

RESEARCH

Open Access

# Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

Mar González-Porta<sup>1</sup>, Adam Frankish<sup>2</sup>, Johan Rung<sup>1</sup>, Jennifer Harrow<sup>2</sup> and Alvis Brazma<sup>1\*</sup>

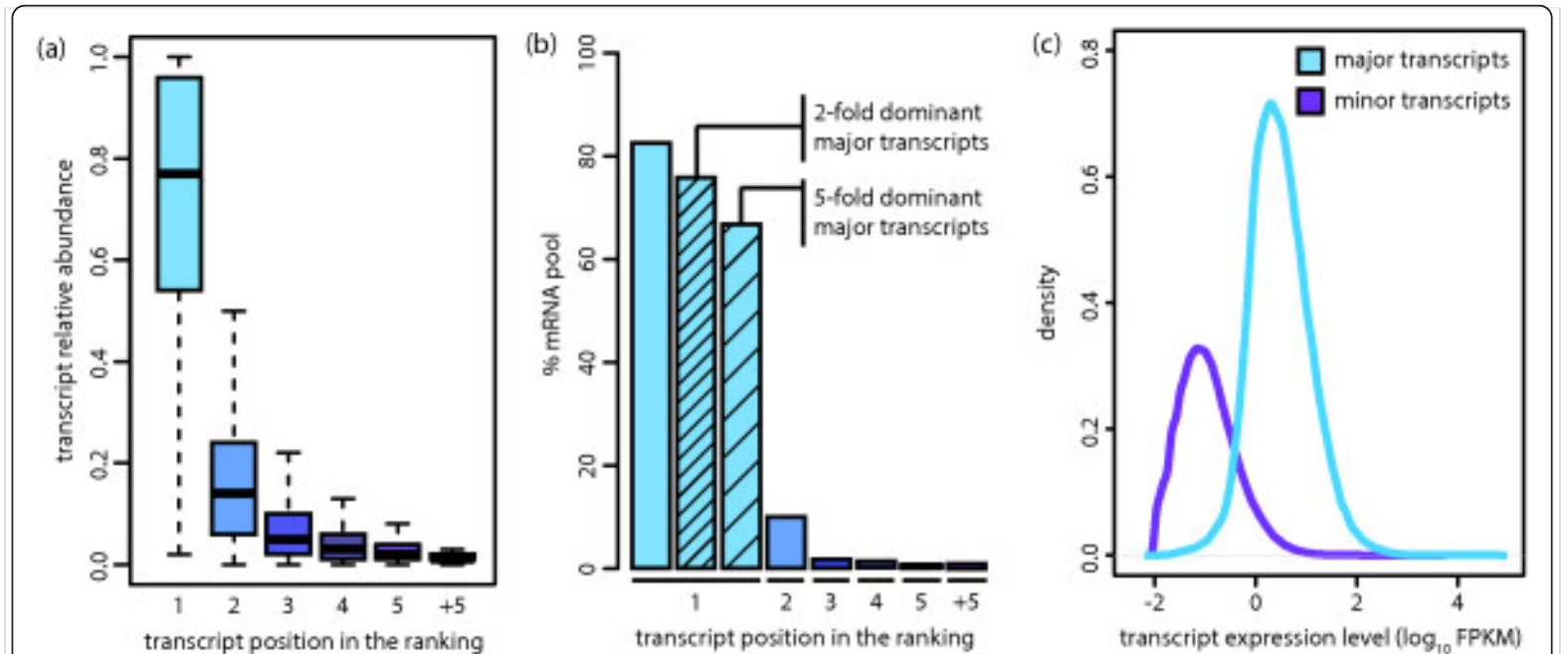
## Abstract

**Background:** RNA sequencing has opened new avenues for the study of transcriptome composition. Significant evidence has accumulated showing that the human transcriptome contains in excess of a hundred thousand different transcripts. However, it is still not clear to what extent this diversity prevails when considering the relative abundances of different transcripts from the same gene.

**Results:** Here we show that, in a given cell, the transcript with the highest relative abundance is at a significantly higher level than others, that is, the total mRNA from protein coding loci, is dominated by a single transcript. We detect a high degree of overlap between the transcripts of alternatively spliced transcripts that are present in the same gene, suggesting that although some minor transcripts may contribute to the proteome. However, we find that the major transcript does not code a protein.

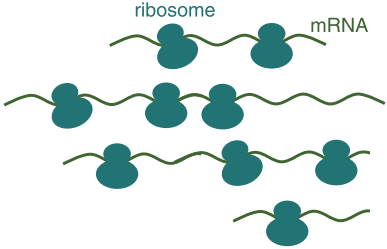
**Conclusions:** Overall, our findings suggest that each gene expresses one dominant transcript per gene and that not all the transcripts contribute to protein diversity. This observation allows us to predict the functional impact of the deletion of a transcript.

**Keywords:** splicing, transcriptome, gene expression



**Figure 1 Most protein coding genes express one predominant transcript.** (a) Relative abundance of the subset of transcripts in each position of the ranking for the primary tissues dataset. For each gene, transcripts were ranked based on their relative abundances. There is generally one predominant transcript over the rest. (b) Percentage of the studied mRNA pool explained by each category of transcripts for the BM dataset. The mean percentage for all samples is represented here. Major transcripts represent approximately 85% of the studied mRNA population and were further classified into two-fold and five-fold dominant. (c) Expression distribution for major and minor transcripts in the tissue dataset. We detect a total of 31,902 transcripts expressed above 1 FPKM in at least one tissue and 26,641 different major transcripts.

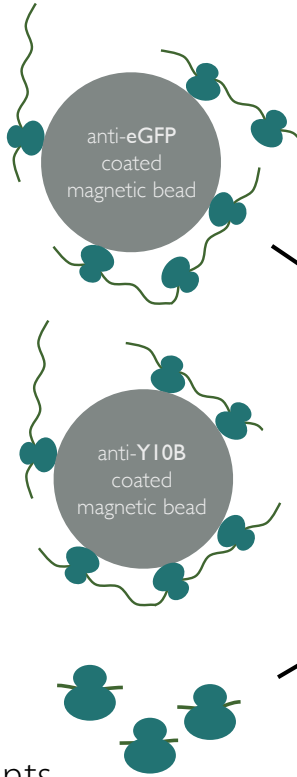
# a) ribosome profiling



**bacTRAP IP**  
L10a ribosomal protein tagged with eGFP

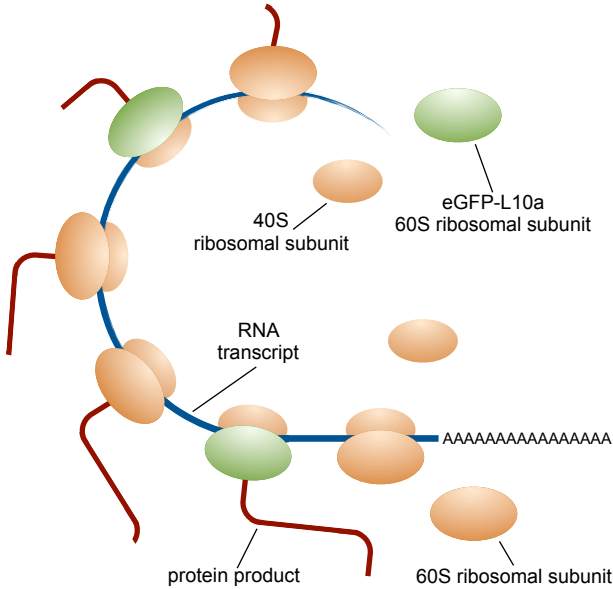
**Y10B antibody IP**  
IP with anti-5.8S rRNA antibody

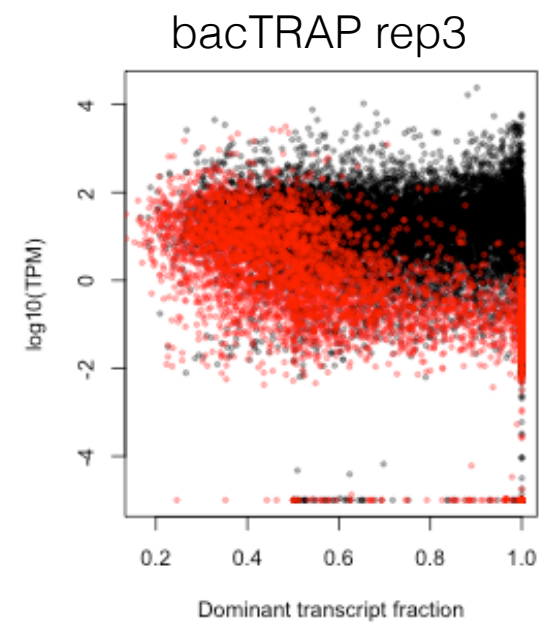
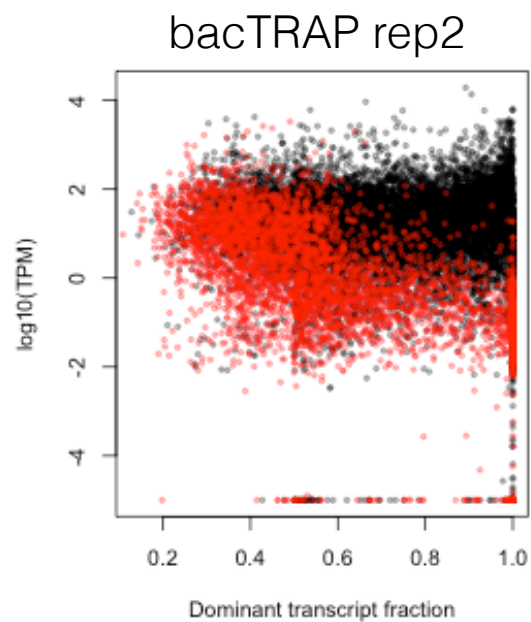
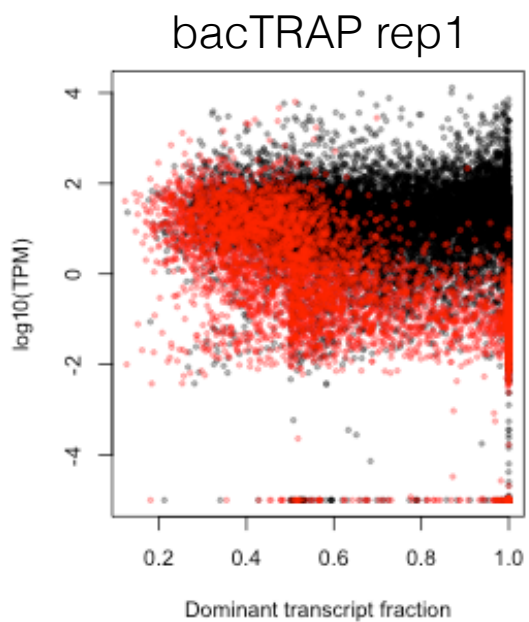
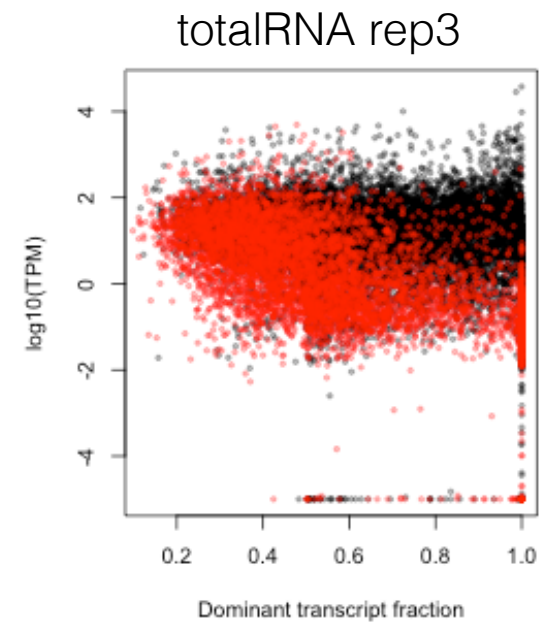
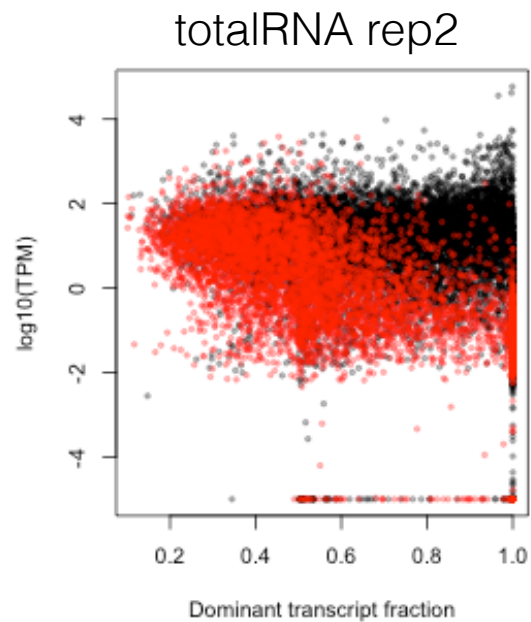
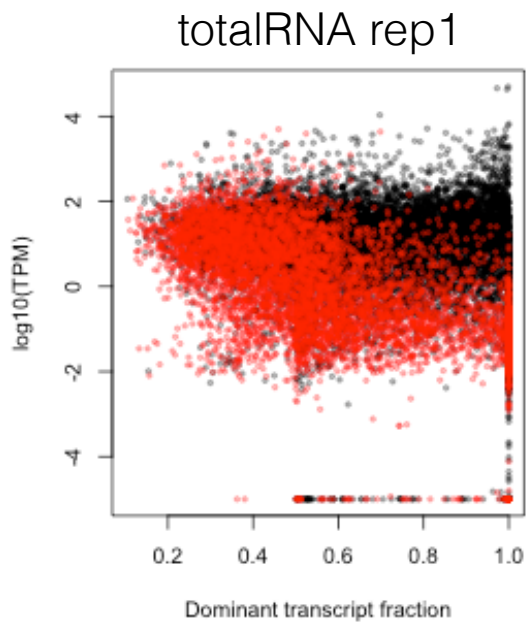
**ribosome footprinting**  
nuclease digestion leaves ribosome-protected fragments



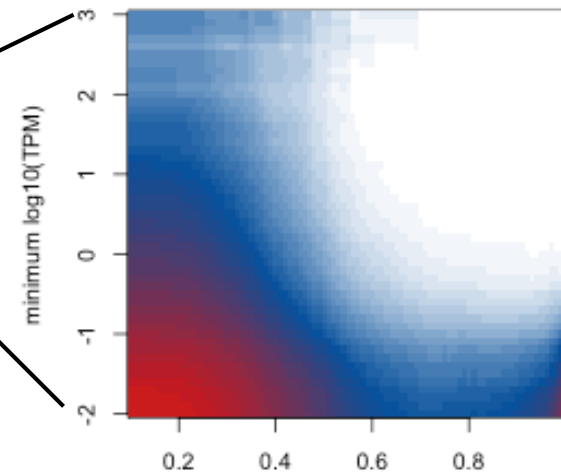
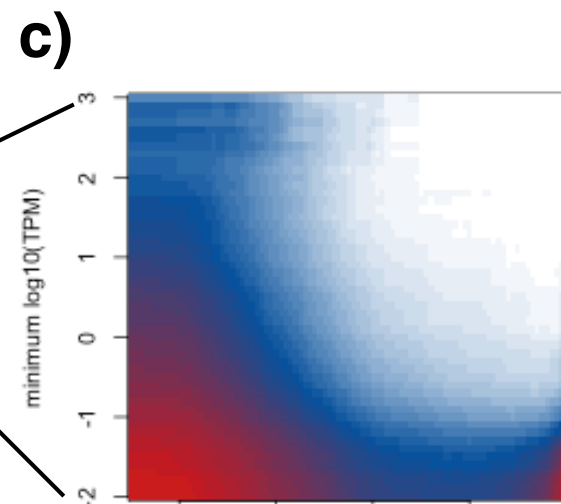
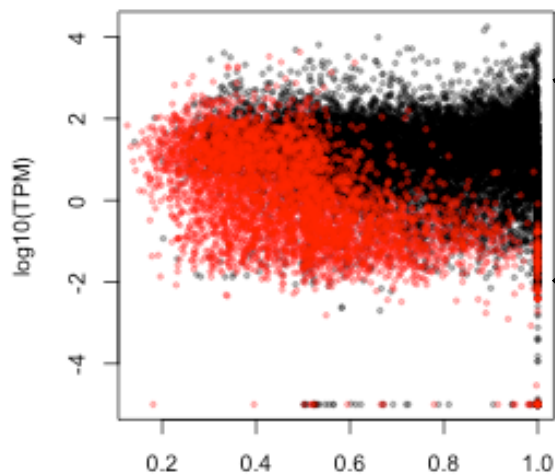
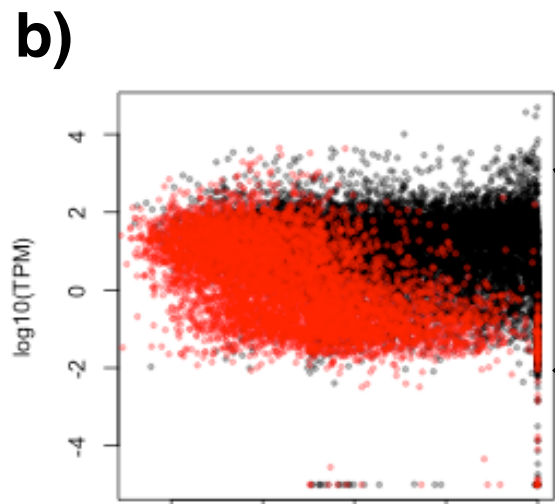
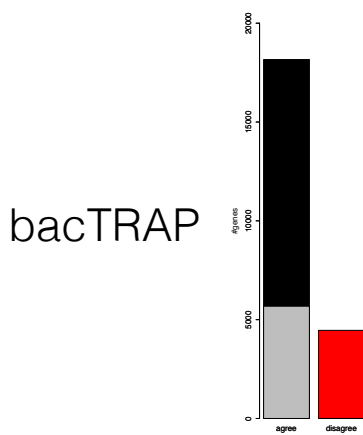
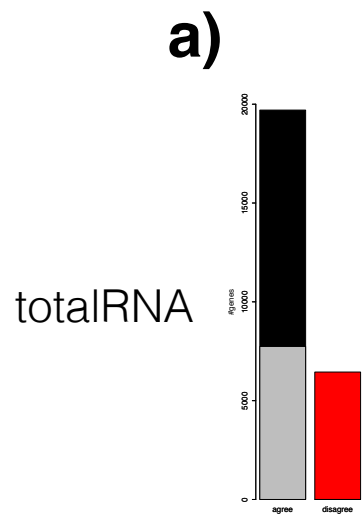
# b) experiment design

experiment type	experiment target
total-RNA	
eGFP-L10a IP	
Y10B IP	
ribosome footprinting MS/MS proteomics	





individual replicates show high degree of consistency in dominant isoform calls by fraction and by expression



**replicates agree on dominant isoform**  
**replicates disagree on dominant isoform**

