

iPhemap: An atlas of Phenotype to genotype relationships of human iPSC models of neurological diseases

Ethan W. Hollingsworth, Jacob E. Vaughn, Josh C. Orack, Chelsea Skinner, Jamil Khouri, Sofia B. Lizarraga, Mark E. Hester, Fumihiko Watanabe, Kenneth S. Kosik and Jaime Imitola

Corresponding author: Jaime Imitola, The Ohio State University

Review timeline:

Submission date:	21 June 2017
Editorial Decision:	22 August 2017
Revision received:	08 September 2017
Accepted:	12 September 2017

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.)

Editor: Céline Carret

1st Editorial Decision

22 August 2017

Thank you for the submission of your manuscript to EMBO Molecular Medicine. At long last, I'm happy to say that we have received the three reports from the two referees we asked to review your metanalysis.

You will see from the comments below that both referees are supportive of publication and only comment on a few items that would make the paper more understandable and comprehensive. I will not detail their comments, as they are explicit enough, but would encourage you to address all concerns.

We would welcome the submission of a revised version within three months for further consideration. Please note that EMBO Molecular Medicine strongly supports a single round of revision and that, as acceptance or rejection of the manuscript will depend on another round of review, your responses should be as complete as possible.

EMBO Molecular Medicine has a "scooping protection" policy, whereby similar findings that are published by others during review or revision are not a criterion for rejection. Should you decide to submit a revised version, I do ask that you get in touch after three months if you have not completed it, to update us on the status.

Please also contact us as soon as possible if similar work is published elsewhere. If other work is published we may not be able to extend the revision period beyond three months.

Please read below for important editorial formatting and consult our author's guidelines for proper

formatting of your revised article for EMBO Molecular Medicine.

I look forward to receiving your revised manuscript.

***** Reviewer's comments *****

Referee #1 (Comments on Novelty/Model System):

The type of iPSC-based disease modeling studies with low n and with poor characterization of cell types used may not provide robust data. Fewer but high quality and sufficiently powered studies could be superior and reduce noise and reporting bias

Referee #1 (Remarks):

The study by Hollingsworth et al addresses an important problem in iPSC based disease modeling in the CNS which is the difficulty of reporting, categorizing and comparing disease phenotypes across various studies, diseases and cell types based on the data published in the current literature. By selecting 93 studies (using QC criteria) the authors identify more than 600 phenotypes that are further classified into a taxonomy of 9 phenotypic clusters. The study also defines a minimal set of reporting criteria in iPSC-based disease modeling studies that is a valuable effort that could be helpful for the field. The overall goal of the study is to define novel association between disease, gene, phenotype, cell type and brain region. Such an effort could potentially reduce redundant work by establishing a database of all the disease studies and phenotypes reported for researchers to further mine. Finally, the authors establish a user-friendly dynamic online database (iPhenmap) that includes additional details that can be accessed by the reader. The online database should be updated with future studies in a timely manner and thereby could become more and more useful if properly maintained and curated.

The major challenge of the work, a challenge that the authors are well aware of, is that the power for most of the studies included is very poor and variable (remarkably low average number of 2-3 patients per study nearly all without isogenic controls!). Furthermore, there is likely a major reporting bias that could skew the data (publishing "positive" phenotypes that are expected based on other work such as for example known results from mouse studies). Despite these limitations, this is a laudable first, tour-de-force effort to address some of those challenges in the iPSC field. The current metastudy also highlights many of the current limitations and shortcomings of iPSC-based disease modeling which could provide an impetus for improving quality and statistical power in future studies.

I did like the included movie that makes it somewhat simpler to follow the work as what is presented in text and figures. On the other hand, I found the paper itself quite difficult to read as it provides very extensive descriptive data across many disorders and studies but often without a definite message of what those data mean. The text jumps forth and back from describing individual, rather random, phenotypes to trying to make broader points on iPSC phenotyping. It could benefit from a clearer outline that guides the reader in a more structured manner through the various points that authors try to make. This is also the case for some of the Figures. For several Figures, the presentation is rather complex. For example, in Figure 4A I am still unclear what the authors try to show and what the main main conclusions are (e.g. with regard to "epoch"). Similarly, Figure 3A is also complex and difficult to read and does not easily allow the reader to come to up with meaningful conclusions.

The authors further present interesting data related to cell type analyzed (for example oligodendrocytes appear to have more phenotypes in average than other neural cell types). However, there is a concern that the definition of an oligodendrocyte (as well as the definition of neurons or astrocytes) is far from standard across the studies. It may be important to discuss the need for minimal criteria to define a given cell type and maturity. For example for oligodendrocytes, the data are likely "contaminated" with by more immature NPC-type cells as true human oligodendrocytes would likely require differentiation periods of > 100 days using current differentiation technologies.

The Graphical abstract is too dense as presented with many panels and text that is too small to read.

The illustration should focus on the conceptual points to better convey the overall effort rather than trying to include too much detail within the graphical abstract.

The authors should more clearly discuss the problem that not all studies look at comparable phenotypes and therefore many associations between two genes may be due to the fact that studies on those genes looked at those phenotypes but not at other phenotypes could have similarly linked those genes. For example in PD related genes or in AD related genes, investigators will automatically steer towards certain phenotypes based on previous work which would likely skew those data across the broader gene and disease classes..

In conclusion, the study is a first effort to catalogue phenotypes from iPSC studies in a systematic manner. While I am not convinced that the current metastudy allows for any major novel conclusions at this point, I think that this is a valuable and bold effort towards this goal, and that the conclusions may sharpen in the future if the dataset continues to increase. The work clearly points to the need for studying and reporting disease phenotypes in a more standardized manner in order to fulfill the great potential of iPSC-based disease modeling.

Referee #2 (Remarks):

Hollingsworth et al conduct a meta-analysis of studies that have assessed the pathophenotypic expression in iPSCs and various neural subtypes differentiated from patients iPSCs. This study is quite comprehensive and very timely after nearly a decade of disease modeling studies using iPSCs for various neurological disorders. The authors have performed a very thorough analysis, which has culminated into iPhemap, a web resource that will be continuously updated, and that will undoubtedly be useful to gain insights into the genotype-phenotype relationship without the noise provided by the different experimental settings and designs of each individual studies. Taken together, this tool will help to further our knowledge and understanding of how, with what and for what iPSCs should be used to study disease associated phenotypes related to neurological disorders.

Minor comments:

Studies using iPSC lines derived from sporadic cases were included in the study, however there is very little mention of how the sporadic cases are being treated in the analyses performed. Wherever possible, it would be better if the distinction between mendelian forms of the disease and sporadic forms could be done in the analyses.

Related to this, in Supporting information Table 10, the disease state of the patients from which the iPSC lines are derived for all sporadic cases should be mentioned.

It would be interesting - provided that the number of studies reporting inter-clonal variations within the same lines is sufficient - to assess the difference in phenotype expression between lines derived from different clones of the same patient line.

The definition of categorical cluster descriptions could be made clearer. For instance, the difference between "impairment of expected cellular functions" and "decreased cellular processes and products" is not obvious. Furthermore, the "Absence of expected normal phenotype" cluster in supporting information Table 2, which is described as "any phenotype that can be described by the complete absence of any phenotype that is expected and found in a healthy version of the same cell" would be better described as complete loss of function.

In figure 3, the color-coding is extremely confusing between the phenotypes and the cell types. Moreover, the colors of the different phenotypes does not seem to be completely identical to that of figure 2 which adds up to the confusion.

Some cell types such as iPSCs, NSCs and astrocytes lack the presence of some of the phenotype classes (Supplemental figure 3). It should be made clearer whether these phenotype classes have been investigated but no difference was observed between the disease line(s) and controls or phenotypes belonging to these particular classes were simply not assessed in those studies.

Figure 4A: While I do like the way the results are presented in this figure, I am missing how this

illustrates the relationship between phenotypes and developmental epochs. Is the data expressed as a cell maturation timeline on the x axis i.e. are the heatmap boxes at the far left representing phenotypes observed in more mature neurons than those further right? What is the organization of the heatmap on the x axis aside from by cell types?

1st Revision - authors' response

08 September 2017

In the next several paragraphs, we will provide a point-by point answer to the reviewers' comments:

Referee #1:

We appreciate the reviewer's positive response about our manuscript, "Despite these limitations, this is a laudable first, tour-de-force effort to address some of those challenges in the iPSC field..." We have addressed the reviewer's comments as follows:

***Question 1.1** On the other hand, I found the paper itself quite difficult to read as it provides very extensive descriptive data across many disorders and studies but often without a definite message of what those data mean. The text jumps forth and back from describing individual, rather random, phenotypes to trying to make broader points on iPSC phenotyping. It could benefit from a clearer outline that guides the reader in a more structured manner through the various points that authors try to make.*

Response:

We appreciate this important reviewer comment. We have significantly restructured our manuscript to reduce the inclusion of specific observations and add more discussion about the meaning of our findings.

***Question 1.2** This is also the case for some of the Figures. For several Figures, the presentation is rather complex. For example, in Figure 4A I am still unclear what the authors try to show and what the main conclusions are (e.g. with regard to "epoch").*

Response:

We thank the reviewer for this comment and have updated **Figure 4** for increased clarity of our findings and moved the heatmap back to **Supplemental Figures**. We have also updated the figure legends and changed epoch to stage.

“Figure 4. Quantification of phenotypes by genes and developmental stage. A) Schematic diagram depicting developmental timeline of iPSC-derived cells included in analysis. B-F) Percent distribution plots of iPSC, NSC, astrocyte, oligodendrocyte, and neuronal phenotypes reported for genes linked to neurodegenerative, neurodevelopmental or other (psychiatric and viral-induced) disorders. Each data point represents a specific disease. One-way analysis of variance (ANOVA) with Bonferroni multiple comparisons tests were performed. G) Distribution of phenotypes by pluripotent, progenitor, and postmitotic cell type for Alzheimer's (AD), Parkinson's (PD), Huntington's Disease (HD), and Rett Syndrome. Two-way ANOVA with Tukey's multiple comparisons test was performed. Data are expressed as mean percentage \pm s.e.m., * $P < 0.05$, ** $P < 0.01$...”

***Question 1.3** Similarly, Figure 3A is also complex and difficult to read and does not easily allow the reader to come to up with meaningful conclusions.*

Response:

We are grateful for this reviewer comment. We agree and have added a simpler distribution plot as **Figure 3A**. We have updated the colors to be more consistent with Figure 2 to reduce any ambiguity, deleted the small numbers, highlighted the most significant associations in the Circos plot and added more description to the figure legend to guide the reader about the significance of the findings.

“Figure 3. Phenotypic classes by patient-derived cell type from 663 annotated phenotypes and phenotype: paper metric. A) Distribution of the phenotype classes within each CNS cell type with total number of phenotypes listed above each respective column. B) Circos plot of phenotype classes by cell type and vice versa are depicted by connecting ribbons, with the

width of each band proportional to the percent composition and the top-most ribbons highlighted. The neuronal ribbons (blue) were found to connect to and be largest for almost every phenotypic class. The outer track indicates the numeric percentage of phenotypic classes comprising each cell type. C) Metric of total phenotypes per cell type with respect to the total number of studies that investigated that particular cell type.”

Question 1.4 *The authors further present interesting data related to cell type analyzed (for example oligodendrocytes appear to have more phenotypes in average than other neural cell types). However, there is a concern that the definition of an oligodendrocyte (as well as the definition of neurons or astrocytes) is far from standard across the studies. It may be important to discuss the need for minimal criteria to define a given cell type and maturity. For example for oligodendrocytes, the data are likely "contaminated" with by more immature NPC-type cells as true human oligodendrocytes would likely require differentiation periods of > 100 days using current differentiation technologies.*

Response:

We acknowledge this important reviewer point. We have added discussion of the need for criterion to define a specific cell type in the MiPSCE section (**Page 19, Line 16**).

“The future inclusion of such data from differentiated cultures may help address the need for a standard set of criteria to define a given cell type, perhaps with thresholds of purity defined by marker expression and physiological measures, which would increase the reliability of comparing reported phenotypes, for it is unclear whether these differences can affect the phenotypes and gene expression of the derived cells (Figure 1).”

Question 1.5 *The Graphical abstract is too dense as presented with many panels and text that is too small to read. The illustration should focus on the conceptual points to better convey the overall effort rather than trying to include too much detail within the graphical abstract.*

Response:

We thank the reviewer for this comment and have simplified our graphical abstract to minimize distraction and guide the reader to the salient points of the work (**Page 2**).

Question 1.6 *The authors should more clearly discuss the problem that not all studies look at comparable phenotypes and therefore many associations between two genes may be due to the fact that studies on those genes looked at those phenotypes but not at other phenotypes could have similarly linked those genes. For example in PD related genes or in AD related genes, investigators will automatically steer towards certain phenotypes based on previous work which would likely skew those data across the broader gene and disease classes.*

Response:

We agree with this reviewer comment and have added discussion of this limitation (**Page 20, Line 14**).

“This potential bias may have influenced our current set of overlapping phenotypes as investigators may have been more inclined to test for phenotypes based on prior work, thereby diminishing the potential for phenotypes to link genes from different diseases. Therefore, expanding which phenotypes are tested for, outside the scope of past studies, will further enrich phenogenetic analyses.”

Referee #2:

We thank the reviewer for their positive comments and valuable suggestions to enhance the clarity and strength of our manuscript. We have addressed their comments as follows:

Question 2.1 *Studies using iPSC lines derived from sporadic cases were included in the study, however there is very little mention of how the sporadic cases are being treated in the analyses performed. Wherever possible, it would be better if the distinction between mendelian forms of the disease and sporadic forms could be done in the analyses.*

Response:

We thank the reviewer for this important point. We have clarified that only one sporadic cell line

was included in our study and the rest were disorders driven by somatic mutations. (Page 12, Line 10).

“In addition, we detected one AD-linked gene, APP, to be most concordant with an AD cell line derived from a sporadic-diseased patient with no known mutation, or “Sporadic” in Supplemental Figure 6A, the only sporadic line included in our analysis.”

Question 2.2 Related to this, in Supporting Information Table 10, the disease state of the patients from which the iPSC lines are derived for all sporadic cases should be mentioned.

Response:

We thank the reviewer for this comment and have specified the disease in **SI Table 10** as “Sporadic (AD).”

Question 2.3 It would be interesting - provided that the number of studies reporting inter-clonal variations within the same lines is sufficient - to assess the difference in phenotype expression between lines derived from different clones of the same patient line.

Response:

We acknowledge this insightful reviewer comment and agree that this analysis would be valuable, but failed to find enough papers including this information in our analysis to make it possible.

Question 2.4 The definition of categorical cluster descriptions could be made clearer. For instance, the difference between "impairment of expected cellular functions" and "decreased cellular processes and products" is not obvious. Furthermore, the "Absence of expected normal phenotype" cluster in supporting information Table 2, which is described as "any phenotype that can be described by the complete absence of any phenotype that is expected and found in a healthy version of the same cell" would be better described as complete loss of function.

Response:

We appreciate this reviewer comment and have edited the descriptions of **SI Table 2** accordingly.
Impairment of expected cellular functions: “This category contains any phenotype that can be described by the presence of a disrupted/changed state of a structure or process that is expected and found in a healthy version of the same cell and cannot be described in terms of increases or decreases. i.e. Impaired structure of adherens junctions (PMID: 24996170).”
Absence of expected normal phenotype: “This category contains any phenotype that can be described by the complete loss of a function that is found in a healthy version of the same cell. i.e. Absence of random X-inactivation (PMID: 21372419).”

Question 2.5 In figure 3, the color-coding is extremely confusing between the phenotypes and the cell types. Moreover, the colors of the different phenotypes does not seem to be completely identical to that of figure 2 which adds up to the confusion.

Response:

We thank the reviewer for this valuable comment. We agree that the colors of phenotype categories and cell types did not match precisely and have corrected this inconsistency in **Figures 2 and 3**.

Question 2.6 Some cell types such as iPSCs, NSCs and astrocytes lack the presence of some of the phenotype classes (Supplemental figure 3). It should be made clearer whether these phenotype classes have been investigated but no difference was observed between the disease line(s) and controls or phenotypes belonging to these particular classes were simply not assessed in those studies.

Response:

We acknowledge this reviewer comment and have clarified this limitation (Page 8, Line 9).
“However, given the small number of studies that modeled patient-derived glial cells, the absence of certain phenotypic clusters may be an artifact of field biases, opposed to the true state of these diseased cells.”

Question 2.7 Figure 4A: While I do like the way the results are presented in this figure, I am missing how this illustrates the relationship between phenotypes and developmental epochs. Is the

data expressed as a cell maturation timeline on the x axis i.e. are the heatmap boxes at the far left representing phenotypes observed in more mature neurons than those further right? What is the organization of the heatmap on the x axis aside from by cell types?

Response:

We thank the reviewer for pointing out this ambiguity and have updated **Figure 4A** for clarity, moved the heatmap back to **Supplemental Figures**, and replaced epoch with stage. We also have updated both of their legends to guide the reviewer through their findings.

“Figure 4. Quantification of phenotypes by genes and developmental stage. A) Schematic diagram depicting developmental timeline of iPSC-derived cells included in analysis. B-F) Percent distribution plots of iPSC, NSC, astrocyte, oligodendrocyte, and neuronal phenotypes reported for genes linked to neurodegenerative, neurodevelopmental or other (psychiatric and viral-induced) disorders. Each data point represents a specific disease. One-way analysis of variance (ANOVA) with Bonferroni multiple comparisons tests were performed. G) Distribution of phenotypes by pluripotent, progenitor, and postmitotic cell type for Alzheimer’s (AD), Parkinson’s (PD), Huntington’s Disease (HD), and Rett Syndrome. Two-way ANOVA with Tukey’s multiple comparisons test was performed. Data are expressed as mean percentage \pm s.e.m., * P <0.05, ** P <0.01...”

We would like to thank the reviewers and editors. Your recommendations certainly have made our contribution more robust to the field of iPSCs as an “eye-opening” work.

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Jaime Imitola

Journal Submitted to: EMBO Molecular Medicine

Manuscript Number: EMM-2017-08191

Reporting 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 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.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

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?	NA
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?	NA
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.	NA
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.	Yes, network graphs meet the power-law distribution assumption by performing regression analyses on logarithmic transformed data, please see Supplemental Text (Page 8).
Is there an estimate of variation within each group of data?	Yes, where applicable (Figure 4) using one-way and two-way ANOVAs, please see Supplemental Text (Page 7).

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://jji.biochem.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?	Depends on the groups analyzed, variance is statistically difference between disease types for neurons and NSCs, as well as between cell types in Alzheimer's and Parkinson's Disease (Figure 4).
---	---

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 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'. 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)).	Dataset provided in Tables and online database.
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. 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 Biomedels (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.	NA

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 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
---	----