Precision design of stable genetic circuits carried in highly-insulated E. coli genomic landing pads

Yongjin Park, Amin Borujeni, Thomas Gorochowski, Jonghyeon Shin, and Christopher Voigt **DOI: 10.15252/msb.20209584**

Corresponding author(s): Christopher Voigt (cavoigt@gmail.com)

Review Timeline:	Submission Date: Editorial Decision: Revision Received: Editorial Decision: Revision Received:	18th Mar 20 20th Apr 20 2nd Jul 20 3rd Jul 20 7th Jul 20
	Accepted:	22nd Jul 20

Editor: Maria Polychronidou

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

1st Editorial Decision

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees who agreed to evaluate your study. Overall, the reviewers are quite supportive. They raise however some concerns, which we would ask you to address in a revision. The reviewers' recommendations are rather clear and I think there is no need to repeat any of the points listed below. Please let me know if you would like to discuss any of the issues raised.

On a more editorial level, we would ask you to address the following:

REFEREE REPORTS

Reviewer #1:

Park et al. present a high-quality synthetic biology manuscript showing an innovative and substantial improvement on the Voigt Lab's Cello gene circuit design system. Anyone familiar with synthetic biology advances will have knowledge of the Cello CAD system, and this manuscript now describes a substantial body of work that allows automated design of gene circuit that are genomically encoded in E.coli (rather than on plasmids). This offers better stability, long-term performance without antibiotic selection, and the ability to construct larger circuits with more orthogonal external inducers.

I am familiar with Cello CAD and the impressive synthetic biology design and experimental advances that were required to achieve the automation of circuit design. I find that this manuscript is similarly impressive in those regards as well as being well written and expertly presented. I struggled to find any criticisms except for typos that will probably be caught in copy-editting stages anyway. So right now I would recommend this publication with no scientific changes required. It is a great paper and one that your audience of synthetic biology readers would appreciate.

Typos:

The appendix sometimes says "Appendix Fig X" and sometimes says "Appendix SFig X" There is inconsistency through the paper and figure legends on whether or not there is a space between a number and its units

The references list has many citation journal names beginning with a '%'

The section 'Sensor array in Landing Pad #3' ends with a reference to Fig 1c but I think this should be Fig 2c.

Reviewer #2:

Summary and general comments

In this manuscript, the authors aimed to design stable genetic circuits in the genome automatically. They constructed three genetic landing pads flanked by double terminators at the E. coli genome with high expression levels and no impact on growth. Circuit design was modularized by putting sensors and genetic circuits in different landing pads. NOT/NOR gate design based on repressors in the genome were optimized and characterized. The Cello software was used for circuit design and prediction. A new user constraint file was constructed for the new genetic landing pads. The genome encoded circuits' cost is lower than the circuits in plasmid, and circuits carried on the genome keep stable for weeks, whereas those on the plasmid break quickly.

Constructing an evolutionary-stable genetic circuit is not a trivial effort, even though many studies have reported genome-encoded genetic circuits. Also, attempt to standardize the design and introduction of genetic circuits onto the genome are worthwhile. This manuscript is important and rigorous. It makes at least two important contributions: 1) it generated landing pads to accommodate the genetic circuits in the genome, and 2) it extended the circuit design automation from the plasmid to the genome.

Major points

1. In this manuscript. the criterion to select landing pad is 1) high expression and 2) no growth defect. In this case, why didn't the authors choose site 13 (position, 3942414), which showed the highest expression level and no growth impact? More importantly, the authors hypothesized that higher expression is better. However, there's no evidence throughout the manuscript to support this notion. Also, the differences between expression levels at different genomic loci are not huge. The manuscript *only* reported 3 landing pads not only limited its application but also lower the level of its generality. Namely, people can put landing pad wherever in the genome; the location has little effects on the circuit performance.?

2. I'm not convinced of the method (OD_600) to measure growth, and the corresponding conclusions of the growth impacts. Dilution once from the overnight seed culture causes a lot of problems on the bacterial physiology. The OD_600 cannot represent the growth in this way, not talking about the "growth rate", which was used in the software.

3. In Figure 5, the evolutionary stability of the genetic circuit was investigated in a changing environment (different signal combination) and it only lasted one-day for each condition. I am very curious about the stability of the circuit in a constant condition for longer time? more days.
4. In the abstract, the authors stated that "these circuits require 8-fold less RNA polymerase when carried on a plasmid". I didn't see the experimental results.

Minor points

1. Procedure for multiple payload insertion into the genome was claimed as "rapid genome engineering methodology", but it is time-consuming to insert the playload one by one. There may/should be more efficient way(s) to do it.

2. "there was a severe growth defect when grown in media lacking thiamine (Figure 1d)."

> "Figure 1d" should be "Figure 1c"

3. "That was found to recover the transcription of the thi operon and overcome the growth defect (v2, Figure 1d)"

>"v2, Figure 1d" should be "v2, Figure 1c".

4. "These sensors produce a 12- to 640-fold induction (Figure 2c, Appendix Figures 8 and 9) with low off states and no evidence of crosstalk (Figure 1c and Appendix Figure 7)."

>"Figure 1c" should be "Figure 2c"

5. Appendix Figure 10b, the first 2 panels on the left column are the same.

Reviewer #3:

In this study, Voigt and colleagues demonstrated a systematic approach to identify and insulate "landing pads" in the E. coli genome for building complex gene circuits. The major considerations include optimal gene expression (using a transponson-guided reporter gene) and minimal growth impact and insulation by using strong transcriptional terminators. They demonstrate construction of a number of gene circuits that perform as designed using the optimal landing pads.

The study builds on the foundation of a number of conceptual and technical advances made in the past several years, some of which having been championed by the Voigt lab. These include large-scale component standardization and quantification, insulation, RNAP flux for component quantification and abstraction, computer aided design and assembly of complex gates. The major distinguishing feature of the current study is the integration of these components (and their quantitative information) for the construction of large-scale circuits, directly in the genome. The work is thoroughly carried out and sets high technical standards for circuit engineering and validation. I have only a few comments that I believe the authors and address in a minor revision.

While I appreciate the technical advance demonstrated in the study, I have reservations about some of the sweeping criticisms the authors made about circuit engineering on plasmids. I am actually skeptical about some of them -- for instance, the growth impact of a circuit in a plasmid is not necessarily greater than that in the genome. It's just a matter of how the circuit is specifically engineered (e.g. plasmid copy number, circuit complexity, and choice of circuit components). Despite the best effort to insulate, an argument can be made that insertion of a large segment of DNA in the chromosome would always create an interruption (to various degrees) to the normal operation of the cell.

My point is not to diminish what the authors have achieved. But I believe they should tone down the argument for advantages of integrating circuits into the chromosomes. What's better really depends on specific engineering objectives, which vary case by case. As noted by the authors, the major objective was "to simplify the process of genome engineering so that it approaches the ease of plasmid manipulations ...". This is an important and valuable objective that the authors have amply demonstrated.

Other points:

1. I do not think the comparison between a circuit in a plasmid carrying a p15 origin (without

selection) and one in the chromosome is fair. As noted by the authors, in the absence of antibiotic selection, many natural plasmids rely on other mechanisms to ensure stable maintenance. One prime example is the F plasmid, a ~100kb plasmid that can be stably maintained in standard lab conditions without positive selection.

Again, I don't think the authors need to make this point to make the work valuable. If the authors do want to make a more general claim, as the writing appears to imply, a more rigorous experimental design is needed (where a circuit is optimized both on a plasmid and in the chromosome).

2. When reporting the growth impact of different landing pads, the authors reported the OD values. It is not clear if what stage of the cultures these values correspond to. These values are useful to report; however, if the authors intend to make strong arguments based on the growth effects (or the lack of), it is more informative to show the growth rates of the cells carrying various landing pads. Two cultures with drastically different growth rates can reach the same OD when fully grown.

3. A major design strategy is the use of strong terminators to insulate. Because of this, it is useful to explain the quantification of these terminators in a greater depth in the results section. For example, Ts was defined in the methods section but it would be helpful to define it in the main text and to elaborate a bit more how it should be interpreted.

Reviewer #1:

1. Typos: The appendix sometimes says "Appendix Fig X" and sometimes says "Appendix SFig X" There is inconsistency through the paper and figure legends on whether or not there is a space between a number and its units. The references list has many citation journal names beginning with a '%' The section 'Sensor array in Landing Pad #3' ends with a reference to Fig 1c but I think this should be Fig 2c.

The typos have been fixed.

Reviewer #2:

1. In this manuscript, the criterion to select landing pad is 1) high expression and 2) no growth defect. In this case, why didn't the authors choose site 13 (position, 3942414), which showed the highest expression level and no growth impact? More importantly, the authors hypothesized that higher expression is better. However, there's no evidence throughout the manuscript to support this notion. Also, the differences between expression levels at different genomic loci are not huge. The manuscript *only* reported 3 landing pads not only limited its application but also lower the level of its generality. Namely, people can put landing pad wherever in the genome; the location has little effects on the circuit performance?

There is a 3rd criterion. The biggest problem with landing pad site selection is the disruption of native gene expression. Although site 13 yields high expression, it disrupts *uhpT* which has a role in glucose-6-phosphate. When this occurred, we would attempt to move the landing pad within a 10kb window. However for site 13, we could not identify a position that did not disrupt endogenous gene expression.

The gene expression levels at different loci (in our hands, 3-fold) are sufficient to alter the predictions of gene circuit design. We do include experiments to this end. Appendix Figure 11 shows how changing the PhIF RBS impacts the response function. Note that the difference in RBS strengths is about 3-fold and this leads to quantitatively different response functions (Appendix Figure 7). A 3-fold change may not make a difference in its function as a "NOT gate," but it changes the response function, which makes an enormous difference in the ability for design automation to connect gates and predict their quantitative response.

2. I'm not convinced of the method (OD_600) to measure growth, and the corresponding conclusions of the growth impacts. Dilution once from the overnight seed culture causes a lot of problems on the bacterial physiology. The OD_600 cannot represent the growth in this way, not talking about the "growth rate", which was used in the software.

We agree and have clarified the Methods. OD_{600} measurements are made with a double-back-dilution protocol. The exception to this is the experiment shown in Figure 1c, we had to use a single back dilution from overnight culture. This is due to the *thiC* KO strain's slow growth in thiamine-deprived media. The *thiC* KO strains grew so slowly that a second dilution was not practical. Note that this observation supports the intent of showing these data to highlight the impact of this mutant on growth.

3. In Figure 5, the evolutionary stability of the genetic circuit was investigated in a changing environment (different signal combination) and it only lasted one-day for each condition. I am very curious about the stability of the circuit in a constant condition for longer time/more days.

We tested the evolutionary stability of genetic circuits shown in Figure 5 maintaining constant inducer concentrations (thus keeping the circuit state) up to 14 days. We have added an Appendix Figure 13 to demonstrate the stability of genome-encoded circuits over the time during this growth experiment.

4. In the abstract, the authors stated that "these circuits require 8-fold less RNA polymerase when carried on a plasmid". I didn't see the experimental results.

We have added Figure 5b to show the reduction in RNAP flux that occurs when the circuit is encoded in the genome. Note that we reduced the number to 4-fold to match the specific circuit in this figure.

5. Procedure for multiple payload insertion into the genome was claimed as "rapid genome engineering methodology", but it is time-consuming to insert the payload one by one. There may/should be more efficient way(s) to do it.

It was our original intent, but attempts to target multiple landing pads have not been successful. We have added a discussion of this in the results.

6. "there was a severe growth defect when grown in media lacking thiamine (Figure 1d)." > "Figure 1d" should be "Figure 1c"

This typo has been corrected.

7. "That was found to recover the transcription of the thi operon and overcome the growth defect (v2, Figure 1d)" >"v2, Figure 1d" should be "v2, Figure 1c".

This typo has been corrected.

8. "These sensors produce a 12- to 640-fold induction (Figure 2c, Appendix Figures 8 and 9) with low off states and no evidence of crosstalk (Figure 1c and Appendix Figure 7)." >"Figure 1c" should be "Figure 2c"

This typo has been corrected.

9. Appendix Figure 10b, the first 2 panels on the left column are the same.

This typo has been corrected.

Reviewer #3:

1. While I appreciate the technical advance demonstrated in the study, I have reservations about some of the sweeping criticisms the authors made about circuit engineering on plasmids. I am actually skeptical about some of them -- for instance, the growth impact of a circuit in a plasmid is not necessarily greater than that in the genome. It's just a matter of how the circuit is specifically engineered (e.g. plasmid copy number, circuit complexity, and choice of circuit components). Despite the best effort to insulate, an argument can be made that insertion of a large segment of DNA in the chromosome would always create an interruption (to various degrees) to the normal operation of the cell. My point is not to diminish what the authors have achieved. But I believe they should tone down the argument for advantages of integrating circuits into the chromosomes. What's better really depends on specific engineering objectives, which vary case by case. As noted by the authors, the major objective was "to simplify the process of genome engineering so that it approaches the ease of plasmid manipulations ...". This is an important and valuable objective that the authors have amply demonstrated.

We have reduced material that could be perceived as overly critical of plasmids and have added some material to the discussion.

2. I do not think the comparison between a circuit in a plasmid carrying a p15 origin (without selection) and one in the chromosome is fair. As noted by the authors, in the absence of antibiotic selection, many natural plasmids rely on other mechanisms to ensure stable maintenance. One prime example is the F plasmid, a ~100kb plasmid that can be stably maintained in standard lab conditions without positive selection. Again, I don't think the authors need to make this point to make the work valuable. If the authors do want to make a more general claim, as the writing appears to imply, a more rigorous experimental design is needed (where a circuit is optimized both on a plasmid and in the chromosome).

We have added a direct comparison of a design where a circuit is optimized for a plasmid and in the chromosome (Figure 5). We have also added a discussion of the situations where plasmids are beneficial and methods to stabilize plasmids.

3. When reporting the growth impact of different landing pads, the authors reported the OD values. It is not clear if what stage of the cultures these values correspond to. These values are useful to report; however, if the authors intend to make strong arguments based on the growth effects (or the lack of), it is more informative to show the growth rates of the cells carrying various landing pads. Two cultures with drastically different growth rates can reach the same OD when fully grown.

The Methods have been clarified. Cells were grown for 5.5 hours in M9 media, corresponding to midexponential phase (OD₆₀₀ of 0.3 to 0.5). The measurements were taken prior to cells saturating.

4. A major design strategy is the use of strong terminators to insulate. Because of this, it is useful to explain the quantification of these terminators in a greater depth in the results section. For example, Ts was defined in the methods section but it would be helpful to define it in the main text and to elaborate a bit more how it should be interpreted.

We have added a detailed description of the definition and the interpretation of terminator strength in the main text. We have also added Appendix Figure 1 to clearly show the assay plasmid design and raw data used to calculate the terminator strength.

Thank you for sending us your revised manuscript. We think that the performed revisions address satisfactorily the issues raised by the reviewers. I am glad to inform you that your manuscript is now suitable for publication.

Before we formally accept the manuscript we would ask you to address a few remaining editorial issues.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ullet

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Christopher A. Voigt Journal Submitted to: Molecular Systems Biology Manuscript Number: MSB-20-9584

orting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner. figure panels include only data points, measurements or observations that can be compared to each other in a scientifically
- meaningful way. graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should
- not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be
- If NCS, the individual data points from each experiment should be protect and any securate rest emproyed around be justified Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name). the assay(s) and method(s) used to carry out the reported observations and measurements an explicit mention of the biological and chemical entity(ies) that are being measured. an explicit mention of the biological and chemical entity(ies) that are being measured.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.). a statement of how many times the experiment shown was independently replicated in the laboratory. definitions of statistical methods and measures:
- common tests, such as t-test (please specify whether paired vs. unpaired), simple χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods
- are tests one-sided or two-sided?
- are there adjustments for multiple comparisons?
 exact statistical test results, e.g., P values = x but not P values < x;
- definition of 'center values' as median or average;
- definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

n the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itse red. If the qu courage you to include a specific subsection in the methods section for statistics, reagents, animal n lels and h

B- Statistics and general methods

Please fill out these boxes Ψ (Do not worry if you cannot see all your text once you press return) 1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? ata from more than 30,000 of E. coli cells were gathered and analyzed to avoid statistical bias 1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used. J/A 2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-۱/A stablished 3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. andomization procedure)? If yes, please describe. For animal studies, include a statement about randomization even if no randomization was used. 4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe 4.b. For animal studies, include a statement about blinding even if no blinding was done 5. For every figure, are statistical tests justified as appropriate? es Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Ve exhibited the distribution of fluorescence measurment obtained from flow cytometry in ppendix figures to present inherent variation in the data.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com http://1degreebio.org

http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-rep

http://grants.nih.gov/grants/olaw/olaw.htm http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm

http://ClinicalTrials.gov http://www.consort-statement.org

http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tume

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/ http://jibiochem.sun.ac.za http://jibiochem.sun.ac.za http://oba.od.nih.gov/biosecurity/biosecurity_documents.html http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	Yes

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog	N/A
number and/or clone number, supplementary information or reference to an antibody validation profile. e.g.,	
Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for	N/A
mycoplasma contamination.	

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

 Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals. 	N/A
 For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments. 	N/A
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	N/A

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	N/A
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	N/A
 For publication of patient photos, include a statement confirming that consent to publish was obtained. 	N/A
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	N/A
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	N/A
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	N/A
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	N/A

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for "Data Deposition".	We provided the "Data Availability" section
Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or figshare (see link list at top right).	
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right). If or JWS Online (see link list at top right), if computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top	N/A
right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines,	
provide a statement only if it could.	