EMBO Molecular Medicine

Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients

Florent Mouliere, Richard Mair, Dineika Chandrananda, Francesco Marass, Christopher G. Smith, Jing Su, James Morris, Colin Watts, Kevin Brindle, Nitzan Rosenfeld

Review timeline:

Submission date:	11 May 2018
Editorial Decision:	25 June 2018
Revision received:	3 September 2018
Editorial Decision:	25 September 2018
Revision received:	11 October 2018
Accepted:	15 October 2018

Editor: Lise Roth

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 25 June 2018

Thank you for the submission of your manuscript to EMBO Molecular Medicine. We have now heard back from the three referees whom we asked to evaluate your manuscript.

As you will see from the reports below, the three referees are supportive of your study, and only ask for minor revisions. Addressing the reviewers' concerns in full will be necessary for further considering the manuscript in our journal. EMBO Molecular Medicine encourages a single round of revision only and therefore, acceptance or rejection of the manuscript will depend on the completeness of your responses included in the next, final version of the manuscript.

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 for Author):

Mouliere and Mair et al. demonstrate in this manuscript that shallow whole-genome sequencing (sWGS) can be used with cell-free tumor DNA from cerebrospinal fluid (CSF) to detect somatic

copy number alterations (SCNAs) without prior knowledge of tumor mutations. In addition, a fragmentation pattern was identified that correlates with detection of SCNAs. This paper provides an improved low-cost screening method for patient samples. The manuscript is largely technically sound, but a few minor comments should be addressed as detailed below.

1. In the figure legend for Fig. 1C, the authors reference sWGS from plasma and urine samples, but this data is not included in the figure. Please include this sWGS data for comparison.

2. The authors demonstrate a relatively clear negative correlation between the amplitude of 10 bp periodic oscillations and the levels of SCNAs (Fig. 2C). The authors should comment on potential explanations for why this trend is observed.

3. The data provided in Supplemental Fig. 2 are very informative and important for interpretation of the paper; thus, this figure should be included in the main text.

4. In Supplemental Fig. 2B, tumor size is missing units.

Referee #2 (Comments on Novelty/Model System for Author):

Mouliere et al conducted a study regarding cell-free DNA cerebrospinal fluid. The manuscript deals with a highly interesting topic since the analysis of cell-free tumor DNA is currently one the hottest topics in the field of oncology and data on brain tumors are still sparse. However, proximal sampling by the analysis of CSF might increase sensitivity.
Overall speaking, the study technically sounds. The manuscript does contain some interesting data and they are on most parts well-presented, however there are some concerns that need to be addressed.
1) The title ". . . using short cfDNA" is somehow misleading, since short cfDNA is not specifically enriched with the presented methods.
2) The authors present data from quite a small cohort of 13 patients and in only 5 of them SCNA could be detected. The author should indicate further analysis approaches for those patients where no cftDNA was detected.
3) In general, the results& discussion section seems a bit minimalistic and could be presented in a more comprehensive way.

4) Moreover, there are some discrepancies regarding the analysis of tumor samples: Line 102: Here the author state that tumor material was availa-ble only from one patient, although two sentences before that the claim that the highest concentration of cfDNA corresponded to three patients with SCNAs in the tumor. Pretty much at the beginning it says that the levels of cftDNA in CSF were not directly correlated to the tumor volume, lateron the detection and confirmation of the SCNA in the CSF sample was influenced by the size of the tumor (in addition to the level of cfDNA, and the glioma grade).

5) The authors observed a shift toward shorter fragment sizes of cfDNA in CSF, which was previously reported to be associated with an enrichment of tumor-derived fragments. Moreover, a negative correlation between the amplitude of the 10 bp periodic oscillations and the levels of SCNAs in these CSF samples was reported. This is somehow contradictory and the author should furthers explain this observation or at least present a hypothesis for it.

6) Figure 2 shows different size distributions of plasma, urine, and CSF. The author should include these data in the results and discussion section.

7) Figure 2 would benefit from comparison of size profiles for a sample with and without SCNA (and different oscillation amplitudes)

Referee #3 (Remarks for Author):

This is an interesting paper that describes a new strategy to detect copy number alterations in glioma via shallow whole-genome sequencing (sWGS) of cell-free DNA in cerebrospinal fluid (CSF). In

addition to analyses of somatic copy number alterations (SCNA), the paper provides some insight in to the fragment length profiles of cell-free DNA in CSF. It reads well and the conclusions are generally well-supported. I am happy to recommend this paper for publication in EMBO Molecular Medicine, and think it will be of interest to the readership of this journal. I have a few comments/suggestions that should be readily addressable.

1. Sequencing data needs to be made available, if possible in an open-access repository.

2. On page 3, lines 90-95, a statement is made regarding the sensitivity of the approach taken in this study to detect glioma, a combination of sWGS and SCNA analysis, and methods reported in the literature. The authors compare the detection rate to numbers achieved with tumor-guided methods, and more expensive approaches. There are several issues with this performance comparison: first, there is no direct comparison made with the approaches reported previously, second, the detection rate (39%) is significantly lower than what was achieved by Wang et al, third, the detection rate is likely strongly dependent on the tumor type and volume, as is also clear from Fig.1B. The authors should qualify the statement on page 3, or perform additional experiments and conduct a formal performance comparison.

3. A rarefaction analysis (detection rate vs seq depth) would be helpful to gain insight in the relationship between the cost of the assay (as determined by the depth of sequencing) and the performance of the assay in detecting glioma. No SCNA signature is detected in 8/13 samples. Is this due to technical limitations, or are these tumor cases that do not display SCNA? Would 10x deeper sequencing uncover additional features?

4. On p5 lines 145-150, a novel fragmentation signature is reported that may provide an alternate way to detect the presence of tumor DNA in CSF. This is quite compelling. The authors should provide a few ideas for the origin of this signature (negative correlation between the amplitude of 10 bp oscillations in the distribution of fragment sizes and the levels of cfDNA). Is a similar feature observed for tumor DNA in other bodily fluids? The latter question can be addressed with an analysis of seq data for plasma DNA in other tumor settings available in public repositories, for example.

1st Revision - authors' response 3 September 2018

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

Referee #1 (Comments on Novelty/Model System for Author): Mouliere and Mair et al. demonstrate in this manuscript that shallow whole-genome sequencing (sWGS) can be used with cell-free tumor DNA from cerebrospinal fluid (CSF) to detect somatic copy number alterations (SCNAs) without prior knowledge of tumor mutations. A In addition, a fragmentation pattern was identified that correlates with detection of SCNAs. This paper provides an improved low-cost screening method for patient samples. The manuscript is largely technically sound, but a few minor comments should be addressed as detailed below.

We thank the reviewer for their comments.

1. In the figure legend for Fig. 1C, the authors reference sWGS from plasma and urine samples, but this data is not included in the figure. Please include this sWGS data for comparison.

The sWGS data for the corresponding plasma and urine samples of this patient exhibited a copy number neutral profiles with no SCNAs, therefore we have not included them in the detailed Fig 1C. These SCNAs plots are now available in Suppl. Fig. 2.

2. The authors demonstrate a relatively clear negative correlation between the amplitude of 10 bp periodic oscillations and the levels of SCNAs (Fig. 2C). The authors should comment on potential explanations for why this trend is observed.

We have developed more this analysis in the text and commented in the discussion so that it now reads: "Thus an overall decrease in the peak fragment size was associated with a reduction in the

amplitude of the sub-nucleosomal peaks. The origin of the 10 bp oscillatory pattern is believed to be due to variable accessibility of the DNA due to its winding around the histone cores (Jiang & Lo, 2016). Alternative nuclease activities in cancer or alternate mechanisms of DNA release may produce an alteration in this fragmentation pattern."

3. The data provided in Supplemental Fig. 2 are very informative and important for interpretation of the paper; thus, this figure should be included in the main text.

We thank the reviewer for this suggestion. This figure is now included in the text as Figure 2.

4. In Supplemental Fig. 2B, tumor size is missing units.

We have added the unit (mm) to the figure.

Referee #2 (Comments on Novelty/Model System for Author):

Mouliere et al conducted a study regarding cell-free DNA cerebrospinal fluid. The manuscript deals with a highly interesting topic since the analysis of cell-free tumor DNA is currently one the hottest topics in the field of oncology and data on brain tumors are still sparse. However, proximal sampling by the analysis of CSF might increase sensitivity. Overall speaking, the study technically sounds. The manuscript does contain some interesting data and they are on most parts well-presented, however there are some concerns that need to be addressed.

We thank the reviewer for their comments.

1) The title ". . ." using short cfDNA" is somehow misleading, since short cfDNA is not specifically enriched with the presented methods.

We have changed the title to ***"Detection of cell-free DNA fragmentation and copy number alterations in cerebrospinal fluid from glioma patients".***

2) The authors present data from quite a small cohort of 13 patients and in only 5 of them SCNA could be detected. The author should indicate further analysis approaches for those patients where no cftDNA was detected.

In addition to the work from Wang et al, there are various other sensitive methods based on tumor-guided assay to detect cftDNA in CSF that are quoted in the manuscript. In addition, the authors have another study under revision in a different journal which suggests that tumor-derived mitochondrial DNA could be used as a surrogate to cftDNA when cftDNA analysis fails in glioma xenografts (Measurements of plasma cell-free tumor mitochondrial DNA improves detection of glioblastoma in patient-derived orthotopic xenograft models; Mair R, Mouliere F, et al Under Review)

3) In general, the results & discussion section seems a bit minimalistic and could be presented in a more comprehensive way.

We have now expanded some aspects of the results and discussion. We have added a new figure, and developed figure 3. We try to stick to the report format as well.

4) Moreover, there are some discrepancies regarding the analysis of tumor samples: Line 102: Here the author state that tumor material was available only from one patient, although two sentences before that the claim that the highest concentration of cfDNA corresponded to three patients with SCNAs in the tumor.

The study uses single region tissue samples from tumors in all cases. However we have multiple regions of tissue from within the same tumor for one patient (GB1). This is now corrected in the document, and specified in the figure 2 of the revised manuscript.

Pretty much at the beginning it says that the levels of cftDNA in CSF were not directly correlated to the tumor volume, later on the detection and confirmation of the SCNA in the CSF sample was influenced by the size of the tumor (in addition to the level of cfDNA, and the glioma grade).

The first part was referring to cfDNA and not cftDNA. This is now corrected in the text.

5) The authors observed a shift toward shorter fragment sizes of cfDNA in CSF, which was previously reported to be associated with an enrichment of tumor-derived fragments. Moreover, a negative correlation between the amplitude of the 10 bp periodic oscillations and the levels of SCNAs in these CSF samples was reported. This is somehow contradictory and the author should furthers explain this observation or at least present a hypothesis for it.

The shift towards shorter size was previously reported for plasma, however this has not been reported previously within the CSF. The reduction is the amplitude of the 10bp periodic oscillations is not contradictory with the global shortening as the 2 are distinct fragmentation features of the circulating DNA.

We have expanded figure 3 and added the following text in the manuscript to highlight this point: "Thus an overall decrease in the peak fragment size was associated with a reduction in the amplitude of the sub-nucleosomal peaks. The origin of the 10 bp oscillatory pattern is believed to be due to variable accessibility of the DNA due to its winding around the histone cores (Jiang & Lo, 2016). Alternative nuclease activities in cancer or alternate mechanisms of DNA release may produce an alteration in this fragmentation pattern."

6) Figure 2 shows different size distributions of plasma, urine, and CSF. The author should include these data in the results and discussion section.
The size distribution of plasma, urine and matched CSF samples were available only for patient GB1. This is now specified in the text.

7) Figure 2 would benefit from comparison of size profiles for a sample with and without SCNA (and different oscillation amplitudes)
Supplementary figure 3 detailed the size profile of the CSF samples for each patients and specify if SCNAs are detected or not. We have also expanded Figure 3 to highlight the differences in fragmentation between the size profiles of cfDNA when SCNAs are detected in CSF.

Referee #3 (Remarks for Author):

This is an interesting paper that describes a new strategy to detect copy number alterations in glioma via shallow whole-genome sequencing (sWGS) of cell-free DNA in cerebrospinal fluid (CSF). In addition to analyses of somatic copy number alterations (SCNA), the paper provides some insight in to the fragment length profiles of cell-free DNA in CSF. It reads well and the conclusions are generally well-supported. I am happy to recommend this paper for publication in EMBO Molecular Medicine, and think it will be of interest to the readership of this journal. I have a few comments/suggestions that should be readily addressable.

We thank the reviewer for their comments.

1. Sequencing data needs to be made available, if possible in an open-access repository.

The data will be available in the ega.box.1048

2. On page 3, lines 90-95, a statement is made regarding the sensitivity of the approach taken in this study to detect glioma, a combination of sWGS and SCNA analysis, and methods reported in the literature. The authors compare the detection rate to numbers achieved with tumor-guided methods, and more expensive approaches. There are several issues with this performance comparison: first, there is no direct comparison made with the approaches reported previously, second, the detection rate (39%) is significantly lower than what was achieved by Wang et al, third, the detection rate is

likely strongly dependent on the tumor type and volume, as is also clear from Fig.1B. The authors should qualify the statement on page 3, or perform additional experiments and conduct a formal performance comparison.

The corresponding section of the manuscript now reads as follow: "Despite the small cohort size and limited sensitivity of sWGS, which generally detects the presence of cftDNA only down to concentrations of ~5% of total cfDNA (Heitzer et al, 2013; Adalsteinsson et al, 2017), our rate of detection was similar to previous studies performed with more expensive genome-wide sequencing methods (De Mattos-Arruda et al, 2015; Pentsova et al, 2016). Targeted methods that track tumor mutations previously identified in each patient by DNA sequencing of tissue biopsies have a higher rate of cftDNA detection (Wang et al, 2015) but the requirement for invasive biopsies makes these methods impractical and expensive in a clinical setting."

3. A rarefaction analysis (detection rate vs seq depth) would be helpful to gain insight in the relationship between the cost of the assay (as determined by the depth of sequencing) and the performance of the assay in detecting glioma. No SCNA signature is detected in 8/13 samples. Is this due to technical limitations, or are these tumor cases that do not display SCNA? Would 10x deeper sequencing uncover additional features?

Published works that employ whole exome sequencing (and thus possess higher loci coverage) do not exhibited better detection rates than our work with sWGS (De Mattos-Arruda et al, 2015; Pentsova et al, 2016)). It is likely that the detection rate in CSF depends upon the contact of the tumor with the CSF spaces (Wang et al, 2015).

4. On p5 lines 145-150, a novel fragmentation signature is reported that may provide an alternate way to detect the presence of tumor DNA in CSF. This is quite compelling. The authors should provide a few ideas for the origin of this signature (negative correlation between the amplitude of 10 bp oscillations in the distribution of fragment sizes and the levels of cfDNA). Is a similar feature observed for tumor DNA in other bodily fluids? The latter question can be addressed with an analysis of seq data for plasma DNA in other tumor settings available in public repositories, for example.

We have expanded on this analysis (figure 3) and added notably the following text to the manuscript: "Thus an overall decrease in the peak fragment size was associated with a reduction in the amplitude of the sub-nucleosomal peaks. The origin of the 10 bp oscillatory pattern is believed to be due to variable accessibility of the DNA due to its winding around the histone cores (Jiang & Lo, 2016). Alternative nuclease activities in cancer or alternate mechanisms of DNA release may produce an alteration in this fragmentation pattern."

Moreover, we have another study under consideration in a different journal focusing in part on the analysis of 10bp periodic oscillations for plasma samples of other cancer types. We have therefore chosen to focus upon the analysis of CSF from GB patients for this work for avoiding duplication.

2nd Editorial Decision 25 September 2018

Thank you for the submission of your revised manuscript to EMBO Molecular Medicine. We have now received the enclosed report from the referee that was asked to re-assess it. As you will see, the reviewer is now supportive, and I am pleased to inform you that we will be able to accept your manuscript pending minor editorial amendments.

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

Referee #2 (Remarks for Author):

The authors adressed most of my concern!

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Florent Mouliere, Kevin Brindle, Nitzan Rosenfeld
Journal Submitted to: EMBO Molecular Medicine
Manuscript Number: EMM-2018-09323

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 $\chi 2$ 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?	we have included all the samples from available patients in this proof of principle study
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	not relevant
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	no samples were excluded from analysis
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.	viewing the small size of the cohort, no randomisation was performed
For animal studies, include a statement about randomization even if no randomization was used.	not relevant
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.	low number of samples processed as a single batch
4.b. For animal studies, include a statement about blinding even if no blinding was done	not relevant
5. For every figure, are statistical tests justified as appropriate?	qualitative analysis
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	qualitative analysis
Is there an estimate of variation within each group of data?	qualitative analysis
Is the variance similar between the groups that are being statistically compared?	qualitative analysis

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-tun

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/

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).	not relevant
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	not relevant

* 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.	not relevant
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	not relevant
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.	not relevant

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	Addenbrookes committee : REC - 15/EE/0094
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.	Informed consent was obtained from all patients included in this study
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	not relevant
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	no restriction on the availability of the human data
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	not applicable
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.	not relevant
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.	not relevant

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	Sequencing data will be available at EGAS00001003255
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).	Sequencing data will be available at ega_box_1048
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).	All available data are already added to the manuscript.
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) 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.	All computational methods used are already published, or a link to a public Github is avaialble (for CNAclinic).

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.	not relevant