# OSCARS
## Open Science Clusters' Action for Research & Society

# Funded Project

# FAIRY: FAIRification of YEASTRACT+ to Support Yeast-based Bioeconomy and Health

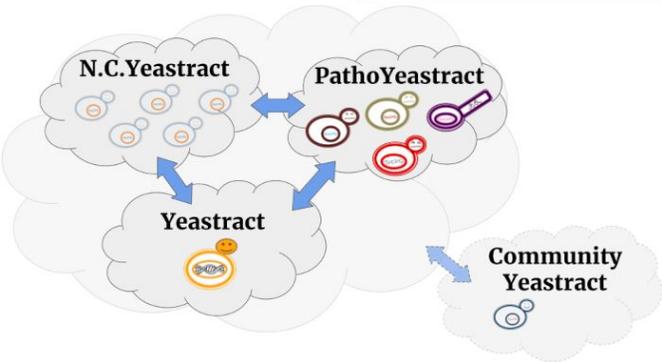Presenter: Jorge dos Santos Oliveira, INESC-ID, 0000-0001-6425-2176

## Implemented by

inesc id lisboa | TÉCNICO LISBOA | iBB Institute for Bioengineering and Biosciences

Funded by the European Union

OSCARS

## What problem(s) are you going to solve?

- YEASTRACT+ is a leading resource for yeast transcriptional regulation, but currently **lacks quantitative gene expression data**
- Many yeast **species remain poorly characterised**, impeding the understanding of their regulatory mechanisms and broader bioeconomic applications
- YEASTRACT+ currently does not provide a **programmatic API** for data access
- Current **data** is only **linked** to a few reference databases (e.g. SGD, GO)
- YEASTRACT+ is not fully compliant with **FAIR principles**

OSCARS

# What are you planning to do to solve the problem?

- Rely on EOSC cloud computing for **raw data curation** (RNA-seq data processing)
- Enrich regulatory information with **quantitative data**
- Develop a new **responsive portal** leveraging new curated data
  - With **programmatic access API** exposing data and functionalities
- Train machine learning models on cross-species regulatory data to:
  - Assess **data quality** (One class classification outlier detection)
  - **Infer** transcriptional **regulatory associations** in poorly-characterised yeasts
- Update and **Expand the body of data** gathered in the database, expanding it to further yeast species

**What will be the results and how do you plan to make them available to the broader community?**

FINAL PRODUCT: An expanded, quantitative and fully FAIR YEASTRACT+ database

Results will be made available through:
- Data and functionalities through our website (https://yeastract-plus.org/)
- Newly curated data available through API + flat files for bulk processing
- Publication of the upgraded YEASTRACT+ in the NAR Database Issue
- Participation in bioinformatics, systems biology and yeast biology scientific meetings

OSCARS

## Who is doing it?

INESC-ID
- Pedro T. Monteiro 0000-0002-7934-5495
- Jorge S. Oliveira 0000-0001-6425-2176
- Emanuel Gonçalves 0000-0002-9967-5205
- João Miranda, project grantee
- Maria Pereira, project grantee

IST - University of Lisbon
- Miguel C. Teixeira 0000-0002-5676-6174
- Isabel Sá-Correia 0000-0003-2208-5183
- Nuno Mira 0000-0001-7556-0385
- Diogo Couceiro, PhD student
- Luís Coutinho, project grantee

inesc id lisboa

IST TÉCNICO LISBOA

iBB Institute for Bioengineering and Biosciences