

High-resolution definition of humoral immune response correlates of effective immunity against HIV

Galit Alter, Karen G. Dowell, Eric P. Brown, Todd J. Suscovich, Anastassia Mikhailova, Alison E. Mahan, Bruce D. Walker, Falk Nimmerjahn, Chris Bailey-Kellogg and Margaret E. Ackerman

Review timeline:

| | |
|---------------------|------------------|
| Submission date: | 18 July 2017 |
| Editorial Decision: | 13 October 2017 |
| Revision received: | 9 January 2018 |
| Editorial Decision: | 12 February 2018 |
| Revision received: | 26 February 2018 |
| Accepted: | 27 February 2018 |

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

1st Editorial Decision

13 October 2017

Thank you again for submitting your work to Molecular Systems Biology. As you will see below, both reviewers appreciate that the presented findings seem interesting. They raise however a series of concerns, which we would ask you to address in a revision.

The reviewers' recommendations are rather clear so I think that there is no need to repeat the points listed below. Please do not hesitate to contact me in case you would like to discuss/clarify any of the points listed by the referees.

REVIEWER REPORTS

Reviewer #3:

Alter et al. applied their previously-described "systems serology" antibody biophysical property profiling approach to a large cohort of HIV-infected individuals. The resulting high-dimensional data was used to build models that differentiate between the humoral immune responses of elite controllers, viremic controllers, and treated and untreated disease progressors. The analysis yielded a number of intriguing differences between subject groups provide rich datasets for further analyses. The data/models did an impressive job in discriminating between study groups, as well as in predicting antibody effector functions.

A more thorough explanation of analyses and figures are needed in many areas. Many of the figures were not explained in adequate detail for scientists without systems biology/statistics expertise. For

example, Figure 2B-D were not detailed outside of the figure legend (what are confusion matrices or LOD scores?).

The manuscript would also be improved by more clearly highlighting the predictive variables in the models, providing further biological interpretation, and bridging these with some of the main biological findings.

The authors noted that aviremic individuals had an enrichment of env-specific antibodies with bisected glycans compared to viremic individuals, while this was not the case with the total antibody response. The authors tout this as an example of the differences that are revealed by biophysical modeling of the study groups. However, this "example" is not readily apparent in any of the model datasets (or at least not pointed out in sufficient detail). I assume that the bottom panel of Figure S2 is in fact explaining this viremic/aviremic model (please also be consistent with group names), but S2 is never mentioned in the manuscript text. Further, no glycan variables appear to contribute to the model (Figure S2C?). In Figure 3, they present this observation using P-values - which elicits concerns that they have not adequately performing a multiple-hypothesis testing correction given the large number of variables that they are testing. Further, the authors suggest relevance of these env-specific bisected-glycan antibodies by citing other studies that have shown such antibodies exhibit enhanced FCgammaR3 recognition and ADCC activity. Thus, are the env-specific bisected-glycan antibodies correlated with ADCC activity and FCgammaR3 ligation propensity data collected in this study? Are these measures different between the viremic and aviremic groups?

Similarly, another interesting result is that ECs have elevated levels of p24 antibodies, which is readily apparent in the volcano plot. However, is this true in the classification model? I cannot tell, as the legend to Figure S2C (controller/progressor) is not legible.

These two biological observations, while interesting, do not appear to be "Examples of differentiating humoral responses features identified by classification models" as the Figure 3 title states. Rather they are examples of observations possible with the high-dimensional data, which confirm or support previous observations / hypotheses. That is fine and still important - however, the authors should reframe these findings accordingly, or better highlight how the models point to these observations.

A number of figure axes labels are illegible (Figure S2C, S3D). The figure 6 legend should note that the plots are shown with labeled variables in Figure S3.

As a more general point, the manuscript should do a better job of emphasizing that the tested variables are CORRELATED with viral control / titer, but are not necessarily causal. For instance, it is possible that the immune signatures reflect reduced viremia rather than cause it.

Finally (and crucially), the authors do not provide adequate data and analysis code. This paper is ultimately a computational analysis of large datasets, but neither the datasets or the computer code are available at least as far as I can tell from the manuscript. All of the underlying datasets and computer code should be made available as supplementary material in a usable format: not PDF images, but CSV / text files and computer code files. If this cannot be accommodated in the journal supplementary format rules, then they should be posted on DataDryad or GitHub or some similar repository designed for hosting code and data.

Reviewer #4:

This is an interesting paper that presents new high-dimensional humoral immunoprofiling approaches and that applies those to elucidate hallmarks of effective HIV-1 viral control. The authors use a combination of new experimental and computational approaches to integrate humoral data on functional, effector function, complement fixation and phagocytosis. Overall, I find the methods and results interesting. This being said, I do have a few questions/concerns that I present below:

Major comments:

Figure 4: I wonder how robust the networks/findings are in this figure? What happens if you were to modify the $|r| > 0.4$ and $p < 0.05$ thresholds? Would the findings be consistent? It would be good to perform a quick sensitivity analysis. Perhaps, use different values of r to see how it changes. Also, how many of the p -values would survive multiple testing. I understand that this is mostly visual and exploratory, but it would be good to understand how strong these interactions are. Perhaps this could be discussed to highlight the fact that many of those are likely false positives.

Redundancy across assays: As shown on Figure 5, it is clear that some of the assays/variables are highly correlated. I was wondering if the prediction accuracy presented in Figure 2 could be improved if one were to downselect the variables to reduce correlation. While the Elastic Net can deal with collinearity, its performance can be greatly affected by it. Also, scientifically speaking it would be great to come up with a minimal set of variables/assays that capture the breadth of antibody function while minimizing redundancy across variables. I understand that this is difficult to do based on a single dataset, but the authors might have some other data that could be used for that. I would at least discuss it.

Figure 5: As discussed above, this is an interesting figure/result. I would propose to use the mean squared error (MSE) for the y -axis as it is more standard for evaluating prediction accuracy. Related, for 5b, report r^2 instead of r .

Use of biological replicates in Figure 5. I don't really understand why the correlation between biological replicates would be an upper bound. Are these technical or biological replicates? If biological, I would argue that the correlation should probably be lower as there might be substantial variability across subjects. I think this needs to be clarified.

Minor comments:

Figure 4: The legend in this figure should say $|r| > 0.4$ and not $r > |.4|$.

Code and data availability: Given the importance of the computational analyses performed, it would be great to share the data and code for full reproducibility.

We thank the reviewers for their thoughtful comments, questions, and suggestions. Responses to each point made in review are described below, and modifications made the manuscript have been tracked in the resubmission. We believe that the manuscript has been substantially improved based on addressing these comments, and we further note that the revised supplemental materials now includes raw data, code, and model outputs.

Reviewer #3:

Alter et al. applied their previously-described "systems serology" antibody biophysical property profiling approach to a large cohort of HIV-infected individuals. The resulting high-dimensional data was used to build models that differentiate between the humoral immune responses of elite controllers, viremic controllers, and treated and untreated disease progressors. The analysis yielded a number of intriguing differences between subject groups provide rich datasets for further analyses. The data/models did an impressive job in discriminating between study groups, as well as in predicting antibody effector functions.

A more thorough explanation of analyses and figures are needed in many areas. Many of the figures were not explained in adequate detail for scientists without systems biology/statistics expertise. For example, Figure 2B-D were not detailed outside of the figure legend (what are confusion matrices or LOD scores?).

In the revised manuscript we have taken care to better introduce terms associated with our analysis approach that may be unfamiliar to readers, as well as to discuss the figures in greater depth.

See for example, page 4-5: "Given the rich data collected, a machine learning approach was employed to identify combinations of a humoral response features able to predict subject class and antibody effector function. Cross-validated classifiers trained to distinguish subject groups identified minimal sets of antibody features that accurately and robustly discriminated among all four subject groups, between viremic and aviremic subjects, and between controllers and progressors (**Fig.2**). Subject class was determined by a class score (LOD, or Log_2 Odds Ratio) that defined the relative likelihood of a given subject's assignment to one as compared to other classes. Differentiation across EC, VC, TP and UP groups was accomplished with approximately 60% accuracy, as compared to the approximately 25% accuracy expected at random based on class size for this 4-group differentiation model, and as observed when scrambled study data was used as a modeling input (**Fig.2a**). Approximately 75% accuracy was observed for the 2-way classifications aimed at differentiation of subjects by progression (EC and VC versus TP and UP) or by viremia (EC and TP versus VC and UP), as compared to the approximately 50% accuracy expected by chance, or again, observed when study data was permuted prior to learning (**Fig.2a**). Confusion matrices (**Fig.2b-d**), which compare actual and predicted classifications, indicated that with the exception of UPs, who could not be confidently differentiated from TPs or VCs, most classes were predicted well. While there was little evidence of systematic confusion, among misclassified VCs and TPs, most were modeled to be UPs, consistent with the difficulty noted in classification of this group, and suggestive that they may have a less distinct humoral profile as a group. Further, class score can be considered a measure of model confidence (**Fig.2b-d**). Whereas in the 2-way classifications misclassified subjects often had marginal class scores, in the 4-group differentiation models, incorrect predictions were often made confidently. For example, misclassified ECs tended to be confidently predicted as either TPs or VCs, suggestive of the existence of some subjects with profiles that are considerably more consistent with these classes."

The manuscript would also be improved by more clearly highlighting the predictive variables in the models, providing further biological interpretation, and bridging these with some of the main biological findings.

Based on this thoughtful suggestion we have expanded the manuscript text to more fully describe and interpret the features identified as predictive of group and functional activity.

See for example, page 6: "For example, numerous bisected glycoforms contributed to the differentiation of aviremic from viremic subjects when the glycosylation profiles of HIV envelope-specific antibodies were used to predict viral load status (**Appendix Fig S3**), suggesting a difference in the level of

antibody bisection between groups.... Whether this glycosylation state difference associated with viral load may be a cause or effect of viremic status remains to be determined, but potential mechanistic relevance to antiviral humoral immunity is suggested by previous studies demonstrating that antibodies with bisected glycans exhibit enhanced FcγR3 recognition and ADCC activity (31), a mechanistic link that was further explored within this data set in models of antibody function. ”

Page 6: ”Analysis of the Fab-specificity of features contributing to classification of ECs but not observed in modeling viremia status (**Appendix Figures S1, S2, and Dataset EV2**), pointed toward recognition of the virus capsid (p24/Gag) as a differential marker between subjects with immune-mediated as opposed to pharmacologic viral suppression.”

The authors noted that aviremic individuals had an enrichment of env-specific antibodies with bisected glycans compared to viremic individuals, while this was not the case with the total antibody response. The authors tout this as an example of the differences that are revealed by biophysical modeling of the study groups. However, this "example" is not readily apparent in any of the model datasets (or at least not pointed out in sufficient detail).

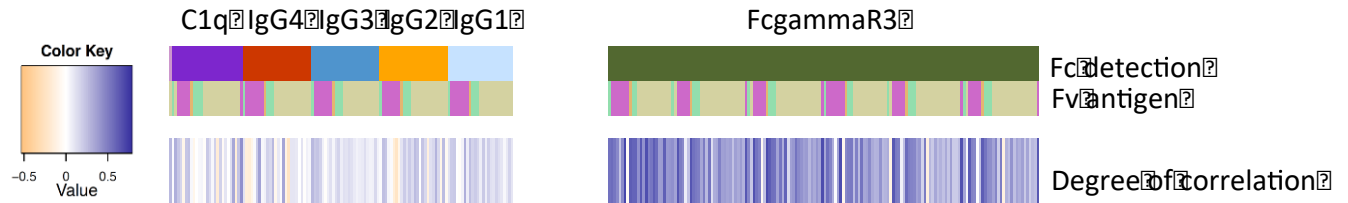
We apologize for the lack of clarity regarding how the individual examples selected for inclusion in Figure 3 relate to functional and class models. The manuscript text and figure set has been significantly expanded to specifically highlight the features that are uniquely selected by the model and make the linkage more clear, as indicated by the example above, and in greater detail below.

I assume that the bottom panel of Figure S2 is in fact explaining this viremic/aviremic model (please also be consistent with group names), but S2 is never mentioned in the manuscript text. Further, no glycan variables appear to contribute to the model (Figure S2C?). In Figure 3, they present this observation using P-values - which elicits concerns that they have not adequately performing a multiple-hypothesis testing correction given the large number of variables that they are testing. Further, the authors suggest relevance of these env-specific bisected-glycan antibodies by citing other studies that have shown such antibodies exhibit enhanced FCγR3 recognition and ADCC activity. Thus, are the env-specific bisected-glycan antibodies correlated with ADCC activity and FCγR3 ligation propensity data collected in this study? Are these measures different between the viremic and aviremic groups?

A number of issues are raised here, which have been addressed as follows:

- In the revised manuscript we have made more explicit reference to Appendix Figure S2 (which relates to the classification results presented in Figure 2b). This figure had been poorly referenced as “Figs.S1,2” in the original submission.
- We have ensured that groups are referred to consistently throughout the manuscript, figures, and Appendix (ie: replacing “nonviremic” with “aviremic”).
- We have provided detail on the basis for inclusion of the specific examples selected for Figure 3 and provided the “missing link” the reviewer is looking for in terms of relating these individual measures to the models as follows:
 - For Figure 3a: new text in the results and a new supplemental figure (Appendix Figure S3) describing the use of glycan data to predict aviremic vs. viremic classes that raises the hypothesis that glycan bisection may differ among groups is included to provide the basis for investigating this parameter in the context of these groups.
 - For Figure 3b: an expanded description of the linkage between the model of subject class whose in response to the more direct reviewer comment on this matter below.
- Given the basis for the plots in Figure 3 is now (better) provided, we have not addressed the multiple hypothesis testing concern raised as we hope it is now clear that these features were not selected on the basis of their p-values, but were investigated based on modeling results.
- Beyond the literature supporting the biological relevance of a difference in antibody bisection, the importance of this difference to the sample set evaluated within this study is now supported by drawing attention to:
 - Figure 6d, which reports the contribution of IgG bisection to the model of ADCC activity (see also Appendix Figure S6), and the correlation of bisected glycoforms with ADCC

- activity (see also Appendix Figure S7).
- Figure 6d, which also illustrates the predominance of Fcγ3R ligation propensity assessments to contribute to ADCC activity models.
- The correlation of Fcγ3R ligation with ADCC, as compared to the correlation between IgG subclass levels or C1q ligation and ADCC is illustrated for the purpose of review, below in **Response to Review Figure 1**.
- Appendix Figure S2, in which the HIV gp120-specific positive coefficient with greatest weight in viremic/aviremic classification is an Fcγ3R measurement.



Response to Review Figure 1: Correlation heatmap depicting relative relationships between ADCC activity and Fc Array measurements among ECs.

Collectively, these relationships support both the relevance of bisection to ADCC activity, and viremic/aviremic group differentiation, within this study, and provide the basis for the inclusion of this feature in Figure 3, and the justification for the description in the figure legend as an example of “differentiating humoral response features identified by classification”.

Similarly, another interesting result is that ECs have elevated levels of p24 antibodies, which is readily apparent in the volcano plot. However, is this true in the classification model? I cannot tell, as the legend to Figure S2C (controller/progressor) is not legible.

The elevation of p24-specific antibodies among ECs evident in the volcano plot is indeed reflected in the models of subject class. In the 4 way classification models, of the 16 features with positive coefficient weights for the EC group, 9 represent p24/Gag-specific antibody measurements. In fact, 2 of the top 3 positive coefficients are p24/Gag-specific.

We apologize for the lack of label clarity. Though the supplemental pdfs supplied can generally be clearly read under magnification, we recognize that the small print required to fit the labels is a limitation to the utility of the supplemental graphs when printed. Given the difficulty in reading labels on supplemental figures at print size, and the potential interest in and utility of the data underlying these graphs, we have now included tables reporting the features and coefficient weights for models in the supplemental data.

Relevant to the question raised by the reviewer, we have used these tables to generate a list (**Response to Review Table 1**) of the feature coefficients for the EC group in order of decreasing magnitude, with p24/Gag-specific antibody features indicated in yellow.

| | EC |
|---------------------------|-------|
| IgG1.IIIb.pr55.Gag | 0.379 |
| FcγRIIIa.Chiang.Mai.gp120 | 0.379 |
| C1q.p24.HXBc2 | 0.319 |
| FcγRIIb.HIV1.Vif | 0.306 |
| IgG3.HIV1.Integrase | 0.256 |
| FcγRIIIa.R131.p24.HXBc2 | 0.241 |
| FcγRIIIb.SH.p24.HXBc2 | 0.170 |
| SNA.HIV1.Nef | 0.167 |
| FcγRIIIb.SH.IIIb.pr55.Gag | 0.130 |

| | |
|-----------------------------|--------|
| FcgRIIIb.IIIb.pr55.Gag | 0.116 |
| FcgRIIIb.SH.HIV1.p7 | 0.113 |
| IgG4.HIV1.Rev | 0.086 |
| IgG3.93TH975.gp120 | 0.078 |
| C1q.p24.IIIb | 0.068 |
| LCA.p24.HXBc2 | 0.040 |
| FcgRIIIa.F158.IIIb.pr55.Gag | 0.025 |
| IgG1.gp120.SF162 | -0.001 |
| FcgRIIIa.V158.gp140.UG21 | -0.015 |
| VVL.gp140.CN54 | -0.027 |
| C1q.IIIb.pr55.Gag | -0.041 |
| VVL.HIV1.Nef | -0.053 |
| PNA.gp140.CN54 | -0.076 |
| MBL.HIV1.Vif | -0.111 |
| SNA.gp41.HXBc2 | -0.115 |
| C1q.gp120.RSC3 | -0.139 |
| IgG1.gp140.CN54 | -0.175 |
| IgG3.HIV1.Nef | -0.188 |
| MBL.93TH975.gp120 | -0.188 |
| C1q.gp120.IIIb | -0.194 |
| LCA.gp41.HXBc2 | -0.204 |
| VVL.HIV1.Rev | -0.257 |
| FcgRIIIa.gp120.SF162 | -0.310 |
| VVL.p24.HXBc2 | -0.336 |
| C1q.gp120.ZM109F | -0.338 |
| FcgRIIIa.SOSIP | -0.930 |

Response to Review Table 1: Feature coefficients for classification of the EC group in 4-way classification models. Antibodies specific for p24/Gag are indicated with yellow shading. Line indicates the boundary between features with positive and negative coefficients. (Related to Figures 2 and S1.)

These two biological observations, while interesting, do not appear to be "Examples of differentiating humoral responses features identified by classification models" as the Figure 3 title states. Rather they are examples of observations possible with the high-dimensional data, which confirm or support previous observations / hypotheses. That is fine and still important - however, the authors should reframe these findings accordingly, or better highlight how the models point to these observations.

We hope that the expanded description of how these selected differences relate to classification models, as described above, provides suitable justification for the Figure 3 title.

A number of figure axes labels are illegible (Figure S2C, S3D). The figure 6 legend should note that the plots are shown with labeled variables in Figure S3.

We apologize for the label sizes on supplemental figures at print size. To address this and related comments, and given potential interest in and utility of the data underlying these graphs, we have now included tables reporting the feature labels and coefficient weights for all models in the supplemental data.

We have also now added a note to the Figure 6 legend that Appendix Figure S6 (as renumbered from S3 in revision), as well as the model output tables, report the full text labels for the features used in these models.

As a more general point, the manuscript should do a better job of emphasizing that the tested variables are CORRELATED with viral control / titer, but are not necessarily causal. For instance, it is possible that the immune signatures reflect reduced viremia rather than cause it.

In the revised manuscript, we have emphasized that our findings are associative and that causality cannot be determined from this study.

See for example:

Page 4: "...suggesting the value of comprehensive IgG Fab, Fc, and functional profiling to inform the identification of humoral mechanisms of action and **associative** relationships to viral suppression."

Page 4: "By modeling their interactions to define the underlying principles by which antibodies collaborate and/or compete to direct the overall antiviral activity of the humoral response, we define new biomarkers **associated with** antiviral antibody effector function, aviremia, and non-progression."

Page 6: "Whether this glycosylation state difference **associated with viral load** may be a cause or effect of viremic status remains to be determined, ..."

Page 10: "Thus, **while observations using this approach are necessarily only associative**, an integrated Fc-functional and deep biophysical profiling offers a unique opportunity for the identification and development of strategies that may enhance monoclonal therapeutics broadly."

Page 10: "Moreover, **though this study was designed only to assess associations**, ..."

Finally (and crucially), the authors do not provide adequate data and analysis code. This paper is ultimately a computational analysis of large datasets, but neither the datasets or the computer code are available at least as far as I can tell from the manuscript. All of the underlying datasets and computer code should be made available as supplementary material in a usable format: not PDF images, but CSV / text files and computer code files. If this cannot be accommodated in the journal supplementary format rules, then they should be posted on DataDryad or GitHub or some similar repository designed for hosting code and data.

As requested, the revised manuscript includes supplemental files that include 1) the raw experimental data (dataset EV1), 2) model outputs (dataset EV2), and 3) the code used to generate the results reported (modeling scripts EV3).

Reviewer #4:

This is an interesting paper that presents new high-dimensional humoral immunoprofiling approaches and that applies those to elucidate hallmarks of effective HIV-1 viral control. The authors use a combination of new experimental and computational approaches to integrate humoral data on functional, effector function, complement fixation and phagocytosis. Overall, I find the methods and results interesting. This being said, I do have a few questions/concerns that I present below:

Major comments:

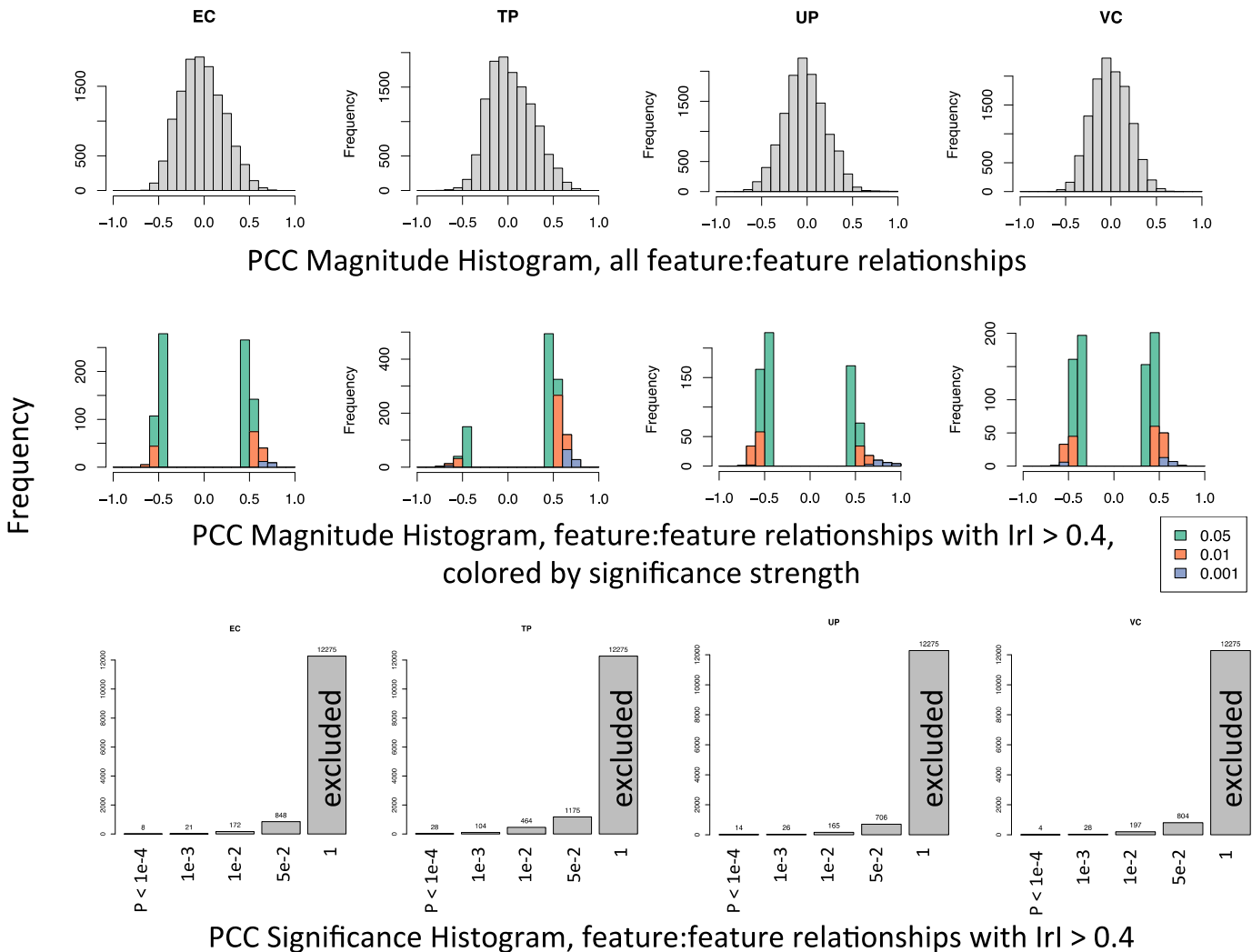
Figure 4: I wonder how robust the networks/findings are in this figure? What happens if you were to modify the $|r| > .4$ and $p < 0.05$ thresholds? Would the findings be consistent? It would be good to perform a quick sensitivity analysis. Perhaps, use different values of r to see how it changes. Also, how many of the p -values would survive multiple testing. I understand that this is mostly visual and exploratory, but it would be good to understand how strong these interactions are. Perhaps this could be discussed to highlight the fact that many of those are likely false positives.

We agree that such networks can be a useful visualization but can represent only a snapshot of the relationships between parameters. We provide below some sensitivity analysis below, have added

more significance information to the figure, and have amended the text to better call attention to the limitations inherent to use of this type of analysis that the reviewer has noted as follows:

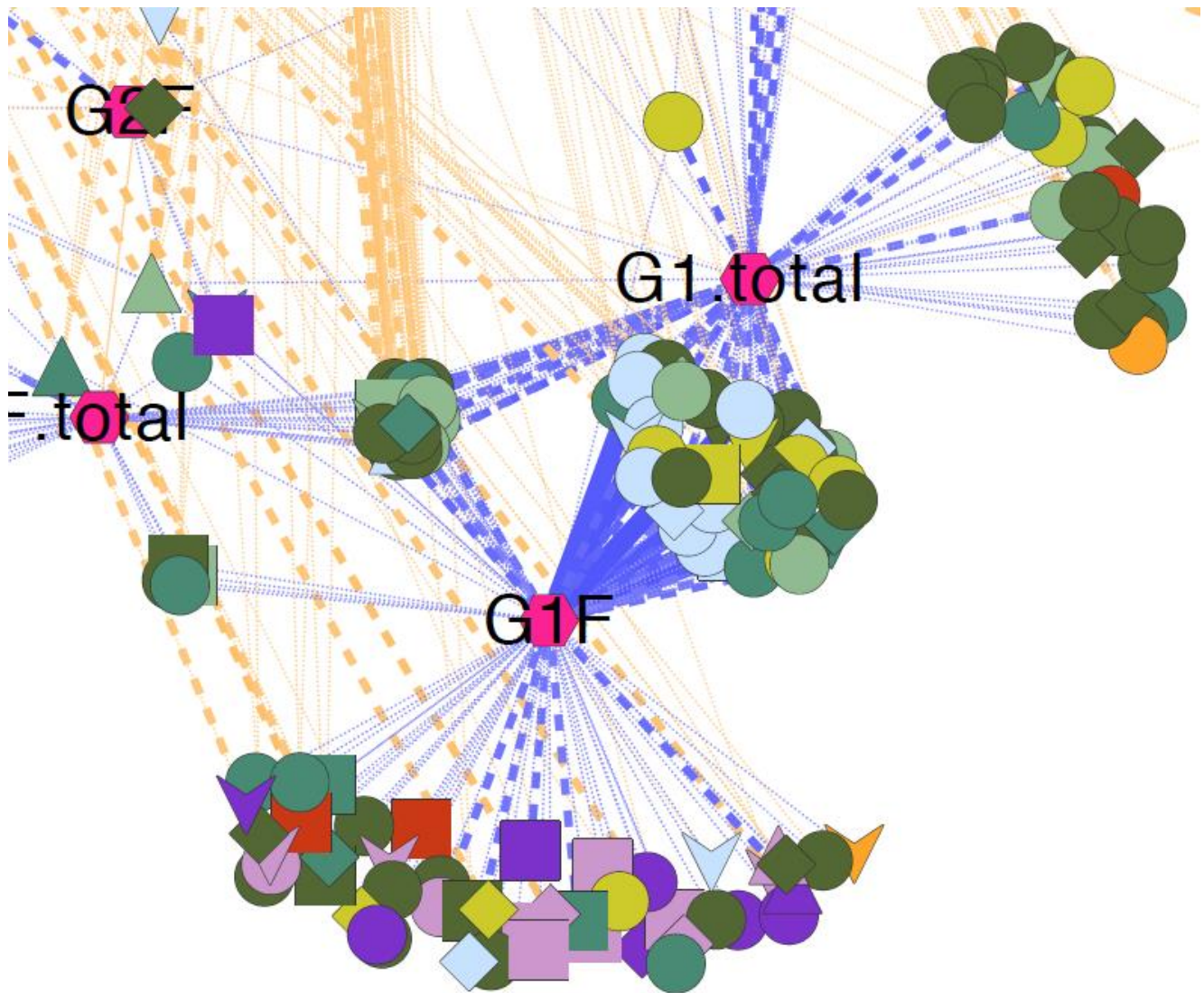
See page 7: “Balancing the competing desires of discovery and confidence, plots depicting relationships between glycoforms and Fc Array measurements with correlation coefficients exceeding an absolute value magnitude of 0.4 and an uncorrected p-value of 0.01, representing the top 7% of correlations in strength and significance, were generated.”

Additionally, we note the correlation strength is indicated by color scale in the figure. While values with $|r| < 0.4$ are excluded from the visualization, we consider this to be a reasonable threshold based on balancing discovery and confidence. This choice results in presentation of feature:feature relationships in the top ~7% across the data set (**Response to Review Figure 2**).



Response to Review Figure 2: PCC magnitude and significance histograms for all feature:feature relationships, as compared to those selected for inclusion in Figure 4’s network plots.

Additionally, we have revised Figure 4 such that line style now indicates the significance of the correlation, with thin dotted lines indicating $p < 0.05$, thick dashed lines indicating $p < 0.01$, and thick solid lines indicating $p < 0.001$ (**Response to Review Figure 3**).

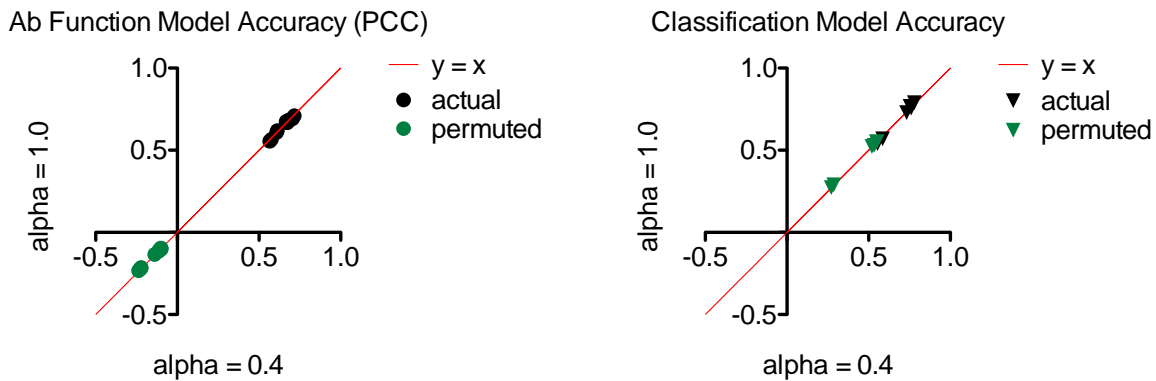


Response to Review Figure 3: Zoomed in view of network plots showing depiction of correlation strength and significance.

Redundancy across assays: As shown on Figure 5, it is clear that some of the assays/variables are highly correlated. I was wondering if the prediction accuracy presented in Figure 2 could be improved if one were to downselect the variables to reduce correlation. While the Elastic Net can deal with collinearity, its performance can be greatly affected by it. Also, scientifically speaking it would be great to come up with a minimal set of variables/assays that capture the breadth of antibody function while minimizing redundancy across variables. I understand that this is difficult to do based on a single dataset, but the authors might have some other data that could be used for that. I would at least discuss it.

We agree that many of the antibody characteristic variables are highly correlated (as is expected given that most antibody types can interact well with most Fc receptors), and that correlated variables can adversely impact model performance. In a previous study we reported an exploration of several different dimensionality reduction approaches, and several different modeling approaches, and found that neither had a strong impact on models of antibody activity performance-wise (Choi et al, PLoS Comp Bio 2015 - <http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004185>).

Here, to investigate the sensitivity of the Elastic Net approach to collinearity among features, we evaluated alpha setting of 1.0 and 0.4, ranging from lasso-like to ridge-like, respectively, and flanking the setting of alpha = 0.8 selected for models presented in the manuscript. We find that model performance was not highly sensitive to this parameter (**Response to Review Figure 4**).



Response to Review Figure 4: Scatterplot of model accuracy outputs for antibody functions (left) and classification models (right, includes 2-way and 4-way classifications). Performance with actual data is shown in black, performance with permuted data is shown in green.

Beyond these technical considerations, as the reviewer suggests, development of a minimal (and non-redundant) feature set that could encapsulate the breadth of antibody function is of high interest given the multiple antibody functions that have correlated with vaccine-mediated protection in non-human primate models across multiple vaccine approaches and regimens. As the reviewer suggests, we agree that analysis of multiple data sets, particularly those from vaccine studies (as opposed to the natural infection cohort evaluated here) would be important in order to have high confidence in the value of a reduced feature set for this purpose. Our study relates only these HIV infected subject populations, including rare controllers, highlighting key features of antibodies that may enhance their activity; we now mention that future studies aimed at further exploring the unique antibody features/profiles in in other cohorts would be of high interest.

See page 9: “However, while novel platform technologies, such as the recently described Systems Serology approach (Ackerman et al, 2017) can provide a comprehensive glimpse at the broad view of the landscape of functional responses of pathogen-specific antibodies, approaches to specifically define minimal physical biomarkers, **which can be tailored to maximize information and reduce experimental redundancy**, which are more easily interrogated in validated assays, and that are associated with desirable clinical outcomes, **need to be further implemented to better define cross-cohort, cross-regimen, and cross-pathogen principles of protective humoral immunity**.”

Figure 5: As discussed above, this is an interesting figure/result. I would propose to use the mean squared error (MSE) for the y-axis as it is more standard for evaluating prediction accuracy. Related, for 5b, report r^2 instead of r .

We agree that MSE is a standard metric for prediction accuracy. However, we sought to explicitly compare the modeling results with experimental replicates, for which correlation coefficients are more typically reported. We also desired to enable comparisons between functional assays, whereas interpretation of MSEs across functions is complicated by the different activity ranges and magnitudes for each assay. Given this intention, we have retained the original y-axis label in Figure 5, and note that MSE values (\pm SD over cross-validation replicates and folds) are provided for each model in Figures S1, S2, S3, and S4. Because these figures include MSE over varying values of lambda, they also provide useful information regarding interpretation of the MSE magnitudes.

As requested, r^2 values are now reported (Figures 5b, S3, and S4).

Use of biological replicates in Figure 5. I don't really understand why the correlation between biological replicates would be an upper bound. Are these technical or biological replicates? If biological, I would argue that the correlation should probably be lower as there might be substantial variability across subjects. I think this needs to be clarified.

We apologize for the confusion regarding our description. Consistent with the figure labeling, correlations between biological replicates are shown. Given the confusion this comparison and its description raised, we have tempered our description/statement as follows:

See page 7: "Prediction accuracy, the correlation between modeled and observed activity, was generally as good as the degree of correlation observed between assay replicates, which **provides a reasonable benchmark for how well a model might be expected to perform.**"

Minor comments:

Figure 4: The legend in this figure should say $|r|>0.4$ and not $r>|.4|$.

We apologize for this embarrassing error. It has been corrected.

Code and data availability: Given the importance of the computational analyses performed, it would be great to share the data and code for full reproducibility.

As requested, the revised manuscript includes supplemental files that include 1) the raw experimental data (dataset EV1), 2) model outputs (dataset EV2), and 3) the code used to generate the results reported (modeling scripts EV3).

Thank you for sending us your revised manuscript. We have now heard back from the two reviewers who were asked to evaluate your study. As you will see below, they are both satisfied with the modifications made and think that the study is now suitable for publication.

Before we formally accept your manuscript for publication, we would like to ask you to address the editorial issues below.

REVIEWER REPORTS

Reviewer #4:

I am happy with all changes to the paper. As far as I am concerned, it can be published as is.

Reviewer #3:

The authors meticulously addressed all of our concerns, and did a commendable job providing the raw data, analysis code, and output figures.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Ackerman, Margaret E

Journal Submitted to: Molecular Systems Biology

Manuscript Number: MSB-17-7881

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 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 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)

| | |
|---|-------------|
| 1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size? | p11-12 |
| 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? | p11-12 |
| 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. | p11-12 |
| 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. | p4, p11-12 |
| 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? | p15, p24-25 |
| Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. | p15 |
| Is there an estimate of variation within each group of data? | p15 |
| Is the variance similar between the groups that are being statistically compared? | p15 |

C- Reagents**USEFUL LINKS FOR COMPLETING THIS FORM**<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-tur><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://ijb.biochem.sun.ac.za>http://oba.od.nih.gov/biosecurity/biosecurity_documents.html<http://www.selectagents.gov/>

| | |
|--|--------|
| 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). | p12-14 |
| 7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination. | p12-13 |

* 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. | p11-12 |
| 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. | p11-12 |
| 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)). | p15, primary data is included as Dataset EV1 |
| 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 <i>Shewanella oneidensis</i> 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 | p15, primary data is included as Dataset EV1 |
| 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, 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) 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. | p17, model code is included as Computer Code EV1 |

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. | NA |
|---|----|