THE EMBO JOURNAL

Absolute proteomic quantification reveals design principles of sperm flagellar chemosensation

Christian Trötschel, Hussein Hamzeh, Luis Alvarez, René Pascal, Fedir Lavryk, Wolfgang Bönigk, Heinz G. Körschen, Astrid Müller, Ansgar Poetsch, Andreas Rennhack, Long Gui, Daniela Nicastro, Timo Strünker, Reinhard Seifert, U. Benjamin Kaupp

Review timeline:

Submission date:	26th Jun 2019
Editorial Decision:	6th Aug 2019
Revision received:	26th Sep 2019
Editorial Decision:	11th Nov 2019
Revision received:	19th Nov 2019
Accepted:	22nd Nov 2019

Editor: Daniel Klimmeck

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 6th Aug 2019

Thank you again for the submission of your manuscript (EMBOJ-2019-102723) to The EMBO Journal. Please accept my sincere apologies for getting back to you with I agree very unusual delay at this time of the year as to protracted referee input. Your manuscript has been sent to four reviewers, and we have received reports from all of them, which I enclose below.

As you will see, the referees acknowledge the innovative approach taken and thoroughness of your work, although they also express a number of major issues that will have to be addressed before they can support publication of your manuscript in The EMBO Journal. While referee #4 raises a number of complementary points and controls required in his/her view regarding the signaling analyses concerning equal versus clustered distribution of channels and soluble versus bound messengers, two major key issues are raised: 1) the mass spectrometric quantification approach lacks sufficient annotation and description at this stage both at the biochemical and instrumentation-bioinformatics sides, which significantly weakens the impact of the study (ref#2 - a mass spectrometry expert; see also ref#4). 2) All referees stress that the results remain largely inaccessible and difficult to follow for a broader audience, which might lack a detailed view of the previous literature and insights into the biophysics analyses presented. The referees thus state that an extensive re-work of the manuscript would be required to present the findings in a much shorter and focussed manner in order to be compelling.

I judge the comments of the referees to be generally reasonable and given their overall interest, we are in principle happy to invite you to revise your manuscript experimentally to address the referees' comments.

We agree however that in particular the robustness of the absolute proteomics is key to this manuscript and would need to be conclusively addressed to achieve the level of robustness and clarity needed for The EMBO Journal.

REFEREE REPORTS:

Referee #1:

The authors present an interesting manuscript that seeks to quantify signaling proteins in the flagella of sea urchin sperm. As flagella, and many mammalian cilia, serve as chemosensory signaling structures the authors seek to use this model to detail a defined signaling pathway and how tight control is maintained on its function. Ultimately this information is important for understanding signaling processes across different organisms and cell types. The authors use mass spec methods to quantify the signaling proteins that control chemotactic steering of sperm and identify several new proteins that may function in these flagella. A major claim is that flagella contain many fold more signaling protein than downstream cellular messengers. This finding is novel and provides new insight into cilia/flagella function. The authors use several methods to measure production of cGMP, Vm, pH and Ca2+ in flagella and cells. Overall, I think this is a well-written and interesting paper.

Minor Issues,
Several of the figures could use a bit more detail in the figure legends. It should be made clear in either the results or legends what the sample is/size is for each figure. Figure 2c would be helped if the color corresponding to light energy was cleared defined.

Referee #2:

In their manuscript "Absolute proteomic quantification reveals design principles of flagellar chemosensation", Christian Trötschel and co-authors applied an established targeted MS (SRM/QconCAT) approach to determine absolute quantities of selected flagellar components to mechanistically decipher chemosensation in the sea urchin Arbacia punctulate. This is a very carefully executed and detailed study discussing the high flagella protein levels obtained in context to the connections of small second messenger and signaling molecules, like cGMP, Calcium, which is clearly interesting and novel. Overall, the manuscript is well written and the results well presented and the conclusions are supported by the data. Unfortunately, the data provided for the proteomics part of the paper is very limited and immature and needs to be improved. This is crucial for this manuscript, since most of the claims made are based on the absolute protein copy numbers determined. Therefore, I recommend publication if the comments below are addressed.

Comments:
Page 20: For all mass spectrometry-based proteomics studies aiming at absolute protein quantification it is crucial to prepare a representative peptide sample, in particular the complete extraction of all proteins is essential. This is not too challenging for soluble proteins, but for membrane proteins, like targeted in this study, this is not straightforward and needs to be checked. The authors use a DDM containing lysis buffer, which is quite unusual for proteomics analysis. Did the authors use this because of its good solubility properties for flagella proteins? And was it protein extraction efficiency tested? Also why was a procedure of 30 minutes on ice used? Was this tested for the samples used? Usually boiling and/or strong ultrasonication are used to assist effective protein extraction. Along this line, the authors state that only the supernatant was employed for further analysis. Does this mean that there was an insoluble flagella pellet left that could indicate incomplete protein extraction?

Page 20: In general, the proteomics analyses performed should be described in more detail to better understand the actual methods used. For instance, how were the peptides selected (please include the mudpit/GeLC-MS results in an additional supplementary table and upload the raw-files to PRIDE (https://www.ebi.ac.uk/pride/archive/)) and how were they validated by SRM? Were any heavy reference peptides used here to confidently identify the target peptides? Please describe the validation in more detail.

Page 21: The use of a QconCAT is an elegant way for protein quantification. The authors mention that the QconCAT was spiked into the flagella protein extract followed by SDS-PAGE. Why is was an SDS-PAGE used at this stage? The QconCAT and the flagella proteins would run to different positions in the gel and digested separately that will certainly introduce additional variability to the

samples. An in-solution digestion would be a more elegant way to proceed from this point. Also, multiple sample preparation methods are described in the paper cited. The authors need to either include a clearer reference or a better description of the in-gel digestion approach. Also, the next sentence mentions a new LC system (waters) and does not fit in the context. Overall, this paragraph needs substantial reworking to be understandable and useful to the reader.

Page21: The authors employed 3 different QconCAT concentrations for the analysis but did not mention this in the results part. How were they used for the quantification?

Page 21: In the next paragraph, the authors suddenly switch to a different MS approach providing some general instrumental details but lack to describe any parameters used to analyze their samples. The reference provided references another manuscript for the SRM analysis and it is not clear what was done. The authors should provide details on LC gradient, MS parameters used and in particular how the data was validated and quantified (transition, q-values for identifications scoring...). Also, the SRM data should be uploaded to Panorama (https://panoramaweb.org/project/home/begin.view?) or transition plot of all quantitated peptides need to be included in the supplementary section of the manuscript to allow inspection by others.

Page 22: It is good that the authors include isotope impurities in their protein abundance calculations. Unfortunately, only a very limited amount of the quantitative data analysis is provided in supplementary table 2. In table 1, a bit more information is provided, but it would be interesting to show the absolute quantities of all peptides obtained to better judge the accuracy of the absolute quantification results. At least the ratios, summed transition intensities (peak areas), number of transitions used and absolute quantities for the 3 different spiked in QconCAT concentrations should be shown.

Page 22: The authors should describe the basis for removing "Inaccurately identified peptides" outliers from quantification.

Referee #3:

This is an exciting approach turning traditional enzymology where substrate concentrations are in vast excess to that of enzymes in solution and instead provides a more realistic overview inside cells. This is an excellent example where quantitative proteomics together with optochemical approaches with enzymology can offer a new perspective to compartmentalization. In the crowded cellular environment, receptors are in vast excess to signalling molecules. This has far reaching implications for second messenger signalling where the receptors in their abundant concentrations function more as kinetic traps for substrate.
I support publication but the paper needs to be rewritten to emphasize the major conclusions. The abstract in particular- has been made quite inaccessible "Our results pose "paradigmatic questions about recording signalling events in small compartments" for instance. Would help the reader if the authors described the implications more clearly. I also have the following major points for the authors to address:

1) Another example of "inaccessible" writing- On P.3- "It is thought that the exquisite sensitivity of cilia or flagella rests on the inventory and topographical arrangement of signalling molecules and the compartmentalization of cellular reactions" . A reference is needed to qualify "It is thought ...." or is this the authors' viewpoint? If so, it needs to be clarified.
2) Again, reference is needed for "A concept of physiological significance has emerged that cilia and flagella are specialized cell organelles"
3) What are the lower limits for binding constants in the complexes? How sensitive is the protein quantitation to dissociation constants. The concentration of nucleotides will impact how much remains bound? What kind of variability in quantitation is obtained under conditions of receptor overstimulation?

This is a completely new view of compartmentalization and protein signaling networks. Quantitation also has major implications for interpretation for deletion mutagenesis, point mutants in signaling networks. This will be of importance to readers of EMBO J.

Referee #4:

The manuscript „ Absolute proteomic quantification reveals design principles of flagellar chemosensation" decribes interesting experiments on see urgin sperm that enlighten the mechanism of chemosensation in this model system.
The manuscript is in principle of high interest for the general readership of EMBO Journal, it demonstrates in detail that the flagella contains more cyclases than cAMP molecules even after GC-activation. The authors develop a new concept of a concerted action of the cGMP-activated channel and a directly linked phosphodiesterase which might be of high relevance for other flagellar systems including human sperm.
But in its current state the manuscript is hard to read and almost impossible to follow the argumentation unless the reader consults a number of previous publications.
I highly appreciate the data but the presentation and interpretation is not acceptable in its current state.
The lack of transparency is partially owing to the general disadvantage of this model system that it is not genetically tractable is such a way that individual components of the machinery can specifically be deleted.

Details

l.101: here the confusion begins. The authors are stating: Changes in Ca2+i control the swimming path". What is meant here ? I guess that the authors do NOT mean the single free Ca2+ in the whole flagella and also agree that the essential value is the total Ca2+ that enters the flagella most of which is immediately bound. Here it is of importance that most of the Ca2+ is bound to proteins of the dense axoneme. This should be made clear already at this point.

l.129: the QconCat technology is not well described. To name the proteins QconCat is misleading an unnecessary. Synthetic "Labelled protein" versus flagellar "unlabeled protein" everyone would understand. See: https://www.polyquant.com/technology/
for the simplest way to describe the technology.

l.140ff: Numbers of the transporter molecules per flagella do not mean much for the reader unless the surface area is given. I calculated it myself and found it much later in table I.
It would be very helpful if the calculated flagellar area of 30 mirometer2 (3x10E-5) and the volume of 1 femtoliter = 10E-15 l (1 cubic micrometer) would be given here.
Since the diameter is absolutely essential, an EM or tomography figure would be convincing in this respect.
The next question is: are the proteins equally distributed along the flagella or are clustered (but this question my go beyond the scope of the manuscript)
l.150: What are "signaling proteins" in this context? Moreover I cannot follow the argumentation that these molecules (probably GC and ion transporters) are more abundant than H+, Ca2+, cAMP and cGMP because in case of the later four, "concentrations" are considered that include only free but non membrane-bound molecules. On the other hand the transporters are ONLY membrane-bound / -integrated molecules. So, I do not understand what "concentration" (which is a definition for a volume) means for proteins in a two dimensional membrane ?
- the cGMP capture is efficient because the local diffusion dimensions are so small. Along one micometer of the flagella, the "volume" is only 25 attoliter, and in case you like to calculate with "concentrations" in such a volume a single molecule in an attoliter would correspond to 10 micromolar, but I doubt that this makes any sense. In reality the free volume is even much smaller because most of the flagella is occupied by axonemal proteins.
To my understanding the only parameter that counts is the number of signaling molecules AND messengers along the inner surface, separated from those bound to the axoneme in case of Ca2+.
The question of soluble vs bound ions has been discussed for photosynthesis, where the protons in the chloroplast drive the F0F1-ATPase although at pH 7 only a single free proton is in the small volume.

l.156. "The large number of CNGK channels suggest that capturing of cGMP is highly efficient". One might conclude the opposite: you need a large number of channels because the binding is inefficient (??).
l.157. here it is not clear for the reader why HCN1 is a downstream target. It is activated by a cAMP instead of cGMP. This should be clarified. Which activator dominates sNHE and HCN1 activation: high voltage or cAMP ? Please refer to Fig.1 in this chapter.
l.172. please use signaling proteins instead of signaling molecules (because the messengers are also signaling molecules)
l.172. is here the total cAMP,cGMP and Ca2+ meant including or excluding the Ca2+ bound to the axoneme ?
l.174. the terms "high copy number" or "high density" used here are much better than the term "concentration" for proteins in a 2D-membrane.
l.185: this is a very interesting perspective that signaling from the GC to CNGK must be most efficient as well. (see discussion)
l.192. The next critical point is the quantification of the cGMP and the membrane voltage change. The statement "the cGMP inside the cells can be calibrated precisely" is not enough for such a critical parameter. The reader needs to know the validity of the method and obviously in this case the released cGMP is meant here and not the free cGMP. Calculation of the free cGMP by using equations of p.24 is rather indirect and the reliability questionable.
Equally suspicious is the determination of the membrane voltage. How reliably is PeT FluoVolt ? To explain both methods more precisely is absolutely necessary before the results can be validated. The recordings look beautiful in Fig.2 and Fig.4c, but I am still quite skeptical about the linear range, pH and ion sensitivity etc. . Why not saying .....by using the electrochromic fluorescent organic dye XYZ....at line 192 or 206 .? But, how many dye-molecules are integrated in the membrane ? They respond to membrane voltage changes by charge redistribution which also influences the voltage. One needs a calibration curve for these measurements.
l.211. what eqn. 1 are the authors referring to here ?
l.233 .... "that dissociation from CNGK determines the rate of Vm recovery rather than PDEs catalytic power". This can also be explained in a better way. In light of the fast on-kinetics for cGMP binding, the off kinetics must also be fast because otherwise the equilibrium would not be established quickly during the 2 second light pulse. Whether this is determined by the off-kinetics of the CnGK or the on-kinetics of the PDE cannot be determined from this argumentation.
l.235 by this statement the authors want the reader to believe instead of convincing them.
L237 ff. the chapter can be substantially shortened and the essence can be explained more clearly. In l.251 the authors conclude that the hydrolytic power of PDE5 is not rate limiting but in L.260 they conclude that hydrolysis of cGMP fails to explain the fast break down. On the other hand the small number of cGMPs compared with the large number of PDEs only requires cGMP binding and not large turn overs of the PDE. And again simulations that are not shown should be deleted from the text.
l.281: time or light range ?
l.192. .....while CNGK channels are still open .... Please refer to the respective figure to compare repolarization with CNGK closure
l.295. why slightly ? it is twice the size
l.304: Why is this experiment different from Fig. 2b. ?
- conclusions of this paragraph must be made more transparent and consistent.

l. 319. Please refer to Fig.1
l. 348. Here I would appreciate the used dye to be mentioned. Does it record free Ca2+ or total Ca2+ which depends on the binding affinity.
l.360 by using pHrodoRed ??
l.361. how much ? any quantification ?
l.373. this ratio is misleading as explained above and small molecules - especially H+ and Ca2+ are highly buffered.
l.377. most of the reactions are surface reactions.
l.380. surface diffusion is in most cases faster than diffusion through solution space especially in densely packed media
l. 382. This is an important conclusion of the paper and should be more high-lighted (as already mentioned above).
l.386. misleading

l.389. I think another aspect is important: in a local volume of 25 fl along 10 micometer flagella, you cannot deposit many molecules because a very small number would generate
large chemical and electrical gradients.
l.401. what happens with the GMP and AMP ? do they diffuse back into the cell body ?
l.436. which is < 1 molecule within the flagella of 1 fl (10-15 l), this calculation does not make sense

at the end I like to mention that the topic is super interesting but it has to be mediated in a clearer more transparent and more focused way, high-lighting the essential points without diluting them by a lengthy discussion of side aspects. I would reduce the discussion by > 50%.

l.821: can the signal of Fig.S2 be converted to a pHi change or number of protons ?

1st Revision - authors' response 26th Sep 2019

Response to the referees:

We thank the referees for their time to review the manuscript, the positive feedback, and the constructive criticism. We revised the manuscript according to the referees´ suggestions.

Referee #1:

Minor Issues, Several of the figures could use a bit more detail in the figure legends. It should be made clear in either the results or legends what the sample is/size is for each figure. Figure 2c would be helped if the color corresponding to light energy was cleared defined.
We scrutinized all figures and provided the respective information.

Referee #2:

Page 20*: "For all mass spectrometry-based proteomics studies aiming at absolute protein quantification it is crucial to prepare a representative peptide sample, in particular the complete extraction of all proteins is essential. This is not too challenging for soluble proteins, but for membrane proteins, like targeted in this study, this is not straightforward and needs to be checked. The authors use a DDM containing lysis buffer, which is quite unusual for proteomics analysis. Did the authors use this because of its good solubility properties for flagella proteins? And was it protein extraction efficiency tested? Also why was a procedure of 30 minutes on ice used? Was this tested for the samples used? Usually boiling and/or strong ultrasonication are used to assist effective protein extraction. Along this line, the authors state that only the supernatant was employed for further analysis. Does this mean that there was an insoluble flagella pellet left that could indicate incomplete protein extraction?"*
We used DDM for its good solubilization properties at low temperatures and because it has been used in previous work. We tested the extraction efficacy of membrane proteins by two independent approaches that are now detailed in the Methods section and documented in Supplementary Fig. 3 and Supplementary Table 4, respectively. Briefly, about > 90% of the GC, the CK, and of three other proteins, which we tested by Coomassie staining or Western blotting, is solubilized. Neither sonification nor longer incubation times did markedly improve solubilization. We abstained from boiling, because it may cause protein aggregation. There was a pellet that consisted primarily of axonemal proteins like

tubulin (55 kDa) and an intermediate chain of dynein (75 kDa). Moreover, to quantify solubilization by an independent method, the flagella samples from three biological replicates were split: one probe was solubilized with DDM, centrifuged, and the supernatant analyzed; the other probe was first solubilized with DDM, followed by SDS sample buffer (no centrifugation step). For both conditions, the complete workflow from SDS-PAGE to SRM and quantification with Skyline was carried out. Except for the soluble PDE10, the protein content for DDM vs. DDM + SDS solubilization was similar (Supplementary Table 4). Thus, we trust that our procedures are as quantitative and reproducible as possible throughout.

Page 20: "*In general, the proteomics analyses performed should be described in more detail to better understand the actual methods used. For instance, how were the peptides selected (please include the mudpit/GeLC-MS results in an additional supplementary table and upload the raw-files to PRIDE (https://www.ebi.ac.uk/pride/archive/)) and how were they validated by SRM? Were any heavy reference peptides used here to confidently identify the target peptides? Please describe the validation in more detail.*"

Raw files and search results (PepXML and ProtXML files) have been uploaded to PRIDE repository. The data at PRIDE can be accessed with the following information:

Project Name: Arbacia punctulata flagellar proteome

Project accession: PXD015332

Project DOI: Not applicable

Reviewer account details:

Username: reviewer34535@ebi.ac.uk

Password: UW11Bxgn

How peptides were selected from the GeLC and MudPIT data was already described in the original manuscript: "Only peptides without methionine, cysteine as well as a series of arginine and lysine residues (e.g., KK, KR, RK, or RR) were validated in subsequent SRM measurements". As requested, additional details about the peptide validation with the heavy counterpart are now provided:" Importantly, the synthetic isotope-labeled protein was used to confirm the identity of the previously selected 53 peptides for quantification by manual comparison of pairwise retention times and fragment ions between the respective sample and standard peptides."

Page 21: *"The use of a QconCAT is an elegant way for protein quantification. The authors mention that the QconCAT was spiked into the flagella protein extract followed by SDS-PAGE. Why is was an SDS-PAGE used at this stage? The QconCAT and the flagella proteins would run to different positions in the gel and digested separately that will certainly introduce additional variability to the samples.*

The referee's concern was also our concern. It is for this reason that SDS-PAGE samples were allowed to migrate only shortly (i.e. for about 1 cm) into the separation gel to avoid separation, following the procedure described by the cited

reference Pichlo et al. (2014). It should now become clear from the revised text that the gel material that contained all loaded proteins was used for digestion of all proteins at once. We trust that this clarification also convinces the referee.

An in-solution digestion would be a more elegant way to proceed from this point. Also, multiple sample preparation methods are described in the paper cited. The authors need to either include a clearer reference or a better description of the in-gel digestion approach. Also, the next sentence mentions a new LC system (waters) and does not fit in the context. Overall, this paragraph needs substantial reworking to be understandable and useful to the reader."

We described the method in our previous publication (Pichlo et al. 2014) in section *"*Quantification of GC density by SRM MS*"*. We tested (unpublished) in-solution digestion with (1) trypsin in NH_4HCO_3 buffer and (2) trypsin in presence of the surfactant RapiGEST. However, the in-gel digestion protocol was more robust regarding protein sequence coverage and it resulted in a better S/N ratio of the quantification – most likely due to the extensive washing removing potentially interfering small molecules.

We revised the Methods section accordingly: now sample processing and LC-MS/MS are in two separate paragraphs. The LC conditions were essentially identical to those described in Pichlo et al. (2014); yet, we provide now more information to facilitate understanding of the LC-MS/MS procedure.

Page21: *"The authors employed 3 different QconCAT concentrations for the analysis but did not mention this in the results part. How were they used for the quantification?"*

For quantification, an aliquot of 20 µg sperm flagella protein was spiked with either 0.15, 1.5, or 15 pmol QconCAT standard protein to account for different concentrations of sperm target proteins; the idea is to obtain H and L peptide intensities of similar range. For example, 15 pmol and 1.5 pmol were suitable for the GC, whereas low abundant proteins like CatSper required 1.5 and 0.15 pmol standard protein. Ultimately, as many L/H ratios as possible were derived based on manual inspection of peak quality. The L/H ratios for every spike-in sample are documented in Supplementary Table 6.

Page 21: *"In the next paragraph, the authors suddenly switch to a different MS approach providing some general instrumental details but lack to describe any parameters used to analyze their samples. The reference provided references another manuscript for the SRM analysis and it is not clear what was done. The authors should provide details on LC gradient, MS parameters used and in particular how the data was validated and quantified (transition, q-values for identifications scoring...). Also, the SRM data should be uploaded to Panorama (https://panoramaweb.org/project/home/begin.view?) or transition plot of all quantitated peptides need to be included in the supplementary section of the manuscript to allow inspection by others."*

In the new version, selection of candidate peptides and SRM quantification are described in two separate paragraphs. We did not base identification on q-values, because L/H pairs for each peptide are more powerful. To confirm identity, the four known most intense and specific transitions were used in manual pairwise comparisons. SRM data has been uploaded to *Panorama*. Peptide and protein ratios are also described in Supplementary Tables 5 and 6.

The Panorama dataset with ProteomeXchange ID PXD015502 can be accessed for review with this login information from panoramaweb.org:

Link: https://panoramaweb.org/arbacia_flagella.url
Email: panorama+ruhr1@proteinms.net
Password: pNF1ilBk

Page 22: *"It is good that the authors include isotope impurities in their protein abundance calculations. Unfortunately, only a very limited amount of the quantitative data analysis is provided in supplementary table 2. In table 1, a bit more information is provided, but it would be interesting to show the absolute quantities of all peptides obtained to better judge the accuracy of the absolute quantification results. At least the ratios, summed transition intensities (peak areas), number of transitions used and absolute quantities for the 3 different spiked in QconCAT concentrations should be shown."*

Supplementary Table 6 provides information about L/H pair ratios and spiked-in QconCAT standard concentrations for all samples. With respect to information about peak areas, we refer the reviewer and readers to the *Panorama* dataset that contains the Sykline files with the respective peak areas (without isotope impurity correction) for all peptides and samples. The two most intense and specific transitions were always used for quantification; this is now mentioned in the manuscript.

Page 22: "*The authors should describe the basis for removing "Inaccurately identified peptides" outliers from quantification."*

Essentially, the QconCAT standard protein was used to verify the peptides selected for quantification. The following common criteria were used for verification: pairwise identical retention times and similarity of fragment ion mass and intensity for the heavy (standard) and light (sample) peptide. Three sample peptides did not meet these criteria, and were not used for quantification. The approach is described in the new section "Selection of candidate peptides for protein quantification".

Referee #3

1. "I support publication but the paper needs to be rewritten to emphasize the major conclusions. The abstract in particular- has been made quite inaccessible "Our results pose "paradigmatic questions about recording signalling events in small compartments" for instance. Would help the reader if the authors described the implications more clearly."

We completely rewrote the abstract. We also rewrote other parts to render the manuscript more accessible.

2. "Another example of "inaccessible" writing- On P.3- "It is thought that the exquisite sensitivity of cilia or flagella rests on the inventory and topographical arrangement of signalling molecules and the compartmentalization of cellular reactions" . A reference is needed to qualify "It is thought ...." or is this the authors' viewpoint? If so, it needs to be clarified."

We have changed the text and included references.

3. "Again, reference is needed for "A concept of physiological significance has emerged that cilia and flagella are specialized cell organelles"

References have been included.

4. "What are the lower limits for binding constants in the complexes? How sensitive is the protein quantitation to dissociation constants. The concentration of nucleotides will impact how much remains bound? What kind of variability in quantitation is obtained under conditions of receptor overstimulation? "
We are not sure whether we understand the comments correctly. Does the reviewer mean protein complexes or complex between a ligand and a receptor (e.g. CNGK channel and cGMP)? The quantification of proteins by mass spectrometry is completely independent of any ligand. The proteins are denatured and tryptically digested. Dissociation constants are irrelevant for protein quantification. Which variability in quantification does the reviewer refer to: protein or cGMP? What is meant with "receptor overstimulation"?

Referee #4
The manuscript „ Absolute proteomic quantification reveals design principles of flagellar chemosensation" describes interesting experiments on sea urchin sperm that enlighten the mechanism of chemosensation in this model system. The manuscript is in principle of high interest for the general readership of EMBO Journal, it demonstrates in detail that the flagella contains more cyclases than cAMP molecules even after GC-activation. The authors develop a new concept of a concerted action of the cGMP-activated channel and a directly linked phosphodiesterase which might be of high relevance for other flagellar systems including human sperm. But in its current state the manuscript is hard to read and almost impossible to follow the argumentation unless the reader consults a number of previous publications. I highly appreciate the data but the presentation and interpretation is not acceptable in its current state. The lack of transparency is partially owing to the general disadvantage of this model system that it is not genetically tractable is such a way that individual components of the machinery can specifically be deleted.
We turned the lack of genetical tools into an advantage. From sea urchin sperm as a model, we can gain quantitative insights from an easily accessible and highly homogenous population of **intact motile cells**. In k.o. models, the expression of proteins other than the targeted ones are affected for unknown reasons. In fact, k.o. models should be accompanied by quantitative MS to look for potential changes in protein stoichiometry. Moreover, many of the signaling proteins in sea urchin sperm have been cloned and characterized by heterologous expression, and a wealth of quantitative information exists in original research papers (Gauss et al. *Nature* 1998; Strünker et al. *Nature Cell Biol.* 2006; Windler et al. *Nature Comm.* 2018) and reviews (Wachten et al. *CSH Perspectives* 2017; Kaupp & Strünker *Trends Cell Biol.* 2017).

l.101*:"here the confusion begins. The authors are stating: Changes in Ca2+i control the swimming path". What is meant here? I guess that the authors do NOT mean the single free Ca2+ in the whole flagella and also agree that the essential value is the total Ca2+ that enters the flagella most of which is immediately bound. Here it is of importance that most of the Ca2+ is bound to proteins of the dense axoneme. This should be made clear already at this point."*
The changes in free Ca^{2+} concentration is meant. The changes in total Ca^{2+} cannot be measured. Enzymes respond to changes in free $[Ca^{2+}]_i$ as do indicators. We include now the word "free" $[Ca^{2+}]_i$ whenever necessary, although the notation "$[Ca^{2+}]_i$" itself always refers to free Ca^{2+} concentration in the literature.

l.129: *"the QconCat technology is not well described. To name the proteins QconCat is misleading an unnecessary. Synthetic "Labelled protein" versus flagellar "unlabeled protein" everyone would understand. See: https://www.polyquant.com/technology/ for the simplest way to describe the technology."*
We rephrased in more detail and in simpler terms the part describing QconCAT. The acronym QconCAT is established in the MS community. Nonetheless, we replaced "QconCAT protein" by "synthetic labelled protein" or "standard protein".

l.140ff: *"Numbers of the transporter molecules per flagella do not mean much for the reader unless the surface area is given. I calculated it myself and found it much later in table I. It would be very helpful if the calculated flagellar area of 30 mirometer2 (3x10E-5) and the volume of 1 femtoliter = 10E-15 l (1 cubic micrometer) would be given here."*
We listed in Table 1 copy numbers, concentrations, and densities (where applicable) along with the geometric volume and the surface area. Thus, in the original manuscript all necessary information was provided in Table 1. There was a simple rationale why we list protein concentrations: the abundance of proteins relative to each other and relative to the respective second messengers can be compared. The same holds true for copy numbers and densities. Finally, it provides a sense for the relation between copy numbers and concentrations.

l.140ff: *"Since the diameter is absolutely essential, an EM or tomography figure would be convincing in this respect."*
We teamed up with the group of Daniela Nicastro (UT Southwestern, Dallas,) to determine both the total volume and various sub-volumes that are accessible to small messenger molecules or diffusible proteins. A new section along with a new Figure 2 describes this work.

l.140ff: *"The next question is: are the proteins equally distributed along the flagella or are clustered (but this question my go beyond the scope of the manuscript)."*
For three proteins that have previously been examined by immunocytochemistry, the antibodies stain the entire flagellum; yet, this data is not quantitative. Thus, homogenous *vs*. localized distribution is not known.
l.150: *"What are "signaling proteins" in this context?*
The chemotactic signaling proteins shown and described in Fig. 1b and in Table 1 are meant.

l.150: *"Moreover I cannot follow the argumentation that these molecules (probably GC and ion transporters) are more abundant than H+, Ca2+, cAMP and cGMP because in case of the later four, "concentrations" are considered that include only free but non membrane-bound molecules. On the other hand the transporters are ONLY membrane-bound / -integrated molecules. So, I do not understand what "concentration" (which is a definition for a volume) means for proteins in a two dimentional membrane?"*
We explain and answer these issues above (l. 140ff). If readers need different chemical units, they can choose from Table 1. Concentration, does not necessarily imply homogenous distribution over the volume. Thus, even for a membrane

protein contained in a well-defined volume, a concentration can be meaningful – at least for comparison.

l.150: *"- the cGMP capture is efficient because the local diffusion dimensions are so small. Along one micometer of the flagella, the "volume" is only 25 attoliter, and in case you like to calculate with "concentrations" in such a volume a single molecule in an attoliter would correspond to 10 micromolar, but I doubt that this makes any sense. In reality the free volume is even much smaller because most of the flagella is occupied by axonemal proteins."*
It's a common belief that free volumes are "much smaller". We rectify this issue by determining accessible volumes from EM tomograms. In fact, the accessible volumes for proteins or small molecules are 75% to 85% of the total volume (see new section and new Fig. 2).
l.150: *"To my understanding the only parameter that counts is the number of signaling molecules AND messengers along the inner surface, separated from those bound to the axoneme in case of Ca2+. The question of soluble vs bound ions has been discussed for photosynthesis, where the protons in the chloroplast drive the F0F1-ATPase although at pH 7 only a single free proton is in the small volume."*
We appreciate that the referee points out all these interesting biophysical questions. However, we are afraid that discussing local vs. bulk concentrations will impair rather than enhance the clarity of the main message. The issue that only few cGMP molecules are free on average is extensively addressed in the discussion.

l.156. *"The large number of CNGK channels suggest that capturing of cGMP is highly efficient". One might conclude the opposite: you need a large number of channels because the binding is inefficient (??)."*
The efficacy of capturing (and binding) a messenger molecule depends on the concentration of the receptor **and** the value of the dissociation constant K_D. The K_D of CNGK is very low (26 nM), as mentioned in line l.437 of the manuscript. The concentration/density is given in Table 1. The high density ensures that *all* cGMP molecules are *rapidly* captured ("perfect absorber").

l.157. *"here it is not clear for the reader why HCN1 is a downstream target. It is activated by a cAMP instead of cGMP. This should be clarified. Which activator dominates sNHE and HCN1 activation: high voltage or cAMP ? Please refer to Fig.1 in this chapter."*
The meaning of "downstream" is clear from Fig. 1b (which we refer to in the new version). Both, sNHE and HCN1, are activated by the CNGK-mediated hyperpolarization; binding of cAMP shifts their voltage dependence. Therefore, it would be misleading to use the word "dominates". The modulation is exhaustively described in Gauss et al. *Nature* (1998) and Windler et al. *Nature Comm.* (2018). We now avoid mentioning cAMP here, because this information is not important for the argument.

l.172. *"please use signaling proteins instead of signaling molecules (because the messengers are also signaling molecules)."*
We changed "signaling components to signaling proteins".

l.172. *"is here the total cAMP,cGMP and Ca2+ meant including or excluding the Ca2+ bound to the axoneme ?"*

We changed “concentration” to “free concentration”.

l.174. *“the terms "high copy number" or "high density" used here are much better than the term "concentration" for proteins in a 2D-membrane.”*
no response required.
l.185: *“this is a very interesting perspective that signaling from the GC to CNGK must be most efficient as well. (see discussion).”*
no response required

l.192. *“The next critical point is the quantification of the cGMP and the membrane voltage change. The statement "the cGMP inside the cells can be calibrated precisely" is not enough for such a critical parameter. The reader needs to know the validity of the method and obviously in this case the released cGMP is meant here and not the free cGMP. Calculation of the free cGMP by using equations of p.24 is rather indirect and the reliability questionable.”*
The calibration of cGMP concentrations has been extensively reviewed in practical terms in Hamzeh et al. 2019. What we mean is released cGMP by photolysis, which – of course – is the free cGMP concentration before cGMP is bound to the CNGK channel.

l.192. *“Equally suspicious is the determination of the membrane voltage. How reliably is PeT FluoVolt ? To explain both methods more precisely is absolutely necessary before the results can be validated. The recordings look beautiful in Fig.2 and Fig.4c, but I am still quite skeptical about the linear range, pH and ion sensitivity etc. . Why not saying .....by using the electrochromic fluorescent organic dye XYZ....at line 192 or 206 .? But, how many dye-molecules are integrated in the membrane? They respond to membrane voltage changes by charge redistribution which also influences the voltage. One needs a calibration curve for these measurements.”*
We take issue with this comment and the word “suspicious”. The method to record and calibrate V_m changes in sea urchin sperm and the chemoattractant-induced voltage responses have been extensively described for electrochromic and PeT dyes (Seifert et al. 2015 *EMBO Journal*, Strünker et al. *Nature Cell Biology* 2006, Hamzeh et al. *Meth. Cell Biol.* 2019). The use of these dyes and their calibration is comprehensively summarized in Hamzeh et al. 2019, (Fig. 6, p.499-503). The sizes and waveforms of V_m responses recorded with four different dyes (di-8-ANEPPS, FluoVolt, BeRST, and RH 414) are similar (see Strünker et al. *Nature Cell Biol*. (2006), Hamzeh et al. *Meth. Cell Biol*. (2019), and this manuscript). At l.205/206 of the original manuscript, it reads “Changes in V_m were followed with the voltage-sensitive PeT dye Fluovolt (Hamzeh et al. 2019; Miller et al. 2012).”

l.211. *“what eqn. 1 are the authors referring to here ?”*
eqn.(1) on p.25 of the Method section. The revised version includes this referral.

l.233 ff *"that dissociation from CNGK determines the rate of Vm recovery rather than PDEs catalytic power". This can also be explained in a better way. In light of the fast on-kinetics for cGMP binding, the off kinetics must also be fast because otherwise the equilibrium would not be established quickly during the 2 second*

light pulse. Whether this is determined by the off-kinetics of the CnGK or the on-kinetics of the PDE cannot be determined from this argumentation."
The intention of the paragraph from l.224 to l.235 is to show that the recovery rates are independent of the cGMP or CNGK-cGMP concentration and independent of short vs. long flashes. This observation suggests that the decay of the CNGK-cGMP complex, i.e. dissociation of cGMP from the channel, like radioactive decay of an isotope, is rate-limiting.

l.235 *"by this statement the authors want the reader to believe instead of convincing them."*
This comment seems to be out of place, considering that the sentence *verbatim* states that we want to carefully test ("scrutinize") by yet other techniques (simulation) whether the above hypothesis is correct. It prepares the reader for the next subchapter about simulations.

l237 ff. *"the chapter can be substantially shortened and the essence can be explained more clearly. In l.251 the authors conclude that the hydrolytic power of PDE5 is not rate limiting but in l.260 they conclude that hydrolysis of cGMP fails to explain the fast break down. On the other hand the small number of cGMPs compared with the large number of PDEs only requires cGMP binding and not large turn overs of the PDE. And again simulations that are not shown should be deleted from the text."*
The two different conclusions derive from two different "experimental" situations. One simulation (PDE alone) estimates the catalytic rate at which PDE can remove a cGMP molecule; this rate is fast and would be not rate-limiting. The second situation (PDE + CNGK) estimates the kinetics of hydrolysis when rebinding of cGMP to the channel occur. This rate – as expected – is much slower than the experimental τ_{relax}. cGMP does first bind to the *CNGK* channel and not to PDE5 – as has been clearly stated in the original manuscript. All simulations mentioned in the text are shown. We have now added a new panel *a* to Fig. 4 that schematically illustrates the various simulation scenarios to avoid this confusion.

l.281: *"time or light range?"*
l.281 light range

l.292: *....."while CNGK channels are still open .... Please refer to the respective figure to compare repolarization with CNGK closure."*
In fact, we refer to the figure in the original manuscript: ("Fig. 4a, compare blue and red traces").

l.295. *"why slightly? it is twice the size"*
We replace "slightly" by "the difference is 1.5 fold."

l.304*: "Why is this experiment different from Fig. 2b. ? - conclusions of this paragraph must be made more transparent and consistent."*
The experiment shown in Fig. 2b is different from that in Fig. 4c, because one is a hyperpolarizing response (Fig. 4c), the other a depolarizing response (Fig. 2b). In Fig. 4c, the HCN channel is active, in Fig. 2b it's not. Thus, there are no inconsistences.

l. 319. *"Please refer to Fig.1"*
We refer now to Fig. 1.
l. 348. *"Here I would appreciate the used dye to be mentioned. Does it record free Ca2+ or total Ca2+ which depends on the binding affinity."*
We mention now FluoForte. The dye measures free Ca^{2+} concentrations.

l.360. *"by using pHrodoRed ??"*
We mention now pHrodo Red.

l.361. *"how much ? any quantification ?"*
Changes in pH_i were quantified in Seifert et al. (2015). For maximal stimulation, $\Delta pH_i \cong 0.2$.

l.373. *"this ratio is misleading as explained above and small molecules - especially H+ and Ca2+ are highly buffered."*
We clarify that free concentrations are meant.

l.377. *"most of the reactions are surface reactions."*
We include a sentence that highlights the surface reactions.

l.380. *"surface diffusion is in most cases faster than diffusion through solution space especially in densely packed media"*
We mention now that surface diffusion may be enhanced. However, we are reluctant to speculate too much.

l. 382. *"This is an important conclusion of the paper and should be more high-lighted (as already mentioned above)."*
We high-lighted this conclusion by explaining what makes a perfect absorber.

l.386. "*misleading"*
We refer now to free concentrations.

l.389. *"I think another aspect is important: in a local volume of 25 fl along 10 micometer flagella, you cannot deposit many molecules because a very small number would generate large chemical and electrical gradients."*
Does the reviewer mean chemical gradients across membranes or along the flagellum? It is not clear whether the referee wants us to include this idea in the manuscript.

l.401. *"what happens with the GMP and AMP ? do they diffuse back into the cell body?"*
The point of the manuscript is that cAMP and cGMP are captured before they escape to the cell body, in this case the sperm head. We do not expect back diffusion. If molecules reach the head, they are taken care of there.

l.436. *"which is < 1 molecule within the flagella of 1 fl (10-15 l), this calculation does not make sense*
The matrix volume is 1.7 fL and the flagellum length is ca. 40 µm. In that volume, 1 nM equals 1 molecule. We are talking here about a statistical mean concentration. The referee him/herself mentions that "a single free proton" exists in chloroplasts at pH 7.0.

l.436. "*at the end I like to mention that the topic is super interesting but it has to be mediated in a clearer more transparent and more focused way, high-lighting the essential points without diluting them by a lengthy discussion of side aspects. I would reduce the discussion by > 50%."*
The referee considers the discussion lengthy and proposes to shorten it. However, the point addressed above indicates that one of the most important messages and insights - almost no free cGMP molecules exist at rest in the flagellum - apparently didn't come across. The referee does not specify the "side aspects" of the discussion. Nonetheless, we searched point-by-point the discussion for potential shortening. However, we realized that the findings are so fundamental that they require and deserve a thorough discussion of all the various implications that might be interesting to a broad readership.

l.821*: "can the signal of Fig.S2 be converted to a pHi change or number of protons?"*
See answer in l. 361. To calculate the numbers of protons would require knowledge about the intrinsic buffer capacity of the flagellum. For cells it has been estimated to be of the order of a few ten millimolar.

2nd Editorial Decision 11th Nov 2019

Thank you for submitting your revised manuscript for consideration by The EMBO Journal. Please again accept my sincere apologies for the unusual delay with the reassessment of your study as to protracted referee input. Your amended manuscript was sent back to three of the referees for re-evaluation, and we have received comments from two of them, which I enclose below.

As you will see the referees find that their concerns have been sufficiently addressed and they are now broadly in favour of publication.

Thus, we are pleased to inform you that your manuscript has been accepted in principle for publication in The EMBO Journal, pending some minor issues related to formatting and data representation as listed below, which need to be adjusted at re-submission.

REFEREE REPORTS:

Referee #2:

The authors have substantially revised the manuscript and satisfactorily addressed all the comments of this reviewer concerning the mass spectrometric analysis. The readability and clarity of the technical aspects of the manuscript have been considerably improved and the manuscript is now ready for publication.

Referee #3:

I am satisfied with the revision. The authors have vastly improved accessibility of their manuscript.

2nd Revision - authors' response 19th Nov 2019

The authors performed the requested editorial changes.

3rd Editorial Decision 22nd Nov 2019

Thank you for submitting the revised version of your manuscript. I have now evaluated your amended manuscript and concluded that the remaining minor concerns have been sufficiently addressed.

Thus, I am pleased to inform you that your manuscript has been accepted for publication in the EMBO Journal.

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: U. Benjamin Kaupp, Ansgar Poetsch, and Christian Trötschel
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBOJ-2019-102723

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?	N.A.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	We only used sperm.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Criterion was vigorous swimming of >70% of the cells in a sample.
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.	N.A.
For animal studies, include a statement about randomization even if no randomization was used.	N.A.
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.	N.A.
4.b. For animal studies, include a statement about blinding even if no blinding was done	No blinding necessary.
5. For every figure, are statistical tests justified as appropriate?	Statistical test were not applied to the data. Raw data or mean +- s.d. are provided.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	No statistical test used. The distribution of data was not assessed.

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/

Is there an estimate of variation within each group of data?	Standard deviation is provided.
Is the variance similar between the groups that are being statistically compared?	No comparison between groups is done.

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).	Primary antibodies 1. monoclonal rat GCN 3D10 (Pichlo et al. 2014) 2. polyclonal rabbit ETK rb1 (produced by LifeTein, Somerset, NJ, USA) 3. monoclonal rat ApNHE 14E1 (produced by German Research Center for Environmental Health, Monoclonal Antibody Core Facility, Germany) 4. monoclonal rat AP47G4 (Bönigk et al. 2009) 5. monoclonal mouse a-tubulin B-5-1-2 (T5168, Sigma)
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	N.A.

* 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.	Wild type sea urchins from the species Arbacia punctulata and Strongylocentrotus purpuratus were used. Only sperm from males were used for these experiments. The Arbacia animals were provided by the Marine Resources Center located at the Marine Biological Laboratory in Woods Hole and the Strongylocentrotus animals were provided the Monterey Abalone Company, Monterey, CA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	N.A.
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	We confirm compliance to the guidelines

E- Human Subjects

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

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. 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	Provided
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).	Data provided in supplementary figures and tables
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).	N.A.
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.	N.A.

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.	N.A.