

Workshop on Comparative Genomics, Cesky Krumlov, 2011

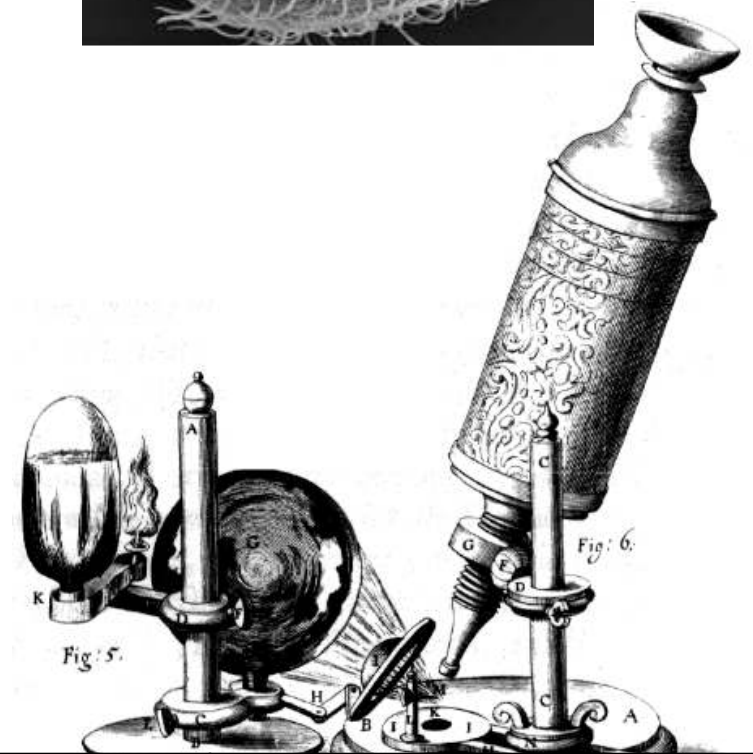
Transcriptome characterization with RNASeq: Applications

Roderic Guigó, roderic.guigo@crg.cat
Center for Genomic Regulation, Barcelona



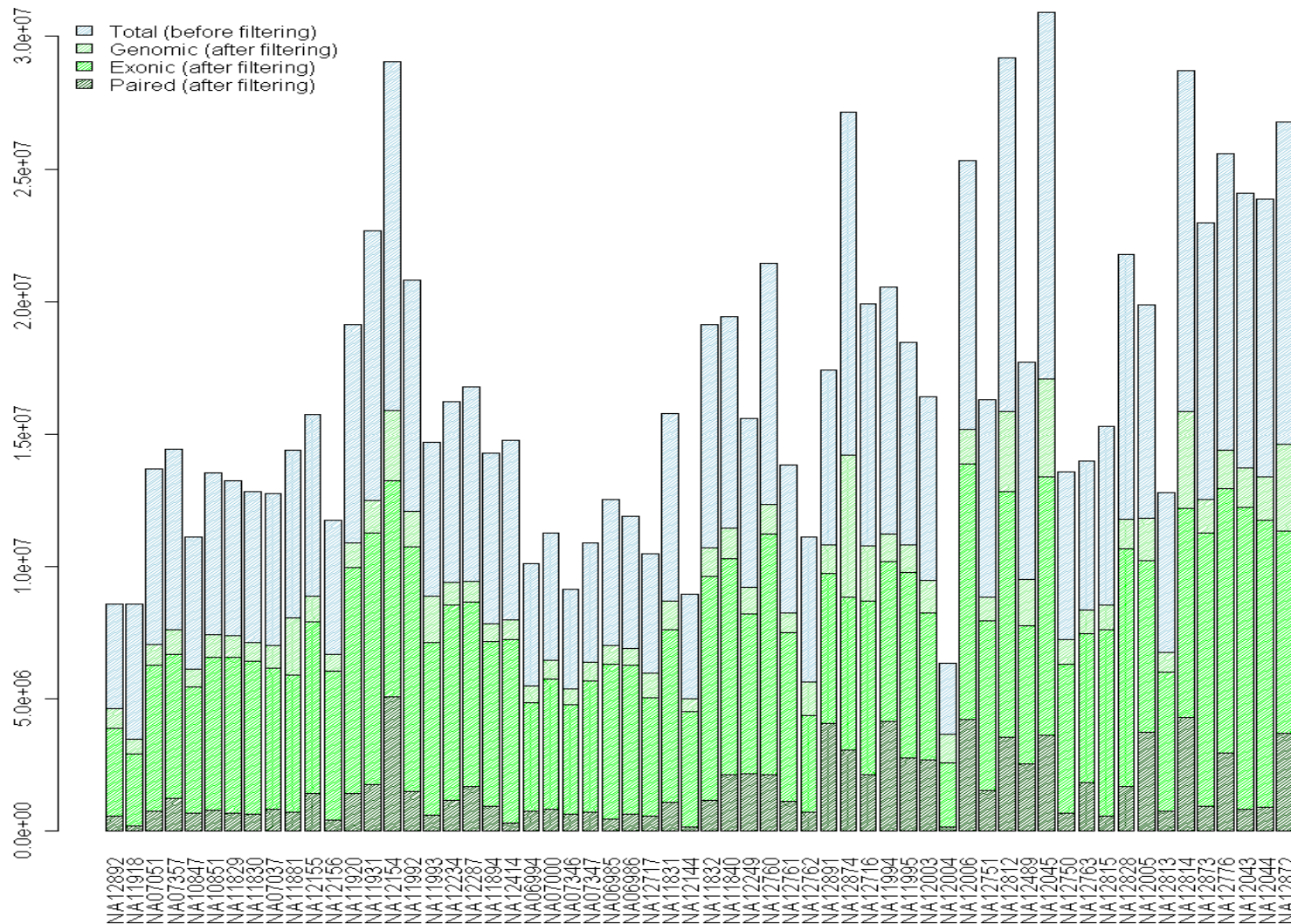
Van Leeuwenhoek

In 1676 his credibility was questioned when he sent the Royal Society a copy of his first observations of microscopic single celled organisms. Heretofore, the existence of single celled organisms was entirely unknown ... The Royal Society arranged to send an English vicar, as well as a team of respected jurists and doctors to Delft, Holland to determine whether it was in fact Van Leeuwenhoek's ability to observe and reason clearly (wikipedia)



1. e-QTL ANALYSIS OF TRANSCRIPT ISOFORMS

Sequence of the transcriptome of 60 CEU individuals for the HapMap project.

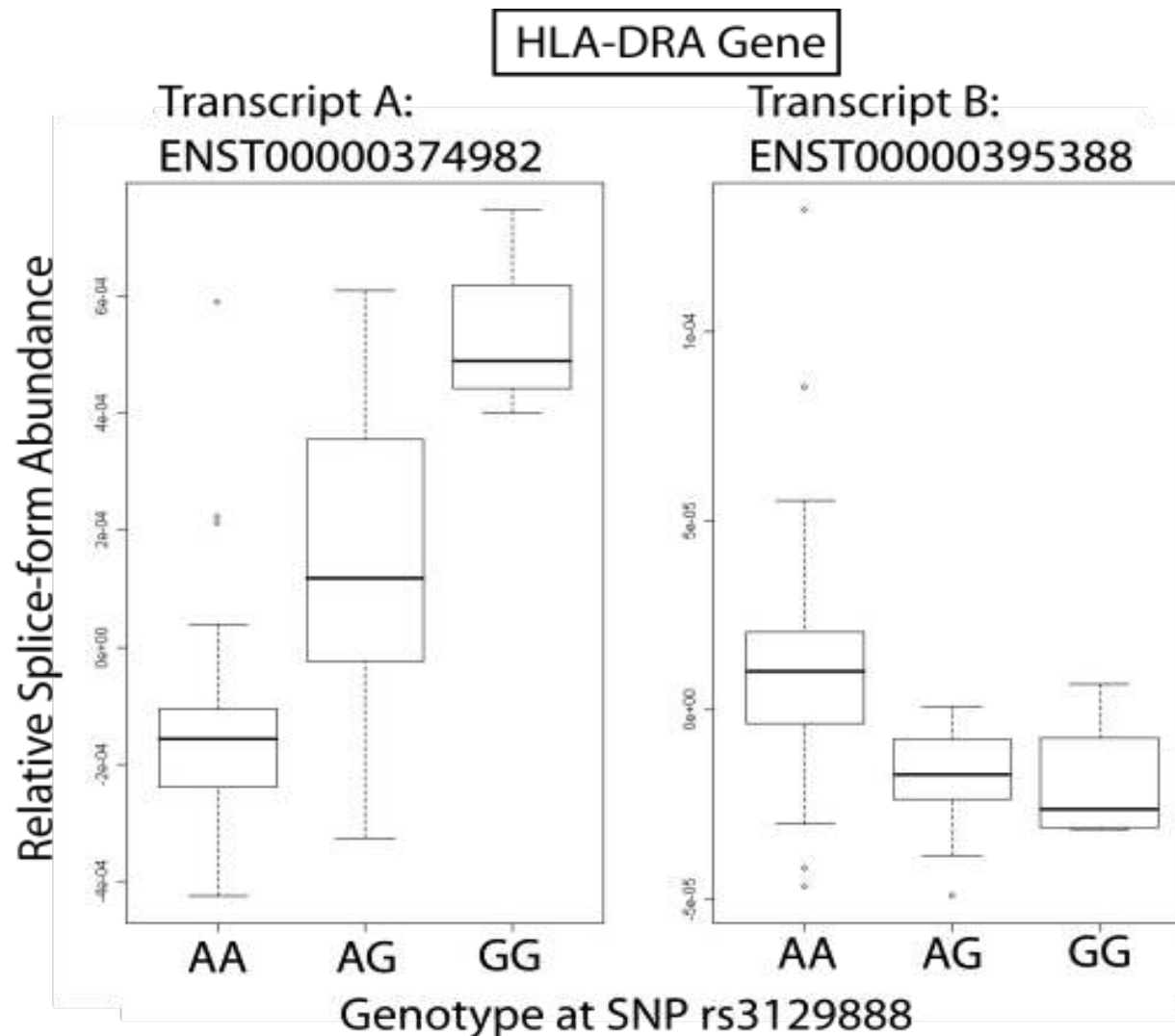


1 Transcriptome genetics using second generation sequencing in a Caucasian population

Stephen B. Montgomery^{1,2*}, Micha Sammeth³, Maria Gutierrez-Arcelus¹, Radoslaw P. Lach², Catherine Ingle², James Nisbett², Roderic Guigo³ & Emmanouil T. Dermitzakis^{1,2*}

- ~10 million reads provide access to the same dynamic range as arrays with better quantification of alternative and highly abundant transcripts.
- Correlation with SNPs leads to a larger discovery of eQTLs than with arrays.
 - eQTLs are discovered with similar frequencies in protein coding and long non coding RNAs
- a substantial number of variants influence the structure of mature transcripts indicating variants responsible for alternative splicing.
- Measures of allele- specific expression allow for the identification of rare eQTLs and allelic differences in transcript structure

Significant eQTL with opposite effect in two different isoforms of the gen HLA-DRA



2. LONG NON CODING RNAs

Long Noncoding RNAs with Enhancer-like Function in Human Cells

Ulf Andersson Ørom,¹ Thomas Derrien,² Malte Beringer,¹ Kiranmai Gumireddy,¹ Alessandro Gardini,¹ Giovanni Bussotti,² Fan Lai,¹ Matthias Zytnicki,² Cedric Notredame,² Qihong Huang,¹ Roderic Guigo,² and Ramin Shiekhattar^{1,2,3,*}

¹The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

²Centre for Genomic Regulation (CRG), UPF, Barcelona, Spain

³Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

*Correspondence: shiekhattar@wistar.org

DOI 10.1016/j.cell.2010.09.001

Nature **458**, 223-227 (12 March 2009) | doi:10.1038/nature07672; Received November 2008; Published online 1 February 2009

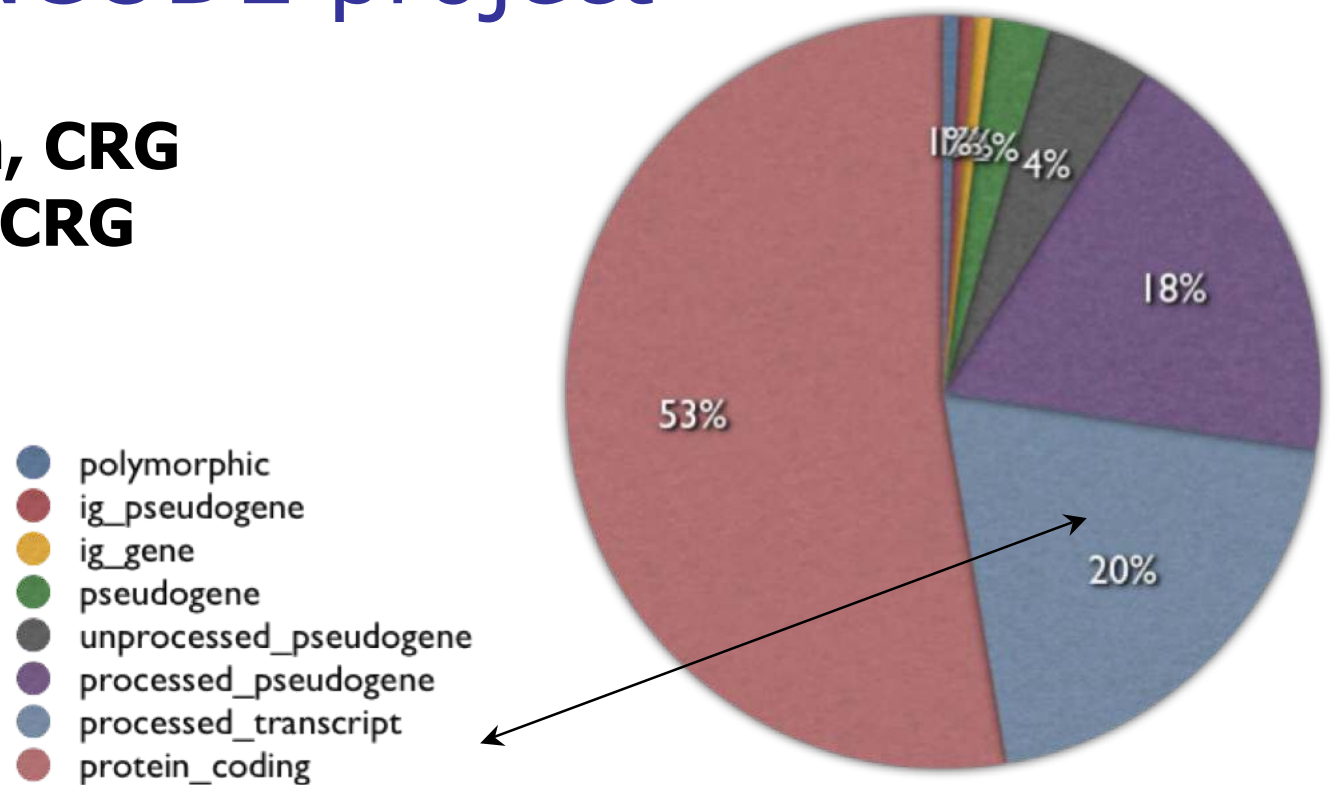
Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals

Mitchell Guttman^{1,2}, Ido Amit¹, Manuel Garber¹, Courtney French¹, Michael F. Lin¹, David Feldser³, Maite Huarte^{1,6}, Or Zuk¹, Bryce W. Carey^{2,8}, John P. Cassady^{2,8}, Moran N. Cabili⁷, Rudolf Jaenisch^{2,8}, Tarjei S. Mikkelsen^{1,4}, Tyler Jacks^{2,3}, Nir Hacohen^{1,9}, Bradley E. Bernstein^{1,10,11}, Manolis Kellis^{1,5}, Aviv Regev^{1,2}, John L. Rinn^{1,6,11,12} & Eric S. Lander^{1,2,7,8,12}

GENCODE annotation.

ENCODE project

Thomas Derrien, CRG
Andrea Tanzer, CRG



- ▶ Processed_transcript: region of the genome which show evidence of expression (cDNA, ESTs...) but without ORFs => good candidates

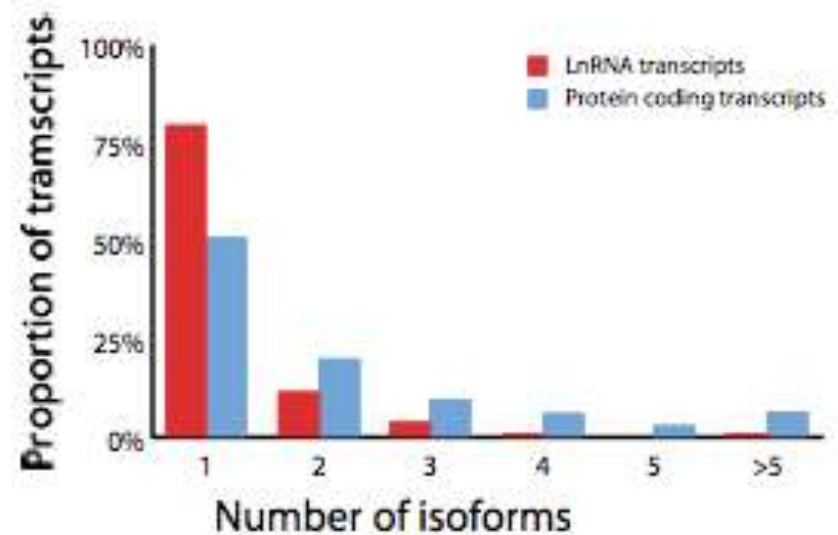
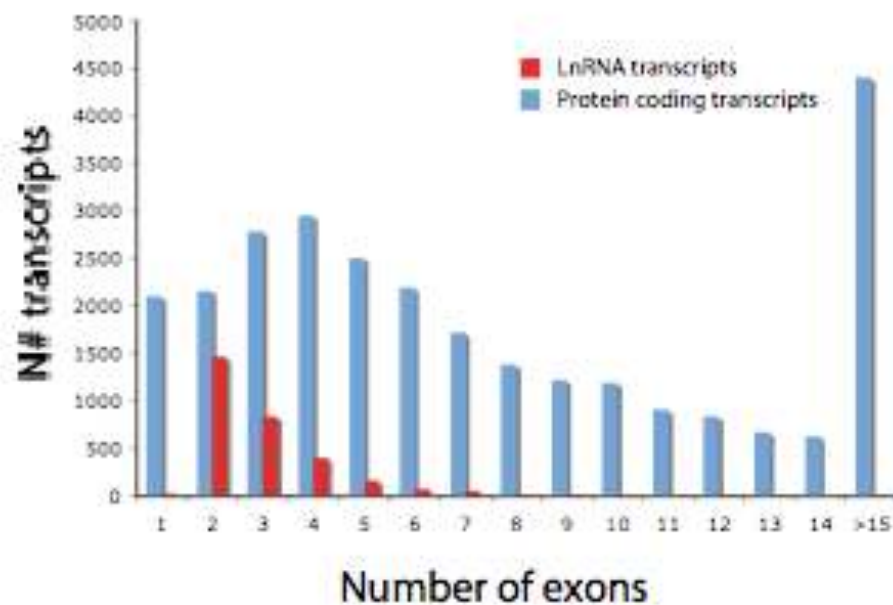
Long non coding RNAs in GENCODE

	Protein Coding	processed pseudogenes	processed transcript		"LnRNAs" filtered (1)
loci	12 422	4 344	4681		2258
transcripts	27 639	4 353	23 392		3019
exons	238888	5 143	99759		8978

Filtering:

- outside pcg loci (Havana, RefSeq, UCSC genes)
- outside 5', 3' UTRs (1kb)
- outside known ncRNAs (miRNAs, snoRNAs)

lncRNAs are spliced and alternative spliced

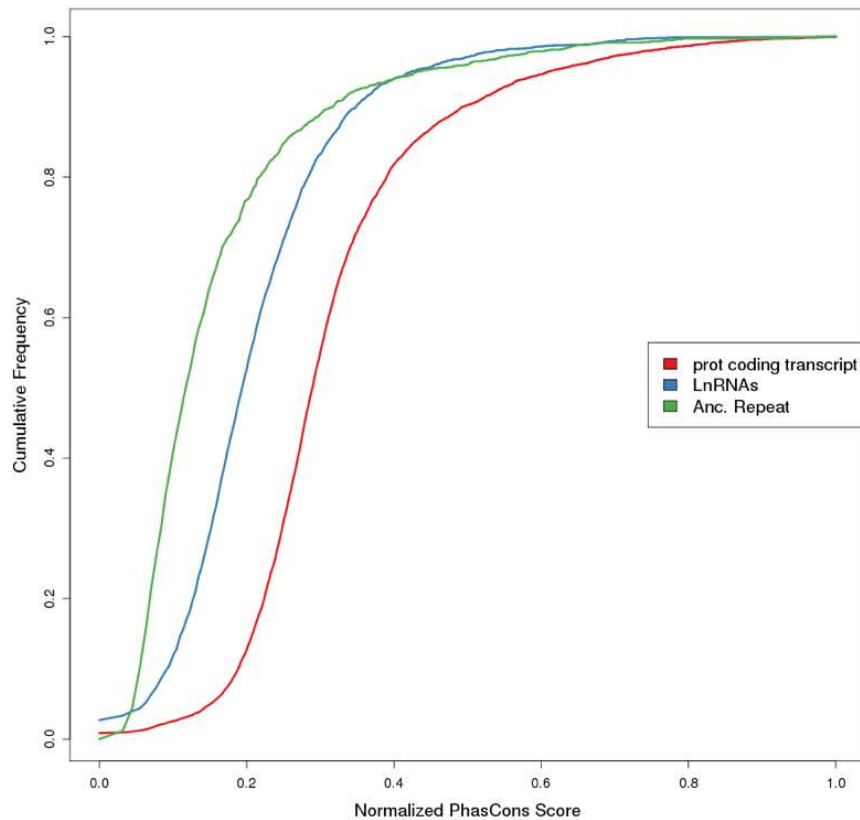


	protein coding	lncRNAs
Size (min/max/avg)	9 /25,252 / 1334	102 /19,269 /843

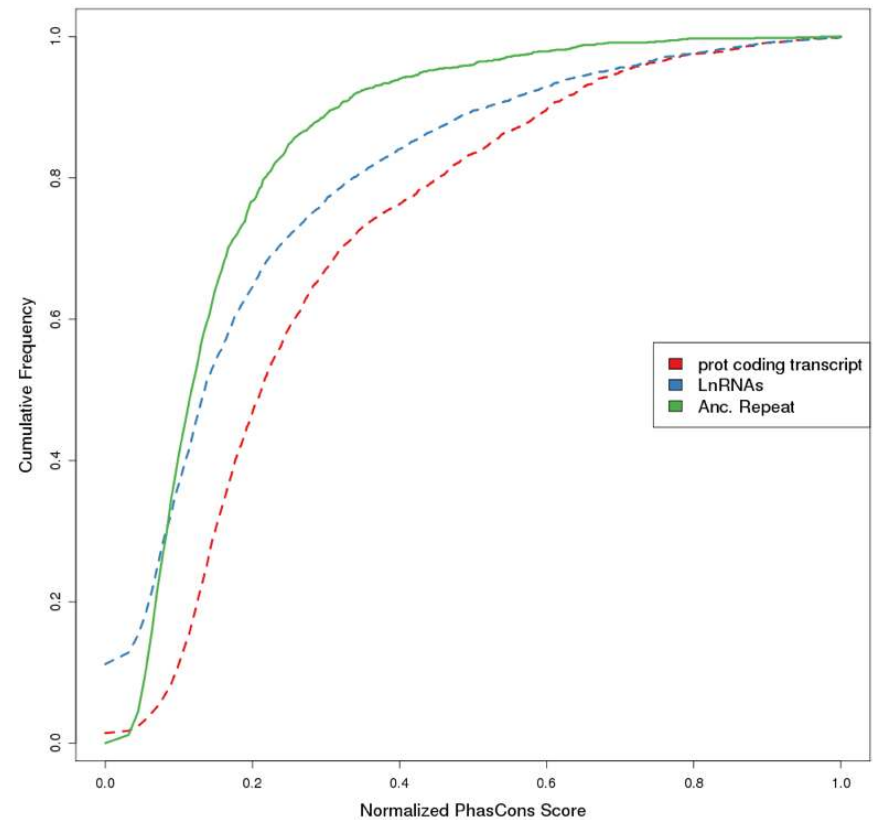
InRNAs are more conserved than background

- ▶ **Normalized PhastCons score**

- ▶ **At the gene level**



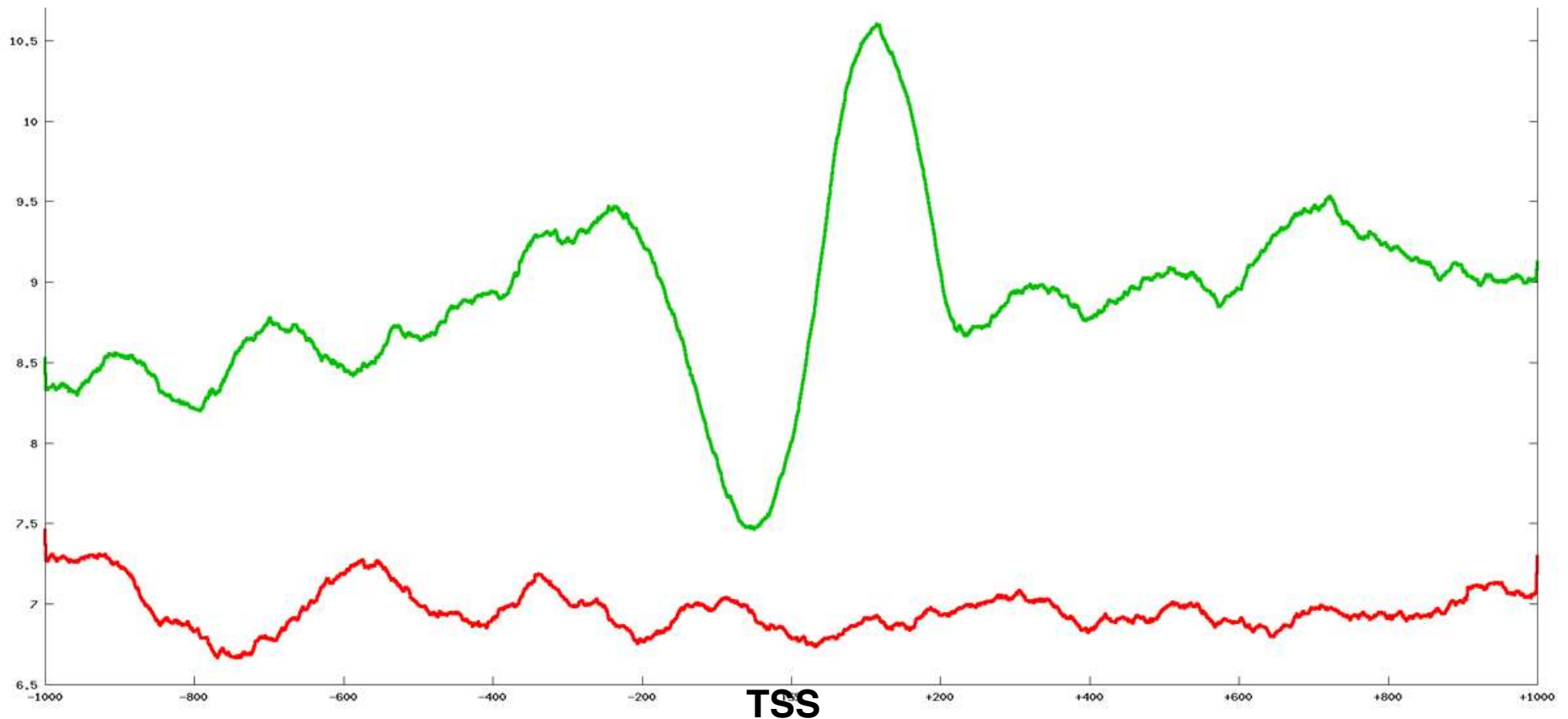
- ▶ **At the promoter level**



InRNAs show characteristic epigenetic marks

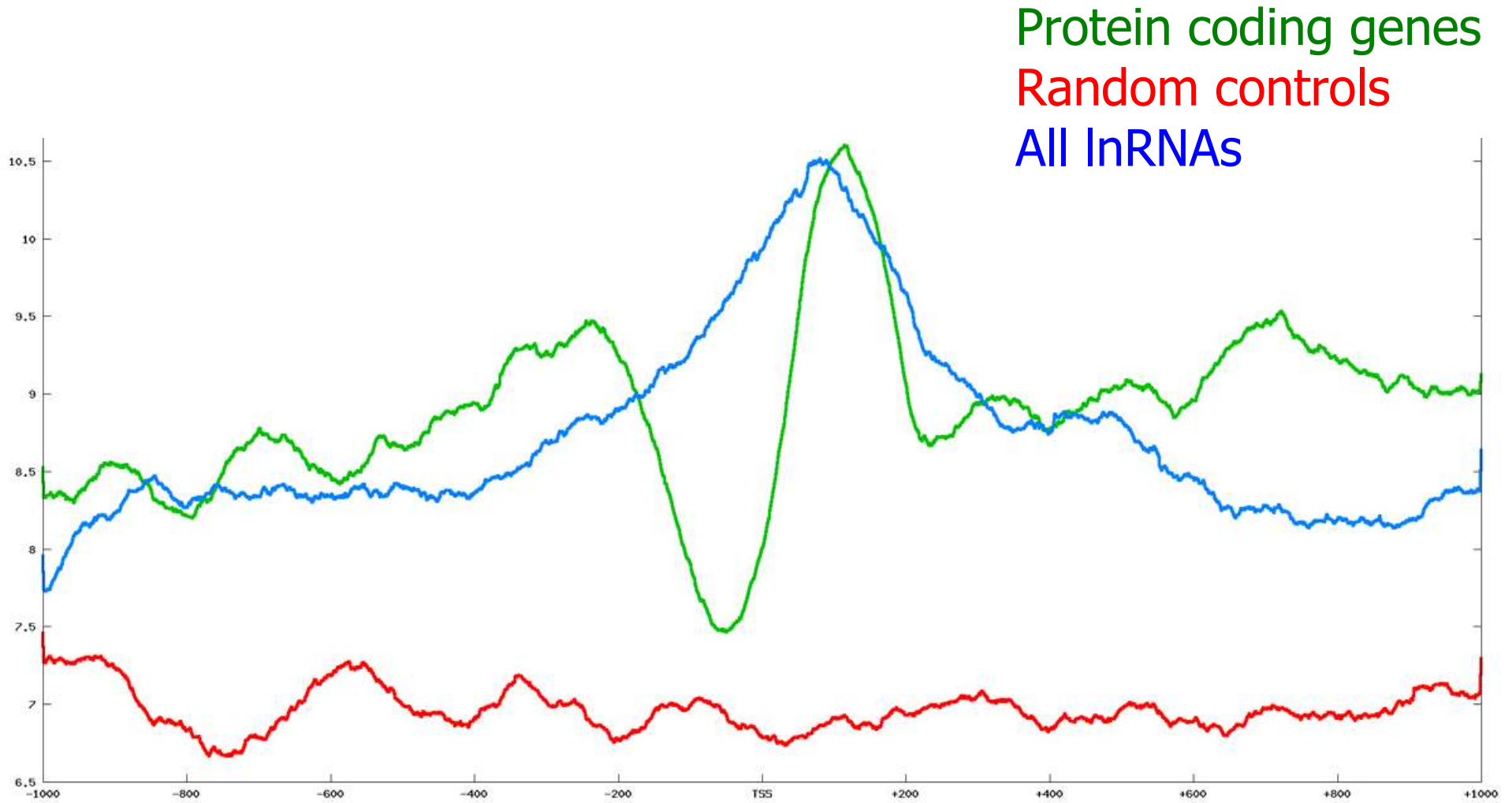
Protein coding genes

Random controls



Nucleosome positioning in CD4+, Schones et al, 2008.

InRNAs show characteristic epigenetic marks



Nucleosome positioning in CD4+, Schones et al, 2008.

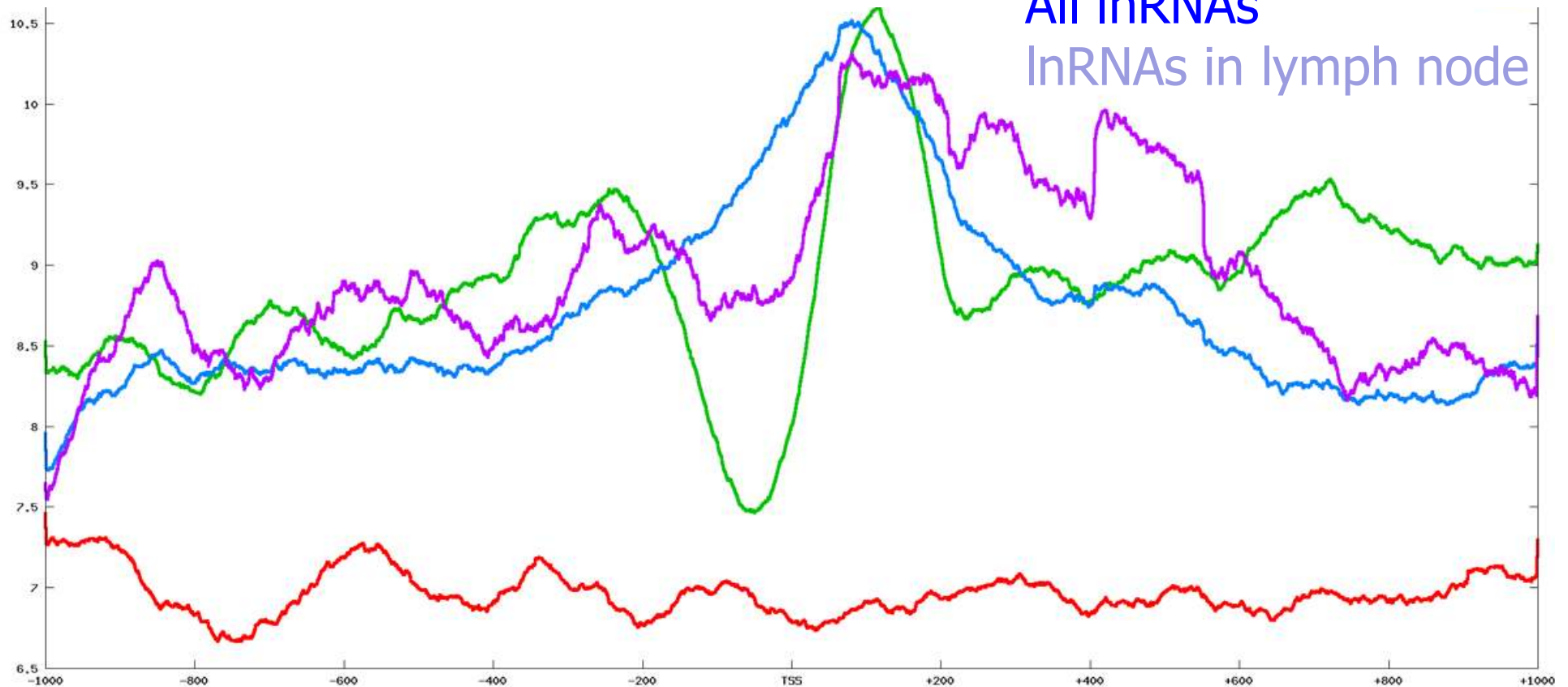
InRNAs show characteristic epigenetic marks

Protein coding genes

Random controls

All InRNAs

InRNAs in lymph node



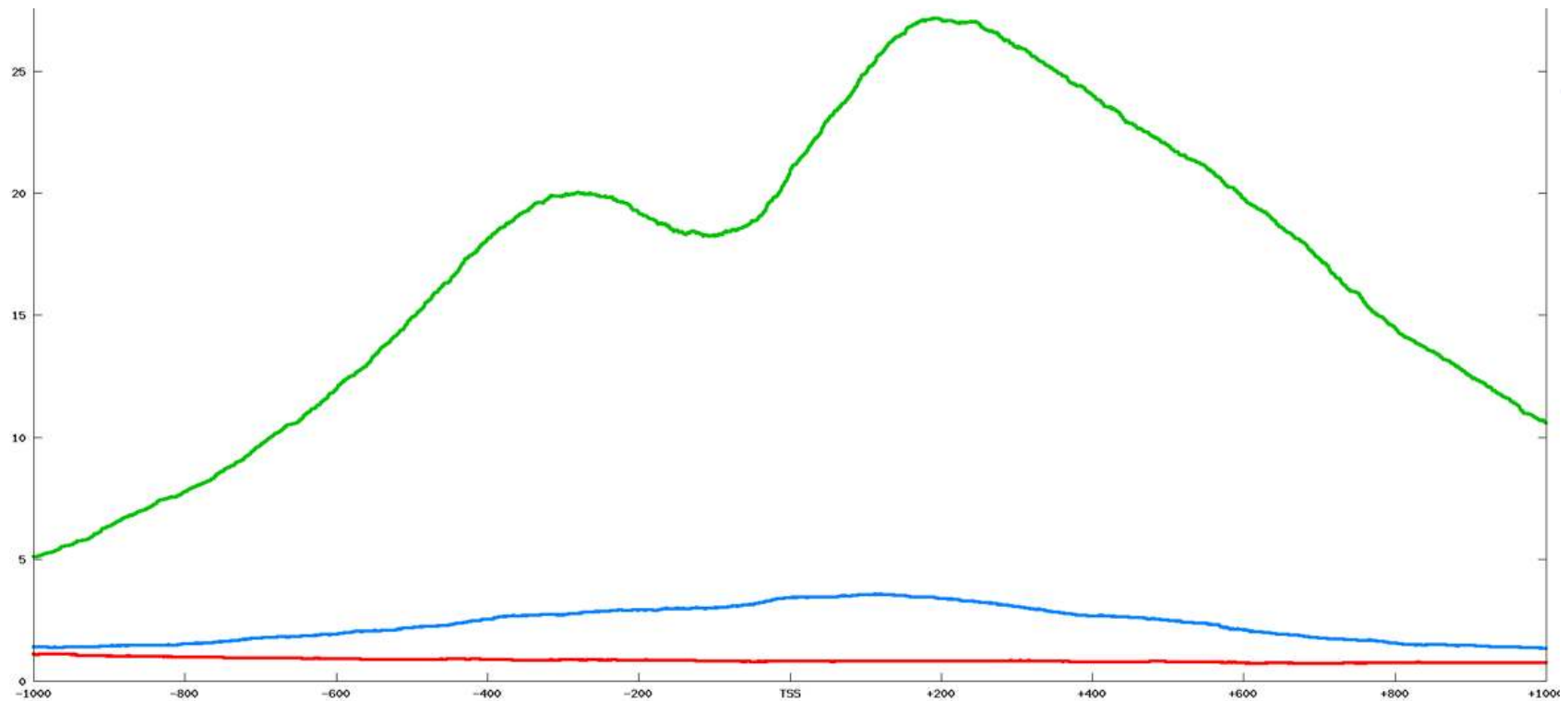
Nucleosome positioning in CD4+, Schones et al, 2008.

InRNAs show characteristic epigenetic marks

Protein coding genes

Random controls

All lncRNAs



H3K4me3 in CD4+, Barski et al, 2008.

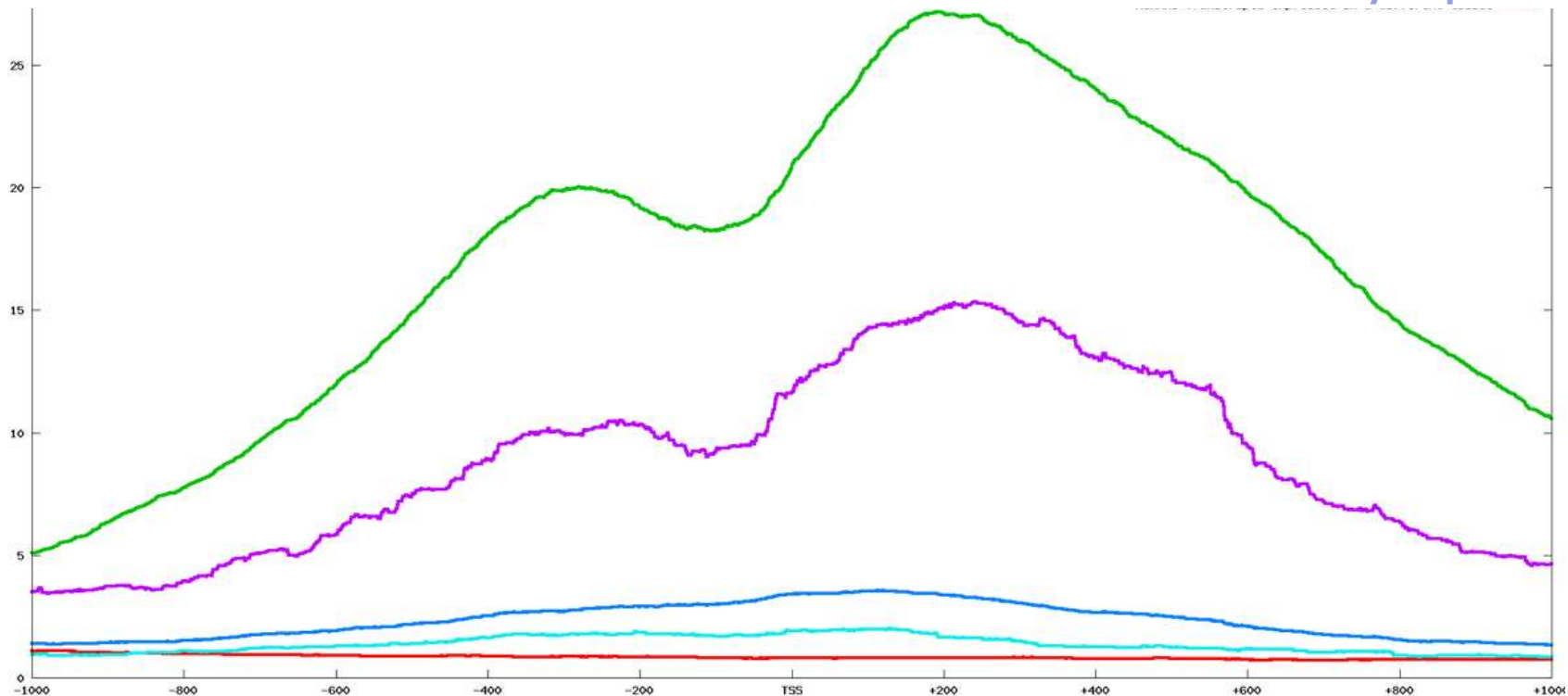
InRNAs show characteristic epigenetic marks

Protein coding genes

Random controls

All InRNAs

InRNAs in lymph node



H3K4me3 in CD4+, Barski et al, 2008.

What is the function of
these lncRNAs?

Using RNASeq to characterize lncRNAs

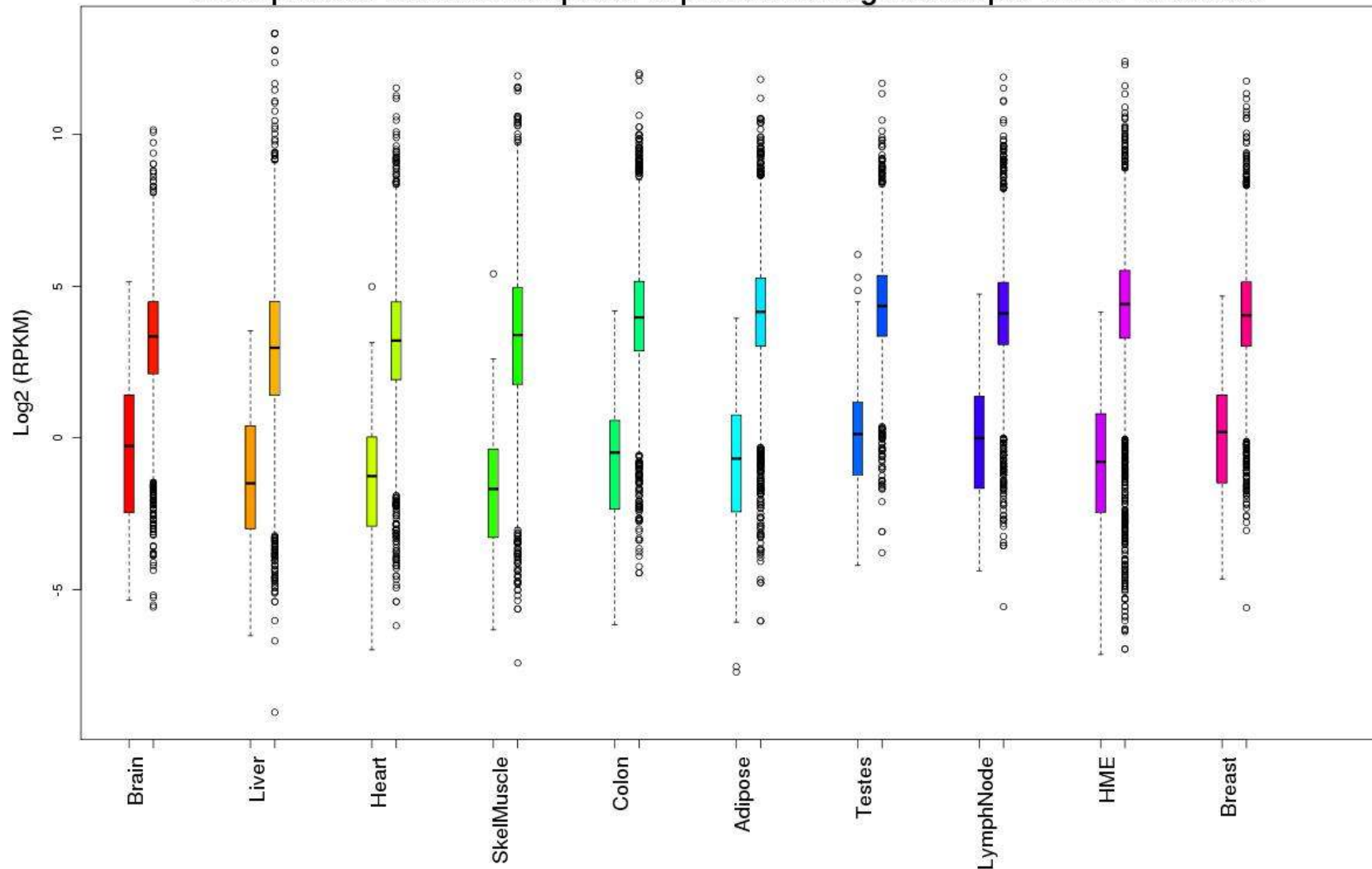
Illumina RNASeq sequences (Wang et al, 2008)

Organism	Origins	# reads	# reads uniquely mapped	Reference
human	ten diverse human tissues + five mammary epithelial or breast cancer cell lines	~400 million reads	~240 millions	Wang E et al., Nature 2008

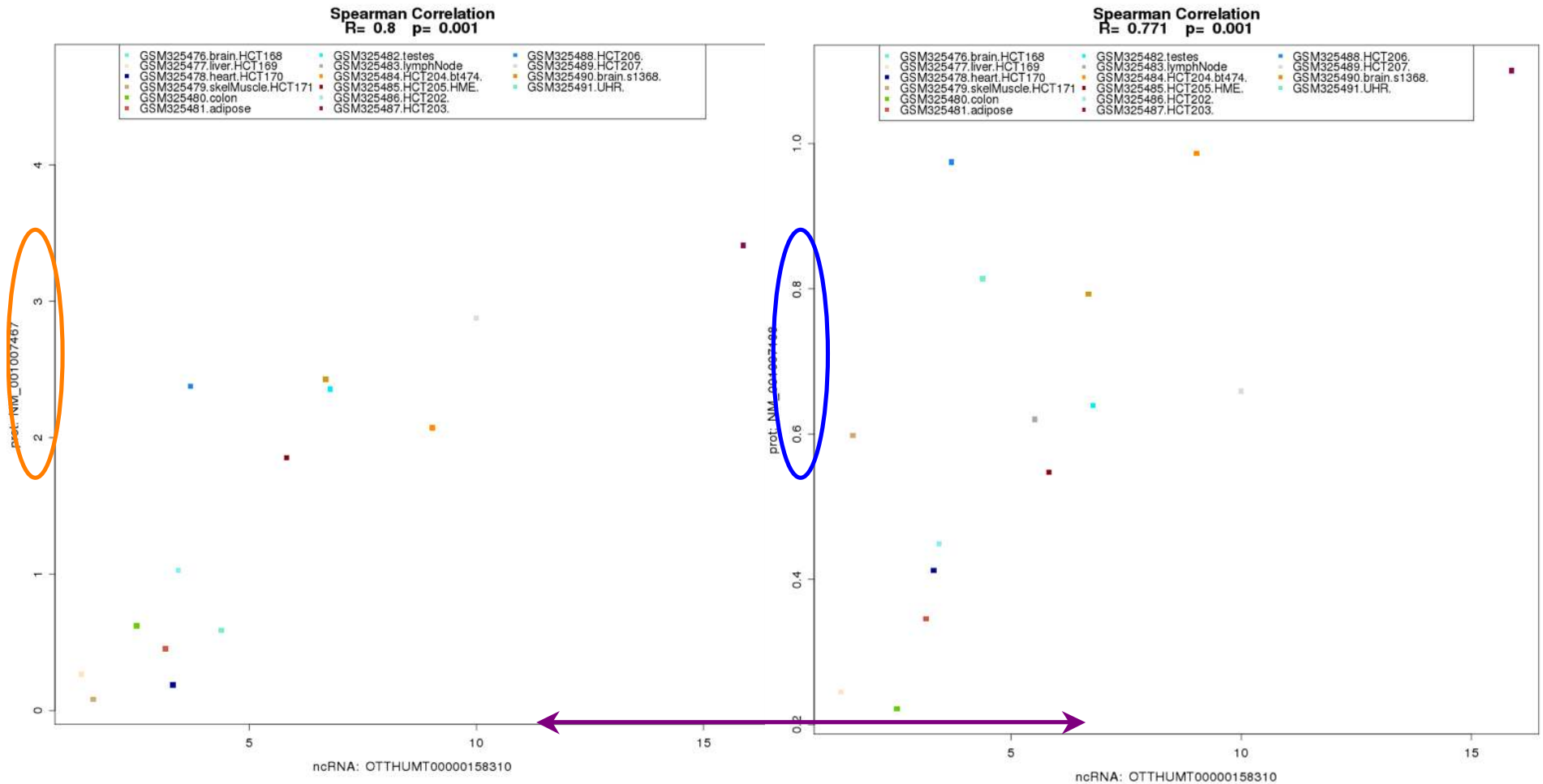
Mapping on genome and transcriptome using GEM (**Paolo Ribeca**)
Quantification with the **Flux Capacitor**

InRNAs show lower expression levels than protein coding genes

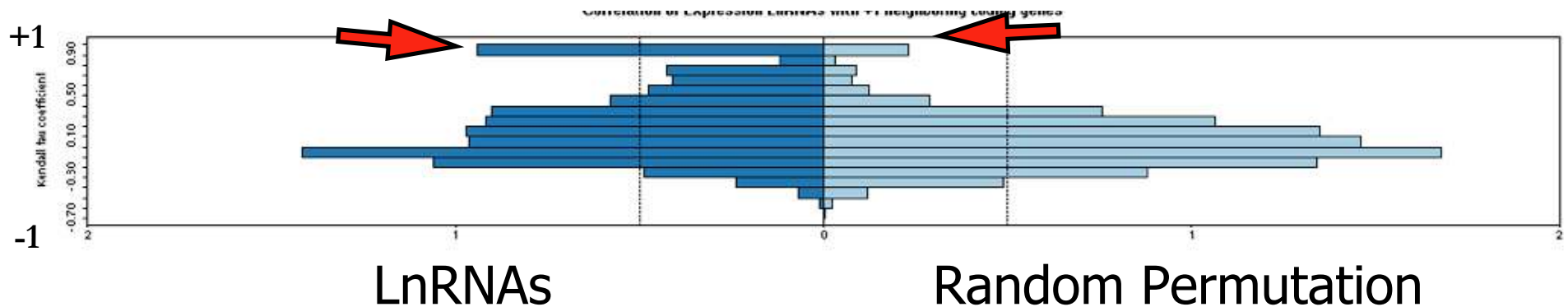
Distribution of the level of expression
for expressed LnRNAs compared to protein coding transcripts within 10 tissues



Correlation of expression between protein coding genes and lncRNAs



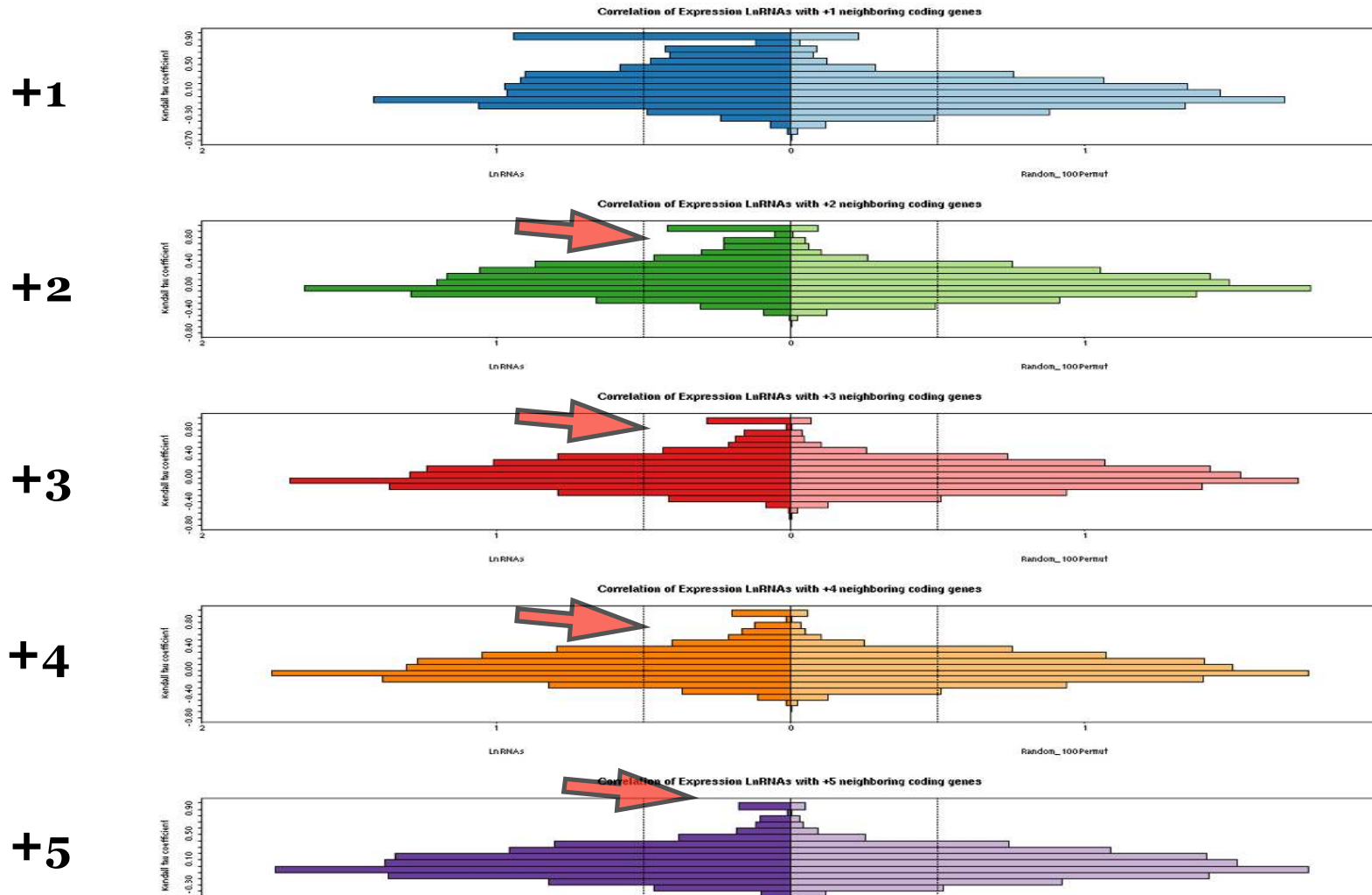
lncRNA expression correlates positively with expression of protein coding genes in the genomic neighborhood



- lncRNAs show positive rather than negative correlation in expression with protein coding genes
- lncRNAs show higher correlation with neighbor (2Mb) protein coding genes than with distal genes.

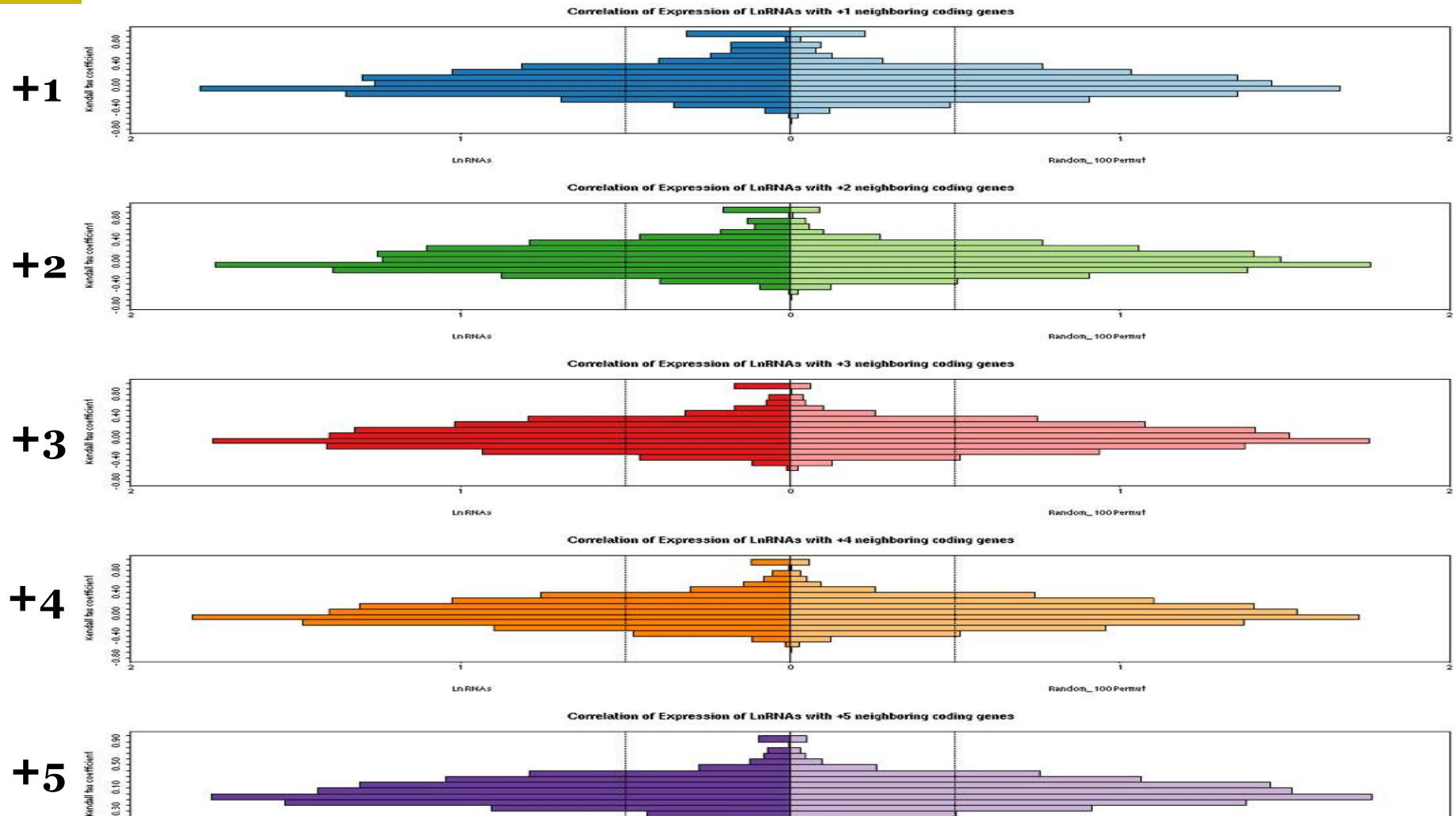
Correlation with protein expression

Correlation of expression between LncRNAs and +1 to +5 coding transcripts

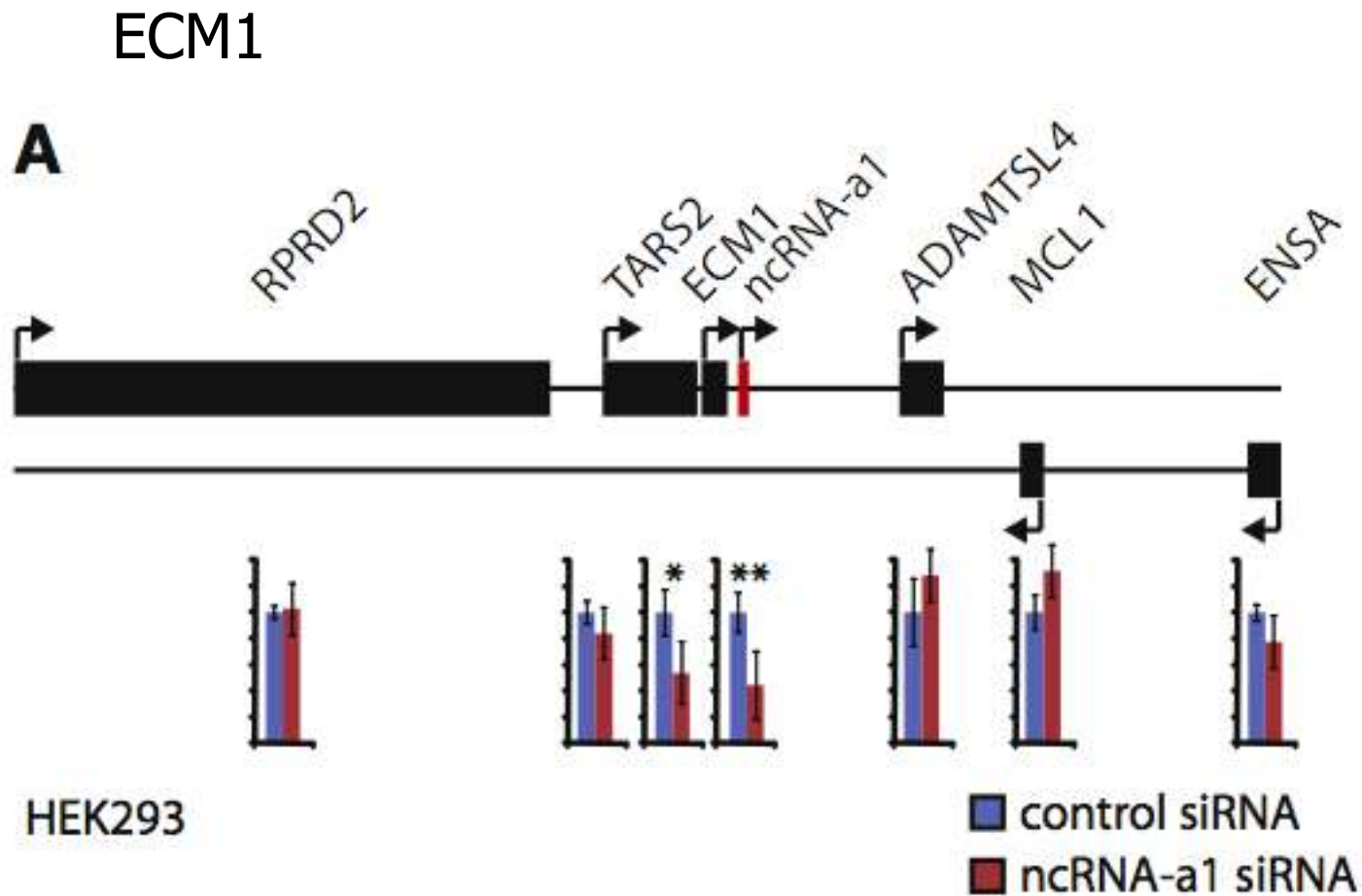


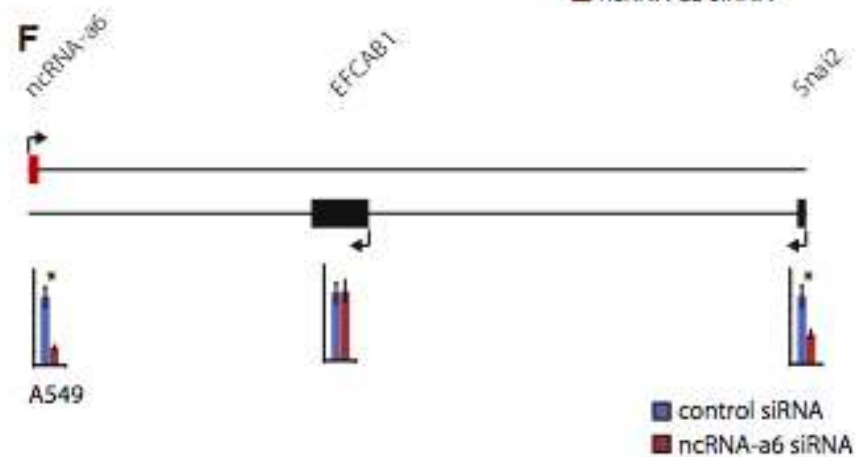
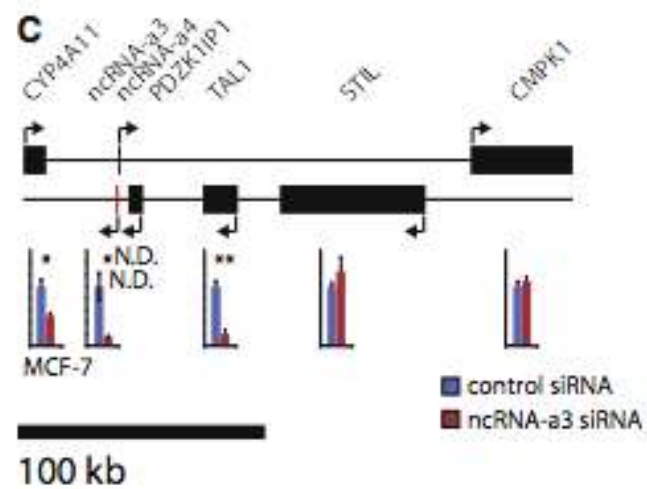
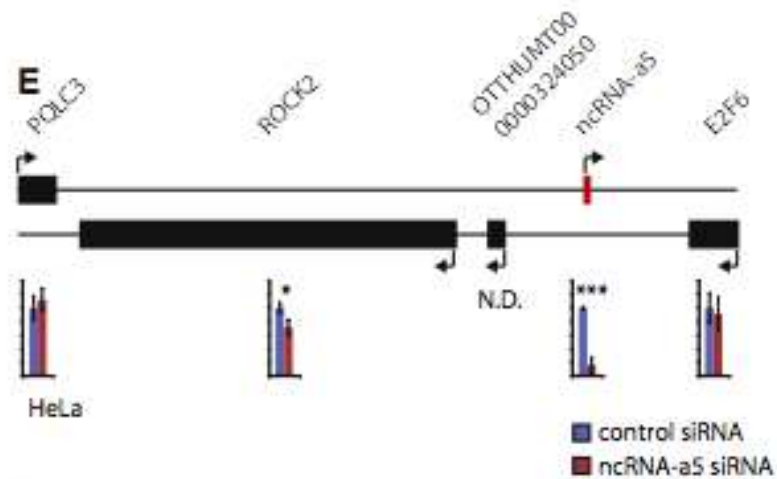
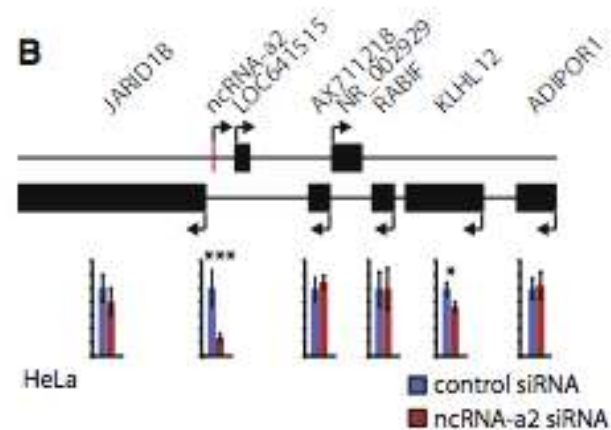
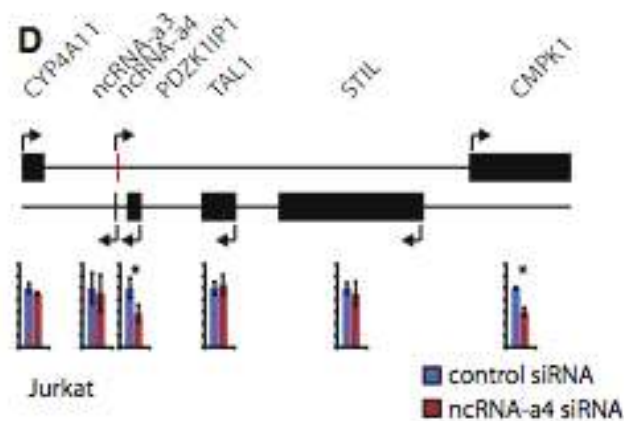
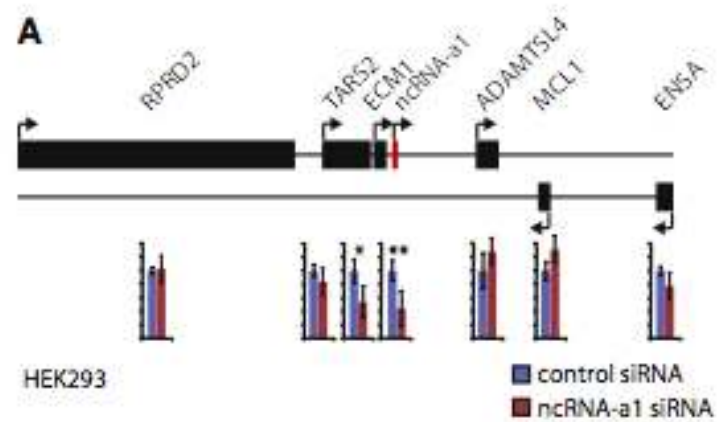
Correlation with protein expression

Correlation of expression between LncRNAs and **+1 to +5 coding transcripts (in opposite orientation)**



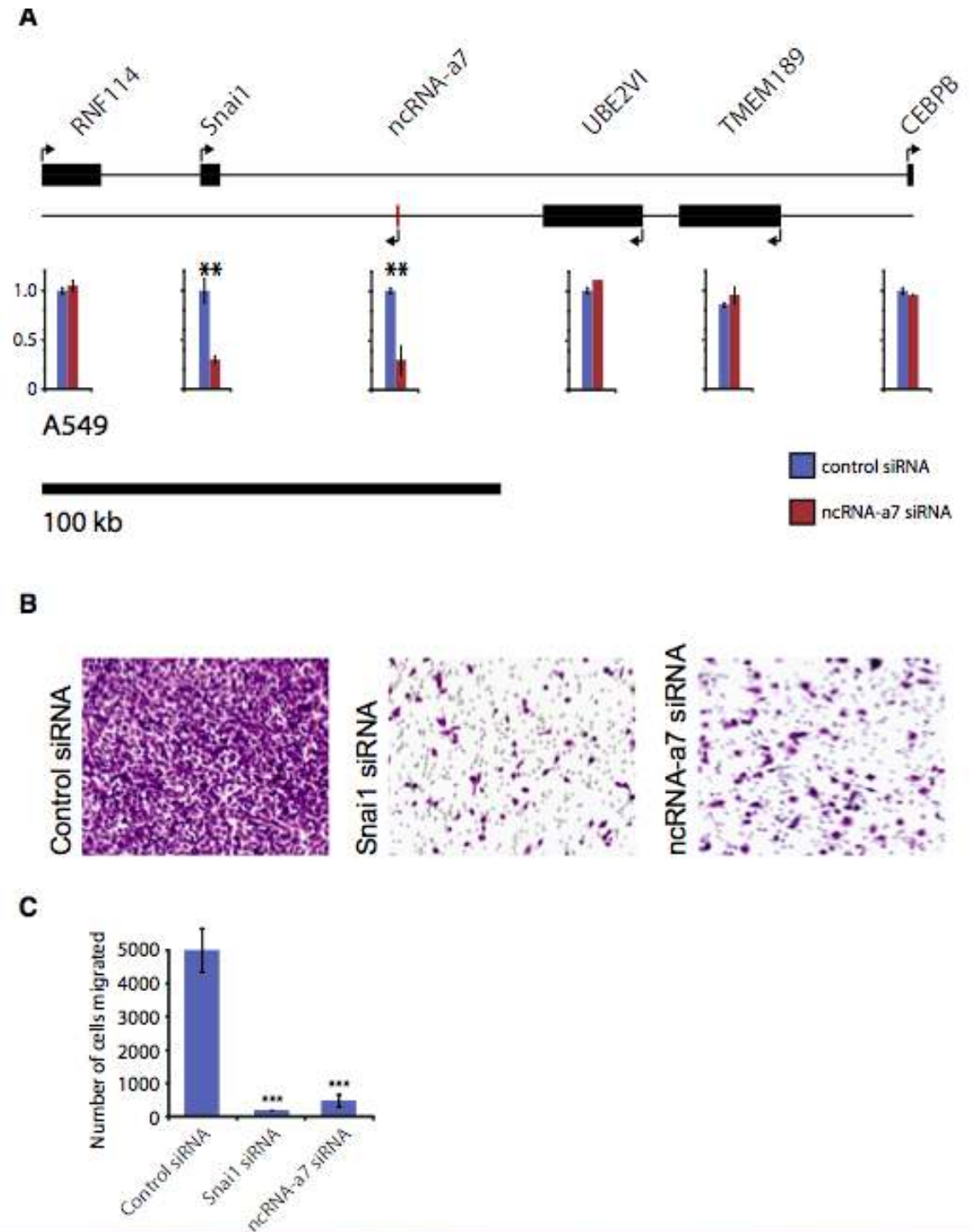
siRNA knockdowns of lncRNA downregulate neighbor protein coding genes: HEK293, A549





Regulation of Snai1 by lncRNA

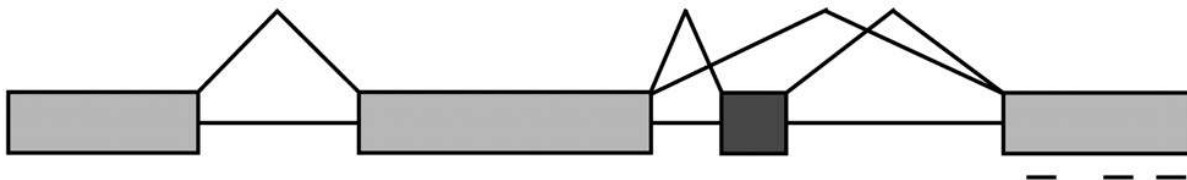
- Snail genes are the regulators of cell adhesion, migration and epithelial-mesenchymal transition
- siRNA against ncRNA-a7 reduces migrating cells to about 10% of the control levels



InRNA Arrays

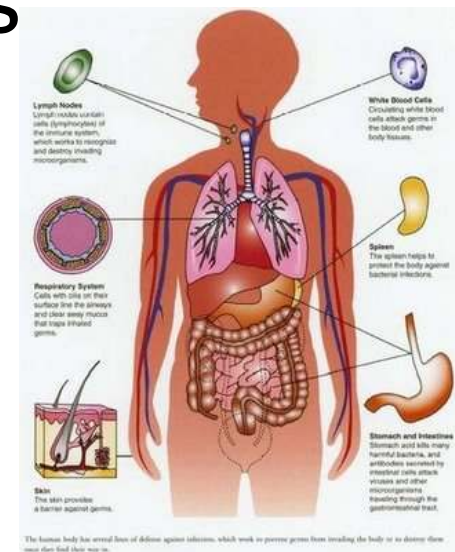
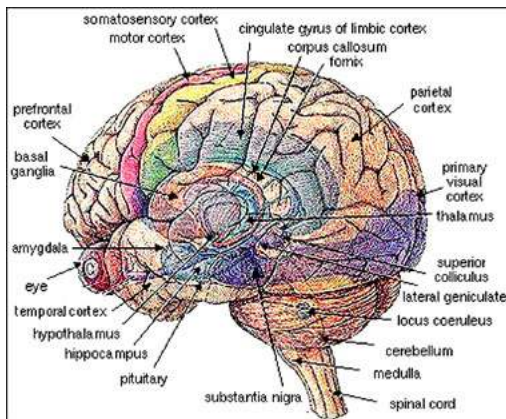
Rory Johnson, CRG

- Custom Agilent 8X60k format
- 9747 lncRNAs
- ~ 6 distinct probes / lncRNA
- Standard controls AND 3231 random coding genes
- COST = EURO 250 per array hyb



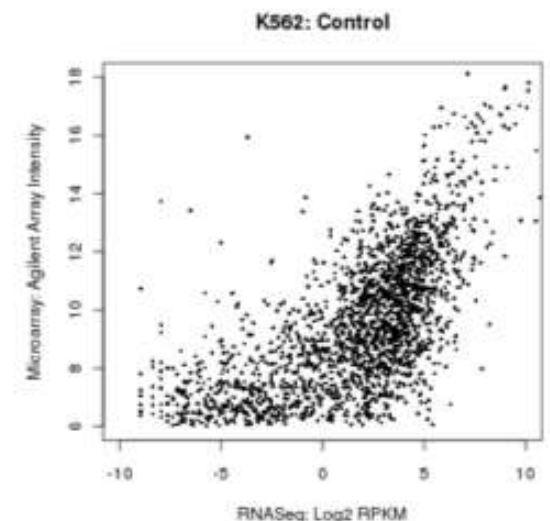
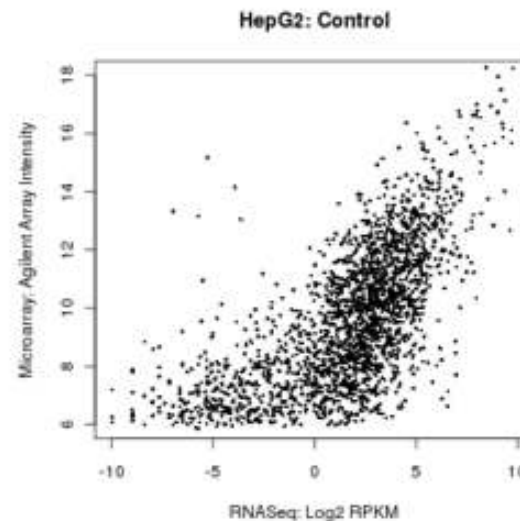
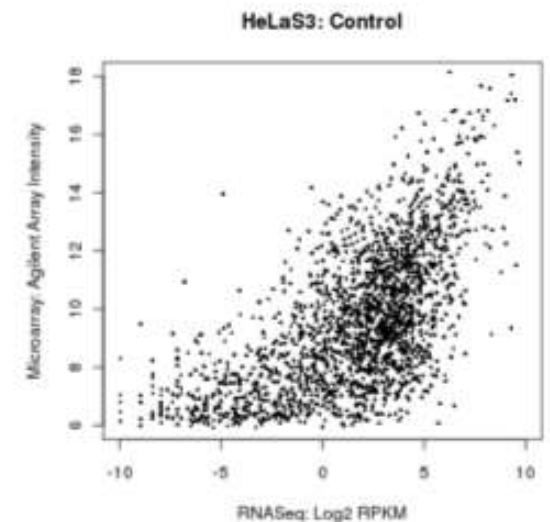
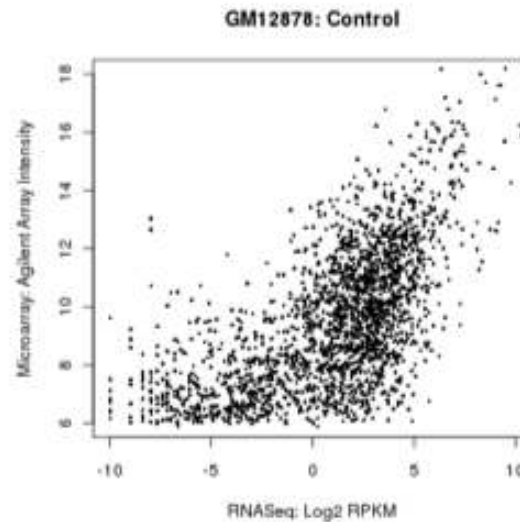
Tissue Samples

- 20 normal human tissues (Ambion)
- ~15 distinct brain regions (Isidre Ferrer)
- 7 cell lines: ENCODE Tier 1 & 2
- Macaque, Chimpanzee brains



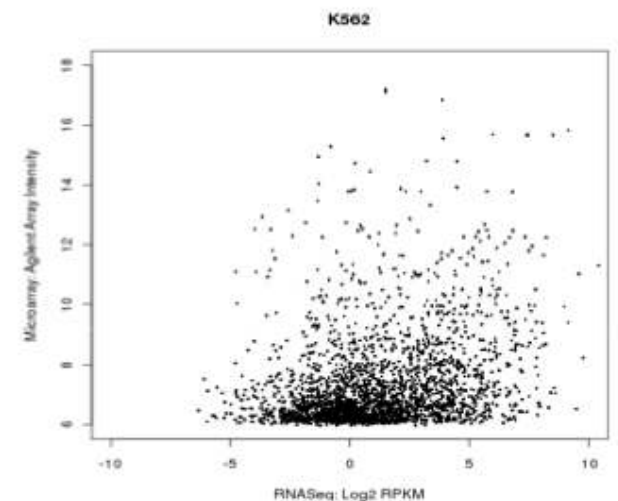
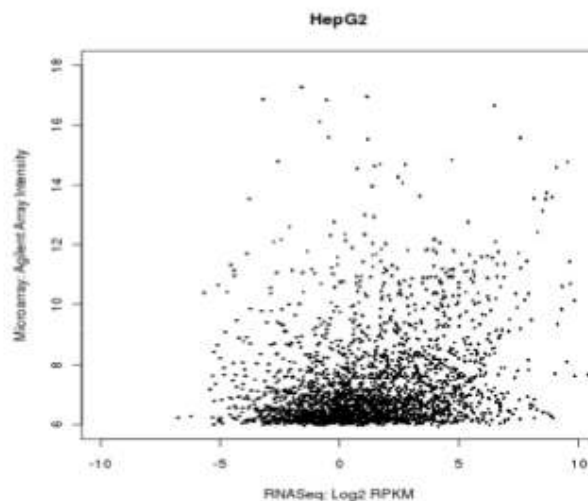
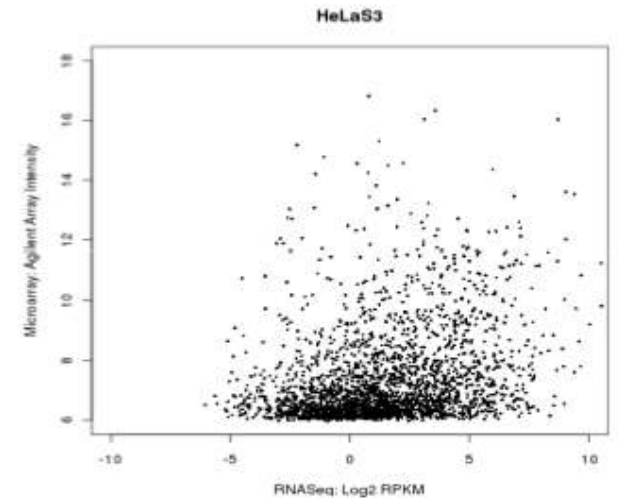
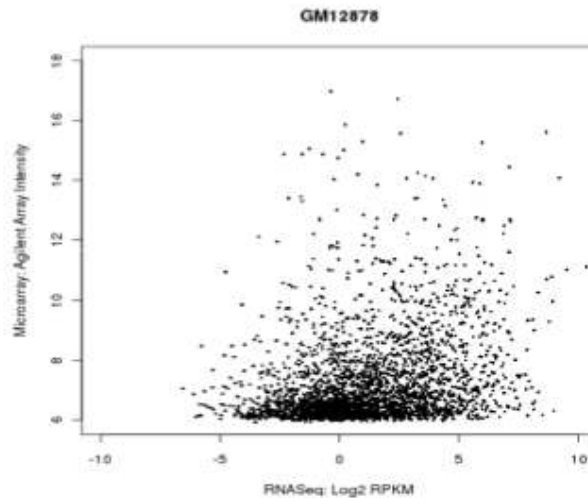
RNASeq vs arrays

- 4 cell lines.
ENCODE
RNASeq data
- **Protein
coding
genes**
(~2500) correlation about
0.65



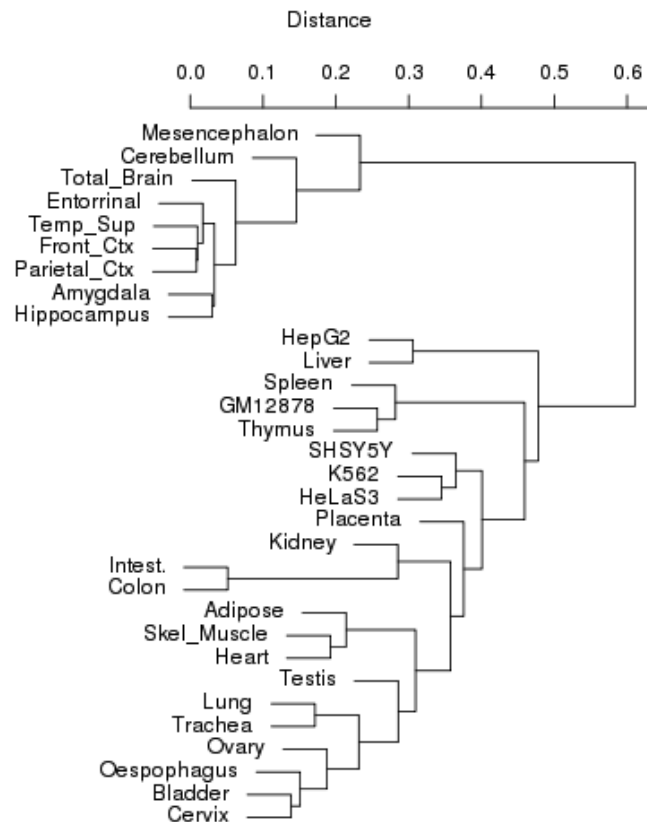
RNASeq vs arrays

- 4 cell lines.
ENCODE
RNASeq data
- **InRNAs**
correlation
about 0.3

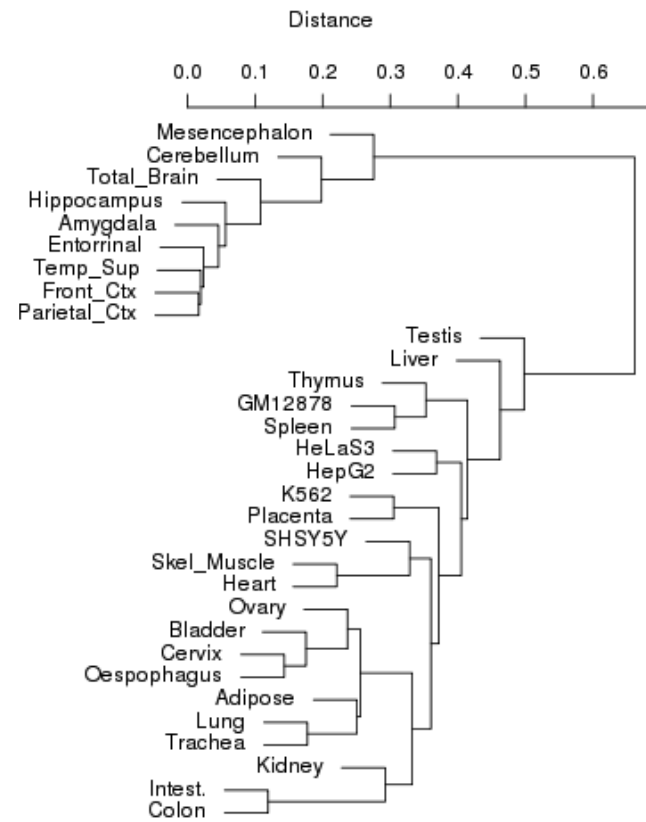


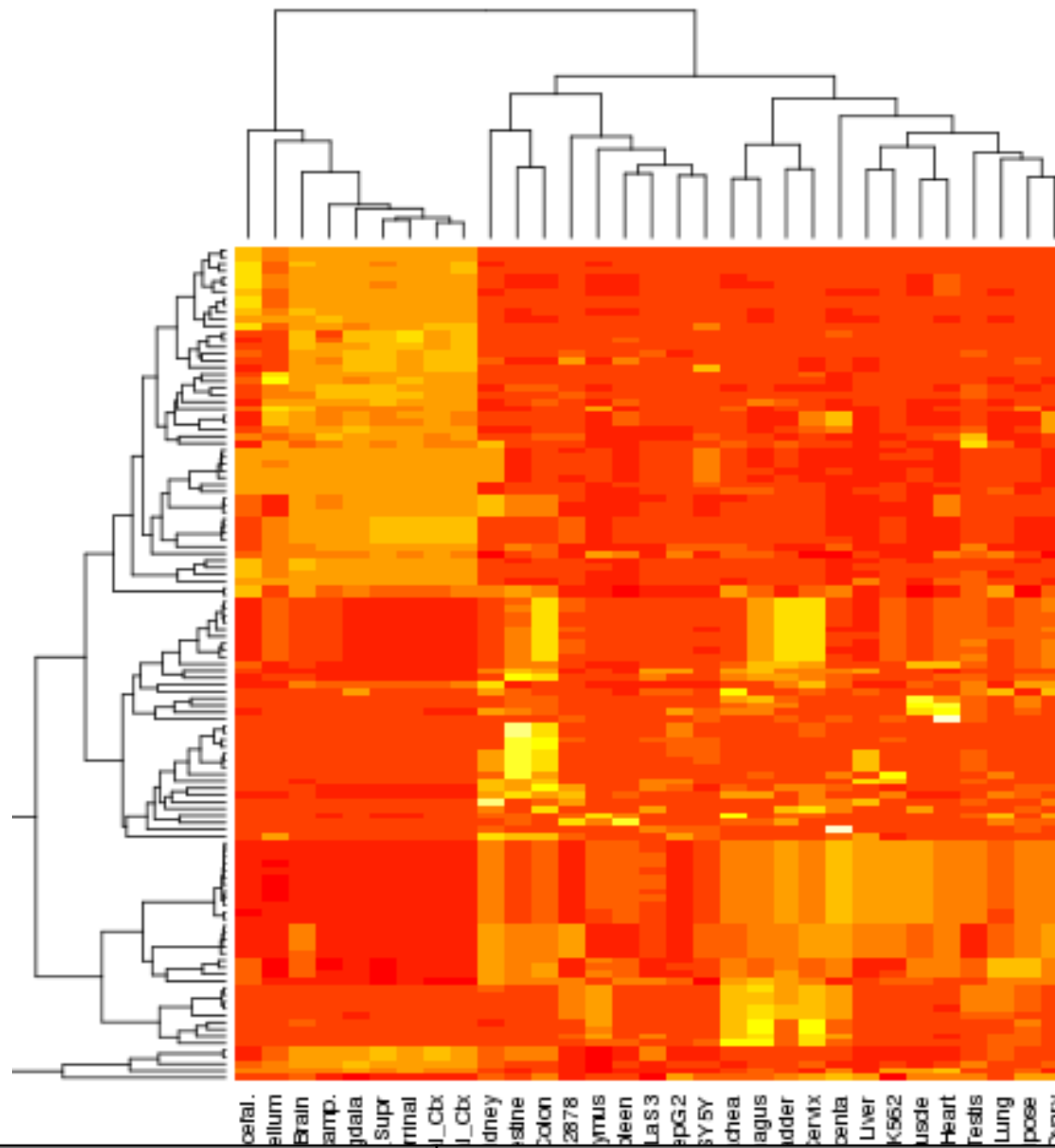
Clustering of tissues according to lncRNA expression reproduces clustering according to coding expression

Protein coding probes



Non coding probes





InRNAs
with
especific
brain
expression

Acknowledgements

- RNASeq
Micha Sammeth
CRG/CNAG
- eQTL
Manolis Dermitzakis
Stephan Montgomery
Geneva
- Long Non Coding RNAs
Ramin Shiekhataar
Ulf Anderson
Malte Beringer
CRG/Wistar
Cedric Notredame
Thomas Derrien
Andrea Tanzer
Giovani Bussoti
Rory Johnson
CRG

