

CASE STUDY

How Ochre Bio turned a manual image processing workflow into an automated pipeline



The challenge

The Ochre Bio team comprises a diverse, interdisciplinary group of scientific experts working on chronic liver diseases. They faced a challenge in moving some of their work forward because of a dependency on a manual image-processing workflow, compounded by working across different locations and time zones.

One of their Taipei-based computational biologists would run this workflow to support a team of wet lab scientists based in New York City (NYC):

- Image pre-processing in Python to divide large images into smaller ones
- Characterizing image tiles using the open-source CellProfiler software
- Aggregating results and generating plots for visualization with R

This was a time-intensive, manual process for the Ochre Bio team:

- **Low visibility:** as with most orgs, infrastructure standardization across different teams posed challenges
- **Bottleneck:** could only be done by a single computational biologist
- **Time constraints:** a multi-step process across different time zones meant increased waiting times

“Code Ocean sped up our internal image pre-processing computational workflow by at least 10x, improving collaboration and productivity between our global teams across time zones and removing the need for complicated and painful cloud computing infrastructure set up. All these ultimately allowed our scientists to focus on interpretation of our causal biological data.”



Dimitris Polychronopoulos
Director of In Silico Biology, Ochre Bio

Key features used



Pipeline Builder



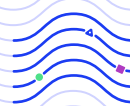
Compute Capsules



No-Code App builder



Data assets



The solution

Ochre Bio automated their previously manual workflow by transitioning it into Code Ocean:

1.

Created separate Compute Capsules for image preprocessing, running CellProfiler, and results visualization

2.

Created a Pipeline to run multiple instances of the CellProfiler Capsule in parallel on AWS Batch

3.

Created a no-code version of the pipeline to share with NYC-based colleagues to use independently

This was possible because Code Ocean integrates multiple different technologies:

- Capsules automatically generate a Dockerfile to enable fast environment creation (or, in this instance, allowed a publicly available Docker Container to be used directly without setup)
- Pipelines automatically generate a Nextflow script to enable parallel computing on AWS Batch
- Code Ocean is installed on Ochre Bio's own AWS instance, meaning results are immediately available and secure inside their own cloud environment
- Capsules support multiple languages, such as R and Python, for scripting visualization steps

The results



Productivity

Automating this process has freed up Ochre Bio's computational biologists to work on other tasks for the org



Democratized Access

Wet lab scientists are now enabled to run computational pipelines with minimal guidance from the dry lab, democratizing access to tools



Performance

Demanding computational steps are now automated and parallelized, making the overall workflow more robust and significantly faster



Visibility

The entire workflow is now visible to the rest of the Ochre Bio team

Key integrations



Amazon S3



AWS Batch



Cellprofiler



Nextflow



R



Python

