GBIO0002 – Genetics and Bioinformatics

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

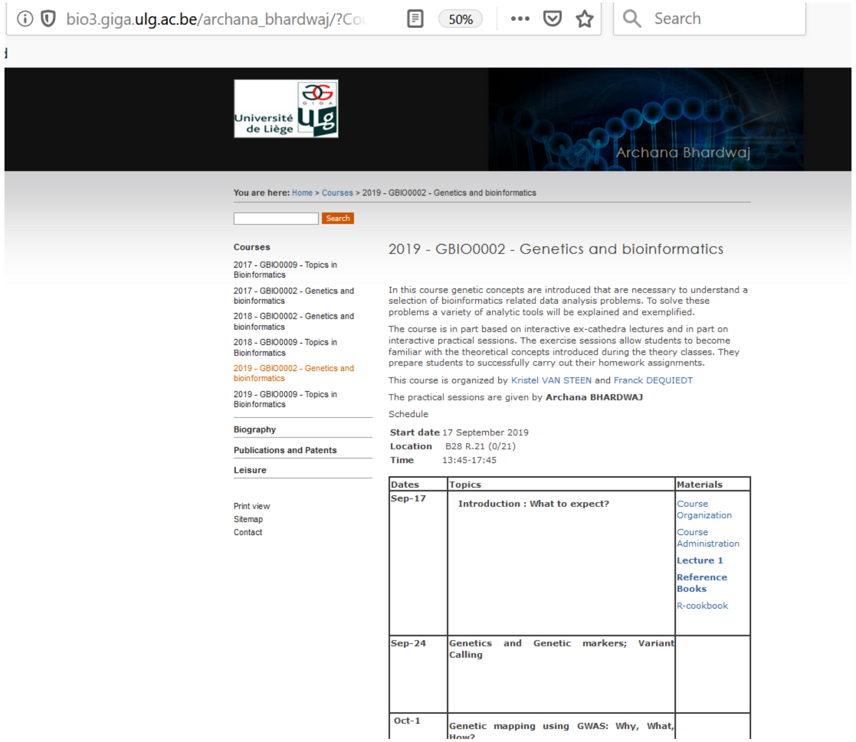
ULg

kristel.vansteen@ulg.ac.be

Administration

- Course website 2019-2020:

http://bio3.giga.ulg.ac.be/archana_bhardwaj/?Courses___2019_-_GBIO0002_-_Genetics_and_bioinformatics


bio3.giga.ulg.ac.be/archana_bhardwaj/?Co
50%
Search
Université de Liège
Archana Bhardwaj
Search
Courses
2017 - GBIO0009 - Topics in Bioinformatics
2017 - GBIO0002 - Genetics and bioinformatics
2018 - GBIO0002 - Genetics and bioinformatics
2018 - GBIO0009 - Topics in Bioinformatics
2019 - GBIO0002 - Genetics and bioinformatics
2019 - GBIO0009 - Topics in Bioinformatics
Biography
Publications and Patents
Leisure
Print view
Sitemap
Contact

You are here: Home > Courses > 2019 - GBIO0002 - Genetics and bioinformatics

2019 - GBIO0002 - Genetics and bioinformatics

In this course genetic concepts are introduced that are necessary to understand a selection of bioinformatics related data analysis problems. To solve these problems a variety of analytic tools will be explained and exemplified.

The course is in part based on interactive ex-cathedra lectures and in part on interactive practical sessions. The exercise sessions allow students to become familiar with the theoretical concepts introduced during the theory classes. They prepare students to successfully carry out their homework assignments.

This course is organized by Kristel VAN STEEN and Franck DEQUIEDT

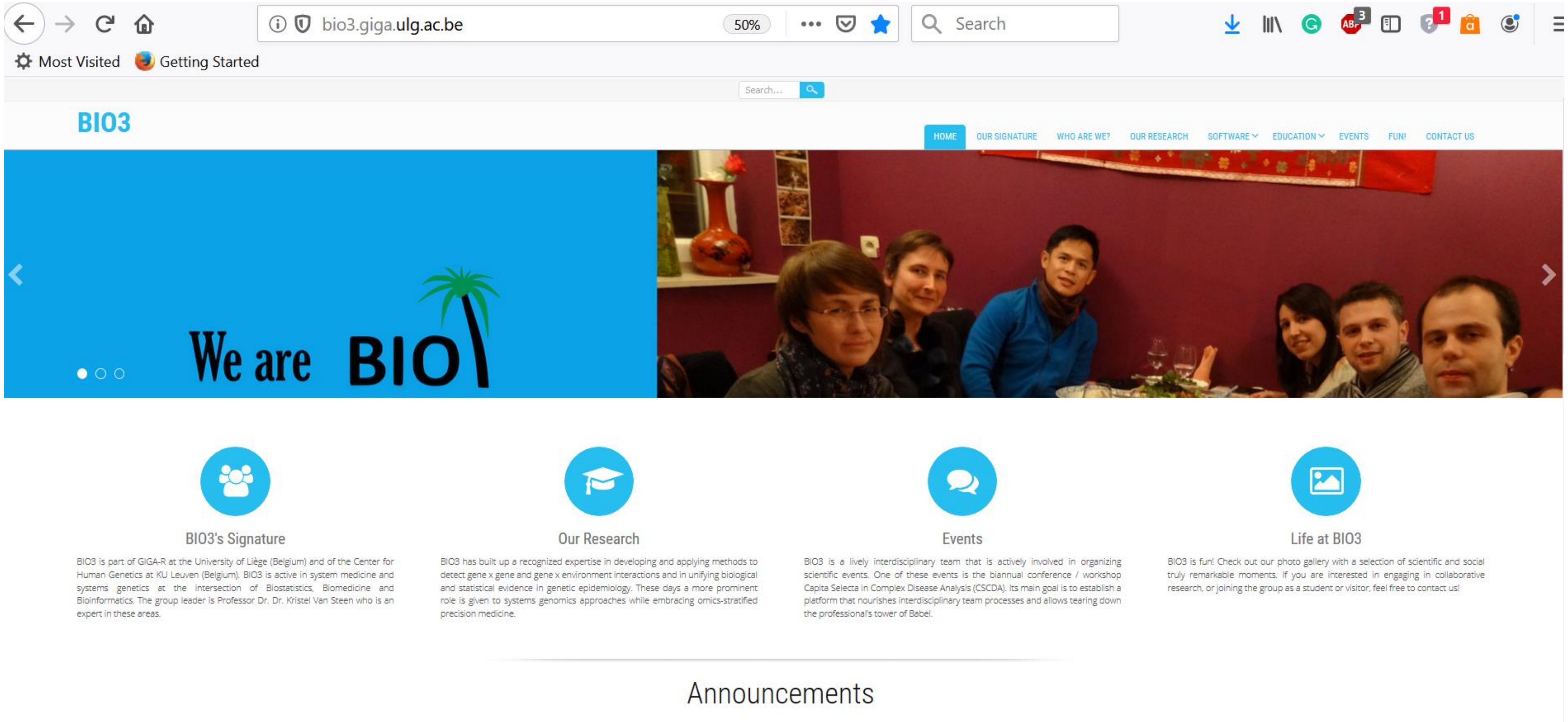
The practical sessions are given by **Archana BHARDWAJ**

Schedule

Start date 17 September 2019
Location B28 R.21 (0/21)
Time 13:45-17:45

Dates	Topics	Materials
Sep-17	Introduction : What to expect?	Course Organization Course Administration Lecture 1 Reference Books R-cookbook
Sep-24	Genetics and Genetic markers; Variant Calling	
Oct-1	Genetic mapping using GWAS: Why, What, How?	

Administration



BIO3's Signature

BIO3 is part of GIGA-R at the University of Liège (Belgium) and of the Center for Human Genetics at KU Leuven (Belgium). BIO3 is active in system medicine and systems genetics at the intersection of Biostatistics, Biomedicine and Bioinformatics. The group leader is Professor Dr. Dr. Kristel Van Steen who is an expert in these areas.

Our Research

BIO3 has built up a recognized expertise in developing and applying methods to detect gene x gene and gene x environment interactions and in unifying biological and statistical evidence in genetic epidemiology. These days a more prominent role is given to systems genomics approaches while embracing omics-stratified precision medicine.

Events

BIO3 is a lively interdisciplinary team that is actively involved in organizing scientific events. One of these events is the biannual conference / workshop Capita Selecta in Complex Disease Analysis (CSCDA). Its main goal is to establish a platform that nourishes interdisciplinary team processes and allows tearing down the professional's tower of Babel.

Life at BIO3

BIO3 is fun! Check out our photo gallery with a selection of scientific and social truly remarkable moments. If you are interested in engaging in collaborative research, or joining the group as a student or visitor, feel free to contact us!

Announcements

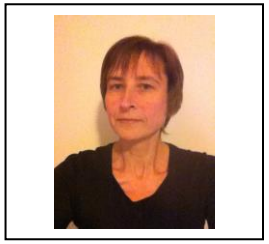
http://bio3.giga.ulg.ac.be/ [research BIO3]

Administration

- Course instructors

Prof. Kristel Van Steen

- Office: level +1, B34 (GIGA tower)
- E-mail: kristel.VanSteen@ulg.ac.be
- http://www.montefiore.ulg.ac.be/~kvansteen



Prof. Franck DEQUIEDT

- Office: level +5, B34 (GIGA tower)
- E-mail: fdequiedt@ulg.ac.be

Teaching Assistant

- Archana Bhardwaj
- Office: level +1, B34 (GIGA tower)
- A.Bhardwaj@uliege.be

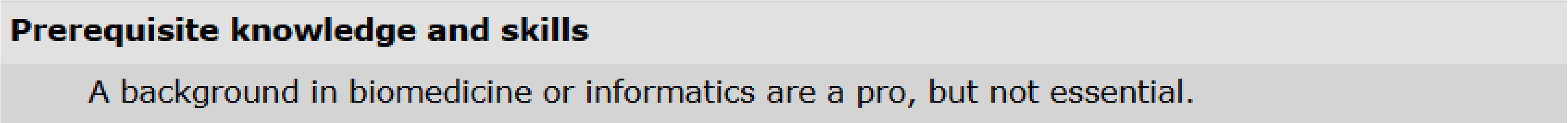
Administration

Complete online form:

https://www.dropbox.com/s/nx7zdxbpcs60r25/List%20of%20GBIO0002%20students%20with%20contact%20details.xlsx?dl=0

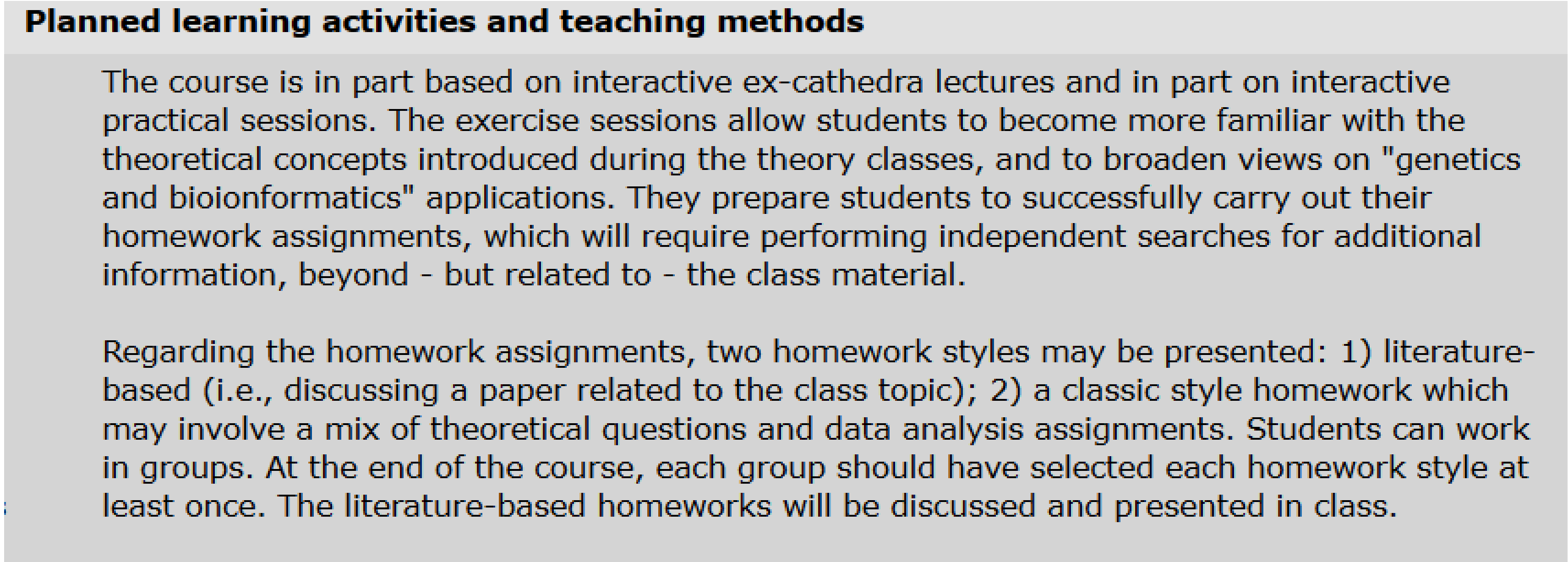
Administration

- Tutor-student commitments (progcours.ulg.ac.be)



Prerequisite knowledge and skills

A background in biomedicine or informatics are a pro, but not essential.



Planned learning activities and teaching methods

The course is in part based on interactive ex-cathedra lectures and in part on interactive practical sessions. The exercise sessions allow students to become more familiar with the theoretical concepts introduced during the theory classes, and to broaden views on "genetics and bioionformatics" applications. They prepare students to successfully carry out their homework assignments, which will require performing independent searches for additional information, beyond - but related to - the class material.

Regarding the homework assignments, two homework styles may be presented: 1) literature-based (i.e., discussing a paper related to the class topic); 2) a classic style homework which may involve a mix of theoretical questions and data analysis assignments. Students can work in groups. At the end of the course, each group should have selected each homework style at least once. The literature-based homeworks will be discussed and presented in class.

What will we be doing?

- General course content

In this course genetic concepts are introduced that are necessary to understand a selection of bioinformatics related data analysis problems. To solve these problems a variety of analytic tools will be explained and exemplified. Different topics typically include:

- The genome and genetic markers [genetics]
- Genome-wide association studies [analytics]
- Sequence technologies [genetics]
- Sequence comparisons [analytics]
- The transcriptome and proteome [genetics]
- Gene co-expression [analytics]

What will we be doing?

- General course content
  - Genetics + Analytics
  - Focus on
    - Understanding key concepts / terminology and their context
    - Interpreting findings / analysis results (NOT CARRYING OUT overly-complicated analyses)

How will we do it?

"Theory" classes

- The "theory" course will be interactive in English/French:
  - In class discussion papers (time permitting → computer!)
  - Interpreting analysis findings: discussing different viewpoints
  - Slides as supporting framework ("syllabus")
- Main instructors:
  - K Van Steen and
  - F Dequiedt

How will we do it?

“Practical” classes

- Application show-cases (computer!)
- “Homework assignments”: time-consuming part of this course and make links to the theory AND practical classes.
- Main tutor: Archana Bhardwaj
- Homeworks: 2 styles
  - Reading assignment with presentation and in-class discussions (graded)
  - Classic homework style (Questions / Answer) assignments (graded)
- Homework assignments result in a “group” slides/report and should be handed in electronically in English
- See also documentation on course website + next slide

Organization of GBIO0002 Homework Assignments
Genetics and Bioinformatics

…

Style 1: Literature project

This involves choosing a paper from the literature that extends or provides additional background on the material of the course (chapter) and then summarizing the paper, its objectives, results while further browsing the internet for additional information or supporting material.

Do not copy the paper, but show you have understood the main ideas of the paper and "discuss" the paper. Such a discussion could include thoughts on what was the key idea, strengths or weaknesses of the methods/experiments, comments on the writing, ways to extend the work, flaws in the argument/data/experiments, etc. Anything is fine, as long as it demonstrates some real thought. Especially for review papers, make sure one subtopic is worked out in more detail, by following up on referenced work or by searching the internet.

A selection of papers will be provided, but if you have another interesting paper to discuss, please send your suggestion to the TA. The course instructors will then decide whether the paper is eligible or not.

All literature projects will be presented and discussed in class. No report is needed. Only slides will do.

Style 2: Classic Q/A

Via representative questions, the idea is to further understand concepts provided in class. Occasionally, simulated or real-life data problems may be provided, that have been analyzed and for which the results require an interpretation. Use the material provided in class but be not afraid to consult the literature. As long as you can answer the given questions, everything is allowed. When you do use the literature, please provide references.

Please follow instructions in class, regarding how to draft your report.

General information regarding homework reports

Style 2 homework assignments may involve writing a short report of no more than the equivalent of 5 single-spaced typed pages of text, excluding figures, tables and bibliography. It should contain an abstract (e.g., depending on the homework style: description of the paper content, description of the problem) and a discussion part (see before). If citations are made to other papers, there should be a bibliography! Only one report per group is needed.

What will be evaluated?

- At the end of the course, you have acquired knowledge about **genetics** (in particular genomics, transcriptomics, technology-related aspects) and about a selection of state-of-the-art, yet basic, **analytic tools**.

- You will be evaluated about key concepts related to **genetics** and the analytic approaches presented during the course (incl. pros and cons, general contexts) and will be presented with a **few analysis results to interpret**.

How will be evaluated?

HW1		HW2		Written Exam	Participation
Genetics	Analytics	Genetics	Analytics		
15	15	15	15	35	5

- No final grade without homeworks; No final grade without exam; Homeworks not handed in in time == ZERO (electronic submission!)
- Written exam in January (terminology, basic analytic contexts, interpretation – see before; multiple choice / open questions; printed course notes as "open book")
- Second term exam: written exam + worst homework on Analytics + worst homework on Genetics

How will be evaluated?

Literature style homeworks

[homework = discuss a paper]

- Discuss the paper in your slides
- Make links
  - with other papers,
  - between the paper(s) and the course,
  - between the paper(s) and additional info outside the course

Evaluation criteria – presentation

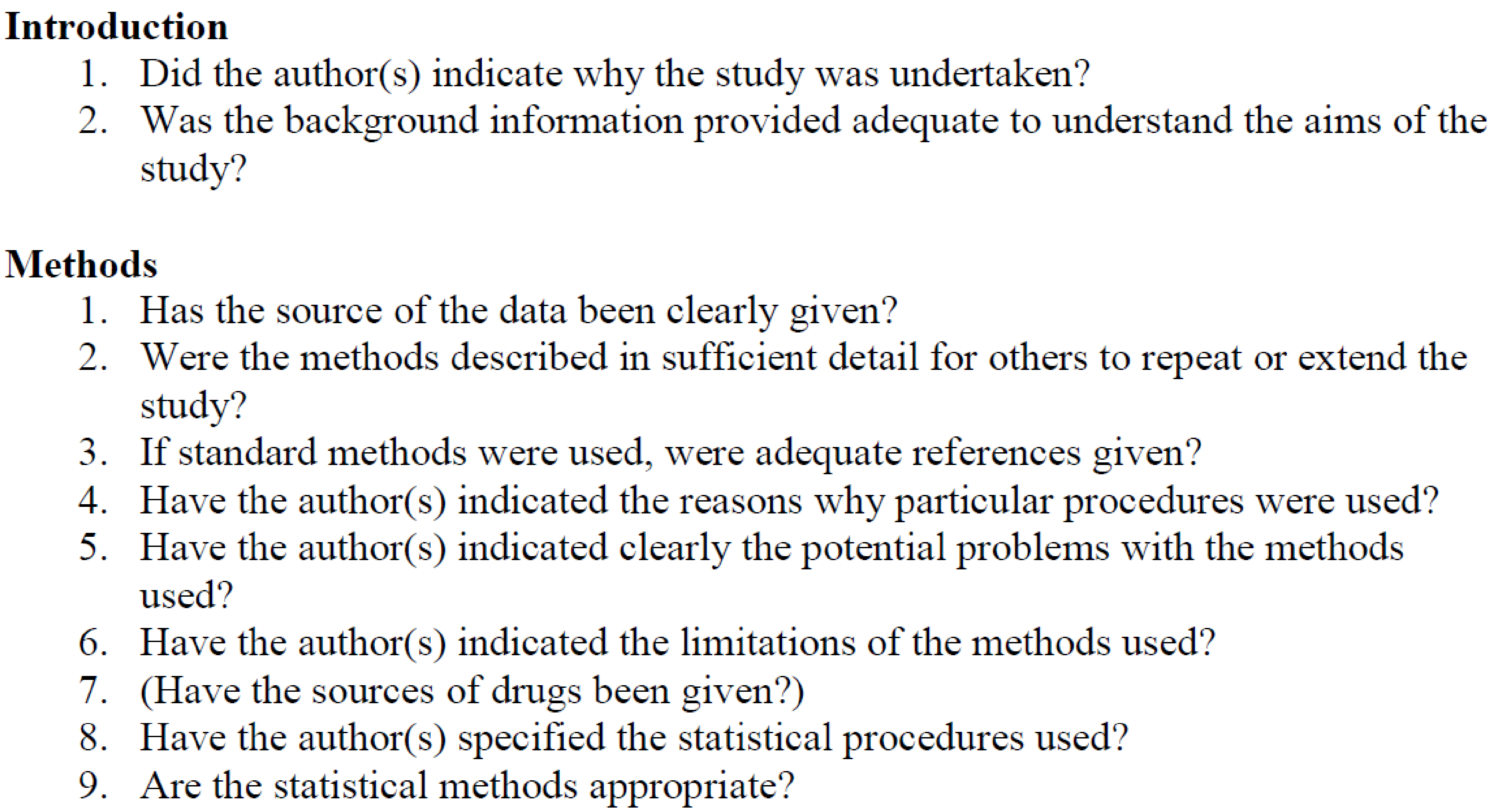
Criterium	Key words
Clarity	Concepts, slides content, slides composition, fellow students do not have questions regarding “new” statements (i.e., not covered in class) made on the slides or during the presentation
Illustrations on slide	Not too much; not only copy and paste from course but novel illustrations; supportive
Presentation Skills	Eager beaver (a person who is very enthusiastic about doing something)
Understanding	Presentation content as presented is understood: adequate reply to questions and comments (incl. those from fellow students)
Group dynamics	Scoring will be done on an individual basis; balanced partitioning of tasks

Evaluation criteria – report

Mainly refers to Q/A style of homeworks or in case of a second term exam one of the worst homeworks was a literature style homework.

- Ability to formulate the research problem and to sketch the context (introductions, data description, tool description, etc)
- Presentation summary of the analysis workflow (methods, analysis section)
- Discussion (of the analysis tools, of the quality of the analysis, validity of results – when put in a broader context, …)
- Creative input (stuffing, conclusion section)
- General structure of the report (sectioning)

Critical evaluation of a paper or report



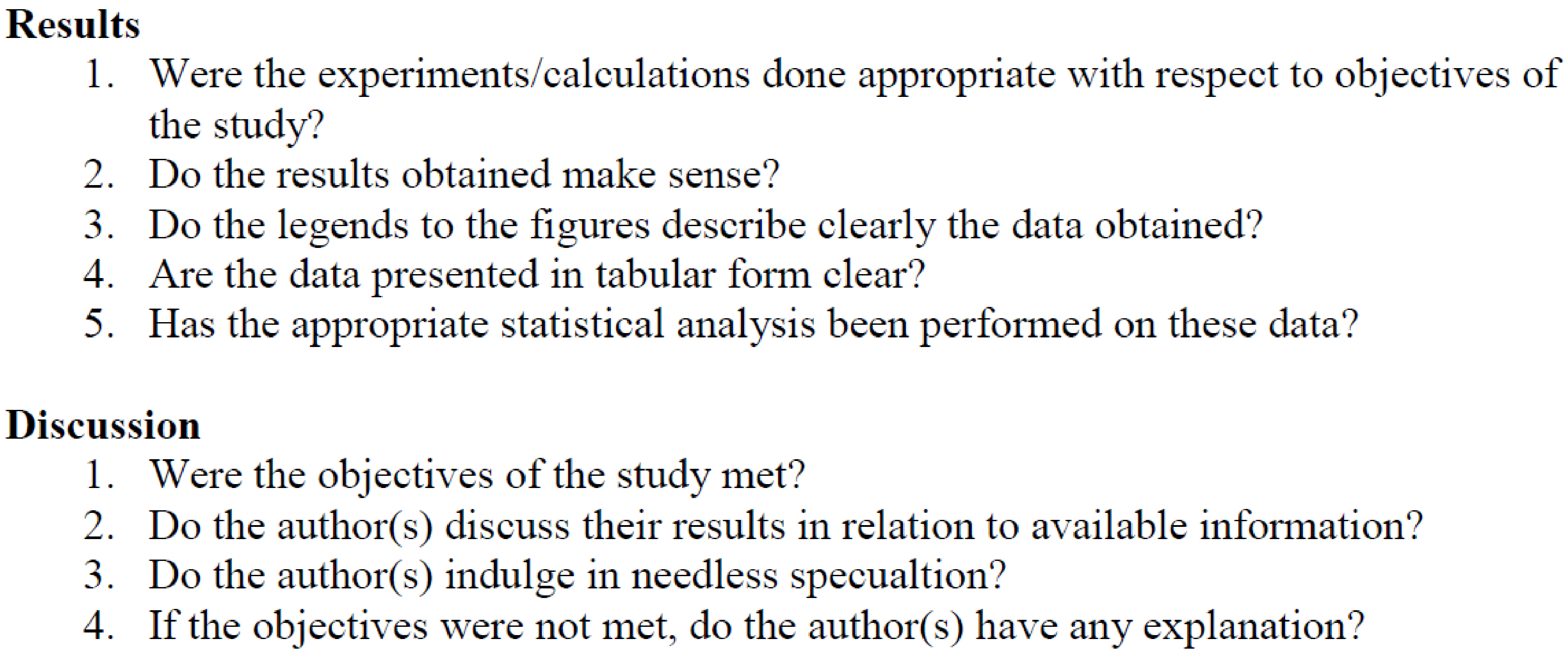
Introduction

1. Did the author(s) indicate why the study was undertaken?
2. Was the background information provided adequate to understand the aims of the study?

Methods

1. Has the source of the data been clearly given?
2. Were the methods described in sufficient detail for others to repeat or extend the study?
3. If standard methods were used, were adequate references given?
4. Have the author(s) indicated the reasons why particular procedures were used?
5. Have the author(s) indicated clearly the potential problems with the methods used?
6. Have the author(s) indicated the limitations of the methods used?
7. (Have the sources of drugs been given?)
8. Have the author(s) specified the statistical procedures used?
9. Are the statistical methods appropriate?

Critical evaluation of a paper or report



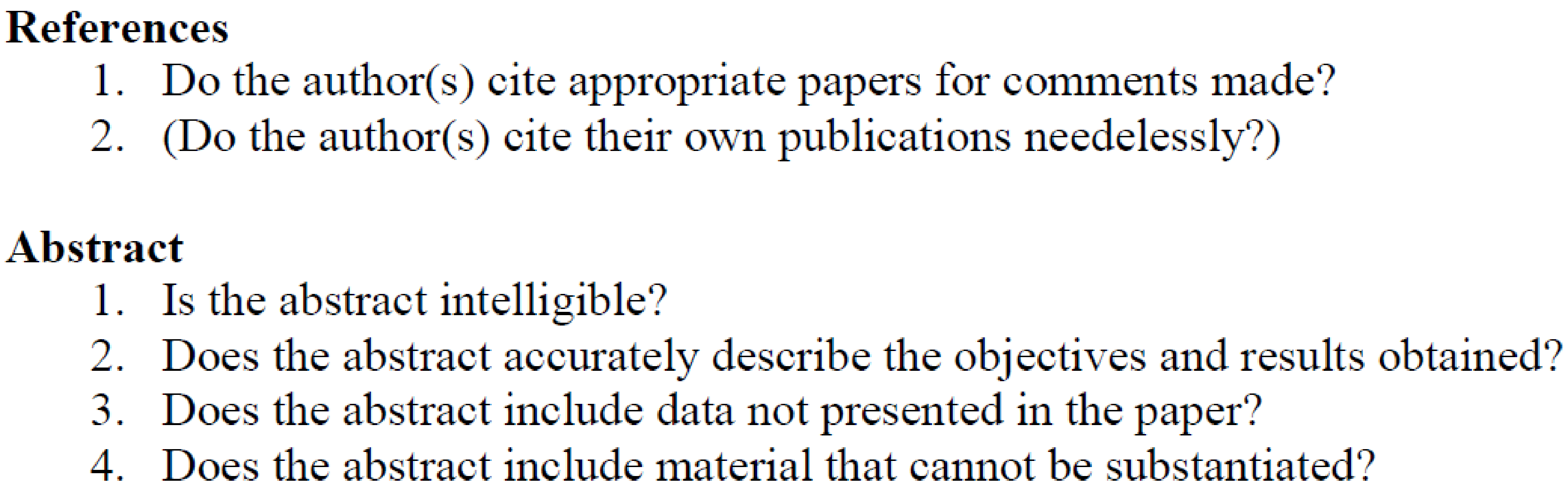
Results

1. Were the experiments/calculations done appropriate with respect to objectives of the study?
2. Do the results obtained make sense?
3. Do the legends to the figures describe clearly the data obtained?
4. Are the data presented in tabular form clear?
5. Has the appropriate statistical analysis been performed on these data?

Discussion

1. Were the objectives of the study met?
2. Do the author(s) discuss their results in relation to available information?
3. Do the author(s) indulge in needless specualtion?
4. If the objectives were not met, do the author(s) have any explanation?

Critical evaluation of a paper or report



References

1. Do the author(s) cite appropriate papers for comments made?
2. (Do the author(s) cite their own publications needelessly?)

Abstract

1. Is the abstract intelligible?
2. Does the abstract accurately describe the objectives and results obtained?
3. Does the abstract include data not presented in the paper?
4. Does the abstract include material that cannot be substantiated?

Effective Reading

Why?

Your teachers give you a pile of papers / book chapter to read.

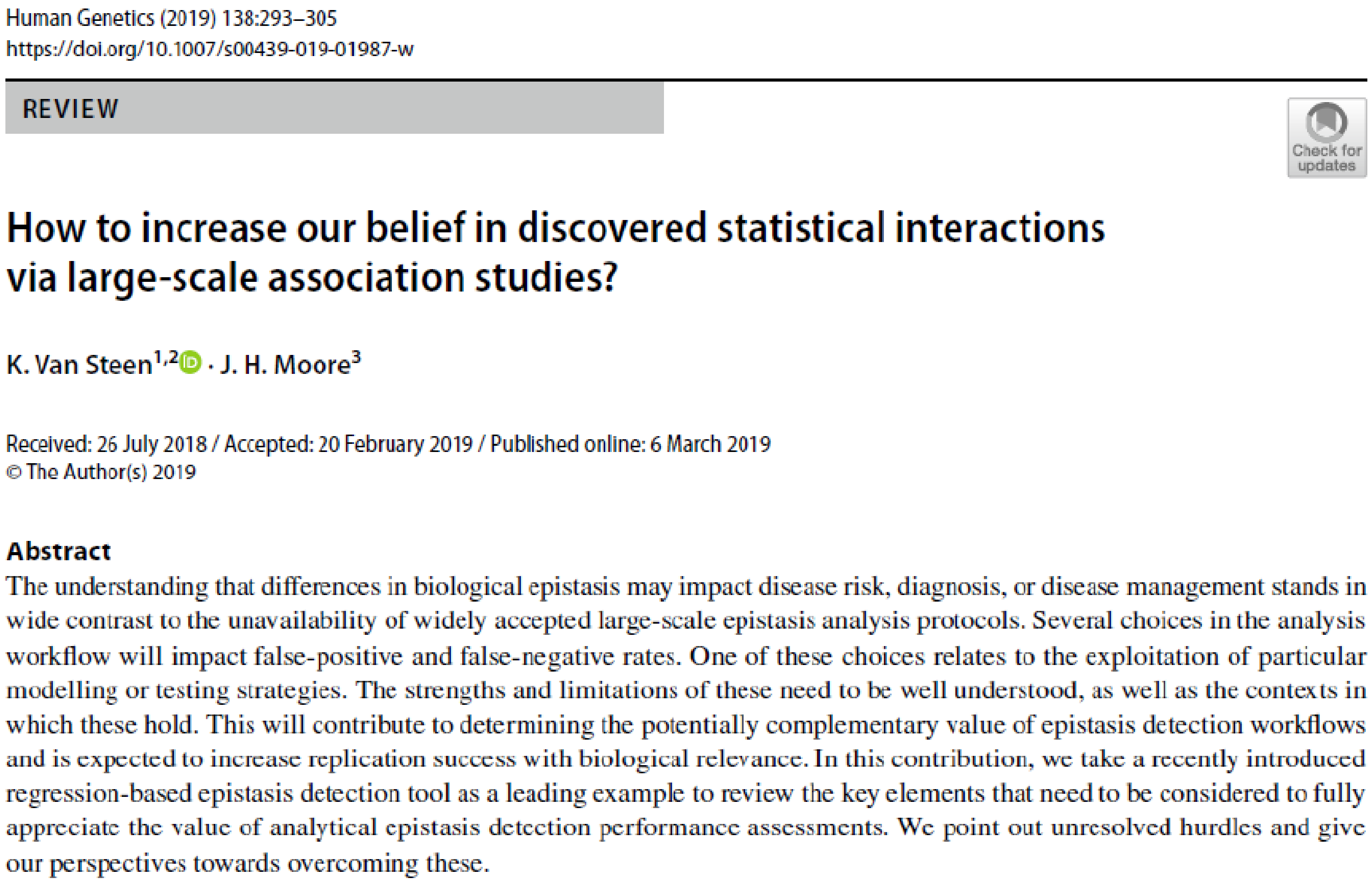
Ouch…

Efficient reading skills will be helpful in multiple ways: knowledge gain, insight in writing styles, structuring thoughts, distinguishing main and secondary issues, …

What are different types of scientific literature?

- Primary (authors carried out the work)
  - Examples: monographs, theses or dissertations, conference papers and reports
  - Peer-reviewed journal
  - Particular format

- Secondary (work of others; target: others in the field)
  - Examples: review journals, monographic books and textbooks, handbooks and manuals
  - More flexible style: still scientific and fully referenced



Human Genetics (2019) 138:293–305
https://doi.org/10.1007/s00439-019-01987-w

REVIEW

How to increase our belief in discovered statistical interactions via large-scale association studies?

K. Van Steen[1,2] · J. H. Moore[3]

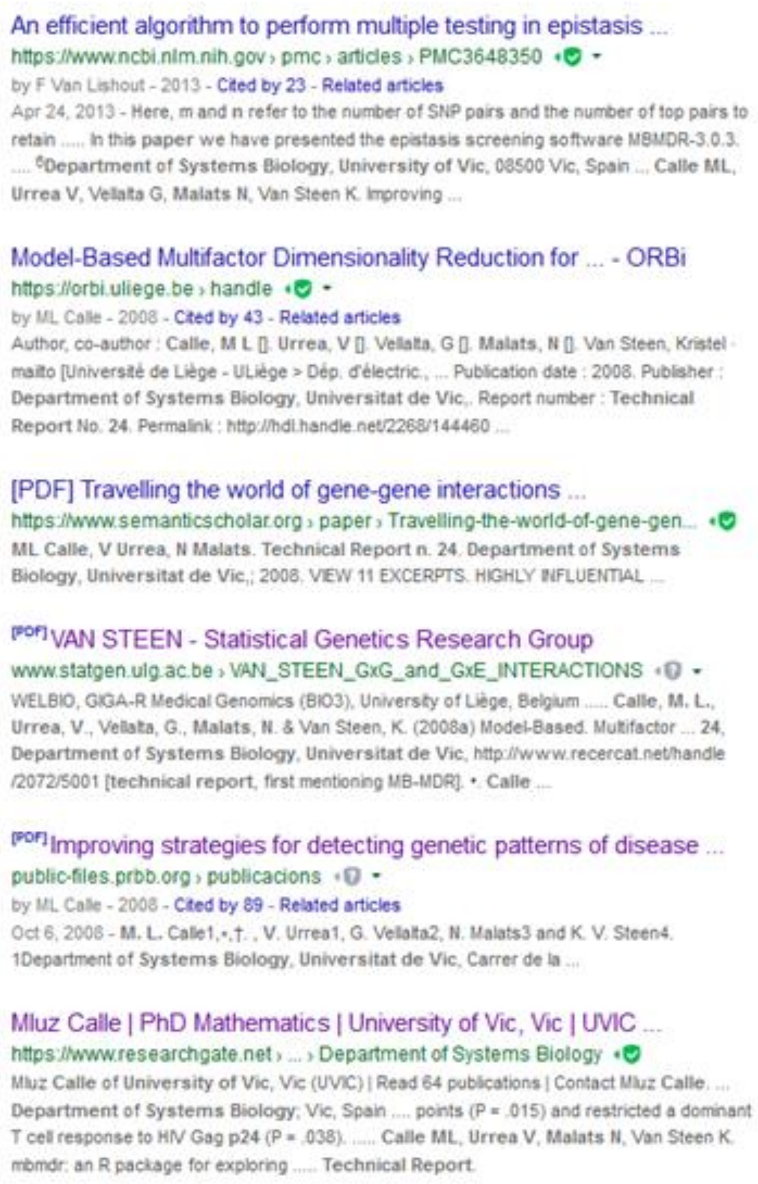
Received: 26 July 2018 / Accepted: 20 February 2019 / Published online: 6 March 2019

Abstract

The understanding that differences in biological epistasis may impact disease risk, diagnosis, or disease management stands in wide contrast to the unavailability of widely accepted large-scale epistasis analysis protocols. Several choices in the analysis workflow will impact false-positive and false-negative rates. One of these choices relates to the exploitation of particular modelling or testing strategies. The strengths and limitations of these need to be well understood, as well as the contexts in which these hold. This will contribute to determining the potentially complementary value of epistasis detection workflows and is expected to increase replication success with biological relevance. In this contribution, we take a recently introduced regression-based epistasis detection tool as a leading example to review the key elements that need to be considered to fully appreciate the value of analytical epistasis detection performance assessments. We point out unresolved hurdles and give our perspectives towards overcoming these.

What are different types of scientific literature?

- Tertiary (work of others; target: interdisciplinary audience, public)
    - Examples: science magazines, newsletters, science articles in newspapers, introductory textbooks and encyclopedias
    - Popular rather than a scientific style; reduced/short bibliography
- Grey (limited distribution, difficult accessing)
    - Examples: technical reports, journals published by special interest groups, abstracts of conference papers and conference proceedings that are only made available to conference participants, working papers, some online documents



An efficient algorithm to perform multiple testing in epistasis ...
https://www.ncbi.nlm.nih.gov › pmc › articles › PMC3648350
by F Van Lishout - 2013 - Cited by 23 - Related articles
Apr 24, 2013 - Here, m and n refer to the number of SNP pairs and the number of top pairs to retain ..... In this paper we have presented the epistasis screening software MBMDR-3.0.3. .... [6]Department of Systems Biology, University of Vic, 08500 Vic, Spain ... Calle ML, Urrea V, Vellalta G, Malats N, Van Steen K. Improving ...

Model-Based Multifactor Dimensionality Reduction for ... - ORBi
https://orbi.uliege.be › handle
by ML Calle - 2008 - Cited by 43 - Related articles
Author, co-author : Calle, M L []. Urrea, V []. Vellalta, G []. Malats, N []. Van Steen, Kristel - mailto [Université de Liège - ULiège > Dép. d'électric., ... Publication date : 2008. Publisher : Department of Systems Biology, Universitat de Vic,. Report number : Technical Report No. 24. Permalink : http://hdl.handle.net/2268/144460 ...

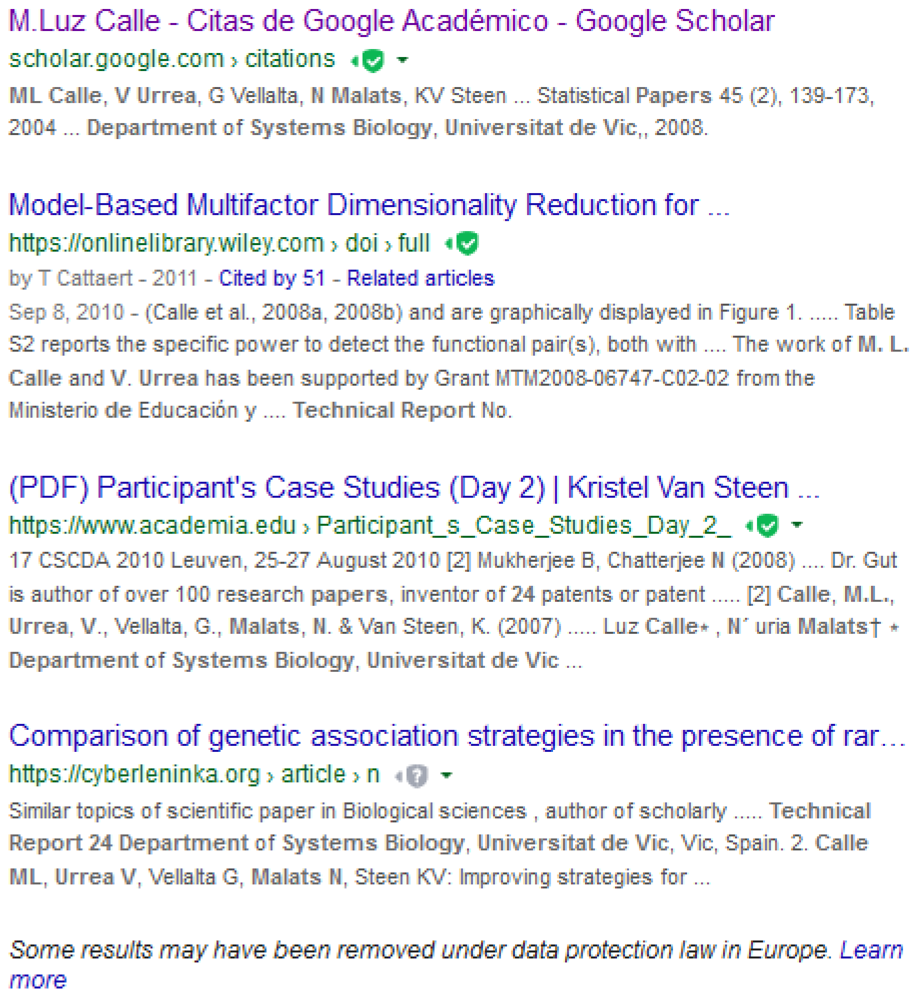
[PDF] Travelling the world of gene-gene interactions ...
https://www.semanticscholar.org › paper › Travelling-the-world-of-gene-gen...
ML Calle, V Urrea, N Malats. Technical Report n. 24. Department of Systems Biology, Universitat de Vic,; 2008. VIEW 11 EXCERPTS. HIGHLY INFLUENTIAL ...

[PDF] VAN STEEN - Statistical Genetics Research Group
www.statgen.ulg.ac.be › VAN_STEEN_GxG_and_GxE_INTERACTIONS
WELBIO, GIGA-R Medical Genomics (BIO3), University of Liège, Belgium ..... Calle, M. L., Urrea, V., Vellalta, G., Malats, N. & Van Steen, K. (2008a) Model-Based. Multifactor ... 24, Department of Systems Biology, Universitat de Vic, http://www.recercat.net/handle /2072/5001 [technical report, first mentioning MB-MDR]. •. Calle ...

[PDF] Improving strategies for detecting genetic patterns of disease ...
public-files.prbb.org › publicacions
by ML Calle - 2008 - Cited by 89 - Related articles
Oct 6, 2008 - M. L. Calle1,∗,†. , V. Urrea1, G. Vellalta2, N. Malats3 and K. V. Steen4. 1Department of Systems Biology, Universitat de Vic, Carrer de la ...

Mluz Calle | PhD Mathematics | University of Vic, Vic | UVIC ...
https://www.researchgate.net › ... › Department of Systems Biology
Mluz Calle of University of Vic, Vic (UVIC) | Read 64 publications | Contact Mluz Calle. ... Department of Systems Biology; Vic, Spain .... points (P = .015) and restricted a dominant T cell response to HIV Gag p24 (P = .038). ..... Calle ML, Urrea V, Malats N, Van Steen K. mbmdr: an R package for exploring ..... Technical Report.

"ML Calle, V Urrea, N Malats. Technical Report n. 24. ...UVIC"



M.Luz Calle - Citas de Google Académico - Google Scholar
scholar.google.com › citations
ML Calle, V Urrea, G Vellalta, N Malats, KV Steen ... Statistical Papers 45 (2), 139-173, 2004 ... Department of Systems Biology, Universitat de Vic,, 2008.

Model-Based Multifactor Dimensionality Reduction for ...
https://onlinelibrary.wiley.com › doi › full
by T Cattaert - 2011 - Cited by 51 - Related articles
Sep 8, 2010 - (Calle et al., 2008a, 2008b) and are graphically displayed in Figure 1. ..... Table S2 reports the specific power to detect the functional pair(s), both with .... The work of M. L. Calle and V. Urrea has been supported by Grant MTM2008-06747-C02-02 from the Ministerio de Educación y .... Technical Report No.

(PDF) Participant's Case Studies (Day 2) | Kristel Van Steen ...
https://www.academia.edu › Participant_s_Case_Studies_Day_2_
17 CSCDA 2010 Leuven, 25-27 August 2010 [2] Mukherjee B, Chatterjee N (2008) .... Dr. Gut is author of over 100 research papers, inventor of 24 patents or patent ..... [2] Calle, M.L., Urrea, V., Vellalta, G., Malats, N. & Van Steen, K. (2007) ..... Luz Calle∗ , N´ uria Malats† ∗ Department of Systems Biology, Universitat de Vic ...

Comparison of genetic association strategies in the presence of rar...
https://cyberleninka.org › article › n
Similar topics of scientific paper in Biological sciences , author of scholarly ..... Technical Report 24 Department of Systems Biology, Universitat de Vic, Vic, Spain. 2. Calle ML, Urrea V, Vellalta G, Malats N, Steen KV: Improving strategies for ...

Some results may have been removed under data protection law in Europe. Learn more

Why is it useful to regularly read scientific documents?

- To gain knowledge (scientific knowledge, opinions, strategies)
- To stay on top of your field as well as linked fields (intro, discussion)
- To learn about journal styles / slang
- To become an expert in sifting through literature
- To learn about written communication

How to read a scientific article?

- Skim the article and identify its structure
- Distinguish the main points
- Generate the questions and be aware of your understanding
- Draw inferences
- Take notes as you read …

Skim the article and identify its structure

- Features of abstracts:
  - Purpose / rationale (why?)
  - Methodology (how?)
  - Results (what was found?)
  - Conclusion (what do the results mean?)

Skim the article and identify its structure

- Features of introductions:
  - Triggering interest
  - Providing enough information to understand the article
    - Broad: What is known?
    - Specific: What is not known?
    - Focus: What are the questions addressed?

Skim the article and identify its structure

- Features of methods:
    - Which experiments / tools were used to address the questions?
    - Most difficult to read especially when not well structured
    - Should provide the reader with information about the design of the experiment such that the validity of them can be evaluated
- Features of results and discussion:
    - Statements of what was found and reference to (visual) data [Figures, Tables] -- results
    - Comparisons to other results, interpretations, opinions -- discussion

Distinguish the main points

- Document level
  - Title, abstract, keywords
  - Visuals (captions)
  - Introduction
- Paragraph level
  - First few sentences in a paragraph
  - We hypothesize, we propose, we introduce, we develop, data suggests, in contrast to, surprising, …

Generate questions and be aware of understanding: <u>active reading</u>

- Before and during reading:
  - Who are these authors? What journal is this? Might I question the credibility of the work? Have I taken the time to understand all the terminology? Have I gone back to read an article or review that would help me understand this work better? Am I spending too much time reading the less important parts of this article? Is there someone I can talk to about confusing parts of this article?
- After reading:
  - What specific problem does this research address? Why is it important? Is the method used a good one/ the best? What are the specific findings? Am I able to summarize them in a few sentences? Are the findings supported by persuasive evidence? Is there an alternative interpretation not addressed? How are the findings unique/new/unusual or supportive of other work in the field? How do these results relate to my work? Applications? Interesting additional experiments to address the questions?

Draw inference: <u>improve understanding and recall information</u>

- Rely on your prior knowledge, world experience, materials provided in the paper, to draw inferences.
  - We learn about some things by experiencing them first-hand, but we gain other knowledge by inference — the process of inferring things based on what is already known.

Take notes as you read

- Details will slip away, eventually …
  - Stuff your (electronic) notebook, keep records of all of your scientific reading with summaries of their importance.
  - Time spent doing this will be regained when writing background, related work or literature review sections.

Be critical of published data/results!

- A lot of data is at your disposal but are they thrust-worthy?
    - Private data collections (curated according to standards?)
    - Public data collections (curated uniformly?)
    - Publications (source or summary data provided?)
    - Computerized databanks (block-chained or not?)

Errors will almost surely exist

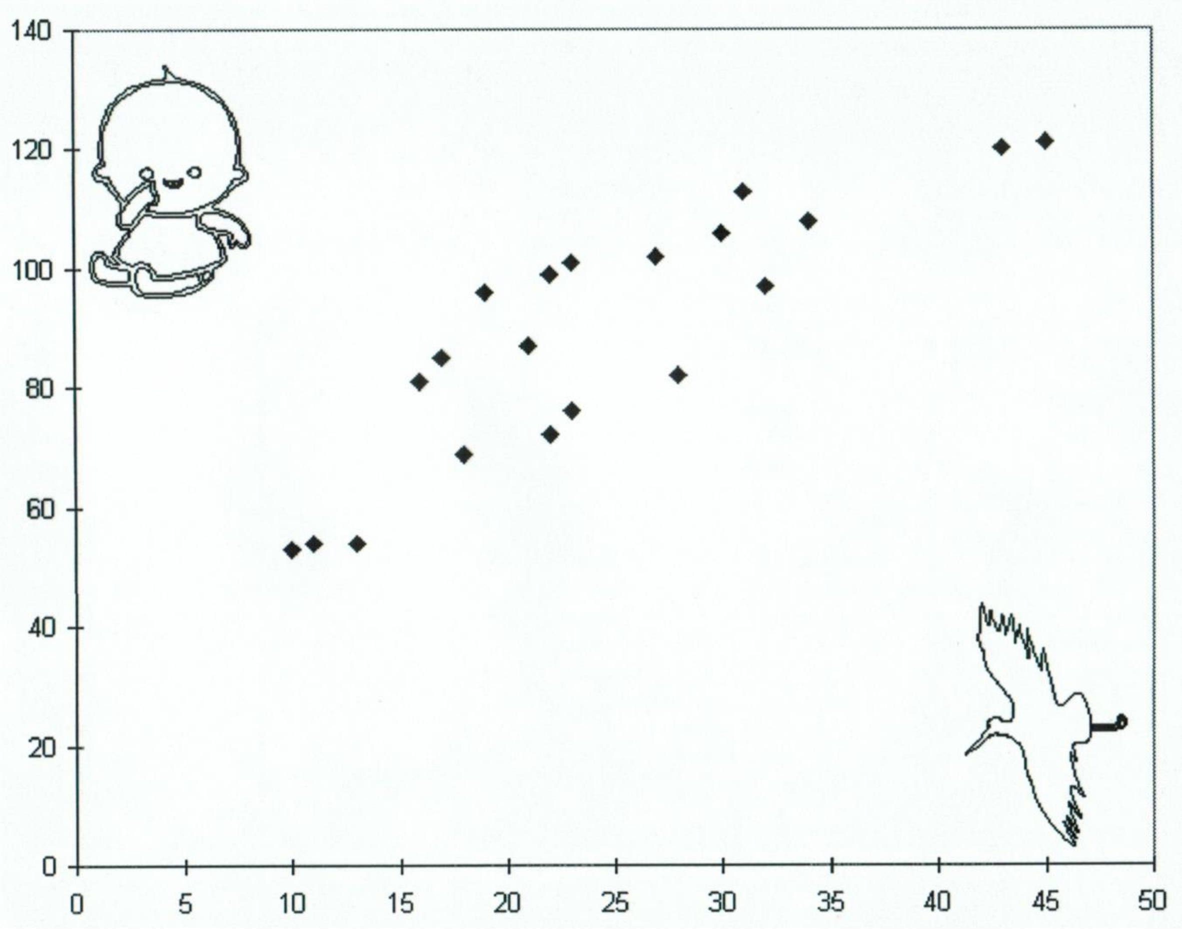
- Apart from sampling errors, measurement error may arise:
  - mistakes in conceptualization
  - structural characteristics of the data collection process
- Relevant questions include:
  - How large are the errors?
  - What is the probability for a given error range?
  - Do errors cluster towards the end of a distribution?
  - In which direction does the error go?

In general: “better” science through “better” data

SCIENTIFIC DATA
#scidata15
Understanding the method is the thing...
Show your workings out
USING ONE MORE NOTE BOOK?
RESEARCH DATA FOR DISCOVERY
TRANSPORTING DATA (esp. across disciplines)
DATA IS THE NEW SPICE
THE CLOUD
Reducing & MANAGING UNCERTAINTY
SCIENCE as an OPEN ENTERPRISE
BUILDING ON REPORTS WE CAN UNDERSTAND
PROF JEREMY G FREY
WHAT IS MY digital AURA?
Royal Society Report tells us how it is
Meta data MORE & MORE
Narrative is key - don't lose the Story!
www.dreamingnewmexico.org
... are RETURNING
Can't see the wood for the trees?
There are risks in SECONDARY analysis
RISKS in Distortion
RISKS in consent and assent
CONTROLLED ACCESS ENCOURAGED...
PRINCIPLES ARE REQUIRED
CONSENT CANNOT ALWAYS BE GIVEN DIRECTLY
TERMINOLOGY
SHARING INDIVIDUAL Patient data in CLINICAL TRIALS...
MR. MATT SYDES
WE HAVE AN Ethical Obligation to consider...
RESEARCH DATA FOR PUBLICATIONS
Publishing establishes PRIORITY
Laser
MANY OF THE MOST INFLUENTIAL PAPERS WERE VERY LIGHT ON DETAIL
Papers form part of a larger eco-system
Reproduceability
is the KEY
HEY- THAT'S MY WORK
THEODORE MAIMAN
I CALL IT A "LASER"
SCIENCE is BASED on trust
DR JOERG HEBER
but trust only gets you so far.... The 'data makes the paper'
BEWARE of WALLED DATA GARDENS
HOW TO MAKE data RE-USABLE?
The Secret's in the metadata
Open data is about MORE THAN DISCLOSURE it must be Fair
Dr. ANDREW HUFTON
Find the right repository for your data...
mmm... reproduce-able
• Findable
• Accessible
• Interoperable
• Reusable
5 PRINCIPLES
• QUALITY CURATION
• COMMITMENT TO LONG TERM PRESERVATION
• FEATURES THAT SUPPORT COLLABORATIVE ANALYSIS
• FEATURES THAT ALLOW YOU TO KEEP DATA PRIVATE TILL READY TO PUBLISH
• INVESTIGATE DATA ARCHIVING OPTIONS AT YOUR INSTITUTIONS...
LIGHTNING TALKS
Open Trials
Not just talk, we build things too
CKAN
CENSORED
It's commercially 'sensitive' So don't share this info
But why?
opening up ALL the world's trial data EVER!
We are currently building a Transparency SCOREBOARD
A case study in Neuroscience
REPRODUCEABLE RESEARCH TOOLS...
EMBEDDING code in your documents
Avoid Embarrassment:
YOUR FUTURE-SELF WILL THANK YOU!
EVA
archived at EBI
CLinvar
MERGING Open Access Database sets
Overcoming Big Data
Active Data Biology: it all goes back to GITHUB
A TOOL BUILT FOR SHARING
GITHUB
A TOTAL TRANSPARENCY
Not in your HEAD, Not in a NOTEBOOK
But ready to share
ALL in one CODE
EMBL EBI
Creating a GLOBAL ARCHIVE & virtuous circle
Minotaur
DATA STORAGE
"We should all see ourselves as innovators"
Philosophical Transactions
1st Scientific paper
Now we can work together on an INDUSTRIAL SCALE
SAY cheese
image BASED data is becoming THE Norm
• real time
• in-field
• molecular measurement
2nd use
Learn to code: it's a basic 21st CENTURY Skill
RESEARCH INNOVATION:
Think Meta
Do THE Right Thing
Be the Change!
e-science allows for MASS Collaboration

(www.nature.com/openresearch/)

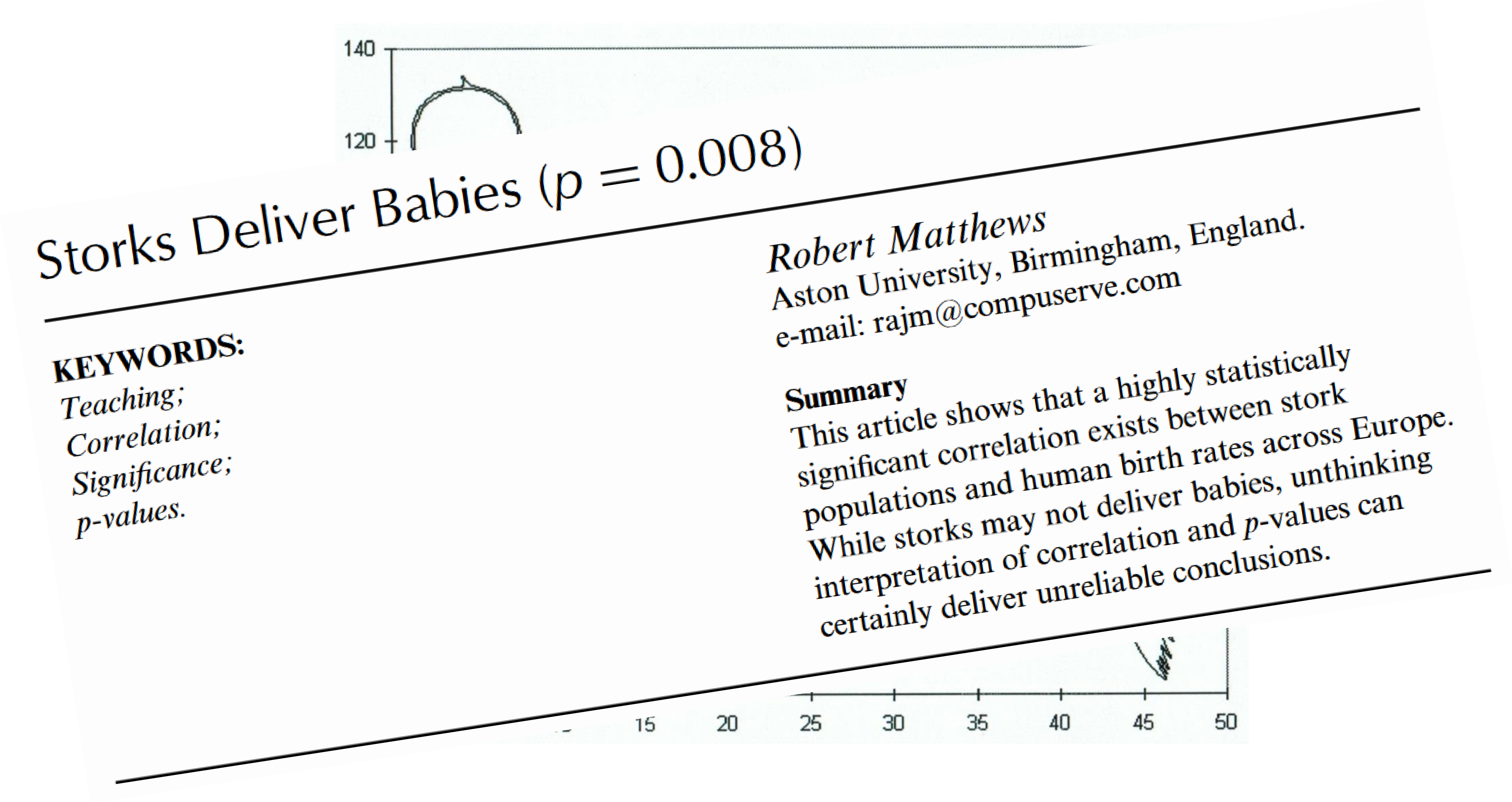
Beware if jumping to conclusions: causation versus association



number of newborns

number of breeding stork pairs

Beware if jumping to conclusions: causation versus association


140
120
15
20
25
30
35
40
45
50

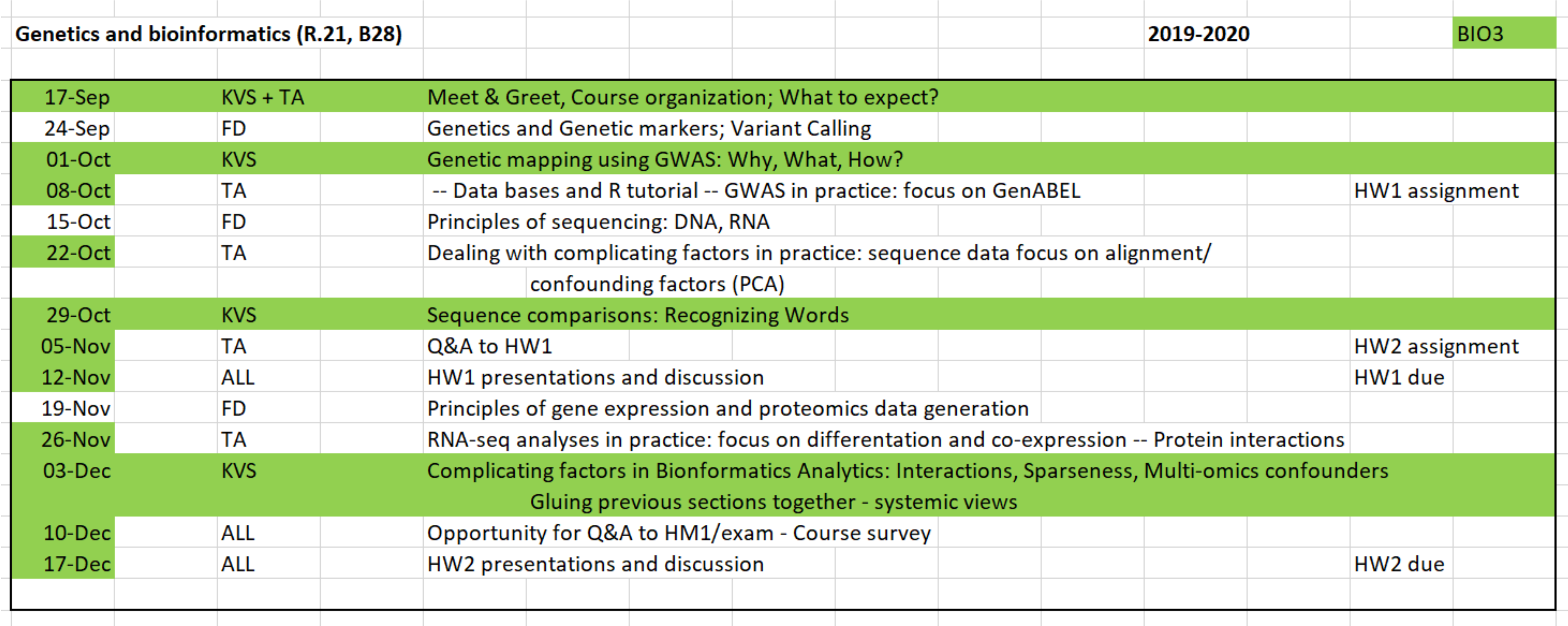
Storks Deliver Babies ($p = 0.008$)

Robert Matthews
Aston University, Birmingham, England.
e-mail: rajm@compuserve.com

KEYWORDS:
Teaching;
Correlation;
Significance;
p-values.

Summary
This article shows that a highly statistically significant correlation exists between stork populations and human birth rates across Europe. While storks may not deliver babies, unthinking interpretation of correlation and *p*-values can certainly deliver unreliable conclusions.

Tentative course layout



Genetics and bioinformatics (R.21, B28)		**2019-2020**	BIO3
17-Sep	KVS + TA	Meet & Greet, Course organization; What to expect?	
24-Sep	FD	Genetics and Genetic markers; Variant Calling	
01-Oct	KVS	Genetic mapping using GWAS: Why, What, How?	
08-Oct	TA	-- Data bases and R tutorial -- GWAS in practice: focus on GenABEL	HW1 assignment
15-Oct	FD	Principles of sequencing: DNA, RNA	
22-Oct	TA	Dealing with complicating factors in practice: sequence data focus on alignment/	
		confounding factors (PCA)	
29-Oct	KVS	Sequence comparisons: Recognizing Words	
05-Nov	TA	Q&A to HW1	HW2 assignment
12-Nov	ALL	HW1 presentations and discussion	HW1 due
19-Nov	FD	Principles of gene expression and proteomics data generation	
26-Nov	TA	RNA-seq analyses in practice: focus on differentation and co-expression -- Protein interactions	
03-Dec	KVS	Complicating factors in Bionformatics Analytics: Interactions, Sparseness, Multi-omics confounders	
		Gluing previous sections together - systemic views	
10-Dec	ALL	Opportunity for Q&A to HM1/exam - Course survey	
17-Dec	ALL	HW2 presentations and discussion	HW2 due

Questions?