
RNA-Seq Analysis on *Saccharomyces paradoxus*

Eric Kramer Rosado
Joan Pau Cebrià Costa
Dylan Dalton Martínez

Principles of Genome Bioinformatics

Màster Universitari en Bioinformàtica per a les Ciències de la Salut

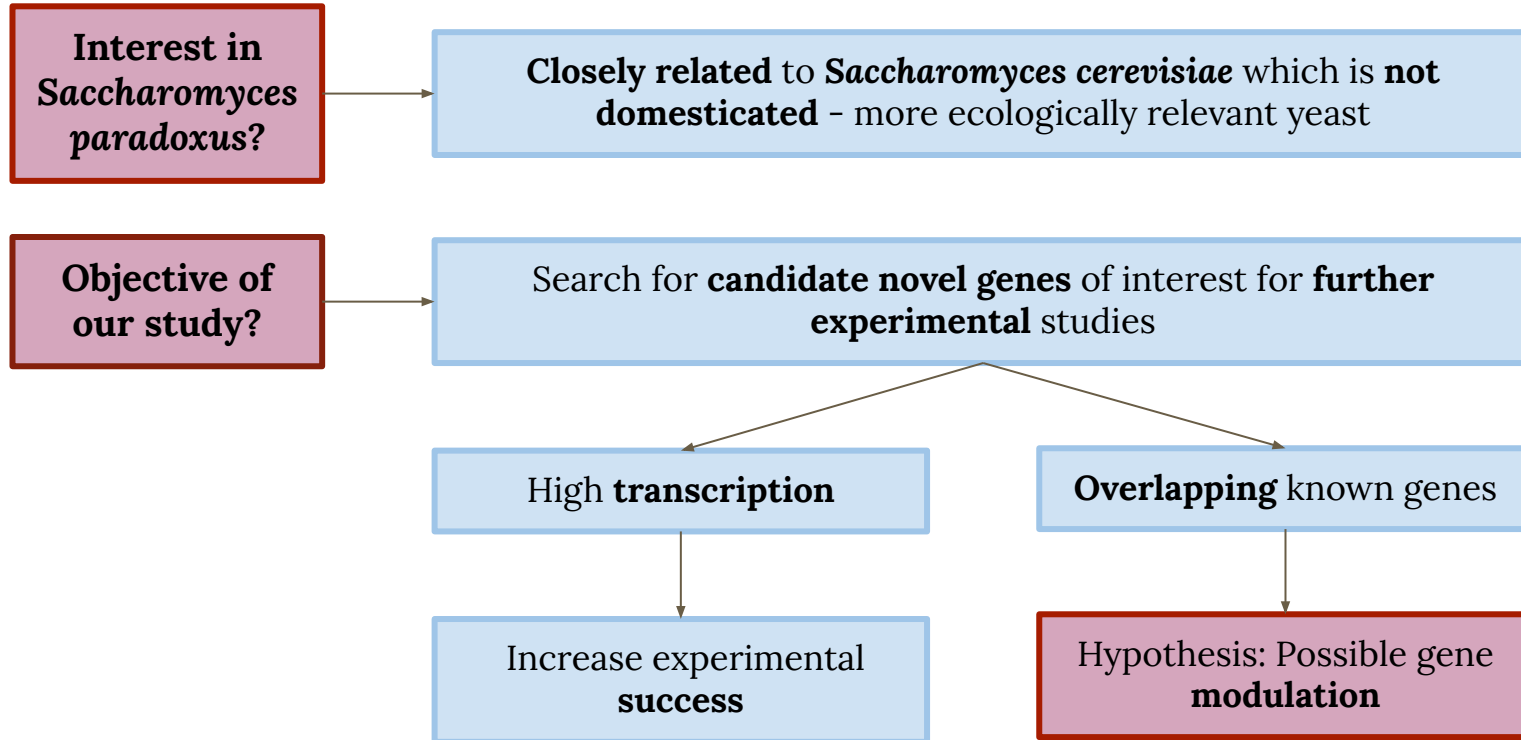


Contents

1. Introduction
2. Material & Methods
3. Results
4. Discussion

5 μm

Introduction



Introduction

Hypothesis:
possible gene
modulation

There is **increasing evidence** for **translation** of both **sense and antisense alternate frame ORFs**. This evidence is **generally ignored** and is usually presumed **as non-functional** although we argue this inference is made too quickly. (Ardern et al., 2020)

Europe PMC Funders Group
Author Manuscript
Essays Biochem. Author manuscript; available in PMC 2015 January 06.
Published in final edited form as:
Essays Biochem. 2013; 54: 91–101. doi:10.1042/bnc0540091.

The functions of natural antisense transcripts

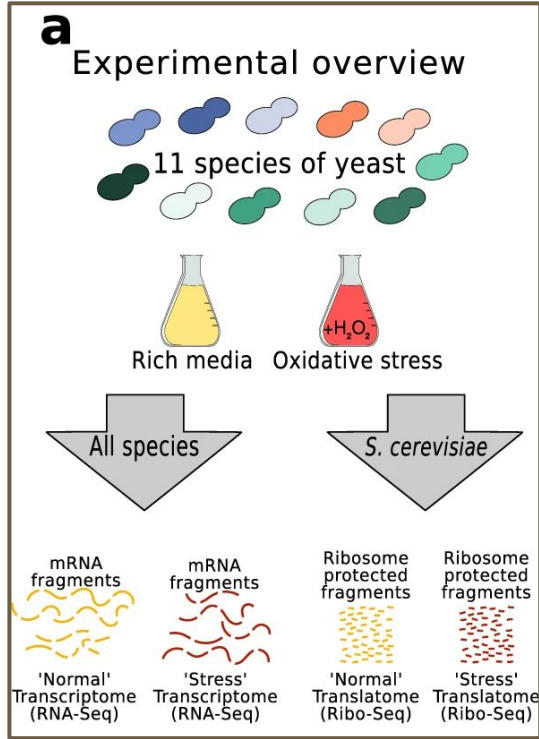
Megan Wight and Andreas Werner¹
Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle
NE2 4HH, U.K.

We here propose a **possible functionality** in that of **modulating gene expression**, a mechanism previously described but never in *S paradoxus* (Wight & Werner, 2013)

Search and study **Novel transcripts** with **High Expression** (TPM)
with **antisense** overlapping

(Wight & Werner, 2013)

Material & Methods: Workflow



(Blevins et al., 2021)



**RAW RNA-SEQ
DATA**

FASTQ-C

Analyze read quality

Trimmomatic

Remove adapters & low q. reads

Hisat-2

Align to Reference Genome

Stringtie

Assemble transcripts using mapped reads

**Novel
Transcripts**

**Known
Transcripts**

Material & Methods: Workflow

Novel transcripts

1st Filter:
Overlapping /
antisense

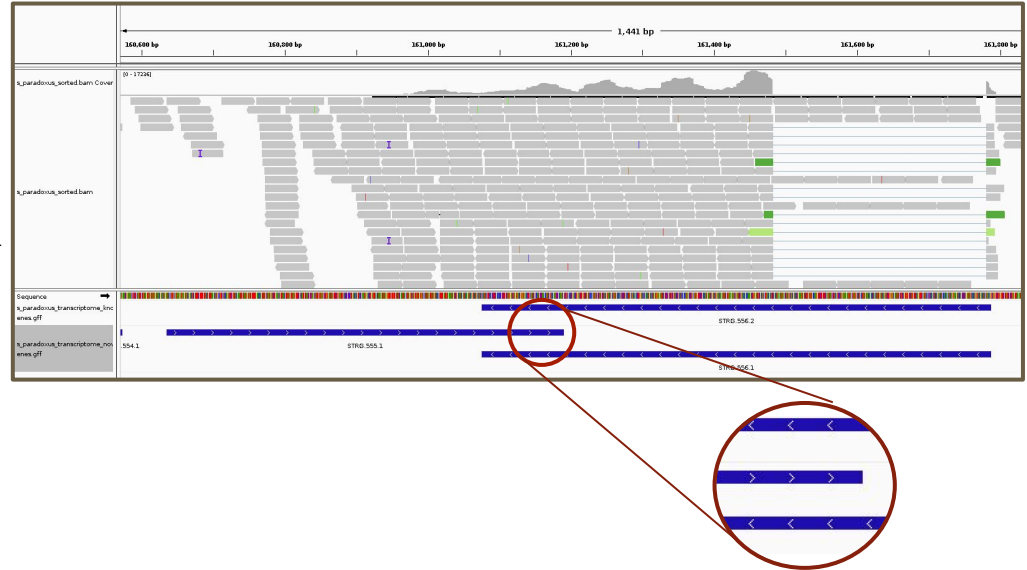
Selection of genes overlapping antisense with previous described ones

2nd Filter:
TPM expression

Selection of those transcripts upper the 3rd quartile

BLAST search

Use *S.cerevisiae* proteins as database



Material & Methods: Workflow

**Novel
Transcripts**

1st Filter:
Overlapping /
antisense

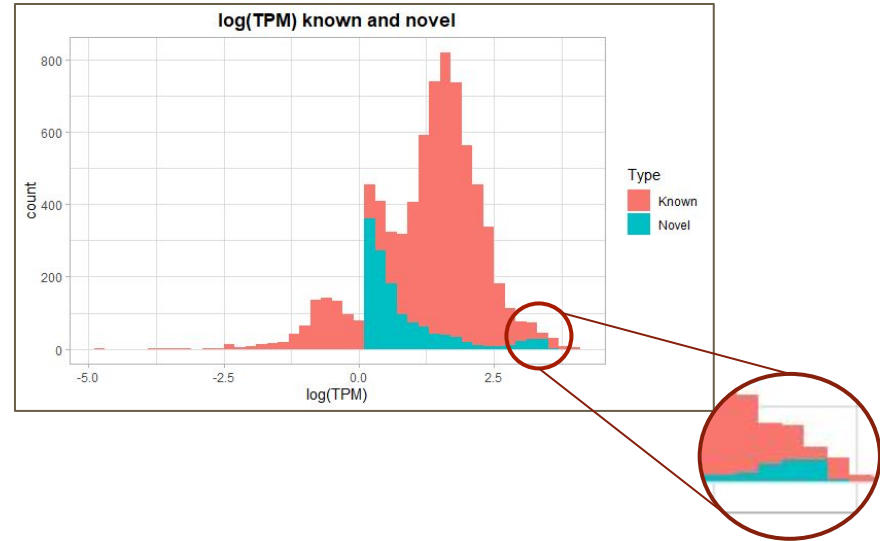
Selection of genes overlapping antisense with
previous described ones

2nd Filter:
TPM expression

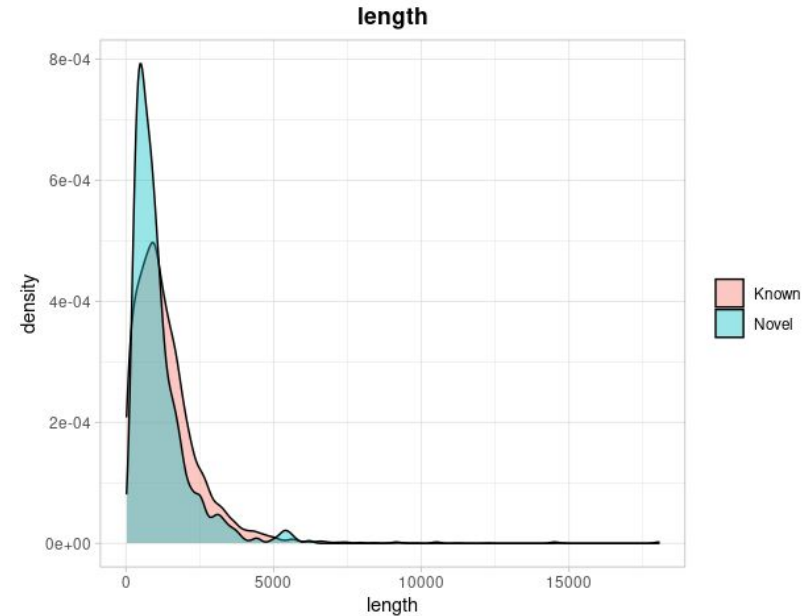
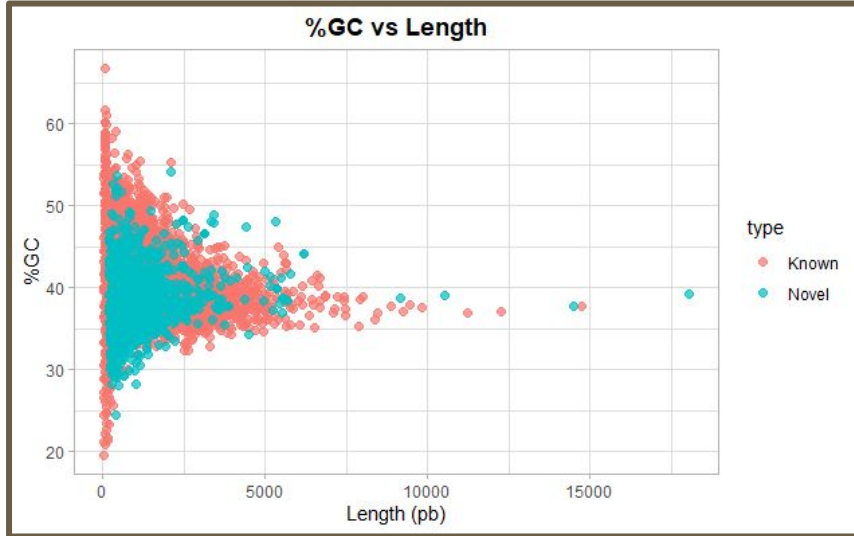
Selection of those transcripts upper the 3rd quartile

BLAST search

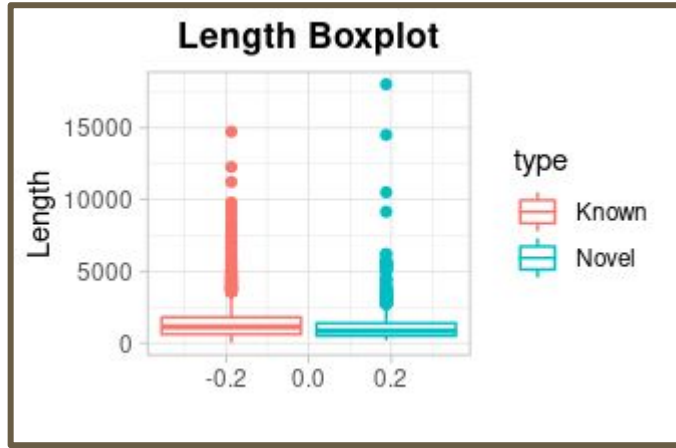
Use *S.cerevisiae* proteins as database



Results: %GC Content & Length

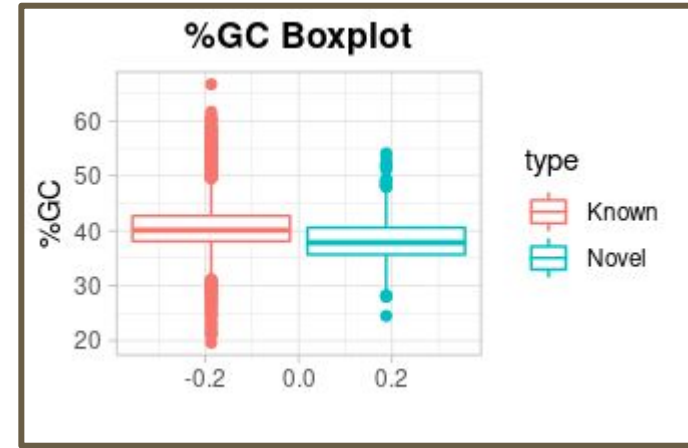


Results: %GC Content & Length



W = 4611844, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

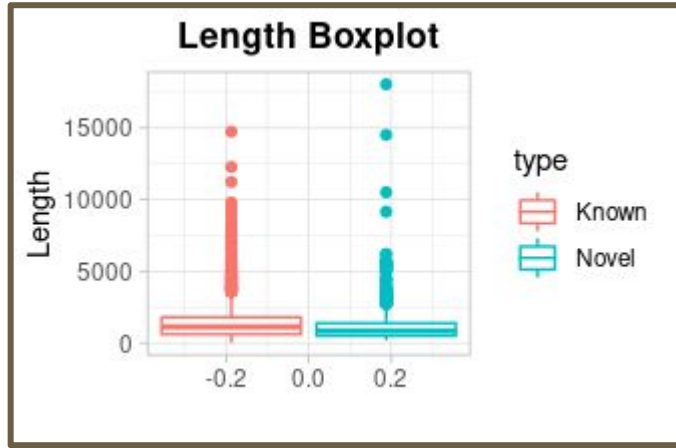
H0: $\mu = \mu$; Length means equal
H1: $\mu \neq \mu$; Length means different



W = 5417348, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

H0: $\mu = \mu$; GC% means equal
H1: $\mu \neq \mu$; GC% means different

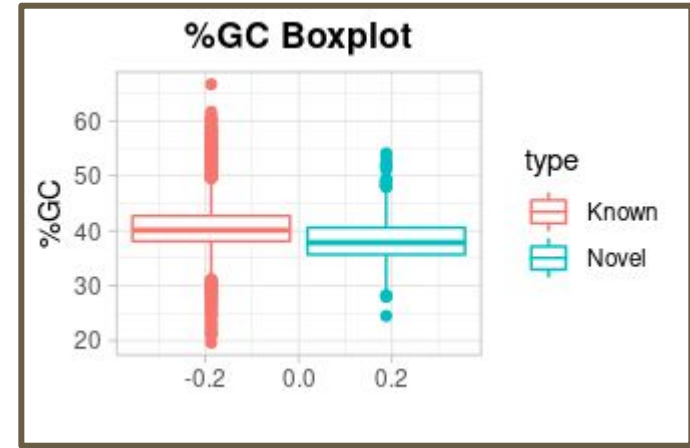
Results: %GC Content & Length



W = 4611844, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

~~H0: $\mu = \mu$; Length means equal~~
H1: $\mu \neq \mu$; Length means different

Significant!



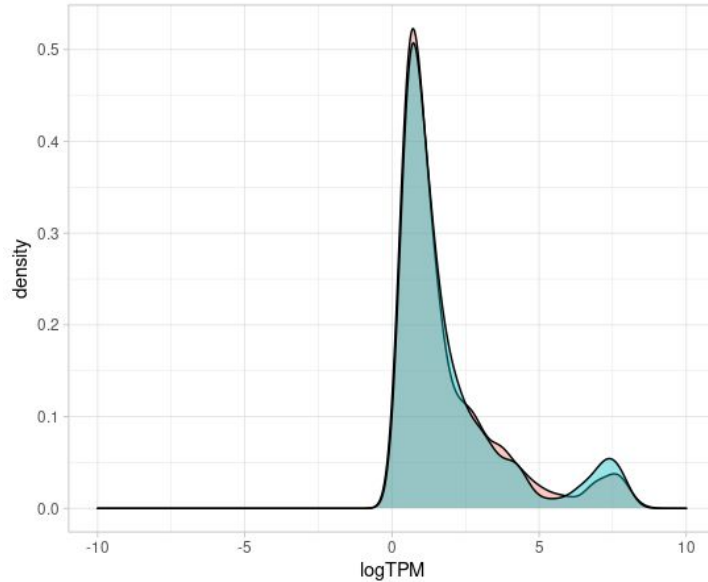
W = 5417348, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

~~H0: $\mu = \mu$; GC% means equal~~
H1: $\mu \neq \mu$; GC% means different

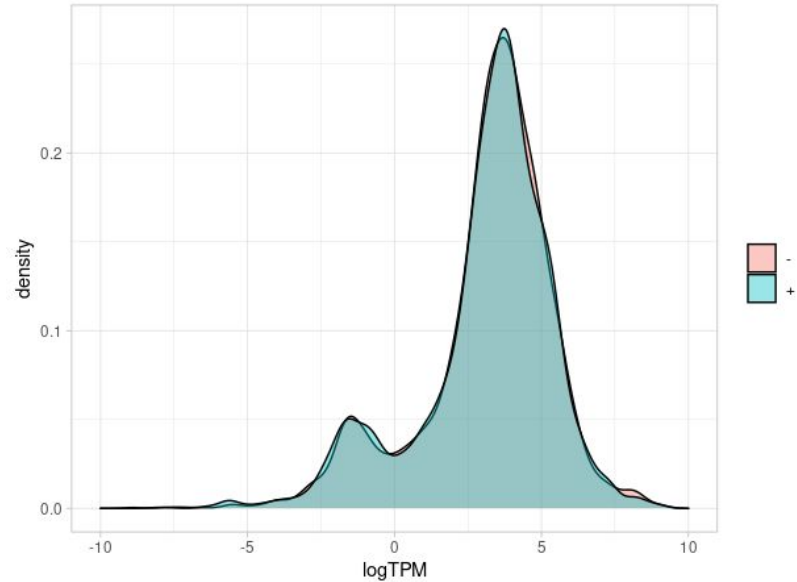
Significant!

Results: Transcripts Per Million (TPM)

No difference between sense and antisense transcription

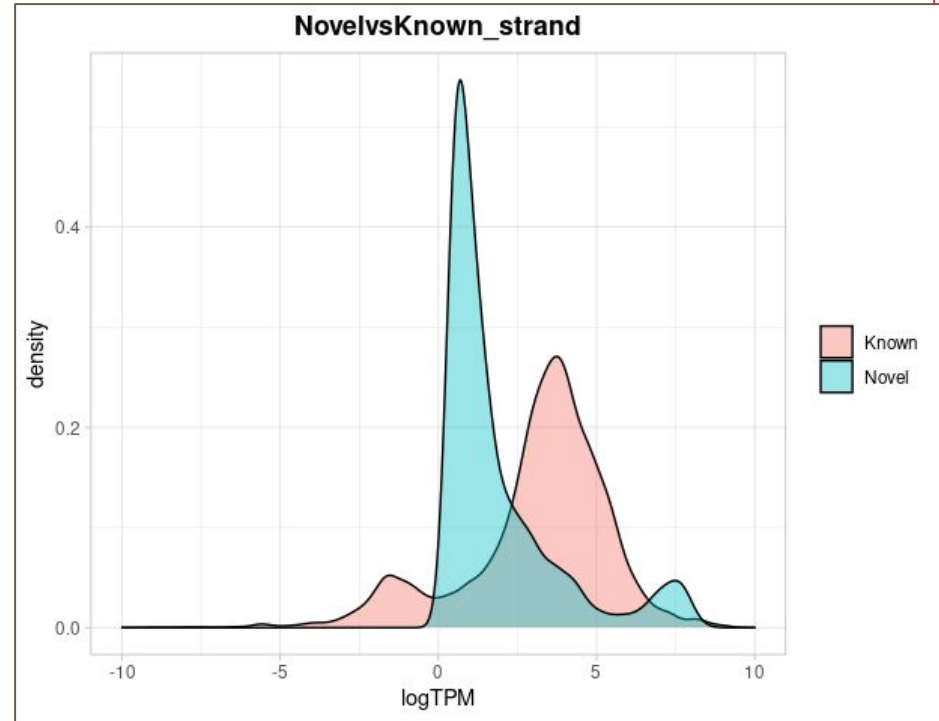
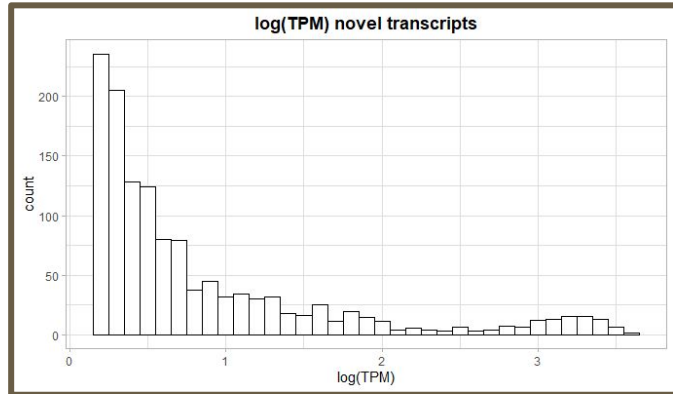
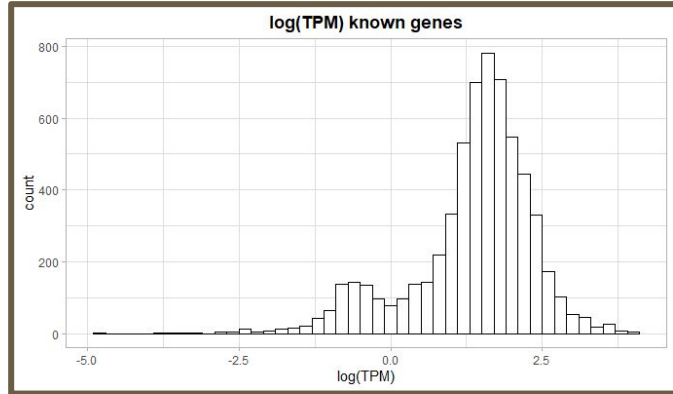


**Novel
transcripts**

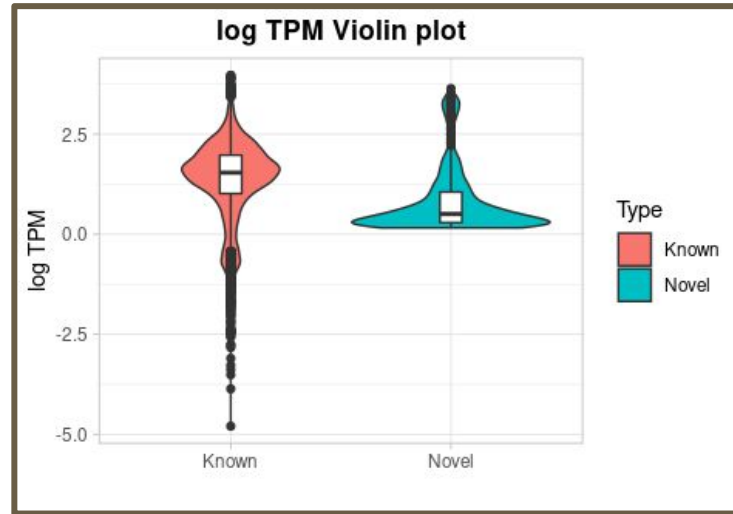


**Known
transcripts**

Results: Transcripts Per Million (TPM)



Results: Transcripts Per Million (TPM)



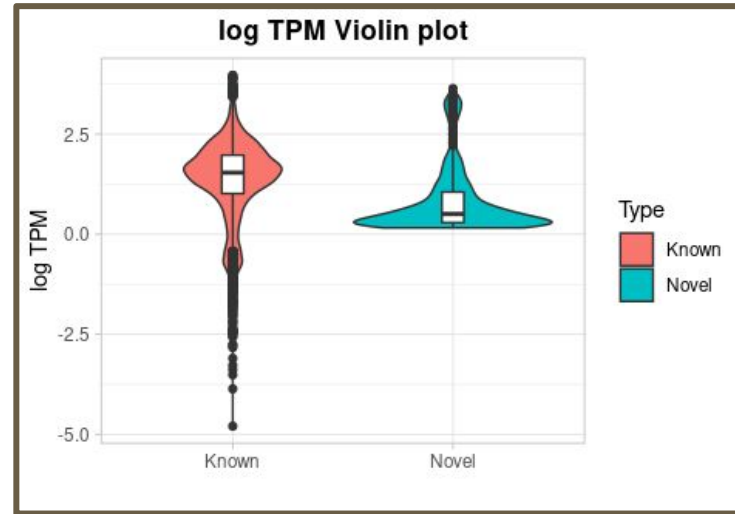
W = 5740918, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

H0: $\mu = \mu$; TPM means equal

H1: $\mu \neq \mu$; TPM means different

Results: Transcripts Per Million (TPM)



W = 5740918, p-value < 2.2e-16

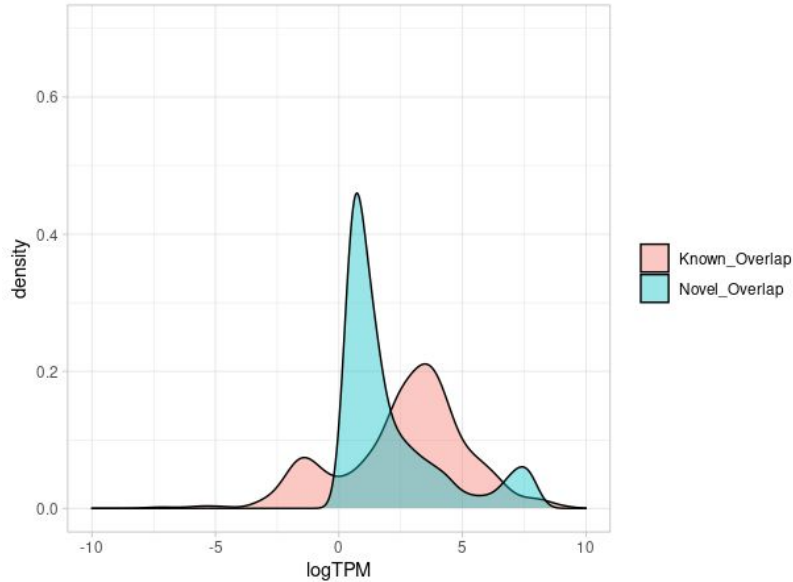
alternative hypothesis: true location shift is not equal to 0

~~H0: $\mu = \mu$; TPM means equal~~

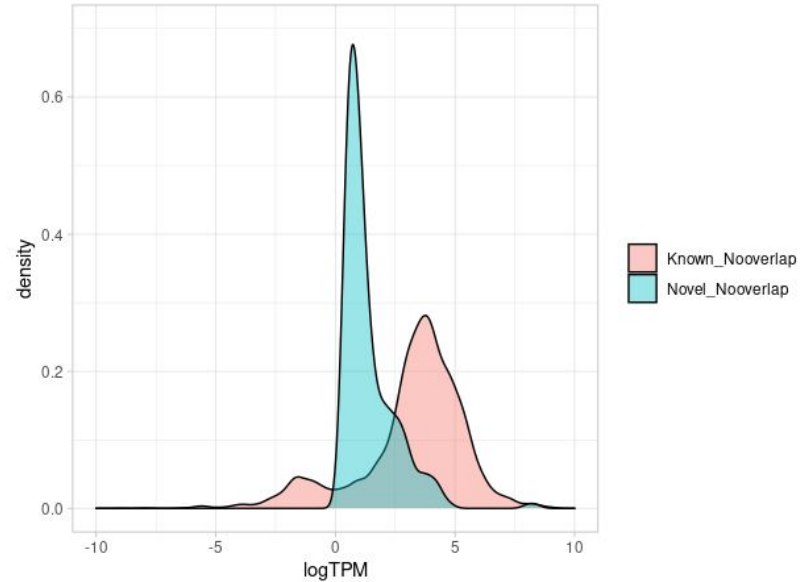
H1: $\mu \neq \mu$; TPM means different

Significant!

Results: Transcripts Per Million (TPM)



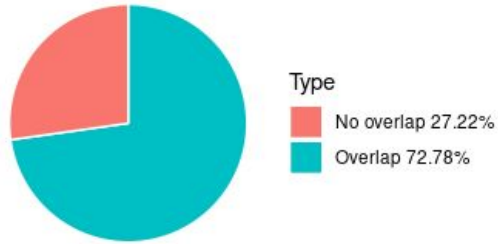
Overlapped



No overlapped

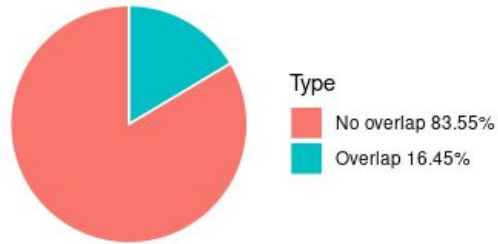
Results: Overlapping

Overlap of novel genes



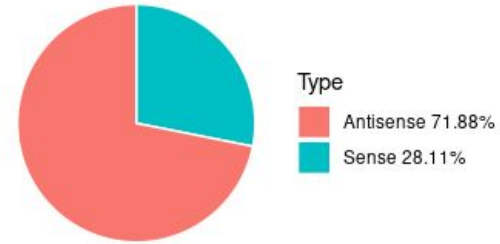
941/1293 novel genes (72.78%)

Overlap of known genes



1017/6180 known genes (16.45%)

Sense vs Antisense in novel genes

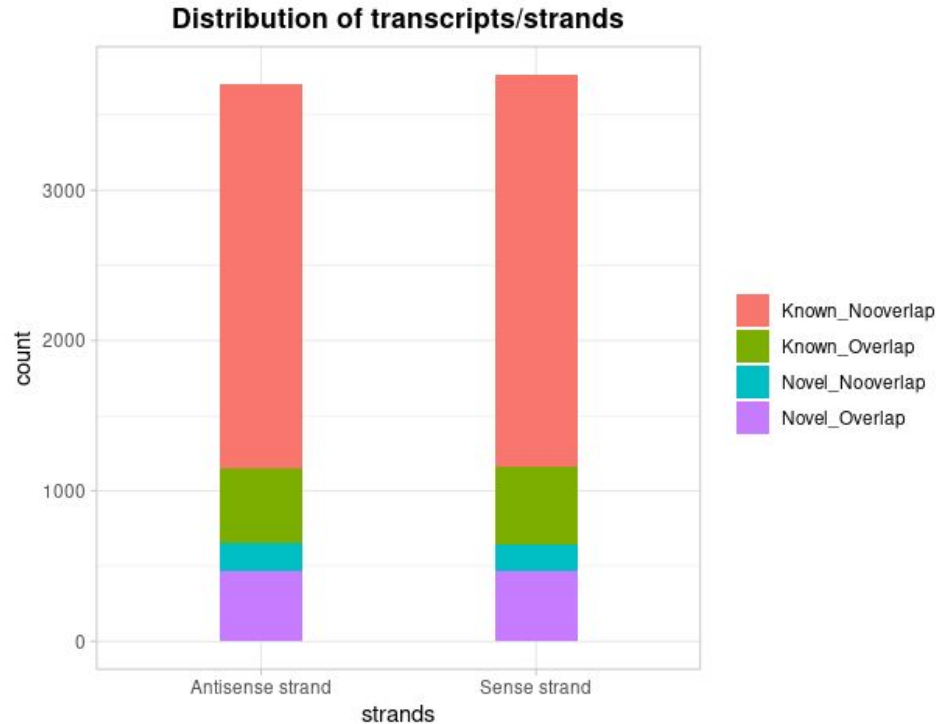


1st Filter:
Overlapping

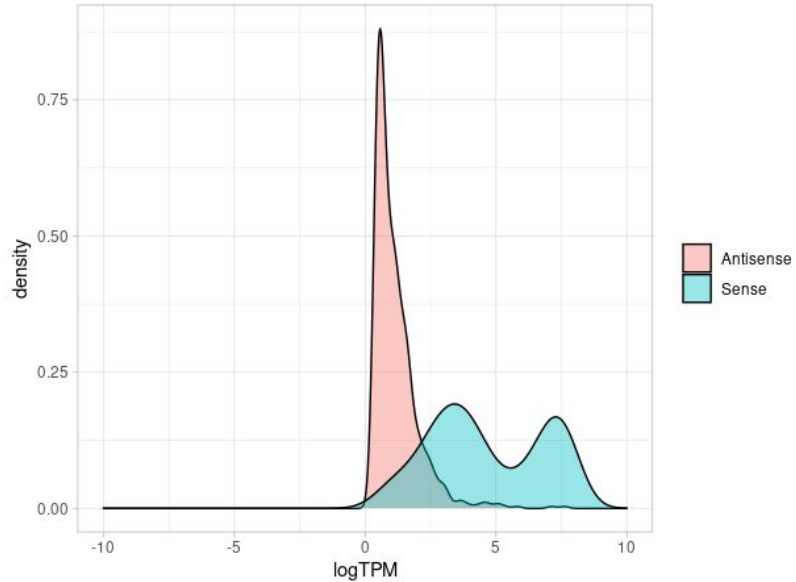
Select novel genes which overlap in
antisense with known genes

Results: Overlapping

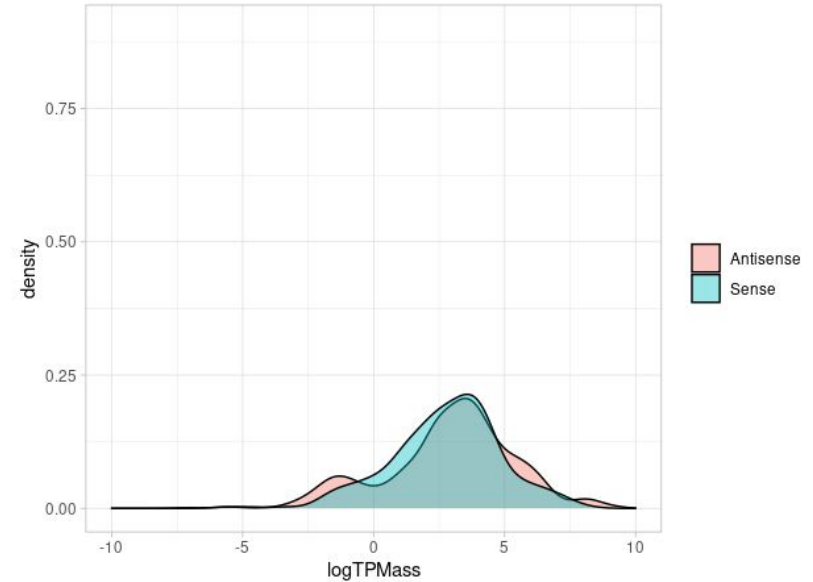
No difference in strand distribution



Results: Overlapping

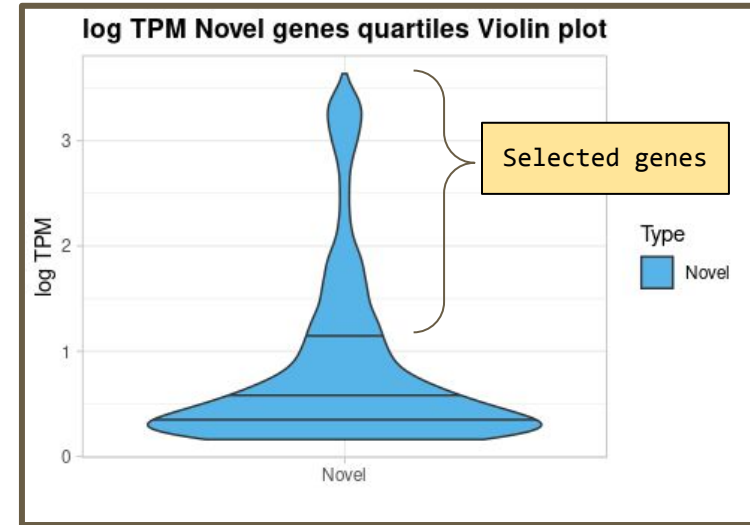
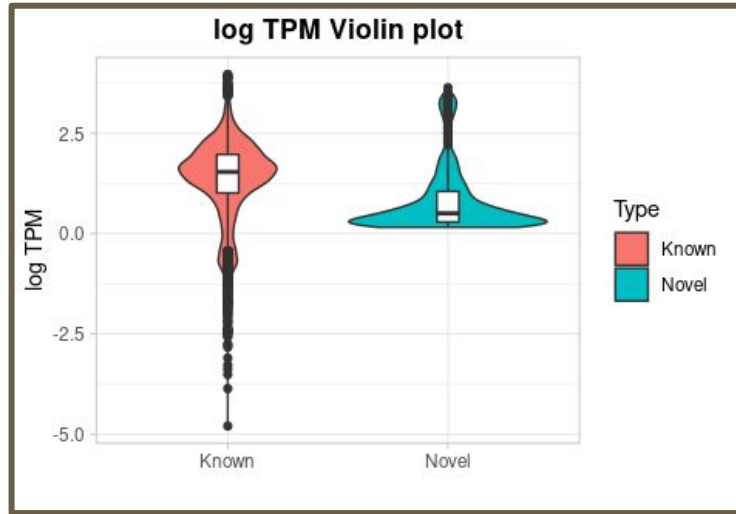


**Novel
transcripts**



**Known
transcripts**

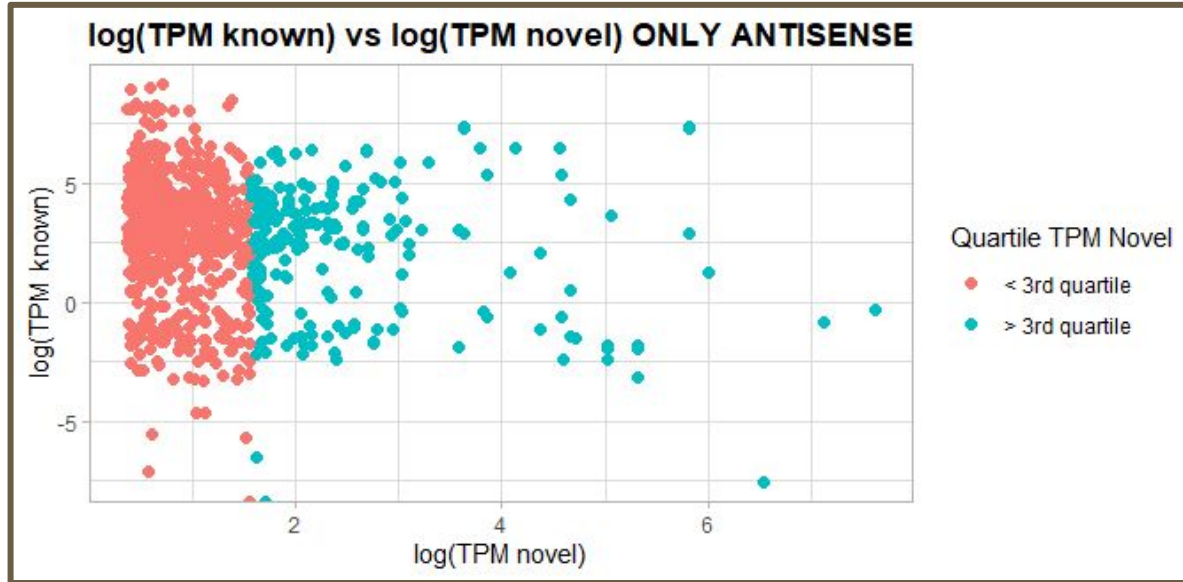
Results: Transcripts Per Million (TPM)



2nd Filter:
TPM Expression

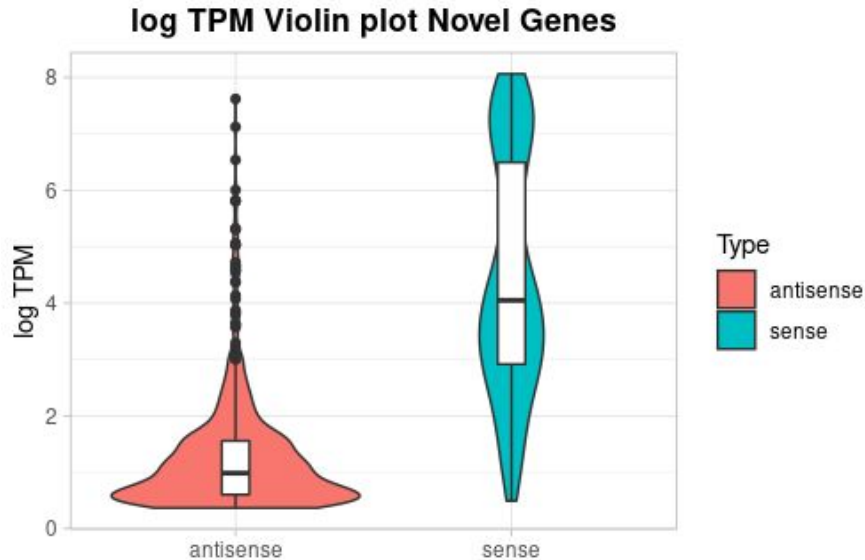
Select genes above third Quartile expression

Overlapping (antisense)



tpm novel expression limits known expression of known

Results: Overlapping sense vs antisense

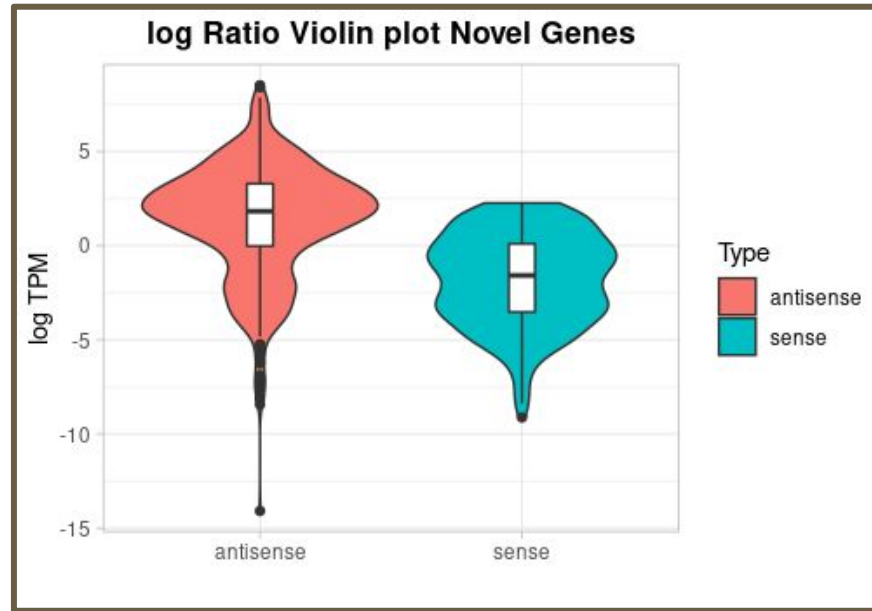


W = 246055, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

H0: $\mu = \mu$; TPM means equal
H1: $\mu \neq \mu$; TPM means different

Significant!

Results: Overlapping sense vs antisense

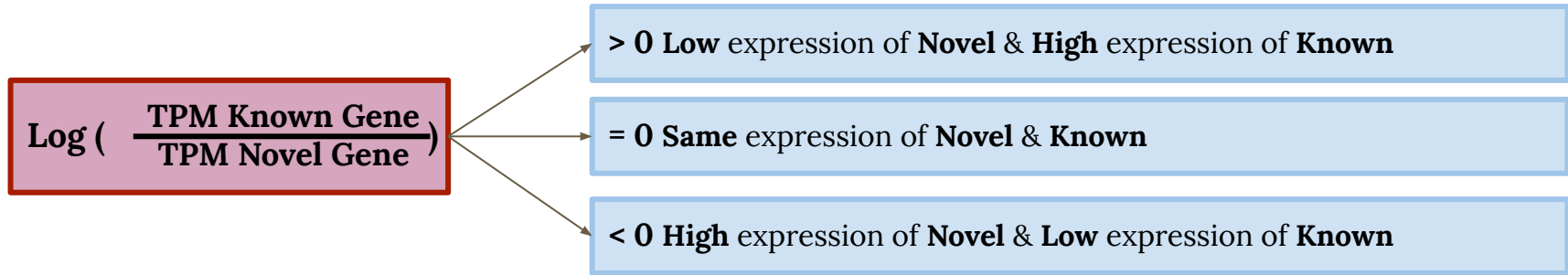


W = 246055, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

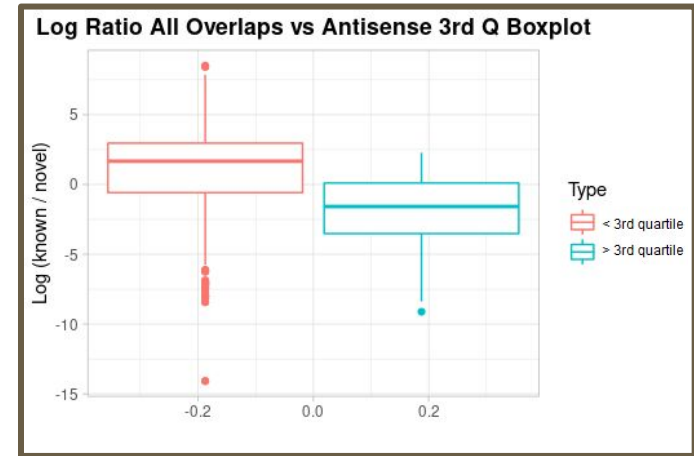
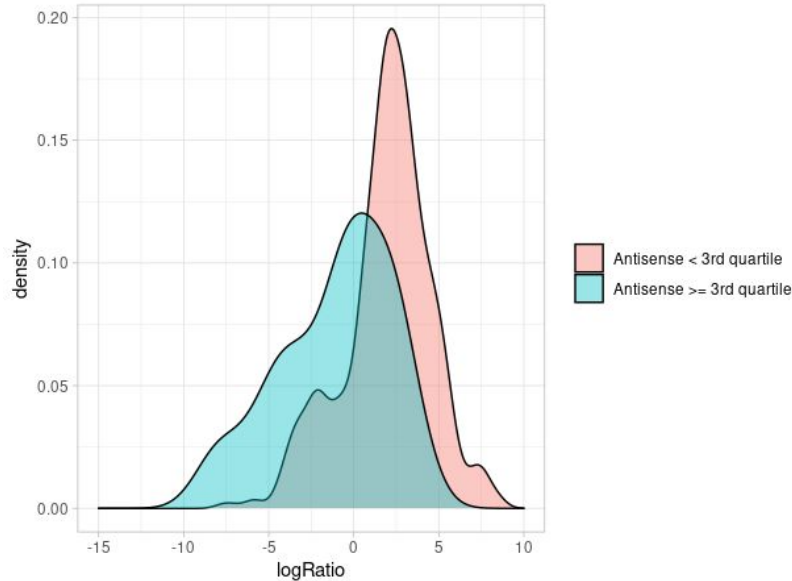
H0: $\mu = \mu$; TPM means equal
H1: $\mu \neq \mu$; TPM means different

Significant!

Results: Overlapping TPM ratio



Results: Overlapping TPM ratio



W = 133386, p-value = 0.0005448

alternative hypothesis: true location shift is not equal to 0

H0: $\mu = \mu$; Ratio means equal

H1: $\mu \neq \mu$; Ratio means different

Significant!

Results: BLAST

Novel Genes

1st Filter:
Overlapping /
antisense

2nd Filter:
TPM expression

52 High Expression Antisense Novel Genes

BLAST search

Use *S.cerevisiae* proteins as database

52 High Expression Antisense Novel Genes

BLAST search

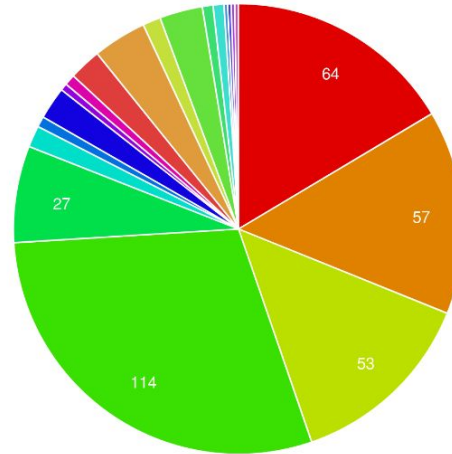
Database: *S.cerevisiae* proteins >60 % Coverage >60% Identity

Query Seq_ID	Subject Seq_ID	Cov	% Iden	E-Value
SPAR_10:112902-113324(-)	YKL163W	82	42.222	1.7
SPAR_10:395945-396731(-)	YJL019W	74	93.333	1.43e-128
SPAR_11:352762-353164(+)	YKL053C-A	64	98.837	8.33e-60
SPAR_11:538727-539831(+)	YKR046C	74	97.048	0.0
SPAR_1:1914-2679(+)	YAR050W	99	65.613	7.38e-24
SPAR_1:2184-2679(+)	YAR050W	99	61.058	7.23e-15
SPAR_12:24337-26797(+)	YLL051C	87	88.343	0.0
SPAR_12:718642-719940(-)	YLR316C	84	43.182	0.014
SPAR_13:161767-164049(-)	YHR216W	69	74.834	3.43e-72
SPAR_13:20547-21697(+)	YML125C	81	96.795	0.0
SPAR_13:266344-271349(-)	YMR008C	74	72.203	0.0
SPAR_13:821071-822734(+)	YCR105W	63	25.067	4.32e-16
SPAR_14:266791-267189(-)	YNL190W	99	99.242	1.36e-82
SPAR_14:586307-588170(-)	YNL012W	90	91.555	0.0
SPAR_16:120217-121227(+)	YPL241C	68	75.546	8.22e-125
SPAR_2:415564-417000(-)	YBR101C	61	96.552	0.0
SPAR_4:258909-260042(-)	YDL115C	94	89.297	0.0
SPAR_4:974094-975130(-)	YDR367W	62	93.458	5.71e-133
SPAR_6:251473-251854(+)	YFR033C	69	68.182	3.27e-05
SPAR_7:65861-66544(+)	YGL230C	65	71.622	2.22e-76
SPAR_8:264668-270203(-)	YHR096C	62	69.019	0.0

Results: Overlapping Ontology Classification

Overlapping Known Genes

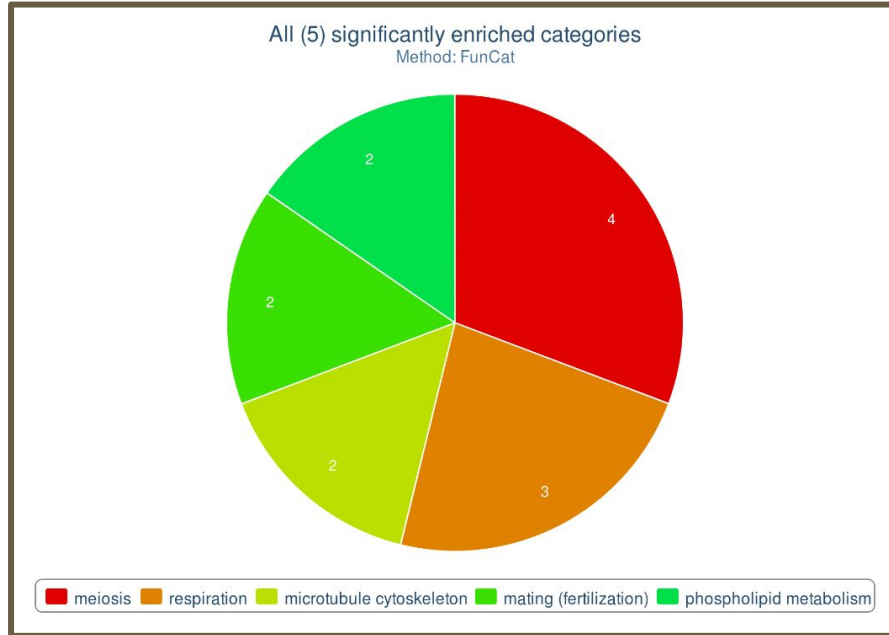
All (20) significantly enriched categories
Method: FunCat



ribosomal proteins translation ribosome biogenesis protein binding RNA binding sugar transport tubulin dependent transport microtubule cytoskeleton polysaccharide binding
 NIK-I-kappaB/NF-kappaB cascade ER to Golgi transport vesicular transport (Golgi network, etc.) vesicle fusion protein modification cell motility cell-cell adhesion regulation of channel activity
 non-directional cell growth other programmed cell death xanthine catabolism

Results: Overlapping Ontology Classification

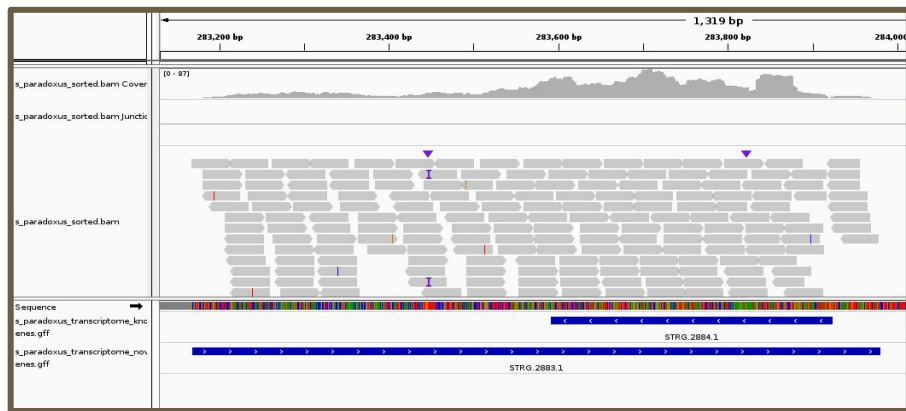
Overlapping Novel Genes



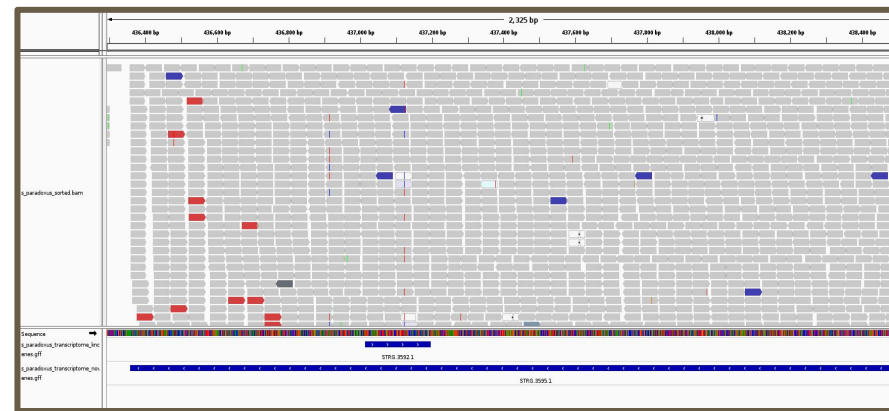
Results: Selection of Genes

Proposed candidate genes for further studies

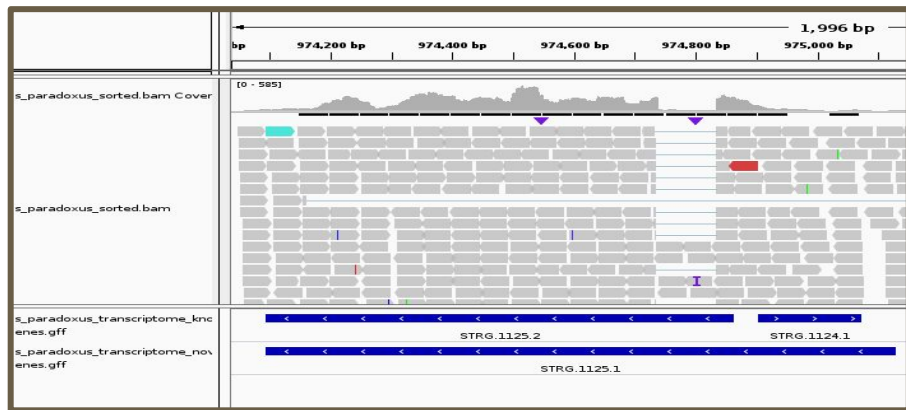
Novel Genes		BLAST	Description	Known Genes	Prediction	Description	Ratio = TPM Known/ TPM Novel
Contig	Transcript ID						
SPAR_10	STRG.2883.1	-		Spar_10.169	YJL077C	Ics3p protein	0.019351
SPAR_4	STRG.1125.1	YDR367W	Kei1p protein	Spar_4.594	YDR366C	Mor1p protein	0.003904
SPAR_7	STRG.1826.2	YGL230C		Spar_7.36	YGL230C	uncharacterized protein	0.023884
SPAR_12	STRG.3595.1	-		Spar_12.250	YLR154W-C	TAR1 protein	0.000910
SPAR_12	STRG.3595.2	-		Spar_12.256	YLR106C	ATPase midasin	0.001938



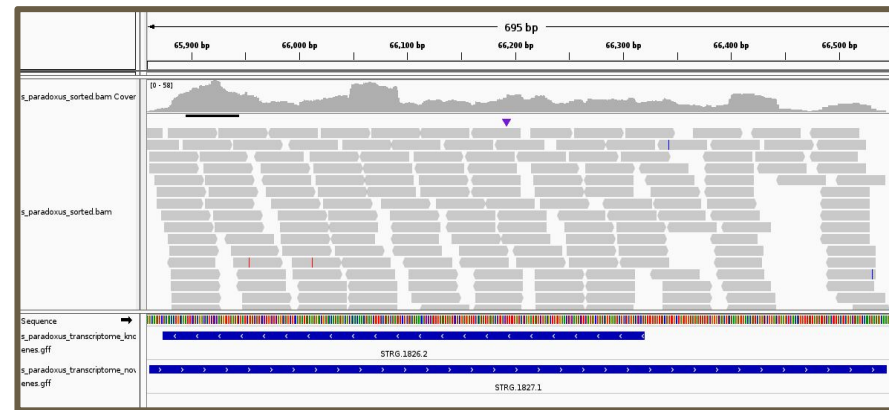
STRG.2883.1



STRG.3595.1



STRG.1125.1



STRG.1827.1

Discussion

- We describe the presence of less abundant novel transcripts in s_paradoxus.
- Novel genes transcripts are shorter and show less percentage of GC and distributed equally in the two strands of DNA.
- These novel transcripts overlap with the canonical genes in the genome.

Discussion

- Expression of the novel transcripts suggest a reduction of the expression of the canonical described ones.
- Novel transcripts impair the expression of canonical genes related to protein formation and interaction.
- Blast and GO analysis of the more expressed novel genes report putative roles in metabolism and meiosis.

Bibliography

- Ardern, Z., Neuhaus, K., & Scherer, S. (2020). Are Antisense Proteins in Prokaryotes Functional? *Frontiers in Molecular Biosciences*, 7, 187. <https://doi.org/10.3389/fmolb.2020.00187>
- Blevins, W. R., Ruiz-Orera, J., Messeguer, X., Blasco-Moreno, B., Villanueva-Cañas, J. L., Espinar, L., Díez, J., Carey, L. B., & Albà, M. M. (2021). Uncovering de novo gene birth in yeast using deep transcriptomics. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-20911-3>
- Dunham, M. J., & Louis, E. J. (2011). Yeast evolution and ecology meet genomics. *EMBO Reports*, 12(1), 8–10. <https://doi.org/10.1038/embor.2010.204>
- Wight, M., & Werner, A. (2013). The functions of natural antisense transcripts. *Essays in Biochemistry*, 54, 91–101. <https://doi.org/10.1042/bse0540091>