

OVERVIEW

Transcriptic is turning biology into an information technology driven by data, computation, and high-throughput robotics with the goal of substantially advancing drug discovery and synthetic biology. As the industry moves towards higher throughput experiments and increased automation, assay development of protocols remains a bottleneck. While automation offers improvements of magnitude in screening capacity, automating experiments requires searching an astronomically large experimental space which can take a trained scientist 2-6 months to develop. A naive search of protocol parameters is inadequate and a synergistic approach combining algorithms with automation is required. We offer an accessible programmatic interface which empowers scientists to efficiently search this space in a lights-out manner. Here, we demonstrate automated learning of a fluorophore's optimal spectra with a > 1000 fold improvement in time to solution over a brute force search. Our platform will enable AI to develop robust assays within days instead of months.

PLATFORM

To create a programmatic cloud lab, protocols are expressed in Autoprotocol, an unambiguous open standard for specifying and transferring experimental protocols. A protocol is composed of instructions, which are compiled down to device-specific commands. This group of commands are then dynamically scheduled with a constraint-programming solver, allowing efficient utilization of devices while respecting biological constraints. To fully unlock device capabilities, custom drivers are written to allow complete collection of metadata and reproducible execution. Upon command execution, data is provided to clients via our API and web application.

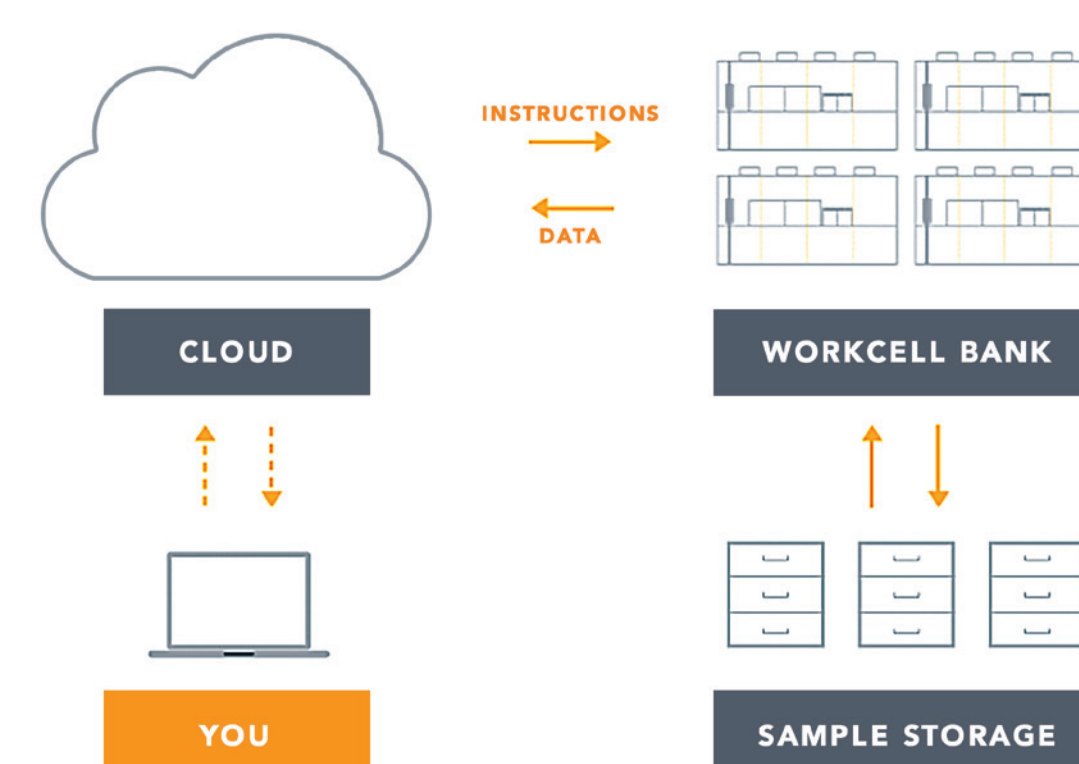


Figure 1. Transcriptic's user workflow.

SCIENCE ACCELERATED

The traditional scientific process is fragmented, slow and not robust. Transcriptic's unified web scientific interface abstracts away execution details allowing scientists and algorithms to harness the power of automation while focusing on experimental design and data analysis. From inventory management to data manipulation, users can expect an intuitive graphical flow or query data through command-line interfaces and APIs.

CLOSED-LOOP LIGHTS OUT BIOLOGY

A closed-loop workflow at Transcriptic follows a typical design-build-test flow. The scientist first designs their experiment structure/model in Autoprotocol-Python, instantiates a run using initial parameters and submits to the Transcriptic Common Lab Environment (TCLE).

Internally, TCLE builds and dynamically schedules the run to an available workcell for execution. TCLE coordinates the dispatching of commands to relevant devices and the generated data is fed back to the platform for post-processing and cleaning.

The structured data is returned to the user via API calls, which can automatically be fed into a desired model for generating new feature values. Finally, a new run can be instantiated either after a manual check or automatically submitted to TCLE, kicking off the next iteration of the execution loop.

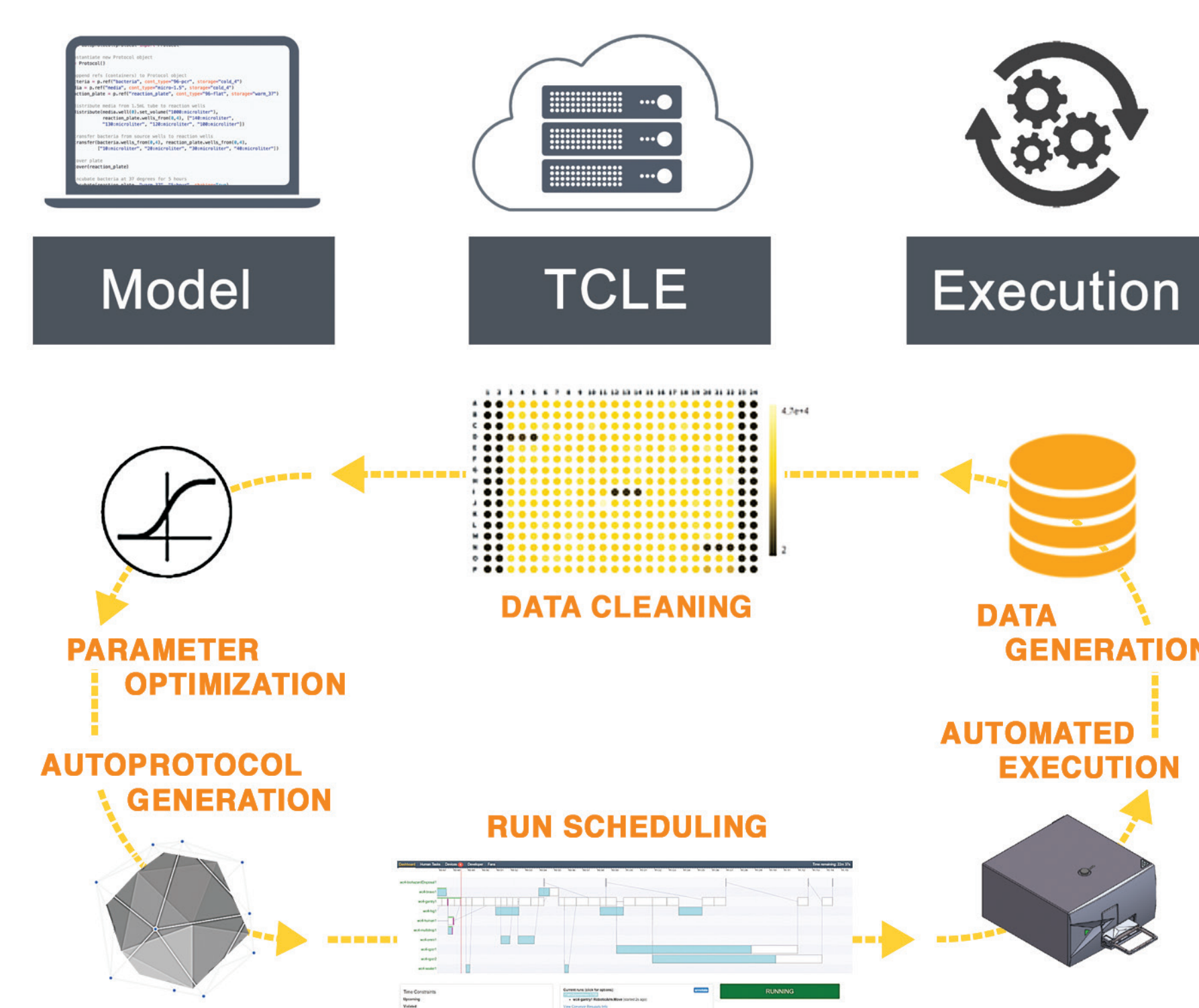


Figure 2. High level representation of data flows in a typical closed-loop experiment executed on the Transcriptic platform.

CASE STUDY

To demonstrate a closed-loop, machine-learning driven workflow, a common fluorophore (Alexa Fluor 594) was optimized for use with our plate readers. We optimized the fluorophore against PBS as a negative control and completed the entire process without human intervention. Each run was automatically submitted to our platform and plates were stored in incubators between runs.

Demonstrating ease-of-use, we developed a webhook-based Python workflow entirely from open-source libraries. The problem was modeled as a constrained non-linear minimization problem, with an objective normalized ratio of positive and negative control read intensities. A constrained trust-region algorithm (COBYLA) was then applied with heuristics such as multiple local starts.

RESULTS

We applied the algorithm and started 5 independent searches. The best performing search is plotted in **Figure 3**, where we see a relatively efficient convergence to (540:nm, 625.6:nm). Since this differs from the manufacturer's recommendation of (590:nm, 617:nm), a line-scan was executed. **Figure 4** reaffirms our conclusion, where we see that for the defined objective, the algorithm finds a better minima. To verify that the algorithm indeed found a global minimum, we conducted a coarse 5:nm scan of the landscape. This reaffirmed our results, showing that some of the traces did indeed get trapped in local minima, but our best trace found the global minimum.

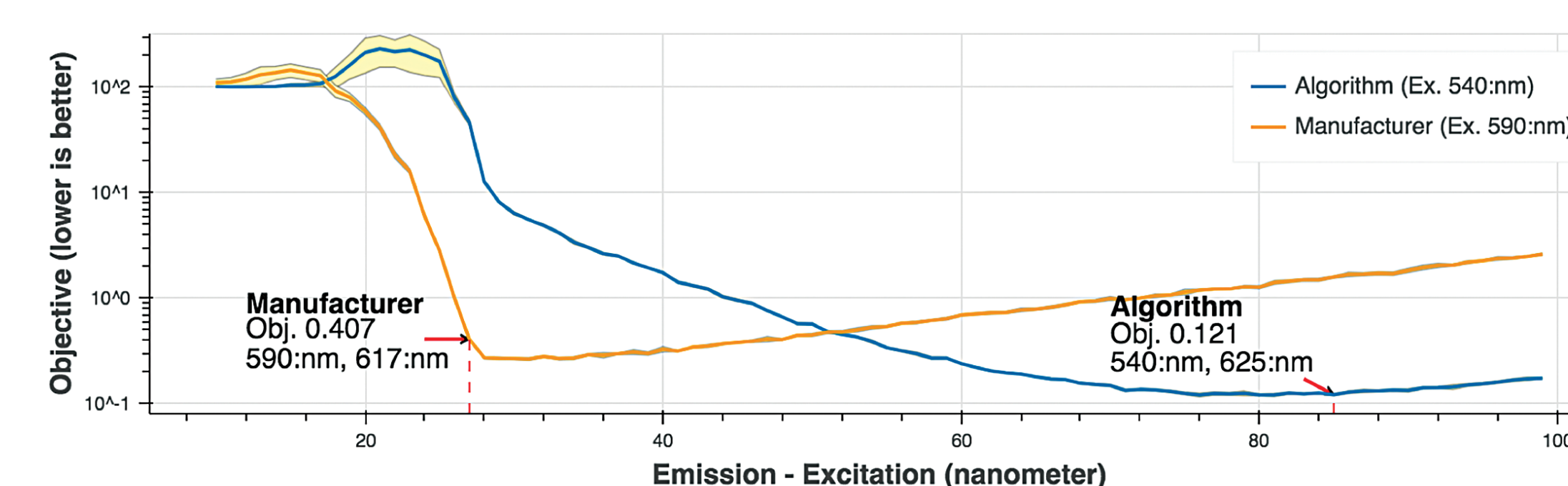


Figure 4. Line-scan comparison of manufacturer vs. optimal excitation wavelength.

Examining the overall time and reads taken, the COBYLA with heuristic approach took a total of 2.5 hours and 106 reads, as compared to a theoretical brute force search, which would have required 803,103 reads and >3000 hours to search the entire space. This translates to a 1000-fold improvement in search time, significantly decreasing the likelihood of reagent degradation effects.

Overall, our approach leads to a better optimum than manufacturer recommendations likely due to device limitations. This demonstrates that our platform, when coupled with machine learning, can account for device/reagent and specific factors during the experimental design process and enable researchers to quickly discover optimal assay parameters specific to their protocols.

CONCLUSION

Here we have demonstrated the power of machine learning on the Transcriptic platform, vastly reducing time to results and automatically learning fundamental physical properties. Applying this approach to other aspects of assay development will lead to dramatic reductions of protocol onboarding time and enable the exploration of automation friendly protocols in days. This will be required to keep pace with the demands of today's science and empower the lab of the future.

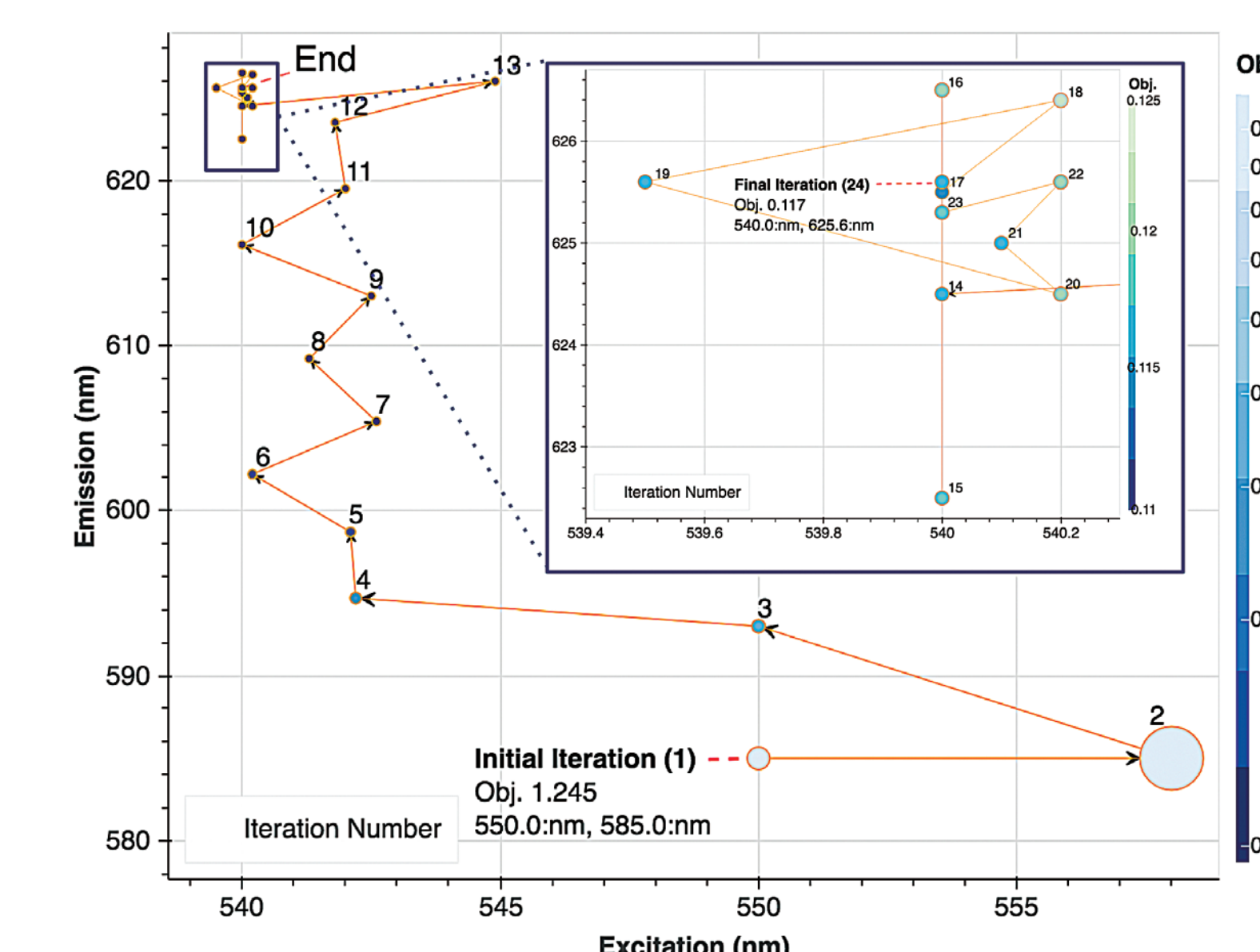


Figure 3. Progress of best-performing trace, progressing from initial (550:nm, 590:nm) to final optimal point (590:nm, 617:nm).

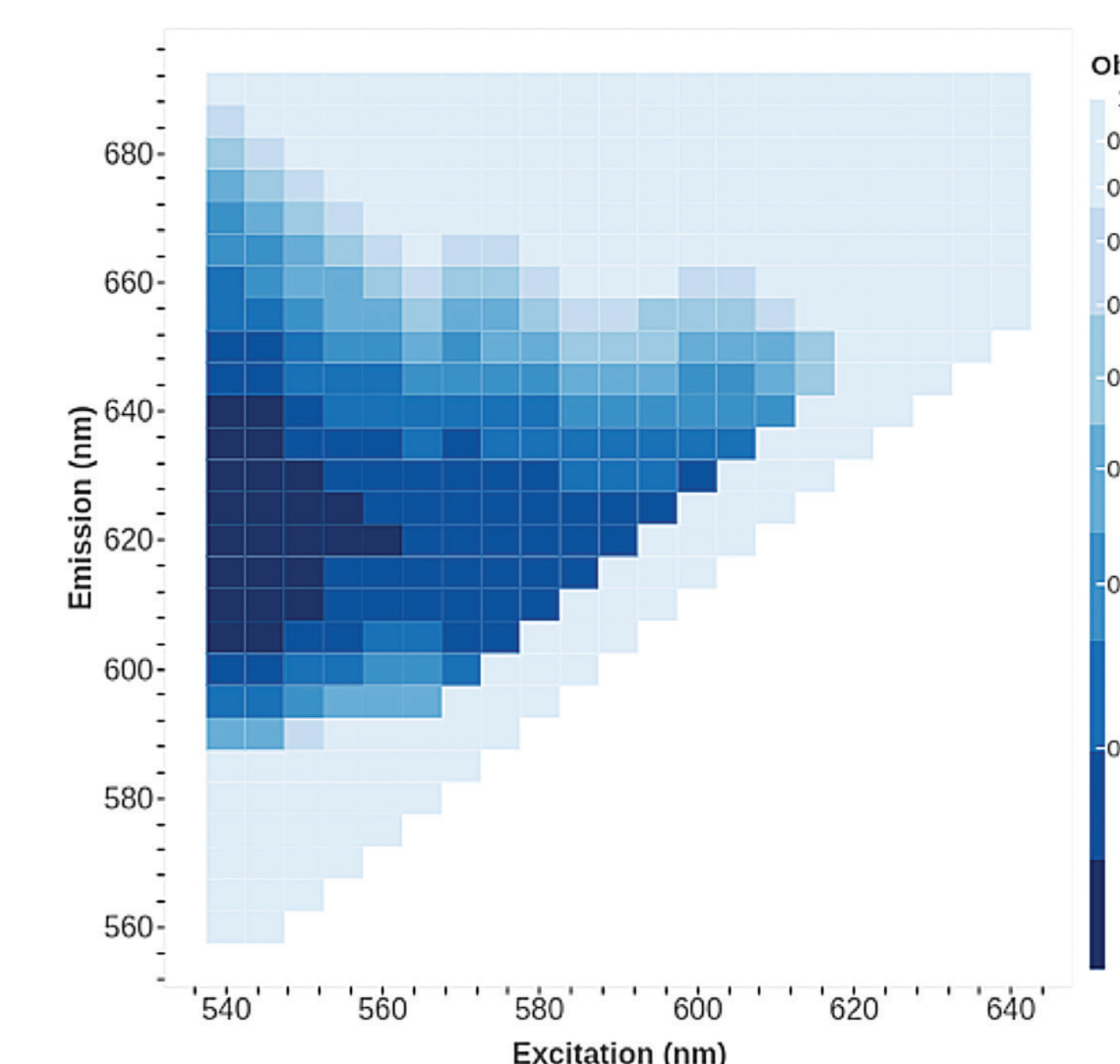


Figure 5. Coarse grid search of the parameter landscape. Objective values above 1 are truncated.