Predicting clone composition from relatives

MANTIS DISCOVERY

MANTIS · MIT CSAIL

OCTOBER 2026

01

The question

Sequencing a cell destroys it, so nobody can measure the same cell twice. In a lineage-linked assay every cell carries an inherited barcode, so the descendants of one ancestor can be recognised as relatives afterwards. We ask when measuring some of a clone’s cells helps predict the state of its other cells, and what a rule that fails is missing: fitting within the study, reuse of a rule from another study, or simply more cells captured.

The practical question underneath is whether measuring a relative is worth what it costs. Every cell spent on a relative is a cell not spent elsewhere, so a rule has to beat the simplest alternative, which is to ignore the relative and predict the same number for every clone.

02

How we are approaching it

- We score a fixed expression signature in four wells and move a two-bin predictor between the two control wells without retuning it.
- In two barcoding data sets, of 657 and 1,801 clones, we predict a clone’s composition in one group from the other, comparing a rule fitted here, a rule frozen from the other study, and simply copying the relative’s value.
- We cap the relatives measured per clone at 1, 3, 5, 10 or all of them, with exact formulas for the expected loss.
- We refit on 10 to 400 training clones and keep between 5% and all of the existing annotation.
- We compare every rule against a constant fitted on the same clones, so a gain means the relative carried information the constant did not.

03

The signature check

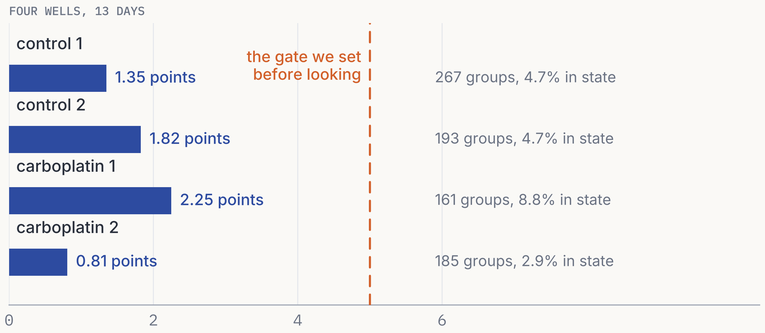


FIGURE 1. Where the project started. In four wells we asked how much more often a cell is in the state of interest when a relative is, against cells paired at random. All four gaps fall under the five-point gate we set before looking.

04

What we have so far

With one relative measured per clone, a rule fitted within the study beats a constant by 11.9% and 11.2% in the two directions. Copying the relative’s value across has a negative mean gain at every cap.

All 24 frozen configurations carried over from the other study lose to a constant fitted on the target, although each of them beats the constant from their own source.

Fitting the rule needs clones. With 10 training clones per fold the mean gain is negative, and it turns positive at about 19.

05

What calibration buys

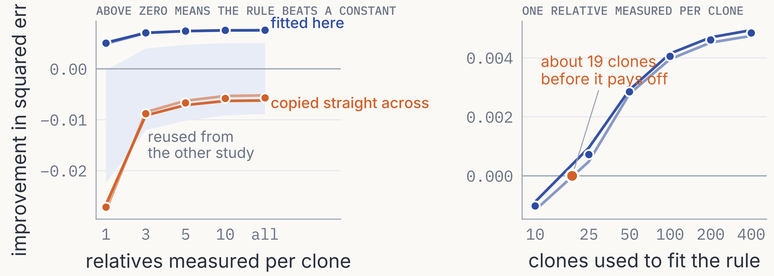


FIGURE 2. How much each rule beats predicting a constant, in squared error, as more relatives are measured per clone. A rule fitted within the study helps from one relative onward. Copying the relative’s value never helps on average, and rules frozen from the other study span the shaded band.

06

What would settle it

- Independent cultures with functional labels, and capture recorded as the experiment runs, rather than a further search for a signature.
- Repeated measurements of the same state, with a model of capture, to fix the noise floor.

07

Open questions

- Because the original capture is unknown and the cultures are not independent, none of this supports a claim about population risk or about why a treatment works.
- We cannot say how much of the remaining error is biological rather than technical.
- Nothing here shows an assay on one cell, or where the sequencing effort is best spent.