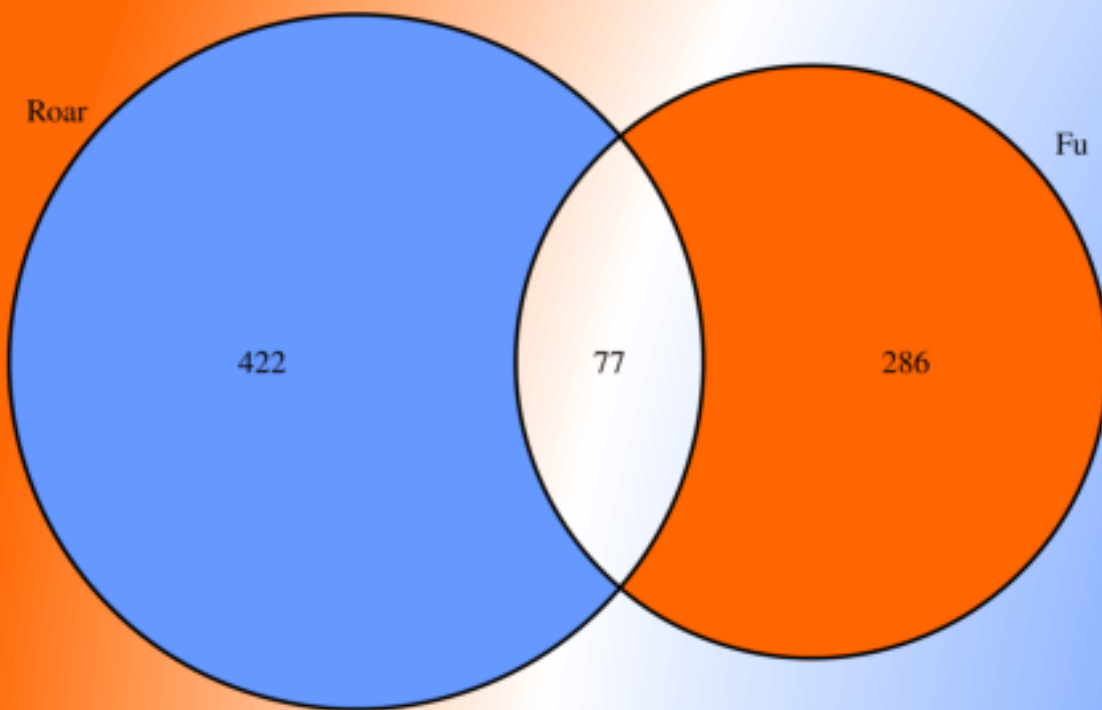


# Roar: detecting alternative polyadenylation with standard RNA sequencing libraries

Elena Grassi<sup>1</sup>, Elisa Mariella<sup>1</sup>, Antonio Lembo<sup>1</sup>, Ivan Molineris<sup>1</sup>, Roberto Santoro<sup>1</sup>, Giorgio Merlo<sup>1</sup> and Paolo Provero<sup>1,2</sup>  
<sup>1</sup>Department of Molecular Biotechnology and Health Sciences, Molecular Biotechnology Center, Via Nizza 52, University of Turin, Italy  
<sup>2</sup>Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Via Olgettina 60, Milan, Italy  
<http://www.bioconductor.org/packages/release/bioc/html/roar.html> - [elena.grassi@unito.it](mailto:elena.grassi@unito.it)

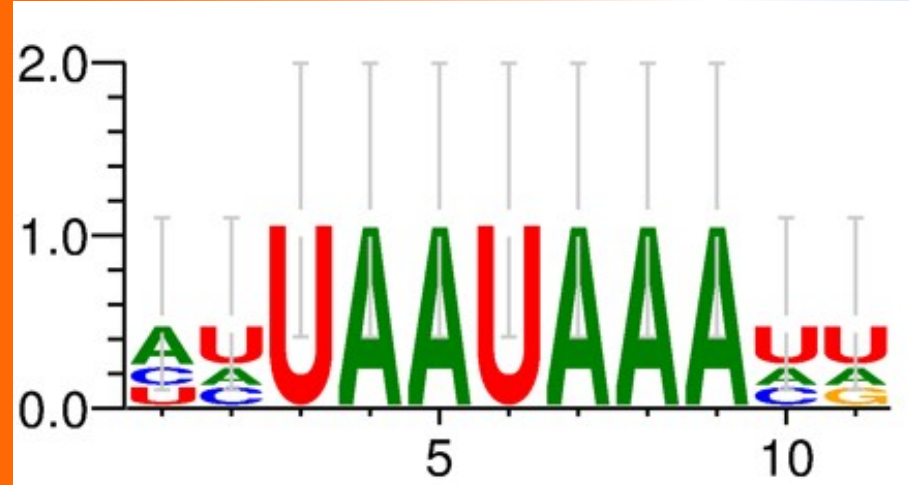
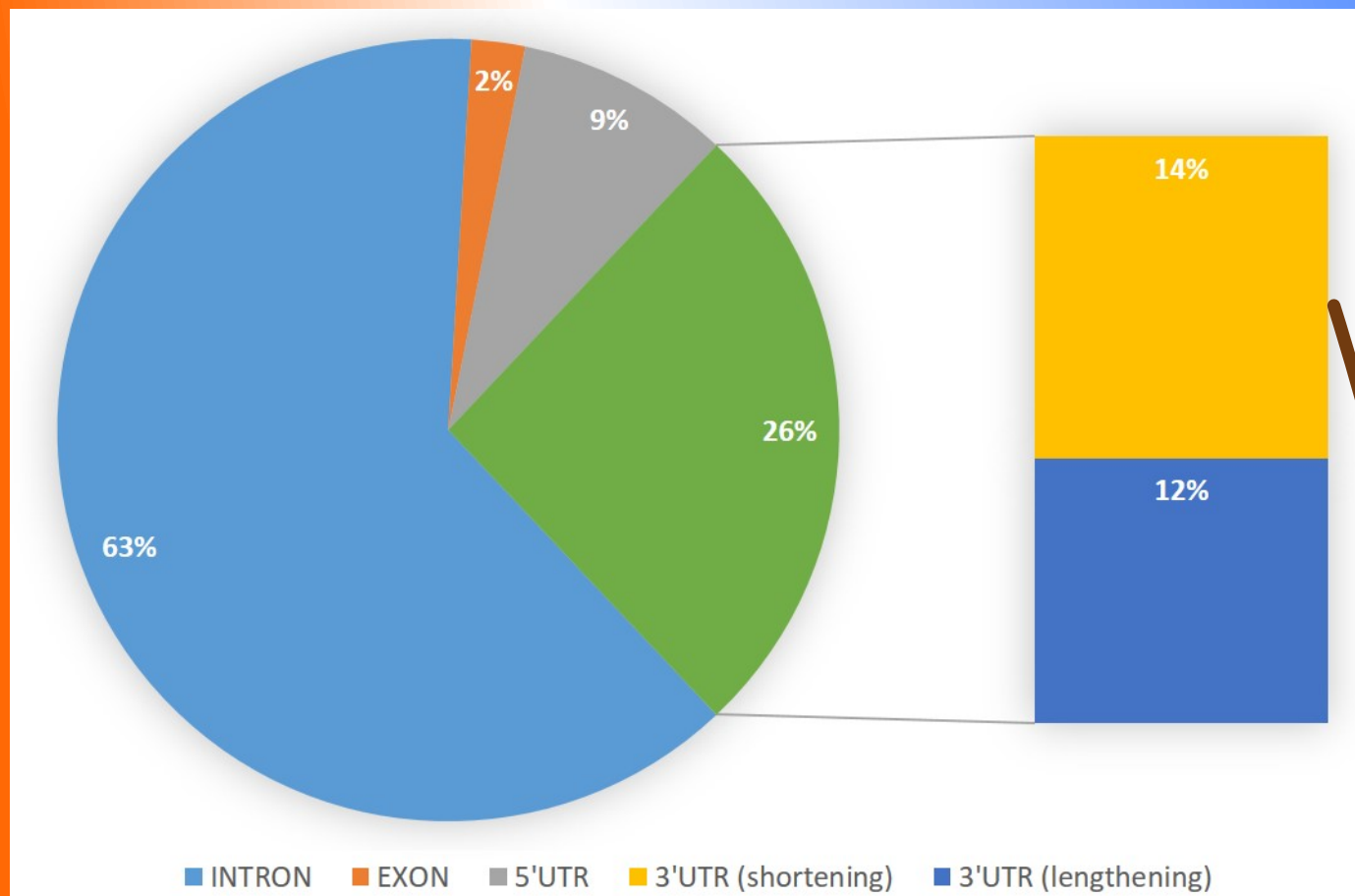


## Validation - in silico



To validate our pipeline we compared roar results (in terms of genes found shortened or lengthened by both methods) with a deep sequencing ad hoc approach using **breast and normal cancer cell lines** and with another one based on microarray analyses using **testis and brain**. All eight overlaps, save one (for lengthened genes), were significant (hypergeometric test, pvalue <0.05).

## Looking for the genetic determinants of APA



We are working on **apaQTLs**: loci that influence the m/M of genes in large cohorts of human lymphoblastoid cell lines. The polymorphisms found in 3'UTRs that determine shortening effects frequently creates a **canonical PAS**.

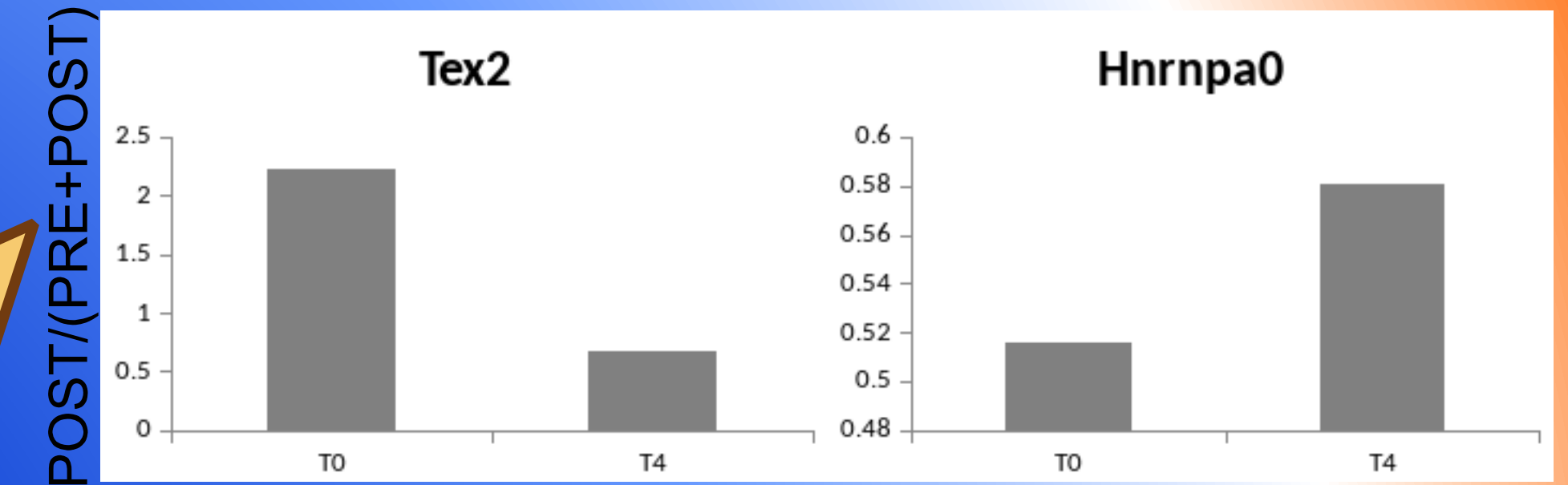


1. we use the **PolyADB** or **APAsdb** database to obtain coordinates for the orange (called PRE) and blue (called POST) portions of transcripts
2. from read counts and fragment lengths we obtain **m/M**. It represents the ratio of short versus long isoforms in a given sample
3. roar is the ratio of the m/M of the two conditions (ie. a **Ratio Of A Ratio**)
4. we use the **Fisher test** to identify transcripts with significant differences, considering only genes expressed in both conditions

What can roar do with **your data**?

## Validation - in vitro

We performed qRT-PCR to validate our predictions of polyadenylation behaviour for 4 genes during the **differentiation** of neural progenitor cells towards the **inhibitory neuron fate**. As expected the vast majority (47 over 52) of these genes tend to express longer isoforms in the mature cells.



## Working on ES differentiating towards different fates

H1 or HUE64 vs	N. shortened	N. lengthened
Neural Progenitor Cells	2001	536
Trophoblast-like Cells	1151	1550
Mesenchymal Stem Cells	1083	733
Mesoendoderm	1387	745
Endoderm	325	67
Mesoderm	150	245
Ectoderm	350	51

The global trends reflect the "lengthening" tendency that has been found during mouse development. We investigated if the lost portions of 3'UTR harbour some enriched miRNA seeds and this indeed is the case: 3 miRNA families (**miR-590/590-3p**, **miR-200bc/429** and **miR-433**) preferentially target (Bonferroni corrected hypergeometric p-values < 0.05) the longest isoforms of genes shortened in H1 versus NPC and two (**miR-124/506**, **miR-134**) those of the genes lengthened in H1 versus the Mesoendoderm differentiated cells.

[1] GEO dataset GSE16256, runs: SRR486237 SRR486238 SRR486239 SRR486240 SRR486241 SRR486242 SRR488136 SRR488137 SRR488684 SRR488685.  
 [2] GEO dataset GSE27003, runs: SRR097788, SRR097789, SRR097790.  
 [3] GEO dataset GSE30352, runs: SRR306857, SRR306858, SRR306838, SRR306839, SRR306840, SRR306841, SRR306842, SRR306843.  
 [5] Ran Elkon, Alejandro P. Ugalde, and Reuven Agami. Alternative cleavage and polyadenylation: extent, regulation and function. Nature Reviews Genetics, 14(7):496-506, June 2013.  
 [6] Yonggui Fu, Yu Sun, Yuxin Li, Jie Li, Xingqiang Rao, Chong Chen, and Anlong Xu. Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. Genome research, 21(5):741-7, May 2011.  
 [7] Antonio Lembo, Ferdinando Di Cunto, and Paolo Provero. Shortening of 3'UTRs correlates with poor prognosis in breast and lung cancer. PLoS one,

