Open Access

What is a Cytosol Landscape?

Sharon Newman[1]

[1]b.next, San Francisco, California, 94107, USA

AbstractKeywords Cytosol, Cytosol Landscape, Discovery Plate, composition, design of experiments

Nucleus Developer Notes

Published Sep 09, 2026

Correspondence to
Sharon Newman
sharon@bnext.bio

Open Access

Copyright © 2026 Newman. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

Open Source

The materials and processes described in this article are distributed under the terms of the CERN Open Hardware License Permissive 2.0, which enables reusers to make and convey any component or device that apply the information made explicitly available under this license, so long as all notices are retained.

1. Integration changes the recipe

We often talk about adding capabilities to a synthetic cell as if they were independent parts: add an energy module, add a sensor, add a control module. But every time we add something to the system, we also perturb it.

Imagine that we begin with a baseline cytosol composition and sweep the concentration of magnesium. We find the magnesium concentration that gives us the highest protein yield, and we call that our optimum. Then, when we try to integrate a new module, we have to add, at minimum, another component - perhaps a protein. Adding such a protein might decrease our yield. But why?

Did the protein degrade? Or did adding it simply move the system away from its previous optimum?

A new protein can change the environment in which the rest of the system operates. It may bind magnesium, consume energy, compete for transcriptional or translational resources, introduce a new buffer, or alter the effective concentrations of other components. Integration also requires finding the recipe that makes both modules happy. We can see this in a simplified example (Figure 1): adding a protein can both shift the optimum we already knew and add a new dimension that we now need to explore.

What if, instead of starting over every time we added a module, we had a map showing where the system works, where it is fragile, and which direction to move next?

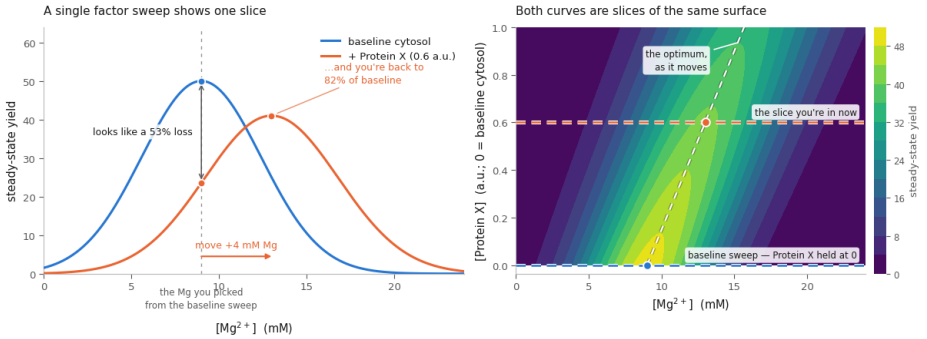


Figure 1. ***Adding a module can move the optimum. left:*** *Steady-state protein yield against magnesium concentration (blue) for a base cytosol composition. When we add a new module or protein, using the same cytosol composition makes the new module appear to perform poorly. Moving along the magnesium axis reveals that the system can still perform well. It now just operates in a different region (simulated data).* ***right:*** *adding a new protein adds a new dimension to the Cytosol landscape that we now also need to consider (simulated data). Note that landscapes may not be as clean as the simulated data.*

2. What is a Cytosol Landscape?

A Cytosol Landscape is a map between **composition** and **performance**.

Each point on the map represents a recipe: particular concentrations of magnesium, ribosomes, tRNA, proteins, energy substrates, and other components. The height of the landscape represents an outcome we care about—protein yield, production rate, productive lifetime, reproducibility, cost, or a combination of these.

A conventional magnesium sweep is already a simple slice through this landscape. We hold nearly everything constant, move along one compositional axis, and measure how the output changes. If we vary magnesium and one protein together, we begin to see a two-dimensional surface. Add more components, and the same idea continues into dimensions we can no longer draw (nor reasonably consider in our minds).

Thinking in terms of landscapes changes the questions we can ask:

- **Integration:** When we add a protein or module, what shifts? Which other components need to change to recover—or improve—performance?
- **Robustness:** Are we operating on a broad plateau, where small measurement errors barely matter, or near a cliff, where a small change causes performance to drop?
- **Cost:** Can we use less of an expensive reagent without reducing performance?
- **Mechanism:** Which resources are becoming limiting as the system changes?
- **Optimization:** Which small set of experiments would be most informative to run next?

The highest point on the landscape (global/local optima) may not even be the best place to build cells. A sharp peak can give excellent yield under perfectly controlled conditions but fail when a reagent concentration is miscalculated and off by 5%. A slightly lower but broader plateau may be far more reproducible—and far easier to extend with new modules.

3. We are not mapping all of Cytosol

The idea of a landscape can sound more complete than it really is. Cytosol contains more than 100 components. A dense map of every possible combination would require an absurd number of experiments. Even with robots, we cannot fill a 100-plus-dimensional space densely enough to know every hill, valley, and cliff. Fortunately, we do not need to. The practical goal is not to create a universal atlas of Cytosol. It is to build **question-driven maps of the regions that matter**.

For example:

- What happens in the neighborhood around our current working cytosol recipe?
- Which components have wide tolerances, and which create steep cliffs?
- How does one energy or control module shift the local landscape?
- Can we reduce the concentration of an expensive component while retaining useful performance?
- Which measurements would help distinguish energy limitation from transcriptional or translational limitation?

A one-component sweep traces a line through the landscape. A two-component sweep draws a surface. When we want to explore several components at once, sampling methods such as Latin hypercubes or Sobol sequences help spread a limited number of experiments across a larger space. These methods don't reveal the whole landscape, but rather, they give us strategically placed landmarks from which we can begin to infer its shape for specific questions

we have. A Cytosol Landscape is a growing collection of compatible experiments, local maps, and models organized around decisions.

4. Discovery Plates: surveying the interesting regions

We built a workflow to make what we're naming Discovery Plates to explore these smaller, question-driven regions. Each Discovery Plate begins with a question and a defined region of composition space. We generate a set of informative recipes, determine whether those recipes are physically possible to make from the available stocks, assemble many low-volume reactions, and measure their performance over time.

We have built a concentration-planning pipeline that translates desired final compositions into feasible master-mix and per-well dispensing instructions. It also surfaces when a theoretically interesting condition cannot actually be made—for example, because a protein stock is too dilute, a required volume is below the pipetting limit, or there is not enough headroom left in the reaction.

Most of our Discovery Plate reactions so far have been run at 2 µL to 10 µL, allowing us to test more conditions with less Cytosol material. We have partnered with LabCraft to use their Plex, an automation robot, to dispense very small volumes on the order of 10 nL, making multidimensional protein and small-molecule titrations possible at a scale that would be extremely difficult to perform by hand.

At such small volumes, physical constraints are tricky. Sticky or viscous reagents such as ribosomes or super concentrated proteins can be difficult to dispense. Reagents can wick at the printhead or clog a cartridge due to static charge or other weird physics at that scale thats hard for us to predict. As with most things, putting theory into actual wet-lab is super tough. The hard part is not only choosing clever conditions, but also ensuring that the recipe we designed is the recipe the robot actually made. This is a big part of what we've been working on.

5. From landscapes to models, and back again

As we sample more components, the resulting landscape becomes harder to visualize and think about with human eyes and brains, and this is when we lean towards models!

Different kinds of models help us extract different kinds of value from the data:

- **Statistical analyses** can identify influential components, interactions, sensitivities, robust regions, and likely cliffs within the region we measured.
- **Surrogate models** can estimate performance between measured points and help us choose informative conditions for the next experimental round.
- **Bayesian optimization** can balance testing promising conditions with exploring uncertain regions.
- **Mechanistic models** can test explanations for why the landscape has its shape—for example, energy depletion, nucleotide imbalance, magnesium availability, or competition for transcriptional and translational resources.

The useful loop is:

Experiment arrow.r landscape arrow.r model arrow.r next experiment

A landscape can tell us *what changed*. A model can help us ask *why it changed*, whether that explanation predicts another perturbation, and what measurement would test it.

Our current modeling effort is deliberately not trying to produce a perfect universal synthetic-cell model. The near-term goal is to build small, experimentally grounded models that can answer concrete questions: What resource is likely to run out first? How does expression burden change the energy budget? What happens when an energy module is introduced? Which Discovery Plate perturbations would distinguish between plausible bottlenecks? We'll share a bit more about our modeling efforts in a future note.

This also changes how we design experiments. If we measure only final GFP yield, many mechanisms can look the same. Time-resolved protein production, mRNA measurements, quantitative fluorescence standards, and carefully chosen controls give models something more informative to explain.

6. What have we done so far?

So far, we've built the tools and generated the first datasets needed to map useful regions of cytosol composition space.

6.1. We have built the map-making machinery

- An Opentrons + LabCraft workflow for assembling low-volume composition experiments.
- A concentration pipeline that converts target compositions into feasible master-mix and titration instructions.
- Space-filling condition generators for exploring several variables at once.
- Initial response-surface and workflows for selecting subsequent conditions, (some in collaboration with FindWhatMatters that we have some DevNotes on).

6.2. We have begun measuring slices of the landscape

Here are some examples of experiments we've run as we learn more about how to appropriately make these large probes into the fitness landscape.

- A 2D sweep of Magnesium and Potassium showing lower magnesium and higher potassium correlates with better signal
- A 3D exploration of magnesium, Ribosomes, and T7RNAP showed that magnesium had the largest effect on steady-state GFP signal in the tested region.
- We then continued on to increasing number of factors we explored including: a 5D exploration of different salts and PMix (top) and a 7D exploration of Amino Acids, ATP, GTP, Creatine Phosphate, Mg, K, tRNA (middle), and an 8D exploration of tRNA, Mg, GTP, ATP, DNA, Creatine Phosphate, PPK, PolyP (for some baseline work related to PPK Energy Module integration) (bottom).

Higher-dimensional experiments are much harder to visualize—and much easier to over-interpret. We have analyzed these datasets using our pipeline, but before drawing or sharing conclusions, we are validating the underlying dispensing, controls, and comparability across experiments. As that foundation improves, these datasets can become inputs to models that help us identify sensitivities, test mechanisms, and choose the next experiments.

6.3. We have learned what is required for the map to be trustworthy

Some experimental rounds produced useful signals. Others exposed limitations in the platform: reagent depletion, sticky proteins, spatial and volume effects, or requested dispense volumes that did not always match actual volumes when trying a new protein. Additionally, many of these were also done before the days of us getting good standards implemented, which also makes it hard for us to combine and compare experiments.

```
<Axes: xlabel='[K] (mM)', ylabel='[Mg] (mM)'>
```

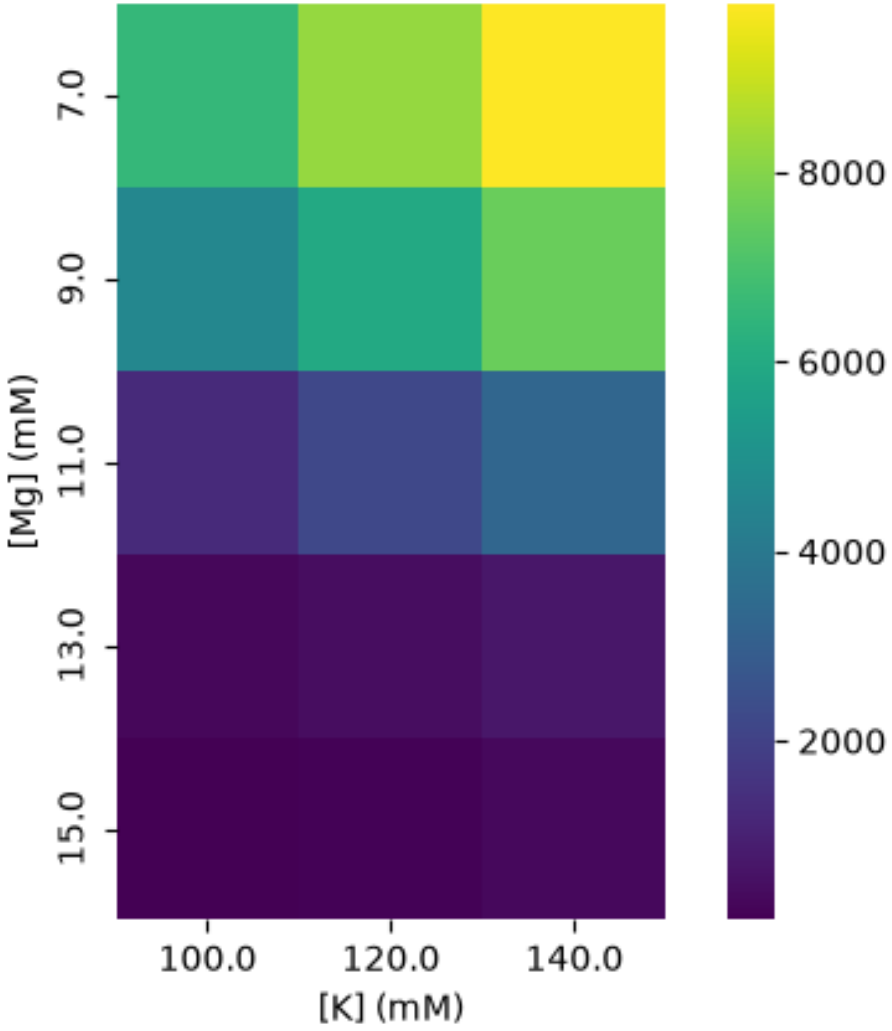


Figure 2. *Steady state fluorescence from a magnesium × potassium sweep on base cytosol.*

Our current priorities therefore emphasize:

1. Characterizing robot reproducibility across the reagents we actually use.
2. Standardizing controls and quantitative measurements across plates and batches.
3. Adding measurements—such as mRNA readouts—that make the data more useful for modeling.
4. Running focused guard-banding experiments around our working cytosol to identify tolerances and cliffs.
5. Connecting each experiment to a decision: a formulation choice, QC specification, model parameter, or next module-integration step.

This work also sits within a growing field. Within cell-free systems, recent work from Bernard-Lapeyre and Meyer combines automated liquid handling with active learning to explore a PURE compositional landscape. Their results argue against a single universal optimum: high-performing compositions can depend on DNA concentration, architecture, and component

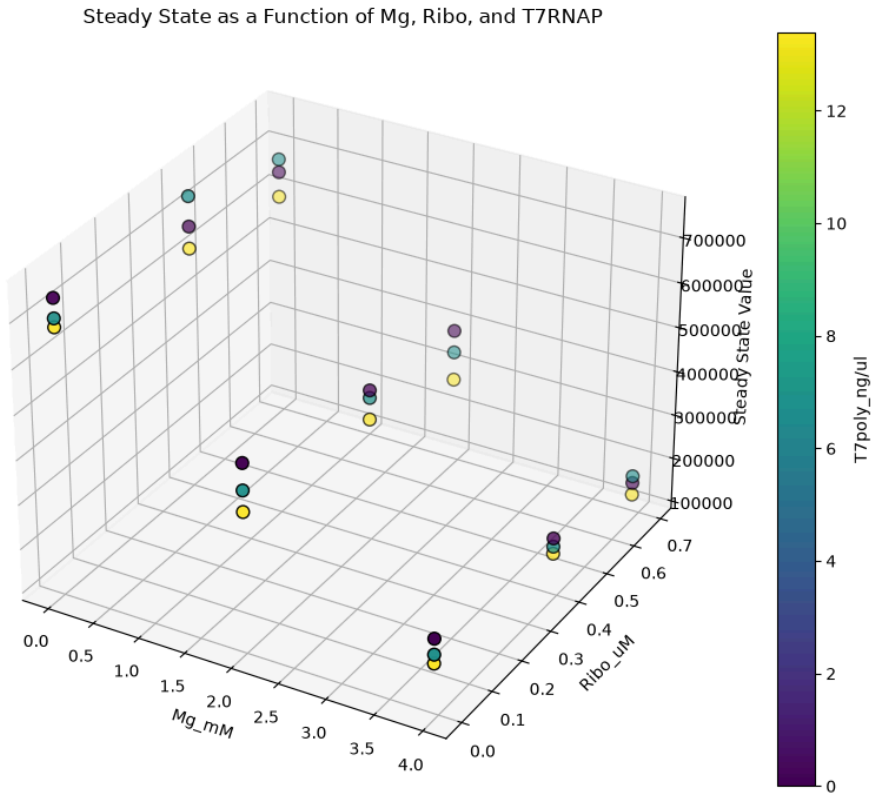


Figure 3. *Steady state fluorescence as a function of magnesium, ribosomes and T7RNAP, replicate means, coloured by T7RNAP.*

batch. That reinforces the value of adaptable, context-specific maps rather than one fixed recipe.

7. Where do we map next?

We're excited to work through these challenges so that we can ask increasingly powerful questions about module burden, energy demand, productive lifetime, and integration. We'd love to chat if you're working on similar things, or think a region is particular interesting to probe! We leave you with a question:

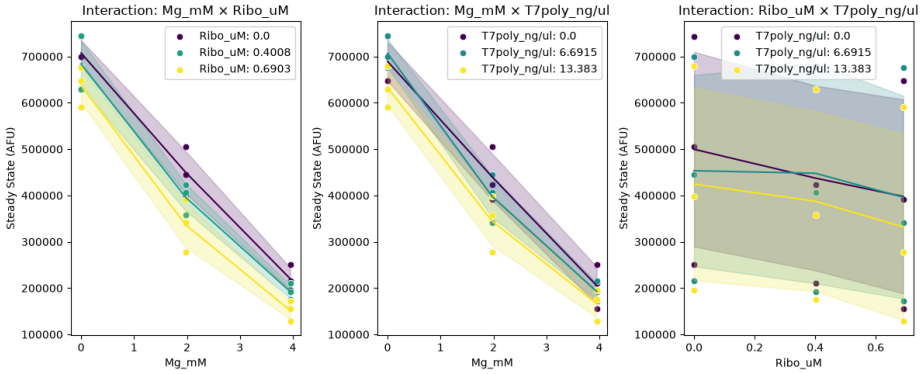


Figure 4. *Pairwise interaction plots for the same three factors*

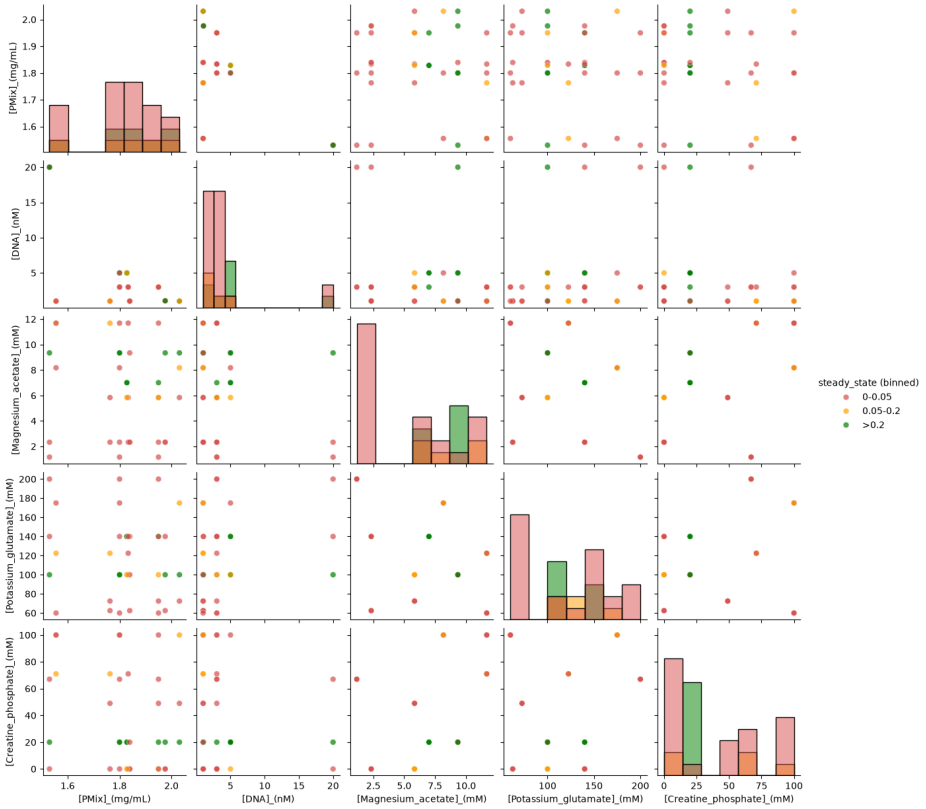


Figure 5. *5-factor Discovery Plate: PMix, DNA, magnesium acetate, potassium glutamate and creatine phosphate, coloured by binned steady state.*

We cannot map the entire landscape. So how much of it do we need to map to make the next integration decision with confidence?

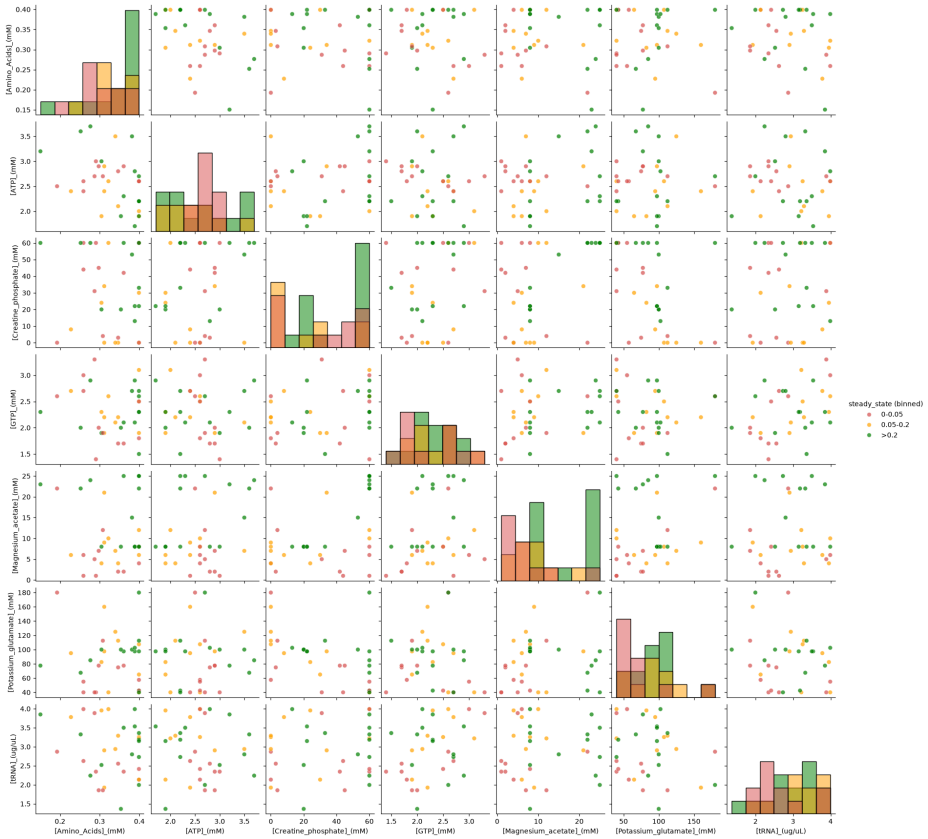


Figure 6. *7-factor Discovery Plate: amino acids, ATP, creatine phosphate, GTP, magnesium acetate, potassium glutamate and tRNA, coloured by binned steady state.*

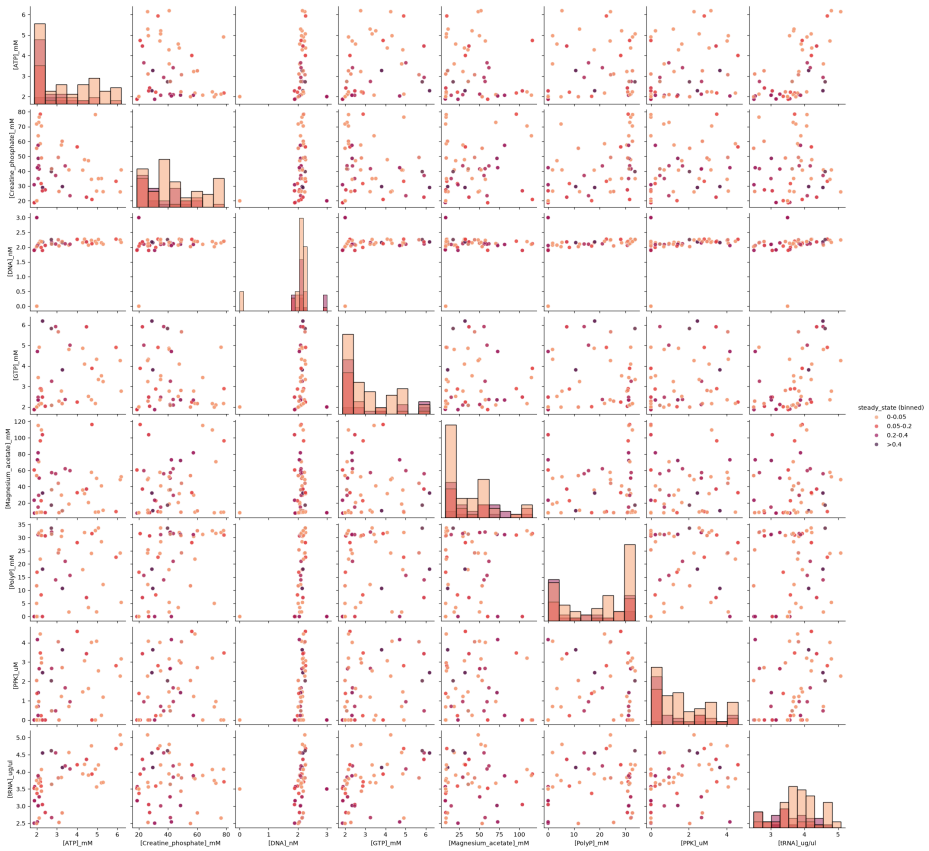


Figure 7. *8-factor Discovery Plate: ATP, creatine phosphate, DNA, GTP, magnesium acetate, PolyP, PPK and tRNA, coloured by binned steady state.*