

Analysis of *Saccharomyces bayanus* RNA-Seq data for the study of *de* *novo* transcripts

by Silvia González and Alejandro Madrid

Context of the study

Saccharomyces bayanus - transcriptomic data

Objectives

To identify transcripts in *Saccharomyces bayanus* that could have originated from previously non-coding genomic regions using RNA-Seq data

To analyse different characteristics of those putative novel transcripts



Blevins et al. *BMC Res Notes* (2019) 12:250
<https://doi.org/10.1186/s13104-019-4286-0>

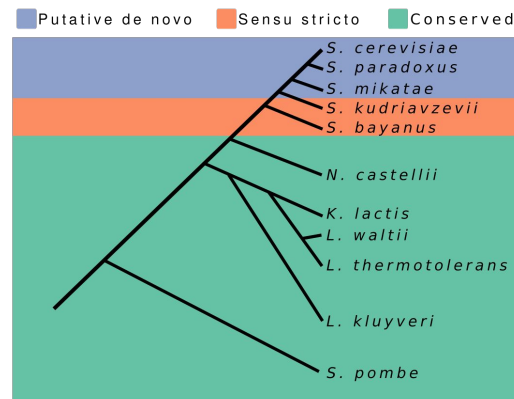
BMC Research Notes

DATA NOTE

Open Access

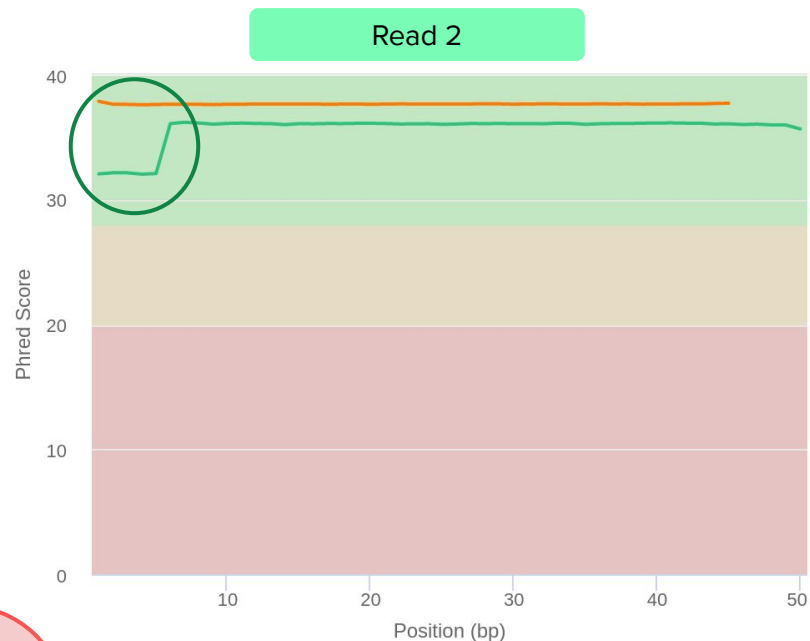
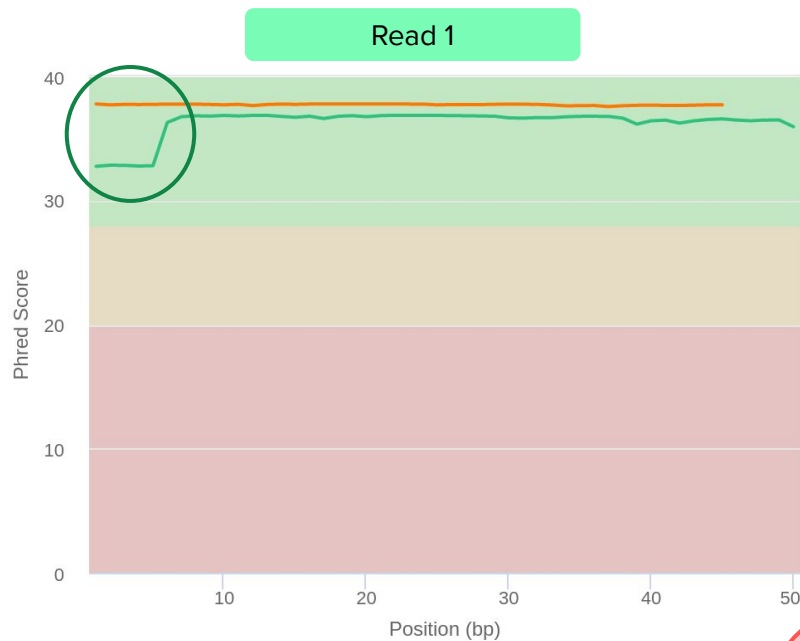
Transcriptomics data of 11 species of yeast identically grown in **rich media** and oxidative stress conditions

William R. Blevins^{1,2*}, Lucas B. Carey^{2,3} and M. Mar Albà^{1,4}



Transcript assembly and preliminary analysis

Quality of the sequencing reads

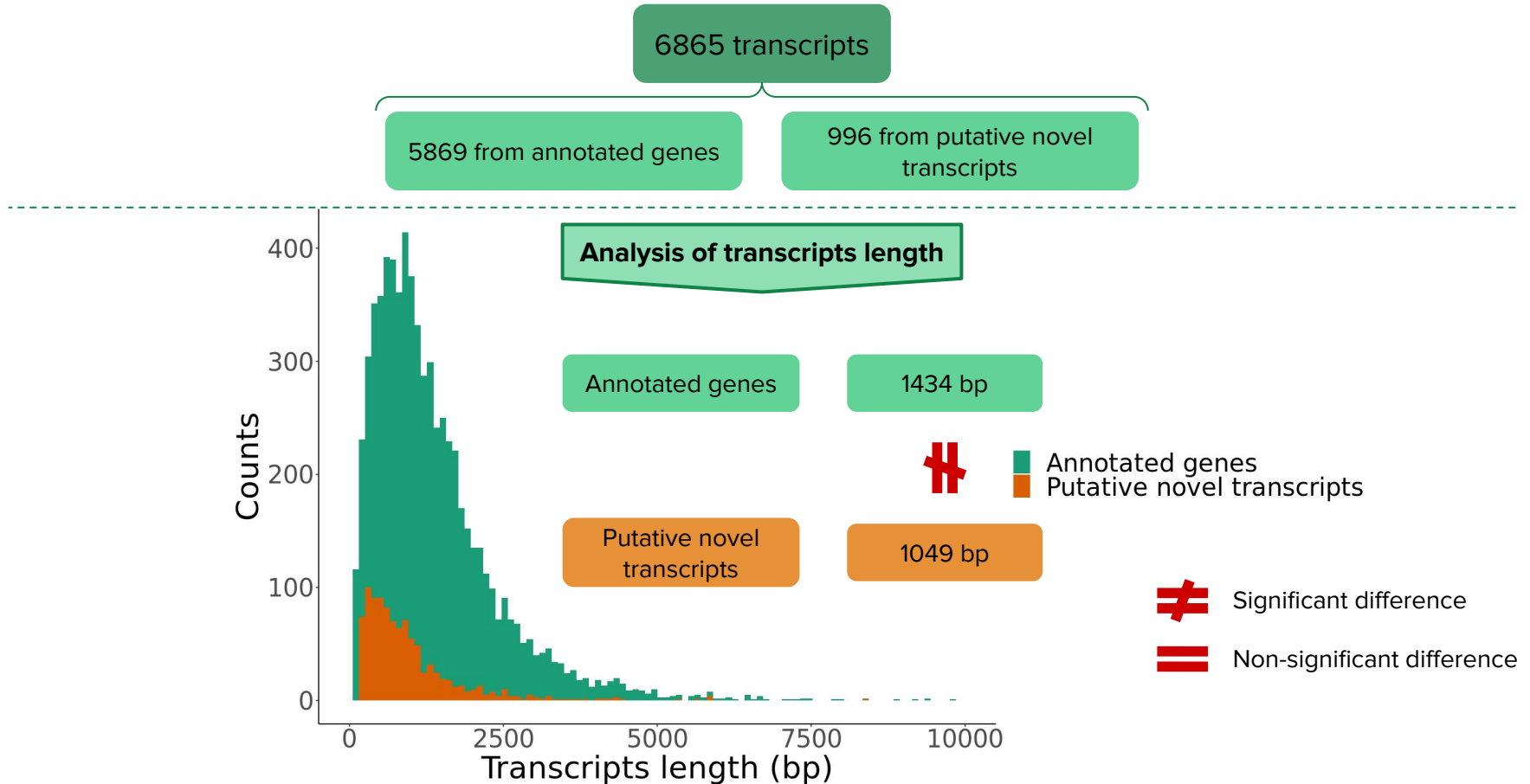


↓33%

Raw reads

Trimmed reads

Analysis of annotated vs putative novel transcripts



Analysis of longest ORFs lengths

Putative novel transcripts

156 bp \equiv 52 aa

1131 bp \equiv 377 aa

Annotated genes

Medians

- Annotated genes
- Putative novel transcripts
- Random sequences from annotated genes
- Random sequences from putative novel transcripts

Annotated genes
Putative novel transcripts

Randomized sequences

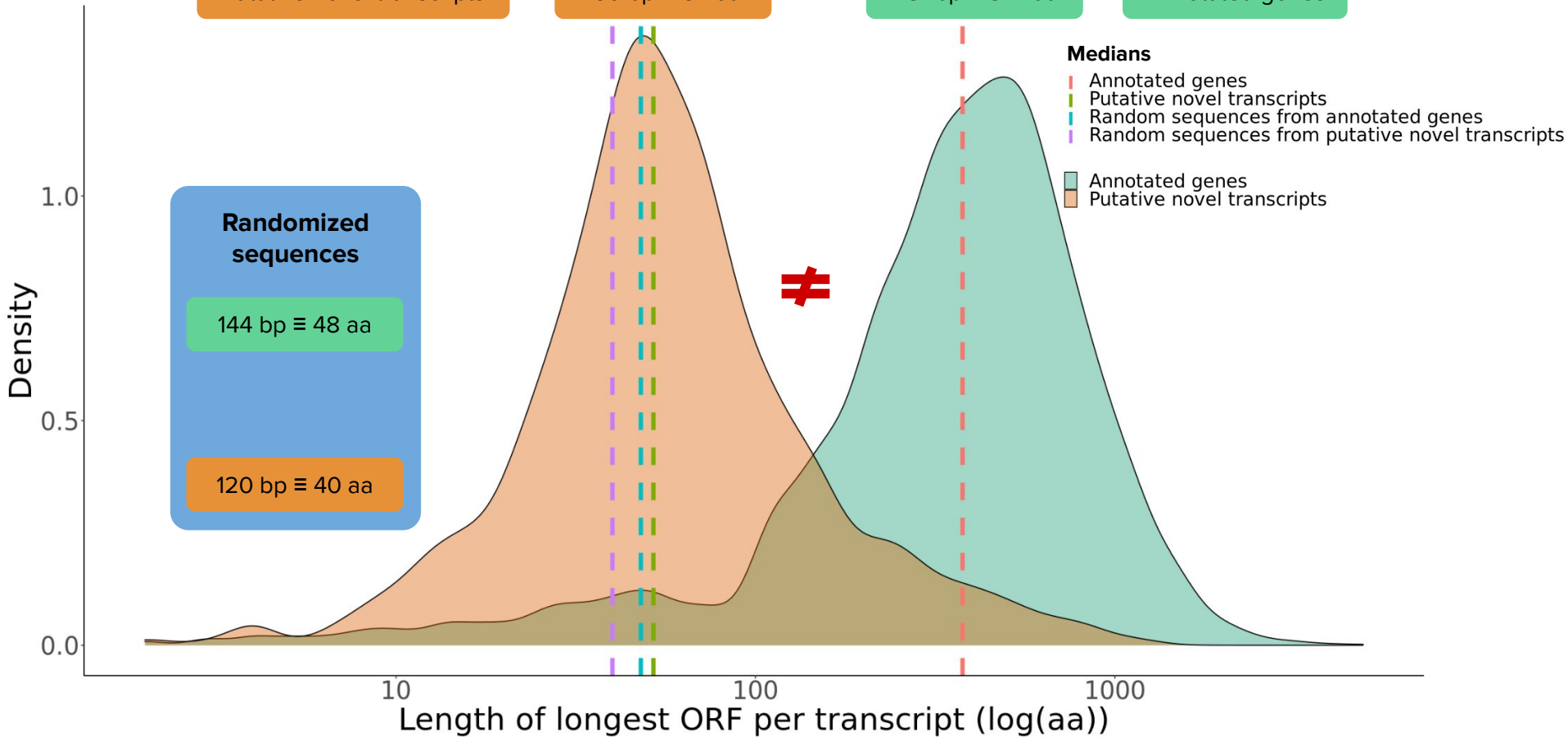
144 bp \equiv 48 aa

120 bp \equiv 40 aa

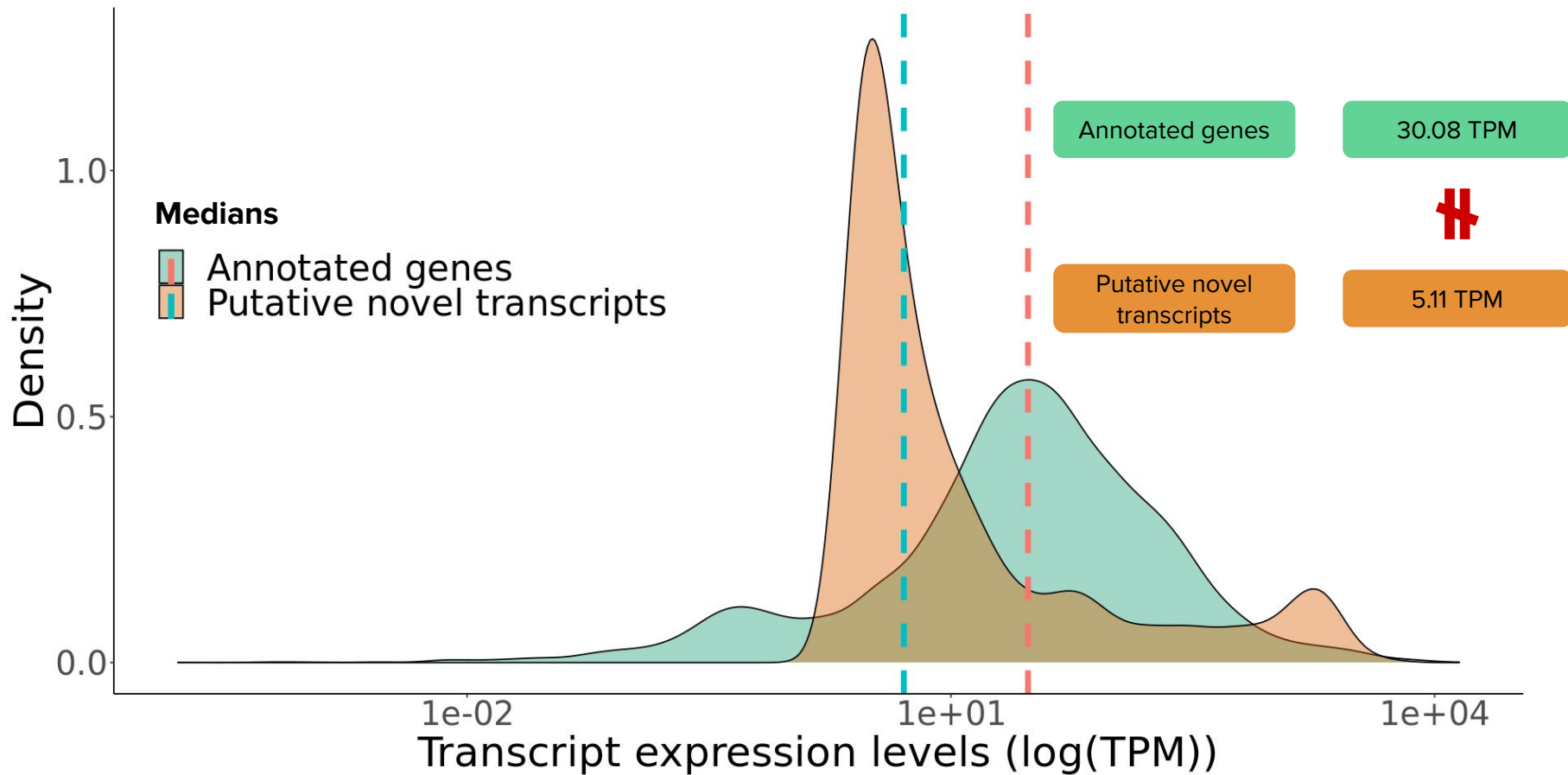
Density

\neq

Length of longest ORF per transcript (log(aa))

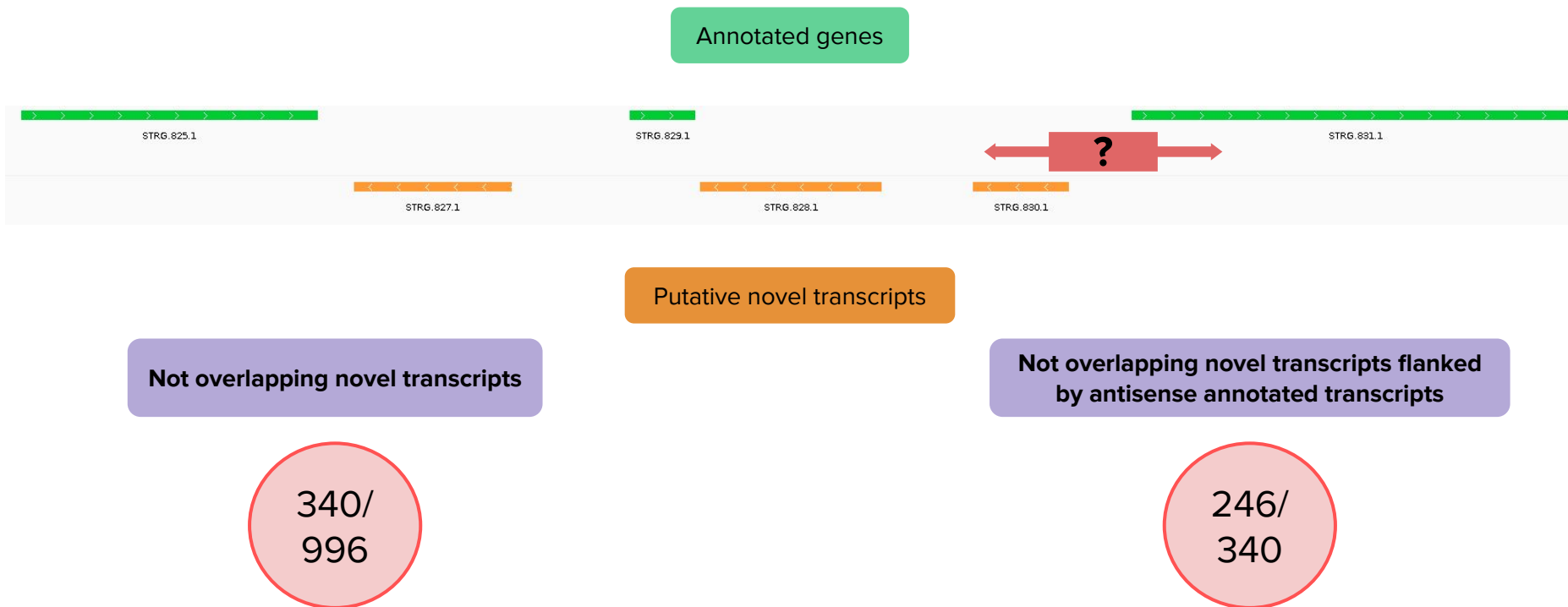


Analysis of expression levels



Further data analysis

Study of **putative novel** transcripts **not overlapping** annotated genes



Study of **putative novel** transcripts **overlapping** annotated genes in **antisense**

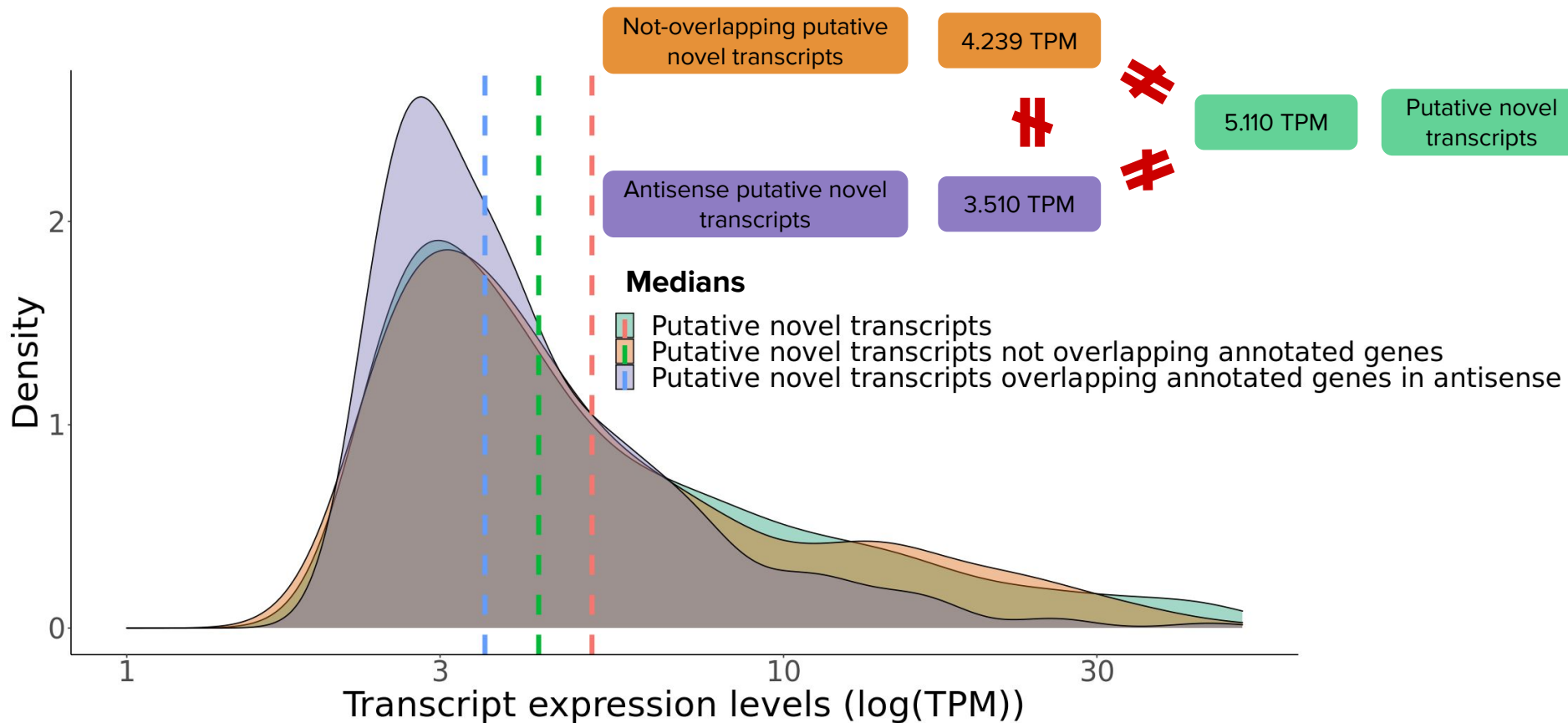
Annotated genes



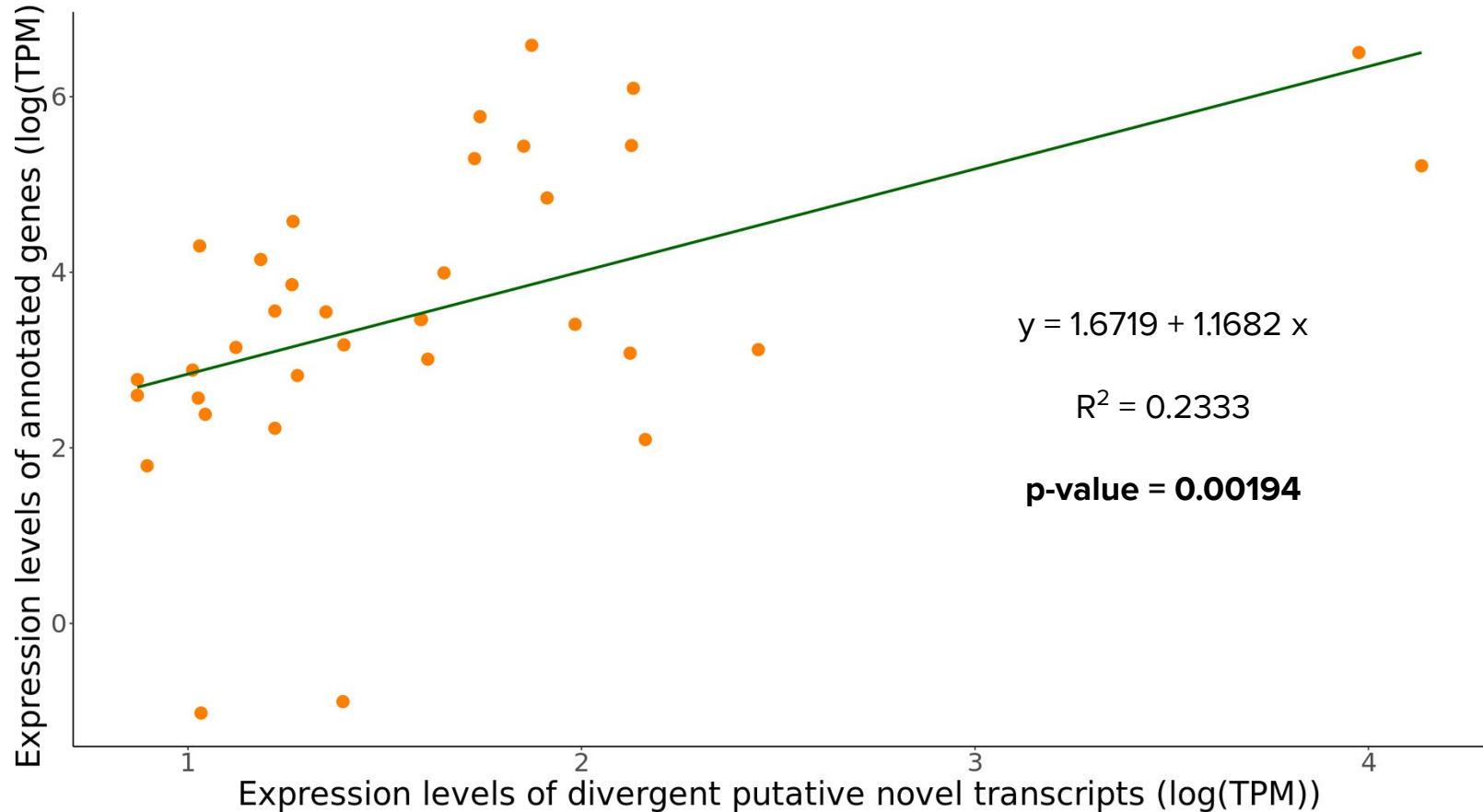
Putative novel transcripts

289/
996

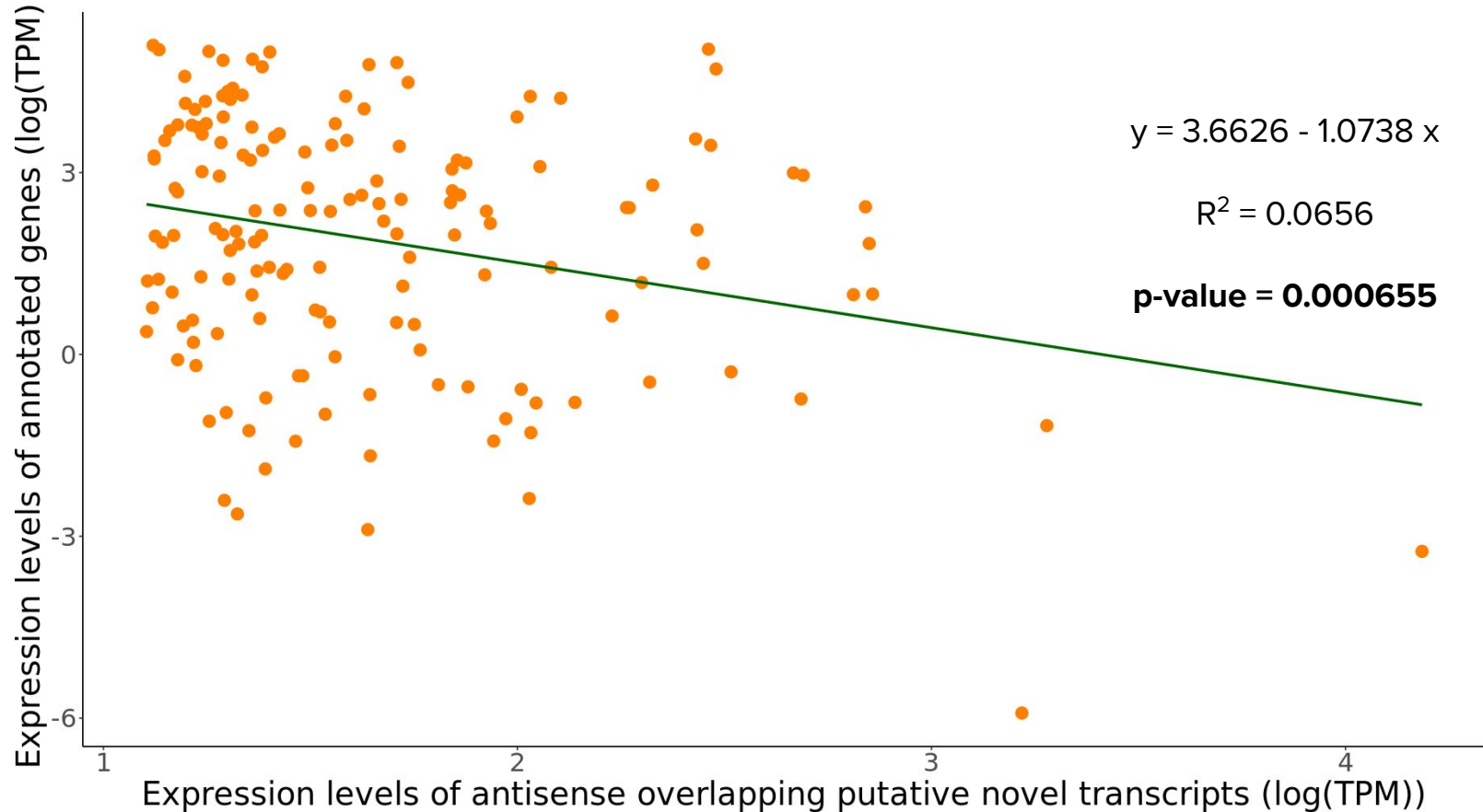
Analysis of expression levels



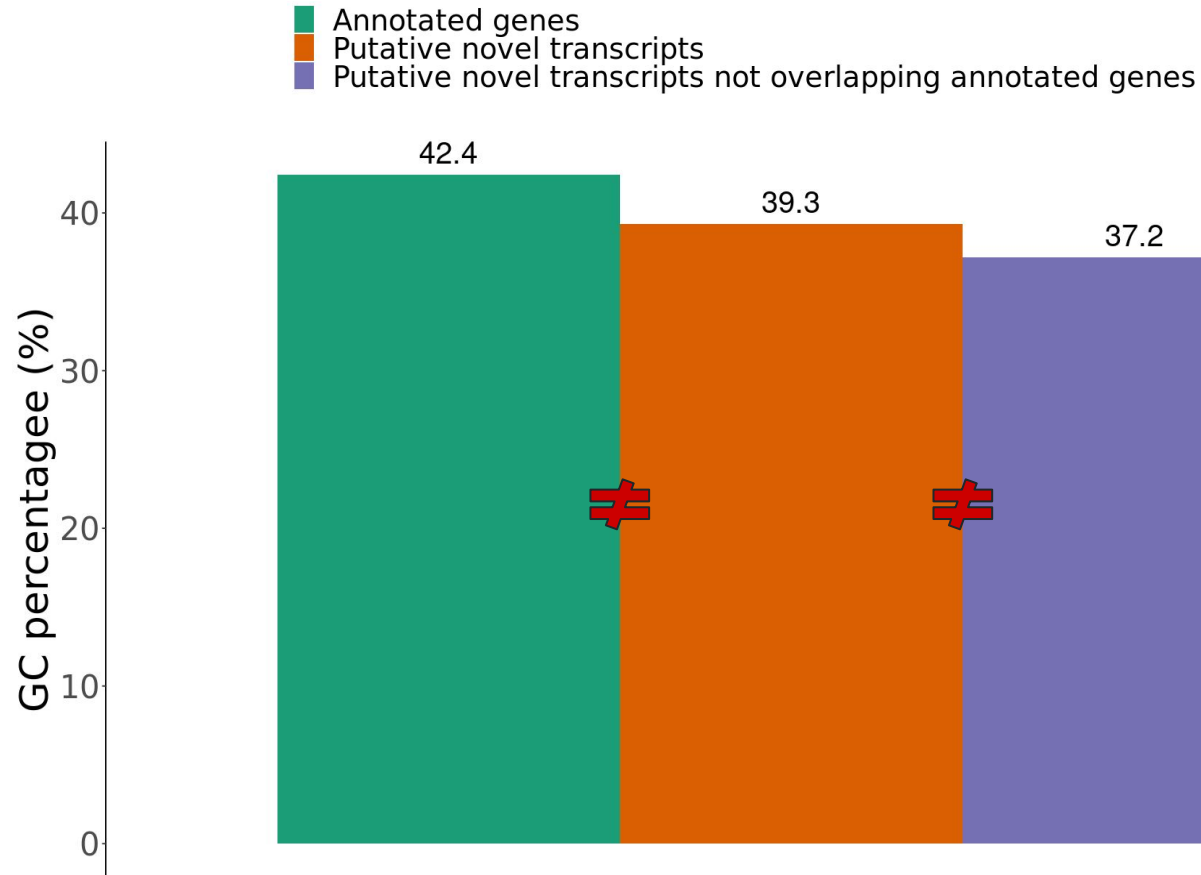
Comparison of expression levels of **annotated** genes and the **putative novel** transcripts **divergent** to them



Comparison of expression levels of **annotated** genes and the **putative novel** transcripts **overlapping** them in **antisense**



Analysis of GC content in transcripts



Analysis of transcripts length

Not-overlapping putative
novel transcripts

483 bp

Antisense putative novel
transcripts

655 bp

740 bp

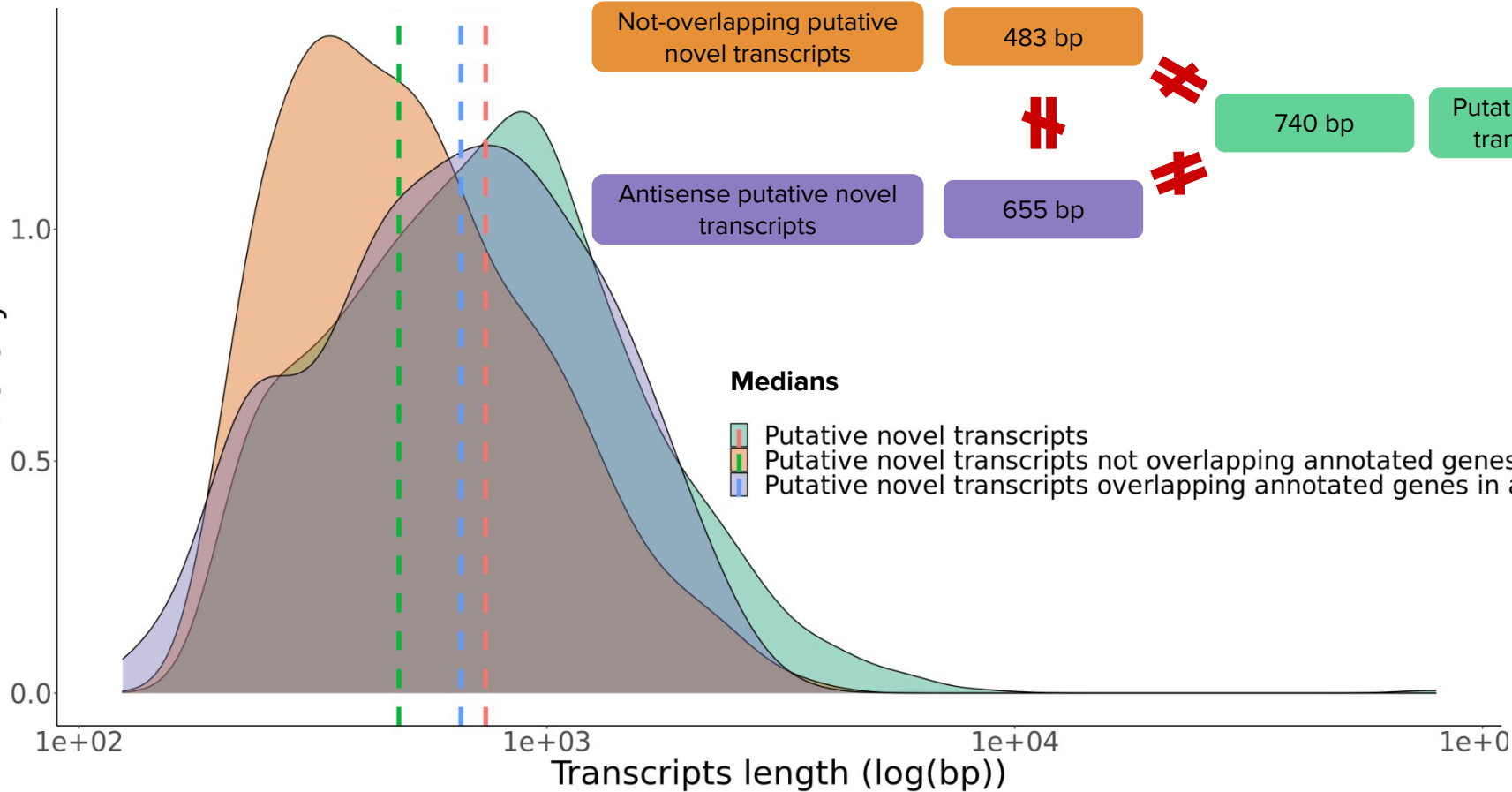
Putative novel
transcripts



Medians

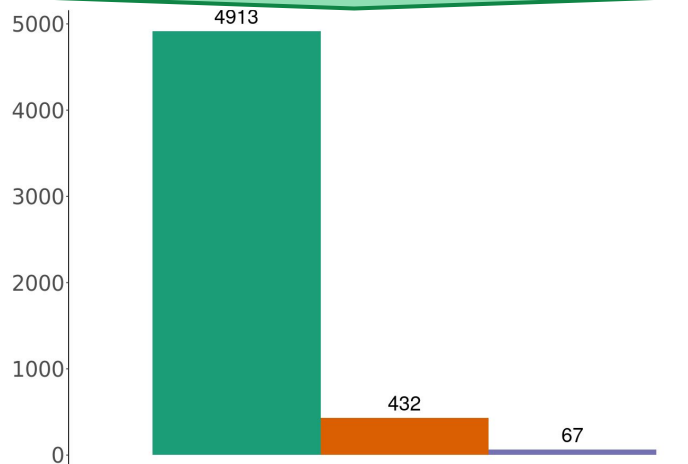
- Putative novel transcripts
- Putative novel transcripts not overlapping annotated genes
- Putative novel transcripts overlapping annotated genes in antisense

Density

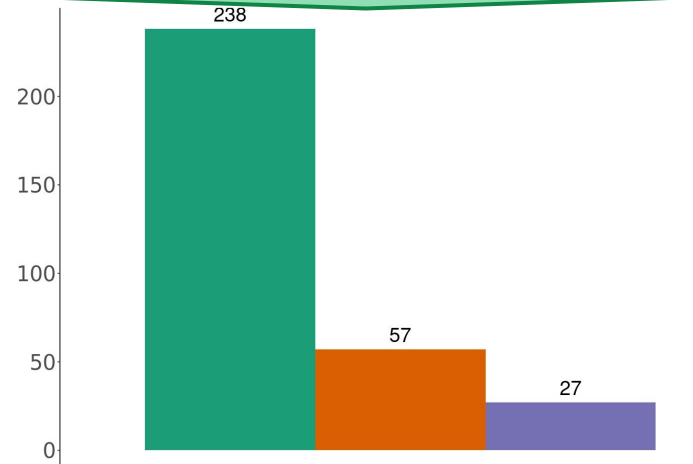


Number of transcripts

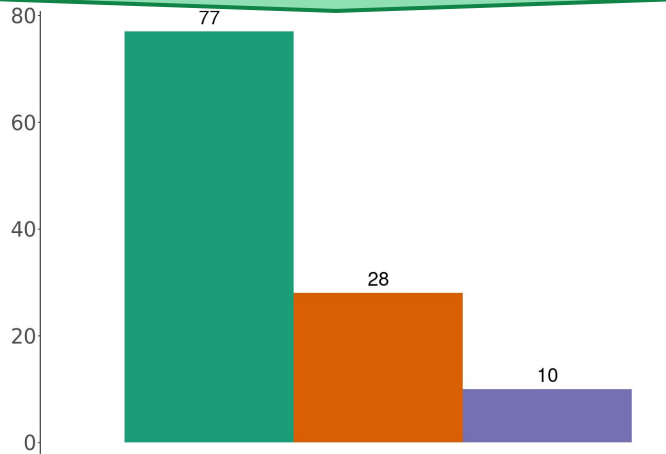
Number of ORFs per annotated transcript



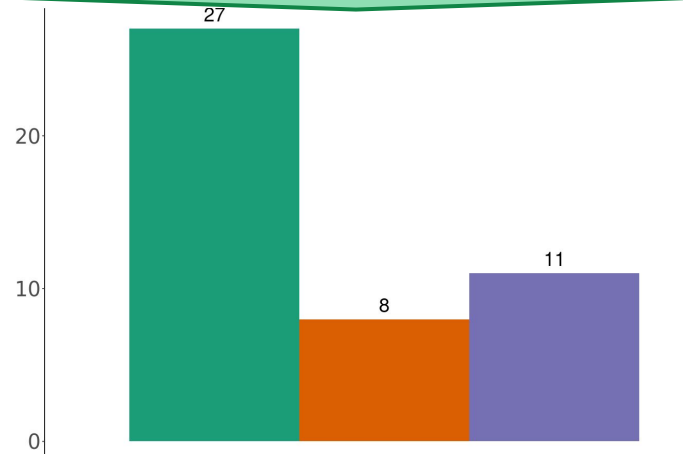
Number of ORFs per putative novel transcript



Number of ORFs per putative novel transcript overlapping in antisense an annotated gene



Number of ORFs per putative novel transcript not overlapping annotated genes



1 ORF
2 ORFs
3 or more ORFs

Analysis of longest ORFs lengths

Not-overlapping putative novel transcripts

132 bp \equiv 44 aa

177 bp \equiv 59 aa

Antisense putative novel transcripts



261 bp \equiv 87 aa

261 bp \equiv 87 aa

Putative novel transcripts

Randomized sequences

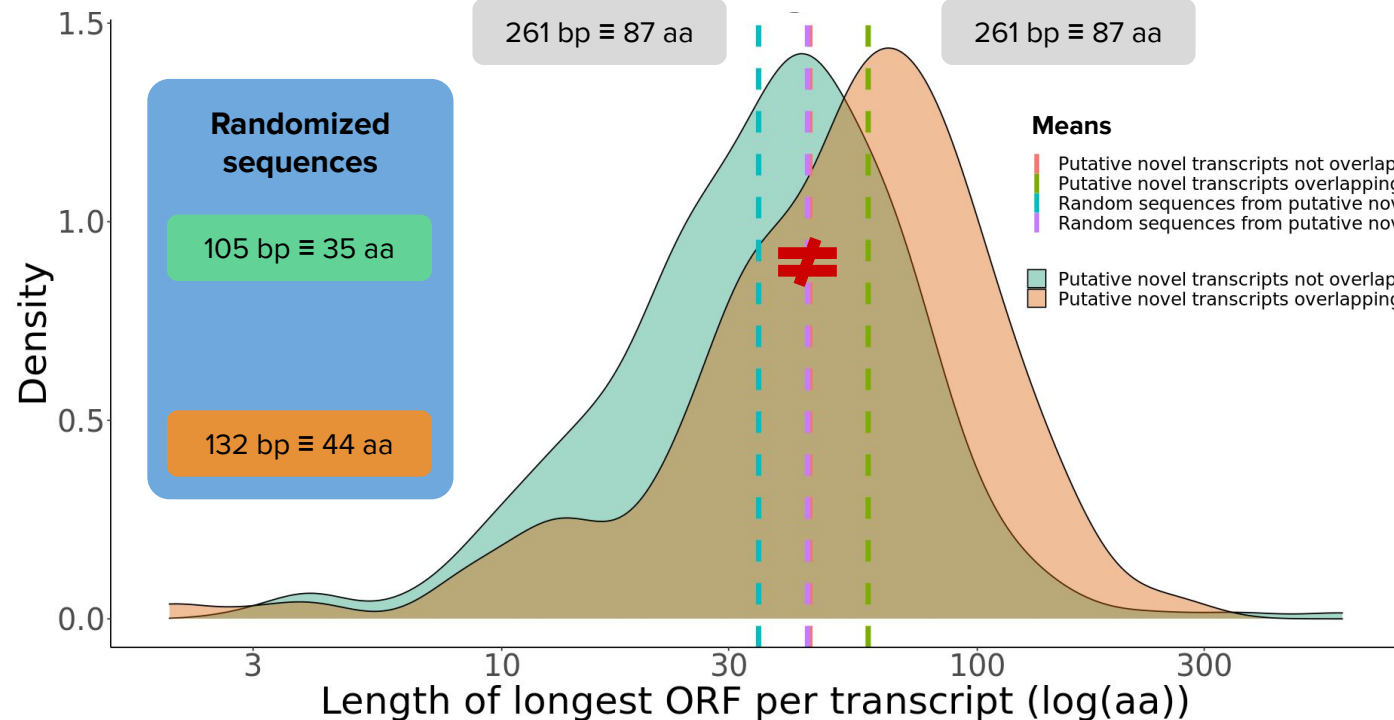
105 bp \equiv 35 aa

132 bp \equiv 44 aa

Means

- Putative novel transcripts not overlapping annotated genes
- Putative novel transcripts overlapping annotated genes in antisense
- Random sequences from putative novel transcripts not overlapping annotated genes
- Random sequences from putative novel transcripts overlapping annotated genes in antisense

- Putative novel transcripts not overlapping annotated genes
- Putative novel transcripts overlapping annotated genes in antisense



BLASTx against *Saccharomyces cerevisiae* proteome - GO enrichment

Biological process

Putative novel transcripts

Hexose transmembrane transport
Nucleotide transmembrane transport
Peptidyl-tyrosine phosphorylation
Cell wall integrity MAPK cascade
...

576/996, 3076 total
matches

Biological process

Not overlapping putative novel transcripts

Phenylalanine transport
Carboxylic acid transport
Electron transport coupled proton transport
Tryptophan transport
...

26/340, 347 total matches

Biological process

Antisense putative novel transcripts

Hexose transmembrane transport
Hexitol metabolic process
Arabinose catabolic process
Peptidyl-tyrosine phosphorylation
...

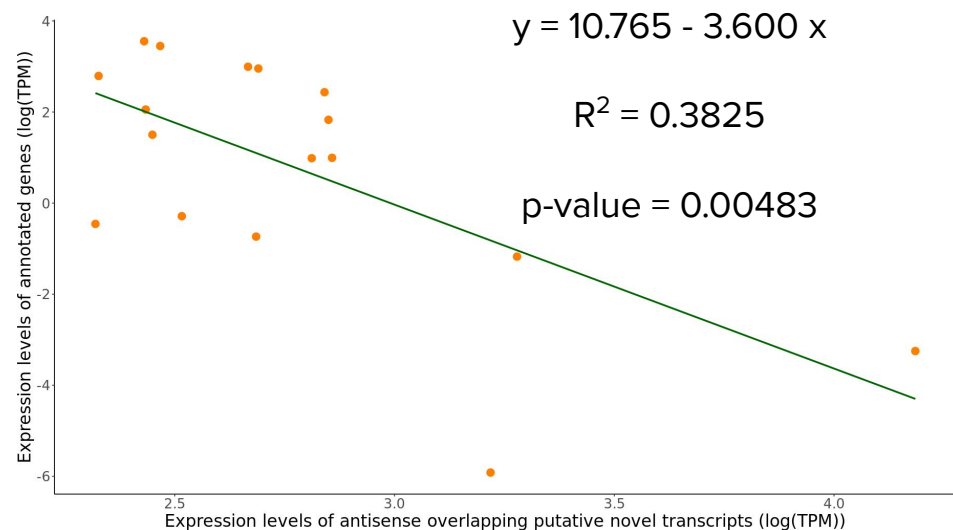
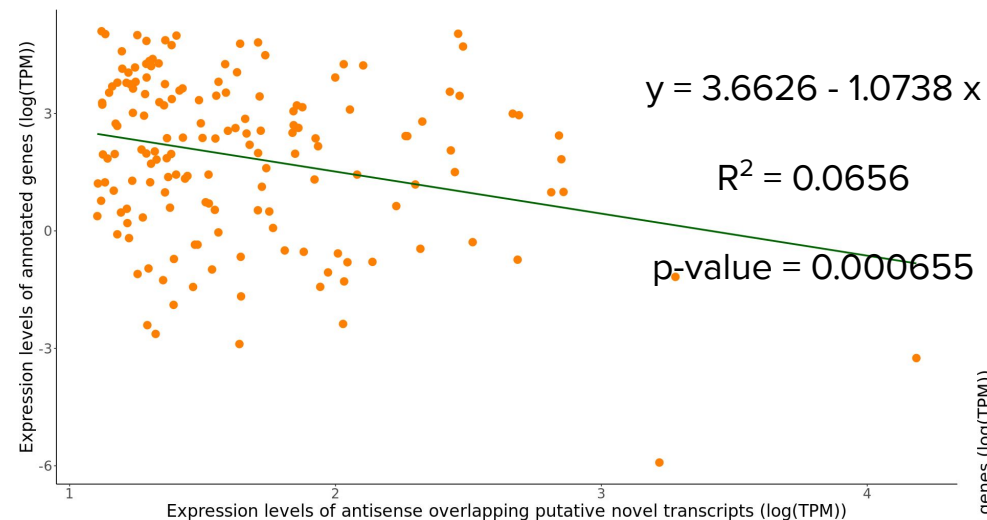
(-) strand: 263/289, 1136
total matches

(+) strand: 0/289, 0 total
matches

Conclusions

1. A significant proportion of **putative novel transcripts** seem to derive from **divergent promoters**.
 2. Novel transcripts **overlapping annotated** genes in **antisense** might have a regulatory function.
 3. Novel transcripts, and particularly those **not overlapping** annotated ones, are notably **enriched in ORFs** compared to known genes.
-

Comparison of expression levels of **annotated genes** and the **putative novel** transcripts **overlapping** them in **antisense**



BLASTn against *Saccharomyces cerevisiae* genome

Putative novel transcripts

379/996, 928 total matches

Not-overlapping putative novel transcripts

85/340, 270 total matches

Antisense putative novel transcripts

85/289, 95 total matches

BLASTx against *Saccharomyces cerevisiae* proteome - GO enrichment

Biological process

Hexose transmembrane transport
Nucleotide transmembrane transport
Peptidyl-tyrosine phosphorylation
Cell wall integrity MAPK cascade
...

Molecular function

Ubiquitin conjugating enzyme activity
Hexose transmembrane transporter activity
Carbohydrate:proton symporter activity
Oxidoreductase activity, acting on metal ions, NAD or NADP as acceptor
...

Cellular component

Cytosolic large ribosomal subunit
Cytosolic small ribosomal subunit
Preribosome
Plasma membrane
...

Detection of SNPs on the analysed sequences

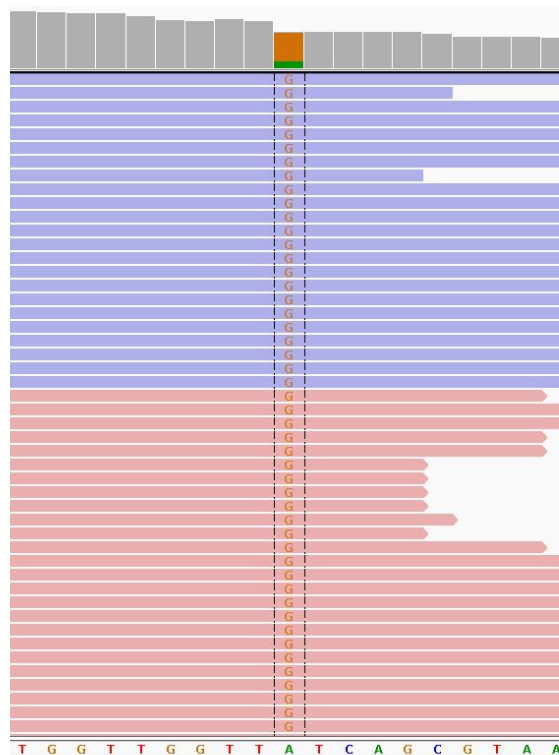
Forward

Reverse

Heterozygous SNPs



Homozygous SNPs



Strand bias = false positive

