EMBO Molecular Medicine

Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities

Jermaine Goveia, Andreas Pircher, Lena-Christin Conradi, Joanna Kalucka, Vincenzo Lagani, Mieke Dewerchin, Guy Eelen, Ralph J. DeBerardinis, Ian D. Wilson and Peter Carmeliet

Corresponding author: Peter Carmeliet, VIB and KU Leuven

Review timeline:

Submission date:	07 July 2016
Editorial Decision:	01 August 2016
Revision received:	08 August 2016
Accepted:	10 August 2016

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

Editor: Roberto Buccione

1st Editorial Decision 01 August 2016

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

You will see that all three Reviewers are quite positive and agree that your manuscript is relevant, interesting, useful and well written.

There are a few suggestions for improvement that I am sure you will have no problem dealing with. We would thus be pleased to consider a revised submission, incorporating the reviewers' suggestions. I will be making an editorial decision on your next, final version.

In the likely event of acceptance, you will be asked to fulfill a number of editorial requirements as listed below. I suggest that you provide the following information and amendments requested with the next, final version of your manuscript to accelerate the process:

1) As per our Author Guidelines, the description of all reported data that includes statistical testing must state the name of the statistical test used to generate error bars and P values, the number (n) of independent experiments underlying each data point (not replicate measures of one sample), and the actual P value for each test (not merely 'significant' or 'P < 0.05'). Yopu may provide the P values as a separate table.

2) Every published paper now includes a 'Synopsis' to further enhance discoverability. Synopses are displayed on the journal webpage and are freely accessible to all readers. They include a short

standfirst as well as 2-5 one sentence bullet points that summarise the paper. Please provide the synopsis including the short list of bullet points that summarise the key NEW findings. The bullet points should be designed to be complementary to the abstract - i.e. not repeat the same text. We encourage inclusion of key acronyms and quantitative information. Please use the passive voice. Please attach this information in a separate file or send them by email, we will incorporate it accordingly. You are also welcome to suggest a striking image or visual abstract to illustrate your article. If you do please provide a jpeg file 550 px-wide x 400-px high.

3) Please note that we now mandate that all corresponding authors list an ORCID digital identifier. You may do so though our web platform upon submission and the procedure takes <90 seconds to complete. We also encourage co-authors to supply an ORCID identifier, which will be linked to their name for unambiguous name identification.

I look forward to reading a new revised version of your manuscript as soon as possible.

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

Referee #1 (Remarks):

Summary
The current manuscript by Goveia et al. provides a meta-analysis of clinical metabolic profiling (CMP) studies in tumor diseases and diabetes. In this respect, the authors compile CMP studies published between 2010 and 2015. They report that the vast majority of all published CMP studies only report on a subset of measured metabolites and also largely lack appropriate meta-data on patient tumor staging etc. Also, most CMP studies rely on a cross-sectional design, thereby not exploring a longitudinal change in metabolite levels during the course of the disease. In addition, the author's meta-analysis demonstrated that most data remain unconfirmed by independent experimental settings. Due to these limitations, the authors finally employed a semi-quantitative meta-analysis by vote-counting. These analyses demonstrated that across all included CMP studies a number of well-established tumor-associated metabolites, e.g. lactic acid and glutamic acid, could be confirmed to be enriched in tumor tissue. In addition, 3-hydroxbutyric acid could be identified a potential novel tumor marker in cancer patients. Overall, the authors conclude the there is a critical need for standardization across future CMP studies.

General Comments
Given the increasing recognition of tumor cell metabolism as a key feature of the malignant phenotype, the identification of tumor-associated metabolites and their potential role as biomarkers and/or therapeutic targets represents an important topic in oncology. In this regard, Goveia et al. provide an interesting and meaningful overview over the validity and usefulness of previously published CMP studies in the field. The manuscript is concise, clearly structured, and comes to clear statements regarding the potential impact of current CMP studies on clinical improvements in tumor diagnostics and therapies. Despite the fact that the chosen approach hardly allows for the discovery of novel metabolite pathways in cancer, the current manuscript may receive broad attention throughout the cancer metabolism community by raising awareness of the weaknesses and limitations of current clinical/experimental approaches. In this respect, the manuscript may serve as an "eye opener" for the cancer metabolite community to increase efforts in data harmonization and reproducibility.

Referee #2 (Remarks):

The manuscript represents an unorthodox and incisive effort to use metabolomic data for meta-analysis. A large portion of the work is a critique of the suitability of the published metabolomic literature for data mining. The authors cognetly discuss the limitations of the published literature for this purpose, and make an important comparison to genomic and epigenomic literature. The authors then seek to get around these limitations using a "vote-counting" method. This methods allows the authors to identify metabolites that are conistently enriched or depeleted in either tumor tissue or

blood from cancer patients, compared to appropriate controls. With this method, the authors identify lactate and glutamate as enriched in tumors and glutamate and 3-hydroxybutyrate as enriched in the blood of cancer patients where tryptophan and glutamine are depleted. While these data are not highly novel, their finding demosntrates the potential of metabolomic meta-analysis, which will be realized more fully when critiques such as this are more widely appreciated.

Issues to be addressed
:
1.The statistical methods are sound, but they are reported only in the Supplement. If these methods could be reported in the main text, it would enhance the paper.

2. The finding of increased lactate is expected and may reasonably be considered to validate the methods. However, the widespread acceptance of the phenomenon of aerobic glycolysis in cancer (which is an acceptance based on strong evidence) may be a source of bias. Studies finding increased lactate may be more likely to be reported or more likely to be published. Studies finding no change in lactate may be less likely to reach the published literature. This source of bias is inherent in meta-analysis and must be discussed.

3. The authors should discuss the possibility that 3-hydroxybutyrate may be elevated in cancer patients due to cachexia.

4. The authors should consider discussing that metabolomic studies are performed with a broader array of technologies than other holistic, non-biased approaches such as transcriptomics. While transcriptomic studies typically rely on either microarray and RNA-seq, metabolomic studies may use a wider variety of analytic methods, complicating direct comparisons between studies.

Referee #3 (Remarks):

This paper by Goveia and coll. entitled "Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities" reports a data mining and semi-quantitative meta-analysis of metabolites comparing healthy and cancer or diabetic patients, to identify distinct metabolite signatures in different pathologies. This study provides the evaluation of the feasibility of this kind of approach and gives some recommendations to improve its clinical impact. This paper is well written and reports important conclusions of clinical importance. Therefore, this work deserves publication in EMM.

Minor issue:
Page 6 lines4-5: the sentence "Surprisingly, ... metabolites" is unclear. Please reformulate.

1st Revision - authors' response 08 August 2016

REFEREE #1

Summary

The current manuscript by Goveia et al. provides a meta-analysis of clinical metabolic profiling (CMP) studies in tumor diseases and diabetes. In this respect, the authors compile CMP studies published between 2010 and 2015. They report that the vast majority of all published CMP studies only report on a subset of measured metabolites and also largely lack appropriate meta-data on patient tumor staging etc. Also, most CMP studies rely on a cross-sectional design, thereby not exploring a longitudinal change in metabolite levels during the course of the disease. In addition, the author's meta-analysis demonstrated that most data remain unconfirmed by independent experimental settings. Due to these limitations, the authors finally employed a semi-quantitative meta-analysis by vote-counting. These analyses demonstrated that across all included CMP studies a number of well-established tumor-associated metabolites, e.g. lactic acid and glutamic acid, could

be confirmed to be enriched in tumor tissue. In addition, 3-hydroxbutyric acid could be identified a potential novel tumor marker in cancer patients. Overall, the authors conclude the there is a critical need for standardization across future CMP studies.

General Comments

Given the increasing recognition of tumor cell metabolism as a key feature of the malignant phenotype, the identification of tumor-associated metabolites and their potential role as biomarkers and/or therapeutic targets represents an important topic in oncology. In this regard, Goveia et al. provide an interesting and meaningful overview over the validity and usefulness of previously published CMP studies in the field. The manuscript is concise, clearly structured, and comes to clear statements regarding the potential impact of current CMP studies on clinical improvements in tumor diagnostics and therapies. Despite the fact that the chosen approach hardly allows for the discovery of novel metabolite pathways in cancer, the current manuscript may receive broad attention throughout the cancer metabolism community by raising awareness of the weaknesses and limitations of current clinical/experimental approaches. In this respect, the manuscript may serve as an "eye opener" for the cancer metabolite community to increase efforts in data harmonization and reproducibility.

GENERAL RESPONSE: We thank referee #1 for these thoughtful comments assessing our meta-analysis as a valuable contribution to the field of cancer metabolism.

REFEREE #2

The manuscript represents an unorthodox and incisive effort to use metabolomic data for meta-analysis. A large portion of the work is a critique of the suitability of the published metabolomic literature for data mining. The authors cognetly discuss the limitations of the published literature for this purpose, and make an important comparison to genomic and epigenomic literature. The authors then seek to get around these limitations using a "vote-counting" method. This methods allows the authors to identify metabolites that are conistently enriched or depeleted in either tumor tissue or blood from cancer patients, compared to appropriate controls. With this method, the authors identify lactate and glutamate as enriched in tumors and glutamate and 3-hydroxybutyrate as enriched in the blood of cancer patients where tryptophan and glutamine are depleted. While these data are not highly novel, their finding demosntrates the potential of metabolomic meta-analysis, which will be realized more fully when critiques such as this are more widely appreciated.

GENERAL RESPONSE: We thank referee #2 for these generally positive comments. We appreciate the comment to report the statistical methods in the main text, which we originally included in the supplement due to space limitations. We also adapted the discussion as suggested and as detailed below. All changes to the text are marked in red.

Issues to be addressed :

1. The statistical methods are sound, but they are reported only in the Supplement. If these methods could be reported in the main text, it would enhance the paper.

RESPONSE: The statistical methods are now presented in the main text (materials and methods section) (not marked in red).

2. The finding of increased lactate is expected and may reasonably be considered to validate the methods. However, the widespread acceptance of the phenomenon of aerobic glycolysis in cancer (which is an acceptance based on strong evidence) may be a source of bias. Studies finding increased lactate may be more likely to be reported or more likely to be published. Studies finding no change in lactate may be less likely to reach the published literature. This source of bias is inherent in meta-analysis and must be discussed.

RESPONSE: As requested, we now discuss such bias as a potential limitation of our study, but also suggest that this may be partially addressed by full data deposition to public repositories.

3. The authors should discuss the possibility that 3-hydroxybutyrate may be elevated in cancer

patients due to cachexia.

RESPONSE: As requested, the possibility that elevated levels of 3-hydroxybutyrate might be caused by tumor cachexia in affected patients is now discussed in the revised text.

4. The authors should consider discussing that metabolomic studies are performed with a broader array of technologies than other holistic, non-biased approaches such as transcriptomics. While transcriptomic studies typically rely on either microarray and RNA-seq, metabolomic studies may use a wider variety of analytic methods, complicating direct comparisons between studies.

RESPONSE: As requested, we now highlight in the discussion that metabolomics studies are indeed performed with a broad array of technologies, a fact that certainly represents a challenge for inter-study comparisons.

REFEREE #3

This paper by Goveia and coll. entitled "Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities" reports a data mining and semi-quantitative meta-analysis of metabolites comparing healthy and cancer or diabetic patients, to identify distinct metabolite signatures in different pathologies. This study provides the evaluation of the feasibility of this kind of approach and gives some recommendations to improve its clinical impact. This paper is well written and reports important conclusions of clinical importance. Therefore, this work deserves publication in EMM.

GENERAL RESPONSE: We appreciate referee #3's comments and assessment of the meta-analysis in terms of clinical importance.

Minor issue:

Page 6 lines4-5: the sentence "Surprisingly, ... metabolites" is unclear. Please reformulate.

RESPONSE: To increase clarity, an additional sentence was included and the sentence referred to was rephrased to read together: "Current metabolic profiling technologies are capable of measuring tens to hundreds of metabolites. However, surprisingly, most individual studies published only a very small subset of all earlier reported metabolites".

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: Peter Carmeliet
Journal Submitted to: EMBO Molecular Medicine
Manuscript Number: EMM-2016-06798

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

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ 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.

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

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

B- Statistics and general methods

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

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	NA
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	The inclusion criteria were pre-established and explicitly reported in table EV7 and in the material and methods section (p15)
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	As described in the materials and methods section (p16), we used a pre-defined data extraction sheet and thereafter used R-scripts to analyze the data without any further human interference.
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
5. For every figure, are statistical tests justified as appropriate?	Yes, and we now moved full description of our statistical analyses from the supplements to the results section of the main text as suggested by the reviewers.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	We use have used parametric (binomial test) and non-parametric (permutation-based) tests to assess statistical significance. The difference between the two procedures was neglectable, thus confirming that the data meet the distributional assumptions of the parametric test.
Is there an estimate of variation within each group of data?	Within-group analysis of variance is not applicable in vote-counting
Is the variance similar between the groups that are being statistically compared?	Within-group analysis of variance is not applicable in vote-counting

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

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

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

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

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	NA
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	We provide full datasets for results of vote-counting (tables EV3-6)
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	We have included these data in table EV8-11
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA