

Accelerating Clone Selection with Transfer Learning and DataHowLab

Clone selection is a critical, time-intensive step in biologics development. Scientists are under significant time pressure to identify a single clone that consistently produces the desired therapeutic protein at high yields while meeting critical quality attributes (CQAs).

Learning from historical clone screenings to assess new process clones and accelerate selection with DataHow's transfer learning

To accelerate clone selection for a new molecule, DataHow leverages its transfer learning capabilities to harness knowledge from historical clone screening data, reducing the reliance on studying extensive experimental data for new process clones at different process conditions.

Assessing the Impact of Transfer Learning via DataHowLab's Hybrid Models

To illustrate the application of transfer learning to clone selection, three cases highlighted from different industrial projects are presented:

Case 1: Accelerating process design via transfer learning. Deducing the optimal clone **AND** optimal conditions with clone data at standard conditions only + historical clone data.

Case 2: Determining which historical clone data improves process understanding of a new clone. The impact of each historical clone is assessed individually and as a collective.

Case 3: Early forecasting and detection of leading clones to accelerate clone selection. Predicting final (day 14) values with data from the first 7 days only.

What is Transfer Learning?

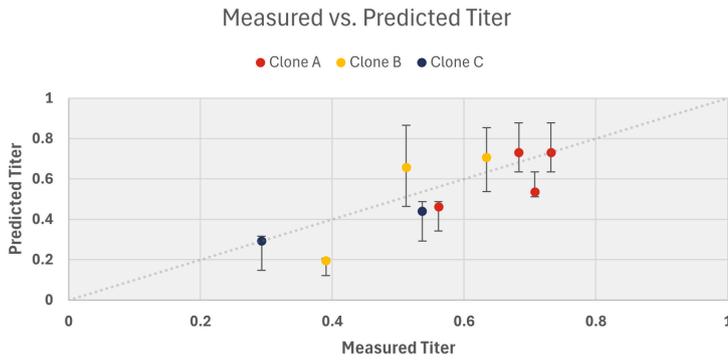
Transfer learning is a machine learning method that leverages data from historical processes to transfer knowledge horizontally to new projects.



This machine learning capability can be applied across molecules (historical insights applied to a new project) and across scales (transferring insight from small to large scale within a project).

This approach accelerates clone selection and transforms historical development data into a central development asset that delivers increasing benefits over time.

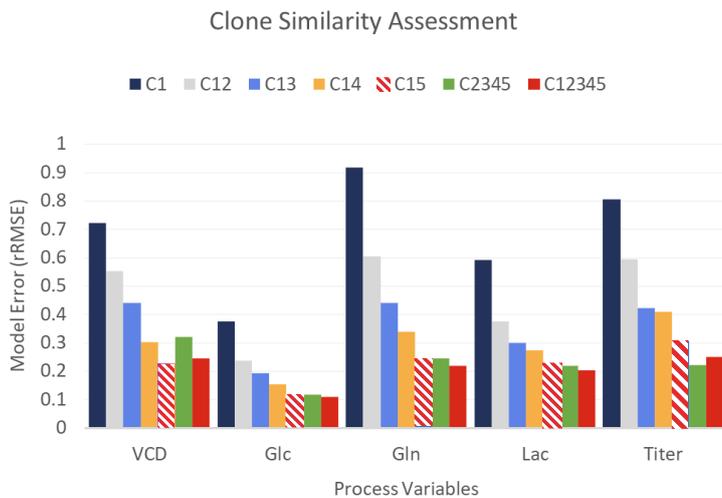
CASE 1 - Accelerating process design via transfer learning



For a new molecule (D), where 3 top clones were tested under **standard conditions only**, the best-performing clone (A) and its corresponding optimal conditions could be identified by leveraging data from the data of 3 older molecules, each of which had up to 5 conditions tested.

By Integrating such approach into the development workflow, a reduction of up to 80% is anticipated.

CASE 2 - Determining which historical clone data improves process understanding of a new clone

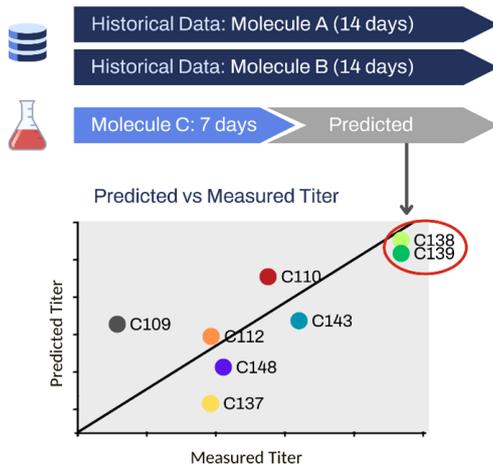


To understand the impact of four historical clones (C2 - C5) to support the development of a new molecule (C1), each was individually included in the training set, along with 7 runs (at different conditions) of clone 1.

Results showed that incorporating data from Clone 5 alone enabled forecasting of Clone 1's behavior under unseen conditions, highlighting the importance of selecting representative clones.

While single clones rarely worsened the model's performance in this case, it is important to perform a similarity assessment to ensure that the right historical data is included.

CASE 3 - Early forecasting and detection of leading clones to accelerate clone selection



The goal in this application was to use the historical data of two molecules, A and B, to identify the best clone for molecule C as early as possible.

For each clone of molecule C, data was collected until day 7, including titer, with the remaining 7 days propagated using a hybrid model.

The model showed good predictive performance against measured values, and the ability to clearly identify the most promising clones, C138 and C139, based on the final titer, the key driver in this process for clone selection.



Historical Data as a Development Asset



Streamlined Clone Selection and Design



Reduction of Experiments



Acceleration of Clone Selection