Synthetic Biology

Letter

pubs.acs.org/synthbio

Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition

³ Noah Davidsohn,^{†,||} Jacob Beal,^{‡,||} Samira Kiani,[†] Aaron Adler,[‡] Fusun Yaman,[‡] Yinqing Li,[†] Zhen Xie,[§] ⁴ and Ron Weiss^{*,†}

s [†]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts

6 02139, United States

7 [‡]Raytheon BBN Technologies, 10 Moulton Street, Cambridge, Massachusetts 02138, United States

8 [§]Bioinformatics Division/Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and

9 Technology, Tsinghua University, Beijing 100084, China

10 **Supporting Information**



ABSTRACT: A long-standing goal of synthetic biology is to rapidly engineer new regulatory circuits from simpler devices. As 11 circuit complexity grows, it becomes increasingly important to guide design with quantitative models, but previous efforts have 12 been hindered by lack of predictive accuracy. To address this, we developed Empirical Quantitative Incremental Prediction 13 (EQuIP), a new method for accurate prediction of genetic regulatory network behavior from detailed characterizations of their 14 components. In EQuIP, precisely calibrated time-series and dosage-response assays are used to construct hybrid phenotypic/ 15 mechanistic models of regulatory processes. This hybrid method ensures that model parameters match observable phenomena, 16 using phenotypic formulation where current hypotheses about biological mechanisms do not agree closely with experimental 17 observations. We demonstrate EQuIP's precision at predicting distributions of cell behaviors for six transcriptional cascades and 18 19 three feed-forward circuits in mammalian cells. Our cascade predictions have only 1.6-fold mean error over a 261-fold mean 20 range of fluorescence variation, owing primarily to calibrated measurements and piecewise-linear models. Predictions for three feed-forward circuits had a 2.0-fold mean error on a 333-fold mean range, further demonstrating that EQUIP can scale to more 21 complex systems. Such accurate predictions will foster reliable forward engineering of complex biological circuits from libraries of 22 standardized devices. 23

24 **KEYWORDS:** synthetic biology, systems biology, genetic circuits

ne of the key challenges in synthetic biology is to 25 accurately predict the behavior of novel biological 2.6 27 systems, thereby enabling faster and more effective engineering 28 of such systems.¹⁻³ This challenge is becoming a critical issue, given the growing gap between the exponential increase in 29 30 length of DNA sequences that can be readily synthesized³⁻⁹ 31 and the much slower increase in the complexity of genetic 32 circuits that have been demonstrated.⁹⁻¹⁴ Accurate predictions 33 of the behavior of genetic circuits are an important ingredient 34 for addressing this gap. As the number of genetic elements in a 35 circuit increases, the number of candidate designs increases 36 exponentially. Accurate predictions help cope with this 37 exponential explosion by dramatically reducing the number of 38 candidate designs that must be considered. Predicting circuit

behavior, however, has been extremely difficult, and without 39 reliable predictions, even relatively simple circuits have typically 40 required extensive and costly tuning to achieve the desired 41 results.^{9,10,12,15} 42

Recently, there have been major steps toward improving the 43 accuracy of genetic circuit predictions. First, genetic elements 44 are now being characterized using calibrated and standardized 45 measurements.^{16–19} Second, several investigations have pro- 46 vided means of combining families of primitive elements of a 47 transcriptional unit, such as promoters and 5'UTRs, in order to 48 more reliably and predictably control constitutive gene 49



Received: June 13, 2014



Figure 1. Stages of EQuIP from data gathering to predictions. (A) Experimental observation of the behavior of regulatory and constitutive elements in cells, where these elements are combined with additional biological circuitry for calibrated measurement (top: stage 1). Behavior of constitutive elements is measured over time. For regulatory elements, the relationship between input, circuit copy number (indicated by different colored lines), and output is measured at a single time point. Data is used to build rate functions for time-dependent regulated production and for loss of protein concentration, which can be mathematically integrated for computational simulation (bottom: stage 2). (B) The behavior of a biological circuit is predicted by linking production functions for each regulatory relation and loss functions for each relevant protein, according to circuit topology, then simulating concentrations over time according to the network of rate functions (stage 3).

so expression levels.^{20–23} For gene regulation, progress has been s1 made toward accurate predictions of single self-regulating s2 negative feedback transcriptional units,²⁴ which may exhibit s3 decreased variability,²⁵ increased variability,²⁶ or oscillatory s4 behavior,²⁷ depending on conditions. For multicomponent s5 regulatory circuits, however, there is a critical need for a s6 fundamentally new approach to prediction. Prior efforts have s7 either focused on characterizing a complex circuit and then s8 predicting the influence of modulating or replacing specific s9 elements^{11,27–29} or else have suffered from reduced precision 60 when parts were first characterized and then subsequently used 61 for predictions of a more complex circuit.

Prior methods for predicting multicomponent regulatory 62 63 circuits have typically relied on explicit biochemical models, ^{10,28} 64 such as Hill functions or chemical reaction networks, that 65 depend strongly on the completeness and correctness of 66 models of relevant cellular mechanisms. Such models are also 67 frequently under-constrained by experimental data and thus 68 require significant parameters to be set by heuristics or untested 69 assumptions, rather than through direct (or indirect) 70 experimental observations. This is a critical problem for 71 predictions: the inherent uncertainty of an under-constrained 72 model means that the same observation can be explained by a 73 number of different sets of parameter values.³⁰ The predictive 74 accuracy of a model is therefore impaired, because even if the 75 model fits observations for one particular use, if the wrong 76 parameters are chosen it is likely to fail on future predictions. In 77 sum, although there has been much recent progress in characterization and prediction of genetic parts and circuits, 78 79 even the behavior of a "simple" circuit such as a two-repressor 80 cascade cannot generally be predicted accurately and reliably. 81 We thus focus on the prediction of combinational genetic 82 circuits (i.e., circuits without feedback or state), as both an

important goal in its own right and as a step toward prediction ⁸³ of circuits with more complex dynamics such as oscillations and ⁸⁴ bistability. ⁸⁵

We address the current challenges of multicomponent circuit 86 prediction with a new method, Empirical Quantitative 87 Incremental Prediction (EQuIP), that models expression of 88 each gene using a piecewise function of regulatory inputs, 89 circuit copy number, and time, based strictly on high-precision 90 experimental observations. EQuIP predicts the behavior of a 91 biological circuit by mathematically composing, in accordance 92 with circuit topology, these gene expression models along with 93 models of exponential dilution and decay. As can be expected, 94 accurate circuit predictions require accurate models. To this 95 end, EQuIP ensures that significant observable phenomena are 96 incorporated in each gene expression model whether or not 97 they agree with current mechanism hypotheses and also ensures 98 that every parameter of the model is directly grounded in 99 experimental data. Given current limitations in the under- 100 standing of biochemical mechanisms and in the ability to 101 determine relevant parameter values through experiment, the 102 flexibility of piecewise approximation is highly valuable. A 103 piecewise function can directly approximate unmodeled or 104 poorly modeled mechanisms and can substitute simple 105 empirical functions for mechanisms whose parameters cannot 106 be determined from observable data. EQuIP thus combines 107 mechanistic models (i.e., derived from the underlying molecular 108 processes) and phenotypic models (i.e., models aiming to 109 capture observed behavior with minimal assumptions), using 110 mechanistic models where the underlying parameters can be 111 adequately determined and phenotypic approximation where 112 they cannot. This combination greatly improves the accuracy 113 with which the behavior of biological circuits can be predicted. 114



Figure 2. EQuIP characterization via time-series and dosage-response assays. (A) Biological circuit architecture for calibrated measurements, using fluorescent reporter proteins to quantify induced expression of repressor, regulated expression of output, and constitutive expression as an indicator of relative circuit copy number. EBFP2, a blue fluorescent protein, is input (IFP); EYFP, a yellow fluorescent protein, is output (OFP); mKate, a red fluorescent protein, is constitutive. (B) Time series characterization shows a linear increase in the fraction of cells constitutively expressing a fluorescent reporter, beginning a short time after transfection, until reaching saturation at approximately 70% transfection efficiency. Dotted lines show ± 2 standard deviations. (C) Progression of mean fluorescence is similar for constitutive and activator-driven fluorescence, implying little impact from transcriptional activation delays. Normalized expression is computed by dividing by mean MEFL for t = 48-72. (D) Relations between input, copy number, and output for TAL14, TAL21, and LmrA: data from 12 different inducer dosages is segmented into subpopulations by constitutive fluorescence (plus marks) and grouped by subpopulation across dosages (colored lines). Insets show histograms of constitutive fluorescent protein expression used for segmenting the subpopulations; only colored bins have sufficient samples and separation from untransfected cells and are therefore included in the input/output curves. Extrapolation beyond the range measured in each transfer curve experiment is shown with dashed lines (see Supporting Information Section 6).

f1

To provide circuit predictions, EQuIP operates in three main 115 116 stages (Figure 1). In the first stage (Figure 1a top), we measure 117 gene expression of constitutive and regulated elements using calibrated flow cytometry assays, with all measurements 118 119 converted to equivalent standardized units. These measure-120 ments are taken at various combinations of time and regulatory 121 inputs sufficient to precisely characterize the expression dynamics of all relevant circuit components. In the second 122 123 stage, these measurements are used to compute two sets of rate 124 functions (Figure 1a bottom). A regulated production function 125 is a mapping from regulatory input (e.g., concentration of a 126 transcriptional repressor), circuit copy number, and time, to the gene production rate. A loss function specifies the rate at which 127 molecule's concentration decreases due to dilution or 128 а 129 degradation. We can combine these two functions to describe 130 the time evolution of a regulated gene product. In the third 131 stage of EQuIP, the rate functions for multiple elements are 132 combined to simulate the time evolution of more complex 133 regulatory circuits (Figure 1b). Time evolution simulations are 134 carried out by composing the rate functions according to the 135 circuit topology and computing the integral with respect to

time for various combinations of input and circuit copy 136 number. As currently formulated, EQuIP can be applied to 137 combinational circuits with relatively strong expression and low 138 cross-interference, in conditions similar to those under which 139 the devices were characterized. In the remainder of the paper, 140 we present details of EQuIP and then demonstrate that its 141 circuit simulations accurately predict experimental observations. 142 In particular, we characterize three regulatory relations 143 (transcriptional repressors TAL14, TAL21, and LmrA, each 144 acting on a corresponding promoter) in mammalian HEK293 145 cells and use those characterizations to precisely predict the 146 behavior of all six two-repressor cascades that can be made 147 from these repressor/promoter pairs, as well as three feed- 148 forward circuits constructed from the same elements. We 149 conclude by evaluating the contribution of the various 150 components of EQuIP to its precision in predicting composite 151 circuit behavior. 152

RESULTS AND DISCUSSION

The goal of EQuIP is to predict the behavior of regulatory 154 circuits, which we test in this paper through prediction of two-

153

Letter



Figure 3. EQuIP produces close agreement between computational predictions and experimental data for two-stage cascades. (A) Biological circuit architecture for repressor cascades using TAL14, TAL21, and LmrA (repressor choice indicated as R1 and R2). Fluorescent reporters quantify input, output, and relative circuit copy number as in Figure 2. (B, C) Plots compare input/output relations for predictions (circles) and experimental data (pluses) for two of the cascades (others shown in Supporting Information Section 6). The predictions include all points that use extrapolation for less than 10% of the simulation steps. Insets as in Figure 2D.

156 repressor transcriptional cascades and feed-forward circuits. We 157 model a transcriptional circuit using two types of functions 158 (regulated production and loss), each taking current concen-159 trations as input and yielding rates of concentration change as 160 output. A transcriptional circuit, such as a two-repressor 161 cascade, is then simulated by integrating a network of regulated 162 production and loss functions, as described above. To restrict 163 the scope of the problem, we consider only transient 164 transfections of combinational circuits comprising orthogonal 165 regulatory elements (see Supporting Information Section 2) 166 and use each repressor/promoter pair at most once in any 167 given circuit.

The first stage of EQuIP is to gather experimental data 168 169 characterizing the regulated production dynamics for each 170 repressor/promoter pair and the loss dynamics for each protein (Figure 1A, top). We characterize these dynamics with two 171 172 experiments (Figure 2): a time series assay, which provides the 173 mechanistic components of both production and loss models, and a dosage-response assay measured at a single time point, 174 which provides the phenotypic components of the production 175 model. To obtain precise and commensurate units in our 176 models, we apply the TASBE protocol for calibrated flow 177 cytometry, which allows us to use Molecules of Equivalent 178 Fluorescein (MEFL)³¹ as a consistent proxy unit for protein 179 180 concentration (see Supporting Information Section 3).

 f_2

All time-series and dosage-response characterization assays use circuits built with the same template (Figure 2A). The purpose of this circuit topology is to measure the behavior of a repressor/promoter pair at various levels of repressor concentration. We chose to regulate concentration of repressor 185 by doxycycline/rtTA induction^{32,33} as indicated by EBFP2. The 186 concentration of the output gene product is indicated by EYFP, 187 and constitutive mKate serves as a transfection marker and an 188 indicator of relative circuit copy number.³⁴ All promoters that 189 we characterize in this paper are hybrid promoters that also 190 require Gal4 activation (see Supporting Information Section 2 191 for discussion of this modular approach to mammalian 192 promoter design). 193

For the time series experiment, we measured constitutive and 194 transcriptional activator driven expression for 72 h post- 195 transfection. We found that the fraction of cells with observable 196 (*i.e.*, above autofluorescence) constitutive expression of a 197 fluorescent protein increases linearly over time, beginning 198 some short time after transfection, and finally saturates at 199 approximately 70% of cells (Figure 2b). This observation is 200 consistent with typical lipofection protocols and a model 201 whereby plasmids enter the nucleus during mitosis (per the 202 standard lipofection hypothesis³⁵) in an unsynchronized 203 population of mammalian cells. Given the expected stability 204 of our fluorescent and repressor proteins, this mechanism plus a 205 constant rate of constitutive production can be used to create a 206 quantitative model of transfection and fluorescent protein 207 production. Fitting against both the percentage of expressing 208 cells and mean constitutive mKate in expressing cells (Figure 209 2C) gives a mean initial delay of 25 h and cell division on 210 average every 20 h (which correlates well with independent 211 hemocytometer measurements as well; for detailed discussion 212 of growth rate measurements, see Supporting Information 213



Figure 4. Comparison of the precision of EQuIP predictions for two-stage cascades. (A) Ratio of highest and lowest output means for all induction/ copy-number subpopulations for given cascades (*e.g.*, the highest and lowest plus symbols in Figure 3B and C) and respective mean prediction errors. (B) Prediction errors for Hill function models fit to partial implementations of EQuIP vs the full EQuIP method, showing the improvement in both accuracy (mean error) and precision (95% envelope).

214 Section 4). We also note that expression of a fluorescent 215 protein activated by constitutive rtTA3 and VP16Gal4 has no 216 significant time lag compared to expression of mKate from a 217 constitutive promoter, implying that, for this circuit, transcrip-218 tional activation delays are not significant and may be omitted 219 from our production models. In contrast, even a short lag until 220 repressors accumulate to the levels required for them to have significant impact on gene expression may result in strong 221 222 transient expression from their respective promoters, and 223 therefore, we model the time course of repressor accumulation. From this experiment, we derive three mechanistic elements 224 225 of our models. First, the cell division time provides a good 226 approximation of the loss rate, since the proteins used are expected to be relatively stable. Second, the cell division time is 227 also used to create an inverse function that takes an observed 228 229 output expression and calculates the production rate over time 230 that would produce that output expression (Supporting 231 Information Section 5). This inverse function will be used to 232 create the production rate function of a transcriptional unit 233 from dosage-response data. Finally, simulations of gene 234 expression take into account the mean initial delay in 235 determining the length of time to simulate (Supporting 236 Information Section 4).

We characterize each repressor device with a dosage-237 238 response experiment using the same characterization circuit (Figure 2A) as the time series experiment. For these 239 experiments, we measure output as a function of input at a 240 single point in time. We characterized the regulatory relation-241 ship between three transcriptional repressors (TAL14, TAL21, 242 and LmrA; see Supporting Information Section 2) and a 243 corresponding promoter for each, 72 h after transfection 244 (Figure 2D). The observed relationships between input and 245 output fluorescence were strongly affected by the relative 246 number of circuit copies in the cells. Currently, there is not 247 sufficient understanding of the underlying biological processes 248 to create well-constrained models, based entirely on mecha-249 250 nistic principles, that accurately match the experimentally observed input/output relationships in Figure 2D. Instead, we 251 estimate output gene expression for a given input level and 252 relative copy number phenotypically by piecewise interpolation 253 254 or extrapolation of the observed outputs (lines in Figure 2D). 255 This is then transformed into a hybrid phenotypic/mechanistic 256 model of regulated gene production using the inverse function 257 derived from the time-series experiment (Supporting Information Section 5). Hence, this regulated production model retains 258 every feature of the experimentally observed behavior. 259

To validate EQuIP, we constructed all two-repressor 260 cascades comprising TAL14, TAL21, and LmrA following the 261 architecture shown in Figure 3A. Figure 3B and C illustrates the 262 f3 72 h input/output predictions vs experimental data for the 263 TAL14-TAL21 and TAL21-TAL14 cascades, respectively. The 264 experimentally observed output levels for different combina- 265 tions of input and copy number have a wide range: across the 266 six cascades, there is a 261-fold geometric mean difference 267 between the highest and lowest subpopulation output means 268 (e.g., the ratio of highest and lowest plus symbols in Figure 3B 269 and C). The mean error of predicted versus observed output 270 across all input/copy-number combinations is only 1.6-fold for 271 all six transcriptional cascades (Figure 4A). Predicting the 272 f4 output across many subpopulations also provides a prediction 273 of the distribution of output expression for the overall 274 population. The accuracy with which EQuIP predicts 275 population mean and variation is even better than for individual 276 subpopulations: the mean error of predicted versus exper- 277 imentally observed output across all cascades and inductions is 278 1.4-fold for both population mean and population standard 279 deviation (Supporting Information Section 6). 280

With such accuracy, EQuIP may guide circuit design and 281 debugging. For example, EQuIP correctly predicts which 282 combinations of repressors are best matched to provide the 283 greatest differential expression between fully induced and 284 uninduced states in the cascades. Specifically, EQuIP predicts 285 that TAL14-TAL21 and TAL21-TAL14 cascades will have 286 significantly stronger gain than all cascades involving LmrA 287 (due to TAL21 and TAL14 having a better match in their 288 dynamic ranges) and that TAL21-TAL14 will have approx- 289 imately twice the gain of TAL14-TAL21, and these predictions 290 are borne out by our experiments (Supporting Information 291 Section 6). 292

We further evaluate the contribution to accurate prediction 293 of different aspects of the EQuIP method. The first two stages 294 of EQuIP (Figure 1A) consist of a sequence of data gathering 295 and processing steps to produce the model for each device. 296 Figure 4B evaluates the relative contribution of each step in this 297 sequence to the final high prediction accuracy by comparing 298 with a typical prior method—ordinary differential equations 299 using Hill equations fit to device characterization data—applied 300 to the data produced by each step in the data processing 301



Figure 5. EQuIP produces close agreement between computational predictions and experimental data for feed-forward circuits. (A) Biological circuit architecture for feed-forward circuits using TAL14, TAL21, and LmrA (repressor choice indicated as R1 and R2). Fluorescent reporters quantify input, output, and relative circuit copy number as in Figure 2. (B) Plot compares input/output relations for predictions (circles) and experimental data (pluses) for one of the circuits (others shown in Supporting Information Section 6). The predictions include all points that use extrapolation for less than 10% of the simulation steps. Insets as in Figure 2D. (C) Ratio of highest and lowest output means for all induction/copy-number subpopulations for given feed-forward circuits (*e.g.*, the highest and lowest plus symbols in B) and respective mean prediction errors.

302 sequence. As described in Supporting Information Sections 3, 303 10, and 11, EQuIP first converts flow cytometry data to 304 calibrated MEFL, then uses the constitutive marker to separate 305 the subpopulation of successfully transfected cells, segments 306 into bins by constitutive fluorescence level, then finally builds a 307 piecewise model using the means of each bin. We thus created 308 models generated from (1) population means in arbitrary units 309 (our baseline model), (2) population means in calibrated 310 MEFL, (3) population mean MEFL for transfected cells only, 311 and (4) per-bin MEFL for transfected cells only. Each model is 312 an ODE using Hill equations parametrized by curve fit against 313 the observed data for each of the three repressors. We then 314 compare the accuracy of EQuIP and these four models in 315 predicting population means across the full range of 316 doxycycline inductions of the various cascades. The results in Figure 4B indicate that the most important contributions to 317 EQuIP's improved precision versus the baseline model come 318 319 from calibration of measurement units and modeling with 320 piecewise functions (full details in Supporting Information 321 Section 7). The intermediate steps are prerequisites for 322 piecewise models but do not appear to markedly improve 323 prediction quality on their own. In all cases, the inaccuracies in 324 prediction appear to derive primarily from the insufficient 325 constraints that the experimental data provides for fitting Hill 326 equations. Thus, although the fit often appears good (as shown 327 in Supporting Information Section 7), this may not accurately represent the true system and may not correlate well with the 328 predictive accuracy. Indeed, all three intermediate models show 329 no statistically significant differences in performance (Support- 330 ing Information Section 7). 331

Finally, to validate the generalizability of the EQuIP method, 332 we tested its efficacy in predicting the behavior of a more 333 complex feed-forward circuit. This circuit, shown in Figure 5A, 334 f5 is similar to the cascade except that it adds a second path for 335 repressing the output directly through Dox induction of the 336 second repressor. We applied EQuIP to predict the behavior of 337 all six possible feed-forward circuits 72 h post-transfection. We 338 then constructed the three with the greatest variety of predicted 339 behaviors and observed them experimentally. As before, we find 340 that the predictions have a high mean accuracy, as illustrated by 341 the example comparison of predicted and observed expression 342 levels in Figure 5B. Across all three feed-forward circuits, there 343 is an overall mean error of only 2.0-fold across a 333-fold 344 geometric mean difference between the highest and lowest 345 subpopulation output means (Figure 5C). Such a small 346 degradation in accuracy in comparison to the cascades is 347 expected for a more complicated circuit and is an indicator that 348 EQuIP is likely to scale to even more complex circuits. 349

For synthetic biology to become a full-fledged engineering 350 discipline, it must be possible to accurately predict the behavior 351 of novel biological circuits from that of their constituent parts. 352 Our results with EQuIP on six transcriptional cascades and 353 354 three feed-forward circuits are the first demonstration that 355 highly accurate prediction of circuit behavior is possible and 356 provide a quantitative benchmark for future efforts to be 357 compared against. One of the most remarkable features of our 358 results is that they are accomplished entirely through readily 359 accessible observations of fluorescence, without any detailed 360 biochemical analysis or modeling. This does not argue against 361 the value of detailed biochemical modeling: indeed, we were 362 able to use mechanistic biochemical models for the temporal 363 components of our production and loss functions. Rather, our 364 results demonstrate the power of using precise measurements 365 to inform a composable model that describes precisely those 366 phenomena that can be experimentally observed, no more and 367 no less.

Our results show a number of opportunities where improved 368 369 modeling or observation could further increase the quality of predictions. For example, Figure 3B and C and Supporting 370 371 Information Figures 11 and 12 demonstrate that EQuIP 372 generally provides its best predictions when neither circuit copy 373 number nor induction are particularly high. One improvement 374 would be to obtain additional input/output data at higher input 375 levels or to genetically engineer regulatory devices with 376 stronger responses at lower input levels, thus decreasing the amount of extrapolation required (Supporting Information 377 Section 5). Another issue is that extreme fluorescence intensity 378 values can be affected at the low end by cell autofluorescence 379 380 and at the high end by PMT saturation (though this affects only the <5% of our predictions that are in these ranges); this can be 381 382 addressed by assays that vary plasmid dosages and instrument 383 settings for different levels of induction. Other possible areas 384 for improvement are understanding the effect of large circuit 385 copy numbers (which may influence cell behavior or create 386 metabolic load), and accounting for retroactivity effects (e.g., 387 from one TALER regulating multiple downstream promoters) 388 and cellular resource sharing (e.g., VP16Gal4 as a driver for 389 multiple hybrid promoters). With appropriate characterization 390 experiments, such effects should be able to be incorporated as 391 new empirical rate terms, similar to the production and loss 392 models already in the EQuIP framework.

More generally, the specific EQuIP implementation 393 394 presented here can be applied only to circuits that are 395 combinational (meaning there is no feedback) and in which 396 cells do not exhibit strongly divergent behaviors under the same conditions (e.g., noise-induced bistability). Our results, 397 398 however, provide a basis for extending EQuIP to larger and 399 more complex circuits, and to circuits that include feedback (via 400 initial state assumptions) and divergent populations (using 401 distributions rather than means), as discussed in Supporting 402 Information Section 5.1. The ability to predict circuit behavior 403 is highly valuable for engineering biological systems, as it allows 404 efficient selection of circuit elements and offers guidelines for 405 optimization of devices to obtain a desired function. 406 Accordingly, EQuIP supports the synthetic biology goal of 407 creating libraries of modular, standard, and well-characterized 408 components for rapid development of complex systems. Our 409 framework may also be used for studying natural systems, 410 although accurate predictions may initially be more difficult due 411 to the complexity of many natural regulatory interactions. 412 EQuIP thus forms a basis both for advances in design tools and 413 for new investigations in systems biology. Combining these 414 advances with emerging libraries of biological devices will usher 415 in a new era of exponential growth in our ability to engineer 416 biological systems.

417

METHODS

Culture Conditions. HEK 293 FT cells (Invitrogen) were 418 cultured in DMEM medium (CellGro), supplemented with 419 10% FBS (PAA Laboratories), 2 mM L-glutamine (CellGro), 420 1% Strep/pen (CellGro), 1% non-essential amino acids 421 (NEAA) (HyClone), and 10 000× Fungin (Invivogen) at 37 422 °C and 5% CO₂. Cells were passaged in a 100 mm dish by 423 removing culture media, adding 2 mL 0.05% trypsin, waiting at 424 room temperature for 2 min and then resuspending the cells in 425 5 mL of cell culture media and diluting to desired concentration 426 with additional cell culture media. 427

Transfection. Transfections were carried out with Meta- 428 fectene Pro (Biontex Laboratories). Cells were seeded 1 day 429 prior at 2×10^5 cells per well in a 24 well plate. 500 ng of DNA 430 was mixed into 60 µL of DMEM (without supplements). 1.5 431 μ L of Metafectene was then added and the tube was gently 432 mixed and kept at room temperature for 15 min to form the 433 DNA-liposome complex. Fresh media was added to the cells 434 directly prior to transfection (500 μ L of DMEM with 435 supplements). The DNA-Metafectene solution was then 436 added dropwise to the well. Induction of the circuit was 437 performed at this time as well by the addition of a small 438 molecule (*i.e.*, doxycycline). The media was subsequently 439 changed daily with the appropriate amount of inducer. Each 440 circuit was realized with each transcriptional unit encoded on a 441 separate plasmid, for a total of 6 plasmids (5 plasmid circuits 442 add a blank plasmid), and cotransfected: see Supporting 443 Information Section 2 for details of promoter design and 444 Supporting Information Section 8 for plasmid sequences. The 445 DNA for the circuits transfected were in the ratio 1:3:3:1:1:1; 446 where the transcriptional units that contained the "TRE" 447 promoter were the ones that were transfected at 3× the amount 448 of the others for signal matching purposes. For the time series 449 experiment, we measured EBFP2 with 2000 nM Dox for rtTA 450 activation. EYFP was measured for Gal4 activation with 0 nM 451 Dox, and we used TAL21 for R1. We also assumed that the 452 time dynamics are not significantly affected by choice of 453 repressor. Dose-response data was taken with a logarithmic 454 series of Dox dosages. Feed-forward circuits data was taken in a 455 separate experiment with a slight variation on the protocol. See 456 Supporting Information Section 9 for more details. The cell-to- 457 cell variation due to intracellular variation in copy number was 458 typically small in our experiments (Supporting Information 459 Section 10). 460

Flow Cytometry. Flow cytometry data was taken at 72 h 461 post transfection. Cells were again trypsinized as previously 462 described. The cells were then centrifuged at 150g for 10 min at 463 4 °C. The supernatant was removed and the cells were 464 resuspended in 1× PBS that did not contain calcium or 465 magnesium. A BD LSR Fortessa was used to take flow 466 cytometry measurements with the following settings: EBFP2, 467 measured with a 405 nm laser and a 450/50 filter, EYFP, 468 measured with a 488 nm laser and a 530/30 filter, and mKate, 469 measured with a 561 nm laser and a 610/20 filter. Flow 470 cytometry data was analyzed as described in Supporting 471 Information Sections 3, 11, and 12. EQuIP also included 472 internal cross validation and checks to determine the quality of 473 the data collected and identify potential experimental problems 474 (Supporting Information Section 13). 475

Cloning. Creation of the plasmids used for this project was 476 carried out using the Gateway system from Invitrogen. We used 477 a multisite cloning strategy with two entry vectors. One entry 478

479 vector contained the promoter and the other contained the 480 transcription factor or gene. The destination vector was 481 modified from its original sequence to contain an insulator 5' 482 to L4 and a polyadenylation signal 3' to the R1 site.

483 ASSOCIATED CONTENT

484 **S** Supporting Information

485 Details of repressor/promoter design, sequences of all 486 constructed used, details on calibration of flow cytometry 487 data, measurement and estimation of cell division rates, details 488 of modeling and prediction, details of experimental results, and 489 additional experimental and analytical method details. This 490 material is available free of charge via the Internet at http:// 491 pubs.acs.org.

492 **AUTHOR INFORMATION**

493 Corresponding Author

494 *E-mail: rweiss@mit.edu.

495 Author Contributions

⁴⁹⁶ ^{II}N.D. and J.B. contributed equally. N.D. designed and ⁴⁹⁷ performed repressor characterization and two-stage cascade ⁴⁹⁸ experiments, analyzed data, and wrote the manuscript. J.B. ⁴⁹⁹ designed experiments, developed and applied computational ⁵⁰⁰ analysis and prediction techniques, and wrote the manuscript. ⁵⁰¹ S.K. designed and performed feed-forward circuit experiments, ⁵⁰² analyzed data, and wrote the manuscript. A.A. and F.Y. ⁵⁰³ developed and applied computational analysis and prediction ⁵⁰⁴ techniques. Y.L. and Z.X. built initial versions of some DNA ⁵⁰⁵ constructs. R.W. designed experiments, helped develop the ⁵⁰⁶ computation framework, analyzed data, and wrote the manu-⁵⁰⁷ script.

508 Notes

509 The authors declare no competing financial interest.

510 **REFERENCES**

511 (1) Lucks, J. B., Qi, L., Whitaker, W. R., and Arkin, A. P. (2008) 512 Toward scalable parts families for predictable design of biological 513 circuits. *Curr. Opin. Microbiol.* 11, 567–573.

514 (2) Alon, U. (2007) Network motifs: Theory and experimental 515 approaches. *Nat. Rev. Genet.* 8, 450–461.

516 (3) Purnick, P. E. M., and Weiss, R. (2009) The second wave of 517 synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* 518 *10*, 410–422.

(4) Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang,
820 R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie,
821 M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N.,
822 Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall823 Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., 3rd,

524 Smith, H. O., and Venter, J. C. (2010) Creation of a bacterial cell

s25 controlled by a chemically synthesized genome. *Science* 329, 52–56.
(5) Carr, P. A., and Church, G. M. (2009) Genome engineering. *Nat. Biotechnol.* 27, 1151–1162.

528 (6) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, 529 C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules 530 up to several hundred kilobases. *Nat. Methods 6*, 343–345.

531 (7) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009)

532 Golden gate shuffling: A one-pot DNA shuffling method based on 533 Type IIs restriction enzymes. *PloS One 4*, e5553.

(8) Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R.
(2013) Rapid, modular, and reliable construction of complex
mammalian gene circuits. *Nucleic Acids Res.* 41, e156.

537 (9) Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., and 538 Collins, J. J. (2012) Iterative plug-and-play methodology for 539 constructing and modifying synthetic gene networks. *Nat. Methods 9*, 540 1077–1080. (10) Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. 541 (2012) Genetic programs constructed from layered logic gates in 542 single cells. *Nature 491*, 249–253. 543

(11) Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., 544 Levskaya, A., Marcotte, E. M., Voigt, C. A., and Ellington, A. D. (2009) 545 A synthetic genetic edge detection program. *Cell* 137, 1272–1281. 546

(12) Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, 547
 Y. (2011) Multi-input RNAi-based logic circuit for identification of 548
 specific cancer cells. *Science* 333, 1307–1311. 549

(13) Ausländer, S., Ausländer, D., Müller, M., Wieland, M., and 550 Fussenegger, M. (2012) Programmable single-cell mammalian 551 biocomputers. *Nature* 487, 123–127. 552

(14) Lu, T. K., Khalil, A. S., and Collins, J. J. (2009) Next-generation 553 synthetic gene networks. *Nat. Biotechnol.* 27, 1139–1150. 554

(15) Hooshangi, S., Thiberge, S., and Weiss, R. (2005) Ultra- 555 sensitivity and noise propagation in a synthetic transcriptional cascade. 556 *Proc. Natl. Acad. Sci. U.S.A. 102*, 3581–3586. 557

(16) Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., 558 Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., 559 and Endy, D. (2009) Measuring the activity of BioBrick promoters 560 using an *in vivo* reference standard. *J. Biol. Eng.* 3, 4.

(17) Canton, B., Labno, A., and Endy, D. (2008) Refinement and 562 standardization of synthetic biological parts and devices. *Nat.* 563 *Biotechnol.* 26, 787–793. 564

(18) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, 565 M. B. (2005) Gene regulation at the single-cell level. *Science* 307, 566 1962–1965. 567

(19) BealJ.; Weiss, R.; Yaman, F.; Davidsohn, N.; Adler, A. (2012) A 568
 Method for Fast, High-Precision Characterization of Synthetic Biology 569
 Devices; Computer Science and Artificial Intelligence Laboratory 570
 Technical Report; Massachusetts Institute of Technology, Cambridge, 571
 MA. 572

(20) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., 573 Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., 574 Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression 575 via standard transcription and translation initiation elements. *Nat.* 576 *Methods* 10, 354–360. 577

(21) Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., 578 Mutalik, V. K., and Arkin, A. P. (2012) An adaptor from translational 579 to transcriptional control enables predictable assembly of complex 580 regulation. *Nat. Methods* 9, 1088–1094. 581

(22) Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, 582 Y., Arkin, A. P., Endy, D., and Church, G. M. (2013) Composability of 583 regulatory sequences controlling transcription and translation in 584 *Escherichia coli. Proc. Natl. Acad. Sci. U.S.A.* 110, 14024–14029. 585 (23) Salis, H. M. (2011) The ribosome binding site calculator. 586

Methods Enzymol. 498, 19–42. 587

(24) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, 588 M. B. (2007) Accurate prediction of gene feedback circuit behavior 589 from component properties. *Mol. Syst. Biol.* 3, 143. 590

(25) Becskei, A., and Serrano, L. (2000) Engineering stability in gene 591 networks by autoregulation. *Nature* 405, 590–593. 592

(26) Hooshangi, S., and Weiss, R. (2006) The effect of negative 593 feedback on noise propagation in transcriptional gene networks. *Chaos* 594 16, 026108.

(27) Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., 596 Tsimring, L. S., and Hasty, J. (2008) A fast, robust, and tunable 597 synthetic gene oscillator. *Nature* 456, 516–519. 598

(28) Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, 599 model-guided construction of synthetic gene networks with predicted 600 functions. *Nat. Biotechnol.* 27, 465–471. 601

(29) Lou, C., Stanton, B., Chen, Y.-J., Munsky, B., and Voigt, C. A. 602 (2012) Ribozyme-based insulator parts buffer synthetic circuits from 603 genetic context. *Nat. Biotechnol.* 30, 1137–1142. 604

(30) Strang, G. (2003) Introduction to Linear Algebra. Society for 605 Industrial and Applied Mathematics (SIAM), Bangkok, Thailand. 606 (31) Measuring Molecules of Equivalent Fluorescein (mefl), pe (mepe), 607 and rpe-cy5 (mepcy) Using Sphero Rainbow Calibration Particles; 608

н

611 (32) Das, A. T., Zhou, X., Vink, M., Klaver, B., Verhoef, K., Marzio, 612 G., and Berkhout, B. (2004) Viral evolution as a tool to improve the 613 tetracycline-regulated gene expression system. *J. Biol. Chem.* 279, 614 18776–18782.

615 (33) Markusic, D., Oude-Elferink, R., Das, A. T., Berkhout, B., and 616 Seppen, J. (2005) Comparison of single regulated lentiviral vectors 617 with rtTA expression driven by an autoregulatory loop or a 618 constitutive promoter. *Nucleic Acids Res.* 33, e63–e63.

619 (34) Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., and 620 Benenson, Y. (2011) Synthetic incoherent feedforward circuits show 621 adaptation to the amount of their genetic template. *Mol. Syst. Biol.* 7, 622 519.

623 (35) Lechardeur, D., Verkman, A. S., and Lukacs, G. L. (2005) 624 Intracellular routing of plasmid DNA during non-viral gene transfer. 625 *Adv. Drug Delivery Rev. 57*, 755–767.