molecular
systems
biology

Non-genetic diversity modulates population performance

Adam Waite, Nicholas Frankel, Yann Dufour, Ms. Jessica Johnston, Junjiajia Long and Thierry Emonet

Corresponding author: Thierry Emonet, Yale University

Editor: Thomas Lemberger

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. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision   31 May 2016

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from two out of the three referees who agreed to evaluate your manuscript. Given that their recommendations are very similar, I prefer to make a decision now rather than delaying the process further. As you will see from the reports below, the referees find the topic of your study of potential interest. They raise, however, several points that should be convincingly addressed in revision. The recommendations provided by the reviewers are very clear in this regard.

REFEREE REPORTS

Reviewer #1:

The manuscript by Waite et al. aims to understand the connection between non-genetic variability and population performance. The specific question is how biological function emerges as a coevolution of the distribution of phenotypes and the single-cell performance function. The authors decided to use E. coli chemotaxis as a model system since its biased random walk from 'runs' and 'tumbles' is well characterized. In particular, the tumble bias (probability to tumble) is used as a read-out of phenotype. To investigate performance the authors combine microfluidics, cell tracking, and simulations of swimming cells to show that rare phenotypes of low tumble bias strongly improve the population performance in terms of chemotactic drift up the gradient. This statement is formalized using Jensen's inequality, which says that the average of a convex function (chemotactic drift as a function of tumble bias) is larger than the function value of the average. By expressing a key adaptation enzyme (CheR) from an inducible plasmid, they show that the changes in tumble bias have the predicted effect on the drift. While is it well known that chemotactic performance or drift depends on tumble bias and level of adaptation enzymes, the manuscript formulates the

mapping between phenotype variation and population performance in an elegant and clear way. Also the variability of the single-cell trajectories has not been quantified in the past. The findings of the paper may play a role in predicting the time of an infection based on chemotactic variability. The paper is overall convincing and well written to interest a broad audience.

Major points:

What if a different readout of phenotype or performance would have been used? Trivially, if the 'coherence' of the swimming across the cell population is used as a performance measure, then a broad or long-tailed distribution of phenotype is overall bad performance. There might also be concave functions where the average of the population performance is below the performance of the average cell, e.g. anything related to 'precision'. Hence, it seems very subjective to choose one example which 'works'.

The trajectories are rather short (1 min). If trajectories are longer, will rotational diffusion not destroy good chemotaxis at low tumble bias? So low tumble bias is not universally good as this depends on length scale of interest.

Chemotactic performance also depends on gradient steepness, speed of adaptation, level of CheB enzyme, and level of CheY response regulator, just to name a few. However, the roles of these parameters are not discussed. Please explain chosen selection of what to consider and what not.

Are measured trajectories in 2D or 3D? Same question applies to simulations. Is tumble bias measured in a gradient, and if so, tumble bias should change along gradient. How is this taken care off?

An earlier model of the group, published in Dufour et al. (2014), does not account for the bias in angles after a tumble which may increase the drift at high tumble bias. Is this included in the current model?

Minor points:

In Fig. 1 B-D it is unclear what the 'distance past gate' means in terms of the schematic of the flow chamber in panel A. Please indicate 'distance' and 'gate' in this schematic.

Reviewer #2:

In this paper, the authors use the chemotactic behavior of E. coli as a model to investigate the effect of a broad distribution of phenotypes on the efficiency in performing a specific task by a population of clonal individuals. They use a cleverly designed microfluidic device and automated data collection and analysis to record the motion of thousands of E. coli cells in a channel with a pre-established gradient of chemoattractant. This allows them to correlate a specific phenotype (tumble bias) with the performance (distance travelled up the gradient), using both wild-type cells and a strain that expresses the methyltransferase CheR under the control of inducible promoter. They showed that both diffusive spreading in absence of a gradient and chemotactic drift were higher for low but non-zero tumble bias. They also performed simulations of the chemotactic behavior, in excellent agreement with the experimental results. They then used all their data to construct a phenotype to performance map, which was non-linear and convex, and concluded that highly performing outliers that correspond to very low tumble bias are important in bringing the average performance of the population above the performance of the average cell. Notably, this effect could be mathematically formalized using Jensen's inequality and it is likely to play a role in many biological systems.
This manuscript is insightful and technically solid and the use of E. coli chemotaxis as a model system to address effects of population heterogeneity and cell individuality has an excellent tradition. I am confident that this work will be of interest to a general readership of MSB. However, I do have several critical comments that need to be addressed before the manuscript is accepted for publication.

Major point:

Although it is clear that cell performance in a gradient depends on tumble bias in a non-trivial way, this dependence might be more complex than currently suggested by data presentation and at least partly indirect. Can the authors clearly distinguish direct dependence of performance on tumble bias from dependence on other bias-related behavioral parameters? In this context, titration of the adaptation enzyme CheR may not be an ideal way to regulate tumble bias, since levels of CheR similarly affect kinetics of adaptation, another key determinant of the chemotactic performance. Authors' own previous theoretical work clearly established the importance of the adaptation kinetics (i.e, short-term memory) in chemotactic behavior (Emonet & Cluzel, 2008; Frankel et al., 2014). If I am not mistaken in my interpretation, computer simulations shown in Frankel et al. (2014) suggest that changes in adaptation kinetics due to variation in the levels of adaptation enzymes have even stronger effects on performance than changes in tumble bias. Moreover, at low levels of CheR tumble bias might not only be low but also highly variable in time, which might itself increase efficiency of chemotactic movement (Emonet & Cluzel, 2008).

Although the main message of the story will remain valid even if the correlation between performance and tumble bias is indirect, I think that current focus on tumble bias might be potentially misleading. However, this issue could be relatively easily resolved by titrating the response regulator CheY, because levels of CheY should tune mean tumble bias without affecting the adaptation kinetics or slow methylation-dependent bias fluctuations. I do not ask for redoing all experiments at varying levels of CheY, but I believe that an experiment that confirms that the same (or similar) performance-phenotype relation is observed when tumble bias is controlled by CheY levels would be important for data interpretation.

Minor points:

- Figure 3D: Although for each induction level of CheR taken separately the performance is ordered according to the tumble bias, it appears that the performance is significantly different for a fixed range of tumble bias depending on the induction level. Could it be related to effects of CheR on parameters other than tumble bias (see comment above)?

- For both WT and 10uM IPTG induced CheR - and also for DcheY strain as a control - it would be interesting to show the distributions of distances past the gate for each subpopulation with a given tumble bias, at least at a given time point (e.g. 20 min). Are they self-similar, or is the population spreading inherently different for low tumble biases? Only average travelled distances are shown so far.

- Materials and Methods, subsection on data acquisition (Line 378 and following): As far as I understand, the whole channel seems to be scanned sequentially using 1 min long movies, the whole scan being roughly 10 minutes long. Is it actually the case? This should be described more clearly (e.g. how many areas, are they overlapping). Also, the pixel size of the camera is missing.

- Materials and Methods, line 385: Please give the characteristics of the filters of the YFP block

- Line 419: typing mistake "track length is is inversely proportional... "

1st Revision - authors' response   15 November 2016

Continued on next page.

We would like to thank the reviewers for their thoughtful and constructive comments. We have addressed each comment/concern in a point-by-point manner below. Text in the manuscript that has been altered or added to addresses these issues has been colored blue.

The reviewer is right to point out that we focused on one metric to evaluate cell performance. We want to clarify that we did not choose this metric because it works. We were interested in chemotactic ability, and therefore we thought the most natural metric of performance would be how quickly each phenotype is able to climb up a gradient of attractant. A main point of the paper is to demonstrate how diversity can influence population function, not to show that diversity is necessarily beneficial. We fully agree with the reviewer that the paper would benefit from a broader discussion of how the phenotype-to-performance map of different tasks (such as those that capture the concept of "precision") might display different types of nonlinearities (concave and convex), and how when the map is concave, the average performance of the population should be lower than that of the average phenotype. We have added this text to the Discussion section.

The characteristic time scale of rotational diffusion for a ~1 µm bacterium is about 8 sec (Dufour et al., 2014). Recently, we experimentally confirmed that this value is at most 10 seconds (Dufour et al., 2016). So, an upper bound for track duration of 1 min is long enough to capture the effect of rotational diffusion. To check the effect of our minimum trajectory length of 6 seconds, we increased the minimum trajectory duration to 20 seconds. This did not destroy the performance of low tumble bias cells but instead slightly increased it (Appendix Fig. S4) for reasons that we discuss in the *Analysis of trajectories* section of the *Materials and Methods*. Therefore, the observed high performance of low tumble bias cells is robust to these changes in quantification parameters. We also show experimentally that, when tumble bias goes to zero, performance is degraded in (Fig. 3F) as predicted theoretically (Dufour et al., 2014), highlighting that the reviewer's hypothesized performance decrease is reached in our device, but only at zero tumble bias.

Chemotactic performance also depends on gradient steepness, speed of adaptation, level of CheB enzyme, and level of CheY response regulator, just to name a few. However, the roles of these parameters are not discussed. Please explain chosen selection of what to consider and what not.

The reviewer rightfully points out that we did not sufficiently discuss our rationale for our choice of parameters. We have added text in various parts of the Introduction and Results sections to address this concern. In addition, we have run an entirely new set of experiments where we varied tumble bias by manipulating the expression levels of CheY and CheZ instead of CheR and got similar results (Fig. EV5), showing that, at least in our experimental conditions, performance is dominated by tumble bias and not adaptation time (see also answer to Reviewer #2).

As for investigating the effect of gradient steepness, our device was designed to generate linear gradients and the dimensions of the device reflect compromises between various experimental constraints. For example, a longer device would produce a shallower gradient and provide more time to observe cells, but the gradient would take much longer to equilibrate when setting up the experiment. A shorter device would provide a steeper gradient and shorter setup times but would reduce the amount of time we would be able to observe the cells swimming before they reached the top of the gradient.

Are measured trajectories in 2D or 3D? Same question applies to simulations.

In the experiment, cells swim in 3D (the depth of the observation chamber is 10 µm, which is ~10 times the size of a cell) but are tracked in 2D. We treat the simulation exactly like the experiment, so there the cells also swim in 3D constrained by boundaries in the same dimensions as the chamber but are tracked in 2D.

Is tumble bias measured in a gradient, and if so, tumble bias should change along gradient. How is this taken care off?

The reviewer is correct; the observed tumble bias of a cell will drop relative to its unstimulated tumble bias when that cell senses a gradient of attractant. The magnitude of the drop is a function of gradient steepness (Dufour et al., 2014). Since *E. coli* sense the logarithm of the gradient, the linear gradient in our device is perceived as being steepest at the start and shallowest at the end.

We cannot directly measure the unstimulated versus observed tumble bias in the same cells in our experiment setup. So, we performed a control simulation using an experimentally-derived gradient where all the cells were initialized to be identical and had a known unstimulated tumble bias of 0.23.

We then plotted the average difference between the observed and unstimulated tumble bias as a function of position in the chamber.

As expected, the difference is greatest at the beginning of the gradient where the perceived gradient is steepest (Appendix Fig. S3A). However, this rapidly falls to a small, constant difference (-0.004) within the first millimeter of the observation chamber. The simulation also suggests that the difference between stimulated and unstimulated tumble bias cannot account for the observed differences in performance (Appendix Fig. S3B).

We have added this point to the main text of the manuscript.

Initially, we did not include persistence in our model. We measured the average angular persistence of cells in experiments without a gradient or a gate and found that it was very small (mean change in angle was 81º). We then verified in simulations that including such small persistence did not change the results. The difference between our result and the commonly cited value of 62° (Brown and Berg, 1972. *Nature.*) could be due to differences in experimental conditions. Brown and Berg used a different strain, different growth media, different tracking media, tracked their cells in three dimensions, and their cells were not constrained to a 10 μm depth as was the case in our observation chamber.

We have added this information to the "Simulations" section of the Materials and Methods.

Minor points:

In Fig. 1 B-D it is unclear what the 'distance past gate' means in terms of the schematic of the flow chamber in panel A. Please indicate 'distance' and 'gate' in this schematic.

We have updated Fig. 2A to specify the location of the gate.

Reviewer #2:

In this paper, the authors use the chemotactic behavior of E. coli as a model to investigate the effect of a broad distribution of phenotypes on the efficiency in performing a specific task by a population of clonal individuals. They use a cleverly designed microfluidic device and automated data collection and analysis to record the motion of thousands of E. coli cells in a channel with a pre-established gradient of chemoattractant. This allows them to correlate a specific phenotype (tumble bias) with the performance (distance travelled up the gradient), using both wild-type cells and a strain that expresses the methyltransferase CheR under the control of inducible promoter. They showed that both diffusive spreading in absence of a gradient and chemotactic drift were higher for low but non-zero tumble bias. They also performed simulations of the chemotactic behavior, in excellent agreement with the experimental results. They then used all their data to construct a phenotype to performance map, which was non-linear and convex, and concluded that highly performing outliers that correspond to very low tumble bias are important in bringing the average performance of the population above the performance of the average cell. Notably, this effect could be mathematically formalized using Jensen's inequality and it is likely to play a role in many biological systems.

This manuscript is insightful and technically solid and the use of E. coli chemotaxis as a model system

We agree with the reviewer's comment and to address it we have done an entirely new set of experiments where, instead of modifying CheR levels, we manipulated CheY/CheZ levels. See also answer to the reviewer's comment about CheY below.

It is correct that at low levels of CheR expression fluctuations in the methylation-demethylation kinetics become larger. For that reason our simulations do include noise in the methylation-demethylation reactions.

Regarding the effect of changing adaptation time versus that of changing tumble bias, the situation is complicated by the fact that in wild-type cells adaptation time and tumble bias are inversely correlated, as shown experimentally in single cells (Park et al., Nature 2010). This correlation is also maintained in the model used for our simulations (Frankel et al. 2014). However, we did examine how these two effects affect performance independently of each other in another recent theoretical study (Dufour et al., 2014). We found that, in shallow gradients, adaptation time only starts to matter for very low tumble bias cells (see in particular Figure 2A from Dufour et al., 2014) . Thus, for most of the range of tumble bias explored in our experiments, the dominant factor that effects performance is predicted to be tumble bias.

We agree that our focus on tumble bias as the only important chemotactic phenotype was not the best way to present our results, and that both tumble bias and adaptation time could contribute to

performance. In addition to making this more clear in the Introduction and Results sections (please see the response to reviewer 1, above), we followed the reviewer's suggestion and performed additional experiments in a strain where we could control the amount of CheY and CheZ (see Fig. EV5). This allowed us to experimentally test low tumble bias cells in a manner that did not also affect adaptation time in the way that changing CheR does. When CheY was expressed at low levels relative to CheZ, this strain also showed increased performance with decreasing tumble bias similar to low CheR expression. We could also see a nonlinear performance increase at low tumble bias at early time points in this strain. Thus, tumble bias does appear to be the dominant phenotypic parameter in determining performance in shallow gradients.

Minor points:
- Figure 3D: Although for each induction level of CheR taken separately the performance is ordered according to the tumble bias, it appears that the performance is significantly different for a fixed range of tumble bias depending on the induction level. Could it be related to effects of CheR on parameters other than tumble bias (see comment above)?

We believe the reviewer is referring to the fact that, in Figure 3D, the positions of the high and low CheR populations in tumble bias bin 0.05 – 0.15 (green) do not have overlapping error bars. We do not know the mechanistic basis for this discrepancy, but the difference is also visible in the spatial distributions of the populations (Fig. EV2). It is possible that this is related to effects of CheR on parameters others than tumble bias.

- For both WT and 10uM IPTG induced CheR - and also for DcheY strain as a control - it would be interesting to show the distributions of distances past the gate for each subpopulation with a given tumble bias, at least at a given time point (e.g. 20 min). Are they self-similar, or is the population spreading inherently different for low tumble biases? Only average travelled distances are shown so far.

We followed the reviewer's suggestion and added a figure showing the spatial distribution of each tumble bias bin for each pass through the observation for wild-type cells, the high- and low-CheR experiments, as well as $\Delta cheY$ (we did not split this population into tumble bias bins, as this strain cannot tumble). This has been added as Fig. EV2 and is mentioned in the main text.

- Materials and Methods, subsection on data acquisition (Line 378 and following): As far as I understand, the whole channel seems to be scanned sequentially using 1 min long movies, the whole scan being roughly 10 minutes long. Is it actually the case? This should be described more clearly (e.g. how many areas, are they overlapping). Also, the pixel size of the camera is missing.

We apologize for not being clear. We have added a more detailed explanation of the acquisition procedure to the Materials and Methods section.

- Materials and Methods, line 385: Please give the characteristics of the filters of the YFP block

We have added this information to the Materials and Methods section.

- Line 419: typing mistake "track length is is inversely proportional... "

Thank you. We have fixed this mistake.

Thank you again for sending us your revised manuscript. We are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

USEFUL LINKS FOR COMPLETING THIS FORM

Corresponding Author Name: Thierry Emonet
Journal Submitted to: Molecular Systems Biology
Manuscript Number: MSB-16-7044

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

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-tum

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://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

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 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 altered/varied/perturbed in a controlled manner.

➔ 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 χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
 • 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.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	We performed at least two experiments, each providing information about 5,000 – 10,000 cells.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	For each experiment, we had many diagnostic critera to determine whether the experiment worked. We checked that 1) the overall tumble bias did not change much over time, 2) the gradient shape was consistent, 3) cell counts and speed were consistent, 4) there was no evidence of warping of the PDMS chamber due to evaporation.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
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.	No
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	NA
Is there an estimate of variation within each group of data?	NA
Is the variance similar between the groups that are being statistically compared?	NA

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog 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).	NA
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

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

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
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.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
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.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
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.	NA
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.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. 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	NA
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).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	NA
22. 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, CelIML) 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) 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.	Link to new source code is in manuscript.

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top 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.	No