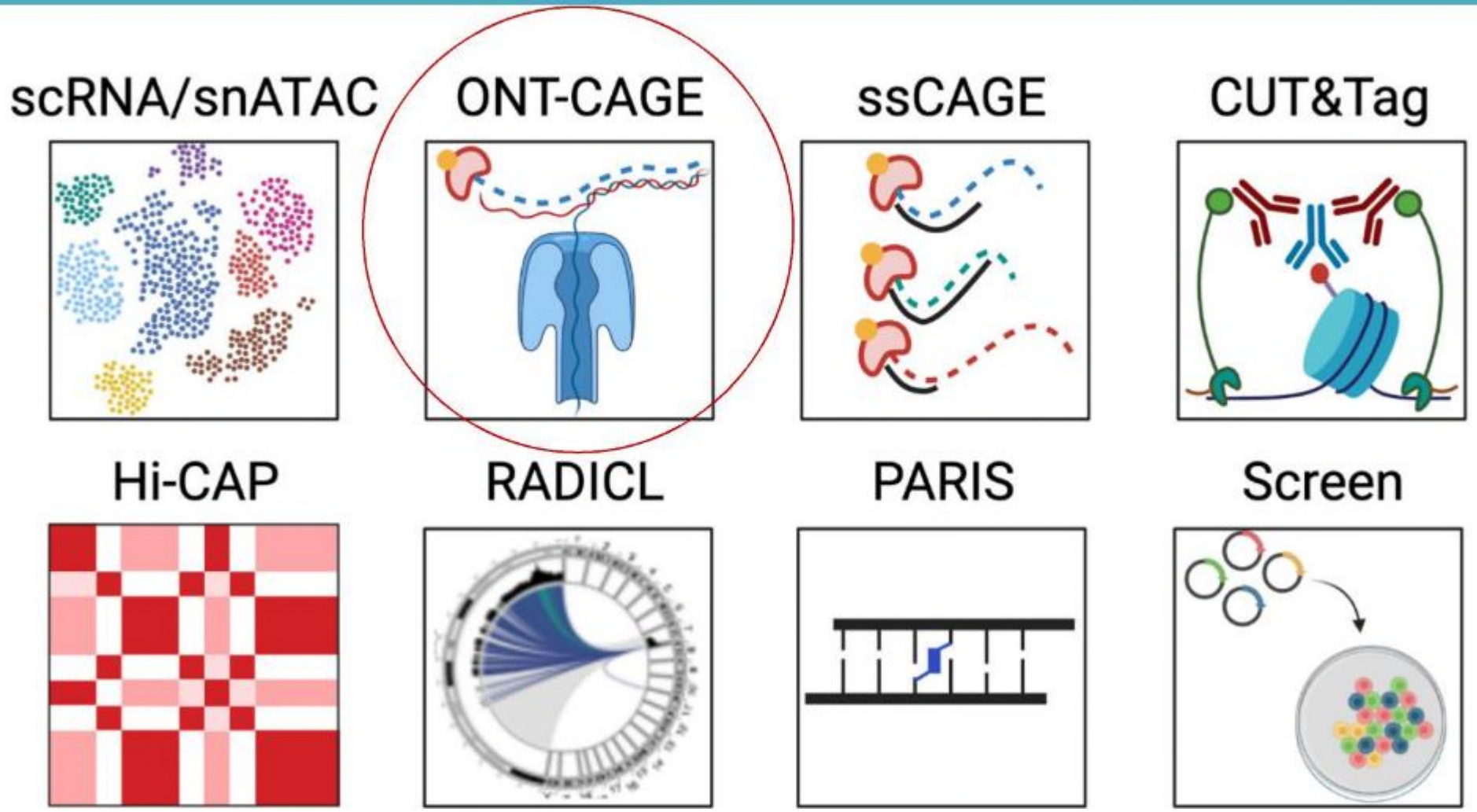


Image: Magda Bienko and Wenjing Kang

What are the interacting RNAs?

- **Full sequence needed to understand key lncRNA variants**
  - **further, we still do not understand the lncRNA diversity**

FANTOM6- Phase 2-1
Complementing the chromatin RNA interactome



October 2022, Banbury meeting discussing transcripts

- **Full set of functional isoforms of protein coding genes mRNAs**
- **Identification of all (functional) lncRNAs**
- **Technologies needed**
- **Consensus: making progress, yet we are far from completion**

Towards completing human genes catalogue(s) including lncRNAs

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature > perspectives > article

Perspective | Published: 04 October 2023

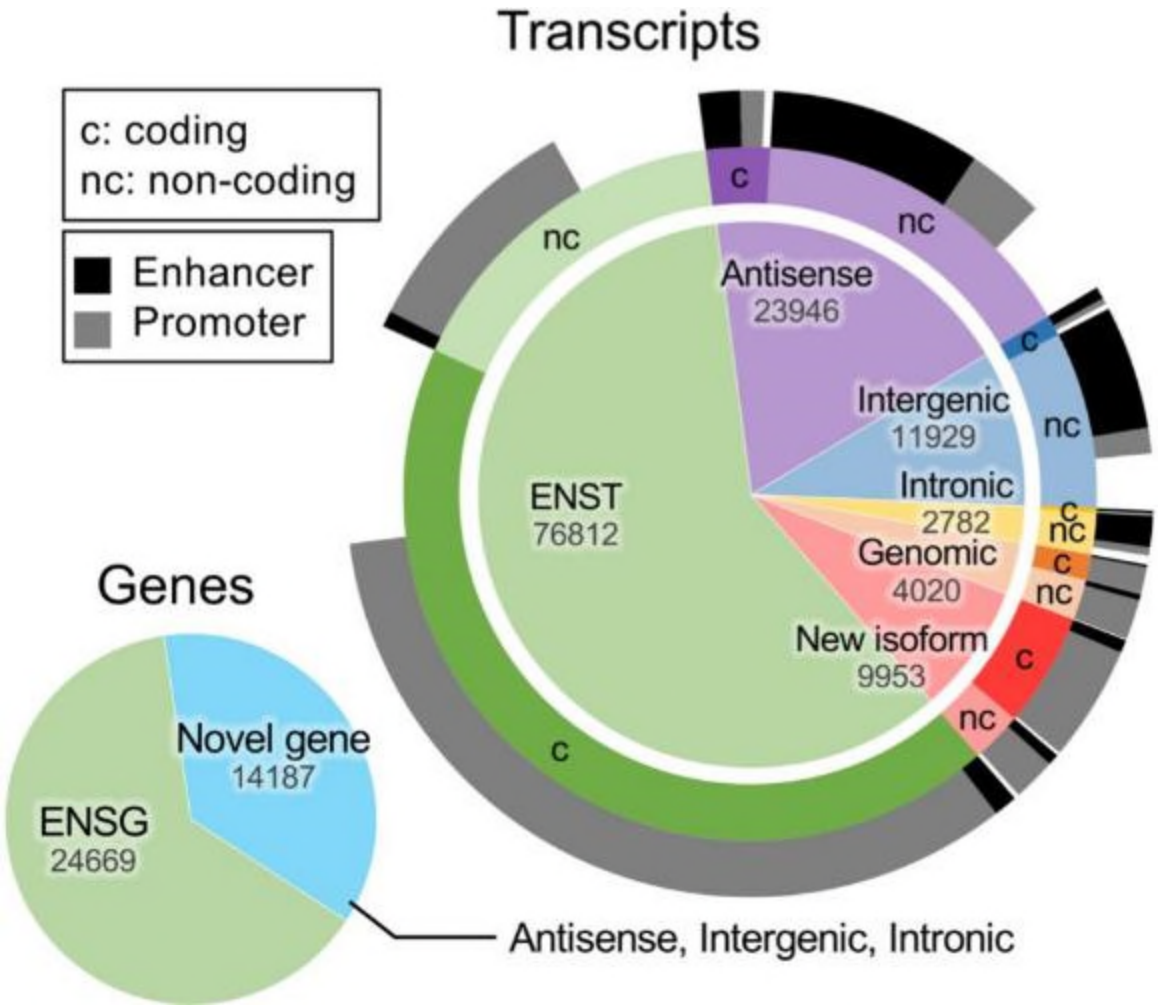
The status of the human gene catalogue

Paulo Amaral, Silvia Carbonell-Sala, Francisco M. De La Vega, Tiago Faial, Adam Frankish, Thomas Gingeras, Roderic Guigo, Jennifer L. Harrow, Artemis G. Hatzigeorgiou, Rory Johnson, Terence D. Murphy, Mihaela Pertea, Kim D. Pruitt, Shashikant Pujar, Hazuki Takahashi, Igor Ulitsky, Ales Varabyou, Christine A. Wells, Mark Yandell, Piero Carninci ✉ & Steven L. Salzberg ✉

Nature **622**, 41–47 (2023) | Cite this article

Amaral *et al. Nature* (2023)

Sequencing full-length cDNAs with ONT



Complementary with GENCODE

- Many RNAs from/with retrotransposon element
- Many different transcripts overlapping promoters, intronic regions, transcripts
- Reliable cap-selected transcripts
- (annotating true 3'-ends with I. Legnini at HT)

FANTOM6 transcriptome manuscripts
- sense/antisense (PARIS),
- RNA structure (icSHAPE)
- Retrotransposons derived RNAs
Takahashi, Bodega et al, in preparation
- Enhancer RNAs landscapes
Yip et al, in preparation

Perspectives on lncRNA discovery

- Extended full-length cDNA (RNA) collection in FANTOM6 (+ FANTOM5 cells)
- Single-cell LONG full-length cDNA method (almost there)
- Redefinition of transcriptomes, splicing variants, lncRNA, RNA biotypes, etc.

- Collaborating with GENCODE

FANTOM6
FUNCTIONAL ANNOTATION OF THE MAMMALIAN GENOME

Changing view of chromatin

- RNA component of chromatin is massively underestimated
  - Emergence of clear patterns of RNA interactions with chromatin
  - Intronic RNAs from protein coding genes (TIR) mostly in cis, but not only.
  - lncRNA mostly in trans
  - Specific pattern involving transposon elements
- There are still surprises from RNAs – more to discover and annotate

- Currently: exploring variability of interactome
  - Multiomics – adding functional data
  - Dynamics of interactions in disease models
  - Interaction with genetic loci
  - Identification of RNAs actionable to regulate a phenotype
  - Mechanistic studies

FANTOM6 + RIKEN + Technopole

April 12-14, 2023
October 2-3, 2023

Also, estblished multiple collaboration agreements with HT, ethical agremeents to process human samples in the FANTOM6

My group organizing the FANTOM6 in collaboration with HT

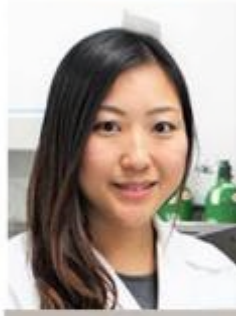
SINEUPs

More unexpected function of lncRNAs and repetitive elements

- *COI disclosure*
- *I am co-founder of Transine Therapeutics (now Harness therapeutics)*

Stefano Gustincich

Hazuki Takahashi

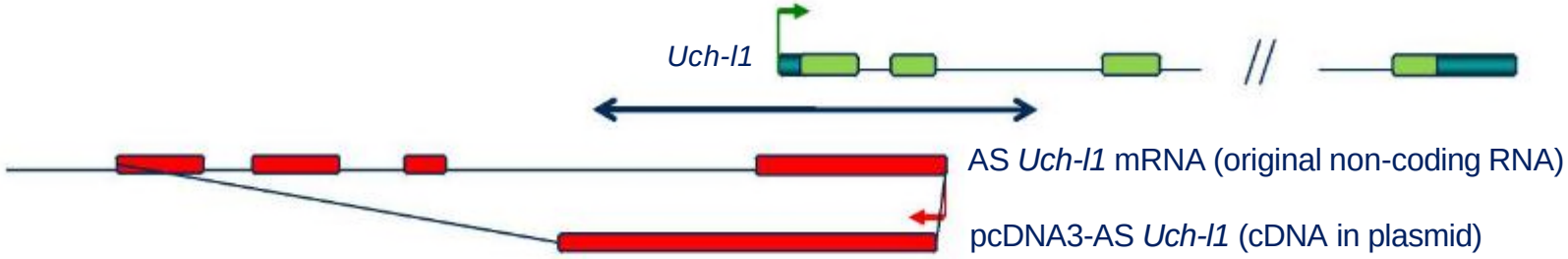


Silvia Zucchelli

Claudio Santoro

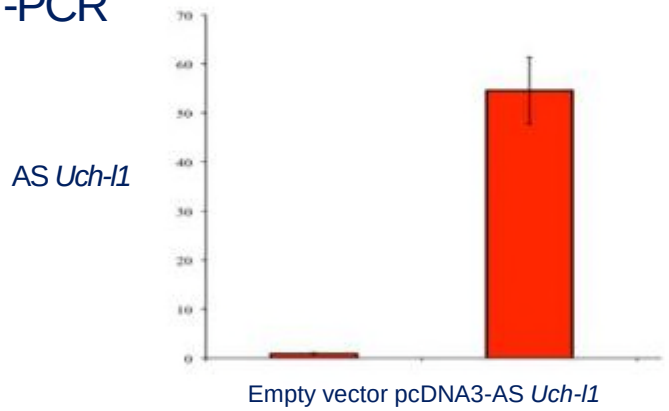
SINEUPs: an important example of antisense RNA function

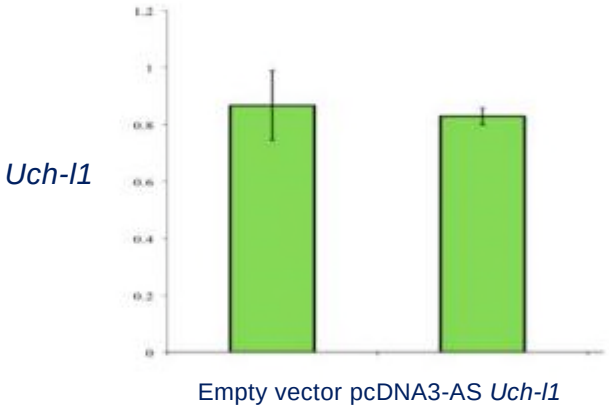
AS Uch-l1 regulates endogenous UCH-L1 protein expression



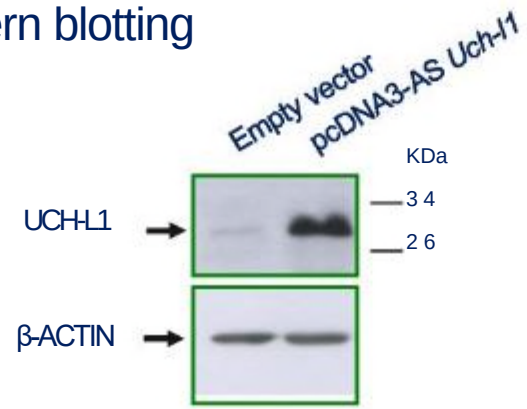
From miRNA Nobel Prize:
the concept is that
antisense = inhibition

qRT-PCR





Western blotting



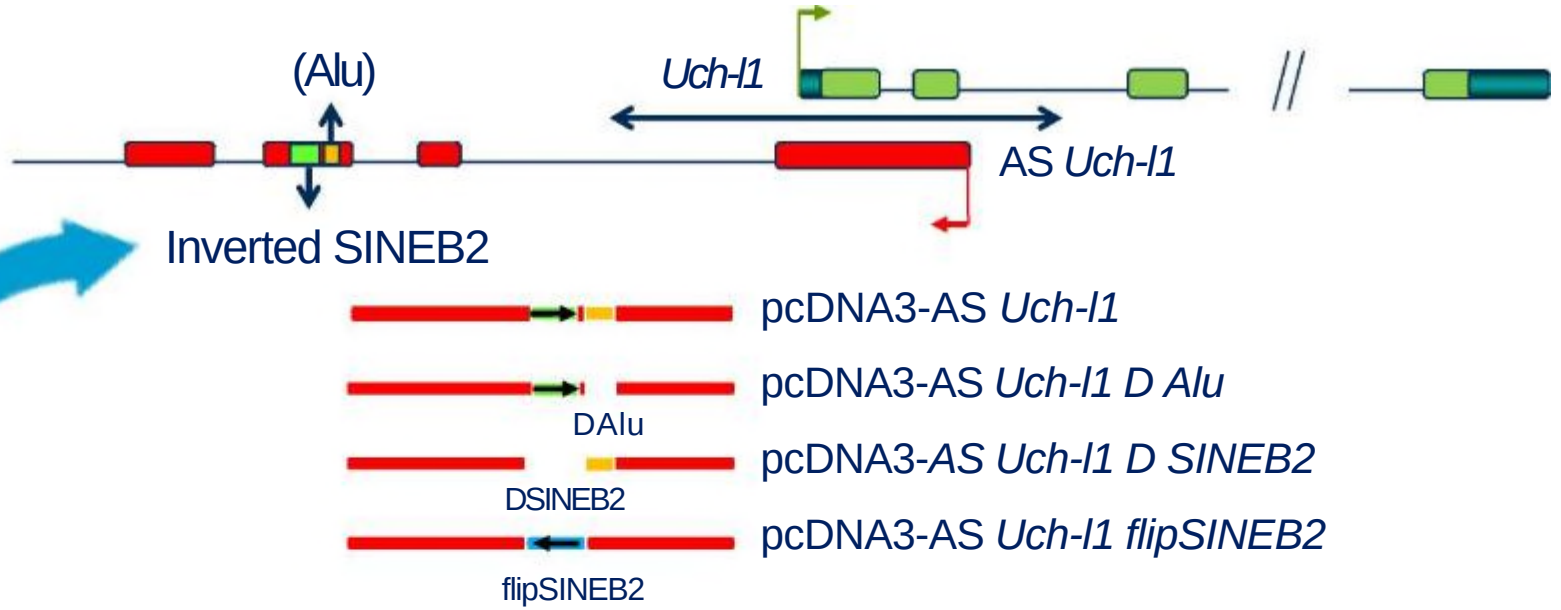
No changes at RNA level

Protein level:
dramatically enhanced!

Carrieri *et al, Nature* 491, 454 (2012)

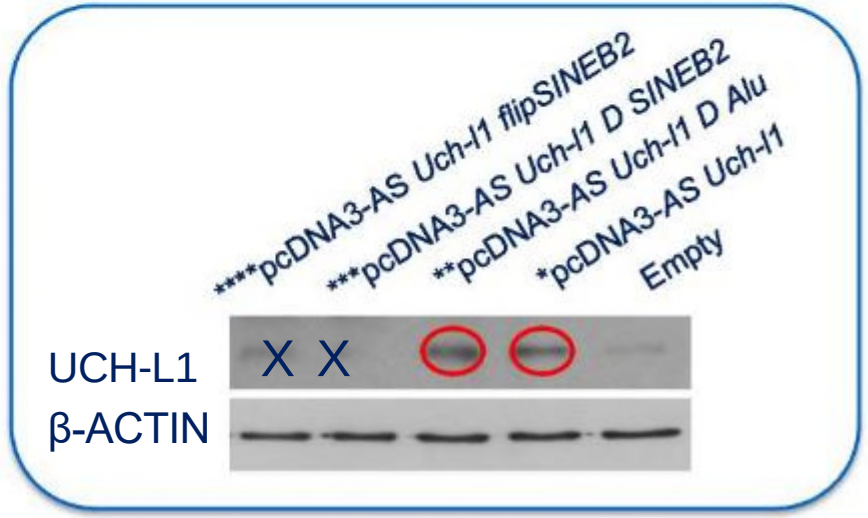
SINEUPs: an important example of antisense RNA function

AS Uch-l1 activity requires the inverted SINEB2



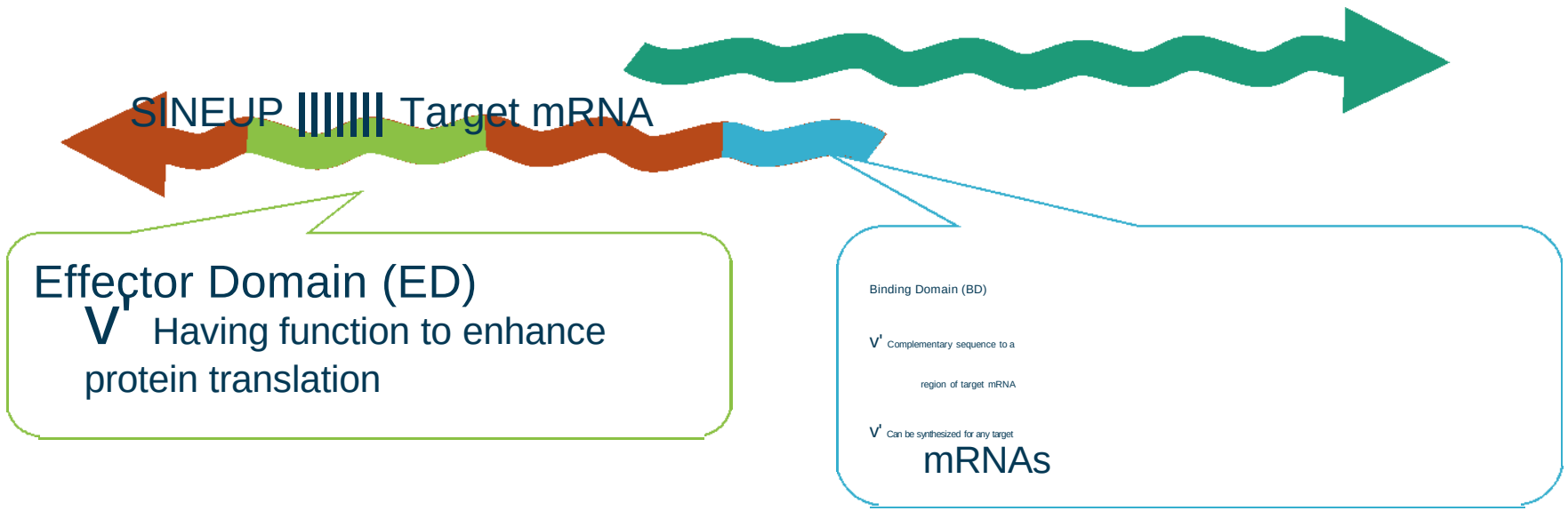
REPEAT MASKER analysis of AS *Uch-l1* sequence

```
  UnnamedSequen      765 GTGC--AGTGCTAGAGGAGGTCAGAAGAGGGCATTGGATCCCCCAGAACT 812
                           --i    iv v     i           i i       ii
C BJ#SINE/B2         156 GTGCCTGGTGCCCGCGGAGGCCAGAAGAGGGCGTCGGATCCCCTGGAACT 107

  UnnamedSequen      813 GGAGTTATACGGTAACCTCGTGGTGGTTGTGAACCACCATGTGGATGGAT 862
                           i ------------ i         i  i        i  ---
C BJ#SINE/B2         106 GGAGTTACA------------GATGGTTGTGAGCCGCCATGTGGGTG--- 72

  UnnamedSequen      863 ATTGAGTTCCAAACACTGGTCCTGTGCAAGAGCATCCAGTGCTCTTAAGT 912
                         -i  i vvi i   v v      v  v       v           vi
C BJ#SINE/B2          71 -CTGGGAATCGAACCCGGGTCCTCTGGAAGAGCAGCCAGTGCTCTTAACC 23

  UnnamedSequen      913 GCTGAGCCATCTCTTTAGCTCC 934
                             ii   i
C BJ#SINE/B2          22 GCTGAGCCATCTCTCCAGCCCC 1
```



Carrieri *et al, Nature* 491, 454 (2012)

Using SINEUPs (antisense lncRNAs)

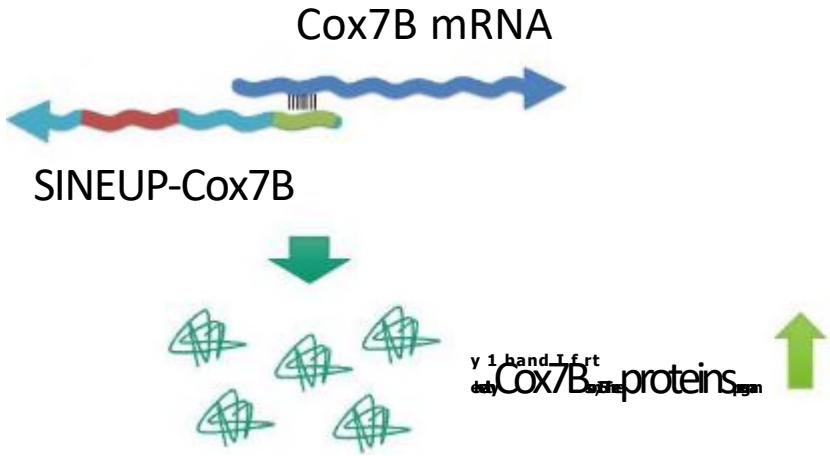


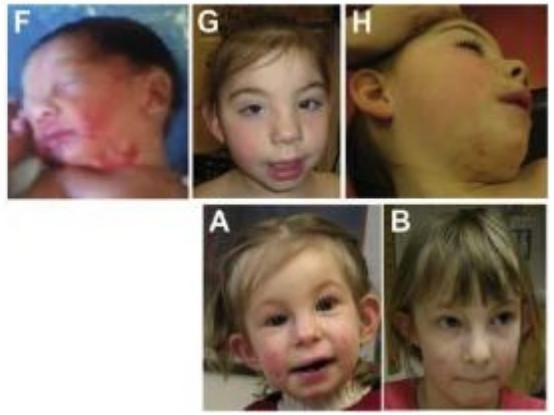
- Customizable design for any proteins
- Up-regulates protein synthesis 2-5 folds (physiological range)
- Acts on endogenous mRNA and exogenous targets

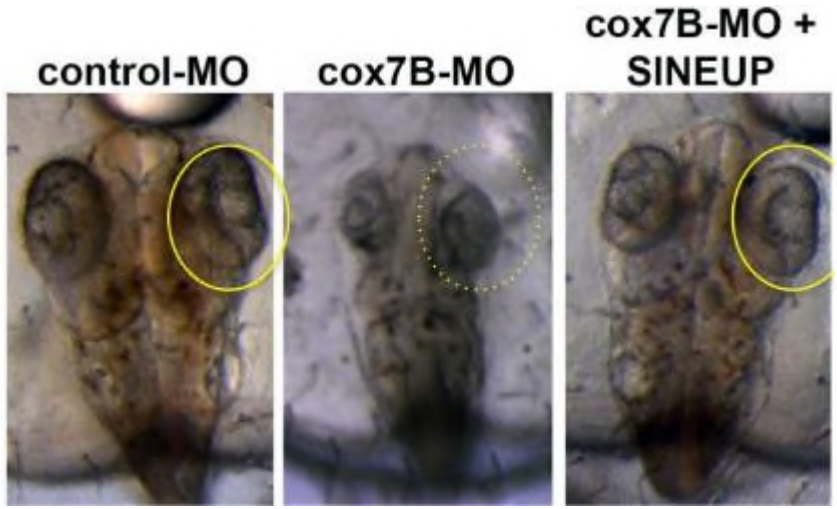
SINEUPs: *in vivo* model of human diseases in Medaka

Microphtalmia with Linear Skin Lesions (MLS)

***SINEUP*-Cox7B**







Embryo: **WT**

Eye: **Normal**

MLS 00% sk
l-blood c
ted that mutat
Diseased
CIV ld l d

MLS + SINEUP
not shown) m
Rescued
ng other com
~50%

e associa ed with
foot A igh diaph
h

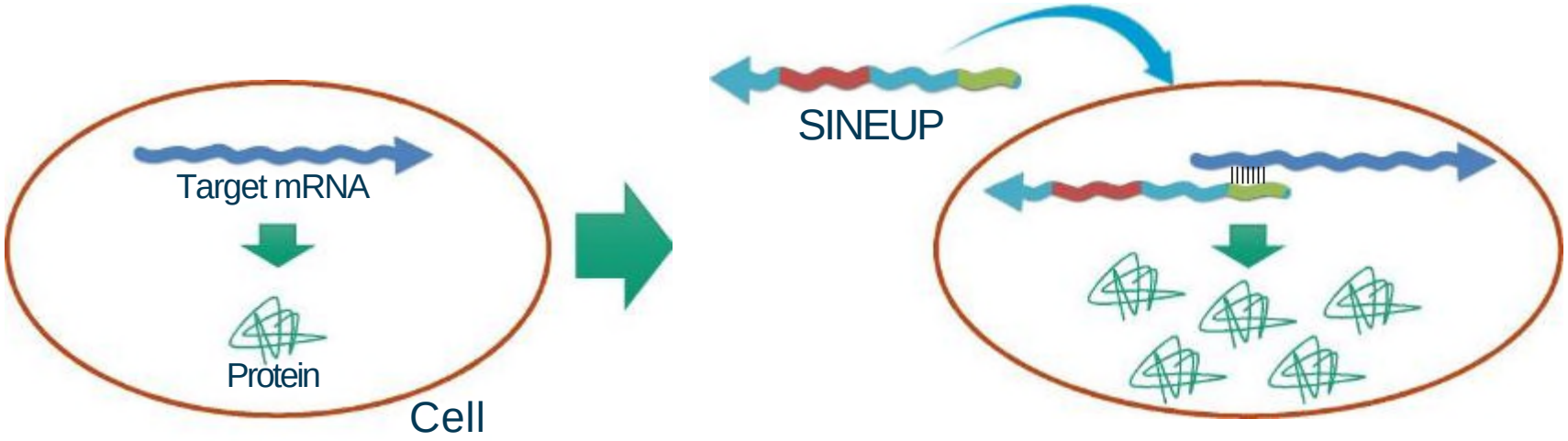
Indrieri A. *et al.*

SINEUPs

The first & only ncRNA to increase protein synthesis

(*Nature* 491, 454, 2012)

Mostly antisense are known for down-regulation



Laboratory

- Important for studying genes

Industry

- Efficient production of proteins
  - Therapeutic antibodies, enzymes.

Therapeutics

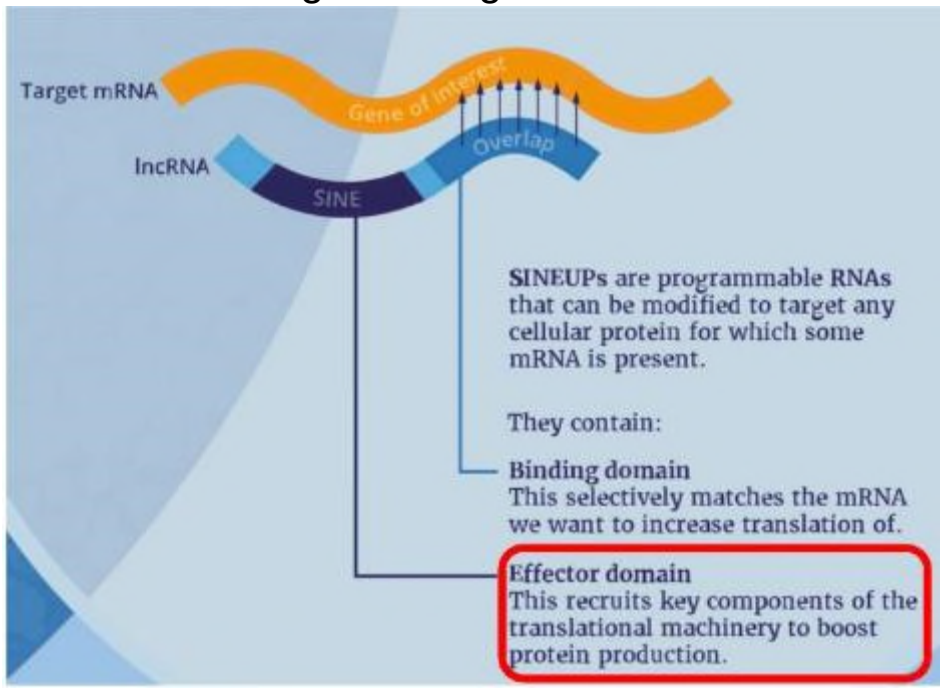
- Diseases caused by low-level proteins
  - Haploinsufficiencies - Neurodegenerative disorders

How many SINEUPs are there?

20 tested SINEs worked as SINEUPs

SINEUPs

SINEs UP regulate target mRNA translation



https://www.transinetx.com/

Harshita Sharma,

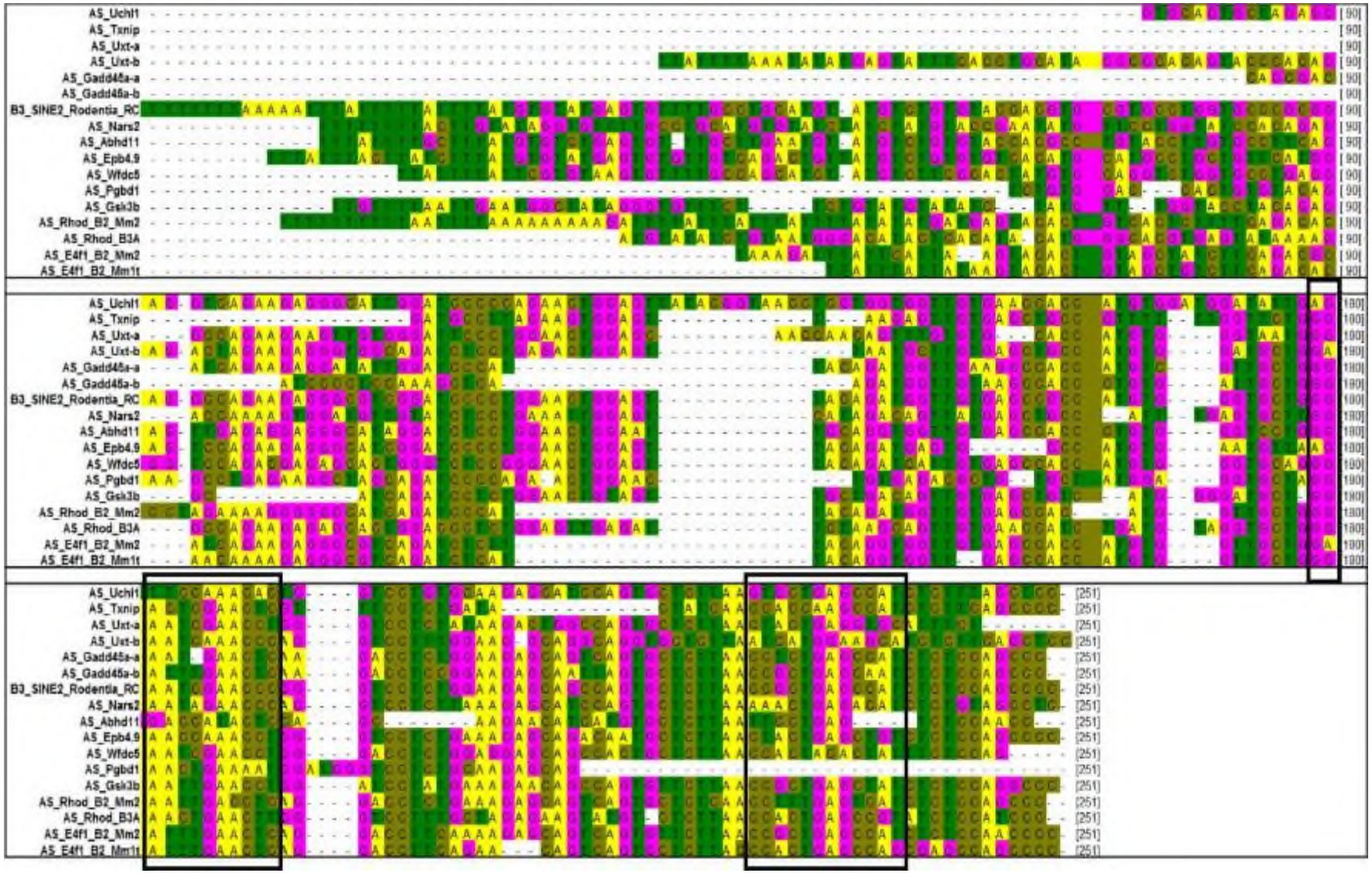
Sequence similarity:
As low as 25%

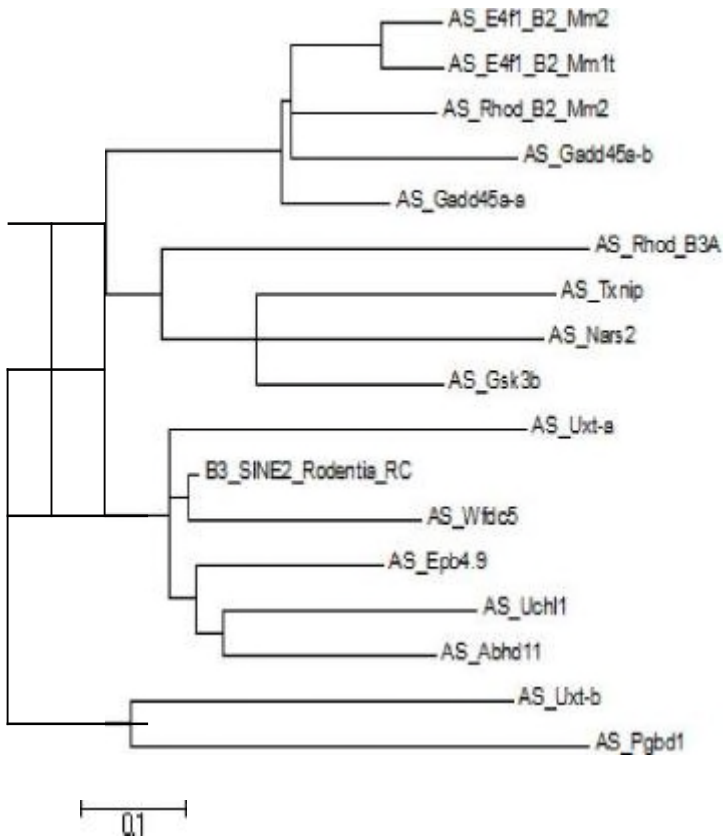
Schein *et al., Scientific Report* (2016)
Sharma *et al., bioRxiv*, May 22, 2023,
https://doi.org/10.1101/2023.05.22.541671

Species	embedded	SINE Class/Family	Length
1.Arabidopsis	consensus	SB4	159
1. Salmon	consensus	SmaI	150
1. Horse	consensus	ERE2	235
1. Mouse	AS-Uchl1	SINEB2/B3	167
1. Mouse	AS-Uxt	SINEB2/B3	131
1. Mouse	AS-Uxt	SINEB2/B3	187
1. Mouse	AS-Txnip	SINEB2/B3	107
1. Mouse	AS-Nars2	SINEB2/B3	214
1. Mouse	AS-Abhd11	SINEB2/B3	200
1. Mouse	AS-Epb4.9	SINEB2/B3	214
1. Mouse	AS-Wgdc5	SINEB2/B3	205
1. Mouse	AS-Pgbd1	SINEB2/B3	121
1. Mouse	AS-Gsk3b	SINEB2/B3	191
1. Mouse	AS-Rhod	SINEB2/B2_Mm2	205
1. Mouse	AS-Rhod	SINEB2/B2_B3A	190
1. Mouse	AS-E4f1	SINEB2/B2_Mm2	190
1. Mouse	AS-E4f1	SINEB2/B2_Mm1t	169
1. Human	pR12A-AS1	FRAM	128
1. Human	ITFG1-AS1	MIRb	202
1. Human	consensus	Ther1	252

Sequence variation and phylogenetic distance between mouse SINEB2 elements

Sequence and length are very variable in SINEs



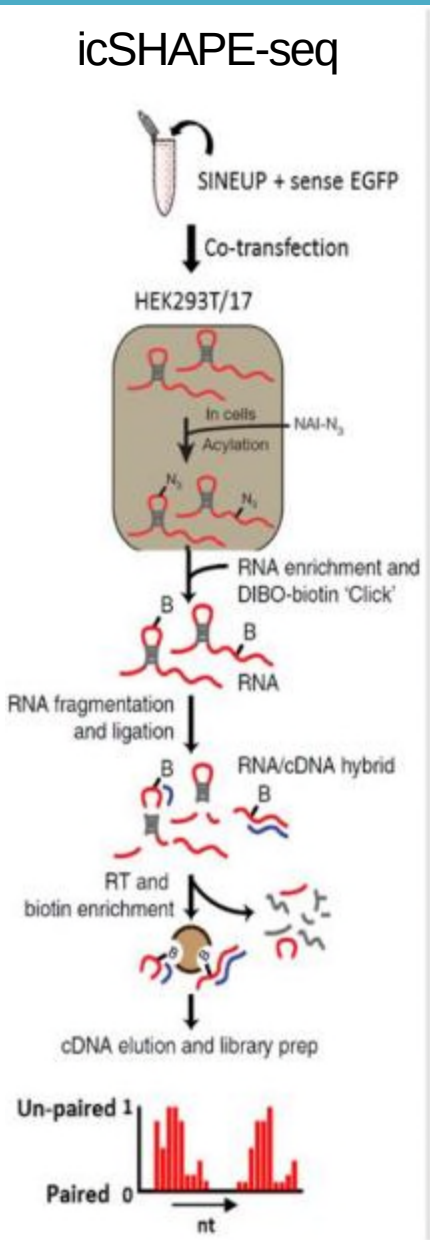


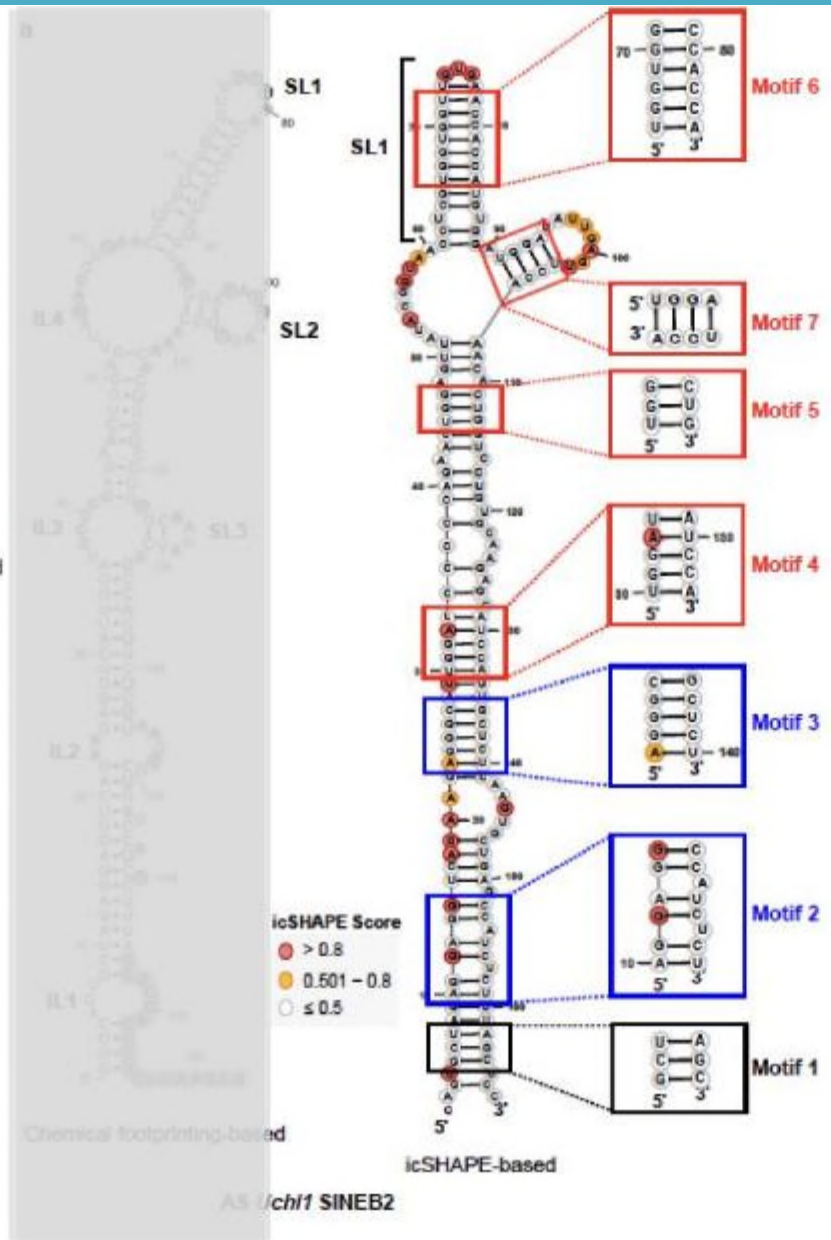
Question:

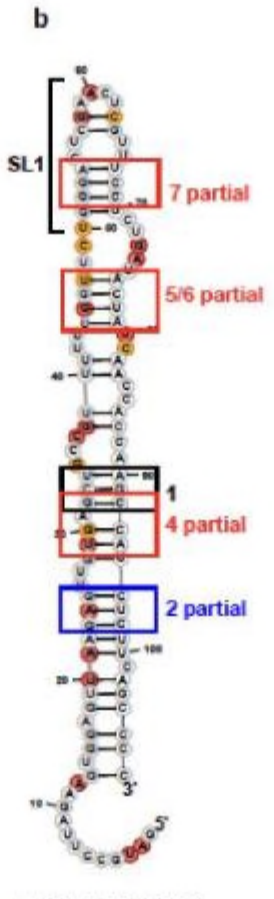
Do RNA structure provide the function more than sequence?

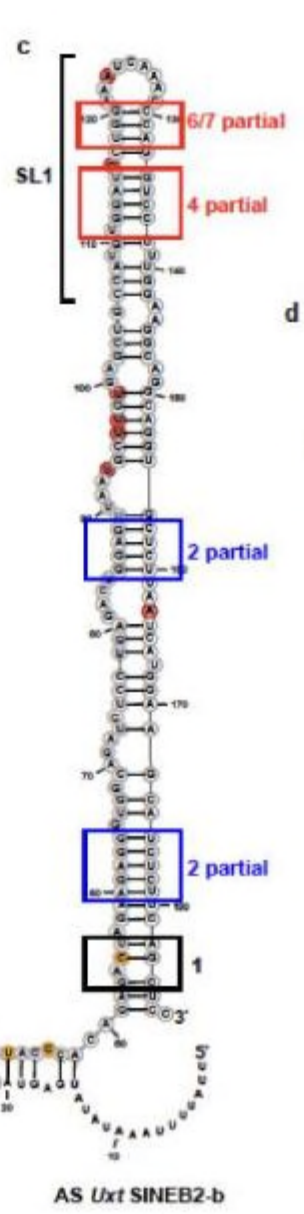
Different sequence SINEs contain similar structure motifs

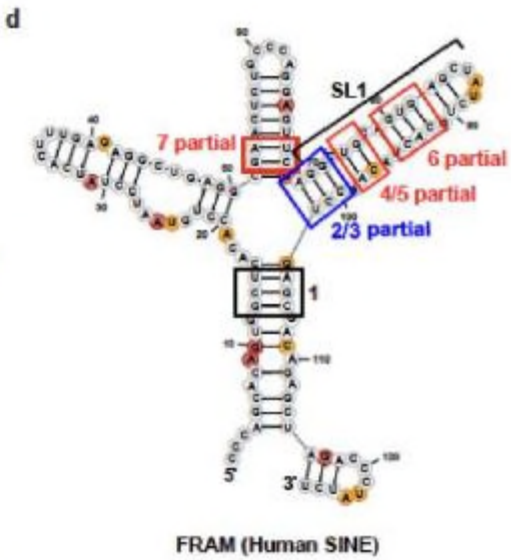
Detecting structure/motifs with icSHAPE









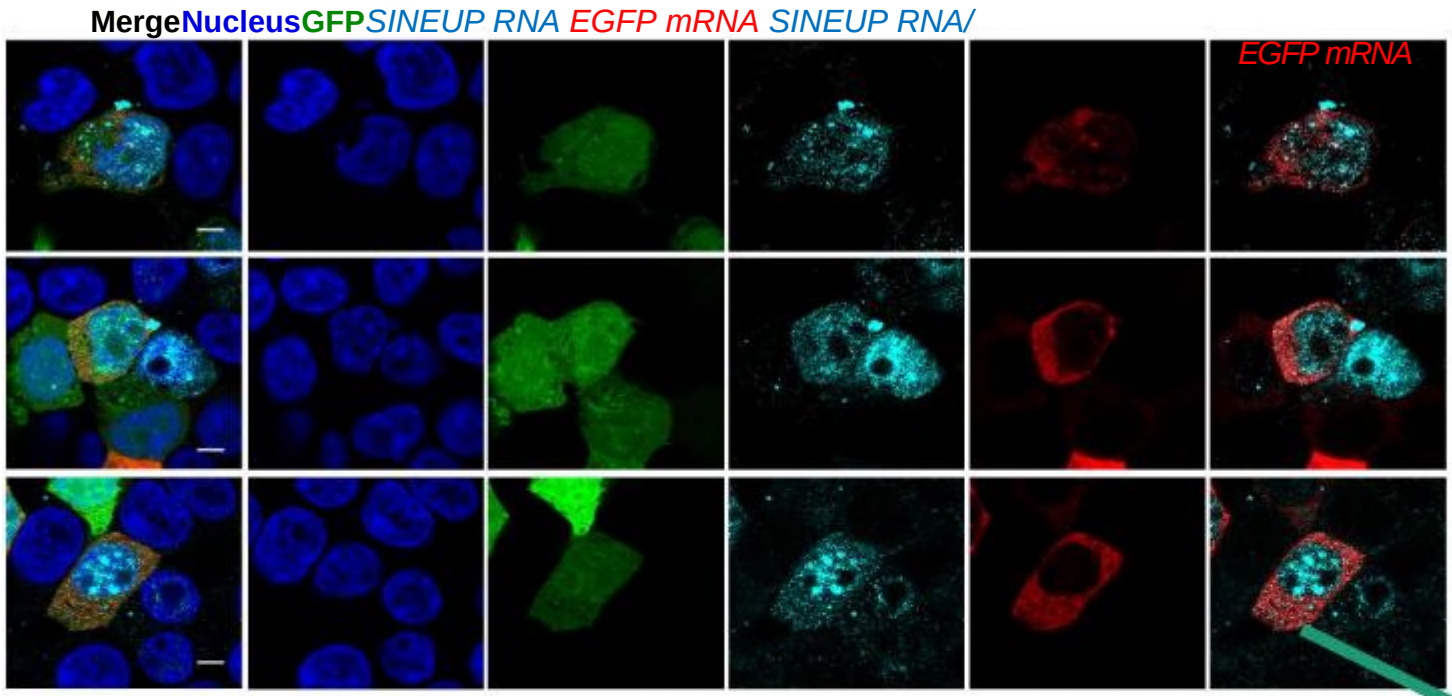


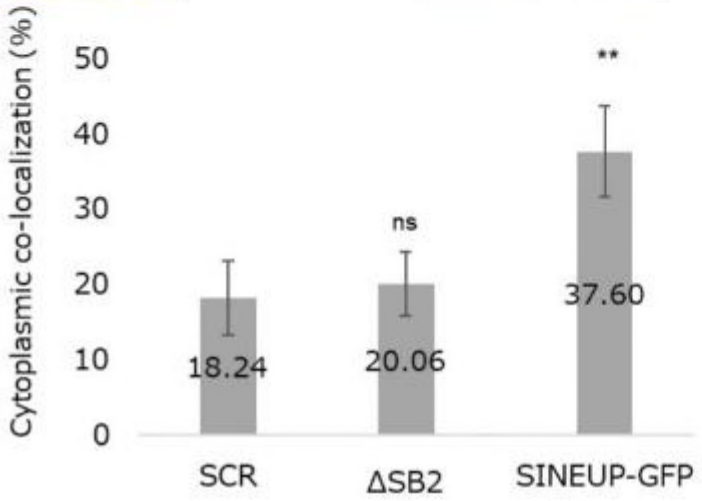
Sharma *et al.*, *bioRxiv*, May 22, 2023, https://doi.org/10.1101/2023.05.22.541671

✓ SL1/Motif6 are common in functional SINEUPs

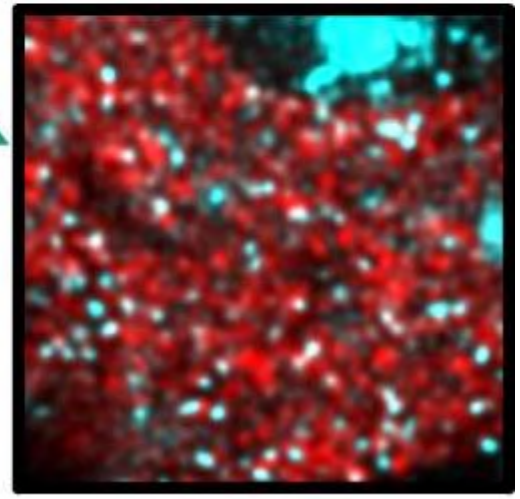
RNA localization is important for the function

Synthetic SINEUP-GFP tend to co-localize with target mRNA in the cytoplasm.





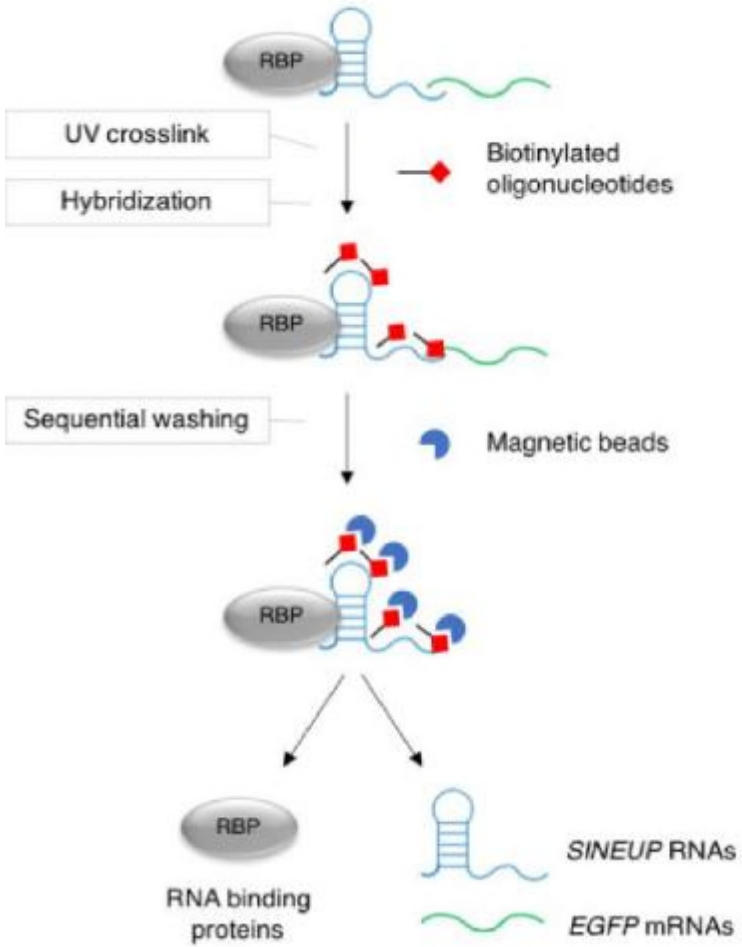
Naoko Toki

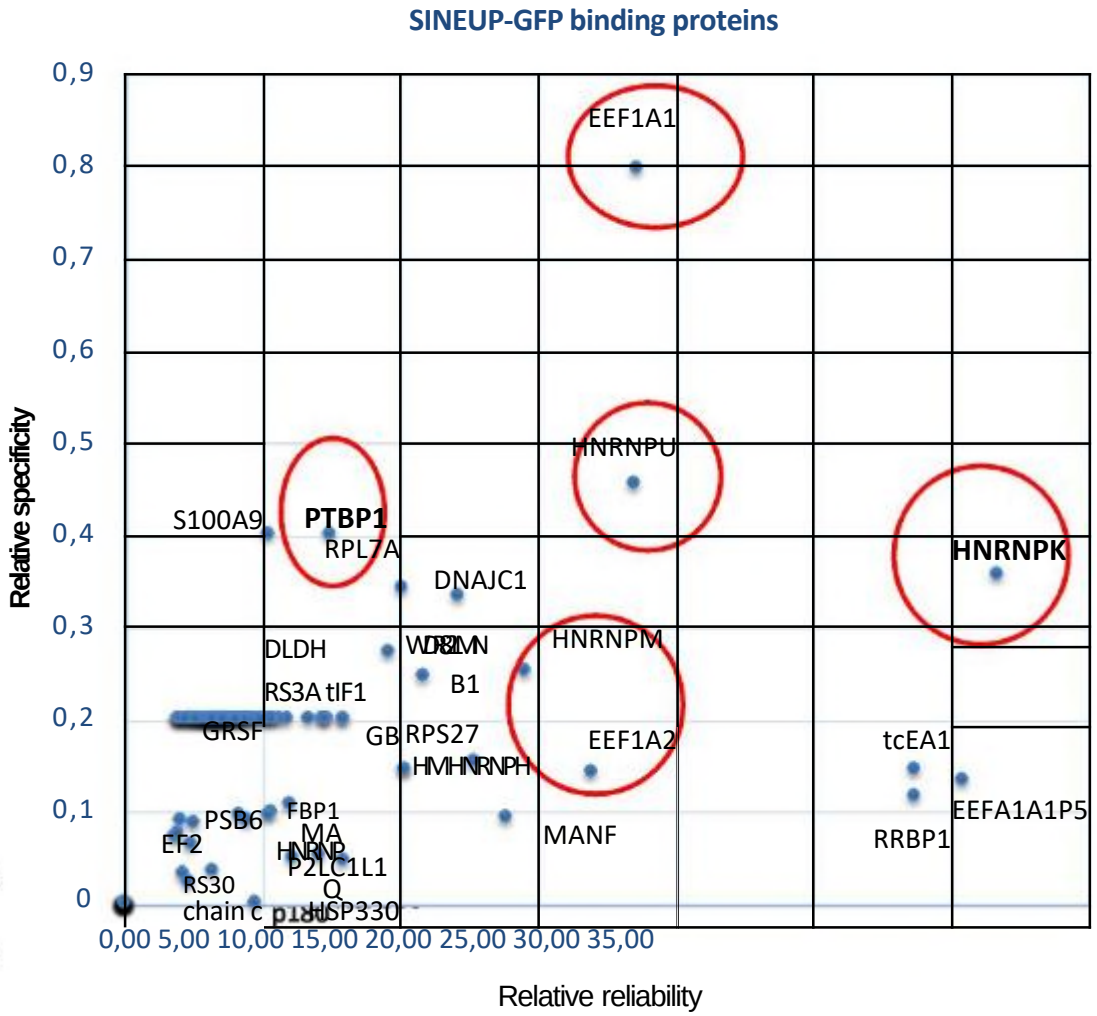


Toki *et al. Nucleic Acids Res.* (2020) biorxiv, (2019) Doi:http://dx.doi.org/10.1101/664029.

Mass Spec analysis to find SINEUP binding proteins

Many heterogenous nuclear ribonucleoproteins (hnRNPs) and translation elongation factors were the candidates of SINEUPs binding proteins.

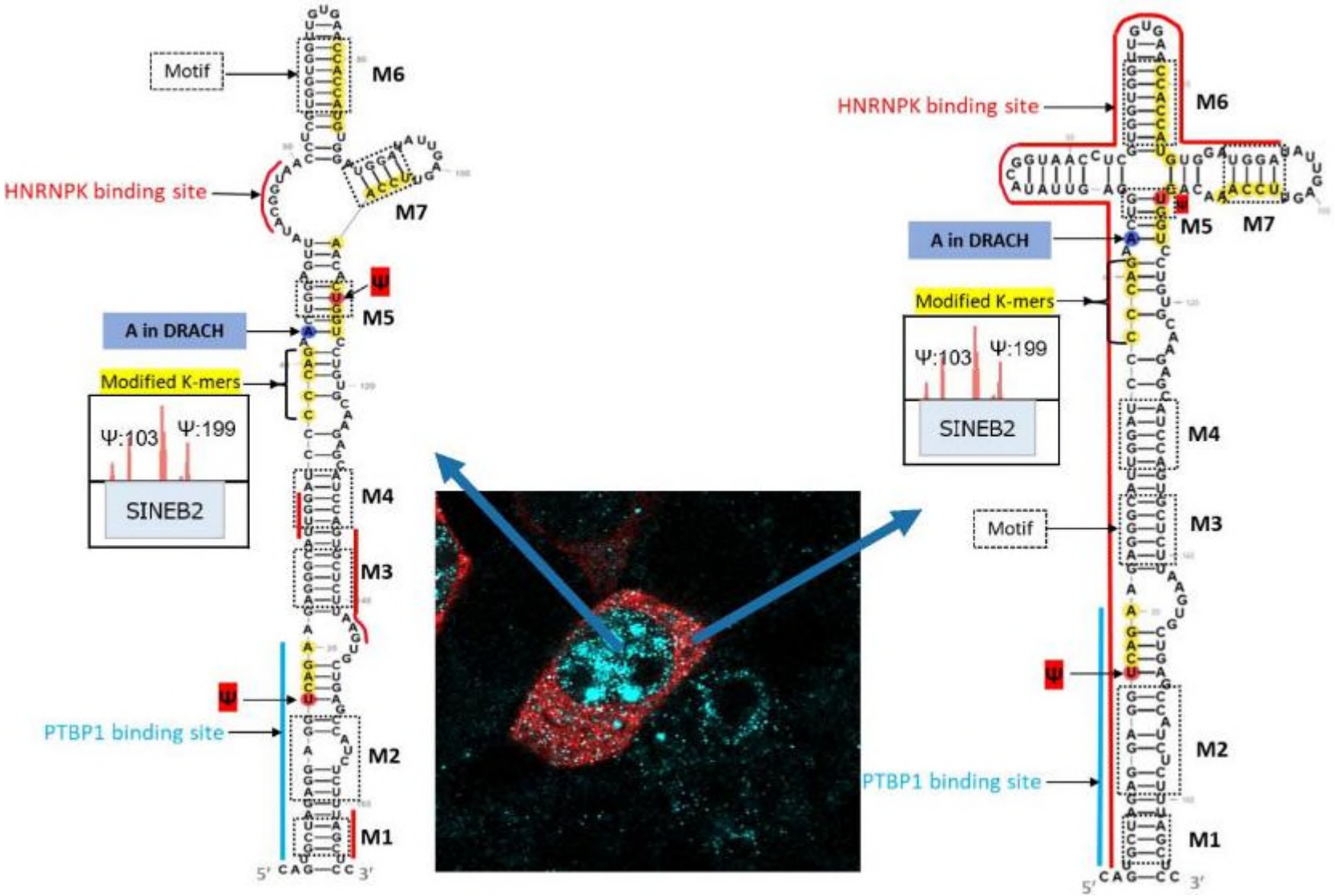




Toki *et al. Nucleic Acids Res.* (2020)
Doi:http://dx.doi.org/10.1101/664029.

SINEUPs structure and RBP binding regions dynamically change in cell compartments

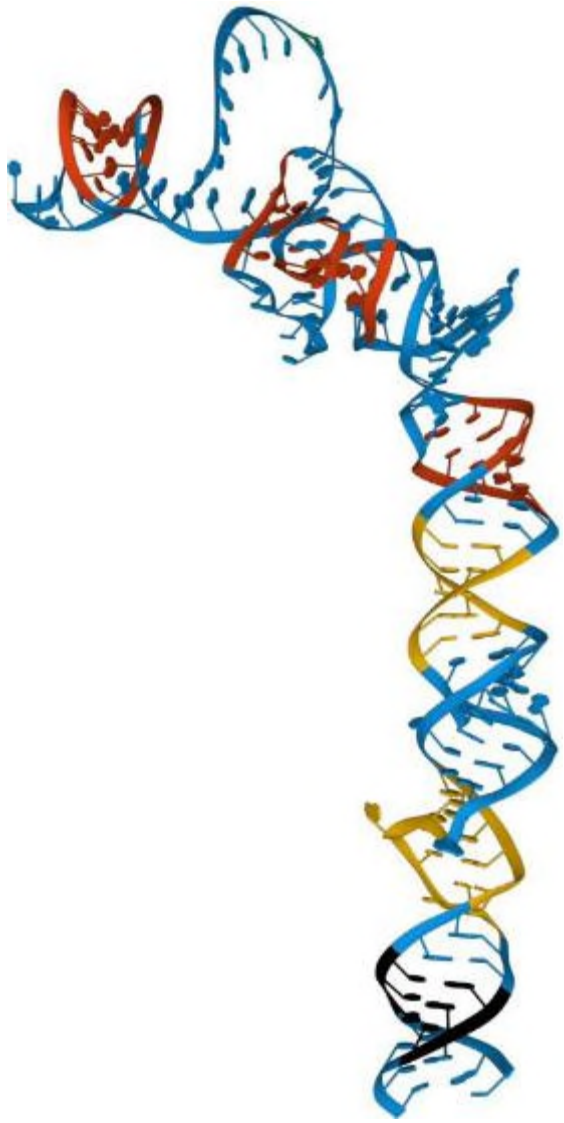
Nuclear structure Cytoplasmic structure



icSHAPE,

Fluctuation regions and HNRNPK binding regions are changed in the cellular compartments.

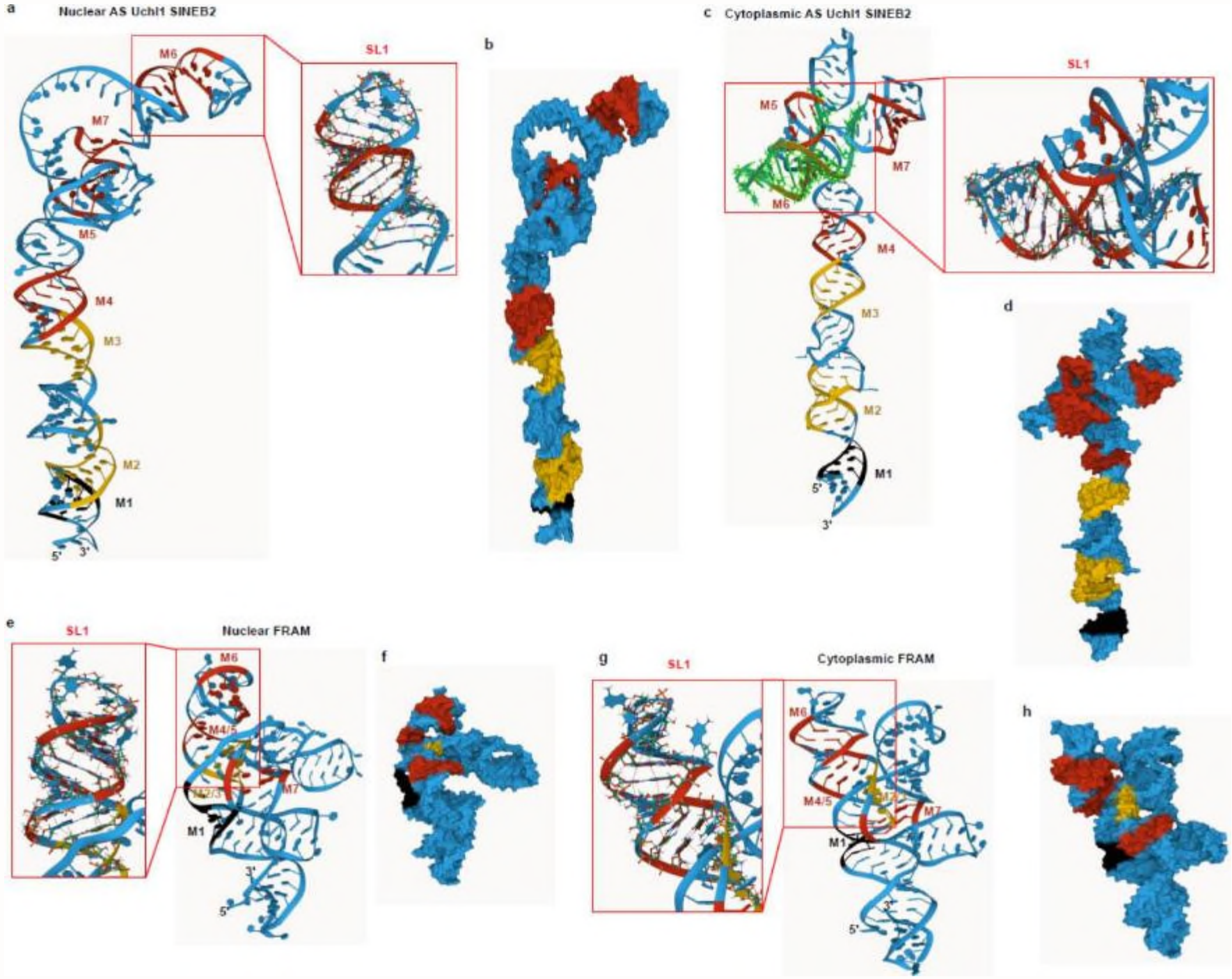
AS Uchl1 SINEB2 3D model based on secondary structure experimental data [animation]



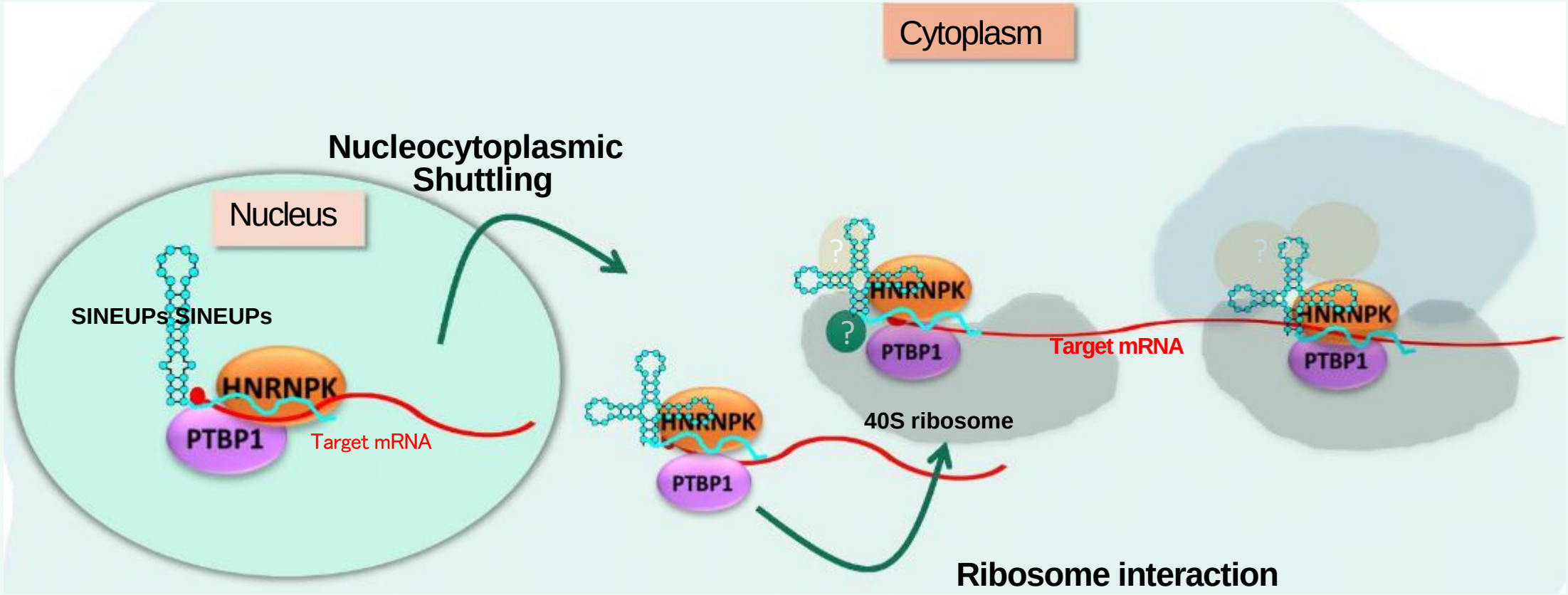
AS Uchl1 SINEB2 3D model: SL1 [animation]

SINE RNA 3D structure models reveal non-canonical intramolecular interactions

- Can be used as preliminary models to complement future experiment-based SINE RNA 3D structure studies



Dynamic changes of SINEUPs structure, RBP binding site and localization



There are still a lot of unknown proteins on the ribosome complex.

Regulating genome activity

- Many lncRNAs have regulatory

+

structural roles or other functions.

- New technologies to answer

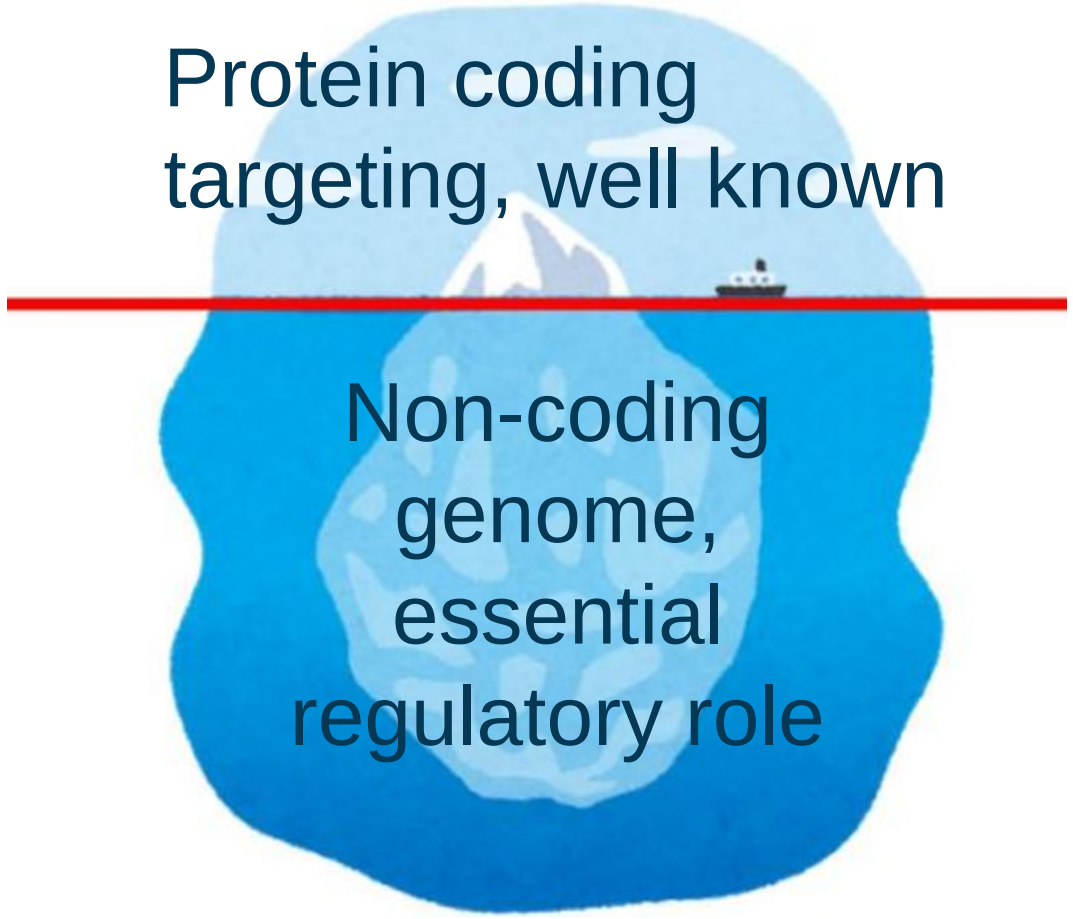
new

questions in gene regulation

  - RNA-chromatin networks to define new RNA role

- Common diseases are caused by

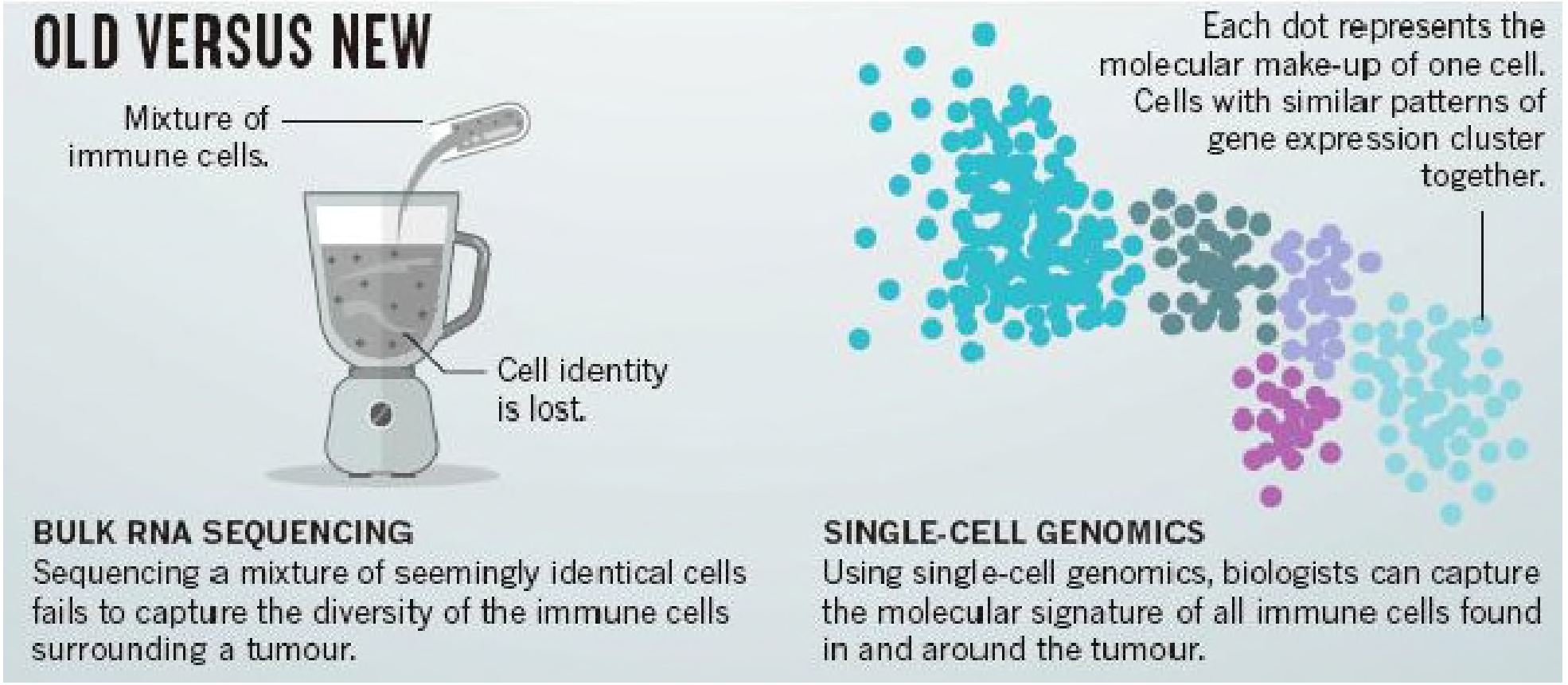
small unbalances of gene regulation.
lncRNAs are the new frontier for drug development.



Future opportunities in gene regulation

Transcriptome for ALL human cells

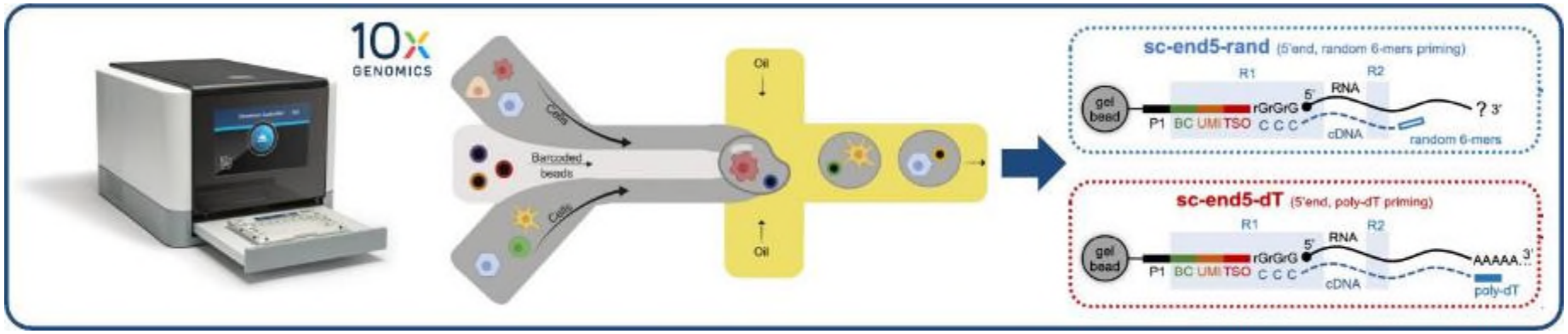
- Promoters, enhancers, lncRNAs as FANTOM5 but in “single cells”: from cell classification to precision genomics



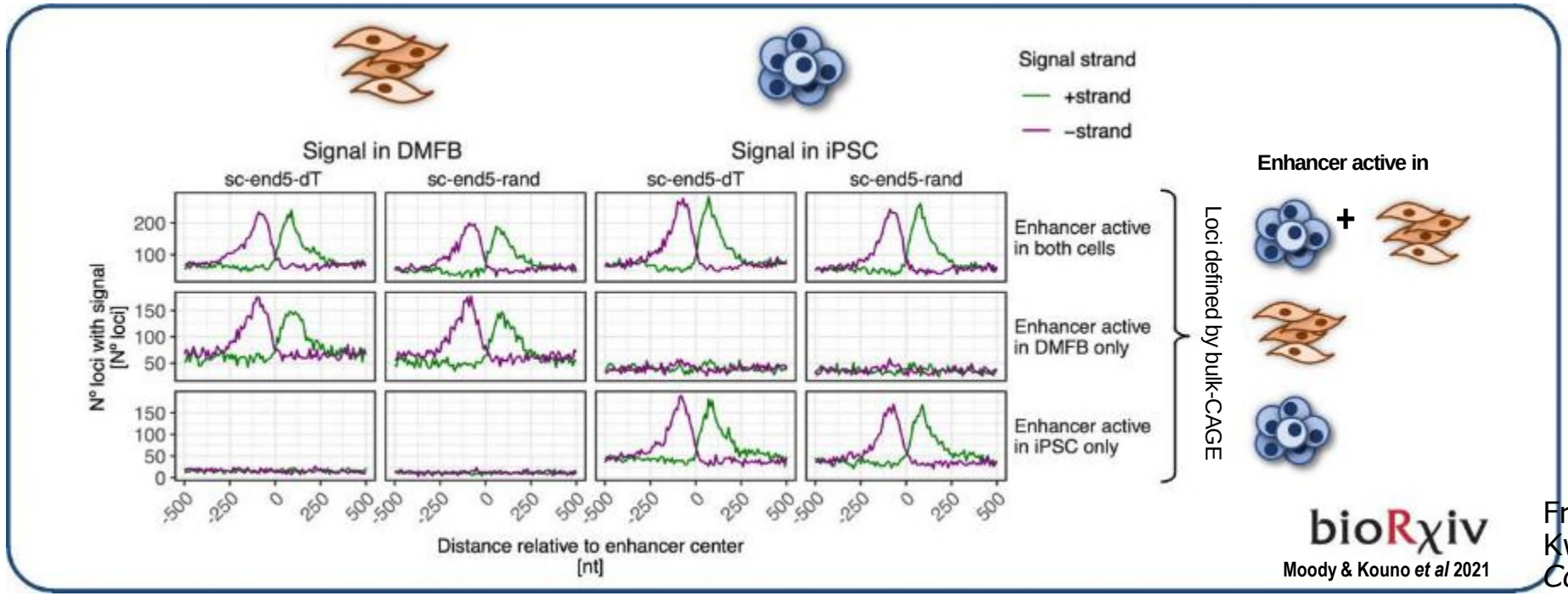
Giladi *et al. Nature* (2017)

We can now profile transcription and regulatory elements in single cells

5'end sc-RNA-Seq : Testing random and oligo-dT priming on 10X Genomics Chromium



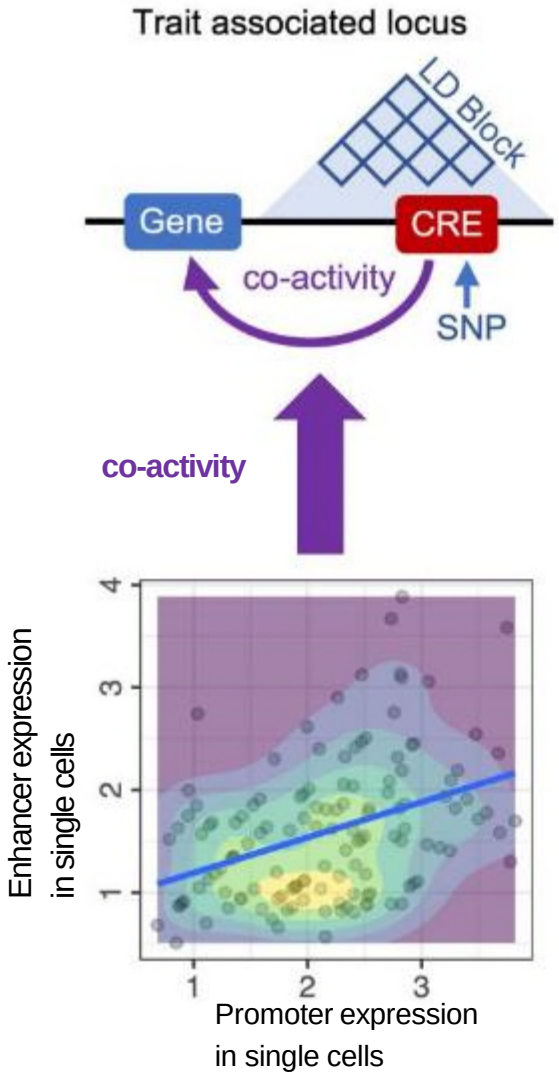
Enhancers : Both random and oligo-dT priming detects eRNA in single cells

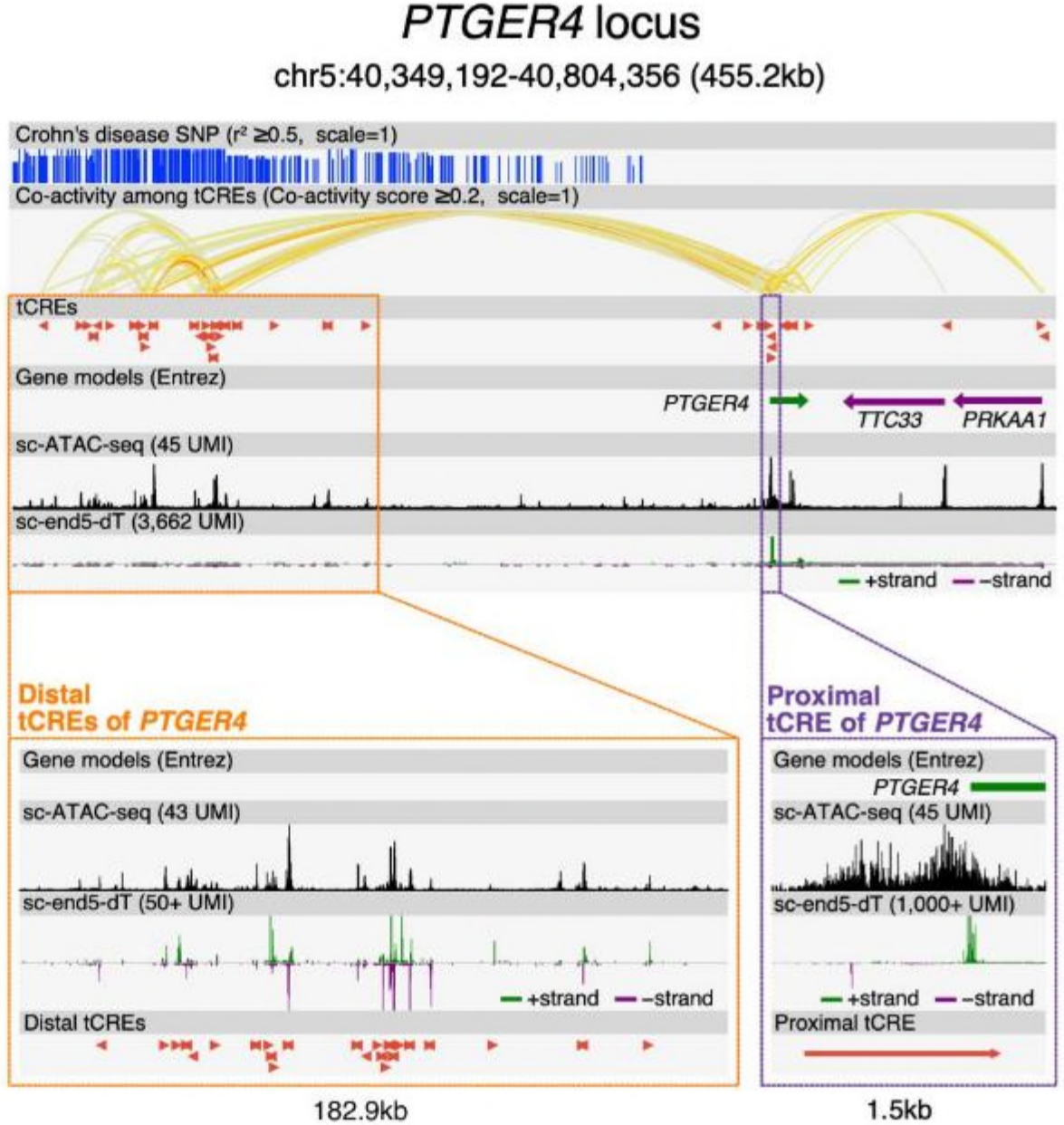


From Kouno T., Moody J., Kwon AT, et al. *Nature Communications* 2019

GWAS interpretation : Linking GWAS variants to candidate genes

We infer how genome sequence variants influence gene expression, in health and diseases, in all human cells



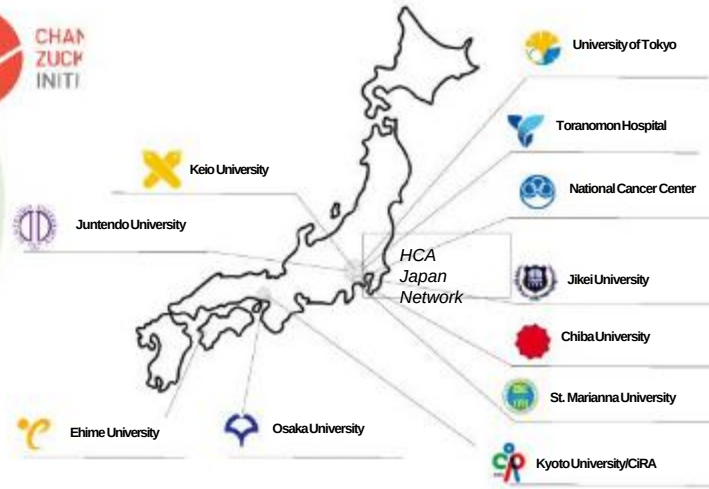


Moody et al. submitted; Biorxiv, https://doi.org/10.1101/2021.04.04.438388.

Human Cell Atlas (HCA)

Comprehensive reference maps of all human cells at single cell level

Domestic collaborations too!



Jay, Chung: leading "Asian Immune Diversity Atlas" The Seed Networks supported by CZI

Single Cell Medical Network in Japan (SC-Med)

HCA meetings

1st: Launch London Oct 2016	2nd: Technology Stanford Feb 2017	3rd: Computation Stockholm June 2017	4th: General Rehovot Oct 2017	5th: General Hinxton Mar 2018	6th: General Cambridge Nov 2018	7th: General Tokyo May 2019	8th: General Barcelona Oct 2019	9th: General Virtual June 2021	10th: General Hybrid Vienna June 2022	11th: General Toronto July 2023	**12th: General Milan Sept 2024**
			1st: Asia Okinawa Nov 2017		2nd: Asia Jeju Nov 2018		3rd: Asia Singapore Nov 2019	4th: Asia Virtual Oct 2020	5th: Asia Virtual Nov 2021	6th: Asia Hybrid Bangkok Nov 2022	7th: Asia Kolkata Nov 2023

- Single cell genomics
- High-resolution 3D imaging analysis
- Bioinformatics

➢ Reformulate our fundamental definition of cell types/development & human biology.

➢ Understand health & diseases at high-resolution.

FANTOM6-Interactome: acknowledgements

Piero Carninci, Wallace Yip, Jay Shin, Hazuki Takahashi, Masaki Kato, Takeya Kasukawa






Advanced Genomics Circuit
Kayoko Yasuzawa
Anika Prabhu
Callum Parr
Tsukasa Kouno
Yan-Jun Lan
Youtaro Shibayama
Fernando Lopez Redondo
Julio Jesus Leon Incio
Masayoshi Itoh

Transcriptome Technology
Mitsuyoshi Murata
Satoshi Takizawa
Kentaro Kaji
Matthew Valentine
Lokesh Pati Tripathi
Xufeng Shu
Harsita Sharma
Hiromi Sueki

Large-scale Biomedical Data Technology
Takeya Kasukawa
Tomoe Nobusada
Imad Abugessaisa
Jessica Severin

Genome Information Analysis
Chung-Chau Hon
Jonathan Moody

IMS Genome sequencing Platform

Comprehensive Genomic Analysis
Yasushi Okazaki
Kokoro Ozaki
Tsugumi Kawashima
Hiroko Kinoshita
Shohei Noma
Chitose Takahashi
Michihira Tagami

Cellular Epigenetics
Jen-Chien Chang

Callum Parr
Hazuki Takahashi
Masaki Kato
Wallace Yip
Hiromi Sueki
Kayoko Yasuzawa
Michihira Tagami

KTH Royal Institute of Technology
Pelin Sahlén

Università degli Studi di Milano
Beatrice Bodega
Valeria Ranzani


Karolinska Institute
Andreas Lennartsson
Magda Bienko
Wenjing Kang


Human Technopole
Aslihan Karabacak Calviello
Rodrigo Pracana
Laura Carpen
Marco Gaviraghi
Ilaria Nisoli
Chiara Medaglia
Magda Bienko
Nicola Crosetto
Lorenzo Calviello
Ivano Legnini



Piero as a person

Special thanks to:

RIKEN
Michiel de Hoon
Chung Chau Hon
Jonathan Moody
Takeya Kasukawa
Tsukasa Kouno
Jessica Severin
Masayoshi Itoh
Harukazu Suzuki
Hazuki Takahashi
Harshita Sharma
Masaki Kato
Naoko Toki
Erik Arner
Alessandro Bonetti
Jay W. Shin
Chi Wai (Wallace) Yip
Matthew Valentine Xufeng Shu
Lokesh Tripathi

Human Technopole
Yoshimi Inaba
Laura Carpen
Marco Gaviraghi
Ilaria Nisoli
Chiara Medaglia
Aslihan Karabacak
Rodrigo Pracana
Magda Bienko
Lorenzo Calviello

Tokyo Metropolitan Institute of Medical Science
Hideya Kawaji

Yokohama City University
Jordan Ramilowski

Harry Perkins Institute of Medical Research
Alistair Forrest

Telethon Kids Institute
Timo Lassmann

Karolinska Institutet
Wenjing Kang
Andreas Lennartsson
Carsten Daub

Univ. Copenhagen
Albin Sandelin
Robin Andersson
Kristoffer Vitting-Seerup

The Roslin Institute
Kenneth Baillie
Tom Freeman
Kim Summers

Univ. Edinburgh
Martin Taylor

Birmingham Univ.
Ferenc Mueller

Imperial College London
Boris Lenhard

Univ. Hosp Regensburg
Michael Rehli

KAUST
Valerio Orlando

Univ. Melbourne
Christine Wells

Univ. Piedmont
Claudio Santoro
Diego Cotella

TIGEM
Alessia Indrieri
Brunella Franco

Università degli Studi di Milano
Beatrice Bodega

IIT Istituto Italiano di Tecnologia
Stefano Gustincich

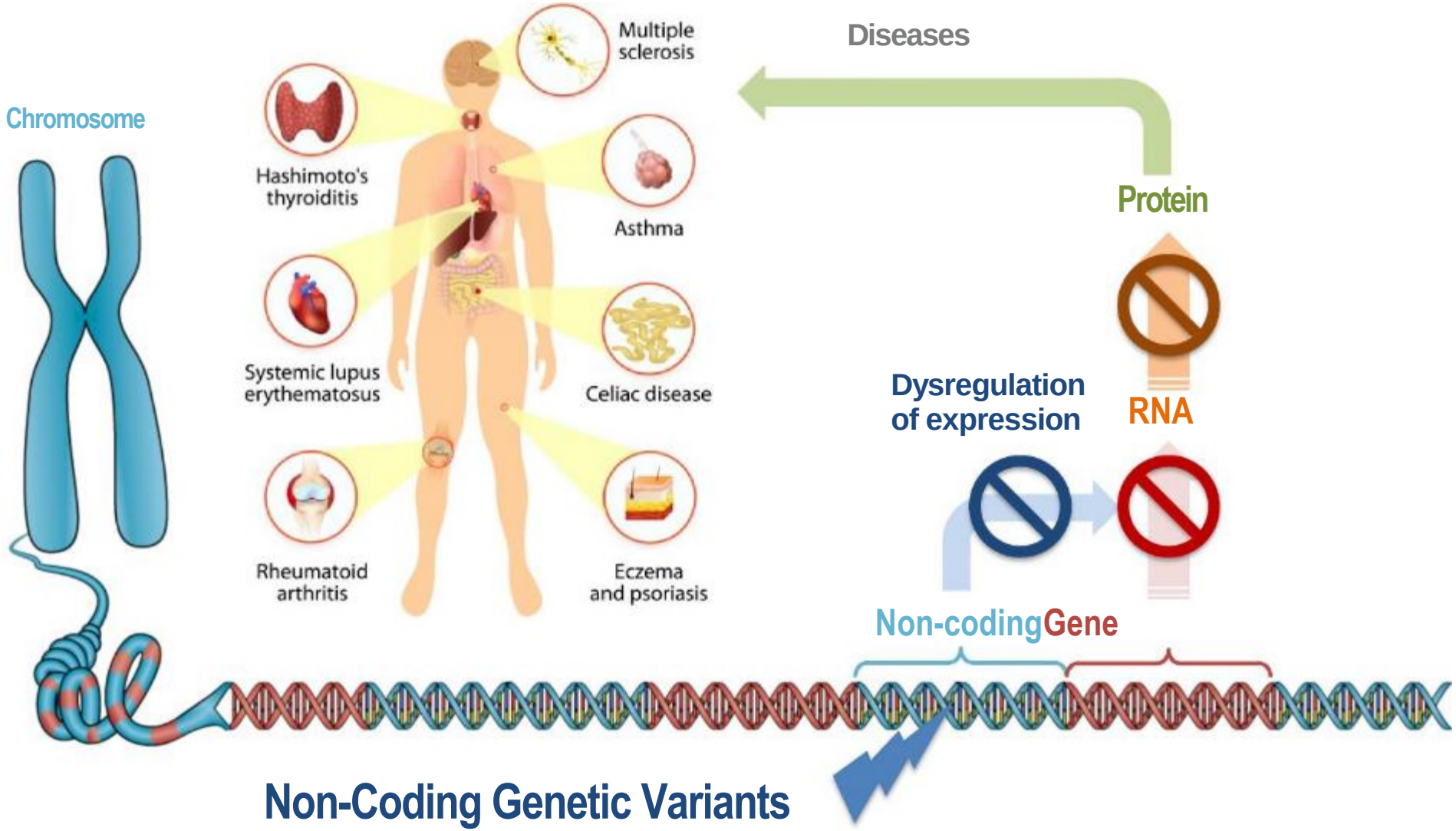
Funding
- Horizon2020,
- ZENCODE-ITN
- MEXT, Japan
- AMED, Japan

Institutional funding
- RIKEN
- Human Technopole

And to many more FANTOM Consortium members, lab members, students, historical collaborators...

Taking advantage of single cell work- single cell CAGE

CAGE approach to map regulatory elements and map Predispositions to Diseases



Slide, Chung Chau Hon

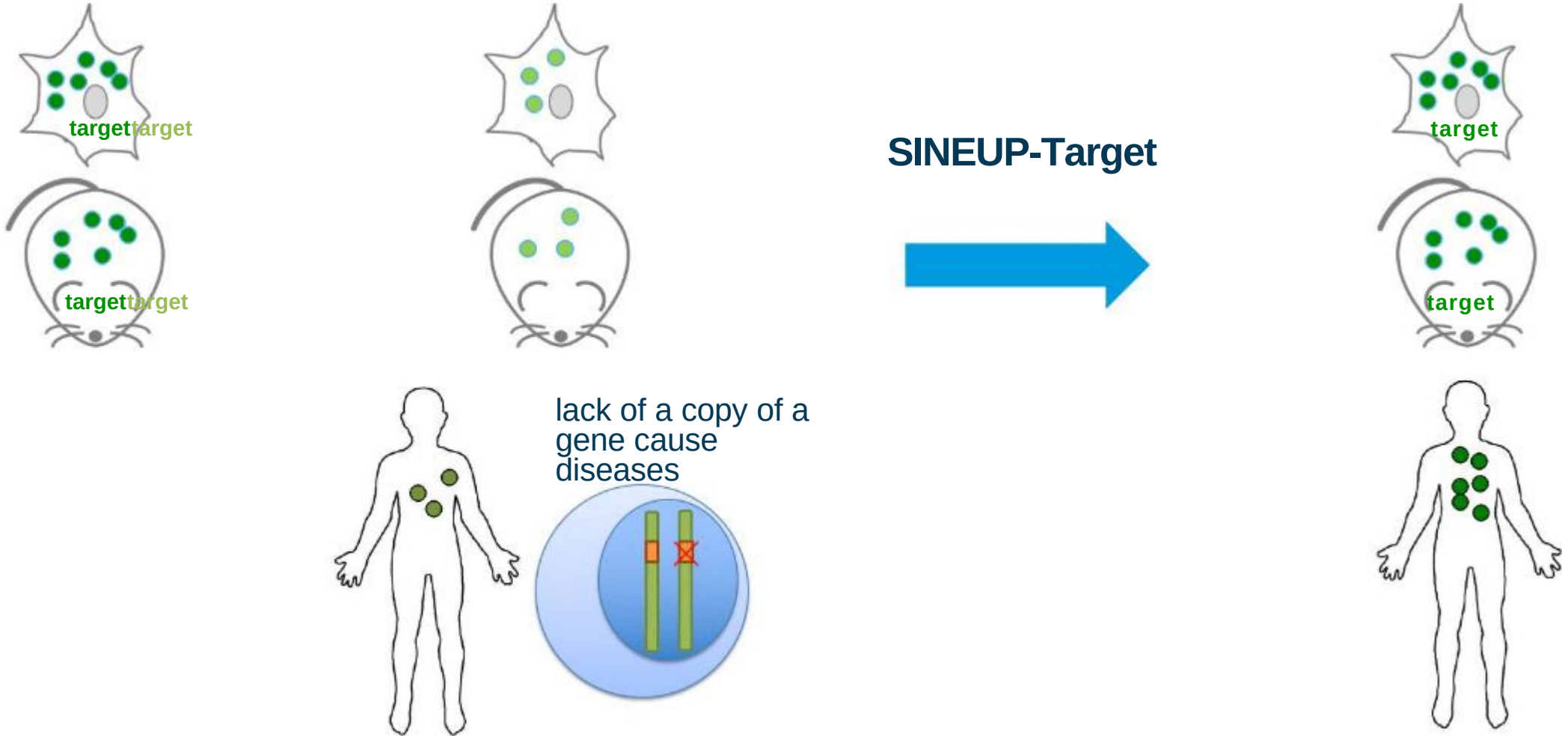
The 7th HCA General Meeting in Japan

Coordination of HCA in Japan and in Asia (HCA executive office at RIKEN)

SINEUPs Therapeutics – haploinsufficiencies & others

At least 300 haploinsufficiency genes known

(Dang *et al*., European Journal of Human Genetics, DOI: 10.1038/ejhg.2008.111, 2008)



Healthy Haploinsufficiencies Therapeutic SINEUPs

Or conditions that require more protein

https://humantechnopole.it/en/

Piero Carninci

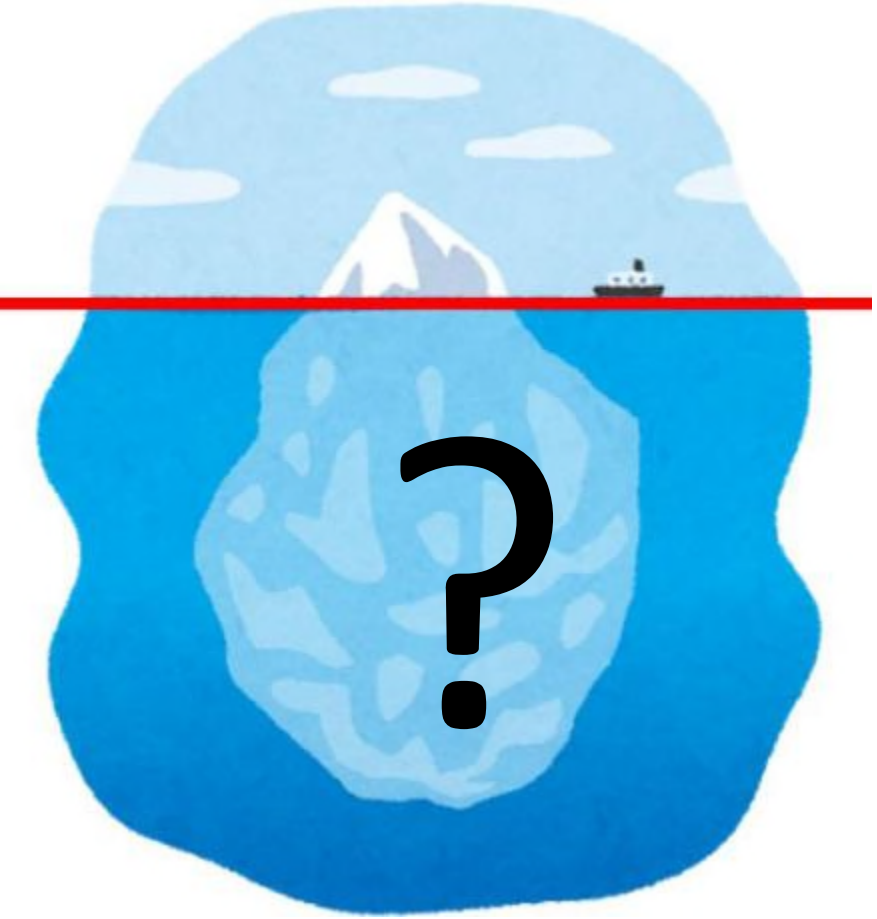
Genomics Research Center, Human Technopole, Milan, Italy

RIKEN IMS, Yokohama

- Creating bridges between Japanese and ... science
- New positions and chances to collaborate

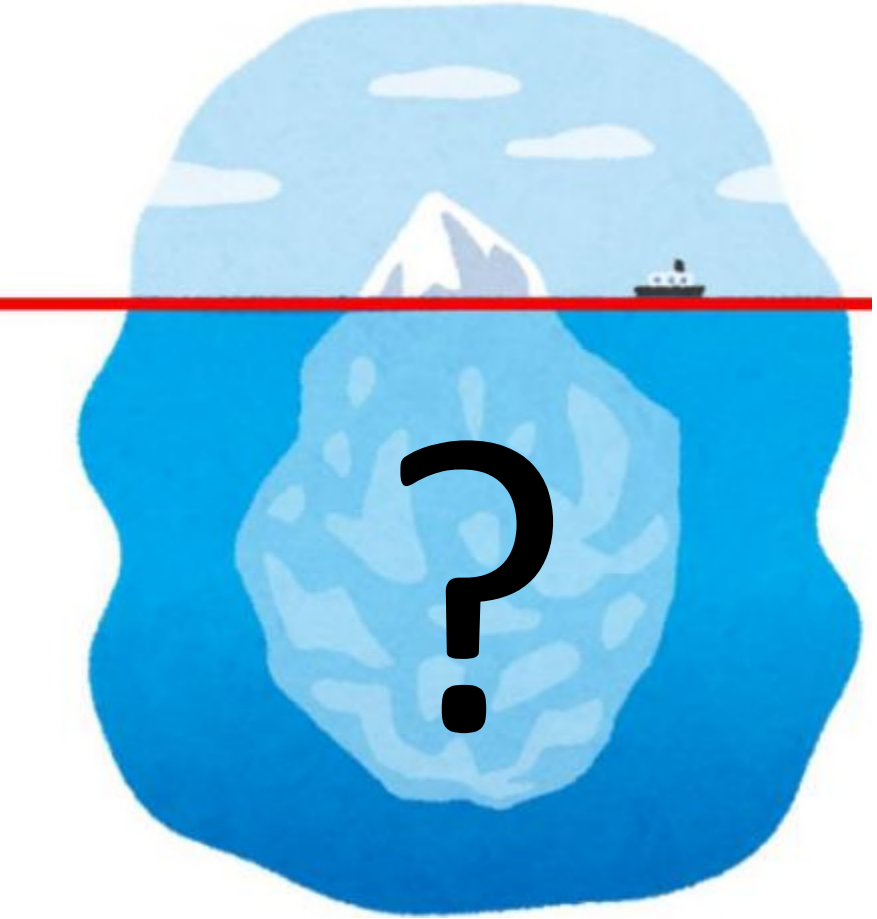
Thank you

Many unanswered questions



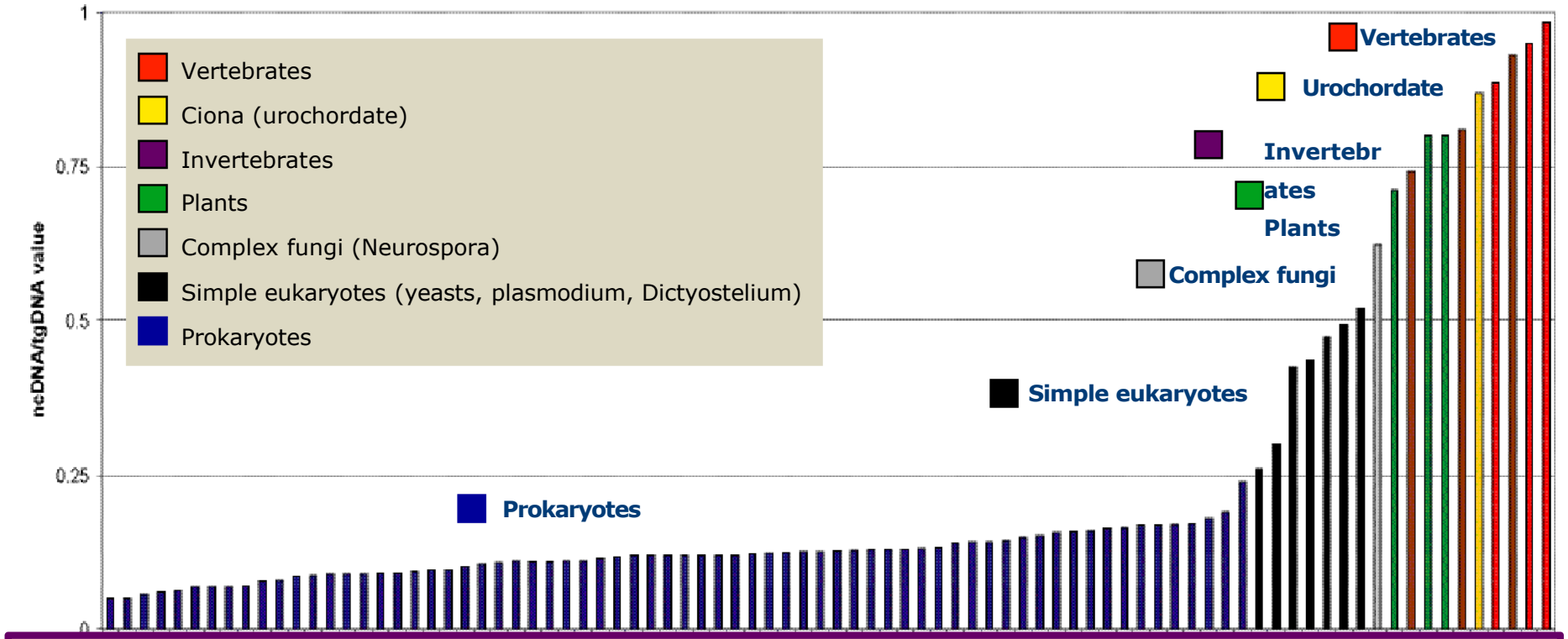
- Cell variation RNA-chromatin interactomes
  - Housekeeping, specific lncRNAs/RNAs
  - Stimulation, differentiation, dependent
- RNAs involved?
  - sequence, structural domains, RBP proteins interactome
  - Relation and cooperation with 3D Chromatin structure
- Diseases:
  - RNAs on GWAS, genetics map, eQTL
- A database of RNA bound to chromatin
  - A database on all RNA variants

Resources for biology



- A broad RNA/chromatin interaction map
- A full-length cell specific, compartment specific map of RNA; and their modifications.
  - How many chromatin regulatory RNA?
    - >100s cell types
  - How many functional RNA variants?
    - >100 cell types + >100s tissues

The proportion of noncoding DNA broadly increases with developmental complexity



Irrespective of the extent of non-coding sequences, it is now evident that the vast majority of the genomes of all organisms is transcribed in a dynamic manner in different cells and tissues at different developmental stages.

J.S. Mattick *Nature Reviews Genetics* 5, 316-323 (2004)
R.J. Taft, M. Pheasant and J.S. Mattick, *Bioessays* 29, 288-299 (2007)

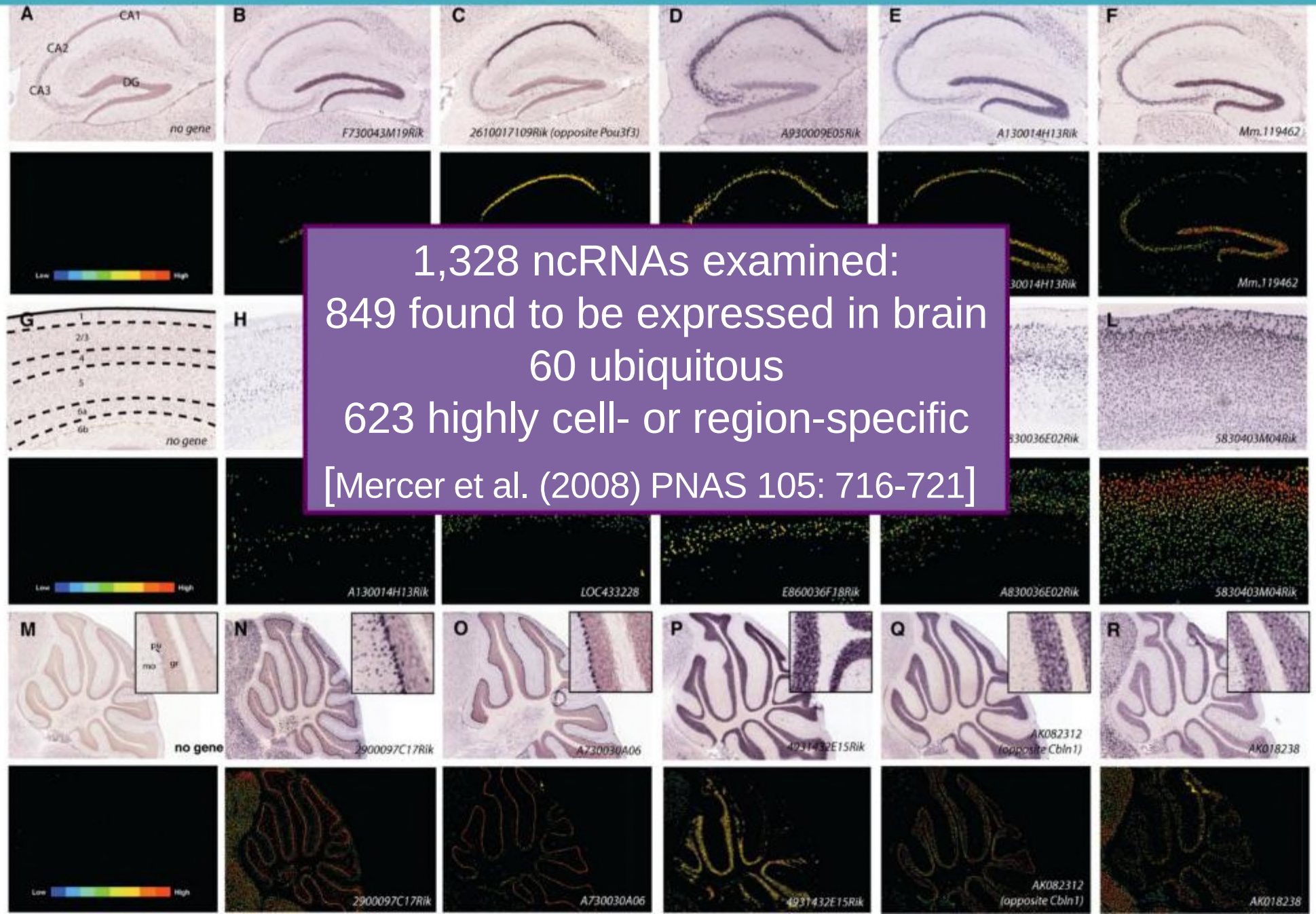
lncRNA essential

- Theoretical basis: regulators of very complex operations.

- More validations:
- Are they expressed and where?
- Can we prove the function?

- Some beautiful and meaningful picture of long-noncoding RNA (lncRNAs) and their
localization.

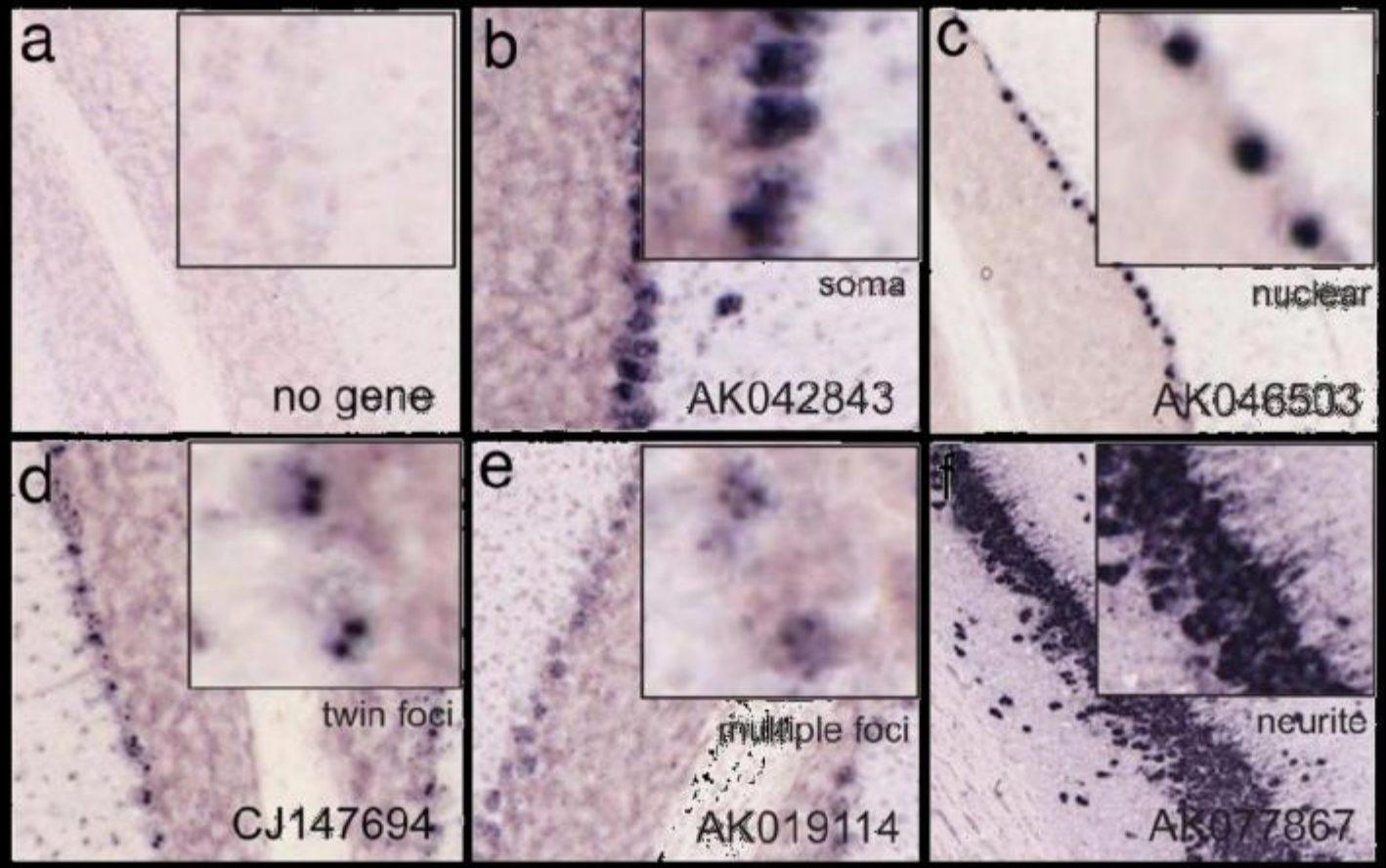
Non-coding RNA expression in mouse brain



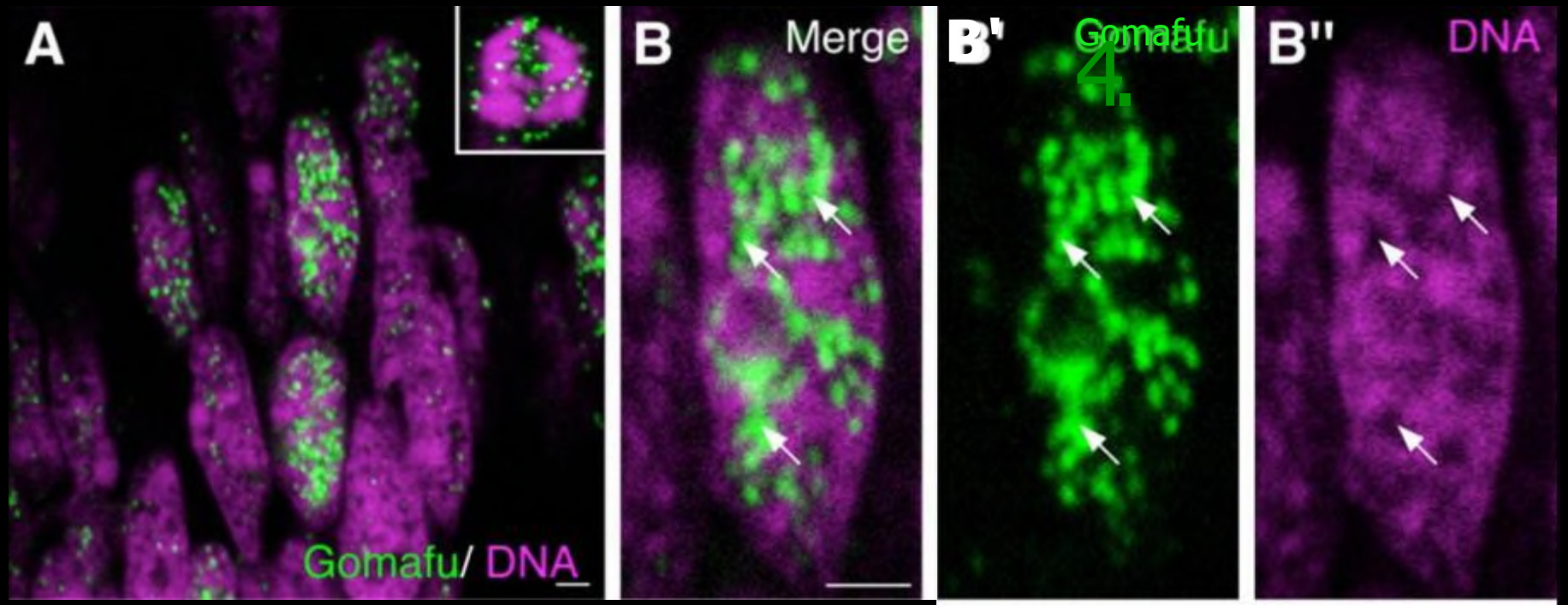
1,328 ncRNAs examined:
849 found to be expressed in brain
60 ubiquitous
623 highly cell- or region-specific

[Mercer et al. (2008) PNAS 105: 716-721]

Subcellular localization of ncRNAs



Subcellular localisation of long ncRNAs in Purkinje cells Subcellular localization ~ putative function

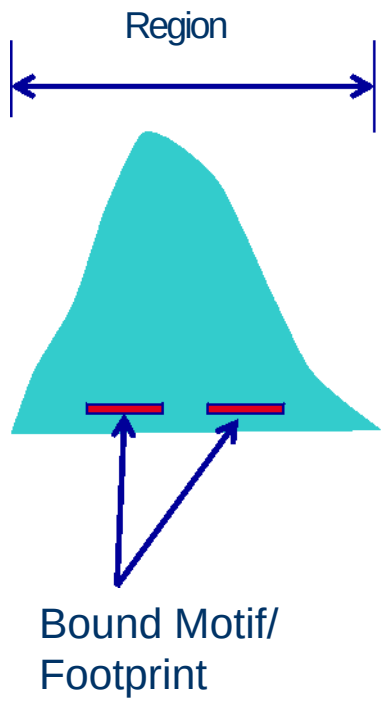


4.

✓ **Subsequent papers confirming the findings**

ENCODE

Element Type	Coverage	Cumulative Coverage
Exons	3%	3%
Chip-seq bound motifs	4.5%	5%
DNaseI Footprints	5.7%	9%
Chip-seq bound regions	8.1%	12%
DNaseI HS regions	15.2%	19.4%
Histone Modifications (*)	44%	49%
RNA	**62%**	**80%**
(* excluding broad marks)		

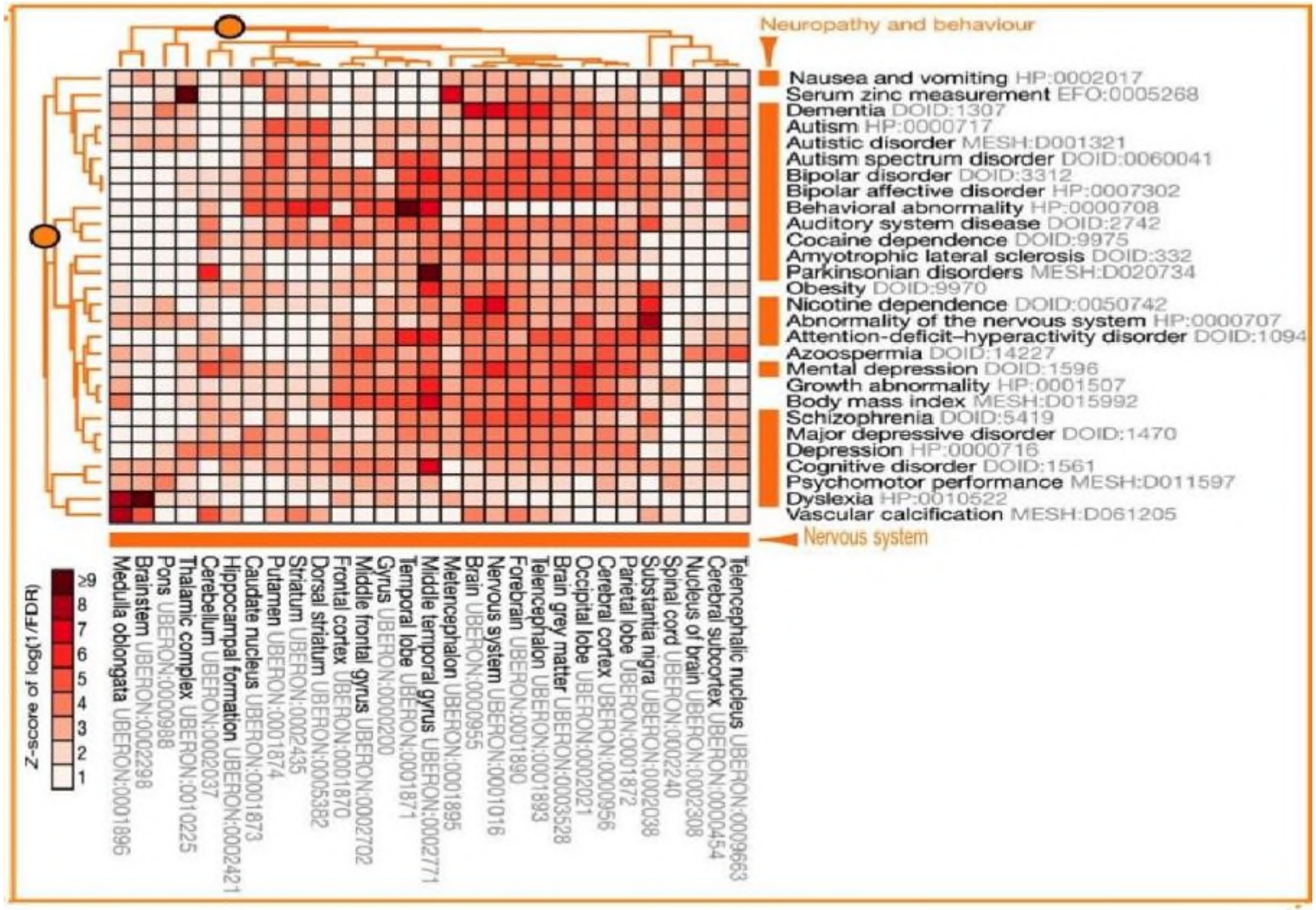


(Union over all experiments and cell types)

In 2012, the ENCODE confirms that genome is broadly transcribed

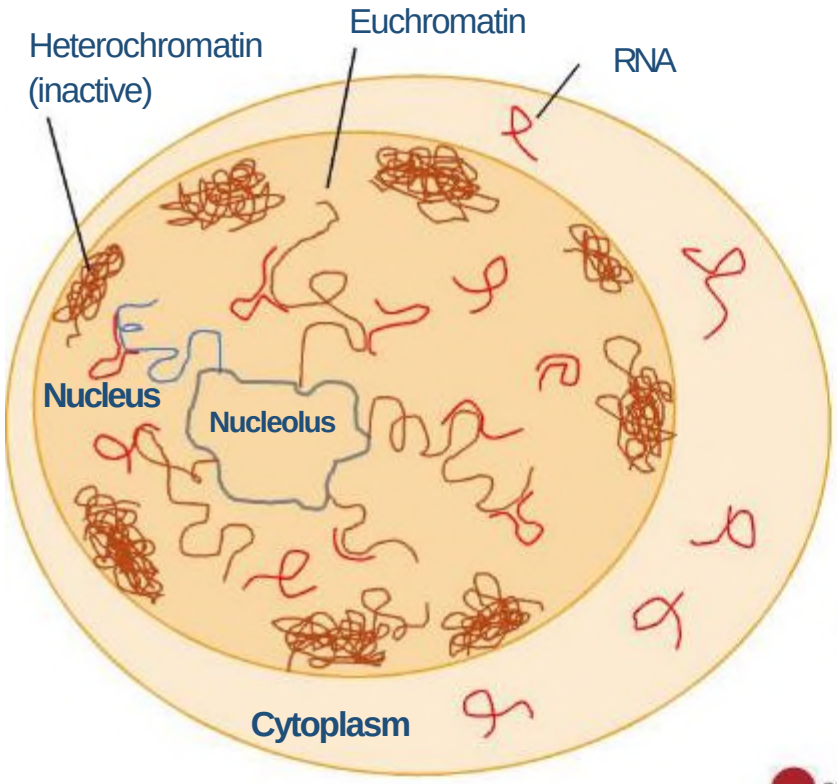
ENCODE Project Consortium, Nature 489, 57 (2012)

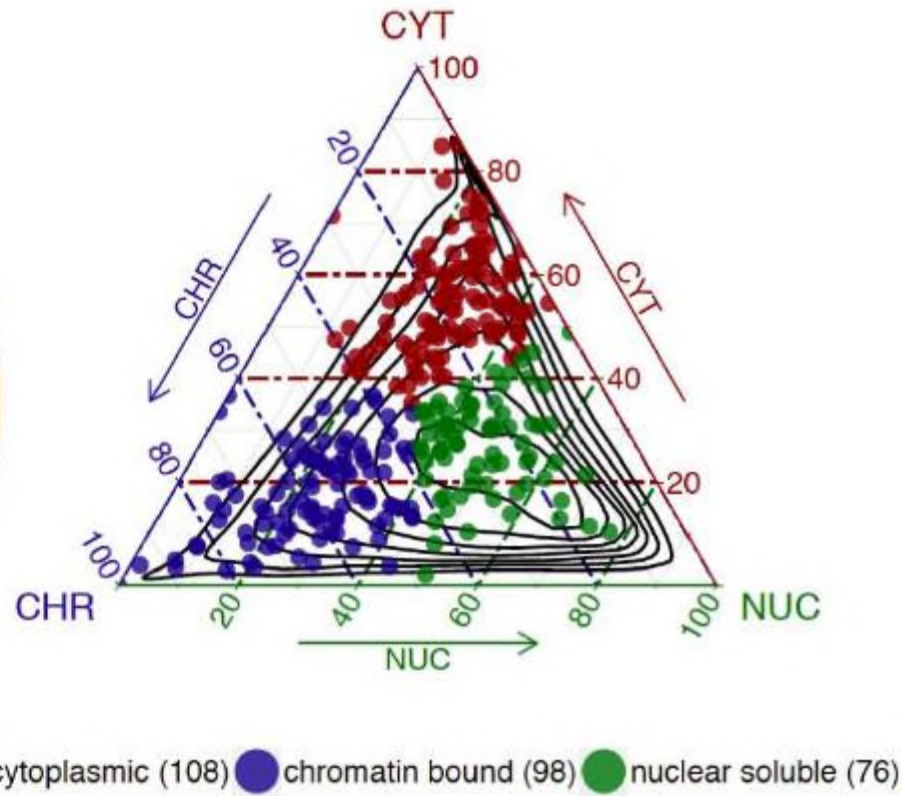
Cell-type-specific lncRNAs implicated in GWAS traits



Hon *et al.*, *Nature* 543, 199 (2017)

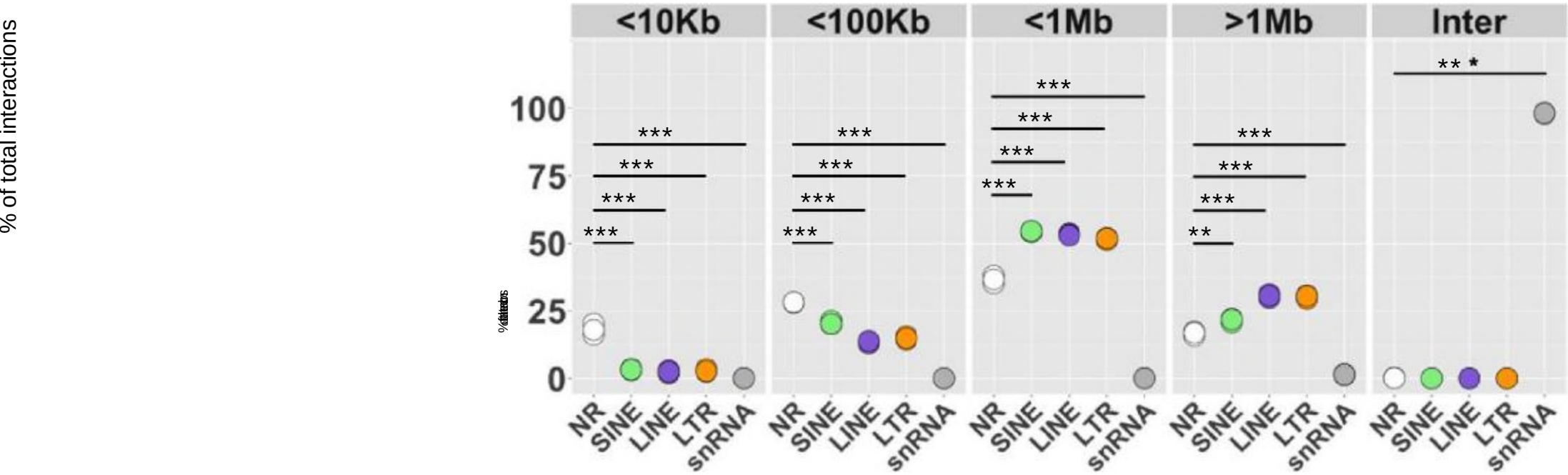
Considerations: localization of lncRNAs
And functional interactions

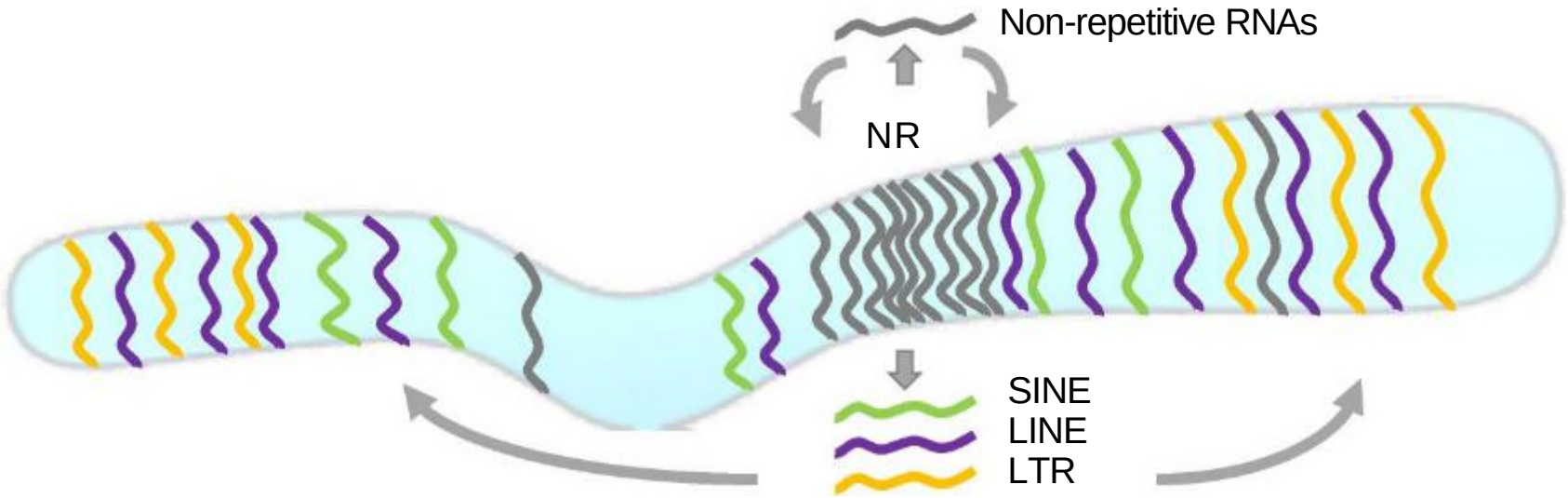




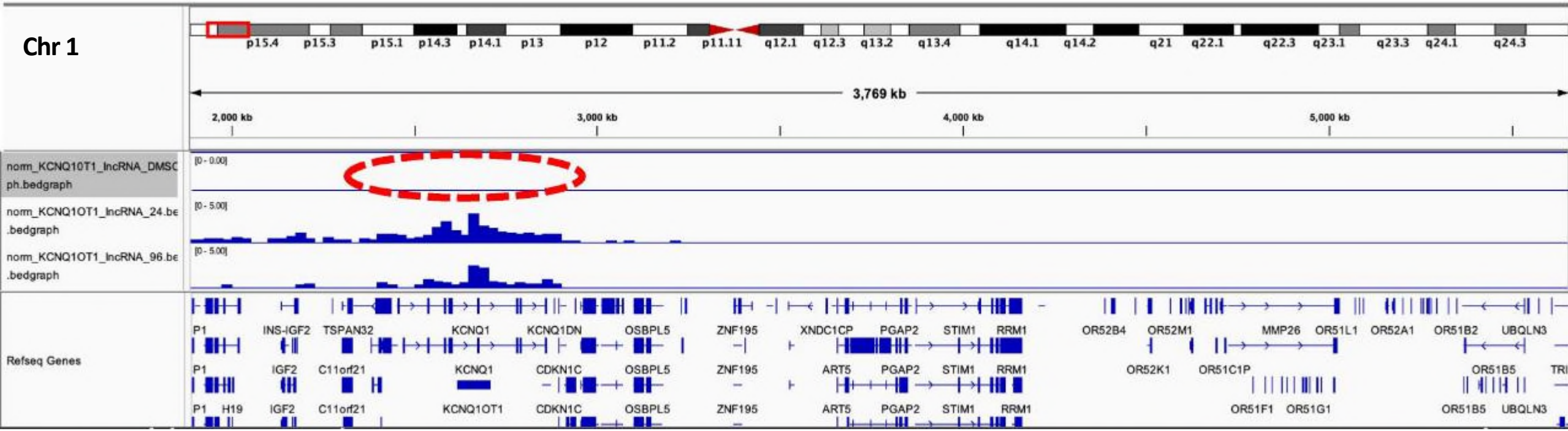
- 64% in the nucleus - 37% in chromatin-bound
  - 27% in the nuclear soluble fraction
- 36% in the cytoplasm

Repeat elements are more interacting at longer distance

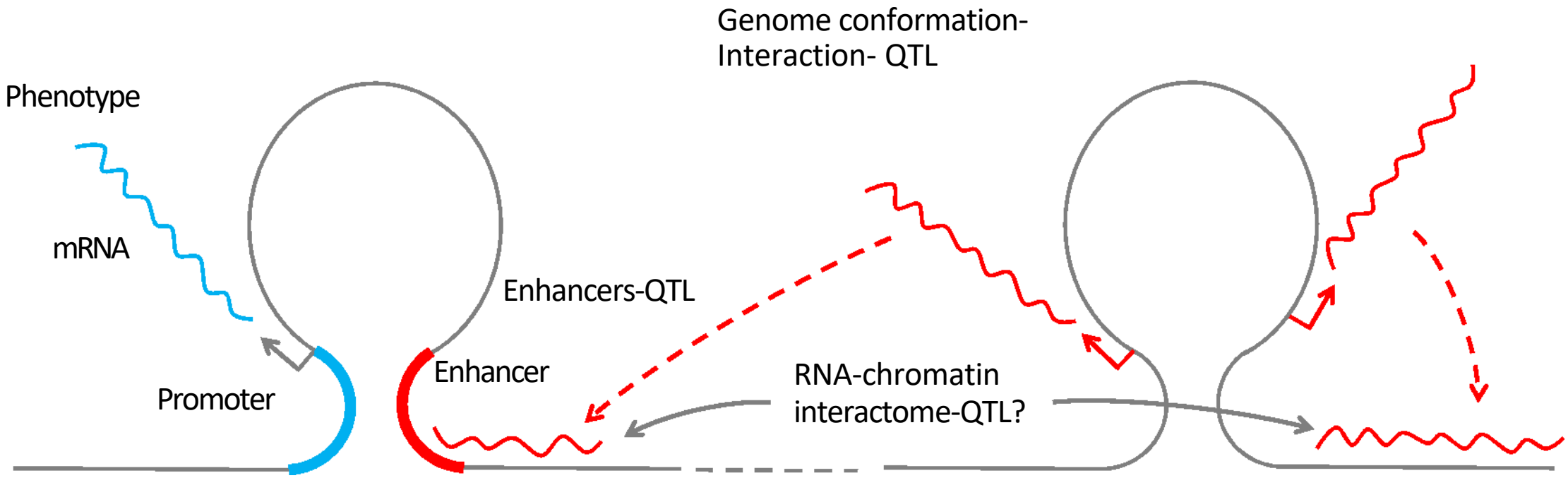




Different patterns of lncRNA interactions during differentiation Kcnq10t1 lncRNA

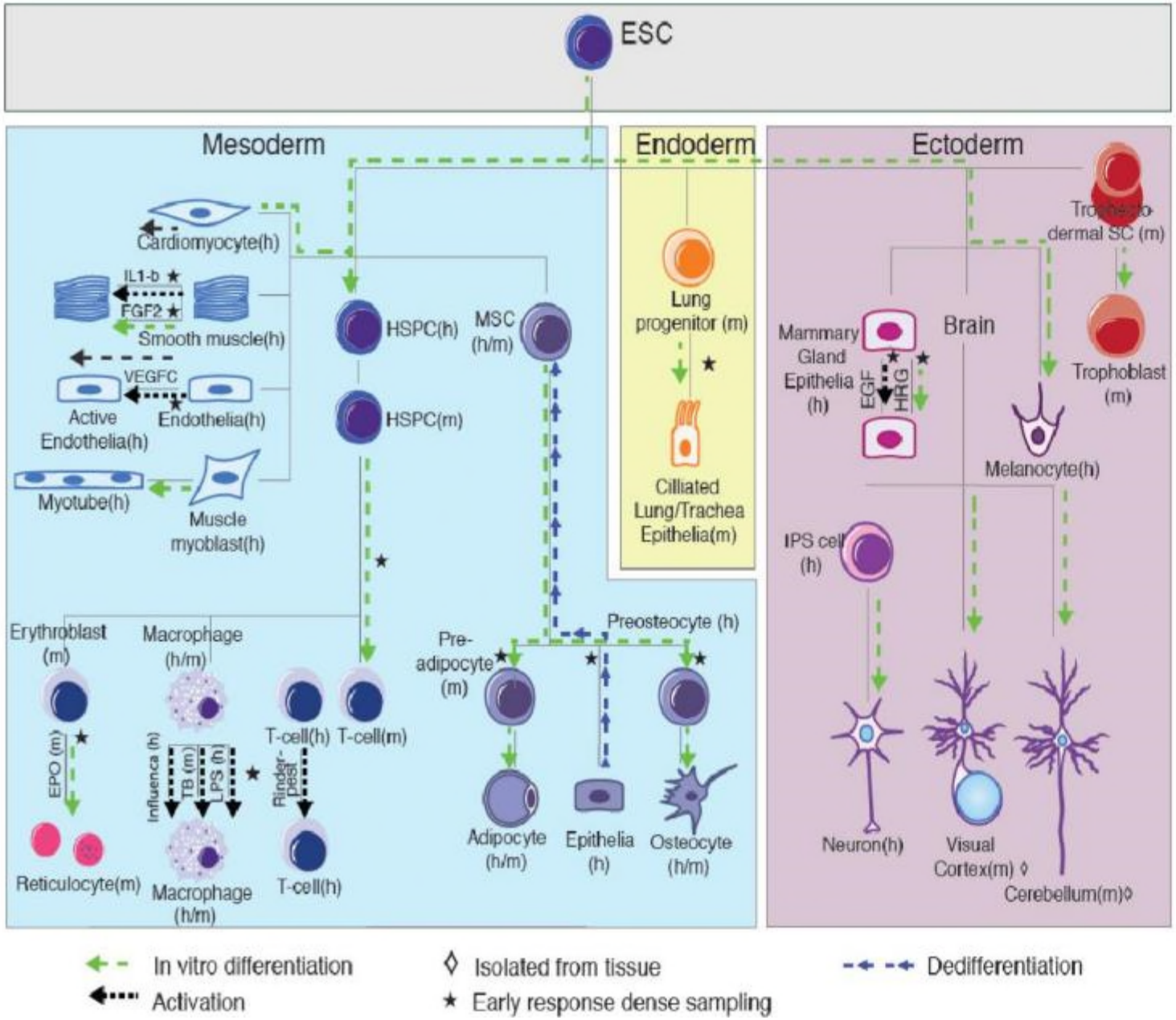


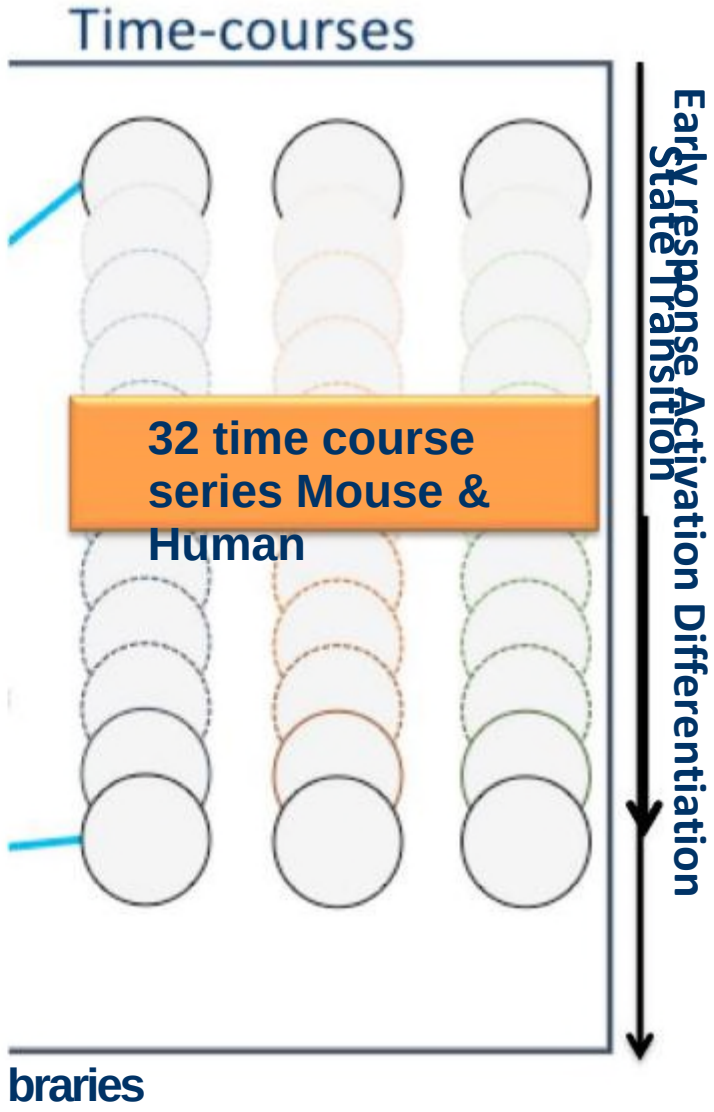
A detailed cell specific map of chromatin: DNA-DNA, RNA-DNA (&proteins) interactions



Broad CAGE map across thousands cells

FANTOM5

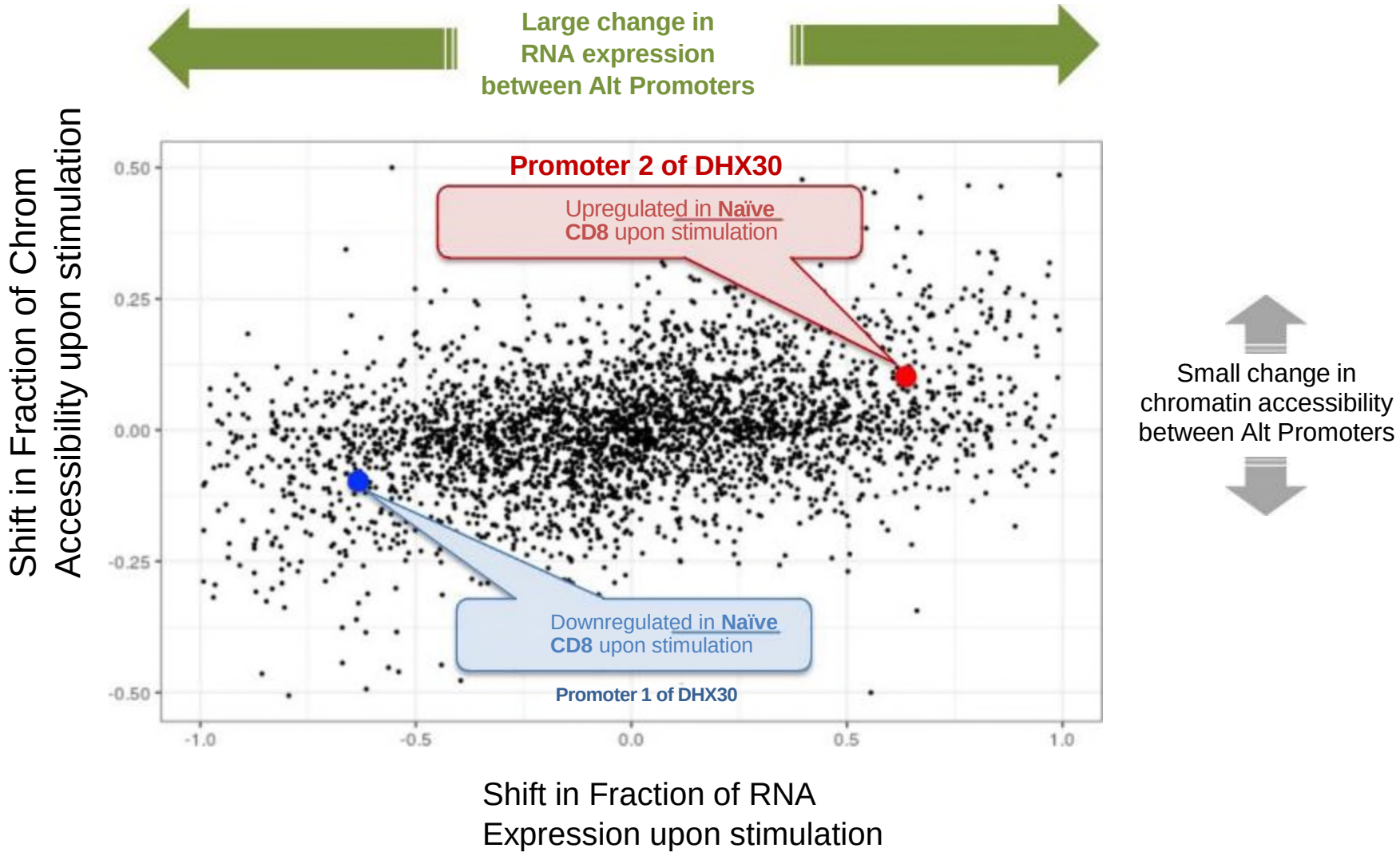




Arner *et al.*, *Science* 347, 1010 (2015)

RNA Expression vs Chrom Accessibility : Alternative promoter usage

123 Genes show significant (fdr <0.05) alternative promoter usage in at least one cell type in PBMC upon stimulation

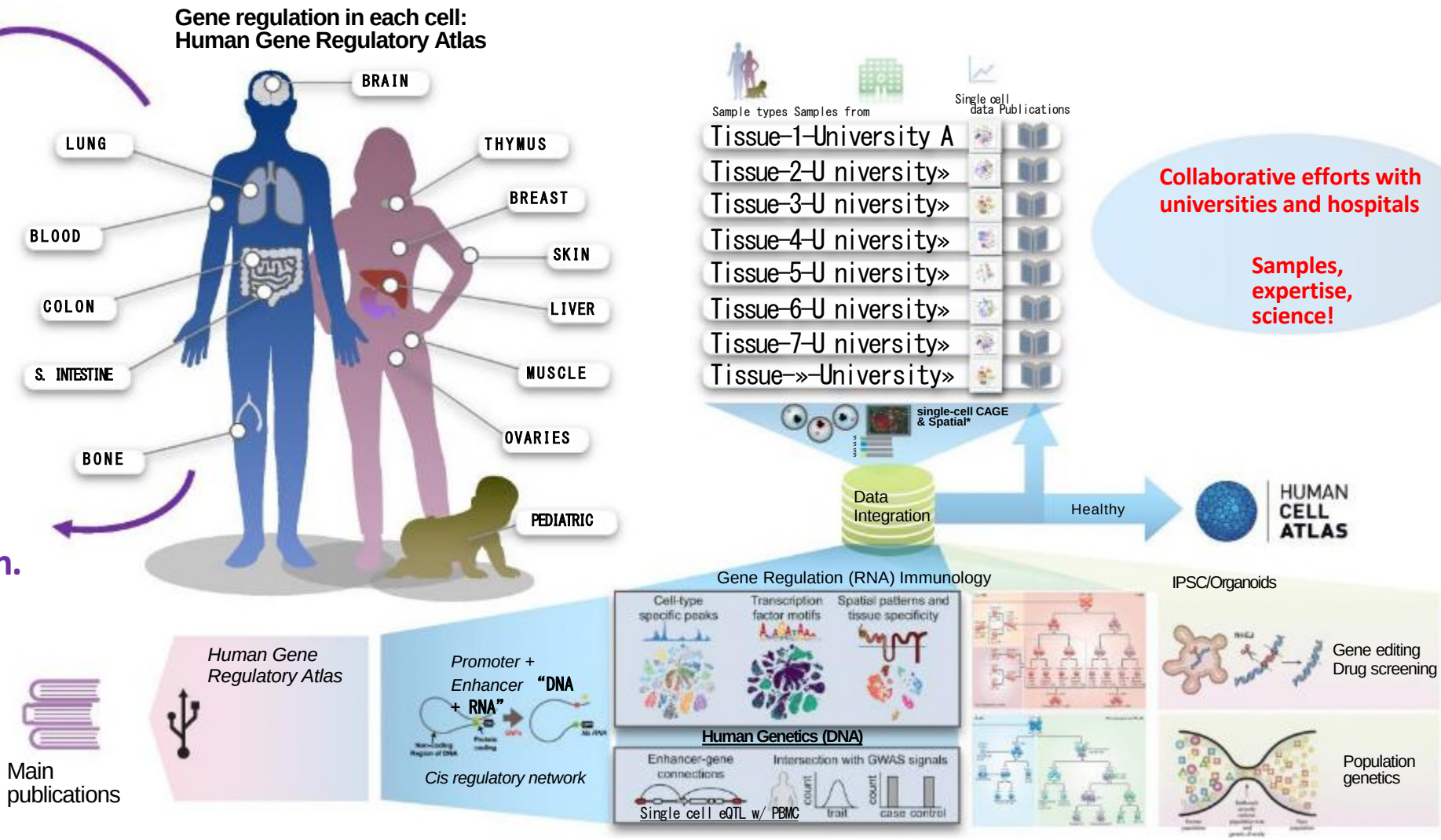


Slide courtesy of Chung Chau Hon and Jonathan Moody

Bringing genomics at single cell level; collaborations with hospitals & universities

Analyze each tissue at single cell resolution.

Analyze gene expression and regulation in health and diseases across genetic variation.



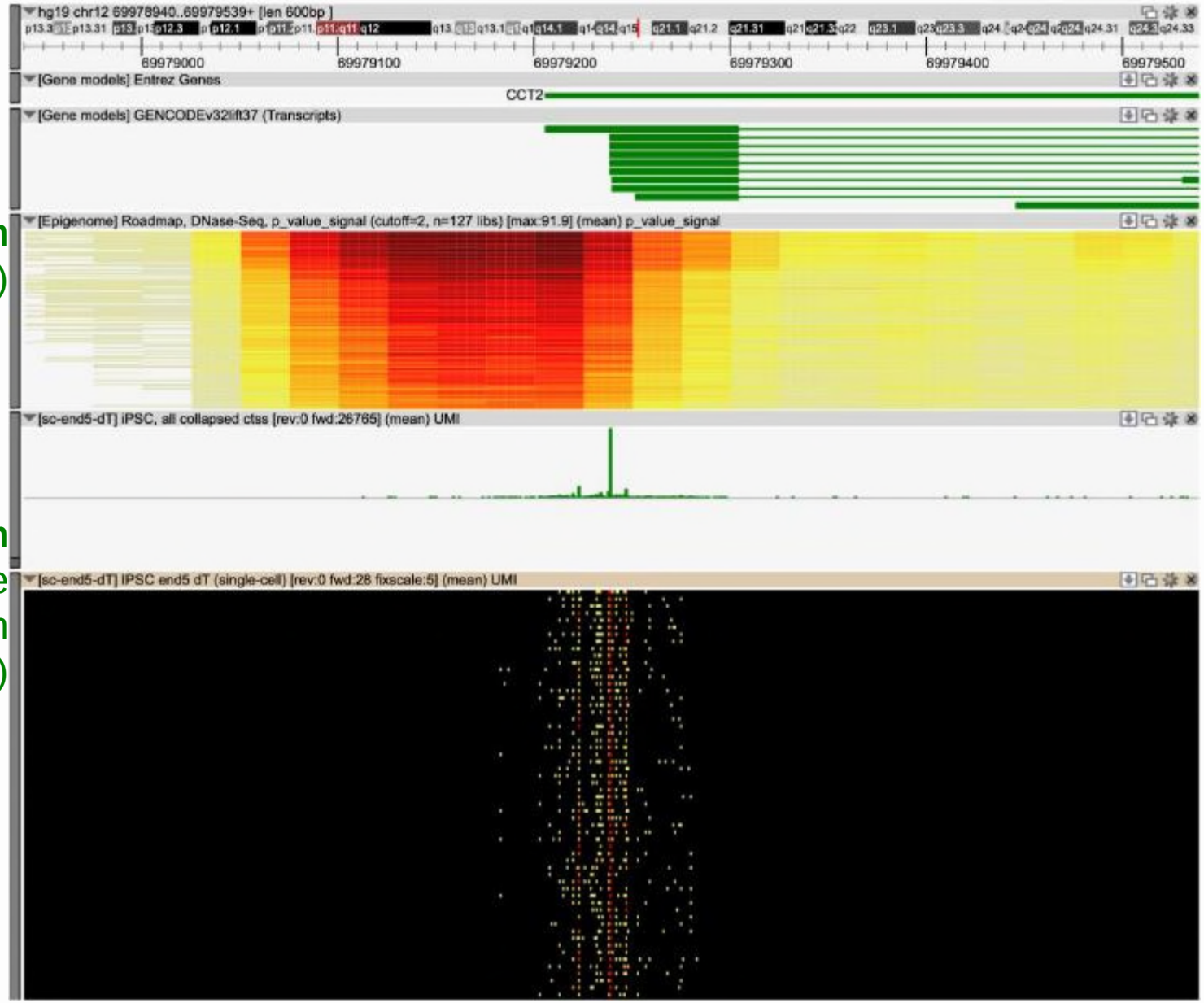
Slide courtesy of Jay W. Shin, RIKEN

At Gene Promoters : Detecting TSS in single cells at nt resolution

DNase-seq
(bulk)

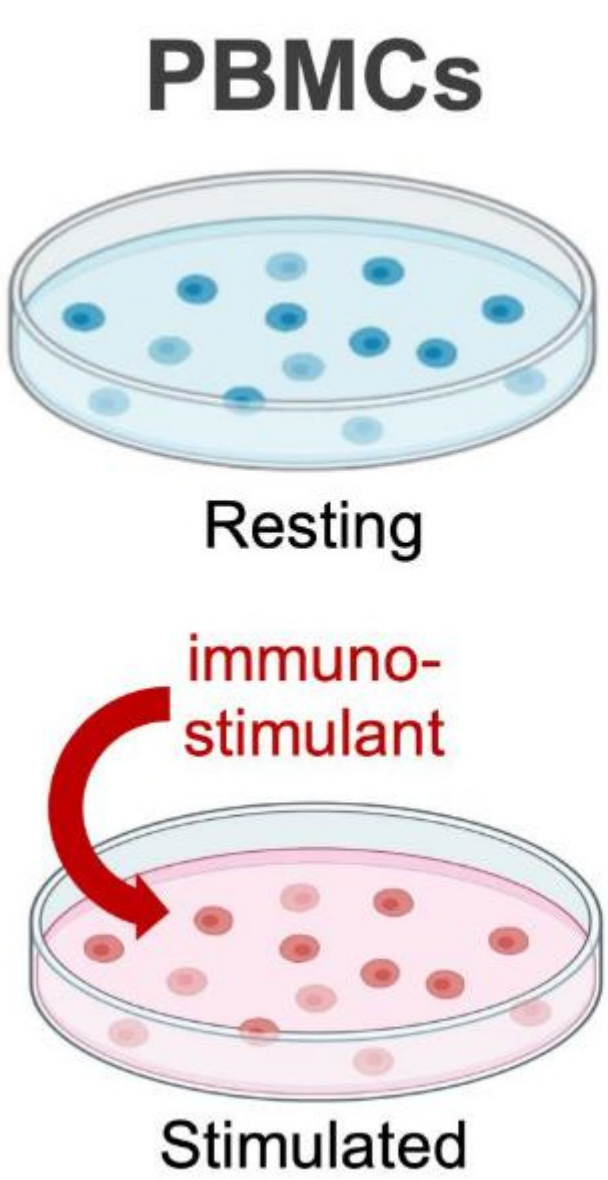
RNA Expression
(pseudo-bulk)

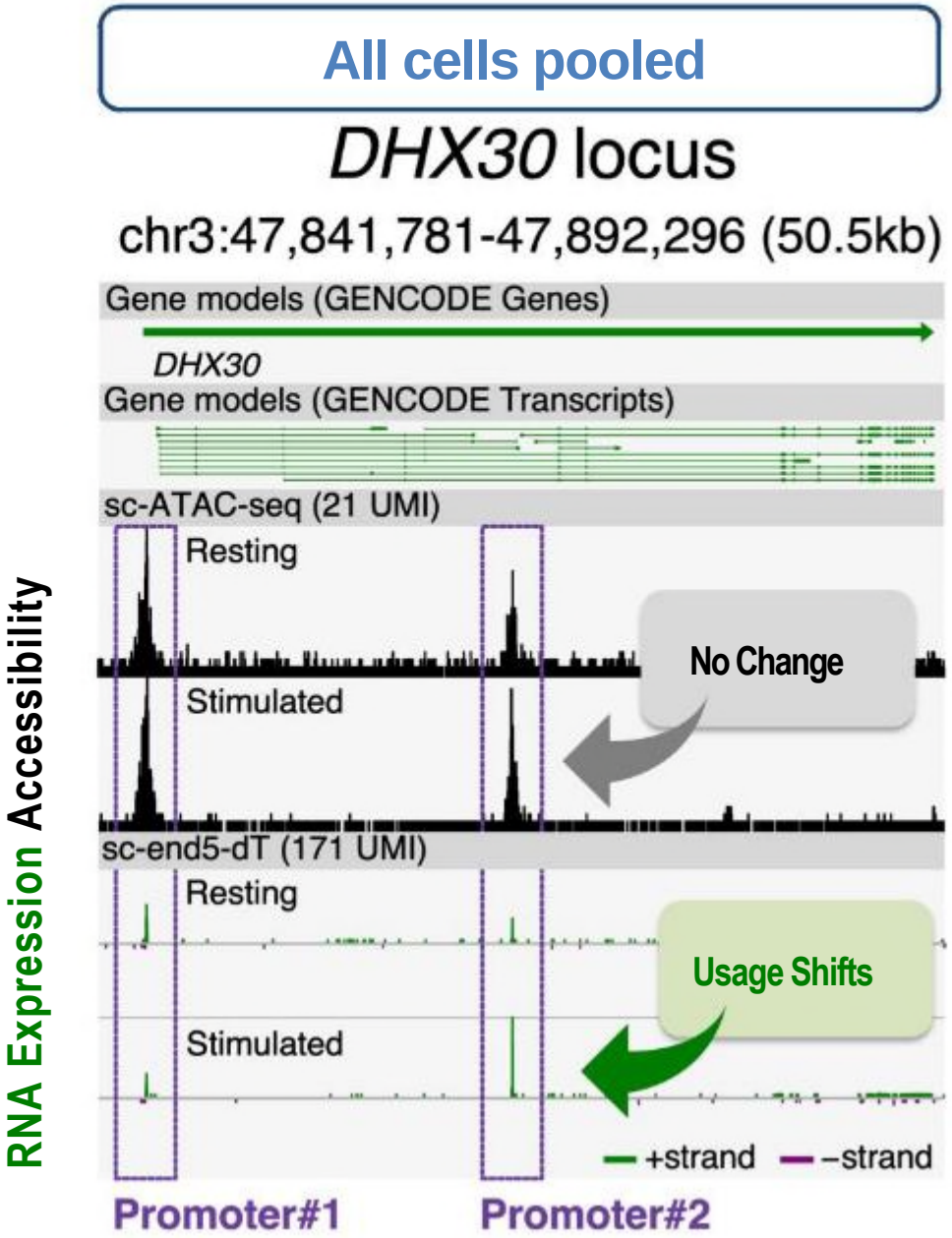
RNA Expression
(single nucleotide in single cells)



Slide courtesy of Chung Chau Hon

At Gene Promoters : Shifts in Alternative Promoter Usage in Single Cells

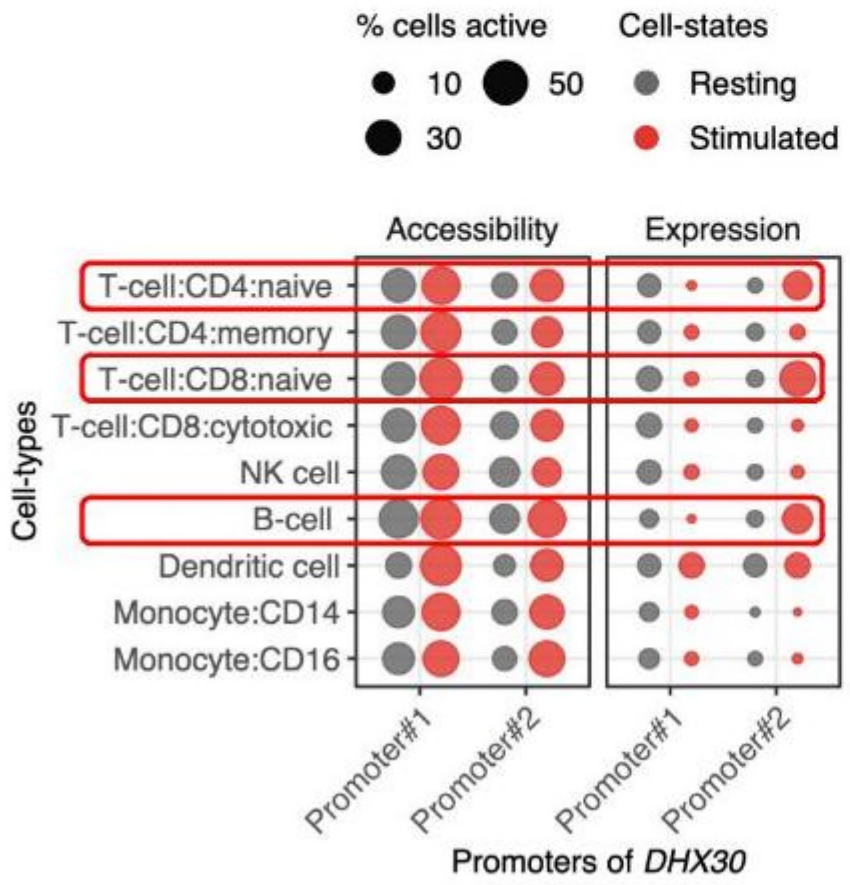




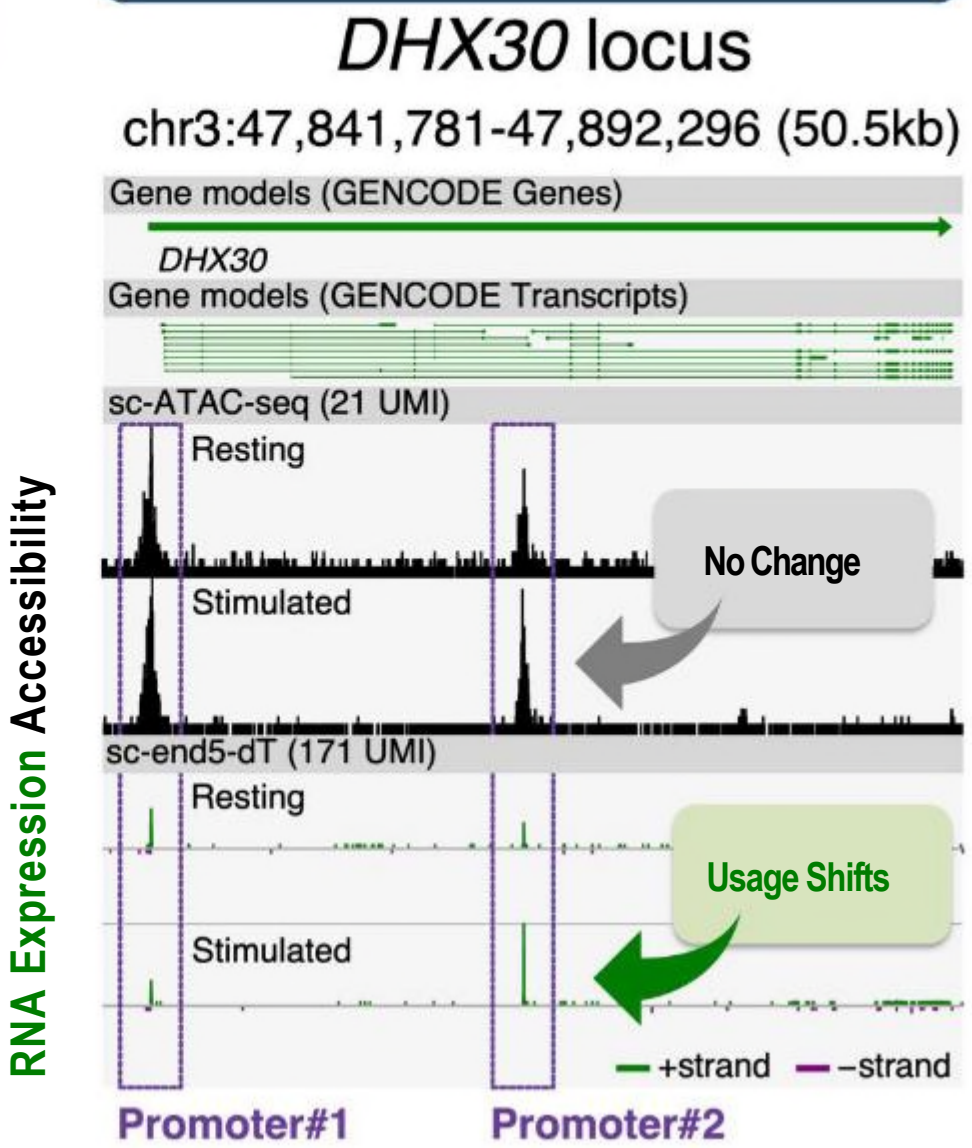
Slide courtesy of Chung Chau Hon

At Gene Promoters : Shifts in Alternative Promoter Usage in Single Cells

Cell-type specific shifts in alternative promoter usage



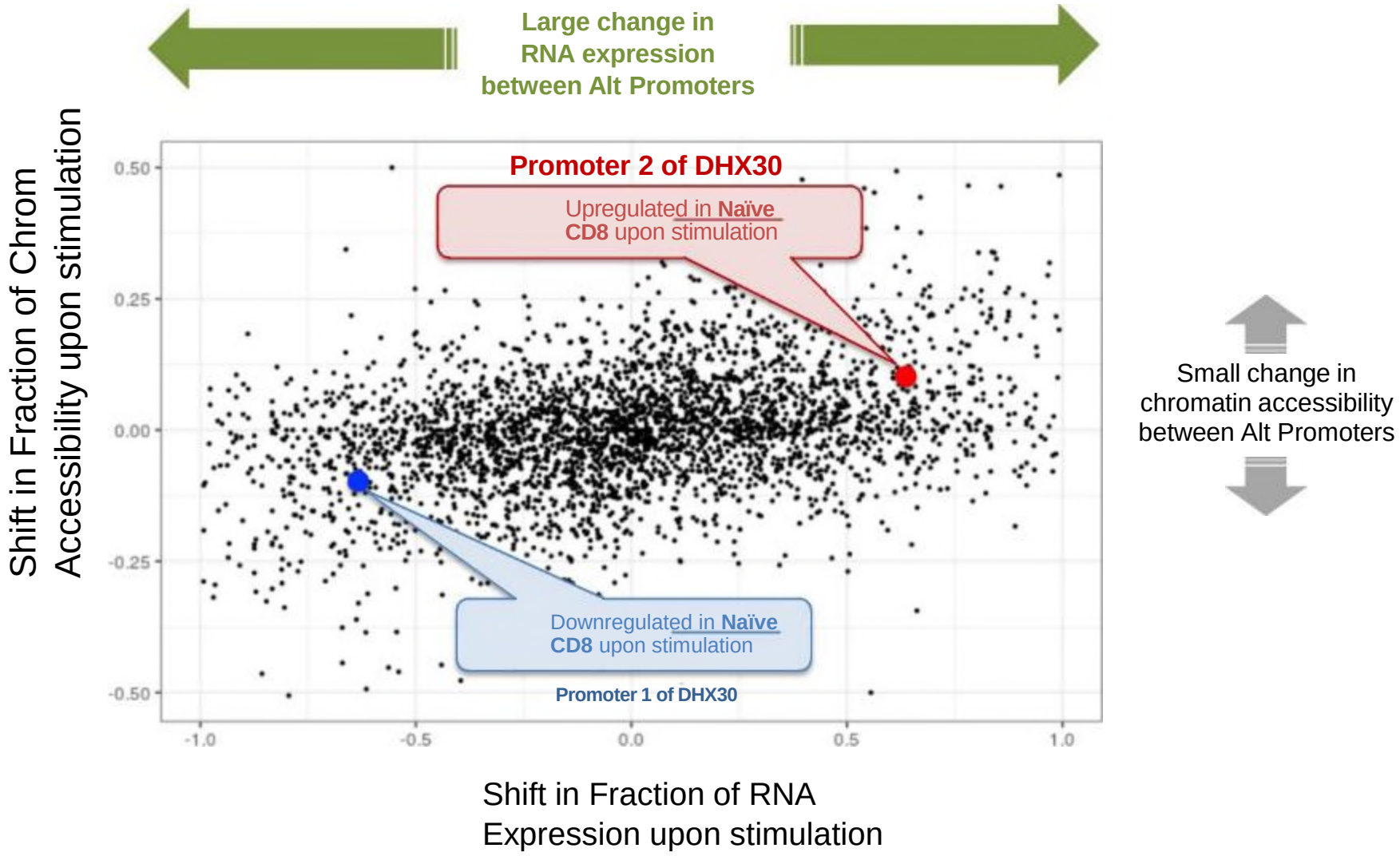
All cells pooled



Slide courtesy of Chung Chau Hon

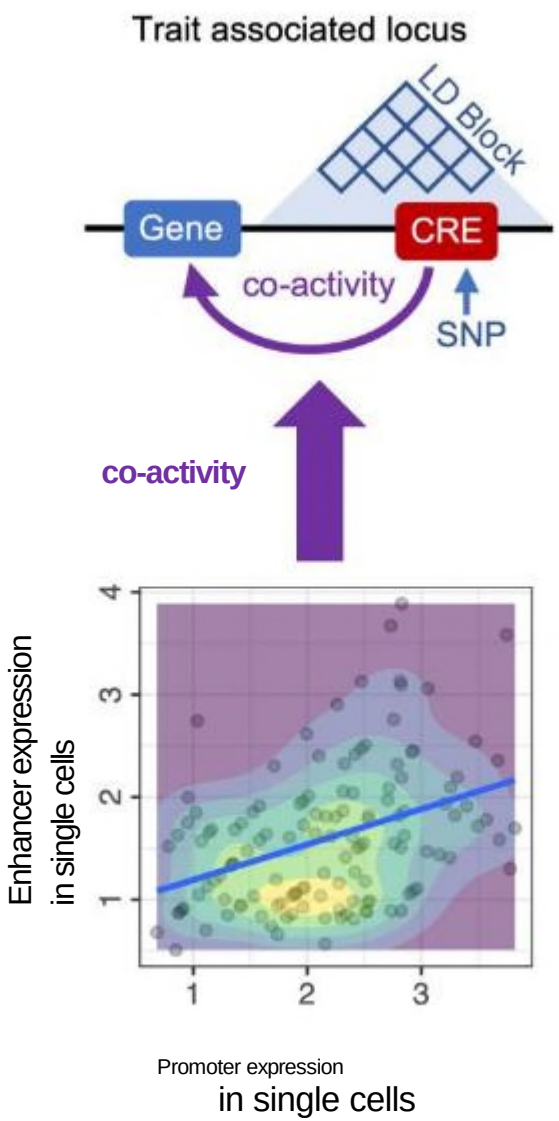
RNA Expression vs Chrom Accessibility : Alternative promoter usage

123 Genes show significant (fdr <0.05) alternative promoter usage in at least one cell type in PBMC upon stimulation

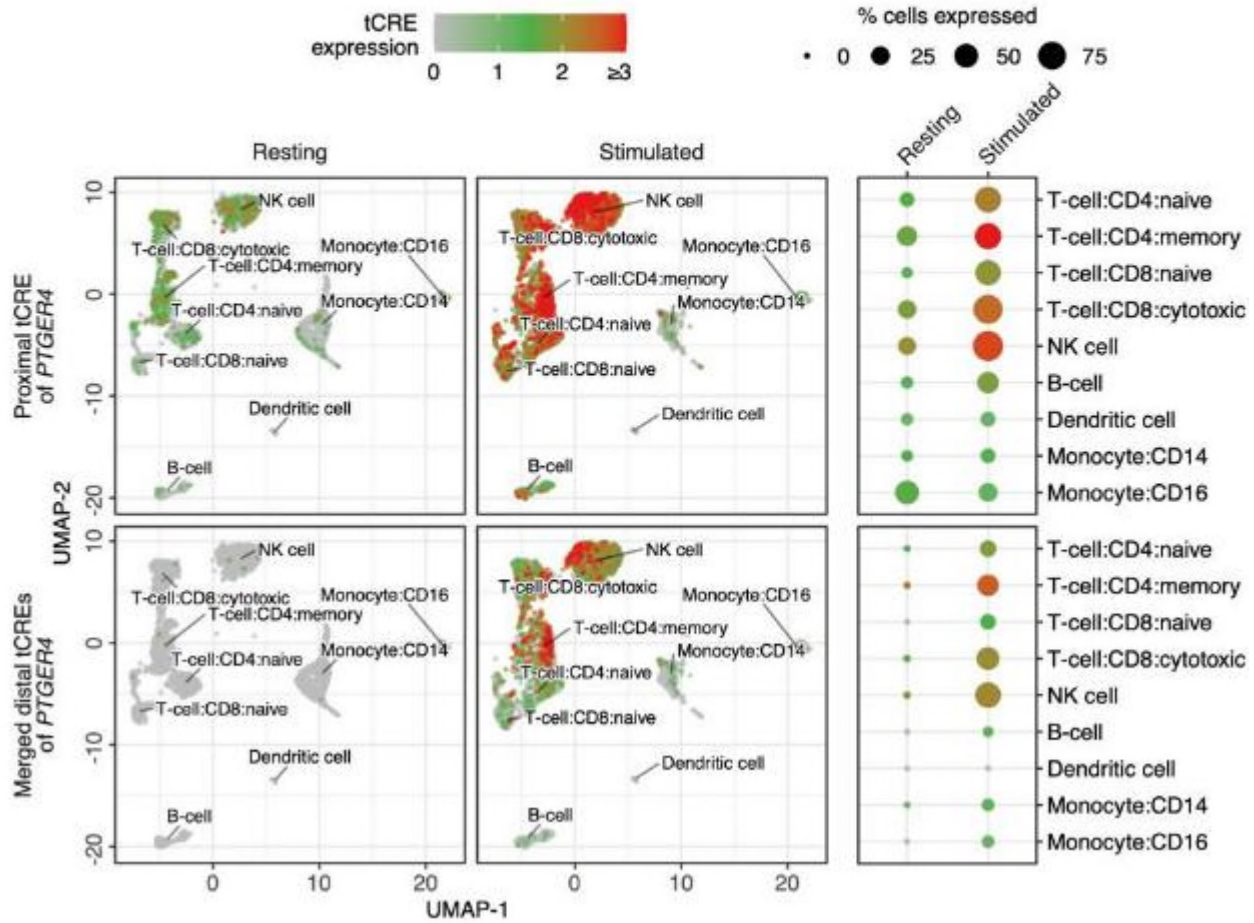


Slide courtesy of Chung Chau Hon and Jonathan Moody

GWAS interpretation : Linking GWAS variants to candidate genes



Cell-type specific co-activation of promoters and enhancers



Slide courtesy of Chung Chau Hon

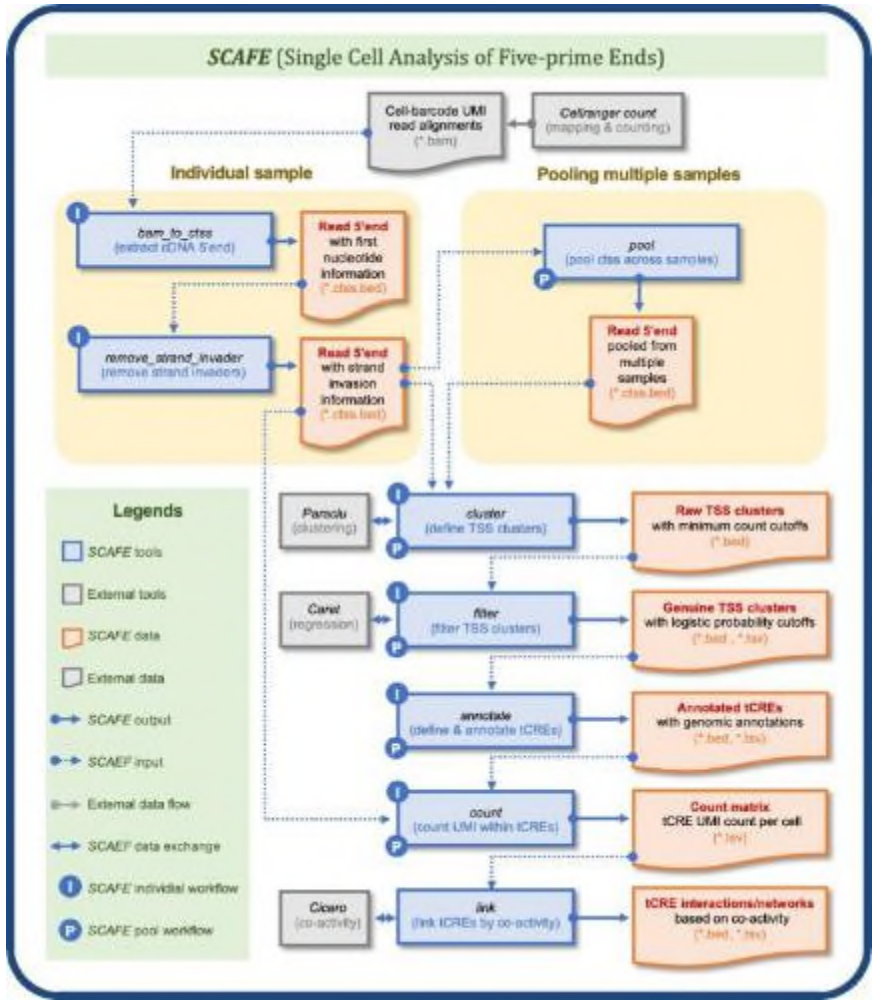
Analyzing tCREs : Gene expression + enhancer activity in one assay

SCAFE : **Software to define, quantify and link tCREs from 10x 5'data**

Make the most of your 10x 5'data.....for FREE!

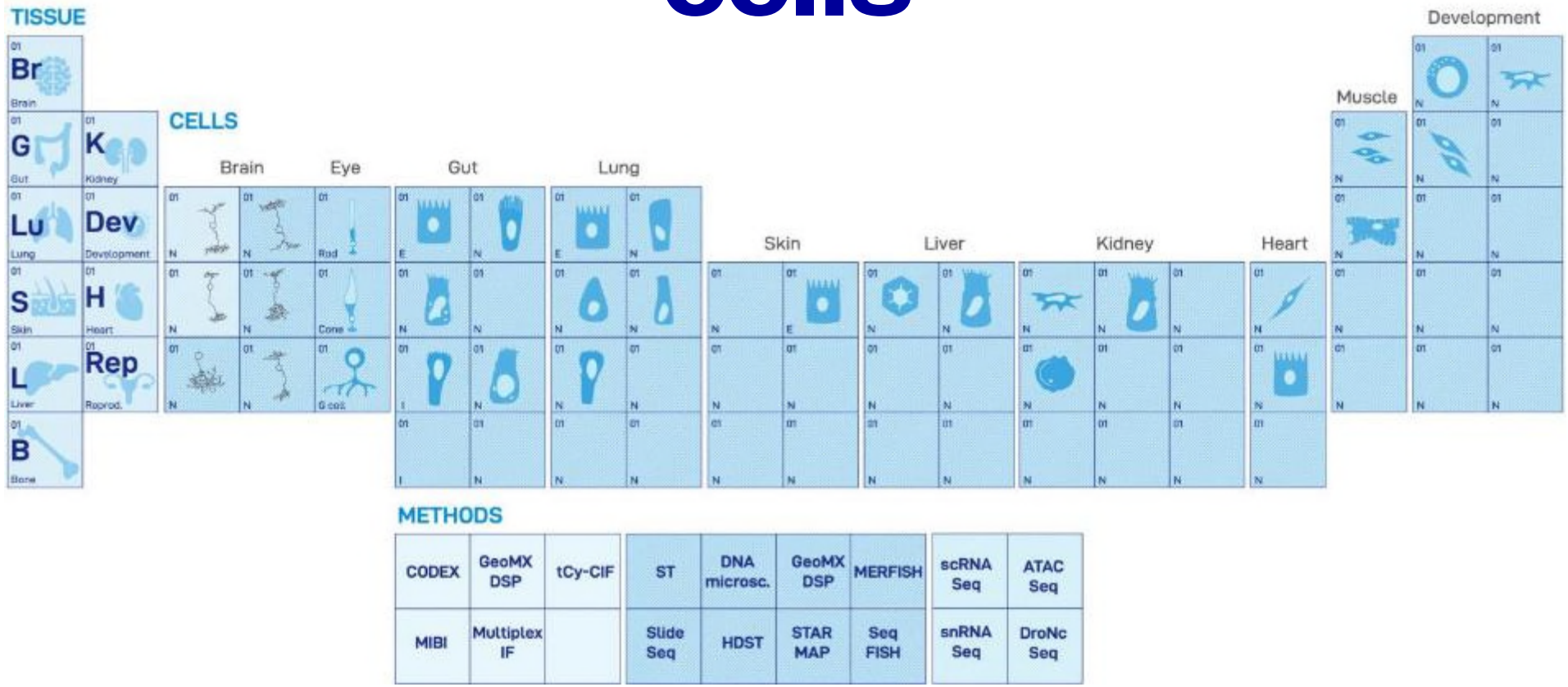
https://github.com/chung-lab/scafe

End-to-end solution for 10x 5'data



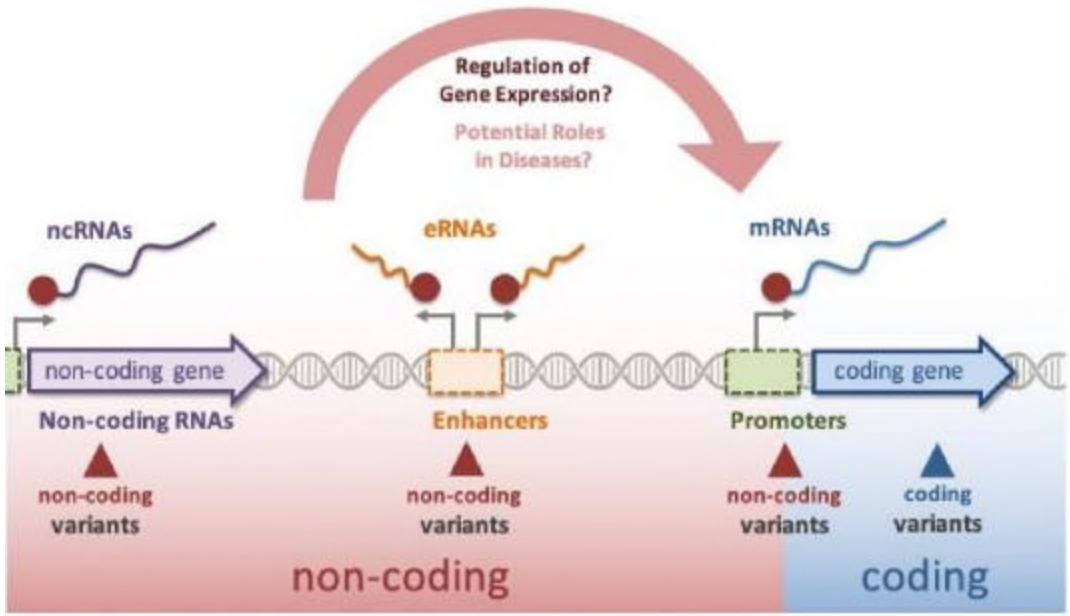
Slide courtesy of Chung Chau Hon

A periodic table of our cells

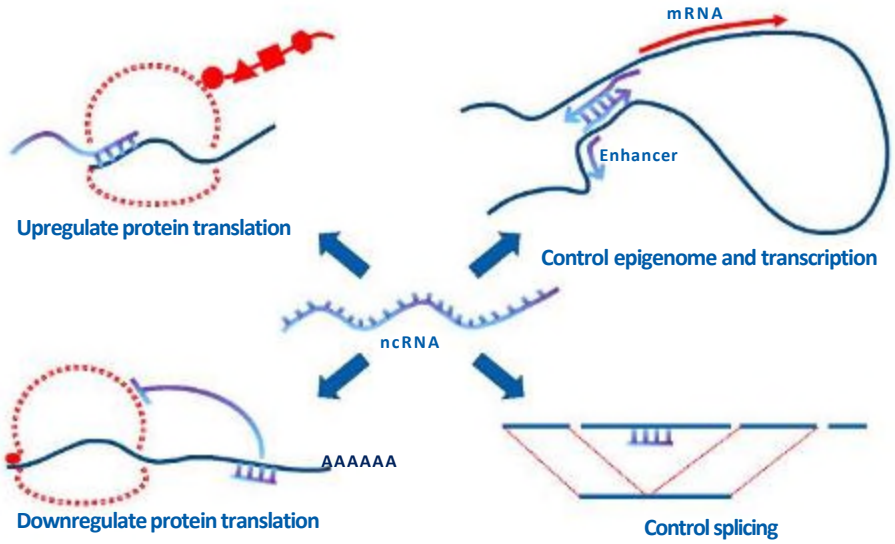


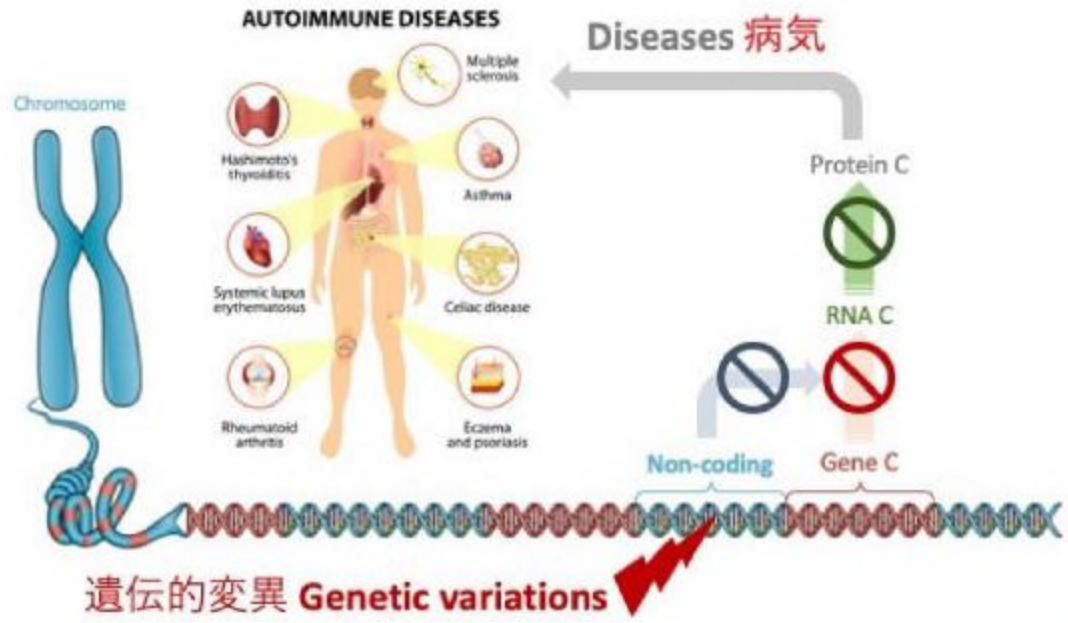
Graphics credit: Ania Hupalowska

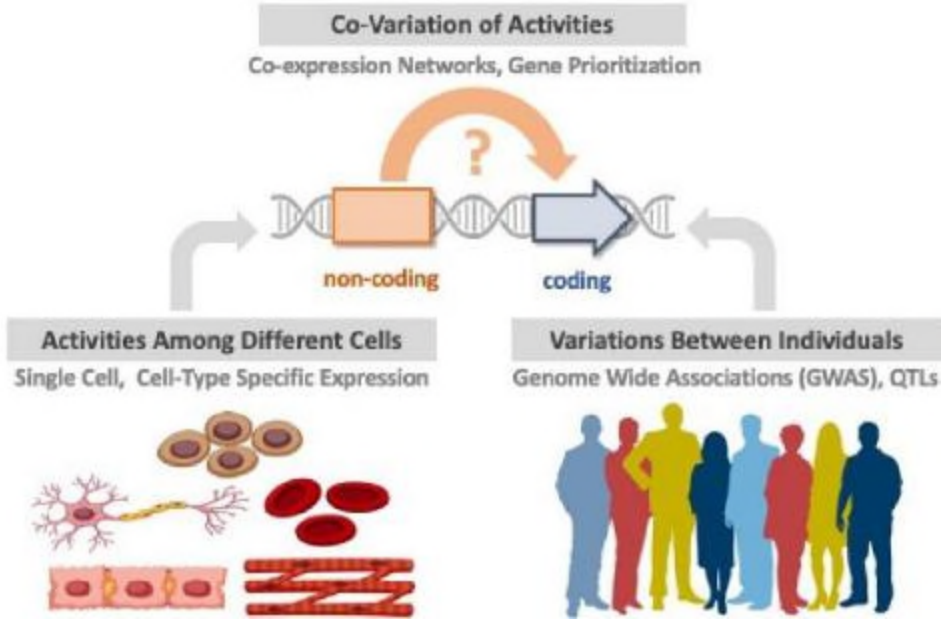
“Non-coding regions of genome”



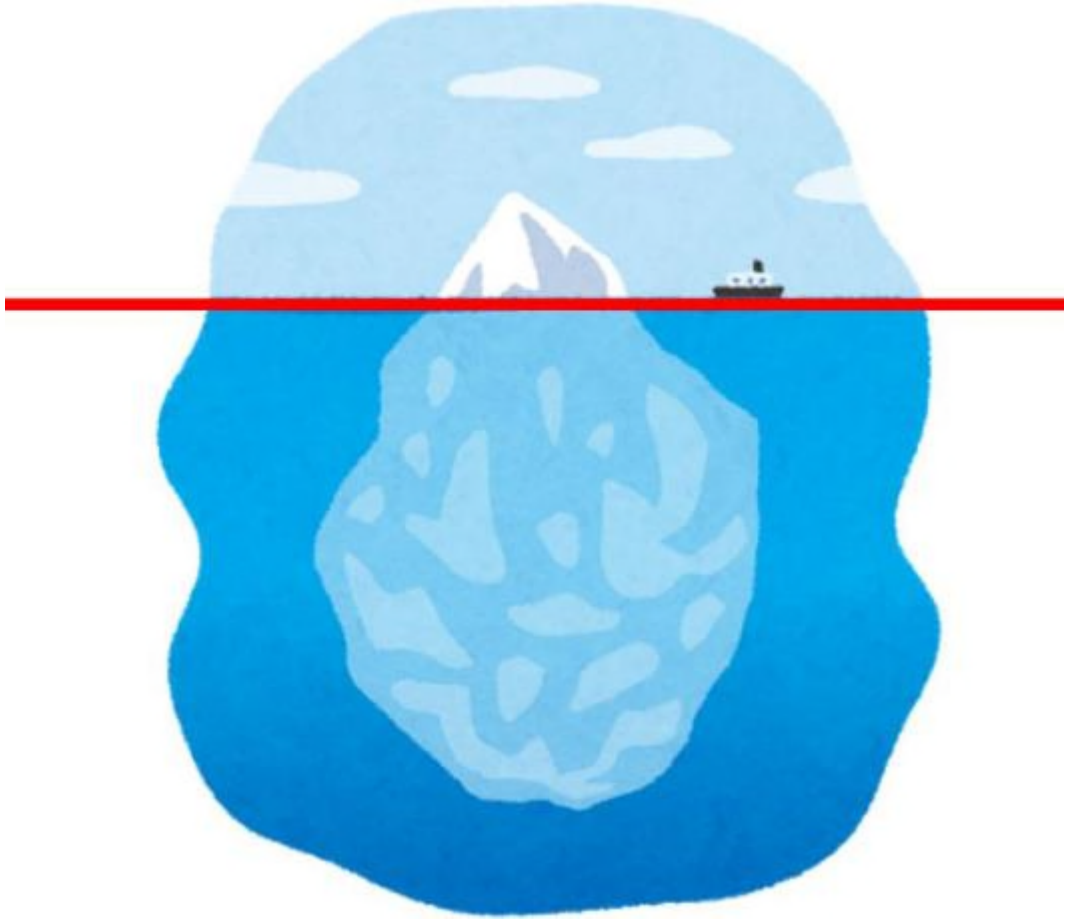
New therapeutics with non-coding RNA







SINEUPs: Is this a family of antisense RNA?



?