Published in final edited form as:
Nat Microbiol. ; 2: 17014. doi:10.1038/nmicrobiol.2017.14.

A tool named Iris for versatile high-throughput phenotyping in microorganisms

George Kritikos[1], **Manuel Banzhaf**[#1], **Lucia Herrera-Dominguez**[#1], **Alexandra Koumoutsi**[#1], **Morgane Wartel**[1], **Matylda Zietek**[1], and **Athanasios Typas**[1,*]

[1]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany

[#] These authors contributed equally to this work.

Abstract

Advances in our ability to systematically introduce and track controlled genetic variance in microbes have fueled high-throughput reverse genetics approaches in the past decade. When coupled to quantitative readouts, such approaches are extremely powerful at elucidating gene function and providing insights into the underlying pathways and the overall cellular network organization. Yet, until now all efforts for quantifying microbial macroscopic phenotypes have been restricted to monitoring growth in a small number of model microbes. We developed an image analysis software named Iris, which allows for systematic exploration of a number of orthogonal-to-growth processes, including biofilm formation, colony morphogenesis, envelope biogenesis, sporulation and reporter activity. In addition, Iris provides more sensitive growth measurements than current available software, and is compatible with a variety of different microbes, as well as with endpoint or kinetic data. We used Iris to reanalyze existing chemical genomics data in *Escherichia coli* and to perform proof-of-principle screens on colony biofilm formation and morphogenesis of different bacterial species and the pathogenic fungus, *Candida albicans*. Thereby we recapitulated existing knowledge but also identified a plethora of additional genes and pathways involved in both processes.

Introduction

High-throughput reverse genetics can provide unprecedentedly rich information on gene function and cellular network organization 1. Both gene-gene and gene-drug interactions rely on accurately measuring the phenotypic change between perturbed and unperturbed

[*]Correspondence: typas@embl.de.

Code availability
The source code for Iris, Iris ColonyPicker, and the Kinetic Data Companion tool is available at http://critichu.github.io/Iris

Data availability
The data that support the findings of this study are available as supplementary tables.

Author contributions
G.K., M.B., L.H.D., A.K., M.W., M.Z., and A.T. conceived and designed the experiments. M.B., L.H.D., A.K., M.W., and M.Z. performed the experiments. G.K. designed and implemented the software used in the analysis. G.K. analyzed the data. G.K. and A.T. wrote the paper.

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

states. Up until now, *fitness-related* measures have been exclusively used for quantifying such changes. For pooled barcoded libraries, this is a necessity, as sequencing reports on the relative abundance of each mutant in the experiment 2. While options increase when using ordered libraries 1, growth has dominated as the simplest phenotypic readout, reporting on effects at multiple levels and scales 3. Concordantly, currently available software tools 4–9 are tailored to measure colony size in high-density arrayed plates as a proxy of growth. Yet many bacterial core programs, such as biofilm, motility, competence and sporulation occur at stages when cells slow down or stop growing completely. Other processes impact cellular physiology but do not have a measureable effect on growth. Unfortunately, current tools cannot quantify such orthogonal-to-growth readouts, and even large-scale efforts to map their underlying genetic determinants have relied on qualitative or semi-quantitative metrics 10–12.

We developed the versatile and extensible image-analysis platform Iris, which can quantify multiple features of high-density arrayed microbial colonies. Iris is open-source, works with a variety of different microbes, and has add-on features that make it compatible with low-throughput or kinetic data. Iris has been successfully used by a number of labs 13–16. Here we illustrate its ability to report both on a series of colorimetric and colony-morphology based assays. By significantly expanding the palette of macroscopic features quantified, Iris sets the foundation for a new era of high-throughput phenotyping in microbes.

Results

Iris: an image analysis platform for microbial colonies

In the last decade, colony growth has dominated microbial high-throughput reverse genetics, serving as the sole quantitative readout of thousands of mutants arrayed on agar surfaces 17–21. To increase the quantitative information extracted from such screens, and to target the unresponsive-to-growth part of the genome, we developed the image analysis software Iris. Iris captures the multitude of phenotypes exhibited by microbial colonies, quantifying features related to colony growth (area, circularity, integral opacity), color and morphology (Fig. 1). The specifics on how each readout works are described in Methods, and the utility of each of the readouts is illustrated in subsequent sections and previous publications 13, 16. Iris binary distribution and source code can be downloaded at: http://critichu.github.io/Iris

Since macroscopic phenotypes heavily depend on the microbe and assay, we reasoned that no matter how inclusive we tried to be, there would always be assays that our current pipeline would be unsuitable for. Therefore, we designed Iris in a modular fashion, enabling future users to add modules. Currently, the first two modules crop and segment the image into single-colony images, which are then fed to the corresponding readout modules, each quantifying a different phenotype (Fig. 1). All modules can be adjusted by the user to fit the assay demands (Methods). Although the software was designed for processing high-density arrayed colonies (from 96- to 6144- colonies per plate), we have also adapted it to targeted approaches, where small sets of colonies are quantified (ColonyPicker; Methods). This meets a long-standing demand for a quantitative tool for microbial macroscopic phenotypes 11, 12, 22–24.

Colony integral opacity recapitulates growth more accurately than colony size

In all published genetic interaction or chemical genomics screens, colony size is calculated as the colony area in pixels. However microbial colonies grow not only in area, but also in height. In addition to colony area, Iris calculates the colony integral opacity, which is the sum of the brightness values for all the pixels within the colony bounds. Thus, colony integral opacity depicts better the 3-dimensional colony growth (Fig. 2a -inset).

To assess how colony integral opacity performed, we re-evaluated the *Nichols et al.* data, where ~4,000 *E. coli* single-gene deletions were screened in >300 conditions 19. As expected, colony size data from previous software and Iris matched nearly perfectly (Supplementary Fig. 1a). On the contrary, colony integral opacity exhibited a better dynamic range, especially for small colonies (Supplementary Fig. 1b). This confirmed previous observations that a few mutants had more translucent or denser colonies. To address this further, we calculated the density of each mutant (colony integral opacity/colony size), and plotted the normalized density for all the data as a function of size (Fig. 2a). Denser colonies could be spotted independent of size, whereas smaller colonies were in general less dense, suggesting that growth defects manifest not only on colony area but also on colony density. Thus integral opacity captures 3-dimensional colony growth better.

We first investigated conditions that changed the overall colony density (Supplementary Table 1). Chemicals targeting the cellular membranes, such as SDS, EDTA, Thiolactomycin, Verapamil and Procaine all resulted in more translucent colonies (Fig. 2B). Although SDS lowers the plate surface tension, resulting into flatter and more spread colonies, none of the other chemicals has surfactant activity, implying a more general role of membrane homeostasis on colony growth (see Supplementary Discussion). In contrast, aminoglycosides (Tobramycin, Streptomycin, Amikacin) and A22 were among the conditions with denser colonies (Fig. 2B). Sub-inhibitory concentrations of aminoglycosides induce biofilm formation in *E. coli* 25. A22 activates the Rcs system 26, inducing colanic acid secretion and making colonies mucoid. In accordance with colanic acid controlling colony density, mutants that consistently produced denser colonies across all conditions, *lon*, *lpcA* and *rfaF*, have been previously associated with mucoid phenotypes due to Rcs activation 27 (Fig. 2c; Supplementary Table 2). As a matter of fact, the Rcs system was differentially activated by distinct LPS core truncations, which was also captured well by colony density (Supplementary Fig. 1c). At the opposite end, mutants with consistently translucent phenotypes across conditions were enriched in genes that are part of core cellular processes (Fig. 2c; Supplementary Table 2).

We also looked for novel conditional phenotypes, as these are what our scoring system captures as gene-condition interactions 4. For example, a knockout of *yciB*, encoding a poorly characterized protein with putative role in cellular morphogenesis, exhibited high salt sensitivity (Fig. 2c), which was previously undetected by the colony size readout. *yciB* cells were aberrantly shaped and often lysed when growing in high salt (Supplementary Fig. 1d). To identify such conditional phenotypes more globally, we calculated integral opacity-based fitness scores (S-scores) for every mutant across each condition as previously described 19 (Supplementary Table 3). By correlating the vector of S-scores for each mutant across conditions, we acquired a gene association network (Supplementary Table 4). Comparing

this with the published network that used colony size S-scores 19 revealed a number of mutants for which we gained association power to other mutants and pathways (Fig. 2d). As an example, the molybdopterin synthesis pathway genes feature higher correlations to genes within the pathway, as well as with genes in related processes (Fig. 2e).

Since colony integral opacity has better dynamic range for measuring cellular fitness, we and others have started using it as the preferred growth readout in chemical genomics screens 14–16.

Quantifiable colorimetric assays- biofilm formation in gram-negative bacteria

Colorimetric assays typically measure the emergence or disappearance of color in or next to a microbial colony as a result of a phenotypic trait. Despite their quantitative potential (good sensitivity and dynamic range), such assays have mainly been used in a qualitative manner in diagnostics and screening. Using Iris, we and others have adapted a number of these assays to high-throughput reverse genetics screens, including membrane permeability/elevated lysis 13, sporulation 16, reporter activity (β-galactosidase) and biofilm formation assays.

Biofilm formation is a key lifestyle decision for microbes, triggered by environmental or nutritional cues, as well as intercellular signaling. Congo Red (CR) binds non-specifically to extracellular matrix components and has been traditionally used to stain *macrocolony* biofilm formation 28. We slightly adapted the CR assay and used it to profile the *E. coli* 19, 29 and *Pseudomonas aeruginosa* PA14 30 deletion libraries. The readout was reproducible within organism (Fig. 3a), orthogonal to growth (Supplementary Fig. 2a) and recapitulated existing knowledge in biofilm formation of the two organisms (Supplementary Table 5). For example, central players in *E. coli* biofilms, such as lipopolysaccharide (LPS), curli fimbriae and flagella were all hits in our screen (Fig. 3b). Curli are extracellular amyloid fibers that constitute the main structural component of *E. coli* K12 lab strain biofilms 22, 31. As expected, mutants in all genes contributing to their synthesis, transport or secretion to the surface abolished CR staining. The notable exception was *csgC*, encoding for a periplasmic protein that inhibits accumulation of intracellular amyloid fiber 32. A number of processes, previously unlinked to biofilm, were also identified in our screen (Fig. 3b, Supplementary Table 5). For example, *E. coli* mutants of the molybdopterin (MPT) biosynthesis pathway were stained more red, with the exception of *moaA* and *moaB*, which have less and no pronounced role in MPT biosynthesis, respectively 33, 34. Both MPT and cyclic-di-GMP (c-di-GMP) biosynthesis draw from the same pool of GTP (Fig. 3b). Failure to synthesize MPT could lead to elevated levels of the ubiquitous second messenger, c-di-GMP, and thus more biofilm 35. Similarly we could recapitulate all known c-di-GMP synthases in both organisms and identified 3 new ones for *P. aeruginosa* PA14 (Supplementary Fig. 2b).

Secondary metabolites, including quorum-sensing molecules, are key for *P. aeruginosa* biofilms. Consistent with literature, blocking quinolone (PQS: Pseudomonas Quinolone Signaling) production inhibited macrocolony biofilm formation 36. In contrast, *pqsL* mutants lead to PQS overproduction 37, and thus to more biofilm (Fig. 3c). A second class of secreted secondary metabolites, the redox-active phenazines, are also known to influence *P. aeruginosa* biofilm morphogenesis 38. Although *P. aeruginosa* PA14 encodes two nearly identical *phz* operons, *phz2* is the only one active in colonies 39. Consistently our

phenotypes stemmed only from mutants in *phz2* (Fig. 3c). However, *phz2* mutants exhibited less CR staining in our assay, in contrast to previous assays where *phz2* mutants were more wrinkled when probed at *later* stages of macrocolony biofilm development 39, 40. This implies that phenazines have a non-monotonic relationship with biofilm formation: their absence slows down matrix secretion at early stages, but increases it later (due to redox imbalance) leading to wrinkled colony formation 40.

P. aeruginosa biofilms are structurally diverse, containing exopolysacchharides, flagella, pili but also extracellular DNA and metabolites 41. In our assay the most contributing factor was type IV pili 42. Type IV pili are used for surface sensing and attachment 43, enabling chemotactic movement on surfaces during early biofilm formation 44. Mutations in nearly all of the subunits of the type IV pili machinery 42, the c-di-GMP effector FimX 45, and the chemotaxis system activating type IV pili extension/retraction 46 resulted in significantly decreased CR staining (Fig 3c, Supplementary Fig. 2b and Supplementary Table 6). Notable exceptions were: *chpB,* encoding for the receptor-specific CheB-like methylesterase which adapts the chemotactic response by deactivating the receptor 46, 47 - thus phenotype is justified; *pilC* encoding for the motor protein of the type IV pilus; and 2/4 minor pilin mutants present in the library, *pilW* and *fimU* (Fig 3C). The *pilC* mutant exhibited elevated CR staining because the isolated transposon insertion mutant 30 leads to overexpression of a minimally truncated functional protein, instead of knocking out the gene (Supplementary Fig. 2c). It is unclear if something similar happens with the minor pilin mutants, but minor pilins were suggested to control c-di-GMP levels, in addition to their structural role in type IV pilus 48. Consistent with a more complex role, a different set of minor pilin mutants exhibited colony structures at later time points of biofilm formation (Supplementary Fig. 2d).

A new algorithm for capturing microbial colony morphology

Colony structure formation, a hallmark of a sedimentary biofilm lifestyle, has been linked to the underlying mechanical forces of colony growth 49, to properties of the produced extracellular material 22, and to responses to oxygen, nutrient or metabolite gradients within the colony 24, 28. Although the wrinkles, valleys, concentric rings, and halos that microbial colonies form have fascinated researchers for decades, we currently lack tools for quantifying them. This is not a trivial undertaking, as texture-detection methods often fail due to technical biases, such as lighting inconsistencies between pictures or organisms. As part of Iris, we devised a new colony structure detection and quantification algorithm. In brief, this algorithm reports both on the “wrinkleness” of the colony, based on the wrinkle frequency and height, and on the size of the surrounding halo (Fig. 4a; Methods).

We benchmarked our new readout by measuring the structure complexity (wrinkleness) and agar invasion (halo size) of two homozygous single-gene deletion *Candida albicans* collections 10, 50. The human opportunistic pathogen *C. albicans* grows *in vivo* as yeast, hypha or pseudohypha. Colony structure on solid media reflects all three single-cell types. Smooth colonies are mostly comprised of yeast cells, while wrinkled ones are mostly comprised of hyphae and pseudohyphae. Invasive filamentation occurs around the colony as hyphae and pseudohyphae penetrate the agar (Fig. 4a). When measuring colony wrinkleness,

Iris successfully identified the main known regulatory network behind biofilm formation 51, 52, but also additional transcriptional regulators (Fig 4b). Processes enriched for low and high morphology scores matched expectations (Fig. 4b & Supplementary Table 7), and data correlated well with previously manually curated morphology scores 10 - automated quantification by Iris exhibited higher dynamic range and accuracy (Supplementary Fig. 3). We also used Iris to measure invasive filamentation, identifying a distinct set of genetic determinants and related processes (Fig. 4c & Supplementary Table 7).

We further profiled colony morphology and CR staining for mutants in the main pathways for biofilm formation in *Salmonella enterica* serovar Typhimurium. *S.* Typhimurium develops red dry and rough (rdar) colonies on CR plates with cellulose and curli fimbriae being the main contributors of these structures 31. Measuring both readouts allowed us to dissect the two pathways, with cellulose contributing mostly to structure formation and curli contributing more to color development (Supplementary Fig. 4).

Overall, our new algorithm captured well colony morphology characteristics in any organism tested. We envision it facilitating future larger-scale endeavors, which until recently were only qualitatively evaluated 10–12.

Kinetics option for Iris readouts

Kinetic data can be more informative than endpoint measurements because: a) they are more robust to sample noise, inoculation variance or asynchronous growth of arrayed colonies; b) can reveal non-monotonic behaviors across time; and c) they are more sensitive for detecting smaller effects. Therefore, we decided to build an add-on R-based software to accompany Iris, which uses the kinetic data to calculate and plot curves for the different readouts across time.

To illustrate its utility, we tracked 3 fitness-related colony features over time for all mutants in an extended version of the *E. coli* single-gene deletion library 19, 29. We asked which readout captures growth better, and how do different media and lighting conditions impact the measurements. We found that adding a dye in the media improved early colony detection (Supplementary Fig. 5a). We used the Gompertz model to fit the time series of all 3 fitness measures and calculate lag phase, slopes, and estimated maximum growth values for all the mutants. As noted before 9, colony growth is linear and not logarithmic. Center integral opacity was the most robust readout when comparing either the Gompertz slope or expected endpoint values for all described measures of mutant fitness (Supplementary Fig. 5b). We then asked whether the endpoint read we use in our screens recapitulates well such kinetic data. Indeed, a well-selected endpoint measurement was a good proxy of growth curve slopes (Supplementary Fig. 5c), thus simplifying data collection and analysis with little compromise in the end result. Nevertheless, correlation dropped when comparing slopes to endpoints of different readouts (Supplementary Fig. 5d).

In conclusion, Iris together with the provided Kinetic Data Companion tool simplify the processing of kinetic data analysis and visualization for colony measurements. This pipeline can be used with any of the other readouts Iris reports on.

Discussion

We have developed Iris, a freely available software for quantifying macroscopic colony phenotypes. In addition to more sensitive measures of microbial growth on plates, Iris features quantification modules tailored for colorimetric and colony-morphology-based assays, reporting on biofilm formation, sporulation, envelope integrity and reporter activity. These new features, which include an algorithm for quantifying colony structure, as well as the modularity and open-source nature of the software, make Iris unique in the field. Our effort concurs well with the advance of high-throughput reverse genetics approaches in microbiology, with currently >20 microbes having arrayed deletion libraries. Iris is also compatible with low-throughput measurements and kinetic data, offering a solution to labs that want to quantify such phenotypes for their targeted studies. Similarly, Iris can deal with diverse microbes, enabling studies of natural isolate collections or interspecies interactions.

We benchmarked Iris by quantifying the different phenotypes our assays captured. In this process, we revealed phenotypes for many genes that remained previously unresponsive, even when probing >300 conditions with growth as a readout 19. These tens of examples of new gene-condition associations can serve as the stepping stone for targeted mechanistic dissection of gene function. In addition to gaining insights into previously uncharacterized genes, our screens revealed previously unappreciated connections of known pathways, such as the potential role of the overall GTP pool in biofilm formation, a different role of phenazines in *P. aeruginosa* early biofilm development, new c-di-GMP synthases involved in *P. aeruginosa* biofilm formation, and transcriptional regulators involved in *C. albicans* filamentation.

In addition to the growth-unrelated readouts, we explored how colony growth is captured more accurately, and its relations with time. Our new readout, colony integral opacity is more sensitive than size for quantifying growth defects. Therefore, we used it to reanalyze the existing *E. coli* chemical genomics data 19, providing novel insights into the functions and associations of many genes. Combining the info from colony size and integral opacity, we measured density-related characteristics of colonies, which linked to the inability to form proper colony structures or to extracellular material secretion (Supplementary Discussion).

In summary, we created a long awaited tool for the microbiology community, which will facilitate quantitative reverse genetics approaches – both at a targeted and a large-scale level. We will maintain and update the software to meet future needs of users. As colony detection and quantification is microbe- and media- dependent, slight deviations of the current available modules can be developed when necessary.

Methods

Software design and implementation

Iris is implemented in Java and uses the image analysis routines of ImageJ 53. This is done via calls to the ImageJ API (Application Programming Interface), or direct calls to the routines of specialized ImageJ plugins. A distribution of the ImageJ code is bundled in the Iris distribution, so the user does not have to install ImageJ separately.

Europe PMC Funders Author Manuscripts

All image analysis steps (see below) are implemented by separate modules. While often specialized for a specific assay, modules are interchangeable and the same module (e.g. implementing plate rotation) can be used by several assays. The collection of modules that were tested to work best for a particular assay is called an assay profile. The user has to choose the assay profile, and select a folder containing the images to be quantified.

The set of algorithms used for each profile is documented online on the Iris webpage: http://critichu.github.io/Iris. By modifying a supplied settings file, users can adapt several parameters for every step of the image analysis pipeline. Image rotation, cropping, segmentation (grid format), and detection parameters can be adjusted either in a per-profile fashion or globally.

Colony array image analysis pipeline

The following steps take place at the whole image level to process high-resolution images of colony arrays on agar plates.

Rotation—The first processing step for the software is to automatically rotate the image so that the colony array is perfectly horizontal. This step is crucial in high-density applications using 1536 or 6144-colony plates. Iris performs accurate image rotation by rotating a grayscale version of the image in 0.5° increments, and calculating the per-row sum of pixel brightness. Colony rows are perfectly horizontal in the rotation that maximizes the variance of brightness sums. To accelerate processing and avoid confounders from the plate borders, only the center part of the image is used to calculate rotation.

Plate border cropping—In the next step Iris detects and crops the plate boundaries. Plastic plate boundaries diffract light towards the camera upon side lighting. Iris detects the elevated brightness level of the plastic plate borders by means of maxima of brightness on the per-row and per-column sums of brightness after grayscale image rotation. For applications where backlight was used the plate borders create a shadow. After inverting the image, the plate borders are detected in a similar fashion as above. Manually cropped plates can be used in the rare cases where automated plate boarder detection fails (see Quality control section).

Image segmentation—A cropped image containing only the colony array is then segmented into image tiles, each holding only one colony. Since colonies are usually brighter than the background, summing brightness per-row and per-column gives brightness sum valleys and peaks. Peaks correspond to the centers of colonies per row or column, while local brightness sum minima correspond to inter-colony space (Fig. 1). Iris uses the minima corresponding to inter-colony space to segment the image into many images containing single-colonies, called tiles. As some colonies can grow larger than their holding tile, Iris then examines each tile and adjusts its borders to accommodate the entire colony. While tile image boundaries are calculated on the grayscale version of the image, these tile image boundaries are applied to the original color image, giving full-color image tiles to be further processed.

Europe PMC Funders Author Manuscripts

Tile processing—Each tile is separately processed by one or several tile processor modules, specialized in quantifying a specific phenotype. This design allows for independent colony phenotypic quantification, but also for easy incorporation of new readouts. To expedite processing in the case of multiple phenotype quantification, colony bound detection occurs at the first tile processor (defined by the selected profile) and is provided as input to the rest tile readers.

Colony phenotype quantification

The following steps take place at a tile (colony) level:

Colony boundary detection—The first step of colony phenotype quantification is to accurately detect the colony boundaries. Colony shapes, sizes, colors, as well as background vary substantially across diverse applications (different microbes, media, dyes or camera settings). Iris uses multiple thresholding algorithms to detect colonies, which differ according to the application (see table at http://critichu.github.io/Iris).

For most applications colonies are brighter than plate background; for those Iris applies grayscale image thresholding algorithms such as the Otsu algorithm. Typically such thresholding algorithms operate on the histogram of image brightness, and attempt to select a threshold best separating the foreground (bright) from background pixels.

For applications such as the biofilm & morphology readouts (CR plates), where the brightness of a colony relative to its background is uncertain (can be brighter or darker than background), or may vary within the assay, Iris employs the Marr-Hildreth algorithm, also known as Laplacian of Gaussian algorithm. This algorithm first applies a smoothing Gaussian filter to the grayscale image. Subsequently a second order derivative of the Gaussian is calculated, whereby zero values denote sharp changes in brightness. These pixel locations are then used as colony boundaries.

Colony circularity can be used to filter abnormal colonies or very rare events in which Iris mistakes agar/lighting abnormalities for colonies (see also Quality Control section). In our experience, colonies detected with circularity under 0.4 should be either discarded or considered for manual inspection. To further increase colony detection robustness, several Iris profiles use multiple thresholding algorithms for each colony and a hierarchical decision process to select the best detection result (after applying internal size and circularity criteria).

In some applications of the CR assay for monitoring of biofilm, *parts* of colonies (rather than the entire colony) can become darker than the plate background as they absorb the CR dye. Accurate colony detection is very difficult in such cases. To circumvent this Iris can calculate the average background color by sampling pixels in the 4 corners of the tile. Subsequently, pixels featuring similar color to the background average are replaced by black. Color similarity is calculated based on their distance in the RGB space. After background color removal, the colony boundaries are detected using the Otsu thresholding algorithm. Since this functionality is only needed rarely for current applications, it is not part of the default profiles, but can be employed for new readouts.

Finally, Iris takes into account only the main colony on the tile, ignoring any other smaller colonies, which could arise from minor contamination. Subsequently, only the colony with the biggest area is kept for downstream analysis

Colony center detection—Applications such as detection of sporulation require the detection of pigmentation in the center of the colony. To robustly detect the colony center, Iris takes advantage of the fact that in high-throughput assays colonies are pinned in rectangular arrays. In this way, the X-axis displacement of colony centers will be equal for all colonies in the same column. The same is valid for the Y-axis displacement of all colonies in the same row. The X-axis displacement for colony centers in each column is then calculated as the median of all colony center X-axis displacements in this column. Colony center Y-axis displacements are calculated in a similar fashion for each row, resulting in robust coordinate calculation for all colony centers.

To initially estimate colony centers per colony Iris calculates the ultimate erosion point of the grayscale colony after applying an image thresholding algorithm (Otsu). This approach is also used in the case of manually cropped single-colony images.

Colony-growth related readouts—Colony size is typically measured by all available software 4–9 as a proxy of growth fitness. Colony size corresponds to the colony area in pixels and is calculated by counting the number of pixels within the colony bounds.

To calculate colony integral opacity, Iris first calculates the average brightness of background pixels in the tile. Background brightness calculation per tile is advantageous in applications where lighting is not uniform across the entire plate. Subsequently, the over-background brightness of every pixel within the colony bounds is calculated as the pixel brightness minus the background average brightness. This is then summed for all within-colony pixels to provide the colony integral opacity. Colony density is calculated as the average integral opacity in the colony, i.e. by dividing the colony integral opacity by the number of pixels in the colony.

A third growth-related readout, center integral opacity, restricts the measurement of colony integral opacity to pixels within a defined area around the colony center. Center area size corresponds to the area size used by a previous study 9, taking into account differences in setup (e.g. image resolution, average number of pixels in colony).

Sporulation—*B. subtilis* sporulating cells turn dark brown in minimal media. To quantify this pigmentation change, Iris assesses the three primary color channels in the cubic RGB representation per pixel. To do so, green and blue channel intensities are added together, and multiplied by a gain factor; the same process is done to the red channel. Subsequently the difference of the red channel product is subtracted from that of the green and blue channels. Since cells turn dark brown, detecting the pigmentation change is aided by detecting changes in pixel brightness, which is also incorporated in the sporulation score formula:

$$color\ \ score{=}g_{RGB}\cdot[g_{R}\cdot R-g_{GB}\cdot(G{+}B)]{+}g_{SD}\cdot(\mathrm{S}{+}D)$$
$$D{=}255\text{-}\ brightness$$

R, G, B, and S denote red, green, blue, and saturation values respectively for each pixel, *g* indicates a gain factor. Pigmentation change during sporulation takes place in the colony center. Iris reports the average value of the color score across pixels in a defined circular area centered at the colony center and a diameter of 24 pixels.

Biofilm formation—In our colony biofilm assay, colonies that form biofilm bind more the CR dye and thus turn dark red. Iris detects this color change similarly as described above for sporulating cells in *B. subtilis*, and uses the same formula as above for each pixel in the total area of the colony.

Chromophore detection—An example of a chromophore assay is the CPRG assay, which can be used to detect mutant cells with higher permeability or elevated lysis frequency by detecting a chromophore reaction on agar plates 13. Chromophore-based assays on solid plates are limited by chromophore diffusion to neighboring colonies. In order to robustly detect low chromophore concentrations, Iris converts the tile image to the cylindrical HSV color space. Subsequently chromophore concentration is quantified as the radial distance between the hue of every pixel and the hue of the chromophore color (e.g. 0° for red color), and summed across colony pixels.

Colony morphology—Microbial colonies are often structured, with ridges extending from the colony center, and valleys separating the ridges. Iris detects colony structure complexity by traversing colony pixels in circles concentric with the colony center. Brightness levels of pixels within such a circle feature valleys and peaks, which coincide with colony structures (Fig. 4a). Since lighting differences can account for brightness peak height, Iris counts the number of peaks in a binary way if they are above a certain brightness threshold.

Invasive filamentation—Colonies of some microbial species, such as *C. albicans,* also extend into the agar, which is readily observable at the image of a colony array. By sequentially applying two image thresholding algorithms of different sensitivity, Iris captures both the extent of the filamentous agar invasion, as well as the over-agar colony size. The Percentile algorithm is first used to detect the extent of the entire colony. Subsequently, the Minimum thresholding algorithm is used to acquire over-agar colony growth. The Minimum algorithm is applied only to the region previously determined by the Percentile algorithm. Subsequently Iris calculates the extent of in-agar growth by subtracting the areas of these two regions.

Quality control

To assess the robustness of the cropping and segmentation algorithms, we used Iris to process > 8,000 high-density array pictures (1536 mutants/plate) with different sources of variability (orientation, lighting, media etc). Failure rate was < 2%, and in all cases the issue was on plate boarder identification. This problem can be easily solved by feeding manually plate cropped plates to Iris.

Detection rates differ per profile and assay. In most non-colorimetric assays misidentified or missed colonies are negligible (<0.01%). In colorimetric assays, color of background and

colony can become sometimes difficult to distinguish. Nevertheless, detection rates of our current biofilm readout module are > 99.7% when looking in tens of millions of *E. coli* and *P. aeruginosa* colonies. Almost all come from false positives (FPs) – i.e. Iris misdetecting an existing colony or mistaking agar abnormalities for colonies. FPs can be detected in batch by using circularity thresholds (misdetected colonies are less circular). False negatives (FNs), i.e. colonies that Iris completely missed, are <0.05%. FNs can be identified by spotting replicate outliers (mutants where only in one replicate experiment have size = 0).

Iris also allows users to control the quality of their results. First, for every input image Iris produces an image file (termed a grid file) that can be used to quickly inspect the segmentation and colony detection results. Users can also visualize each measured phenotype of the Iris output by using a script available on the Iris website (see *Visualizing Iris output* at http://critichu.github.io/Iris). To address such cases, users can rerun Iris with adjusted detection parameters (e.g. increasing minimum circularity), or manually define the misdetected colonies and reprocess them with the Iris ColonyPicker (see below).

Analyzing single colonies using Iris ColonyPicker

Iris is distributed alongside Iris ColonyPicker, which provides a user-friendly way to profile single colonies. After choosing a profile, the user selects an image containing colonies to be analyzed. Defining a rectangular area surrounding a colony will lead to automatic detection of colony boundaries, whereas round selection forces Iris to use this selection as colony boundaries. In both cases, user-selected colonies are analyzed in the background using the tile reader modules of the selected Iris profile.

Kinetic Data Companion tool

Iris is accompanied by a tool to generate growth curves from time series data on any of the phenotypes mentioned. To automatically load all iris files annotated with their post-inoculation timepoint, image file names need to be in the following format: prefix_plateNumber-Hour-Minute.JPG. Prefix can be freely chosen, as long as it does not contain the underscore or dash characters. Detailed instructions on individual functions of the companion tool can be downloaded on the same webpage as the Iris distribution.

Experimental Procedures

Microscopy images—Wildtype and *ΔyciB* cells were grown in LB containing low salt (75mM NaCl) before being transferred to LB with high salt (600mM NaCl) and grown for 6-7 generations before being fixed as described before 54. Fixed cells were imaged on a Nikon Eclipse TE inverted fluorescence microscope with a 60X (NA 1.40) oil-immersion objective. Images were collected using the DS-Qi2 Mono Digital Microscope Camera and the NIS-Elements AR (Advanced Research) software (Nikon).

Rcs –dependent activity—The *rprA* promoter was amplified and inserted into a luxCDABE plasmid backbone without promoter to give p-rprA-lux. The plasmid has a pSC101 origin of replication and confers spectinomycin resistance. Selected Keio collection mutants were transformed with the p-rprA-lux plasmid and were grown into 50 μL of LB medium containing kanamycin and spectinomycin in 384-well plates. Successive dilutions

were used to make sure inoculum was low and we could monitor cells for several generations in exponential phase. OD600 and luminescence were monitored every hour for 14 hours using a filter max F5 (Molecular devices). Here we plot the maximal luciferase activity for each background.

Biofilm formation assay—The *E. coli* KEIO collection 29 was arrayed on solid agar plates in 1536 format, and plates were incubated at room temperature for 30 hours. Medium used was LB 0% NaCl supplemented with 40 μg/ml Congo red (CR) and 20 μg/ml Coomassie blue.

The *S.* Typhimurium single gene deletion library 55 was arrayed on solid agar plates in 384 format, and plates were incubated at room temperature for 48 hours. Medium composition was the same as for *E. coli*.

The *P. aeruginosa* PA14 single deletion transposon library 30 was arrayed on solid agar plates in 384 format, and plates were incubated at room temperature. Medium used was TB (Tryptone Broth) supplemented with 20 μg/ml CR and 10 μg/ml Coomassie blue.

All plates were imaged under controlled lighting conditions (spImager S&P Robotics Inc.) using a 18 megapixel Canon Rebel T3i (Canon Inc. USA).

Colony morphology assay—Two homozygous deletion collections of *C. albicans* -one containing deletions in 674 genes 10 and a second targeting 143 transcriptional factors 50 were arrayed on Spider medium agar plates (a standard hyphal inducing medium) in 96 format. Plates were incubated at 30°C and were imaged after 7 days as described above.

Kinetic assay—The *E. coli* KEIO collection 29 was arrayed on solid agar plates in 384 format, and grown on LB 0% salt medium containing LB with or without Congo Red and Coomassie blue (40 μg/ml and 20 μg/ml respectively) at 37 °C. Plates were imaged hourly after pinning for 14 hours under controlled lighting conditions (spImager S&P Robotics Inc.) using either a translucent background and backlight, or black background and uniform side lighting (front light).

Assessing PilC levels in pilC mutant from PA14 mutant library—*P. aeruginosa* PA14 wildtype and *pilC* transposon mutant were grown for 48 hours at room temperature on LB Miller agar plates. Cells were taken at different timepoints, diluted in PBS, washed and adjusted to an OD of 3. A Western Blot using anit-PilC was performed with 20 μl sample as previously described 56.

Data handling

By design, Iris serves as a tool to extract colony features. Further processing of the results, including detection and removal of potential spatial biases, as well as statistical analysis can be addressed by ad-hoc analysis or dedicated tools (e.g. EMAP 4 and SGAtools 57, 58). Below we describe the data handling approach used for the data presented in this paper.

Reanalysis of existing chemical genomics images—Images from the large scale *E. coli* chemical genomics screen 19 were analyzed with Iris to acquire colony size and colony integral opacity values. Unfortunately ~20% of original the images could not be retrieved. Colony density was calculated as the ratio between colony integral opacity and colony size. Per-condition density median and standard deviation was calculated using all colonies of conditions where at least 8 images could be retrieved.

To compare across mutants, we report normalized density values, whereby differences in the median and standard deviation among conditions were removed by calculating a colony density z-score per image. Normalized density values were then calculated by correcting the density z-scores for spatial biases due to lighting effects. Density z-score values for each of two outer columns in every 1536-array plate were used to fit a linear model, against which they were corrected. The remaining part of the plate was corrected against a second order polynomial surface fitted on the density z-score values. Per-mutant normalized density median and standard deviation were calculated on conditions with at least 2 replicate plates per library plate.

Colony integral opacity was also used as a fitness measure and processed with the EMAP toolbox 4 as previously described 19 to produce mutant-condition phenotypic scores. These scores were then used to calculate gene-gene phenotypic signature correlations (Pearson). These correlations were compared to the equivalent reported correlations calculated using colony size data 19. Gene pair correlation thresholds corresponding to 3 standard deviations were set for both the reported, and the integral opacity-derived correlations. Subsequently, gene pair correlations that differ by more than 1.5 standard deviations between the two datasets were separated in two sets: gene pairs that either gained or lost correlation in the new dataset (Fig. 2d).

A subset of the conditions was not included in the integral opacity-based chemical genomics dataset as not enough replicate images could be retrieved for these conditions. These are 67 conditions of the original 324 conditions (20.6%). We then verified that removal of these conditions from the published size-derived dataset does not lead to increased gene-gene correlations such as those reported among genes of the molybdopterin pathway for the integral opacity data.

Biofilm formation—Biofilm formation per mutant colony was quantified in Iris using the “Biofilm formation” profile. Spatial biases in the *E. coli* 1536 plate format biofilm data were removed by fitting a second order polynomial surface and multiplicatively correcting for the expected values. The polynomial surface correction omitted the outmost 4 rows and columns, whereby each row and column was separately multiplicatively corrected so that their corrected median corresponds to the median of the remaining inner part of the plate. Plate-to-plate biases were then multiplicatively corrected so that the corrected plate medians match.

To remove spatial biases on the *P. aeruginosa* 384 plate format data, an expected ratio per plate position was calculated as the median of ratios of colony biofilm scores to their corresponding plate median. To avoid expected ratios being driven by outliers, these were

further smoothened using a 3x3 sliding window approach. Subsequently, the ratio of each colony was corrected for the expected ratio and multiplied by a plate average calculated as the median value of all observations, thus multiplicatively correcting for spatial and plate-to-plate biases in one step.

For the *E. coli* screen mean and standard deviation values were calculated using 4 technical replicate measurements of two biological replicates per mutant. For the *P. aeruginosa* screen, mean and standard deviation values were calculated using 4 technical replicate measurements per mutant.

For both screens, top and bottom 2.5% of the mutants ranked by their average biofilm score were termed positive and negative outliers. GO enrichments were calculated using an in-house developed software and GO annotations retrieved from the EcoCyc and BioCyc websites 19 for *E. coli* and *P. aeruginosa* respectively. In the case of *P. aeruginosa* enrichments were also done on pathways 59, operon membership 60, and domain content 61.

***C. albicans* colony morphology—**Clones (biological replicates) with discordant growth or morphology phenotypes were flagged and removed from further analysis when manual inspection and literature survey were not enough for deciding which is the right one. The average of acquired morphology values of two technical replicates per clone (2 clones per mutant) were compared to previously reported manual phenotypic scores 10. Outlier GO enrichment analysis was performed with an in-house developed software and annotations downloaded from the *Candida* genome database 62.

Kinetic data—Images were analyzed by Iris to acquire colony size, colony integral opacity, and center integral opacity. Iris profiles used were “Colony opacity”, and “Colony opacity inverted” respectively for images using front light and backlight. Growth curve calculation was performed using the grofit R package. Dataset comparisons, and plotting were done using the statistical computing language R.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to KC Huang for advice on colony boundary detection algorithms. We also thank him, L.E. Dietrich and A. Price-Whelan for critically reading the manuscript and providing feedback. We thank L. Burrows (McMaster, Canada) for providing an antibody against PilC. This work was supported by the Sofja Kovalevskaja Award of the Alexander von Humboldt Foundation to A.T.

References

1. Brochado AR, Typas A. High-throughput approaches to understanding gene function and mapping network architecture in bacteria. Curr Opin Microbiol. 2013; 16:199–206. [PubMed: 23403119]
2. Gray AN, et al. High-throughput bacterial functional genomics in the sequencing era. Curr Opin Microbiol. 2015; 27:86–95. [PubMed: 26336012]

3. Auer GK, et al. Mechanical Genomics Identifies Diverse Modulators of Bacterial Cell Stiffness. Cell Syst. 2016; 2:402–411. [PubMed: 27321372]
4. Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. 2006; 7:R63. [PubMed: 16859555]
5. Lawless C, Wilkinson DJ, Young A, Addinall SG, Lydall DA. Colonyzer: automated quantification of micro-organism growth characteristics on solid agar. BMC Bioinformatics. 2010; 11:287. [PubMed: 20509870]
6. Dittmar JC, Reid RJ, Rothstein R. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. BMC Bioinformatics. 2010; 11:353. [PubMed: 20584323]
7. Young BP, Loewen CJ. Balony: a software package for analysis of data generated by synthetic genetic array experiments. BMC Bioinformatics. 2013; 14:354. [PubMed: 24305553]
8. Wagih O, Parts L. gitter: a robust and accurate method for quantification of colony sizes from plate images. G3 (Bethesda). 2014; 4:547–552. [PubMed: 24474170]
9. Takeuchi R, et al. Colony-live--a high-throughput method for measuring microbial colony growth kinetics--reveals diverse growth effects of gene knockouts in *Escherichia coli*. BMC Microbiol. 2014; 14:171. [PubMed: 24964927]
10. Noble SM, French S, Kohn LA, Chen V, Johnson AD. Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. Nat Genet. 2010; 42:590–598. [PubMed: 20543849]
11. Ryan O, et al. Global gene deletion analysis exploring yeast filamentous growth. Science. 2012; 337:1353–1356. [PubMed: 22984072]
12. Cabeen MT, Leiman SA, Losick R. Colony-morphology screening uncovers a role for the *Pseudomonas aeruginosa* nitrogen-related phosphotransferase system in biofilm formation. Mol Microbiol. 2016; 99:557–570. [PubMed: 26483285]
13. Paradis-Bleau C, Kritikos G, Orlova K, Typas A, Bernhardt TG. A genome-wide screen for bacterial envelope biogenesis mutants identifies a novel factor involved in cell wall precursor metabolism. PLoS Genet. 2014; 10:e1004056. [PubMed: 24391520]
14. Peters JM, et al. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. Cell. 2016; 165:1493–1506. [PubMed: 27238023]
15. Shiver AL, et al. A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S. PLoS Genet. 2016; 12:e1006124. [PubMed: 27355376]
16. Koo BM, et al. Construction and analysis of two systematic, high-quality deletion libraries for the Gram-positive Model bacterium Bacillus subtilis. 2016 in press.
17. Schuldiner M, Collins SR, Weissman JS, Krogan NJ. Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. Methods. 2006; 40:344–352. [PubMed: 17101447]
18. Costanzo M, et al. The genetic landscape of a cell. Science. 2010; 327:425–431. [PubMed: 20093466]
19. Nichols RJ, et al. Phenotypic Landscape of a Bacterial Cell. Cell. 2011; 144:143–156. [PubMed: 21185072]
20. Ryan CJ, et al. Hierarchical modularity and the evolution of genetic interactomes across species. Mol Cell. 2012; 46:691–704. [PubMed: 22681890]
21. Brown JC, et al. Unraveling the biology of a fungal meningitis pathogen using chemical genetics. Cell. 2014; 159:1168–1187. [PubMed: 25416953]
22. Serra DO, Richter AM, Hengge R. Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. J Bacteriol. 2013; 195:5540–5554. [PubMed: 24097954]
23. Ha DG, Richman ME, O'Toole GA. Deletion mutant library for investigation of functional outputs of cyclic diguanylate metabolism in *Pseudomonas aeruginosa PA14* . Appl Environ Microbiol. 2014; 80:3384–3393. [PubMed: 24657857]
24. Okegbe C, Price-Whelan A, Dietrich LE. Redox-driven regulation of microbial community morphogenesis. Curr Opin Microbiol. 2014; 18:39–45. [PubMed: 24607644]

25. Hoffman LR, et al. Aminoglycoside antibiotics induce bacterial biofilm formation. Nature. 2005; 436:1171–1175. [PubMed: 16121184]
26. Cho SH, et al. Detecting Envelope Stress by Monitoring beta-Barrel Assembly. Cell. 2014; 159:1652–1664. [PubMed: 25525882]
27. Majdalani N, Gottesman S. The Rcs phosphorelay: a complex signal transduction system. Annu Rev Microbiol. 2005; 59:379–405. [PubMed: 16153174]
28. Serra DO, Hengge R. Stress responses go three dimensional - the spatial order of physiological differentiation in bacterial macrocolony biofilms. Environ Microbiol. 2014; 16:1455–1471. [PubMed: 24725389]
29. Baba T, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006; 2:2006 0008.
30. Liberati NT, et al. An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A. 2006; 103:2833–2838. [PubMed: 16477005]
31. Zogaj X, Nimtz M, Rohde M, Bokranz W, Romling U. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. Mol Microbiol. 2001; 39:1452–1463. [PubMed: 11260463]
32. Evans ML, et al. The bacterial curli system possesses a potent and selective inhibitor of amyloid formation. Mol Cell. 2015; 57:445–455. [PubMed: 25620560]
33. Hover BM, Tonthat NK, Schumacher MA, Yokoyama K. Mechanism of pyranopterin ring formation in molybdenum cofactor biosynthesis. Proc Natl Acad Sci U S A. 2015; 112:6347–6352. [PubMed: 25941396]
34. Kozmin SG, Schaaper RM. Genetic characterization of *moaB* mutants of *Escherichia coli* . Res Microbiol. 2013; 164:689–694. [PubMed: 23680484]
35. Hengge R. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol. 2009; 7:263–273. [PubMed: 19287449]
36. Wolska KI, Grudniak AM, Rudnicka Z, Markowska K. Genetic control of bacterial biofilms. Journal of applied genetics. 2016; 57:225–238. [PubMed: 26294280]
37. D'Argenio DA, Calfee MW, Rainey PB, Pesci EC. Autolysis and autoaggregation in *Pseudomonas aeruginosa* colony morphology mutants. J Bacteriol. 2002; 184:6481–6489. [PubMed: 12426335]
38. Dietrich LE, Teal TK, Price-Whelan A, Newman DK. Redox-active antibiotics control gene expression and community behavior in divergent bacteria. Science. 2008; 321:1203–1206. [PubMed: 18755976]
39. Recinos DA, et al. Redundant phenazine operons in *Pseudomonas aeruginosa* exhibit environment-dependent expression and differential roles in pathogenicity. Proc Natl Acad Sci U S A. 2012; 109:19420–19425. [PubMed: 23129634]
40. Dietrich LE, et al. Bacterial community morphogenesis is intimately linked to the intracellular redox state. J Bacteriol. 2013; 195:1371–1380. [PubMed: 23292774]
41. Mann EE, Wozniak DJ. *Pseudomonas* biofilm matrix composition and niche biology. FEMS microbiology reviews. 2012; 36:893–916. [PubMed: 22212072]
42. Burrows LL. *Pseudomonas aeruginosa* twitching motility: type IV pili in action. Annu Rev Microbiol. 2012; 66:493–520. [PubMed: 22746331]
43. Persat A, Inclan YF, Engel JN, Stone HA, Gitai Z. Type IV pili mechanochemically regulate virulence factors in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A. 2015; 112:7563–7568. [PubMed: 26041805]
44. Oliveira NM, Foster KR, Durham WM. Single-cell twitching chemotaxis in developing biofilms. Proc Natl Acad Sci U S A. 2016; 113:6532–6537. [PubMed: 27222583]
45. Kazmierczak BI, Lebron MB, Murray TS. Analysis of FimX, a phosphodiesterase that governs twitching motility in *Pseudomonas aeruginosa*. Mol Microbiol. 2006; 60:1026–1043. [PubMed: 16677312]
46. Sampedro I, Parales RE, Krell T, Hill JE. *Pseudomonas* chemotaxis. FEMS microbiology reviews. 2015; 39:17–46. [PubMed: 25100612]

47. Typas A, Sourjik V. Bacterial protein networks: properties and functions. Nat Rev Microbiol. 2015; 13:559–572. [PubMed: 26256789]
48. Kuchma SL, Griffin EF, O'Toole GA. Minor pilins of the type IV pilus system participate in the negative regulation of swarming motility. J Bacteriol. 2012; 194:5388–5403. [PubMed: 22865844]
49. Asally M, et al. Localized cell death focuses mechanical forces during 3D patterning in a biofilm. Proc Natl Acad Sci U S A. 2012; 109:18891–18896. [PubMed: 23012477]
50. Homann OR, Dea J, Noble SM, Johnson AD. A phenotypic profile of the *Candida albicans* regulatory network. PLoS Genet. 2009; 5:e1000783. [PubMed: 20041210]
51. Nobile CJ, et al. A recently evolved transcriptional network controls biofilm development in *Candida albicans*. Cell. 2012; 148:126–138. [PubMed: 22265407]
52. Fox EP, et al. An expanded regulatory network temporally controls *Candida albicans* biofilm formation. Mol Microbiol. 2015; 96:1226–1239. [PubMed: 25784162]
53. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature methods. 2012; 9:671–675. [PubMed: 22930834]
54. Typas A, et al. Regulation of peptidoglycan synthesis by outer membrane proteins. Cell. 2010; 143:1097–1109. [PubMed: 21183073]
55. Porwollik S, et al. Defined single-gene and multi-gene deletion mutant collections in *Salmonella enterica* sv Typhimurium. PloS one. 2014; 9:e99820. [PubMed: 25007190]
56. Takhar HK, Kemp K, Kim M, Howell PL, Burrows LL. The platform protein is essential for type IV pilus biogenesis. The Journal of biological chemistry. 2013; 288:9721–9728. [PubMed: 23413032]
57. Baryshnikova A, et al. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nature methods. 2010; 7:1017–1024. [PubMed: 21076421]
58. Wagih O, et al. SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. Nucleic acids research. 2013; 41:W591–596. [PubMed: 23677617]
59. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic acids research. 2016; 44:D457–462. [PubMed: 26476454]
60. Mao F, Dam P, Chou J, Olman V, Xu Y. DOOR: a database for prokaryotic operons. Nucleic acids research. 2009; 37:D459–463. [PubMed: 18988623]
61. Marchler-Bauer A, et al. CDD: NCBI's conserved domain database. Nucleic acids research. 2015; 43:D222–226. [PubMed: 25414356]
62. Inglis DO, et al. The *Candida* genome database incorporates multiple *Candida* species: multispecies search and analysis tools with curated gene and protein information for *Candida albicans* and *Candida glabrata*. Nucleic acids research. 2012; 40:D667–674. [PubMed: 22064862]
63. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57:289–300.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

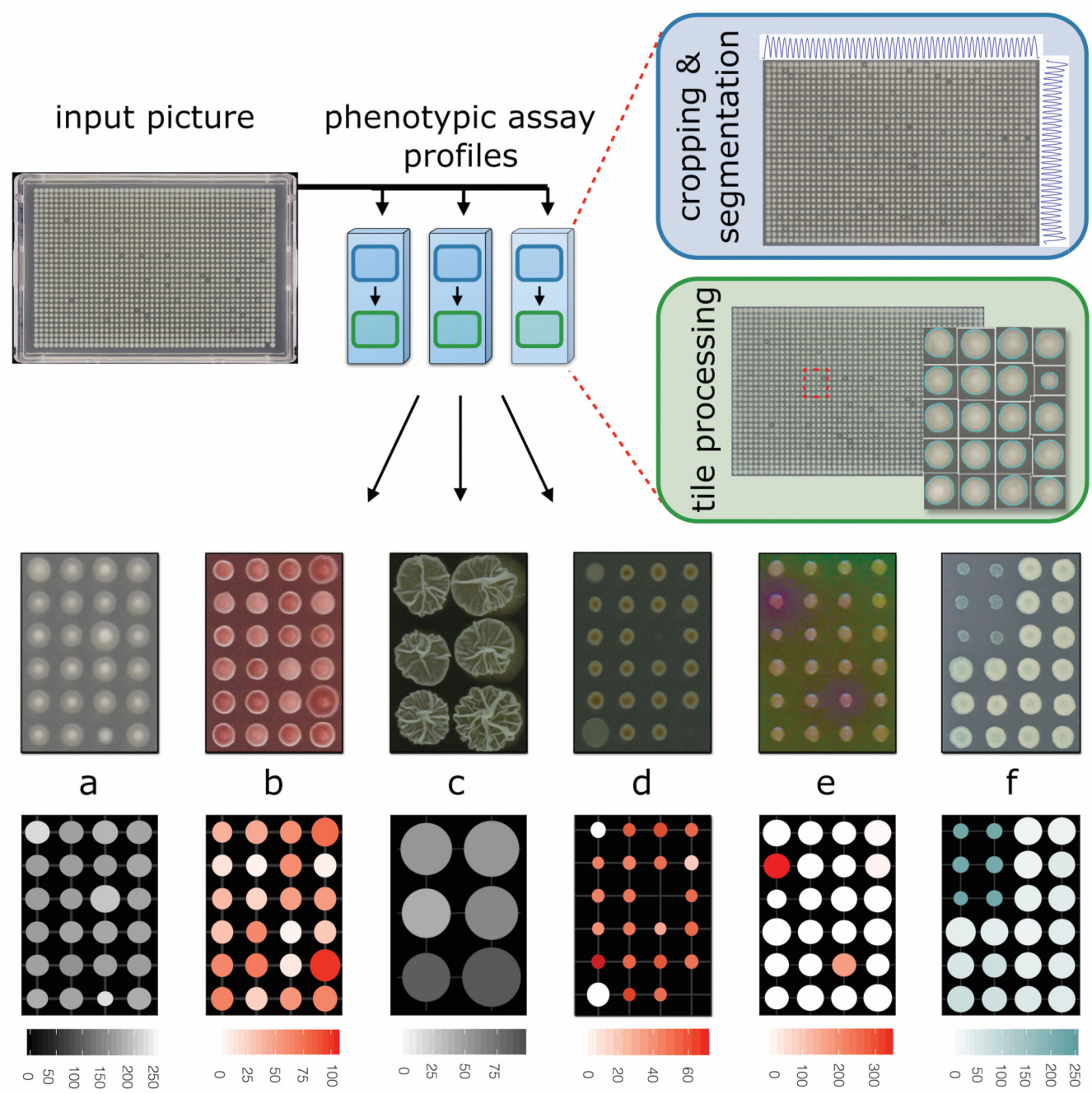

Figure 1. Software design schematic overview.

The colony array is rotated, cropped and segmented to tiles. Tiles are then processed individually for colony identification and are fed to the different phenotype quantification modules: **a.** Colony integral opacity. **b.** Colony biofilm formation. **c.** Colony morphology. **d.** Sporulation. **e.** Envelope integrity. **f.** β-galactosidase activity. Top row: part of quantified colony array (representative images). Bottom row: disk area represents quantified colony area, disc color represents quantified phenotype intensity (in arbitrary units).

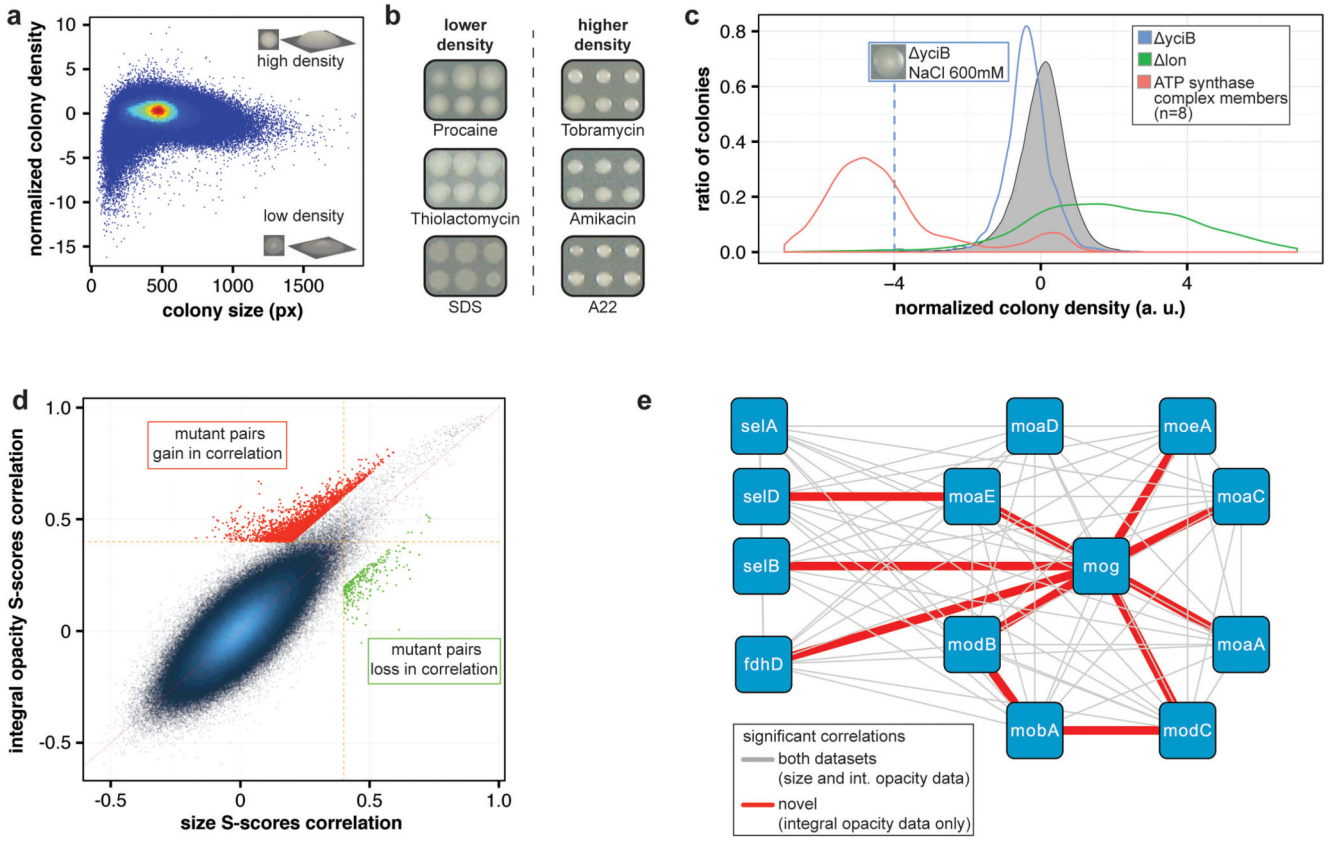

Figure 2. Colony integral opacity is a sensitive metric of growth fitness

a. Colony integral opacity has better resolution for small-sized colonies. Normalized colony density (density=integral opacity/size) versus colony size for all *E. coli* chemical genomics data 19. Inset shows 3D reconstructions of a representative high- and low-density colony. **b.** Examples of chemical conditions with overall higher or lower colony density in *Nichols et al.* 19; a small part of a representative 1536-colony array is shown (n>15) **c.** Density phenotypes hold information not available by size alone. Normalized colony density for all *E. coli* mutants profiled in >250 conditions 19 (shown in gray). Mutants of the ATP synthase complex exhibit severe growth defects and an overall decreased colony density, while the Rcs-inducing *Δlon* strain is mucoid and features overall higher density colonies (n denotes number of mutants in ATP synthase complex present in our screen). A conditional density phenotype example illustrates the defect of *ΔyciB* strain in high NaCl concentration (representative image of n=6). **D.** Re-analysis of fitness data 19 using colony integral opacity as a fitness metric (S-scores are calculated for each mutant-condition interaction and reflect the fitness difference between perturbed and unperturbed state, corrected for data reproducibility) yields more significant correlations among phenotypic signatures of related genes. Red data points indicate mutant pairs for which we gained association power when calculating fitness scores using colony integral opacity instead of colony size (original dataset); green data points indicate opposite situation, **e.** Integral opacity-derived phenotypic signature correlations capture additional interactions (shown in red) among genes coding for molybdenum cofactor biosynthesis and related processes.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

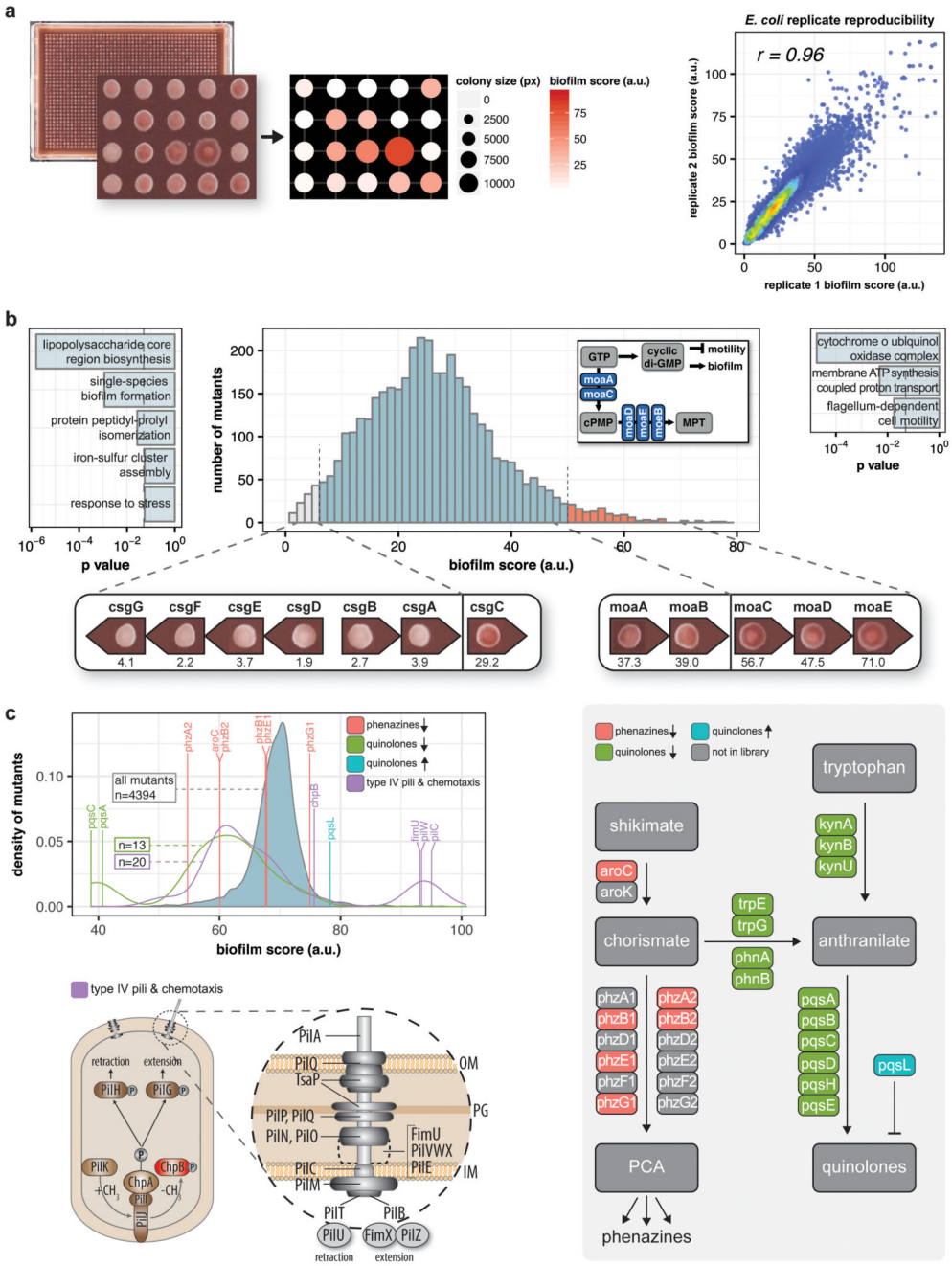

Figure 3. High throughput quantification of macrocolony biofilm formation

a. Colony size and biofilm formation are automatically quantified using Iris. Zoom in a representative *E. coli* 1536-format mutant array plate and quantification of both colony size and biofilm (CR staining). Right: *E. coli* biofilm formation technical replicate reproducibility. **b.** Distribution of colony biofilm formation phenotype in *E. coli* KEIO collection 29 with outlier thresholds colored differently (2.5% each side). Values represent the average biofilm score of technical and biological replicates (n=8; see Methods) in arbitrary units (a.u.). Left, right: outlier GO enrichments (Benjamini-Hochberg-corrected p-

values 63), dotted line is p-value 0.05. Representative images (n=6) demonstrate the deficiency of the curli biosynthesis mutants in biofilm formation, and mutants of the molybdopterin biosynthetic process with increased biofilm formation. **c.** Distribution of colony biofilm formation phenotype in *P. aeruginosa* PA14 transposon mutant library 30. Values represent the average of 4 replicates (Methods). Density plots demonstrate the phenotype of selected mutant groups, and n denotes number of mutants in group. Bottom left: twitching motility mutants (purple in density plot), including mutants of type IV pili machinery and interlinked chemotaxis system, are impaired in biofilm formation (exceptions are *pilC*, 2 of the minor pilins: fimU and *pilW*, and *chpB*; see main text). Right: mutant groups in the phenazine (red in density plot) and PQS (green in density plot) biosynthesis pathways also show decreased biofilm formation. The *pqsL* mutant that overproduces PQS shows increased CR staining. Only *phz2* operon mutants exhibit defects in colony biofim formation. Mutants in gray are not part of the transposon mutant library.

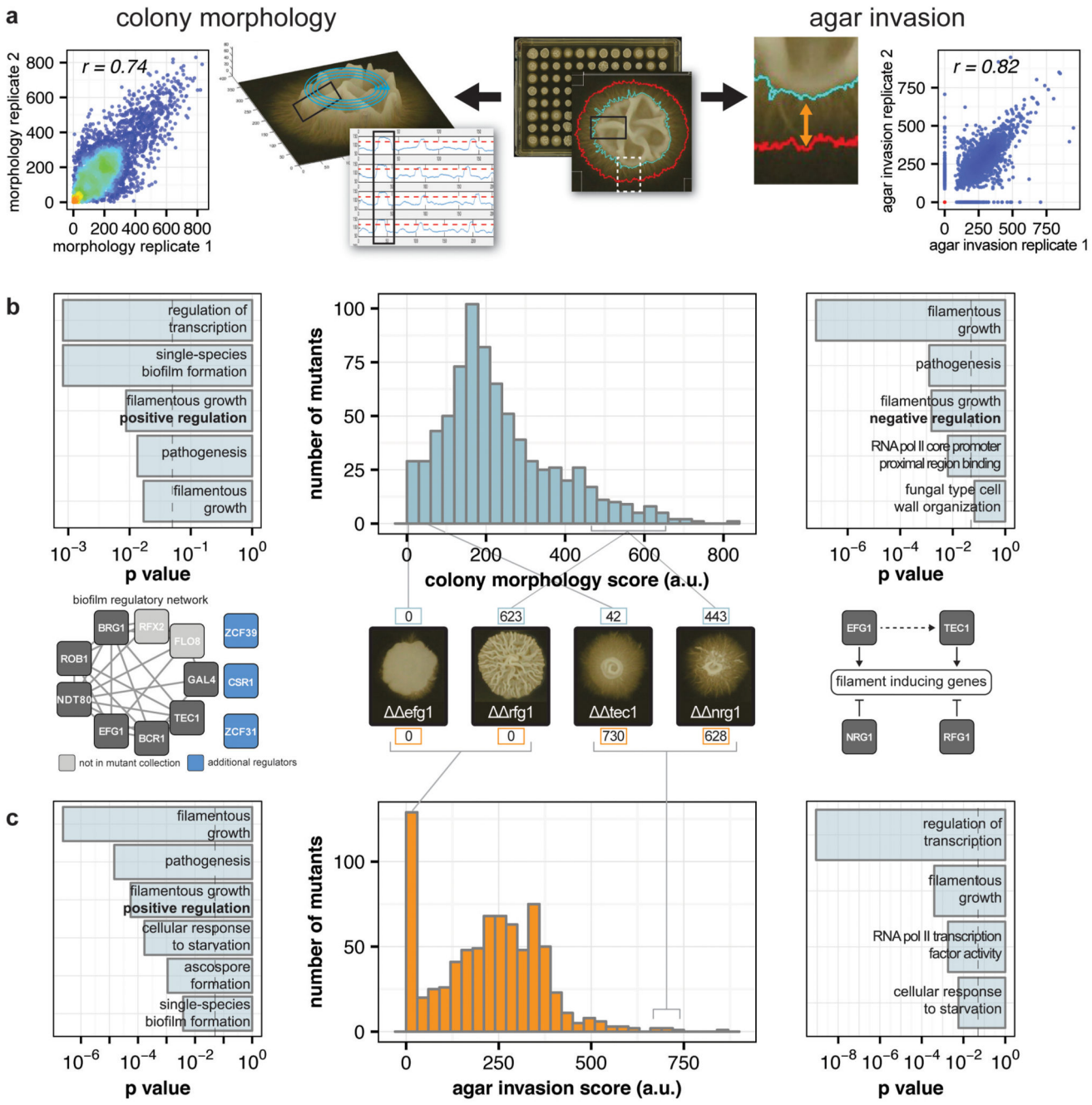

Figure 4. Quantification of colony morphology and invasive filamentation

a. Colony morphology and in-agar growth quantification in *C. albicans* arrayed colonies. Left: colony morphology (structure complexity – wrinkleness) is quantified in Iris by means of an algorithm that measures brightness peaks (over threshold illustrated in red dotted line) by traversing the colony in concentric circles (illustrated by blue circles over a 3D representation of a colony where brightness was used as the z-axis). Right: in-agar growth is quantified by sequentially applying 2 thresholding algorithms of different sensitivities to identify inner colony and halo. Scatter plots illustrate the technical replicate reproducibility

Europe PMC Funders Author Manuscripts

of both readouts. **b.** Distribution of colony morphology in *C. albicans* mutant collections 10, 50 and outlier GO enrichments (BH-corrected p-values 63); dotted line is p-value 0.05. Examples of morphology-impaired outliers include a network of transcription regulators involved in biofilm and colony structure formation, while examples of mutants showing increased colony structure complexity include repressors of filamentous growth Nrg1 and Rfg1. Representative colony pictures of mutants (n=4) from the 96-format colony arrays are shown to illustrate that morphology outliers (Iris-reported value over colony picture) do not necessarily show increased in-agar growth (Iris-reported value under colony picture) and vice-versa. Bottom right: a network with those key activators and repressors involved in *C. albicans* filamentation. Bottom left: Our screen recapitulates all genes annotated as biofilm regulators. Mutants in light gray were not in mutant collection, mutants in blue were not part of the network adapted from 51, 52. **c.** Distribution of in-agar growth in *C. albicans* mutant collections and GO enrichments of outlier mutants (BH-corrected p-values 63); dotted line is p-value 0.05. Values in b. and c. represent the average of technical and biological replicates (n=4; see Methods).