Universitat Pompeu Fabra Barcelona

Universitat de Barcelona

*MSc in Bioinformatics for Health Sciences*

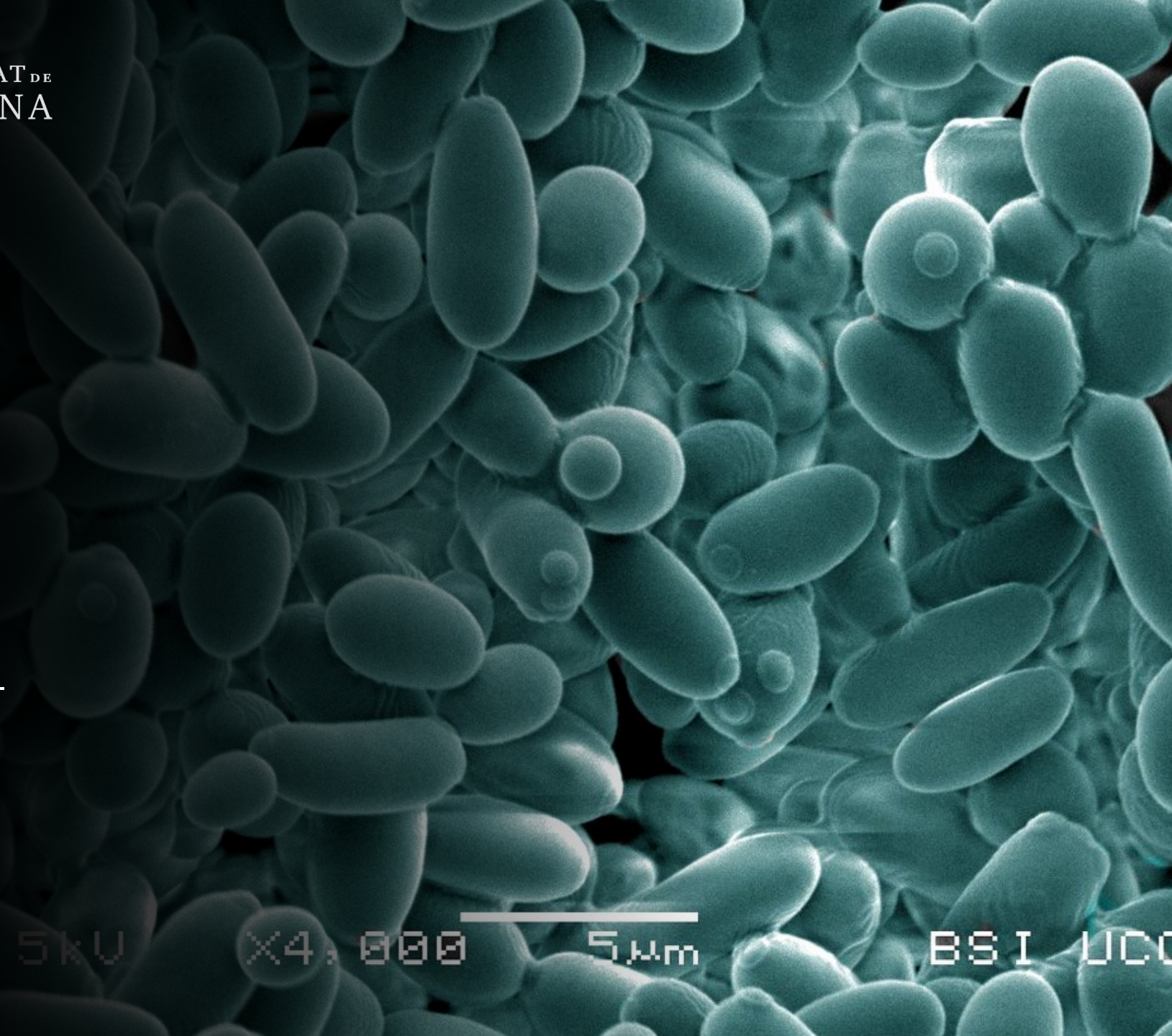
# *Lachancea kluyveri*

## TRANSCRIPTOME ANALYSIS


Clara Suárez Quintana

Júlia Vilalta Mor

Sara Polo Alonso

# INTRODUCTION

*L. kluyveri* → Ascomycetous yeast

- Fruit flies
- Slime fluxes
- Soils
- Food

Only 30 isolates → It is not well studied

*De novo* genes → Gene repertoire across different species

# AIM OF THIS STUDY

To know, annotate and report genes thanks to transcriptome analysis

## WHY?

Possible applications in:

- Industry
- Biomedicine

## HOW?

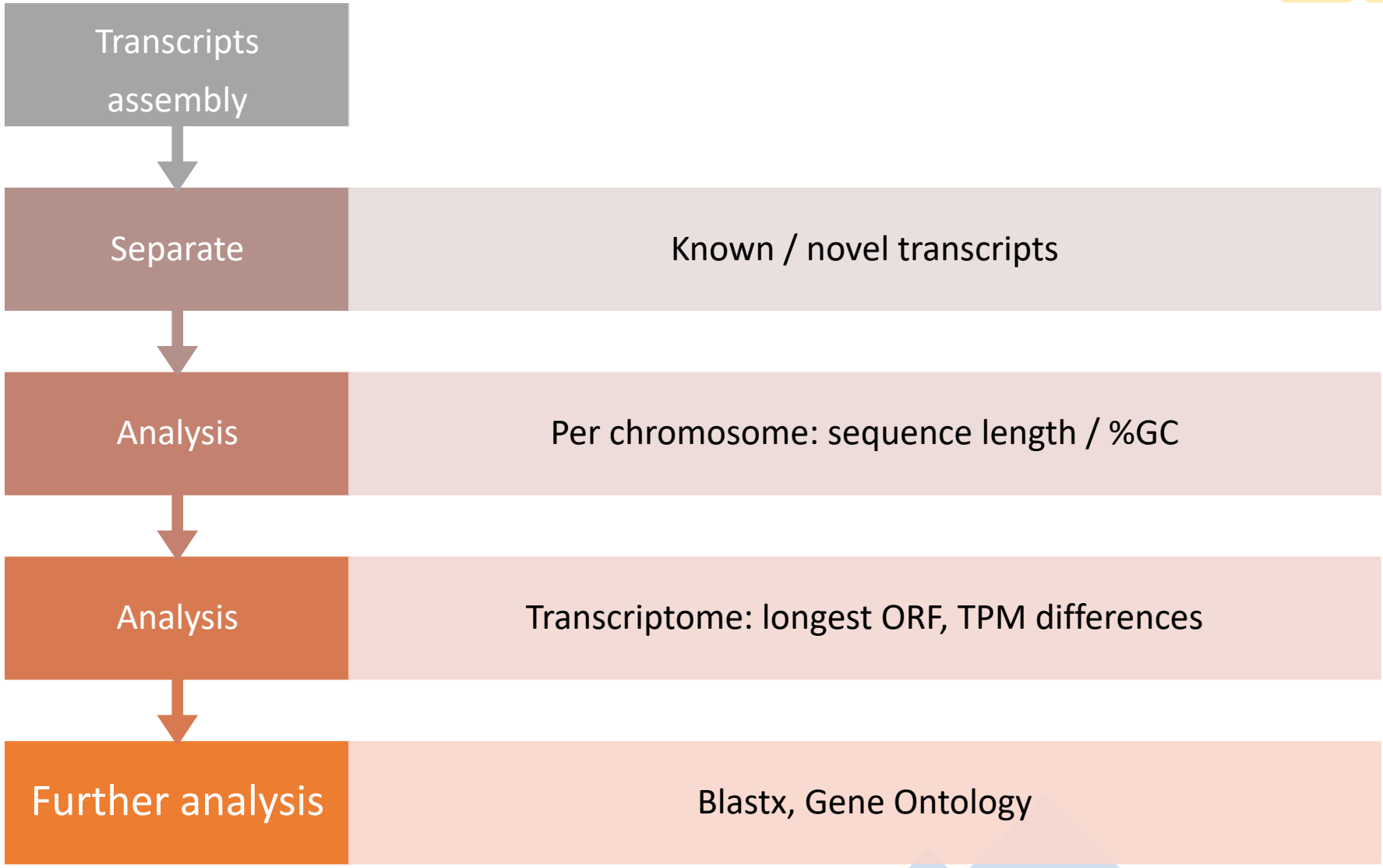
Analysis (novel/known):

- Sequence length
- % GC
- TPM
- Blastx /Gene Ontology

# WORKFLOW

| | |
|---|---|
| Reads quality analysis (QA) | Reads |
| Trimming | Remove adapters/bad quality |
| Basic Analysis | Paired reads |
| Build genome index | |
| Align reads to index | SAM → BAM. IGV. |

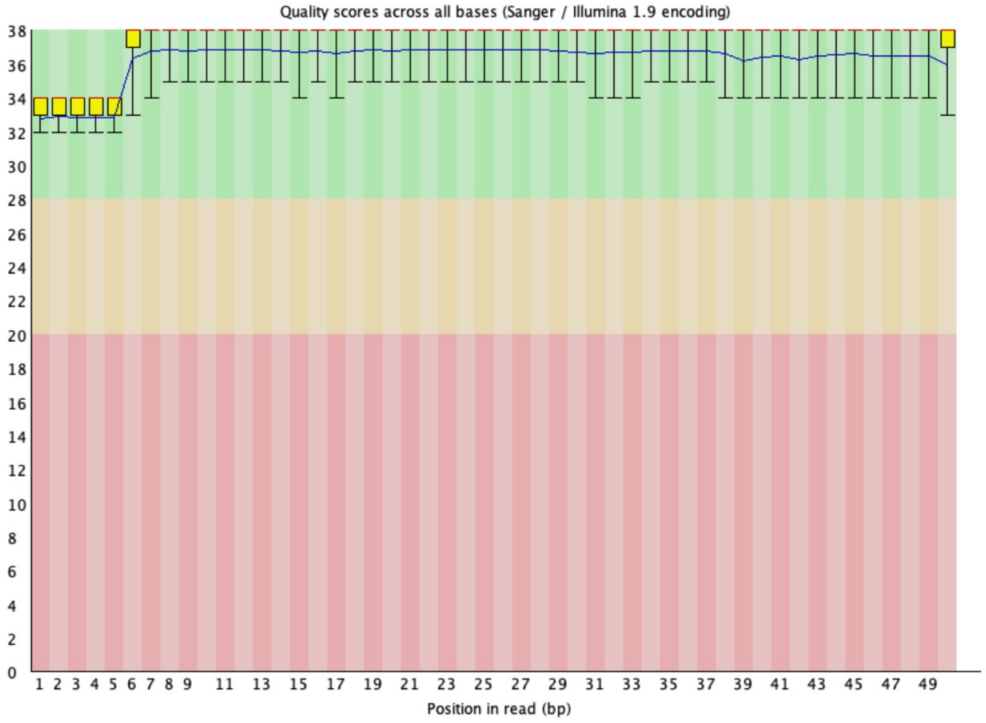
# WORKFLOW


Transcripts assembly
Separate
Known / novel transcripts
Analysis
Per chromosome: sequence length / %GC
Analysis
Transcriptome: longest ORF, TPM differences
Further analysis
Blastx, Gene Ontology

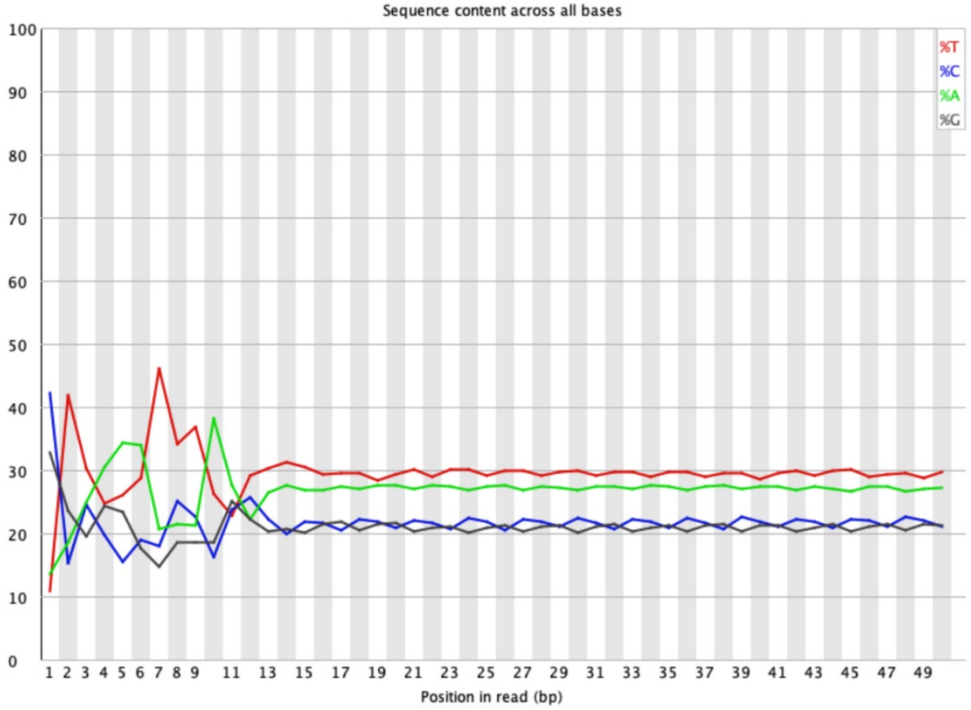
# READS QUALITY (R1)

*Before trimming*

**Per base sequence quality**

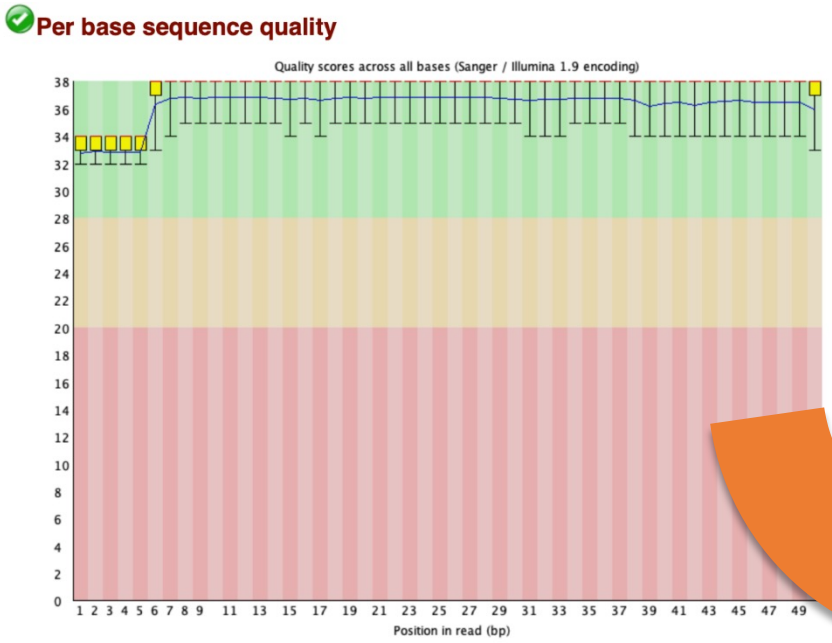


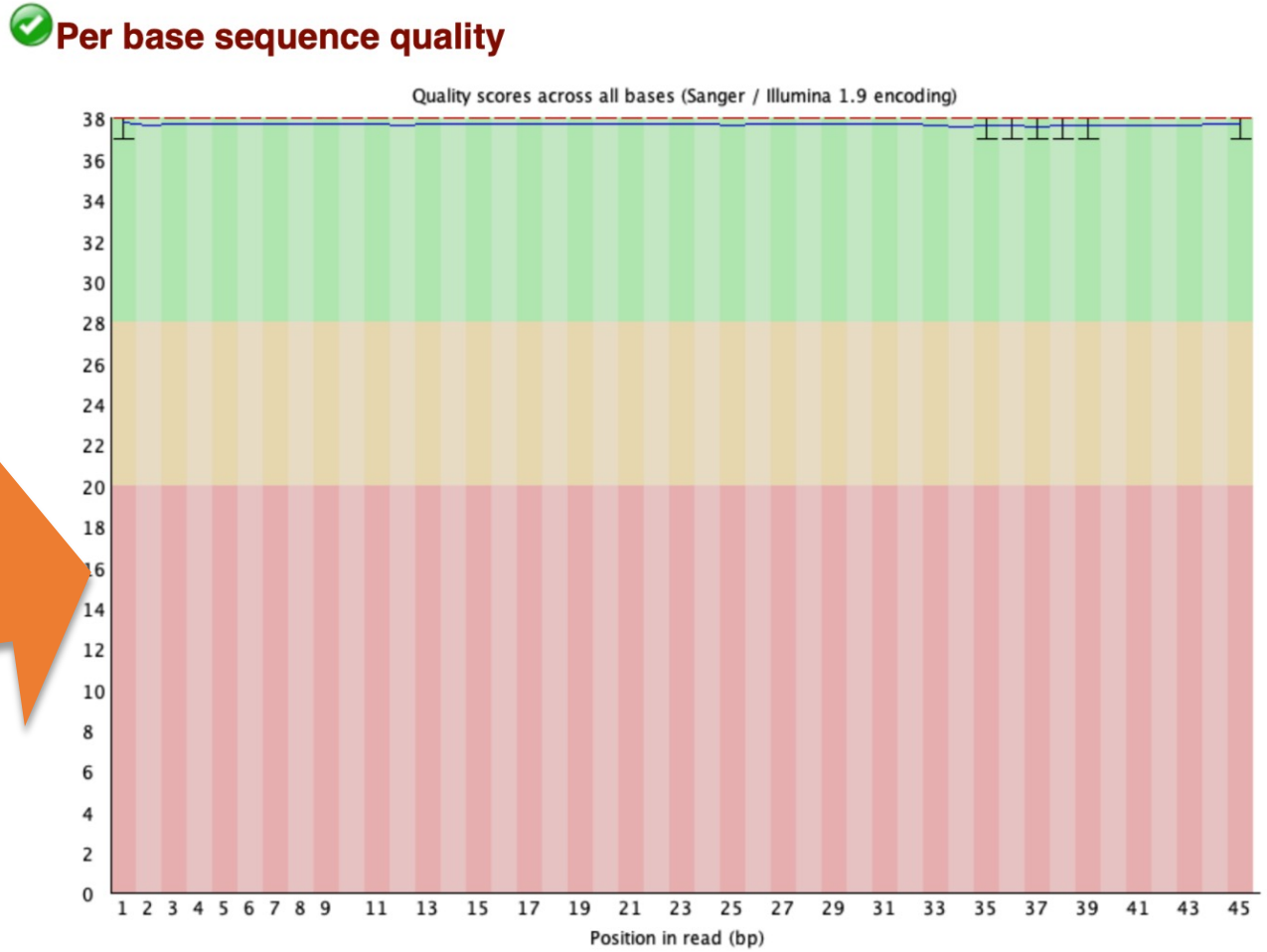

**Per base sequence content**



# READS QUALITY (R1)

*After trimming*



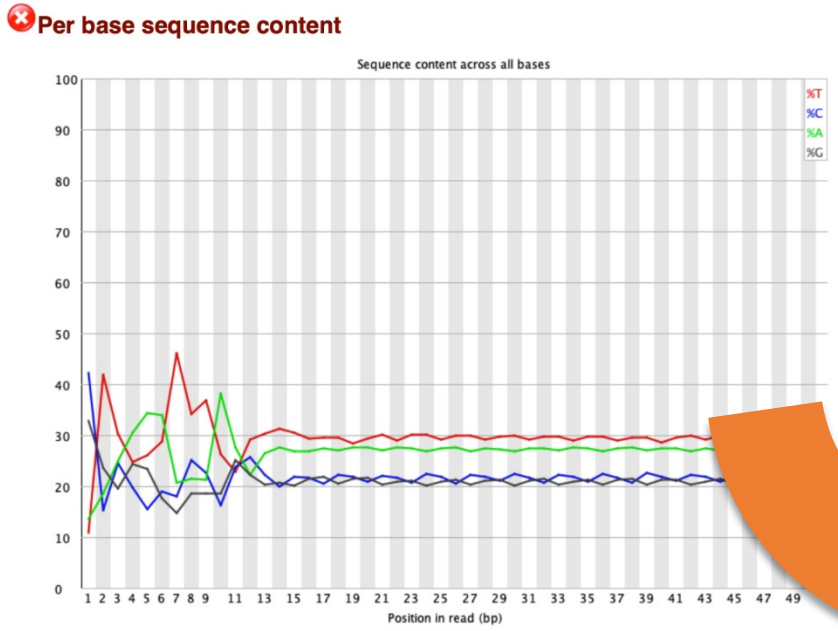



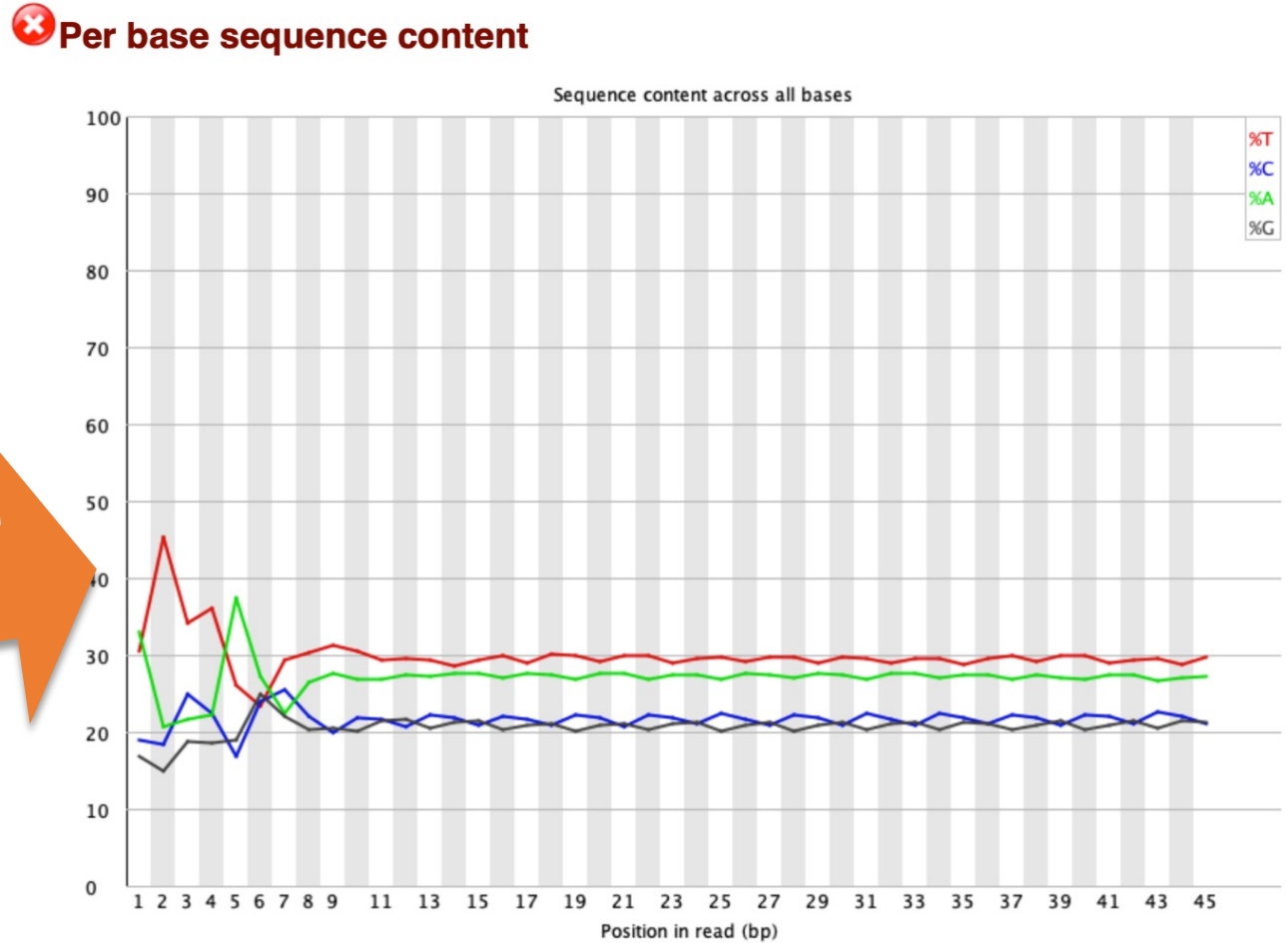

*number of reads*

**24 399 133 → 16 941 329**

# READS QUALITY (R1)

*After trimming*



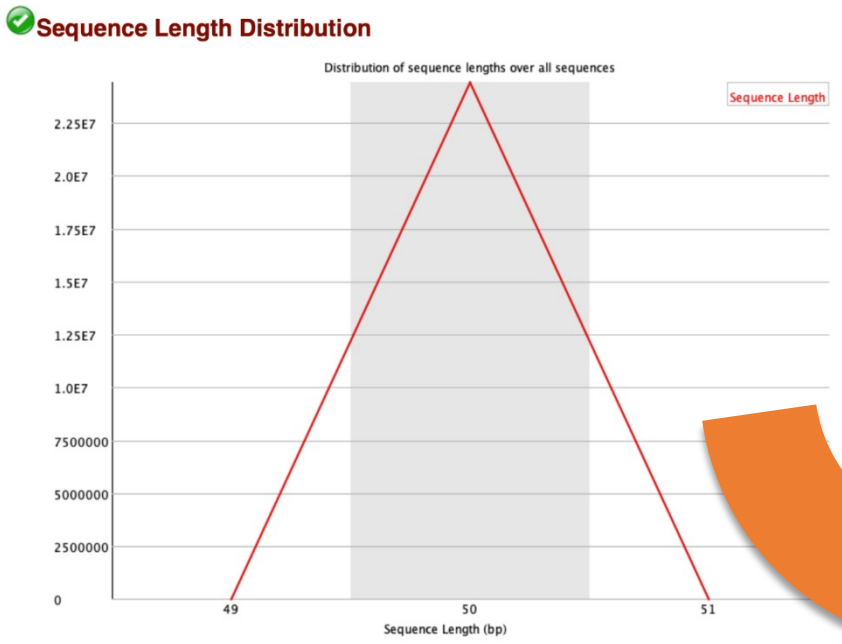



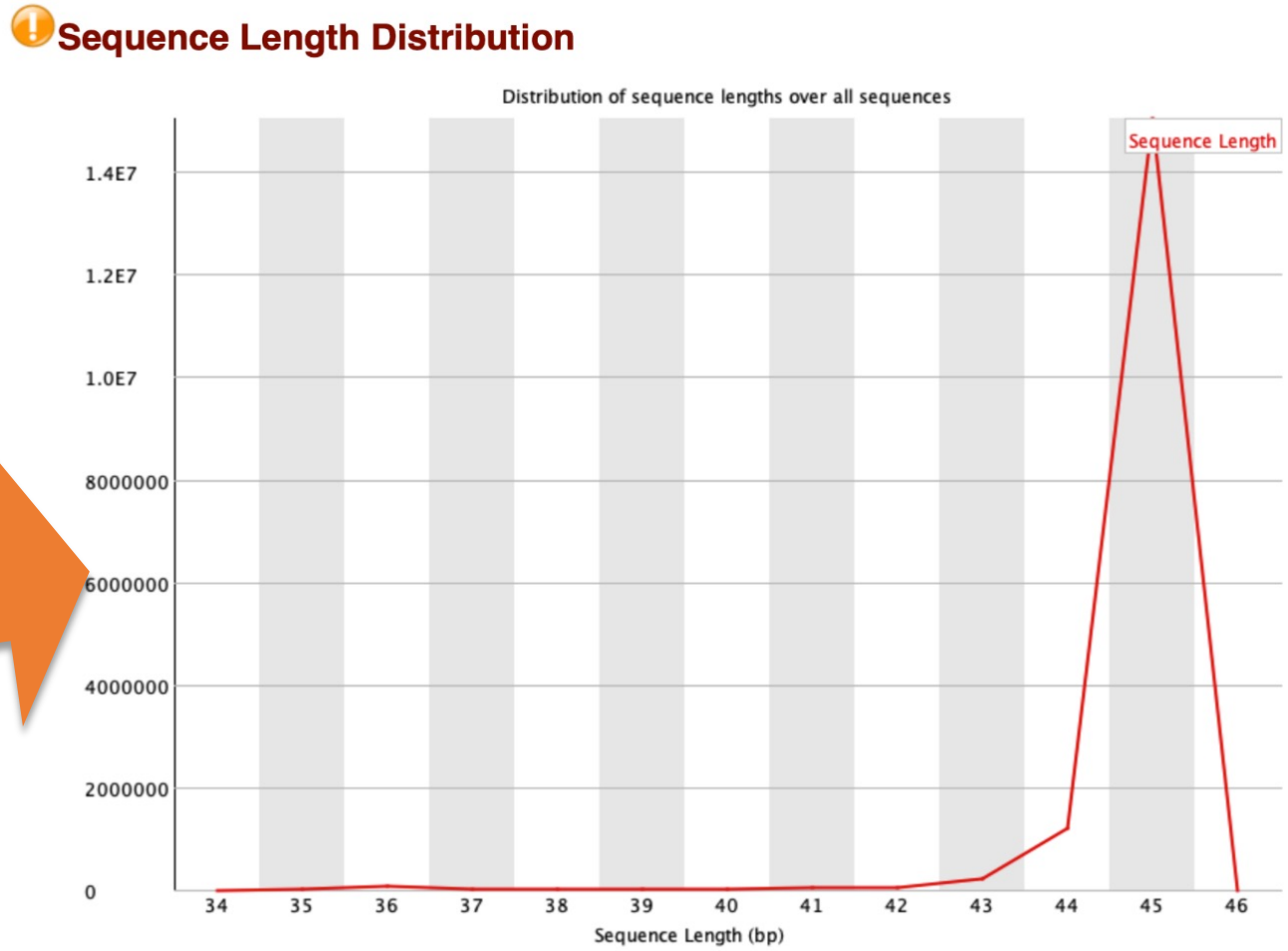

*number of reads*

**24 399 133 → 16 941 329**

# READS QUALITY (R1)

*After trimming*

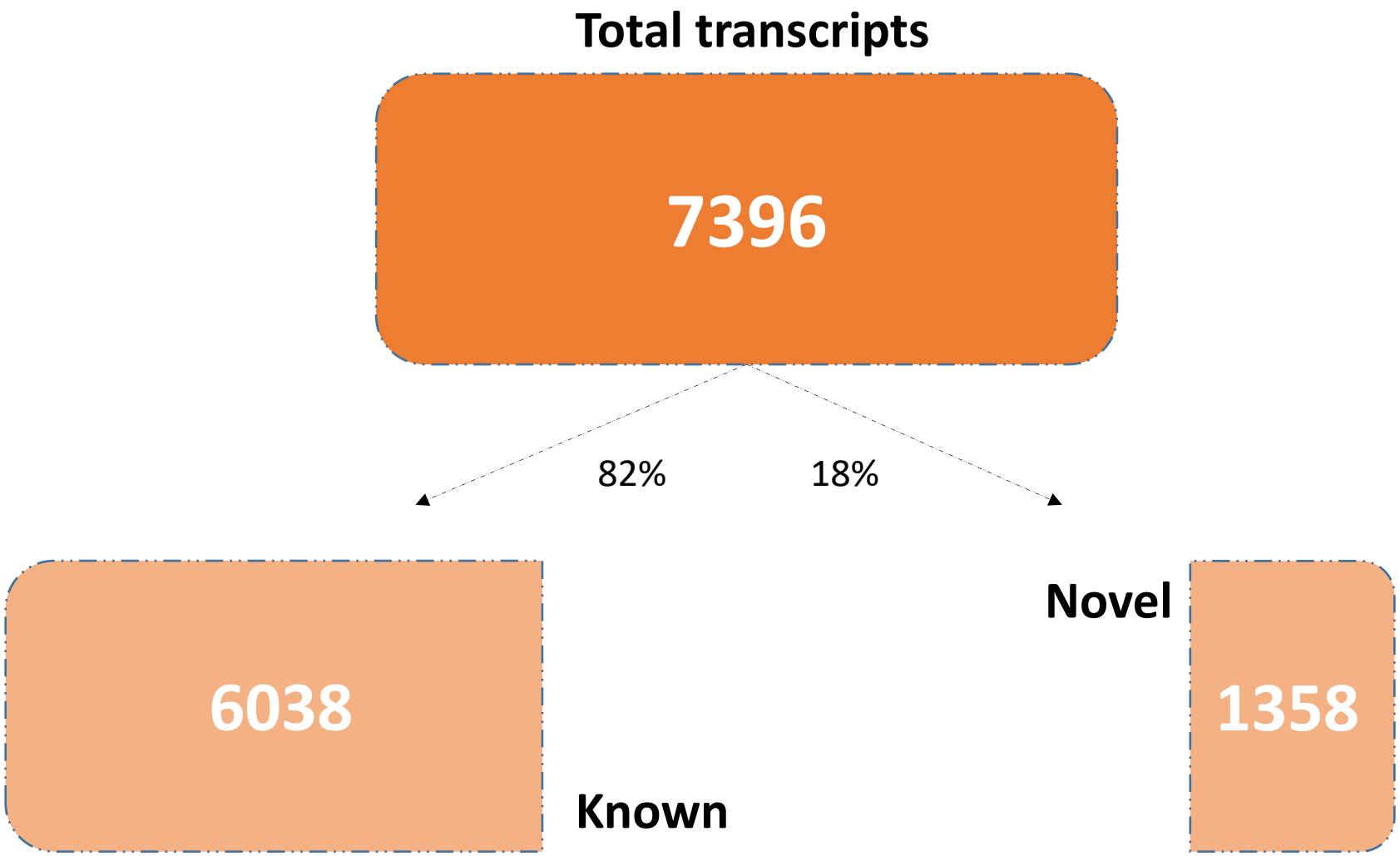



*number of reads*

**24 399 133 → 16 941 329**

# TRANSCRIPTOME ASSEMBLY



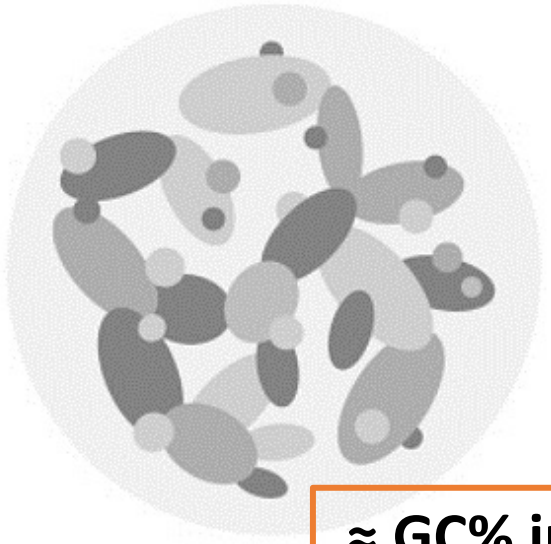
# NOW WHAT?

Conserved genes

vs

Novel genes

**Shorter**
**Higher expression**
**High GC%**



**PROVE**

**≈ GC% in novel and known?**

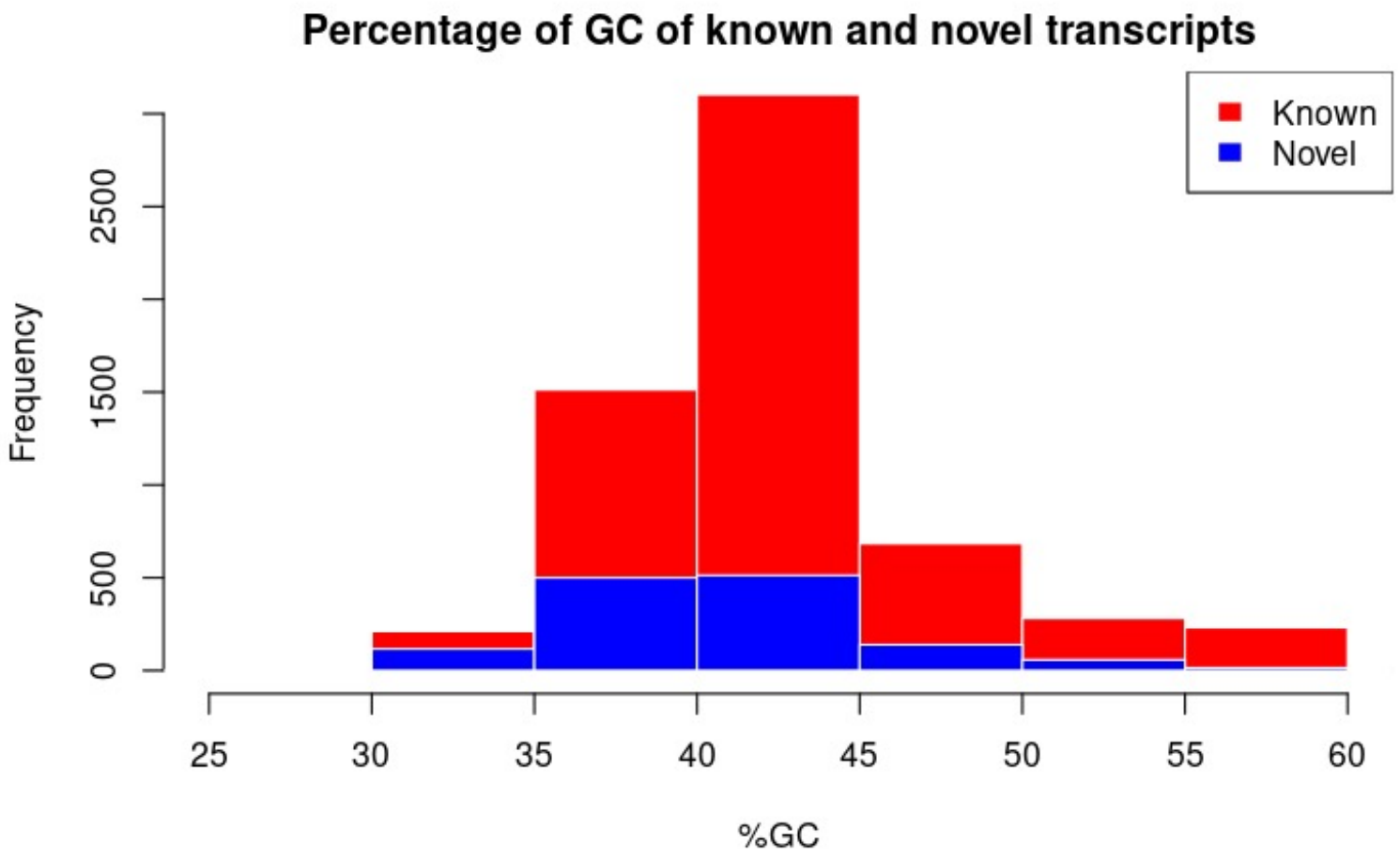
**↓Length novel than known?**

# Percentage of GC

Wilcoxon test

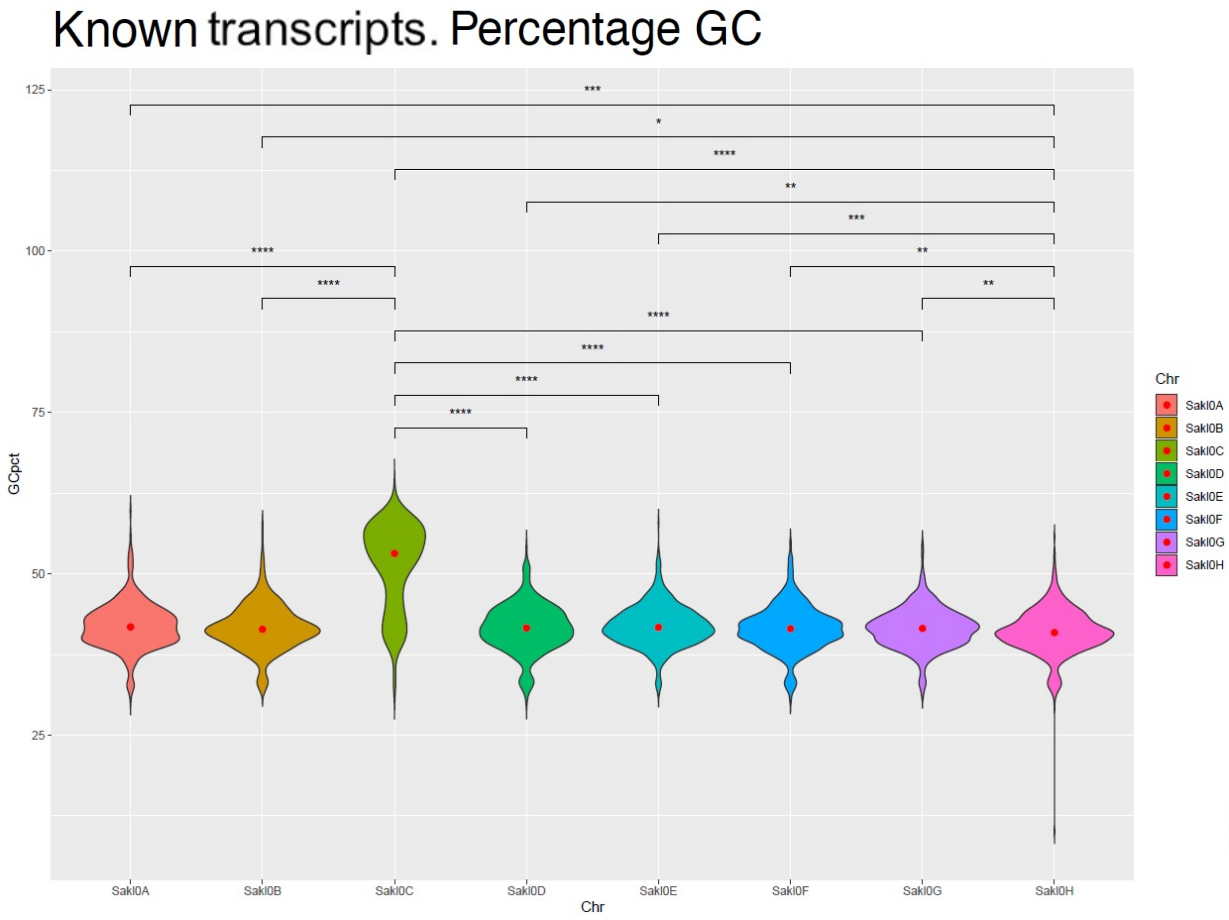
p-value < 2.2e-16

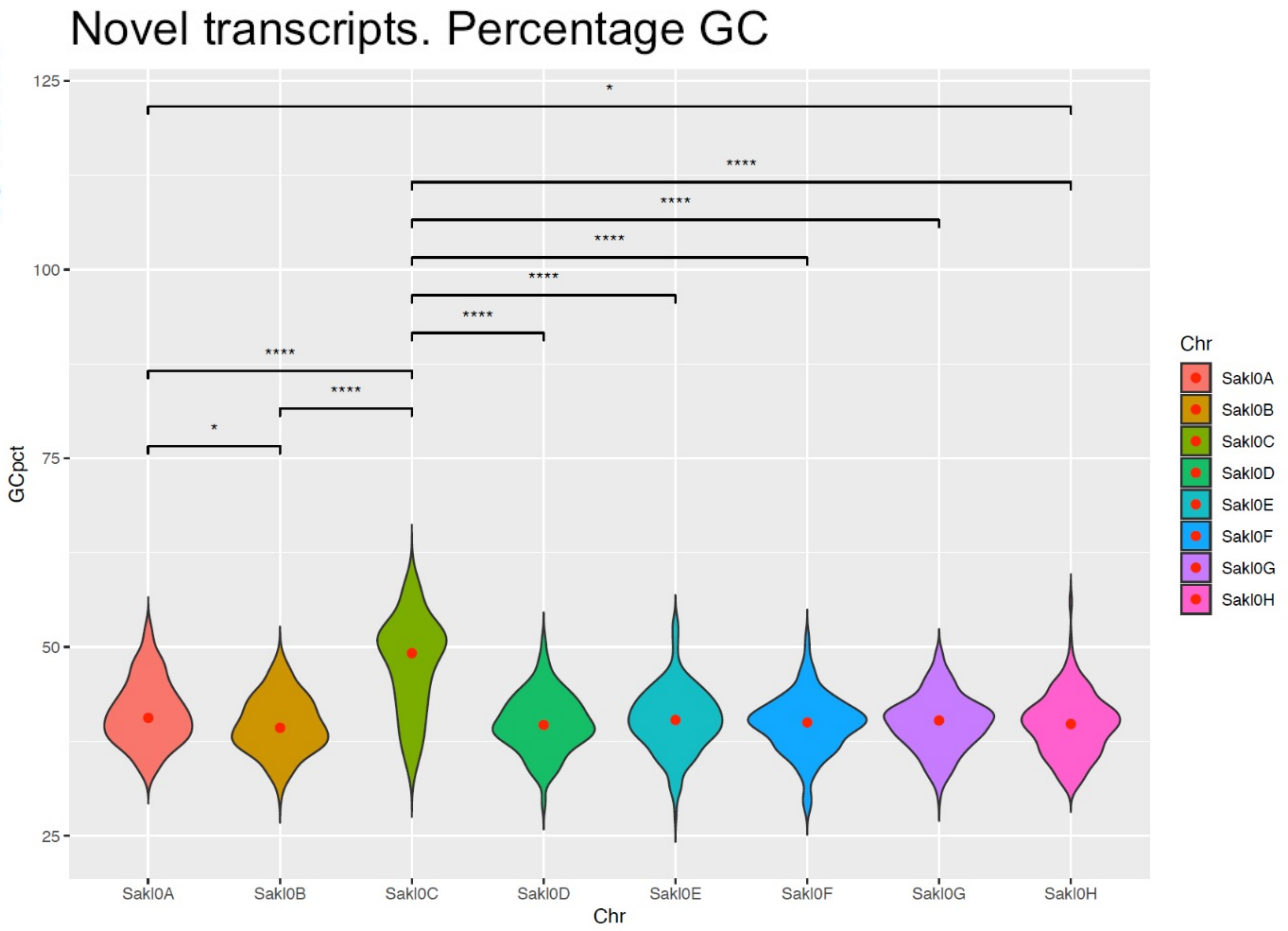
The differences between the distributions are **significant**




Mean %GC
- NOVEL: 40.77
- **KNOWN: 42.56**

# Percentage of GC per chromosome




Test: ANOVA+Tukey

| SYMBOL | MEANING |
|---|---|
| * | P ≤ 0.05 |
| ** | P ≤ 0.01 |
| *** | P ≤ 0.001 |
| **** | P ≤ 0.0001 |

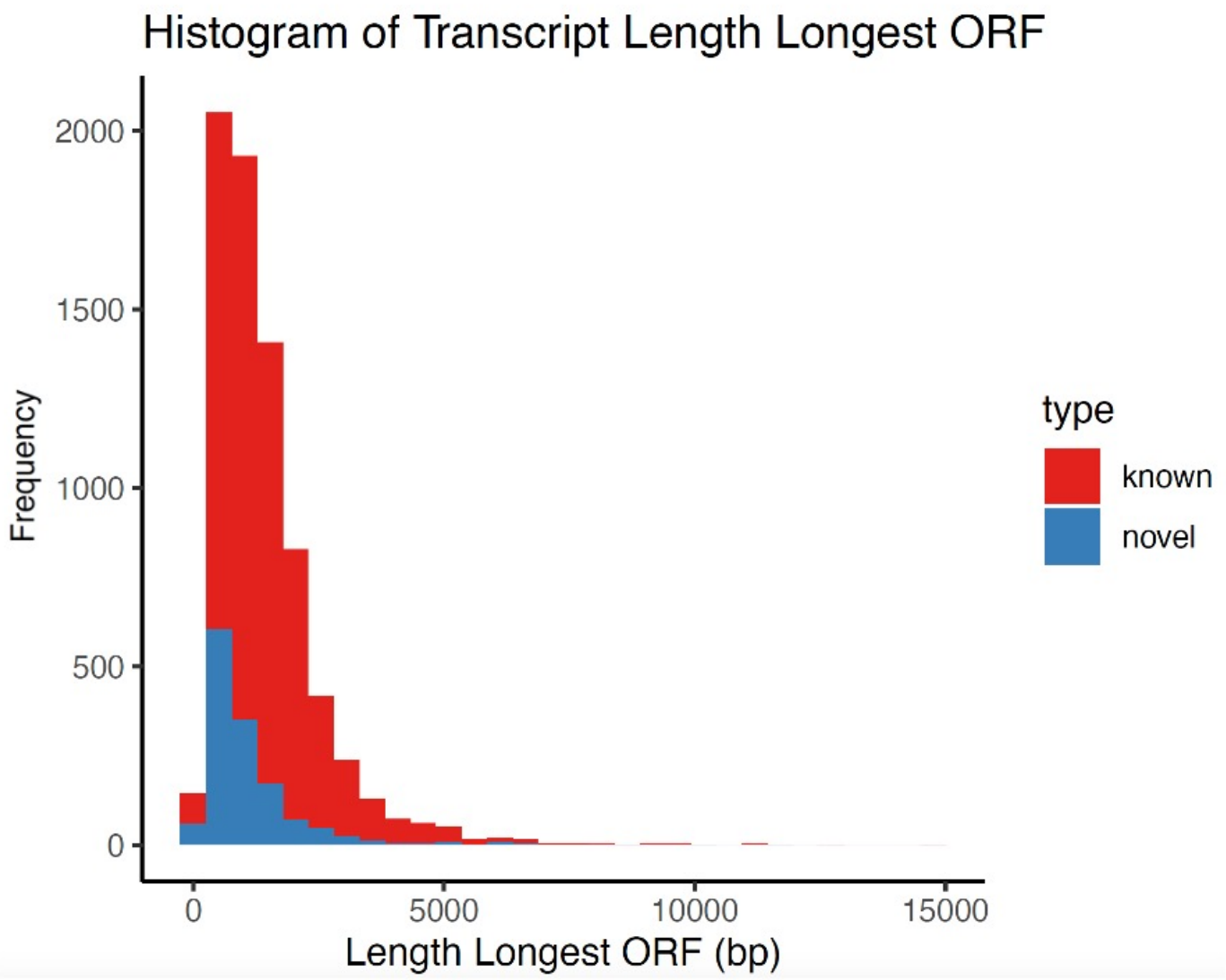
Significant differences
Chr novel-Chr known

# Length of longest ORF

Wilcoxon test

p-value < 2.2e-16

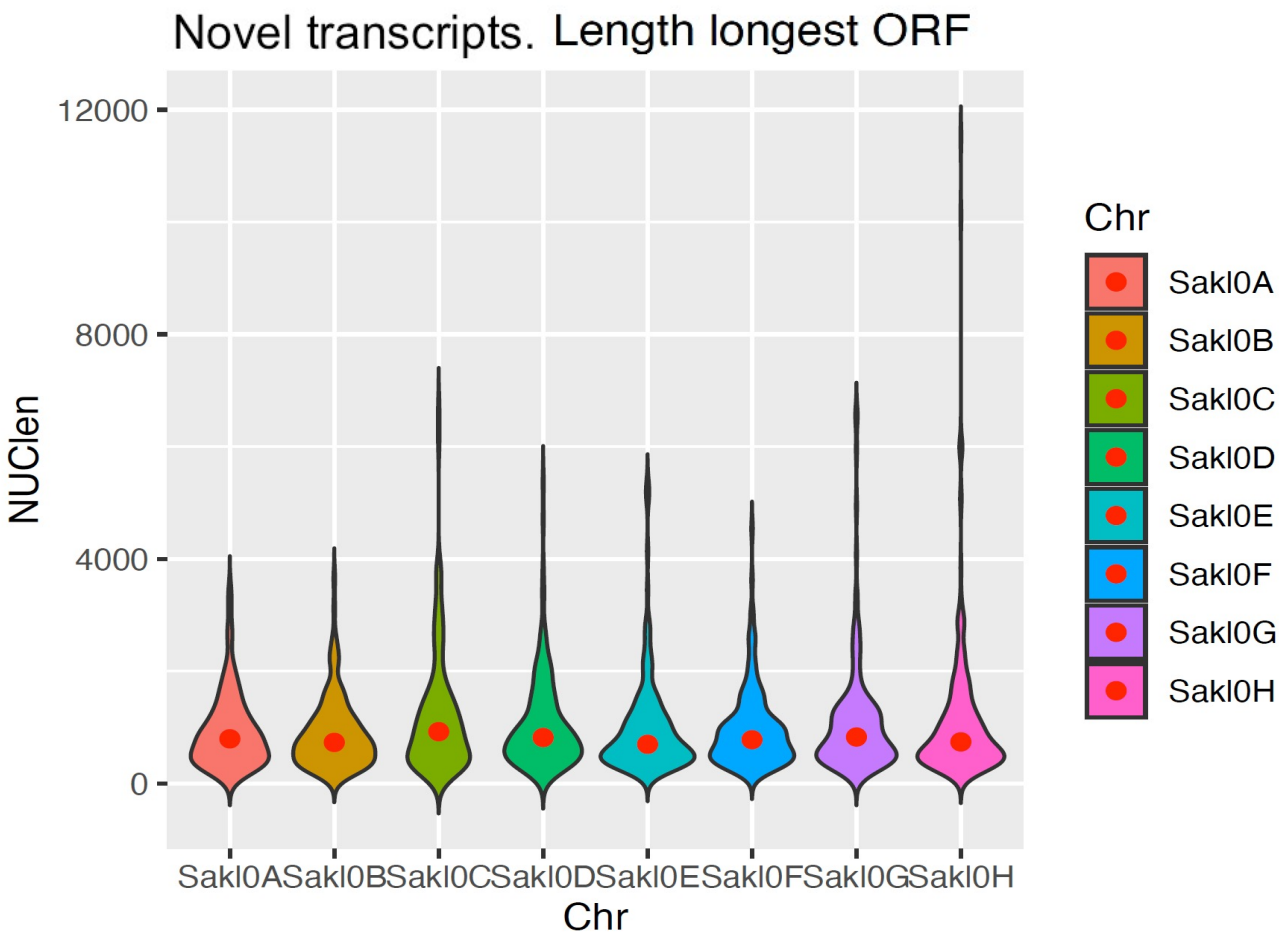
The differences between the distributions are **significant**




Mean length (pb)
- NOVEL: 1043
- **KNOWN: 1476**

# Length of longest ORF per chromosome

No significant differences

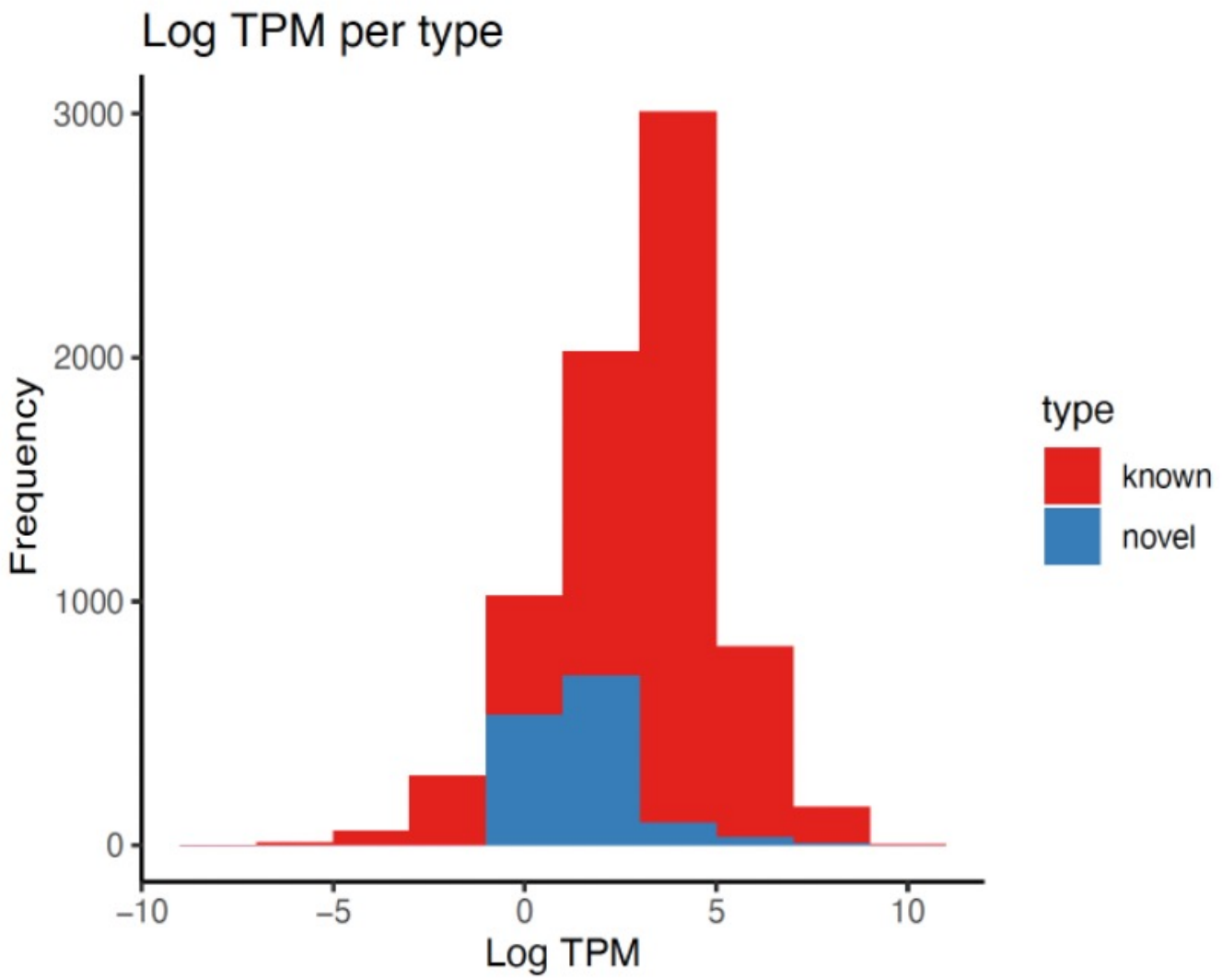


# TPM per type

Wilcoxon test

p-value < 2.2e-16

The differences between the distributions are **significant**

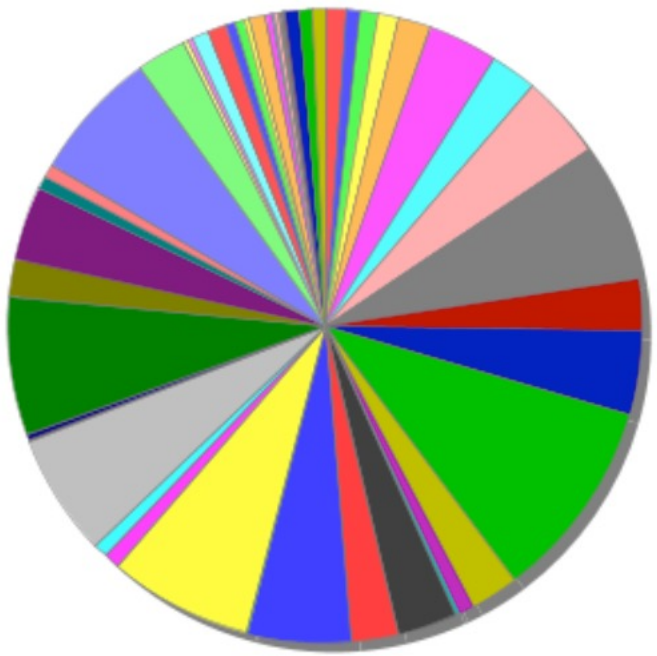


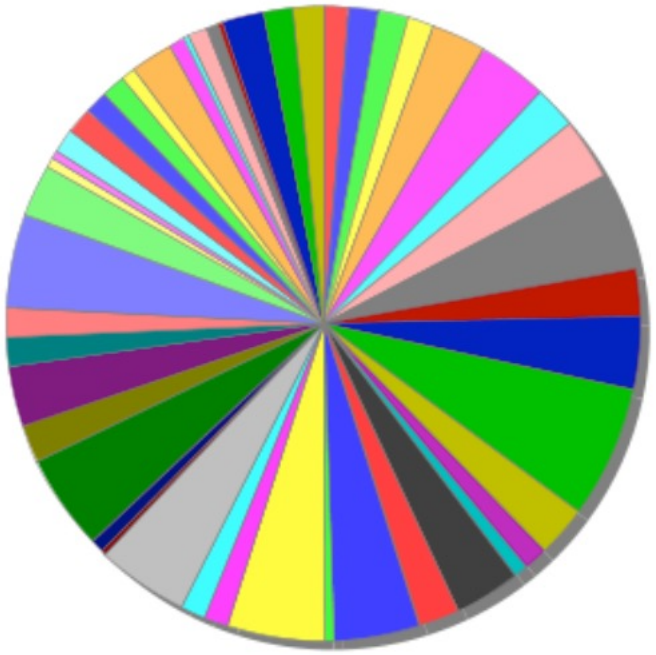
# BLASTX

| | |
|---|---|
| Transcripts | Known/novel |
| Blastx | Database → Proteins *S. cerevisiae* (114 MYA) |
| Results | - Hits (% identity > 90%) → Gene Ontology (biological process, molecular function, cellular component)<br>- No hits (novel) → Analysis |

Biological process

# KNOWN TRANSCRIPTS PANTHER ANALYSIS RESULT





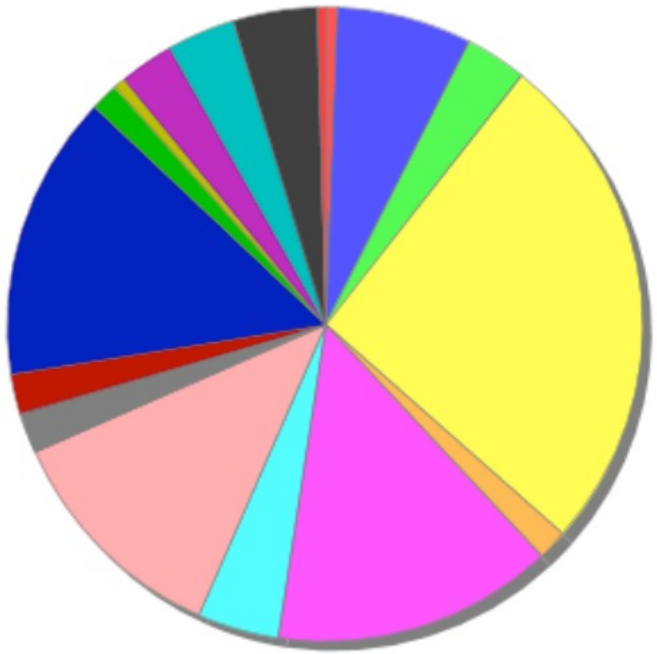
- RNA processing(GO:0006396)
- amide biosynthetic process(GO:0043604)
- cellular amide metabolic process(GO:0043603)
- cellular component assembly(GO:0022607)
- cellular component biogenesis(GO:0044085)
- cellular component organization or biogenesis(GO:0071840)
- cellular macromolecule biosynthetic process(GO:0034645)
- cellular macromolecule metabolic process(GO:0044260)
- cellular metabolic process(GO:0044237)
- cellular nitrogen compound biosynthetic process(GO:0044271
- cellular nitrogen compound metabolic process(GO:0034641)
- cellular process(GO:0009987)
- cellular protein metabolic process(GO:0044267)
- cellular protein-containing complex assembly(GO:0034622)
- cytoplasmic translation(GO:0002181)
- gene expression(GO:0010467)
- macromolecule biosynthetic process(GO:0009059)
- macromolecule metabolic process(GO:0043170)
- maturation of LSU-rRNA(GO:0000470)
- metabolic process(GO:0008152)
- ncRNA metabolic process(GO:0034660)
- ncRNA processing(GO:0034470)
- nitrogen compound metabolic process(GO:0006807)
- nucleosome assembly(GO:0006334)
- organelle assembly(GO:0070925)
- organic substance metabolic process(GO:0071704)
- organonitrogen compound biosynthetic process(GO:1901566)
- organonitrogen compound metabolic process(GO:1901564)
- peptide biosynthetic process(GO:0043043)
- peptide metabolic process(GO:0006518)
- primary metabolic process(GO:0044238)
- protein metabolic process(GO:0019538)
- protein modification by small protein conjugation(GO:0032446)
- protein ubiquitination(GO:0016567)
- protein-containing complex assembly(GO:0065003)
- protein-containing complex subunit organization(GO:0043933)
- rRNA metabolic process(GO:0016072)
- rRNA processing(GO:0006364)
- ribonucleoprotein complex assembly(GO:0022618)
- ribonucleoprotein complex biogenesis(GO:0022613)
- ribonucleoprotein complex subunit organization(GO:0071826)
- ribosomal large subunit assembly(GO:0000027)
- ribosomal large subunit biogenesis(GO:0042273)
- ribosomal small subunit biogenesis(GO:0042274)
- ribosomal subunit export from nucleus(GO:0000054)
- ribosome biogenesis(GO:0042254)
- translation(GO:0006412)
- translational elongation(GO:0006414)

*Saccharomyces cerevisiae* (REF)

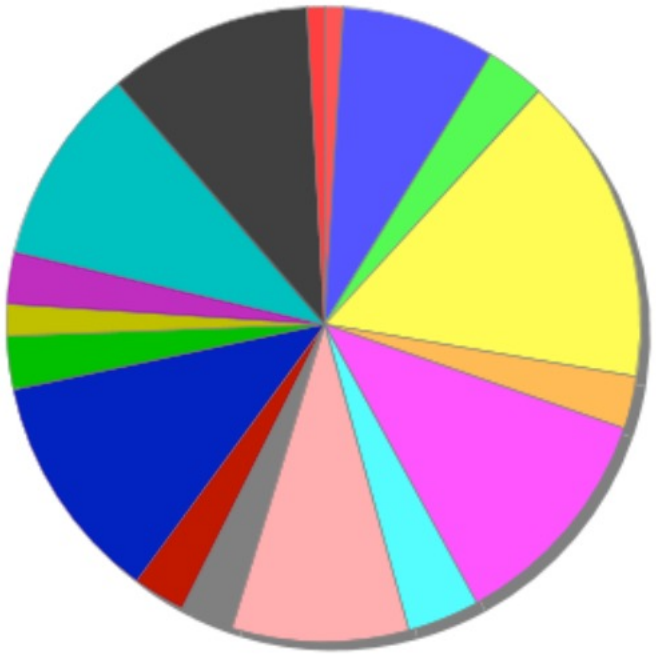
*Lachancea kluyveri* (upload)

Molecular function

# KNOWN TRANSCRIPTS PANTHER ANALYSIS RESULT



*Saccharomyces cerevisiae* (REF)

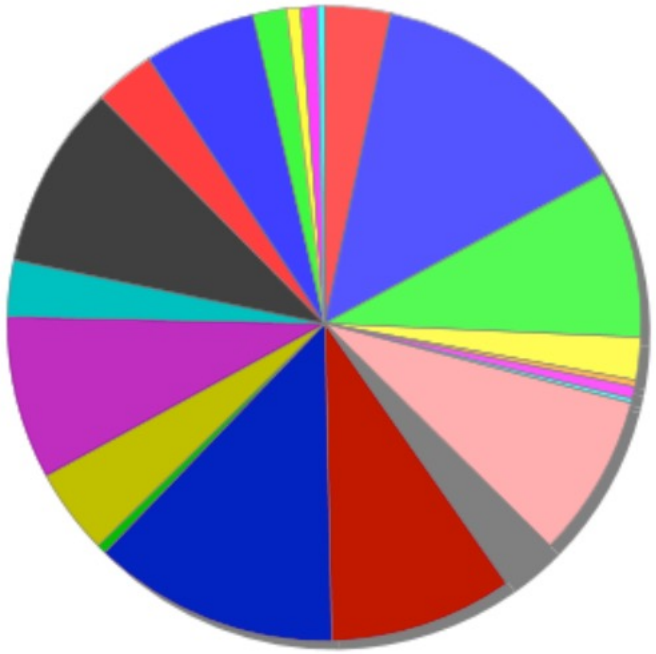


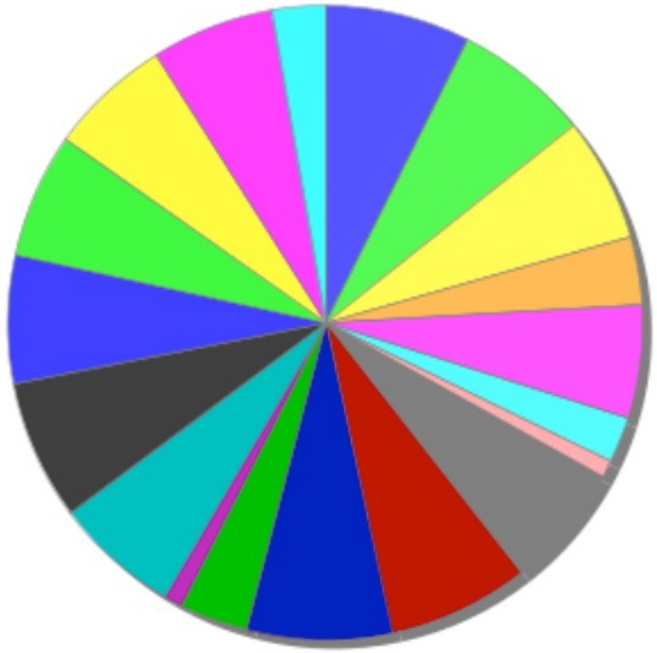
*Lachancea kluyveri* (upload)

- ATP binding(GO:0005524)
- RNA binding(GO:0003723)
- anion binding(GO:0043168)
- binding(GO:0005488)
- carbohydrate derivative binding(GO:0097367)
- heterocyclic compound binding(GO:1901363)
- ion binding(GO:0043167)
- nucleic acid binding(GO:0003676)
- nucleoside phosphate binding(GO:1901265)
- nucleotide binding(GO:0000166)
- organic cyclic compound binding(GO:0097159)
- purine nucleotide binding(GO:0017076)
- rRNA binding(GO:0019843)
- small molecule binding(GO:0036094)
- structural constituent of ribosome(GO:0003735)
- structural molecule activity(GO:0005198)
- ubiquitin-like protein conjugating enzyme activity(GO:0061650)

Cellular component

# KNOWN TRANSCRIPTS PANTHER ANALYSIS RESULT



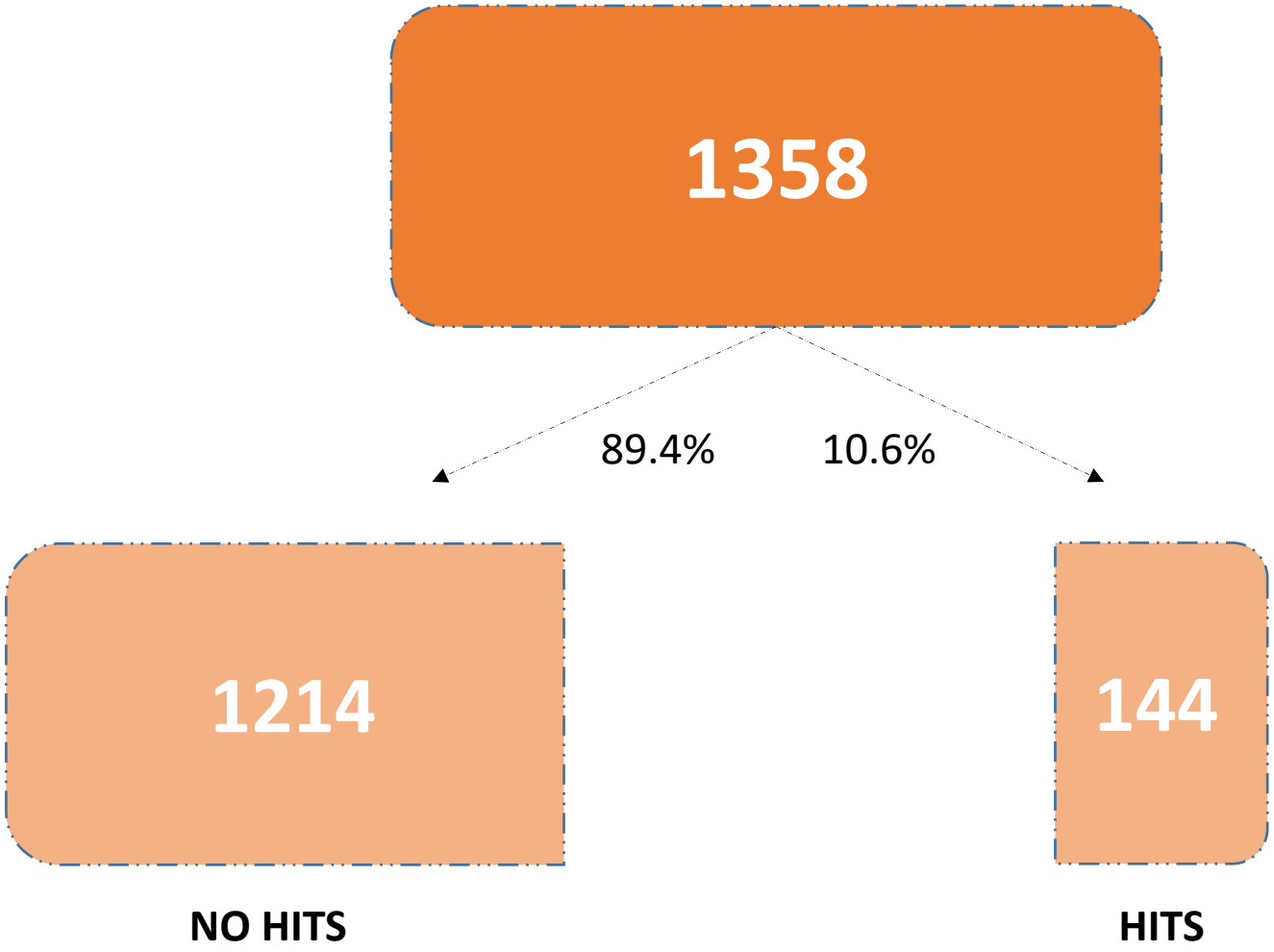


- cell periphery(GO:0071944)
- cellular anatomical entity(GO:0110165)
- cytoplasm(GO:0005737)
- cytosol(GO:0005829)
- cytosolic large ribosomal subunit(GO:0022625)
- cytosolic ribosome(GO:0022626)
- cytosolic small ribosomal subunit(GO:0022627)
- intracellular membrane-bounded organelle(GO:0043231)
- intracellular non-membrane-bounded organelle(GO:0043232)
- intracellular organelle(GO:0043229)
- intracellular(GO:0005622)
- large ribosomal subunit(GO:0015934)
- membrane(GO:0016020)
- membrane-bounded organelle(GO:0043227)
- non-membrane-bounded organelle(GO:0043228)
- organelle(GO:0043226)
- plasma membrane(GO:0005886)
- protein-containing complex(GO:0032991)
- ribonucleoprotein complex(GO:1990904)
- ribosomal subunit(GO:0044391)
- ribosome(GO:0005840)
- small ribosomal subunit(GO:0015935)

*Saccharomyces cerevisiae* (REF)

*Lachancea kluyveri* (upload)

# NOVEL TRANSCRIPTS

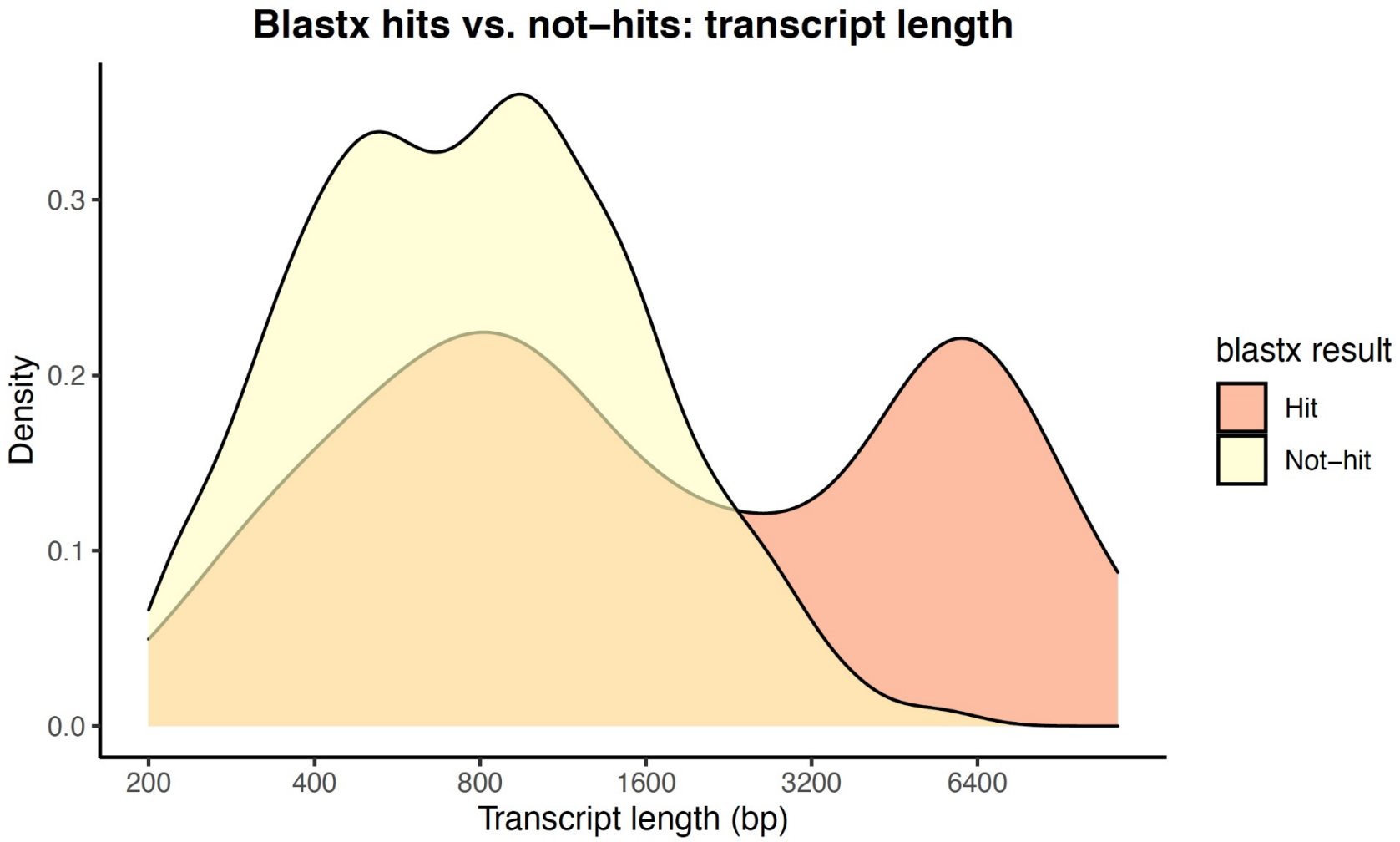


# Length of transcripts

Wilcoxon test

p-value = 3.711e-16

The differences between the distributions are **significant**

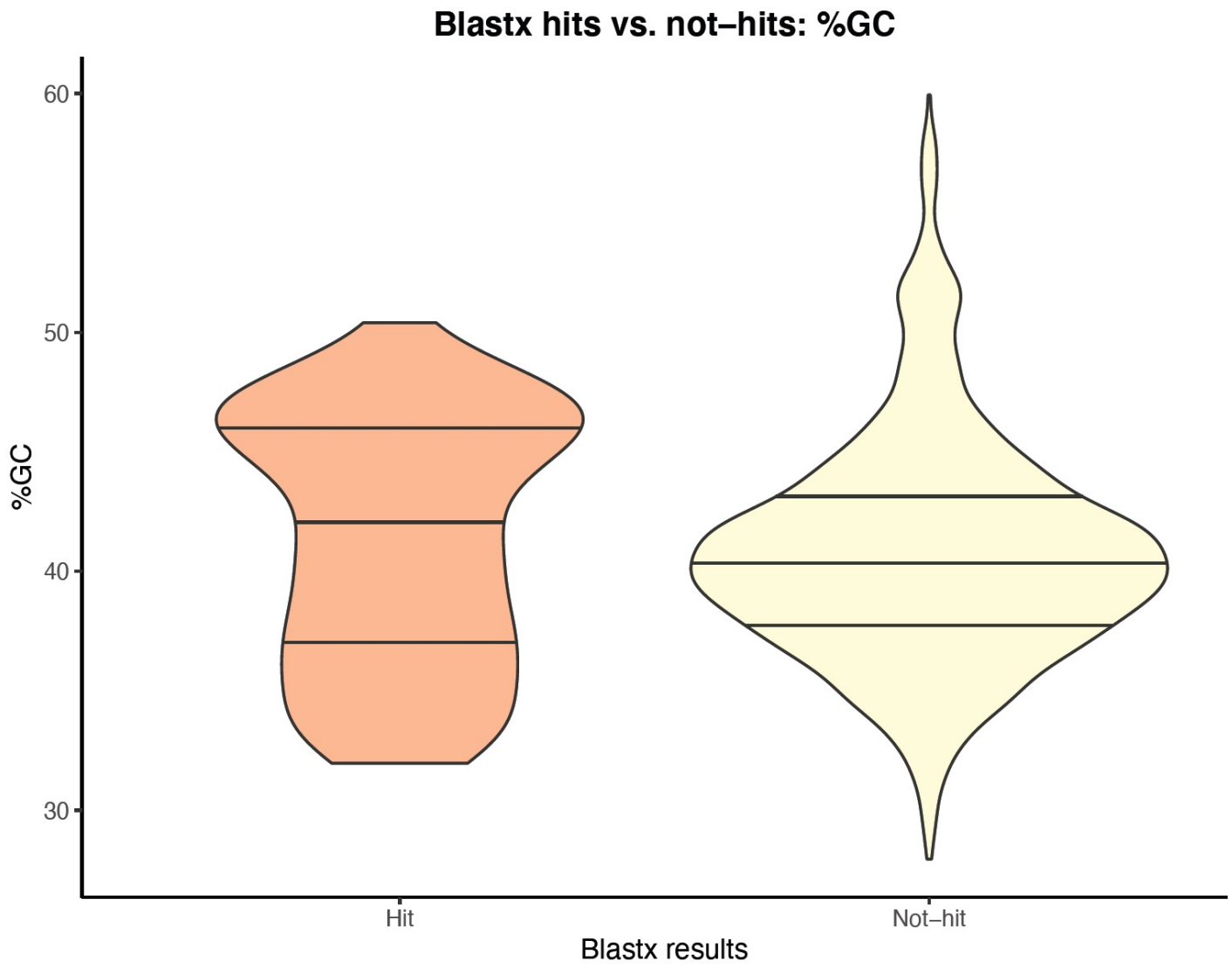


# Percentage of GC

Wilcoxon test

p-value = 0.08194

The differences between the distributions are **not significant**



Biological process

# NOVEL TRANSCRIPTS (HITS) PANTHER ANALYSIS RESULT

*Saccharomyces cerevisiae* (REF)

*Lachancea kluyveri* (upload)

- amide biosynthetic process(GO:0043604)
- cellular amide metabolic process(GO:0043603)
- peptide biosynthetic process(GO:0043043)
- peptide metabolic process(GO:0006518)
- translation(GO:0006412)
- translational elongation(GO:0006414)

Molecular function

# NOVEL TRANSCRIPTS (HITS) PANTHER ANALYSIS RESULT

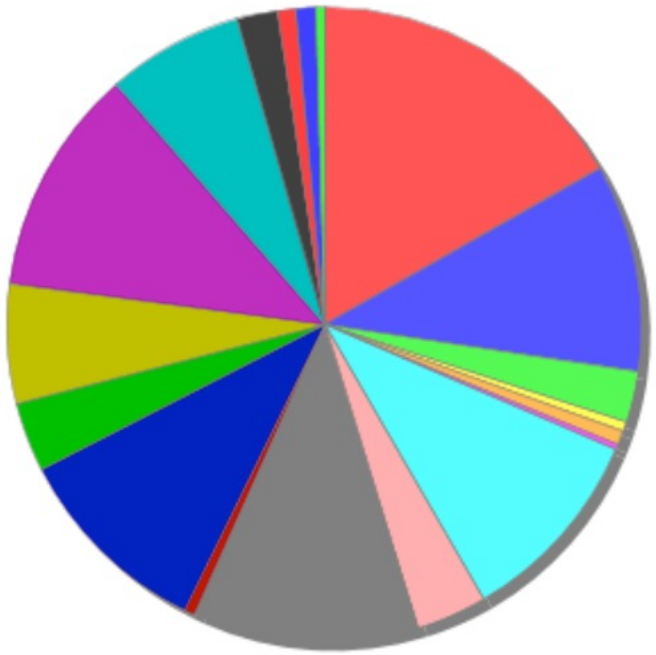
*Saccharomyces cerevisiae* (REF)

*Lachancea kluyveri* (upload)

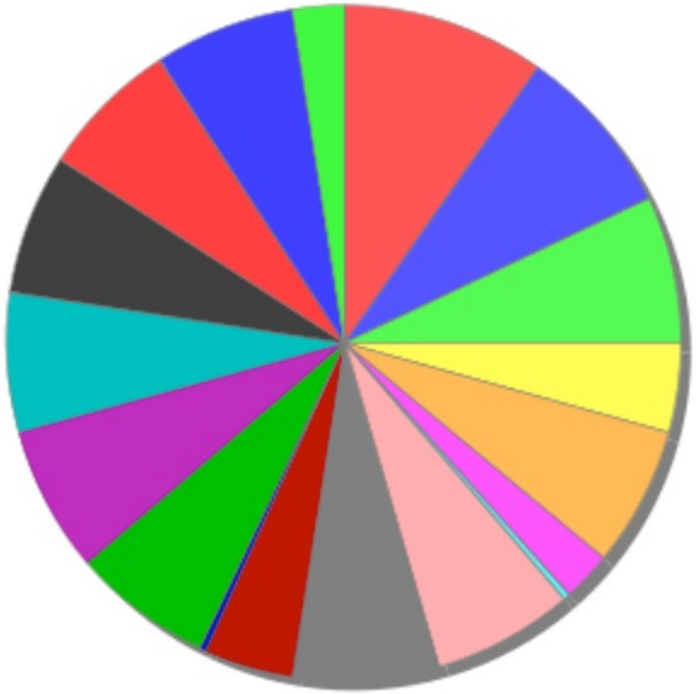
- RNA binding(GO:0003723)
- structural constituent of ribosome(GO:0003735)
- structural molecule activity(GO:0005198)

Cellular component

# NOVEL TRANSCRIPTS (HITS) PANTHER ANALYSIS RESULT



*Saccharomyces cerevisiae* (REF)



*Lachancea kluyveri* (upload)

- cellular anatomical entity(GO:0110165)
- cytoplasm(GO:0005737)
- cytosol(GO:0005829)
- cytosolic large ribosomal subunit(GO:0022625)
- cytosolic ribosome(GO:0022626)
- cytosolic small ribosomal subunit(GO:0022627)
- intracellular membrane-bounded organelle(GO:0043231)
- intracellular non-membrane-bounded organelle(GO:0043232)
- intracellular organelle(GO:0043229)
- large ribosomal subunit(GO:0015934)
- membrane-bounded organelle(GO:0043227)
- non-membrane-bounded organelle(GO:0043228)
- nucleus(GO:0005634)
- organelle(GO:0043226)
- protein-containing complex(GO:0032991)
- ribonucleoprotein complex(GO:1990904)
- ribosomal subunit(GO:0044391)
- ribosome(GO:0005840)
- small ribosomal subunit(GO:0015935)

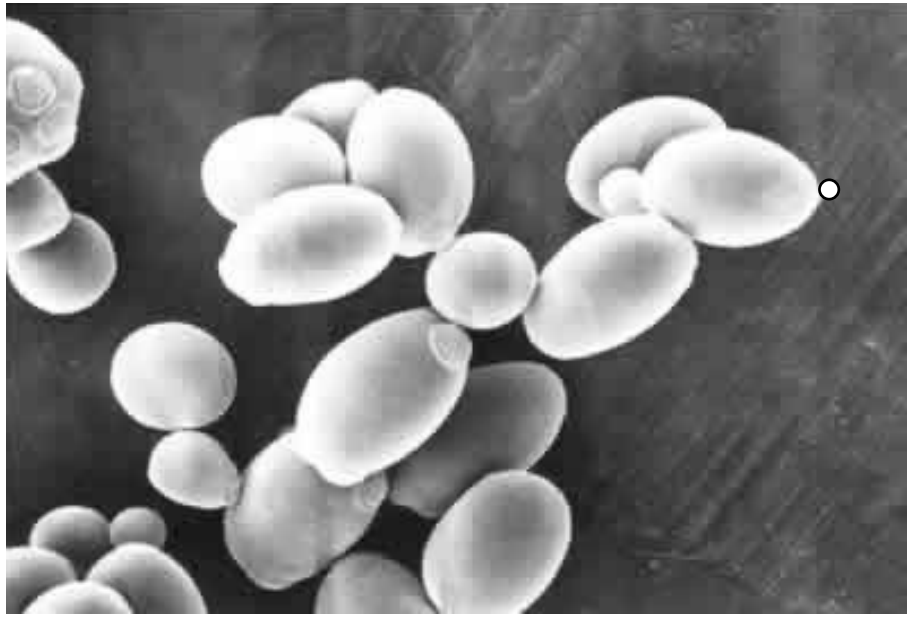
# CONCLUSION

- The difference in GC content between known and novel transcript is **significant**
- The difference in the length of the ORFs between known and novel transcript is **significant**
- The difference in the TPMs between known and novel transcript is **significant**
- The difference in the length of the novel transcripts between hits and not-hits **significant**

# LIMITATIONS

- Evolutionary distance between *S. cerevisiae* and *L. kluyveri*
- Little bibliography about *L.kluyveri*
- Wilcoxon test limitations

# THANK YOU!



Any question?