RESEARCH



Open Access

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

Mar Gonzàlez-Porta¹, Adam Frankish², Johan Rung¹, Jennifer Harrow² and Alvis Brazma^{1*}

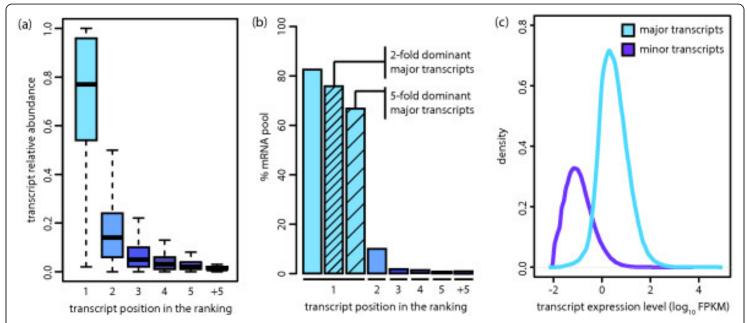


Figure 1 Most protein coding genes express one predominant transcript. (*a*) Relative abundance of the subset of transcripts in each position of the ranking for the primary tissues dataset. For each gene, transcripts were ranked based on their relative abundances. There is generally one predominant transcript 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.

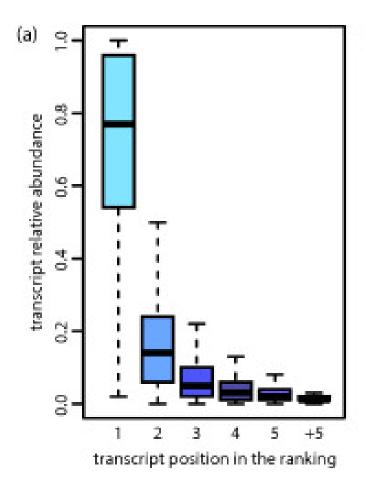


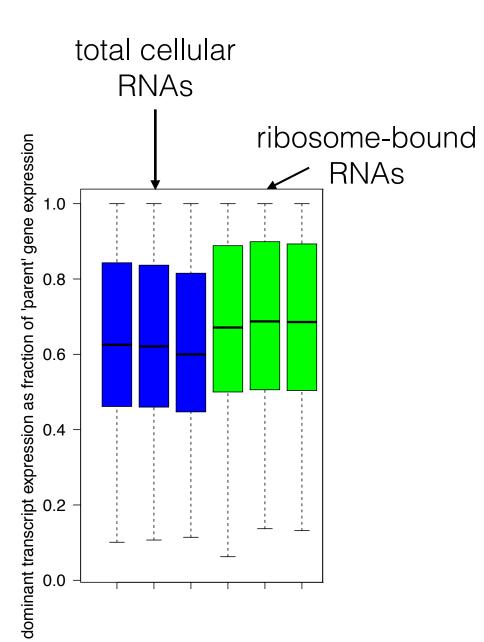
RESEARCH

Open Access

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

Mar Gonzàlez-Porta¹, Adam Frankish², Johan Rung¹, Jennifer Harrow² and Alvis Brazma^{1*}





fucosidase, alpha-L-2, plasma

ENSG0000001036.8



		YEDFGPLFTAK	FFNANQWADIFQASGAK	WPTSGQLFLGHPK	HGGFYTCSDR
001	ENST0000002165.5	1	1	1	1
002	ENST00000367585.1	1	1	0	0
201	ENST00000438118.2	1	1	1	0
003	ENST00000451668.1	0	0	1	0

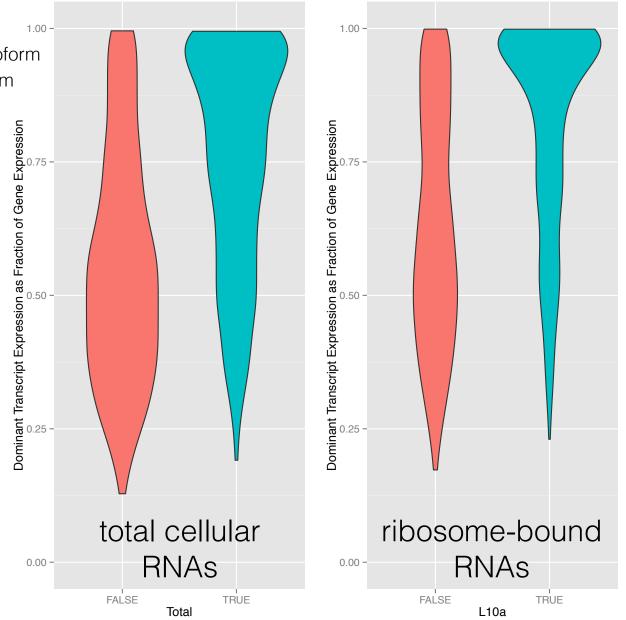
- 4 transcripts, 4 peptides
- 'best' transcript (ENST0000002165.5) supports 100% of these spectra

- 5,522 genes observed in both RNA-seq and MSMS, use the 4,839 of these that have >1 isoform
- only consider (for now) the 2,328 genes for which the MSMS suggest a single dominant isoform (926 genes peptides support only 1 isoform!)
- ~50% of these genes have same dominant isoform reported by both RNA-seq and MSMS
 - Total: 1,141 genes agree with MSMS
 - L10a: 1,158 genes agree with MSMS
 - Y10B: 1,159 genes agree with MSMS

total cellular ribosome-bound RNA-seq / proteomics disagree on major isoform RNA-seq / proteomics agree on major isoform **RNAs RNAs** 4 • og10(Gene Expression [summedTPMs] +1) log10(Gene Expression [summedTPMs] +1) $_{\infty}^{\circ}$ Agreement based on RNA-seq gene expression 1 -1 -0 -0 -FALSE TRUE FALSE TRUE Total L10a

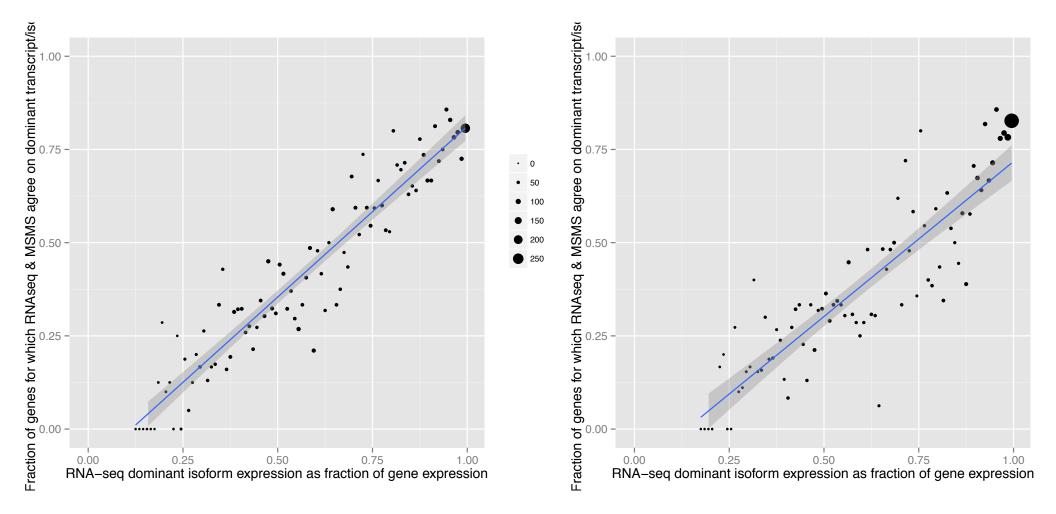
RNA-seq / proteomics **disagree** on major isoform RNA-seq / proteomics **agree** on major isoform

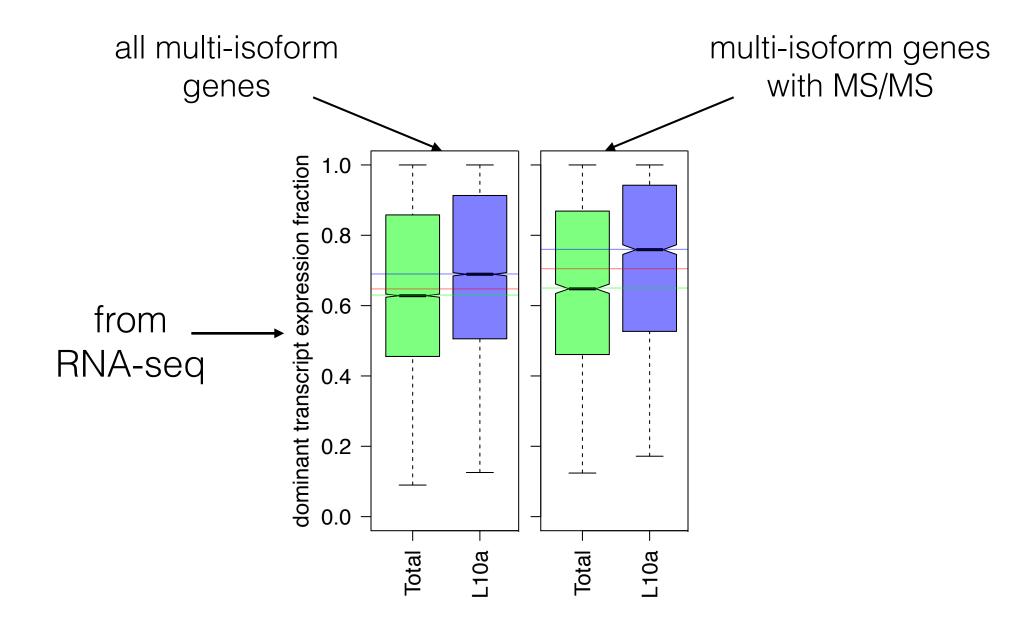
Agreement based on RNA-seq dominant isoform expression as a fraction of the gene's expression



total cellular RNAs

ribosome-bound RNAs





- dominant isoform becomes more dominant in ribosome-associated RNAs
- becomes even more dominant in genes for which we see peptides