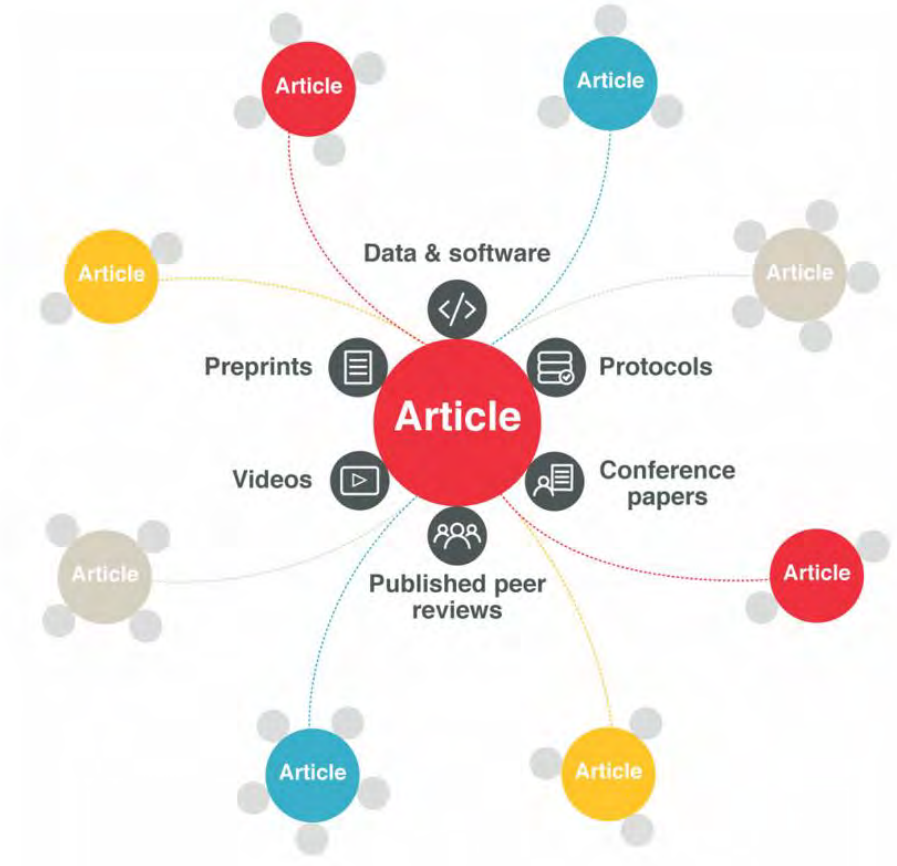
Data Discovery in PMC (and PubMed)

AMIA 2017 | November 7, 2017

Kathryn Funk
Kathryn.funk@nih.gov
Program Specialist | PMC
NCBI | National Library of Medicine

Journal Article as Nexus



PubMed

contains citations, abstracts, and associated metadata

PubMed Central (PMC)

contains full text records of publisher's versions and author manuscripts

https://www.crossref.org/blog/the-article-nexus-linking-publications-to-associated-research-outputs/

Journal Article as Data

PMC is an XML repository…

"the value of PMC lies in its capacity to store and cross-reference data from diverse sources using a common format within a single repository"

https://www.ncbi.nlm.nih.gov/pmc/about/intro/

PMC as Dataset(s)

Collections available for bulk download and text mining

PMC Text Mining Collections

Journal List › HHS Author Manuscripts › PMC5568754

HHS Public Access
Author manuscript
Peer-reviewed and accepted for publication
About author manuscripts | Submit a manuscript

Hum Pathol (N Y). Author manuscript; available in PMC 2017 Nov 1.
Published in final edited form as:
Hum Pathol (N Y). 2017 Nov; 10: 64–68.
doi: 10.1016/j.ehpc.2017.06.007

PMCID: PMC5568754
NIHMSID: NIHMS896153

EBV driven natural killer cell disease of the central nervous system presenting as subacute cognitive decline

Francesca M. Brett,[a,*] Richard Flavin,[b] Daphne Chen,[a] Teresa Loftus,[a] Gascun,[e] Elaine S. Jaffe,[f] Nurul Nor,[g] Mohsen Javadpour,[a] and Dominic

Author information ▶ Copyright and License information ▶

Author Manuscript

Abstract

Brain biopsy in patients presenting with subacute encephalopathy undertaken when a 'treatable condition' is a realistic possibility.

This 63 year old right handed, immunocompetent Caucasian wor rapidly progressive right-sided hearing impairment, a 4 month hi 'dizziness', dysphagia, nausea and vomiting, with the subsequent subacute global encephalopathy. The possibility of CJD was raise

Journal List › PLoS One › v.12(10); 2017 › PMC5646867

PLOS | ONE A Peer-Reviewed, Open Access Journal
View this Article | Submit to PLOS | Get E-Mail Alerts | Contact Us

PLoS One. 2017; 12(10): e0186696.
Published online 2017 Oct 18. doi: 10.1371/journal.pone.0186696

PMCID: PMC5646867

Semi-field assessment of the BG-Malaria trap for monitoring the African malaria vector, *Anopheles arabiensis*

Elis P. A. Batista, Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing,[1,2,*] Halfan S. Ngowo, Data curation, Formal analysis,[2] Mercy Opiyo, Formal analysis,[2] Gasper K. Shubis, Investigation,[2] Felician C. Meza, Investigation,[2] Fredros O. Okumu, Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing,[2,3,4] and Alvaro E. Eiras, Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing[1]

Clive Shiff, Editor

Author information ▶ Article notes ▶ Copyright and License information ▶

PLoS One

Abstract Go to:

Odour-baited technologies are increasingly considered for effective monitoring of mosquito populations and for the evaluation of vector control interventions. The BG-Malaria trap (BGM), which is an upside-down variant of the widely used BG-Sentinel trap (BGS), has been demonstrated to be effective to

Journal List › Med Facts Obs › v.1; 1791 › PMC5072801

NIH U.S. National Library of Medicine | Medical Facts and Observations | Biomedical Journal Digitization

Med Facts Obs. 1791; 1: 40–42.

PMCID: PMC5072801

An Account of an Uncommon Inflammation of the Epiglottis

Communicated in a Letter to Everard Home, *Esq. F.R.S., and by Him to Dr.* Simmons

Mr. Thomas Mainwaring

Author information ▶ Copyright and License information ▶

Med Fac

Full text

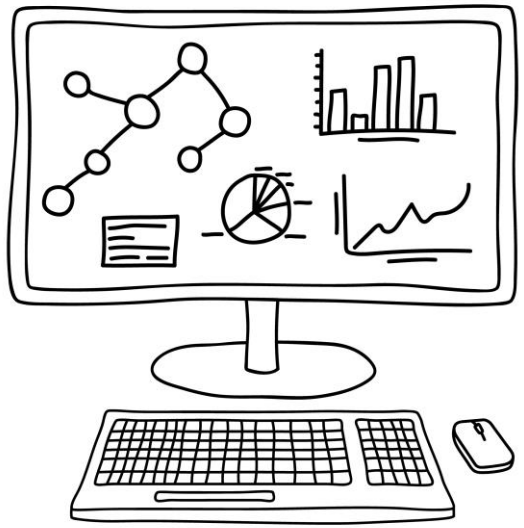
ll text is available as a scanned copy of the original print version. Get a printable copy (PDF file) of the omplete article (773K), or click on a page image below to browse page by page.

40 41 42

https://www.ncbi.nlm.nih.gov/pmc/tools/textmining/

PMC Fact #1: There is a spectrum of openness in PMC

More than 4.5 million articles in PMC are free to read…



… but not all are free of traditional copyright restrictions that restrict reuse

Open Access Subset

1.75 million articles

(or similar license)

ACCESSIBLE VIA

https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/

PMC Fact #2: PMC includes historical journal content

[40]

larly happy for mankind that the virus of rabid animals requires ſo much time to vegetate, (if I may be allowed the expreſſion) as it admits of preventives to be made uſe of, and none, as I have already remarked, ſeems ſo likely to ſucceed as exciſion.

Canterbury,
January 24, 1791.

III. *An Account of an uncommon Inflammation of the Epiglottis. By Mr.* Thomas Mainwaring, *Apothecary in London. Communicated in a Letter to* Everard Home, *Eſq. F. R. S., and by him to Dr.* Simmons.

A Gentleman, about forty years of age, who had been expoſed to the influence of cold, on the 17th of December, 1790, was attacked, in the night, with a violent pain in his throat, and a total inability to ſwallow.

In the morning of the 18th the ſymptoms were a good deal increaſed. The pain was not in the ſituation uſual in ſimilar affections, but lower down, and felt more anteriorly. When he attempted to ſwallow fluids, they paſſed readily

A SPEECH DEFECT CASE TREATED AT COLUMBIA UNIVERSITY.

By Alice C. Hinckley.

Thomas T. is the nine-year-old son of Irish parents. His father was a confirmed alcoholic for many years before his death from pneumonia last fall. His mother is a robust, energetic woman of emotional temperament, who supports her family by doing washing. During the summer she is caretaker of a residence on East Sixty-seventh Street, where living conditions are much more favorable than in the crowded section where the family spend

ON THE TREATMENT OF STRABISMUS BY THE DIVISION OF THE MUSCLES OF THE EYE.

BY P. BENNETT LUCAS,

Lecturer on Surgery at the Hunterian School of Medicine, and Surgeon to the Metropolitan Free Hospital, &c.

The time which has elapsed since I introduced into this country the operation for convergent strabismus, by dividing the tendon of the internal rectus muscle, and the numerous cases since that operation which have fallen under my notice, warrant me in affirming that this unsightly deformity is, in the majority of instances, now under the absolute control of surgical skill.

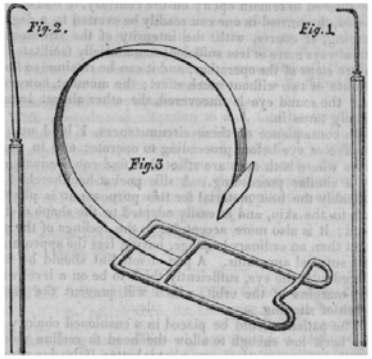
Before strabismus was attempted to be cured by the operation of dividing the tendon of the muscle which gave the eye the wrong direction, a variety of methods of treatment was suggested and put into execution. By some of these the strabismus was occasionally cured; by others it was only alleviated; but the failures which attended the best-directed intentions of the practitioner were, beyond all comparison, the most numerous.

Notwithstanding that the new operation has been attended with the happiest results in those cases to which it is applicable, yet it would be injudicious to assert that it ought to supersede all other methods of treatment,—nay, in some cases the operation will be unattended with benefit; and, in others, the strabismus can be cured by milder means.

I have, in a treatise lately published,* endeavoured to point out those cases of strabismus in which the operation

tion can readily be performed, provided there is no complication of instruments, no busy fingers of many assistants, and no haste to do the thing quickly, as must always be the case when ten or a dozen patients are collected together, and are cut against time!

The only instruments I employ in this operation are, a fine sharp hook, a blunt hook, and a pair of scissors.



MENTAL WELFARE. 47
Book Reviews & Abstracts.
Intelligence Testing : Its Methods and Results. By Rudolf Pintner, Ph.D. University of London Press, Ltd. 406 pp. 7s. 6d. nett.
We may perhaps be sceptical regarding the flood of literature on this subject proceeding from across the Atlantic, but the fact that the author, now Professor of Education in the Teachers' College, Columbia University, U.S.A., was formerly a distinguished student of one of our own Universities, may win for this book a somewhat readier welcome. It will certainly be deserved, for at the present state in the development of the work of mental testing we could not wish for a more suitable volume to put into the hands of those of the general public who wish to* be informed on tue subject. Students in colleges, as well as medical men and teachers who have no special training in this direction, would do well to begin their study by a careful reading of these pages.
Professor Pintner deals in a comprehensive way with the history and theory of mental testing and the general methods employed, concluding with a survey of the results of testing with reference to such special groups

Historical OCR Collection

Supported by wellcome trust

NIH U.S. National Library of Medicine

+370,000 articles

Spanning over 200 years
1791-1996

40 journals

license terms vary by title,
ranging from fair use to CC-BY

https://www.ncbi.nlm.nih.gov/pmc/about/scanning/

ACCESSIBLE IN .TXT FORMAT VIA

FTP

ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/historical_ocr/

PMC Fact #3: PMC is a repository of funded research

Funders added to
http://www.ncbi.nlm.nih.gov/pmc/about/public-access/
as policies are implemented

Author Manuscript Collection

- More than 450,000 author manuscripts collected under NIH and other funder public access policies
- Available for text mining (may also be used consistent with the principles of applicable copyright law)
- Accessible in .txt and XML formats via FTP

https://www.ncbi.nlm.nih.gov/pmc/about/mscollection/
ftp://ftp.ncbi.nlm.nih.gov/pub/pmc/manuscript/

"We developed this resource to increase the impact of NIH funding. Through this collection, scientists will be able to analyze these manuscripts, further apply the findings of NIH research, and generate new discoveries."

- NIH, November 16, 2015

How are these datasets being used?

Datasets in PMC

Discovering datasets and linked data in PMC

How do datasets get in to PMC?

Supplementary Material

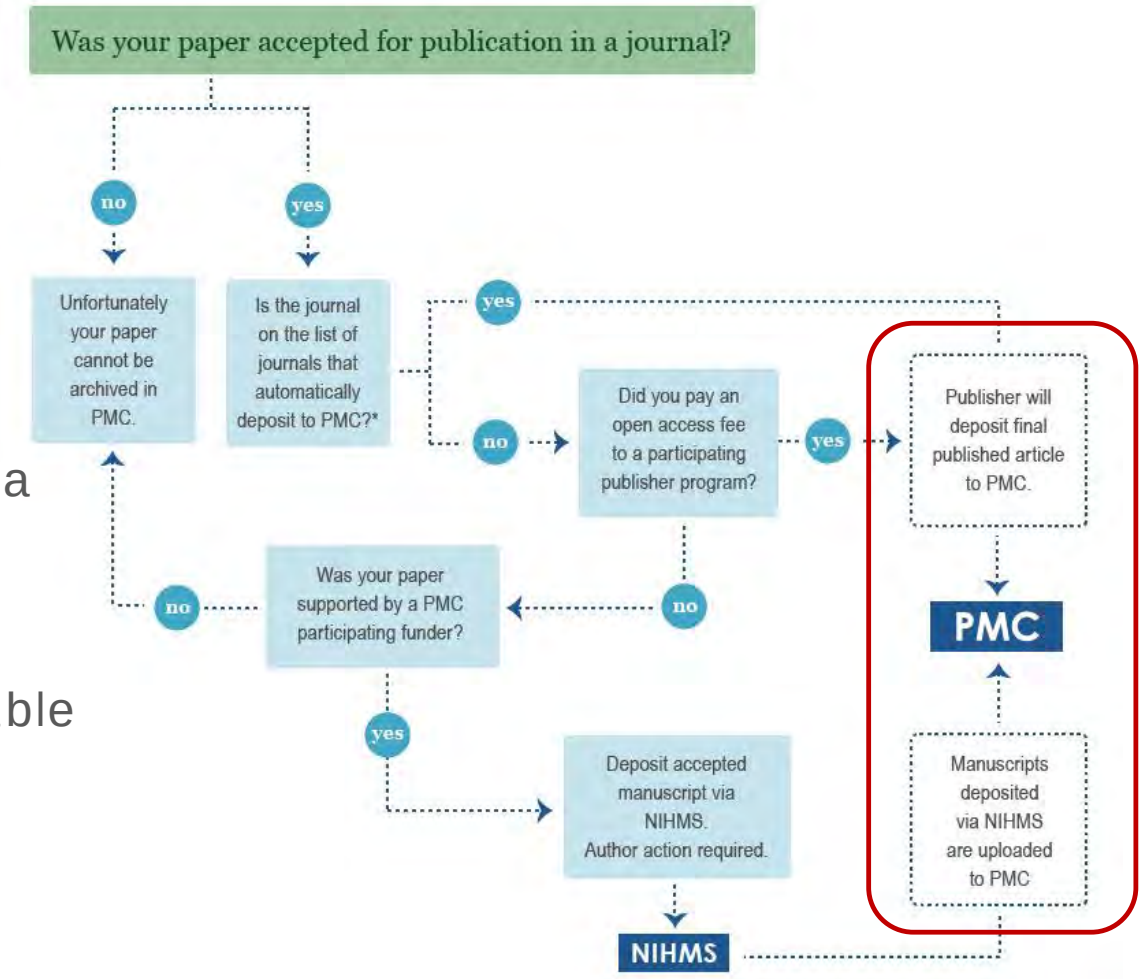
- PMC policy requires all referenced supplementary data files to be deposited with the article or manuscript.

Data Citations

- Authors are also encouraged to make datasets available in a public repository and include the relevant data citation(s) in the paper.

NIHMS Deposit

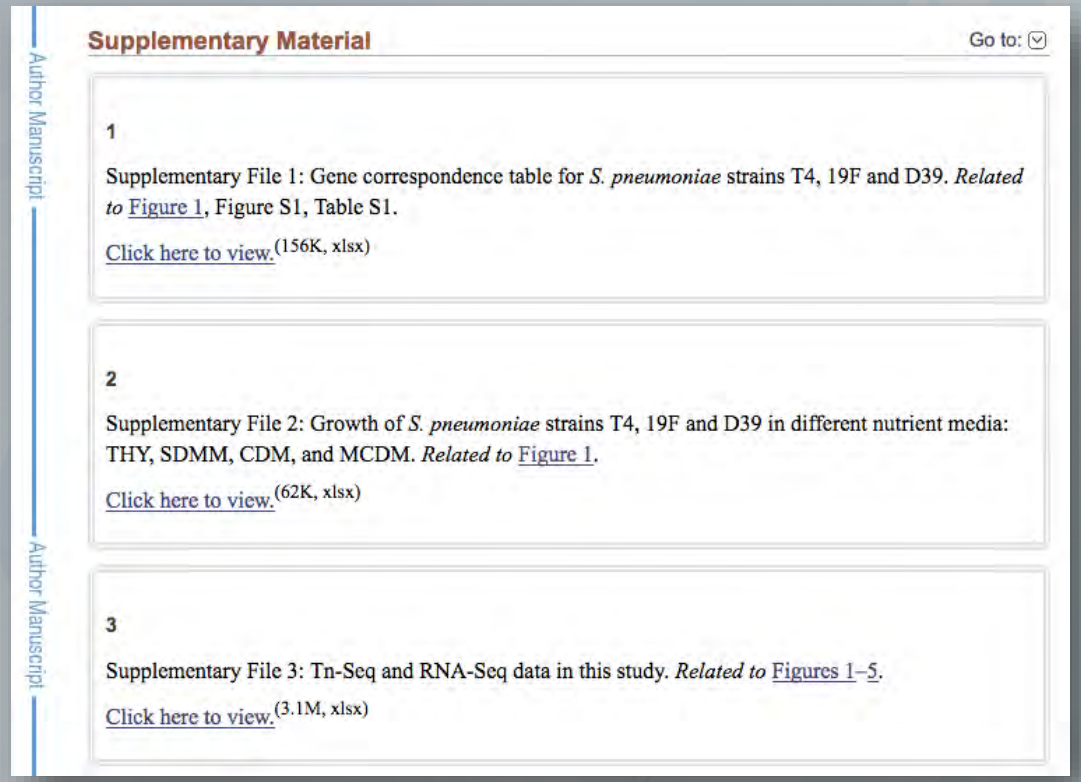
- Funded authors who are required to deposit the author manuscript to PMC can also supply data files at the time of deposit.



https://www.ncbi.nlm.nih.gov/pmc/about/submission-methods/
https://www.ncbi.nlm.nih.gov/pmc/about/guidelines/#supplementary

Supplementary Data in PMC

Supplementary Material Section



Supplementary Material

Go to:

1

Supplementary File 1: Gene correspondence table for *S. pneumoniae* strains T4, 19F and D39. *Related to* Figure 1, Figure S1, Table S1.

Click here to view.(156K, xlsx)

2

Supplementary File 2: Growth of *S. pneumoniae* strains T4, 19F and D39 in different nutrient media: THY, SDMM, CDM, and MCDM. *Related to* Figure 1.

Click here to view.(62K, xlsx)

3

Supplementary File 3: Tn-Seq and RNA-Seq data in this study. *Related to* Figures 1–5.

Click here to view.(3.1M, xlsx)

Author Manuscript

Author Manuscript

PMC Search: has suppdata[filter]

Examples of what you will find in this section:

- Computer code
- Mathematical or computational models
- Audio or video clips
- Oversized tables
- Intervention protocols
- Primary or supplementary data sets
- Expanded methodology sections
- Additional figures

Source: Publication Manual of APA (6th ed.)

Dataset Deposit in NIHMS

- Data files should be supplied with other manuscript files in Step 1 of NIHMS process
- Author must have funding from a PMC-participating funder to deposit
- NIHMS can accept nearly any file type as supplemental; .doc, .pdf, .xls, .csv, .tiff, .jpg, .eps, .ai, mp4 are the most common types received

https://www.nihms.nih.gov/

Dataset Deposit in NIHMS

- Upload as Supplemental file type
- Provide general dataset label (e.g., Dataset 1, Data 2)
- Can provide as many data files as necessary
- Data file(s) must be 2 GB or less

U.S. Department of Health & Human Services

About ▾ Help ▾ Manuscript List Kathryn Funk Sign off

NIH Manuscript Submission System

Test Submission for Dataset Demonstration

Journal: Nature NIHMSID 16339 Provide citation

1. Title Information | 2. Add Funding | 3. Upload Files | 4. Check Files | 5. Set Reviewer & Embargo

Please upload all files associated with your manuscript, including supplemental files, and move to the next step.

Type	Label	File	Size	Date	Remove
Manuscript		Data notes_2017.docx	154.7	2017-09-27 10:01:45	X
Supplemental	Dataset 1	Browse... Data avail statement sample-2017jul.xlsx			X
Supplemental	Dataset 2	Browse... NIH-funded supp data sample-2016jul.xlsx			X

Add another Manuscript, Figure, Table, Supplemental

Choosing a file type
Labeling files
Adding files
Uploading datasets (optional)

All referenced supplementary files are required to be deposited with the manuscript. However, authors also have the option to upload additional supporting data files (under 2 GB) with a file type of "Supplemental" to be made accessible with the manuscript in PMC. These files should be given a descriptive label (e.g., Dataset 1).

Save & Exit | Cancel Submission | ◄ Add Funding | Check Files ►

https://www.nihms.nih.gov/

Data Citations and Statements in PMC

Data Availability Statement

Data Availability Go to:

All relevant data are within the paper and its Supporting Information files. Protein structure files were obtained from Protein Data Bank. Protein structure accession and site mapping are listed in S1 Table (https://doi.org/10.6084/m9.figshare.4910141.v2).

Data Citations

Data Citations Go to:

1. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e6edfc522cf
2. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e47c1be64b8
3. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e47c33f2500
4. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e580cfe8c95
5. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e47be626ee5
6. Ibanez J. M.. Australian Antarctic Data Centre. 2017 https://doi.org/10.4225/15/58e47c03e652a

Examples of what you will find in these sections:

- "Data available upon request"
- References to supplementary material
- NCBI database accession numbers (GEO, GenBank, dbGap, SRA)
- General data repository record DOIs (figshare, Dryad)
- Government data repository links
- Personal servers

Data Citations in PMC

S Gavrilov, RP Harvey & VE Papaioannou. Mus musculus T-box 2 (Tbx2), mRNA. NCBI Nucleotide. 2014:NM_009324.2. NM_009324

```
<!--Data reference: NCBI Nucleotide: pub-id-type="accession" assigning-authority="NCBI"
contains version details-->
<ref id="bib13">
  <element-citation publication-type="data">
    <person-group person-group-type="author">
      <name>
        <surname>Gavrilov</surname>
        <given-names>S</given-names>
      </name>
      <name>
        <surname>Harvey</surname>
        <given-names>RP</given-names>
      </name>
      <name>
        <surname>Papaioannou</surname>
        <given-names>VE</given-names>
      </name>
    </person-group>
    <year iso-8601-date="2014">2014</year>
    <data-title>Mus musculus T-box 2 (Tbx2), mRNA</data-title>
    <source>NCBI Nucleotide</source>
    <pub-id pub-id-type="accession" assigning-authority="NCBI"
      xlink:href="http://www.ncbi.nlm.nih.gov/nuccore/120407038"
      >NM_009324</pub-id>
    <version designator="NM_009324.2">NM_009324.2</version>
  </element-citation>
</ref>
```

"Data citations should facilitate access to the data and any associated metadata, code, or related materials."

https://www.ncbi.nlm.nih.gov/pmc/pmcdoc/tagging-guidelines/article/dobs.html#dob-datacitations

http://jats4r.org/data-citations

Linked Data in PubMed

Discovering datasets and linked data in PMC (& PubMed)

Linked Data in PubMed

Secondary Source ID field

Andrologia. 2017 Oct;49(8). doi: 10.1111/and.12712. Epub 2016 Oct 26.

Characterisation of pomegranate juice effects on human corpus cavernosum.

Gur S[1,2], Rezk BM[3], Abd Elmageed ZY[2], Kadowitz PJ[2], Sikka SC[2], Hellstrom WJG[2].

Author information

Abstract

Pomegranate (POM) juice may benefit the erectile process, but the scientific evidence is lacking. This study evaluates the molecular characterisation and confirmation of POM's action on human corpus cavernosum (HCC) obtained from patients (n = 16) undergoing penile prosthesis implantation. After phenylephrine contraction, the relaxant effects of POM with various inhibitors in the presence and absence of palmitic acid (PA)-induced acute oxidative stress were investigated. Electrical field stimulation (EFS)- and acetylcholine (ACh)-induced relaxation were performed using organ bath preparation. Expression of neuronal nitric oxide synthase (nNOS), endothelial (eNOS), phosphodiesterase (PDE)-5A and cGMP levels were assessed in cells from ex vivo organ cultures of HCC, using RT-PCR, ELISA and immunohistochemistry techniques. POM induced marked relaxation of HCC (maximum response: 97.0 ± 3.1%) and reversed the PA-induced decrease of EFS (20 Hz). nNOS transcription was increased by 7-fold in POM-treated cells without influencing eNOS and PDE5A expressions. We conclude that POM induced marked relaxation of HCC via: (i) nNOS stimulation, and (ii) downstream relaxation stimulated by nNOS and cGMP and bypassing the NO and PDE5. This action provides a rationale for the therapeutic or preventative use of POM in men with erectile dysfunction who do not respond well to PDE5 inhibitors.

© 2016 Blackwell Verlag GmbH.

KEYWORDS: eNOS and PDE5A; erectile dysfunction; human corpus cavernosum; nitric oxide; pomegranate

PMID: 27785815 DOI: 10.1111/and.12712

Secondary source ID

Secondary source ID
GENBANK/NM_000620
GENBANK/NM_000603
GENBANK/NM_033437
GENBANK/NM_001101.2

Linkout: Other Literature Sources

Ann Med. 2017 Dec;49(8):718-727. doi: 10.1080/07853890.2017.1372622. Epub 2017 Sep 6.

Baseline fatty acids, food groups, a diet score and 50-year all-cause mortality rates. An ecological analysis of the Seven Countries Study.

Menotti A[1], Kromhout D[2,3,4], Puddu PE[5], Alberti-Fidanza A[6], Hollman P[3], Kafatos A[7], Tolonen H[8], Adachi H[9], Jacobs DR Jr[4].

Author information

Abstract

OBJECTIVES: This analysis deals with the ecologic relationships of dietary fatty acids, food groups and the Mediterranean Adequacy Index (MAI, derived from 15 food groups) with 50-year all-cause mortality rates in 16 cohorts of the Seven Countries Study.

MATERIAL AND METHODS: A dietary survey was conducted at baseline in cohorts subsamples including chemical analysis of food samples representing average consumptions. Ecologic correlations of dietary variables were computed across cohorts with 50-year all-cause mortality rates, where 97% of men had died.

RESULTS: There was a 12-year average age at death population difference between extreme cohorts. In the 1960s the average population intake of saturated (S) and trans (T) fatty acids and hard fats was high in the northern European cohorts while monounsaturated (M), polyunsaturated (P) fatty acids and vegetable oils were high in the Mediterranean areas and total fat was low in Japan. The 50-year all-cause mortality rates correlated (r= -0.51 to -0.64) ecologically inversely with the ratios M/S, (M + P)/(S + T) and vegetable foods and the ratio hard fats/vegetable oils. Adjustment for high socio-economic status strengthened (r= -0.62 to -0.77) these associations including MAI diet score.

CONCLUSION: The protective fatty acids and vegetable oils are indicators of the low risk traditional Mediterranean style diets. KEY MESSAGES We aimed at studying the ecologic relationships of dietary fatty acids, food groups and the Mediterranean Adequacy Index (MAI, derived from 15 food groups) with 50-year all-cause mortality rates in the Seven Countries Study. The 50-year all-cause mortality rates correlated (r = -0.51 to -0.64) ecologically inversely with the ratios M/S [monounsaturated (M) + polyunsaturated (P)]/[saturated (S) + trans (T)] fatty acids and vegetable foods and the ratio hard fats/vegetable oils. After adjustment for high socio-economic status, associations with the ratios strengthened (r = -0.62 to -0.77) including also the MAI diet score. The protective fatty acids and vegetable oils are indicators of the low risk traditional Mediterranean style diets.

KEYWORDS: 50-year follow-up; Seven Countries Study; all-cause mortality; fatty acids; socio-economic status

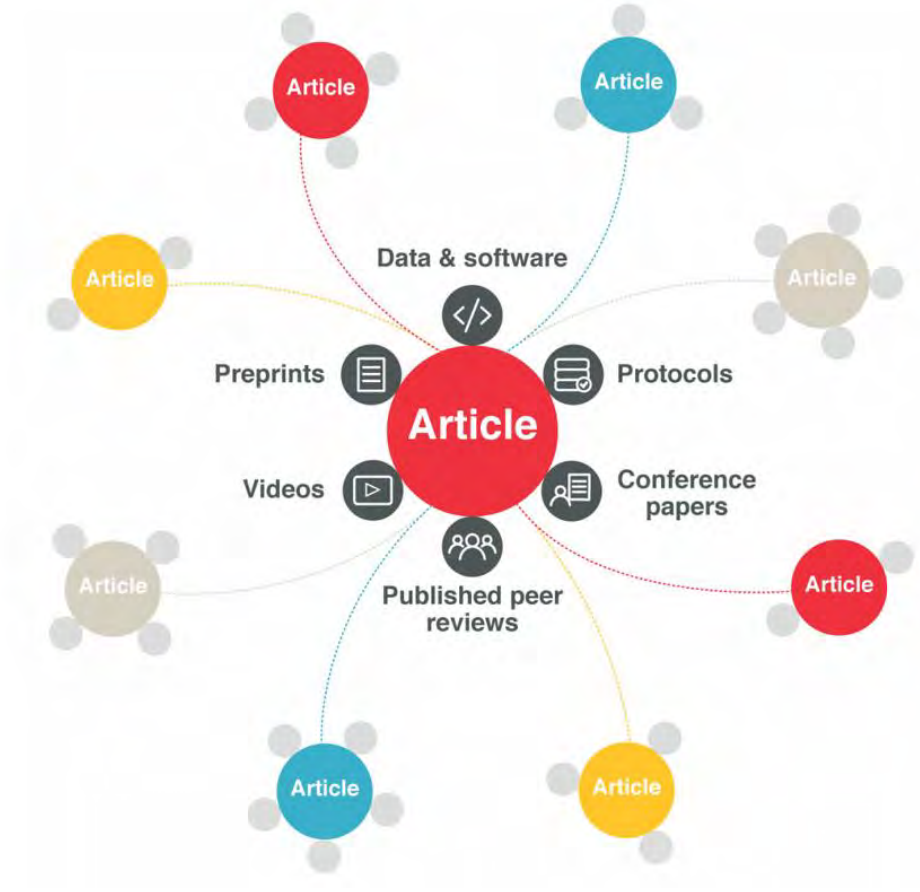
PMID: 28847158 DOI: 10.1080/07853890.2017.1372622

LinkOut - more resources

Full Text Sources
Taylor & Francis

Other Literature Sources
Supporting Data and Materials for the article - figshare

Example PubMed Searches:
hasdatabanklist
genbank[si] OR figshare[si]
loprovfigshare[Filter] OR loprovdryaddb[Filter]



What's next for the article as nexus (or dataset)?

- More central access?
- Easier to find?
- Increased interoperability?
- Encourage more text mining?

Move in a data- and user-driven direction!

https://www.crossref.org/blog/the-article-nexus-linking-publications-to-associated-research-outputs/

Suggestions? (or questions?)

Contact us with your ideas at pubmedcentral@ncbi.nlm.nih.gov

References to articles using the PMC Text Mining Collections

- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*, *6*, e27725. http://doi.org/10.7554/eLife.27725
- Grechkin, M., Poon, H., & Howe, B. (2017). Wide-Open: Accelerating public data release by automating detection of overdue datasets. *PLoS Biology*, *15*(6), e2002477. http://doi.org/10.1371/journal.pbio.2002477
- Goodwin, T. R., & Harabagiu, S. M. (2016). Medical Question Answering for Clinical Decision Support. *Proceedings of the ... ACM International Conference on Information & Knowledge Management. ACM International Conference on Information and Knowledge Management*, *2016*, 297–306. http://doi.org/10.1145/2983323.2983819
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biology*, *13*(3), e1002106. http://doi.org/10.1371/journal.pbio.1002106
- Flórez-Vargas, O., Brass, A., Karystianis, G., Bramhall, M., Stevens, R., Cruickshank, S., & Nenadic, G. (2016). Bias in the reporting of sex and age in biomedical research on mouse models. *eLife*, *5*, e13615. http://doi.org/10.7554/eLife.13615
- Eisen, M. (2016, Sept 28). Exploring the relationship between gender and author order and composition in NIH-funded research. It's not junk [blog]. http://www.michaeleisen.org/blog/?p=1931