



All Sizes Matter!

Saurin Bipin Parikh, Nelson Castillo Coelho, and Anne-Ruxandra Carvunis
 Department of Computational & Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
 Center for Evolutionary Biology & Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

Introduction

Colony-based high-throughput screens have become increasingly popular due to the availability of model organism mutant collections and robotic equipment that help facilitate quicker and larger screens. Such screens have been used to explore gene-gene^{1,2}, gene-environment³ and protein-protein⁴ interactions, and very recently evolutionary biology⁵.

A typical experiment generates multiple agar plates which are used to record growth using imagery that is then analyzed to result in a quantitative output of colony size as a proxy for fitness (fig. 1).

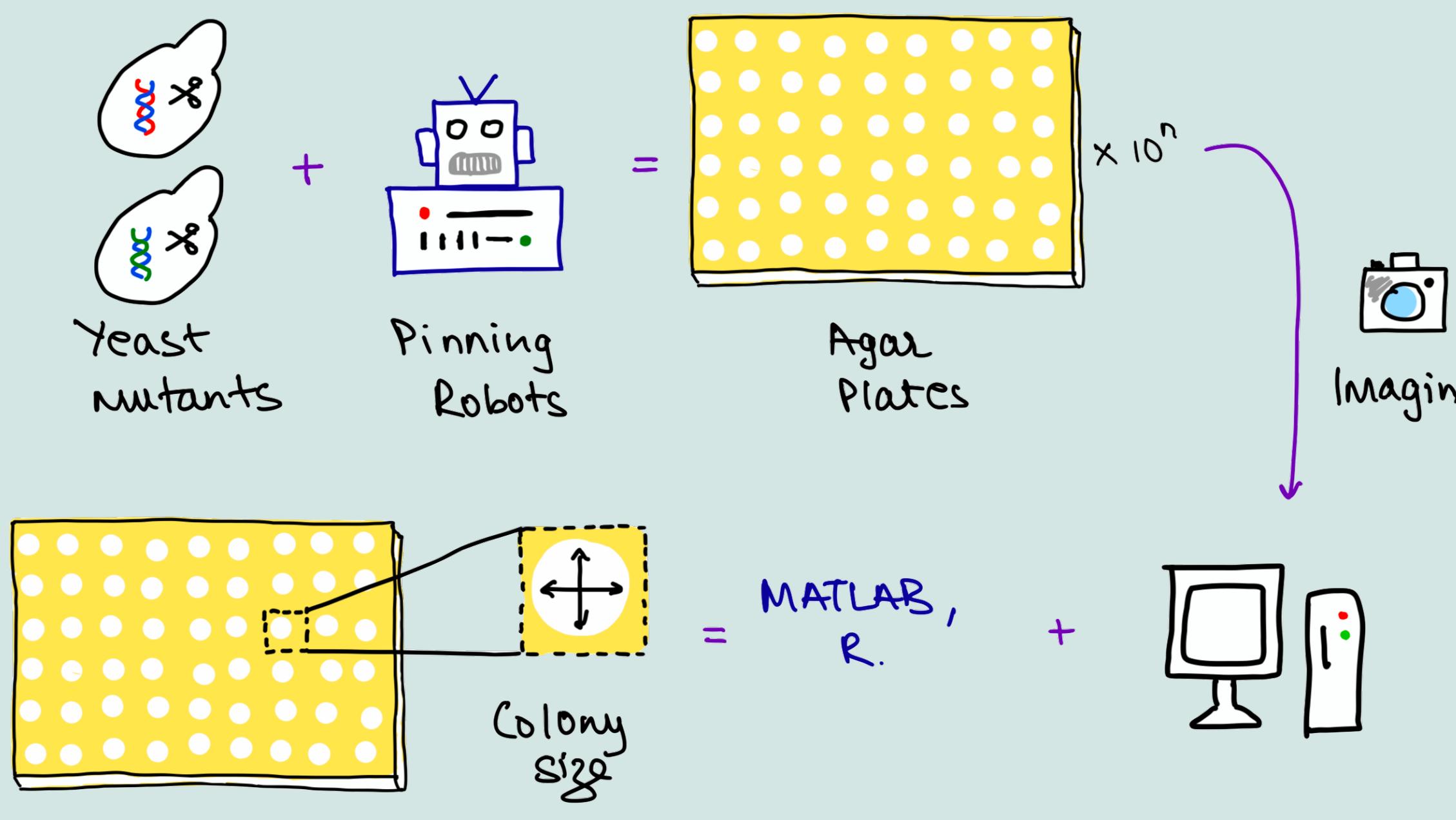


Figure 1. Schematic of a colony-based high-throughput screen.

The environment can rarely be maintained constant within a single plate, let alone across multiple plates, leading to systematic colony growth differences that arise due to technical reasons (fig. 2).

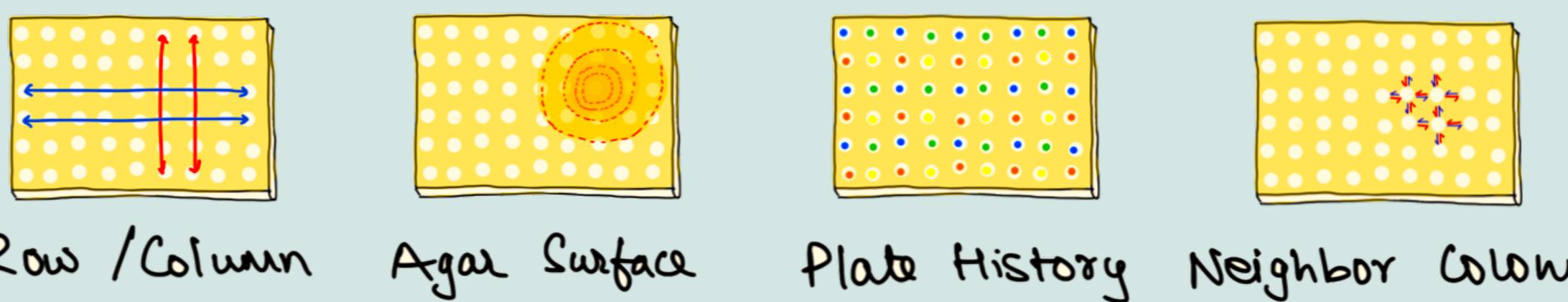


Figure 2. Various spatial biases that affect colony growth.

Existing methods^{6,7,8,9} that correct for these biases assume that most genetic perturbations have minimal effects on fitness and that most screens would contain a normal distribution of fitness effects (fig. 3).

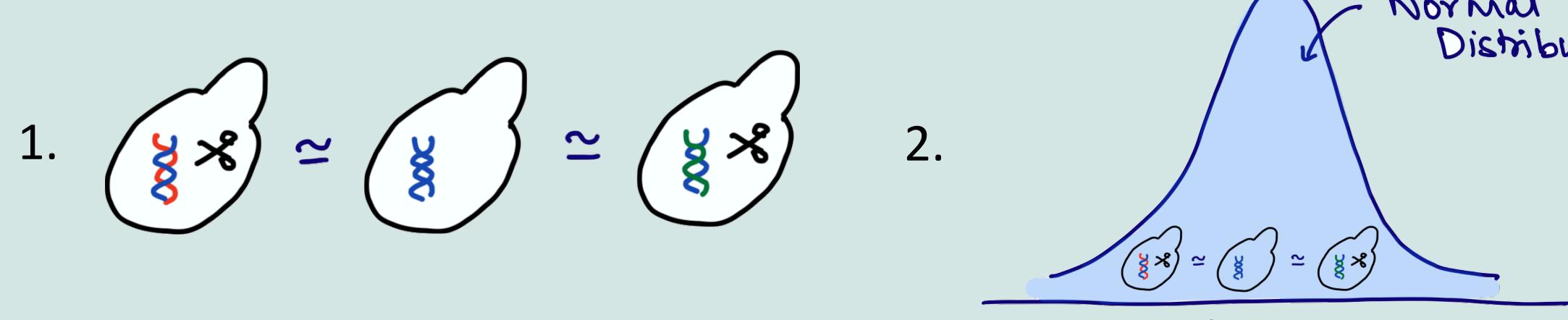


Figure 3. Two assumptions of existing methods for spatial bias correction

Although true for most genome-scale screens these assumptions are restrictive and limit the scope of the scientific examination.

Hence, we developed the LI Detector, an experimental and computational framework that is free from such restrictive assumptions.

References

- Roguev A, et al. Identification and reworking of functional modules revealed by an epistasis map in fission yeast. *Science*. 2008;322(5900):405-10.
- Costanzo M, et al. The genetic landscape of a cell. *Science*. 2010;327(5964):425-31.
- Pearson AB, et al. Integration of chemical genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat Biotechnol*. 2004;22(1):62-9.
- Yang F, et al. Development and application of a recombination-based library versus library high-throughput yeast two-hybrid (RLY2H) screening system. *Nucleic Acids Res*. 2018;46(3):e17.
- Vakirlis N, et al. De novo emergence of adaptive membrane proteins from thymine-rich genomic sequences. *Nat Commun*. 2019; (in press)
- Bear J, et al. Development of ultra-high-density screening tools for microbial "omics". *PLoS One*. 2014;9(1):e85177.
- Wagh O, et al. SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic Acids Res*. 2013;41(Web Server issue):W591-6.
- Young BP, et al. Balony: a software package for analysis of data generated by synthetic genetic array experiments. *BMC Bioinformatics*. 2013;14:354.
- Dittmar JC, et al. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics*. 2010;11:353.

Method

The experimental design (fig. 4A) introduces a reference colony grid on every plate that is utilized by the analytical pipeline (fig. 4B) to identify and correct for spatial bias.

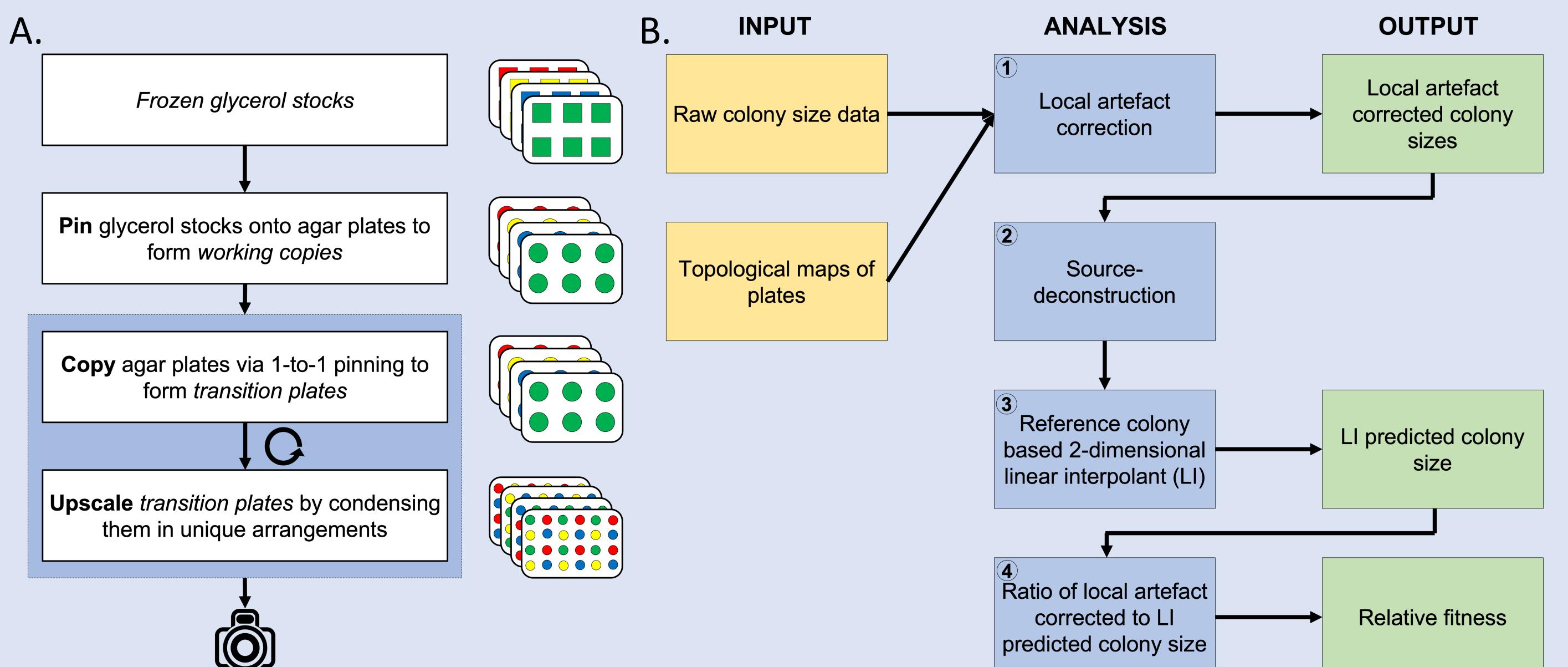


Figure 4. Components of LI Detector framework: A. experimental and B. analytical pipeline.

In order to evaluate this framework, four 384-well density glycerol stocks of a common laboratory yeast strain (FY4) were used to undergo serial pinning on rich media (YPD) as per the above experimental pipeline (fig. 4A). All colonies coming from one randomly chosen stock were mocked as references and those from the other three as queries. The resulting 6144 density agar plates (fig. 5A) had a fourth of all colonies as references (fig. 5B).

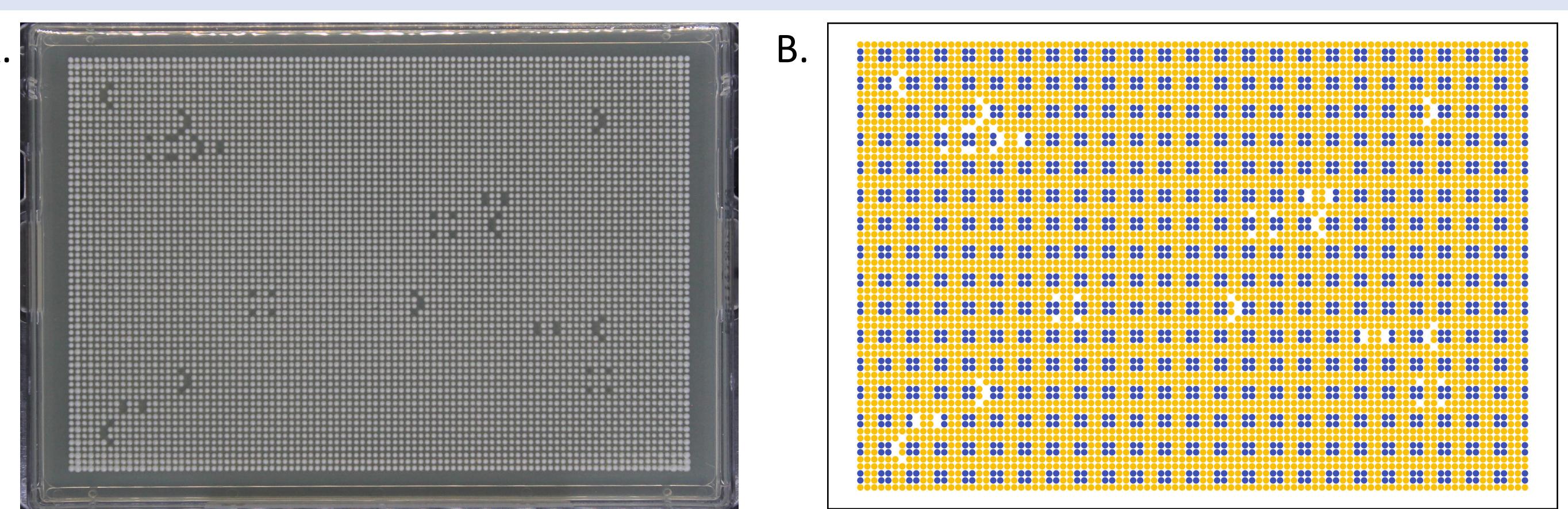


Figure 5. A. Agar plate with 6144 colony density, and B. colony layout.

The concept of virtual plates as condition positive dataset.

Reference Time Point (hours)	0.00	0.15	0.19	0.23	0.34	0.45	0.53	0.67	0.71	0.79	0.88	0.95	1
11.04	0.15	0.19	0.24	0.35	0.47	0.55	0.7	0.75	0.83	0.93	1	1.05	
9.97	0.15	0.19	0.24	0.35	0.47	0.55	0.7	0.75	0.83	0.93	1	1.07	1.13
8.96	0.16	0.21	0.25	0.38	0.5	0.6	0.76	0.81	0.9	1	1.11	1.19	1.25
7.85	0.18	0.23	0.28	0.42	0.56	0.66	0.84	0.9	1	1.11	1.24	1.33	1.4
6.88	0.2	0.26	0.32	0.47	0.63	0.74	0.93	1	1.07	1.19	1.32	1.42	1.49
6.14	0.22	0.28	0.34	0.5	0.67	0.79	1	1.07	1.19	1.32	1.42	1.49	
4.89	0.28	0.35	0.43	0.64	0.85	1	1.27	1.36	1.51	1.68	1.81	1.89	
4.02	0.33	0.42	0.5	0.75	1	1.18	1.5	1.6	1.78	1.98	2.13	2.23	
2.9	0.43	0.55	0.67	1	1.33	1.57	1.99	2.12	2.36	2.63	2.83	2.97	
1.38	0.64	0.82	1	1.49	1.98	2.33	2.96	3.17	3.52	3.92	4.22	4.42	
0.97	1	1.21	1.81	2.4	2.83	3.6	3.85	4.29	4.77	5.13	5.37		
0	1	1.27	1.56	2.3	3.08	3.6	4.58	4.91	5.45	6.08	6.53	6.83	

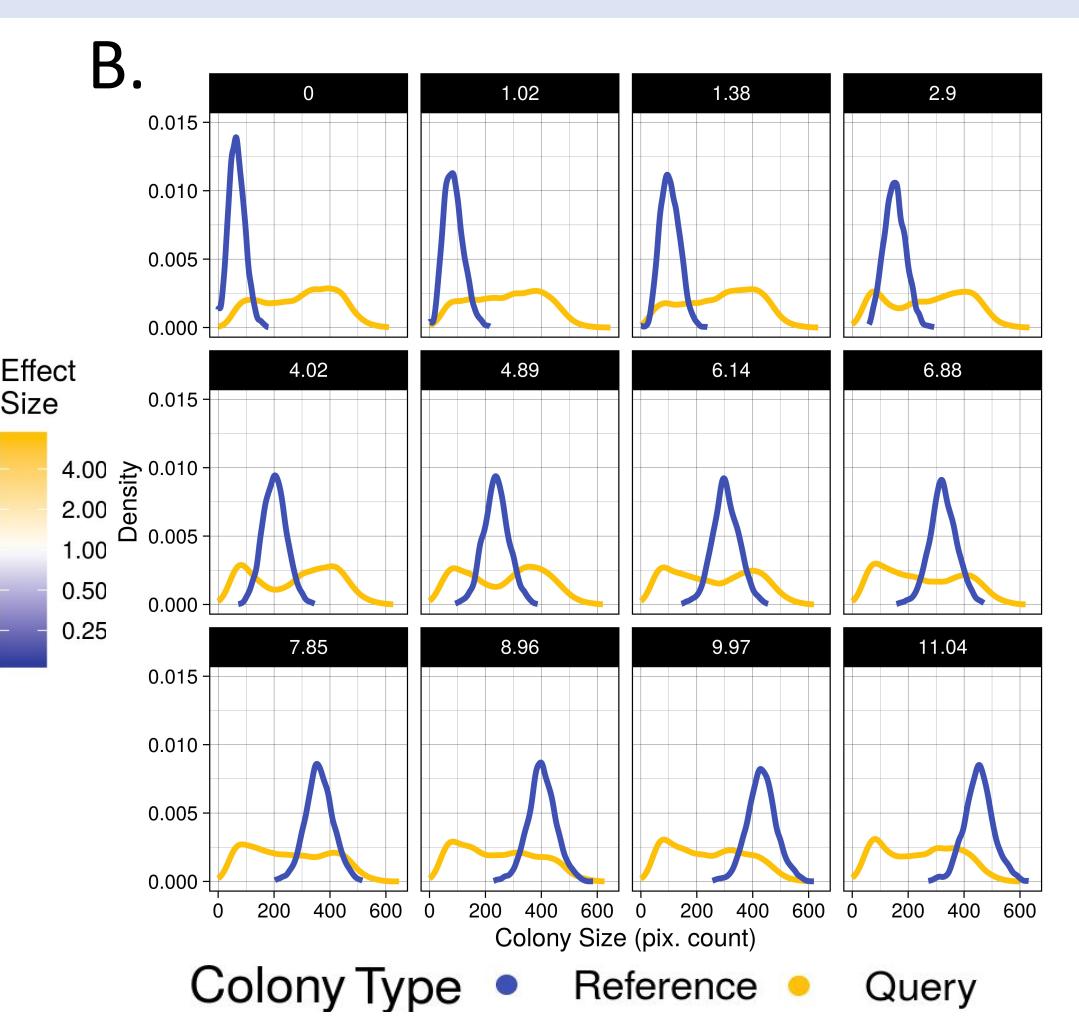


Figure 6. Virtual plates. A. Effect size table for when reference and query colony sizes are taken from two different time points. B. Colony size distributions when query colony sizes are chosen at random from all but the reference colony size time point.

Acknowledgements

This work was supported by the grant R00-GM108865 from the National Institutes of Health and the Searle Scholar Award awarded to Anne-Ruxandra Carvunis.

Results

1. LI Detector removes all source plate bias (fig. 7).

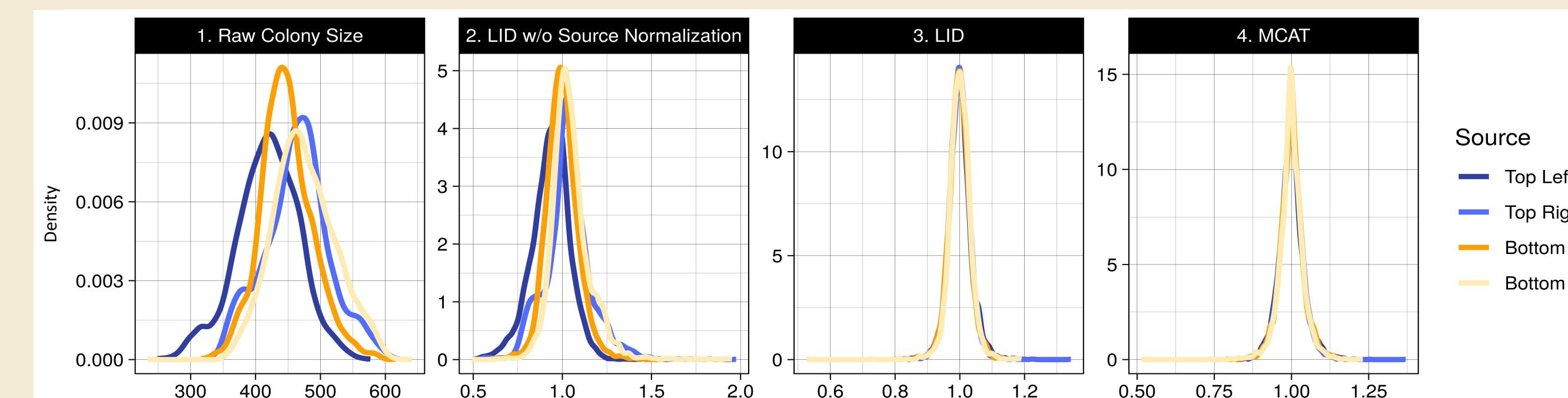


Figure 7. Effect of source normalization. Colors represent source plates used to make a higher density plate. LID = LI Detector, MCAT = Matlab Colony Analyzer Toolkit⁶.

2. Very high specificity and sensitivity (fig. 8).

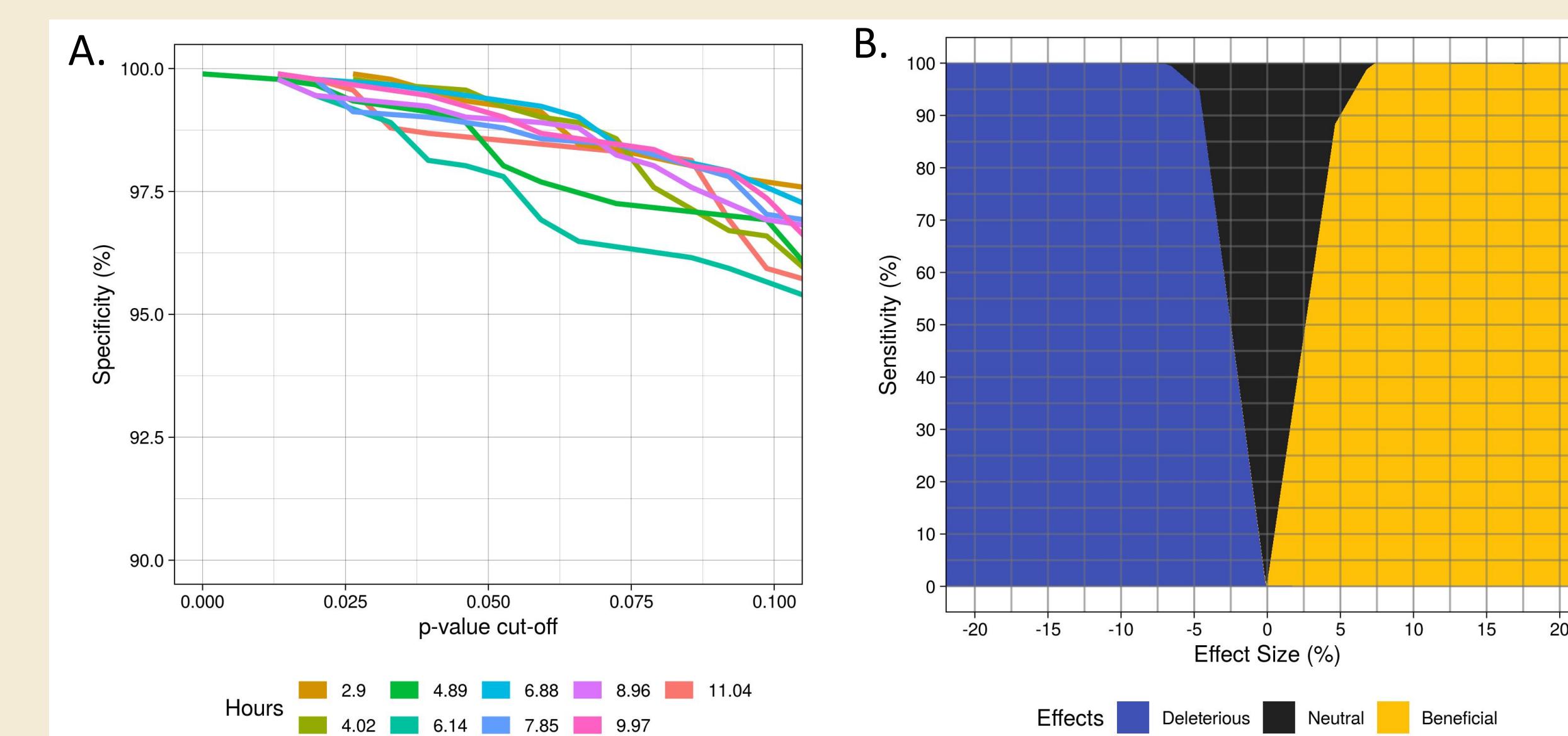


Figure 8. Specificity and sensitivity. A. Relationship between p-value cut-offs and specificity; and B. between effect size and sensitivity.

LI Detector has more than 95% specificity and 90% sensitivity for detecting 5% fitness effects in deleterious or beneficial direction (fig. 8).

3. Sensitivity remains high when colony sizes have a random distribution (fig. 9).

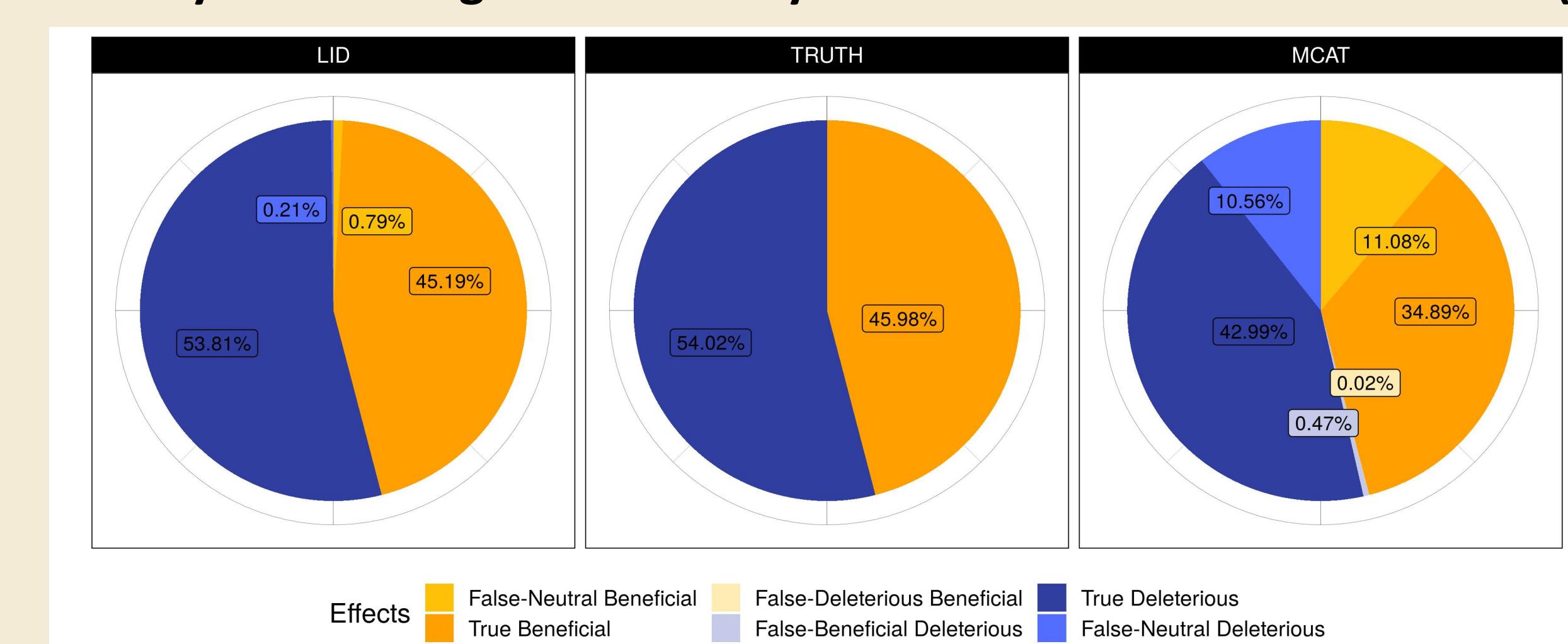


Figure 9. Combined results from all the virtual plates to show a relationship between effect size and sensitivity. LID = LI Detector, MCAT = Matlab Colony Analyzer Toolkit⁶.

Conclusion

LI Detector can resolve very small changes in colony size with high sensitivity and specificity without relying on *a priori* assumptions of underlying fitness distribution. This makes it an ideal framework to develop and refine gene-gene and gene-environment interaction networks of colony forming organisms like yeast. Its ability to detect increases in colony size also makes it an attractive method to investigate questions of evolutionary biology where fitness changes are often small and adaptive. Such applications of CBHTS have not been possible until now.