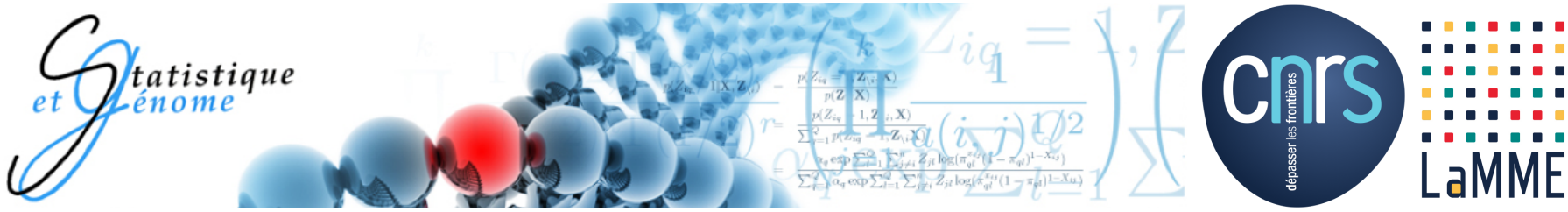


Reproducible Research

An introduction to the "Duke saga"

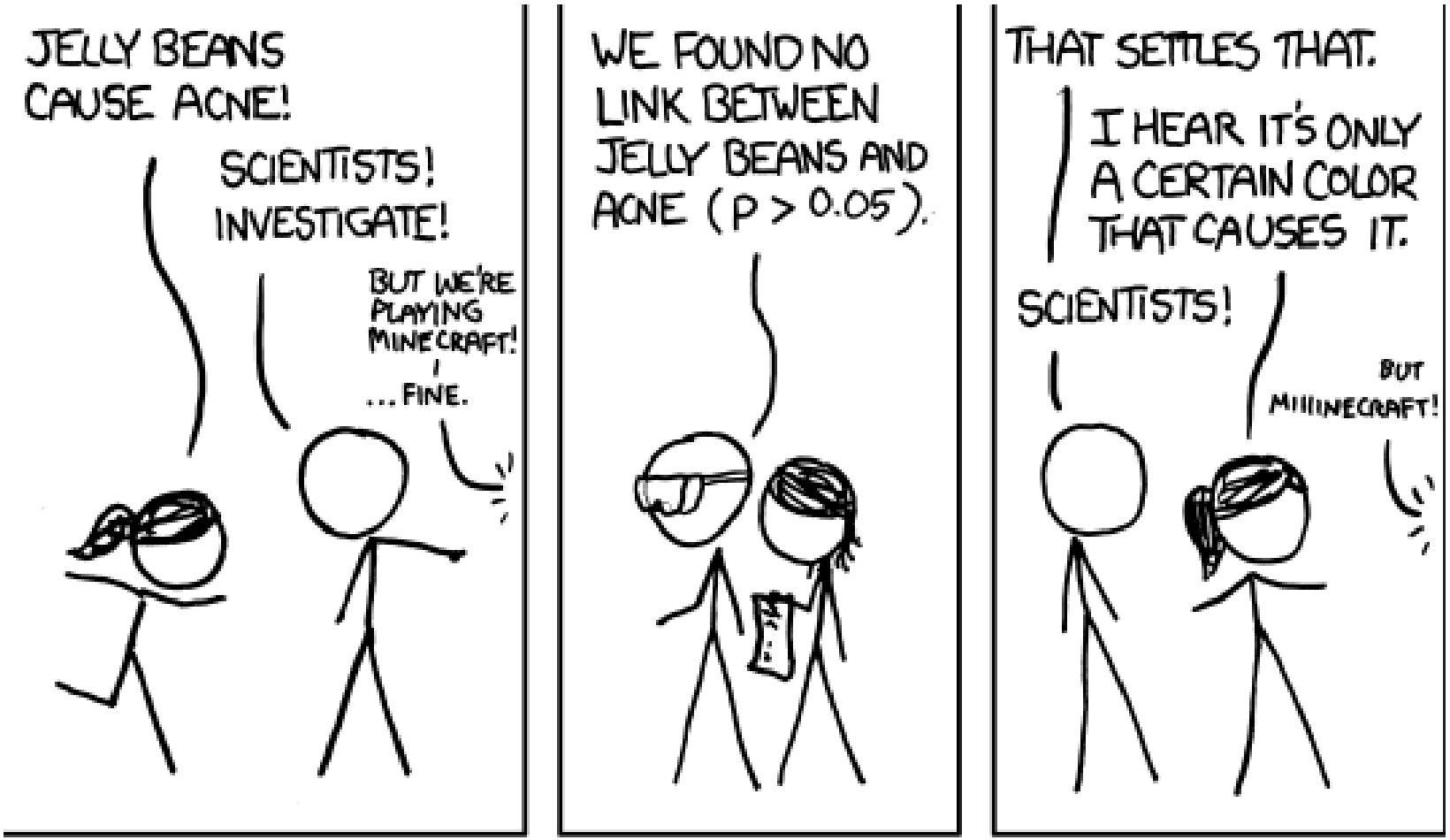
Pierre Neuvial

Laboratoire de Mathématiques et Modélisation d'Evry
Université d'Évry – UMR CNRS 8071 – USC INRA
`http://stat.genopole.cnrs.fr/~pneuvial`

Programming for Genomics — M2 ISG

A multiple testing joke

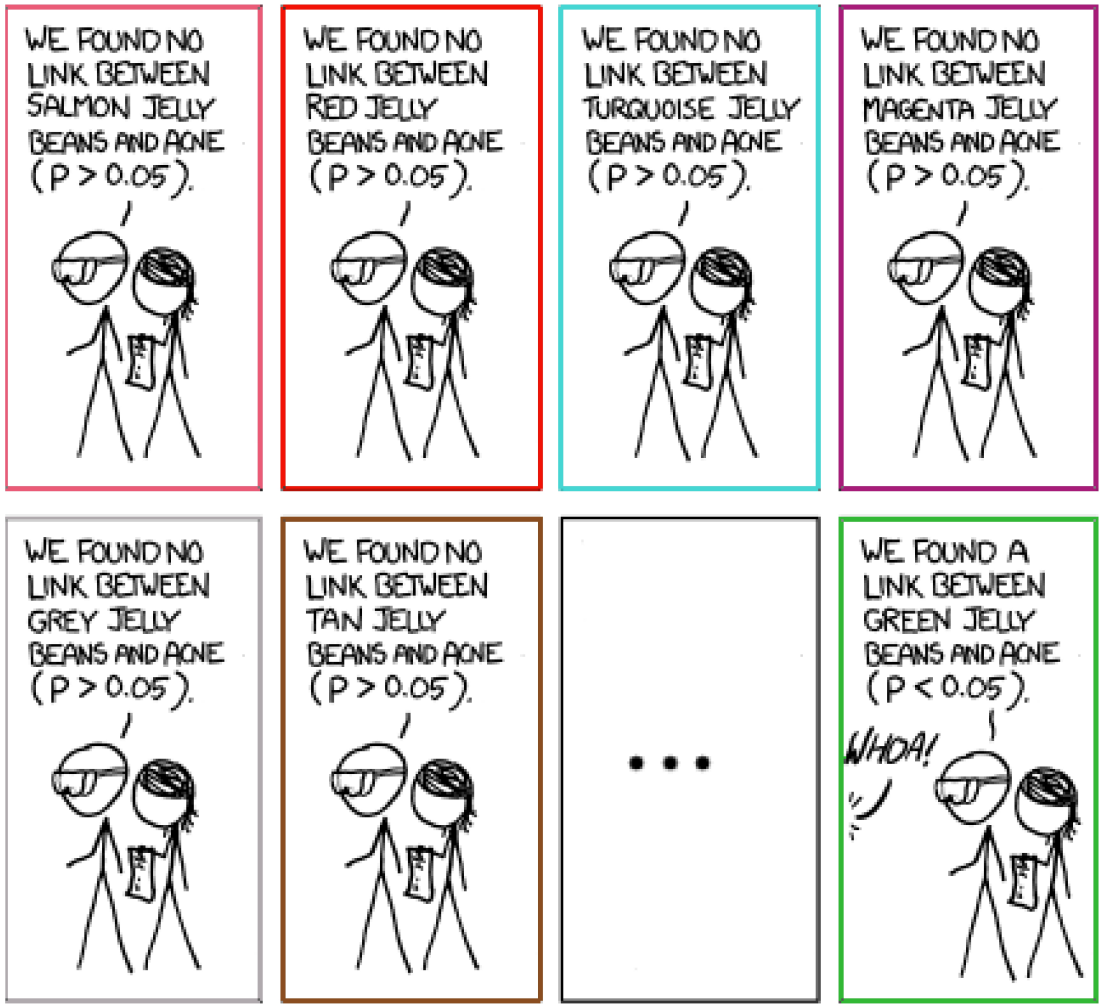
`http://imgs.xkcd.com/comics/significant.png`



Note : Jelly beans = Dragibus ()

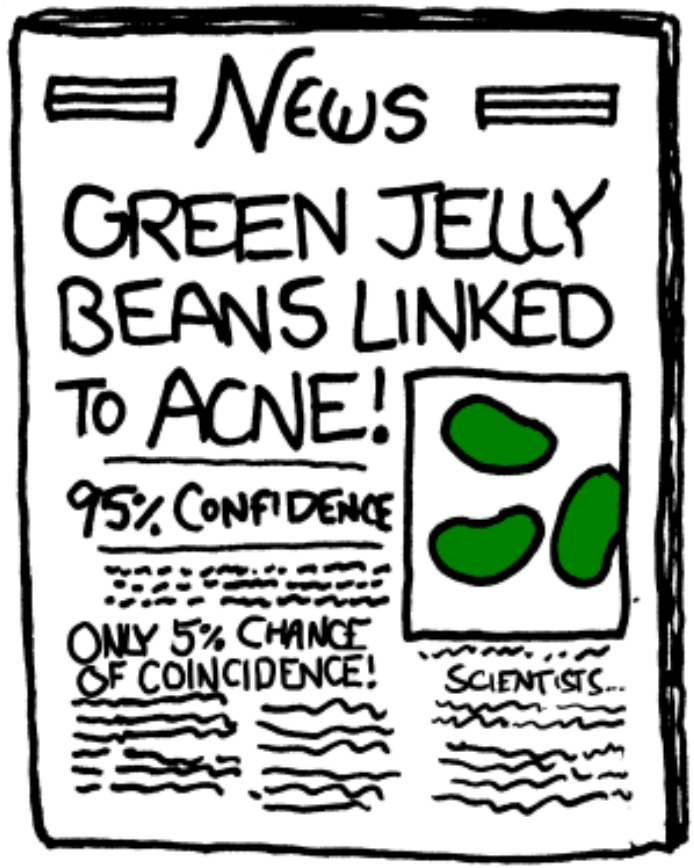
A multiple testing joke

`http://imgs.xkcd.com/comics/significant.png`



A multiple testing joke

`http://imgs.xkcd.com/comics/significant.png` + colors by Y. Benjamini



Can we trust published research ?

Ioannidis, *PLoS Medicine*, 2005

Try many experiments

$\Downarrow$

1000 pure noise

30 perfect signal

$\Downarrow$

publish results with a p-value ≤ 0.05

$\Downarrow$

$\simeq 50$ false discoveries

30 true discoveries

Can we trust published research ?

Ioannidis, *PLoS Medicine*, 2005

Try many experiments

$\Downarrow$

1000 pure noise — 30 perfect signal

$\Downarrow$

publish results with a p-value ≤ 0.05

$\Downarrow$

$\simeq 50$ false discoveries — 30 true discoveries

Can we trust published research ?

Ioannidis, *PLoS Medicine*, 2005

Try many experiments

$\Downarrow$

1000 pure noise — 30 perfect signal

$\Downarrow$

publish results with a p-value ≤ 0.05

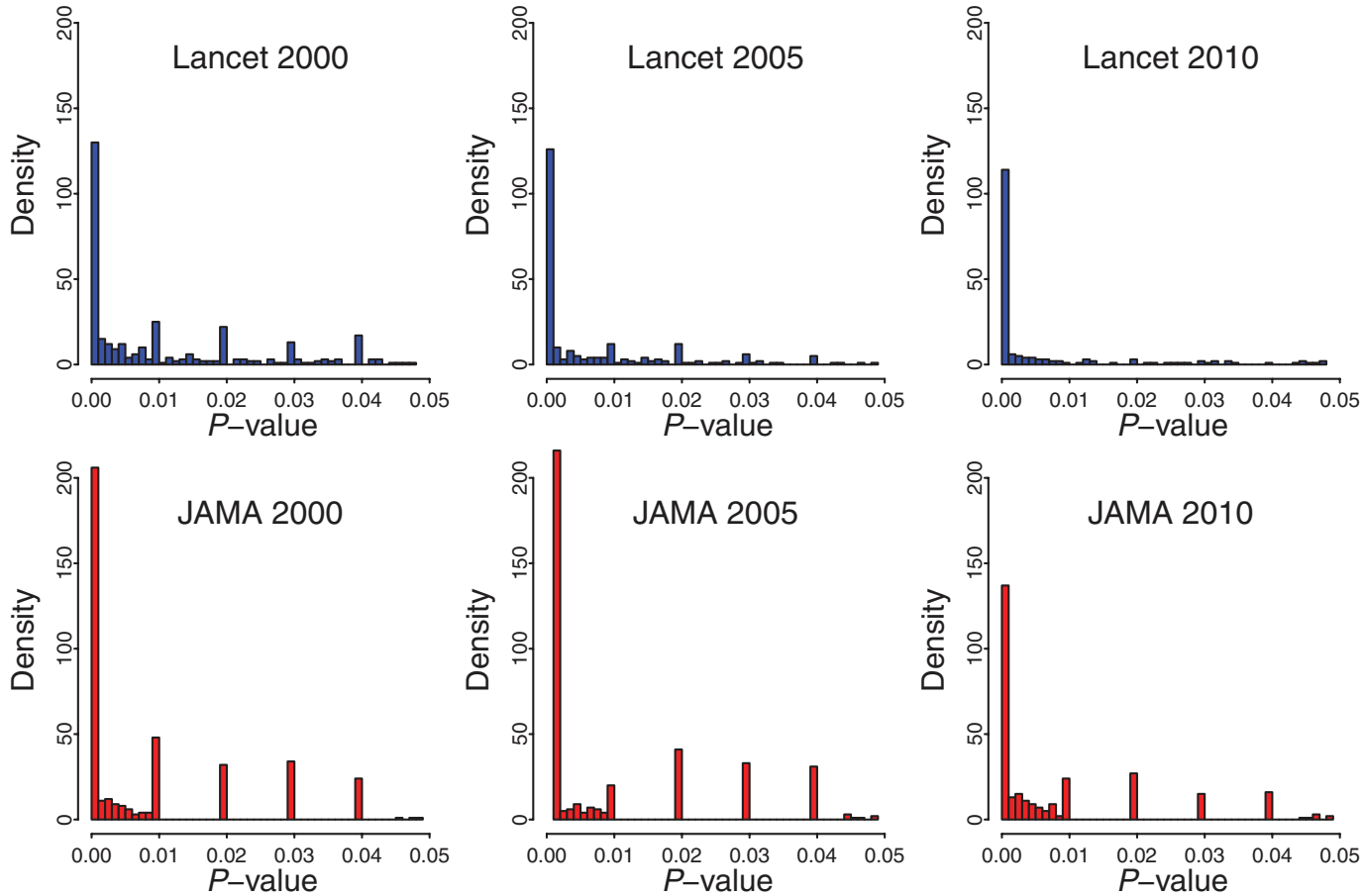
$\Downarrow$

$\simeq 50$ false discoveries — 30 true discoveries

Can we trust published research ?

Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.

$\simeq 14\%$



Can we trust published research ?

- Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.

Can we trust published research ?

- Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.
- Ioannidis, J. P. (2014). Discussion : Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics*, 15(1), 28-36.

Can we trust published research ?

- Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.
- Ioannidis, J. P. (2014). Discussion : Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics*, 15(1), 28-36.
- Gelman, A., & O'Rourke, K. (2014). Discussion : Difficulties in making inferences about scientific truth from distributions of published *p*-values. *Biostatistics*, 15(1), 18-23.

Can we trust published research ?

- Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.
- Ioannidis, J. P. (2014). Discussion : Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics*, 15(1), 28-36.
- Gelman, A., & O'Rourke, K. (2014). Discussion : Difficulties in making inferences about scientific truth from distributions of published *p*-values. *Biostatistics*, 15(1), 18-23.
- Leek J. T., "Why I disagree with Andrew Gelman's critique of my paper about the rate of false discoveries in the medical literature", simplystatistics.org

Can we trust published research ?

- Jager, L. R., & Leek, J. T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1-12.
- Ioannidis, J. P. (2014). Discussion : Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics*, 15(1), 28-36.
- Gelman, A., & O'Rourke, K. (2014). Discussion : Difficulties in making inferences about scientific truth from distributions of published *p*-values. *Biostatistics*, 15(1), 18-23.
- Leek J. T., "Why I disagree with Andrew Gelman's critique of my paper about the rate of false discoveries in the medical literature", simplystatistics.org
- . . .to be continued.

Replication and Reproducibility

Definition from `http://simplystatistics.org`

Replication

"A study is only replicable if you perform the exact same experiment (at least) twice, collect data in the same way both times, perform the same data analysis, and arrive at the same conclusions"

Replication can be

- long (years)
- expensive (millions)
- difficult (experiments)

... particularly in genomics.

Replication and Reproducibility

Definition from `http://simplystatistics.org`

Reproducibility

"A study is reproducible if there is a specific set of computational functions/analyses (usually specified in terms of code) that exactly reproduce all of the numbers in a published paper from raw data"

Reproducibility can be assessed quickly and easily if the authors of the paper publish their code and data.

Reproducibility "should serve as **a minimum standard** for judging scientific claims when full independent replication of a study is not possible".

– R. D. Peng, Science 334, 1226 (2011)

Replication and Reproducibility

M. Bissel, "Reproducibility : The risks of the replication drive", *Nature*, 2013

What can be done to promote reproducibility ?

Although it would greatly benefit science and the scientific community itself, reproducibility (= "minimum standard") is far from being reached.

Provide your code !

e.g. R package, or GitHub, SourceForge, or `http://runmycode.org`

It is the authors who should spend time on their code, not the readers trying to reproduce the results !

Thoughts :

- ask for data/code when refereeing papers
- software development should be budgeted in our grants
- writing "application notes" helps valorizing reproducible work

How bad can things go ? An example from economics

Herndon, Ash, & Pollin (2013)

Table 1: Real GDP Growth as the Level of Public Debt Varies
20 advanced economies, 1946–2009

	Ratio of Public Debt to GDP			
	Below 30 percent	30 to 60 percent	60 to 90 percent	90 percent and above
Average real GDP growth	4.1	2.8	2.8	−0.1

From Reinhart & Rogoff (2010).

HAP 2013 : "Reinhart & Rogoff (2010) have clearly exerted a major influence in recent years on public policy debates over the management of government debt and fiscal policy more broadly. Their findings have provided significant support for the austerity agenda that has been ascendant in Europe and the United States since 2010."

How bad can things go ? An example from economics

Herndon, Ash, & Pollin (2013)

Table 1: Real GDP Growth as the Level of Public Debt Varies
20 advanced economies, 1946–2009

	Ratio of Public Debt to GDP			
	Below 30 percent	30 to 60 percent	60 to 90 percent	90 percent and above
Average real GDP growth	4.1	2.8	2.8	−0.1

From Reinhart & Rogoff (2010).

HAP 2013 identified several problems with the "-0.1" in this table :

- missing and omitted data
- coding errors (in MS Excel)
- "unconventional" weighting scheme

After correction : "-0.1" becomes "+2.2"

How bad can things go ? An example from genomics

Links for the "Duke saga"

- The official "Duke saga starter set"
- The economist : "[An array of errors]"
- The 2009 *Annals of Applied Statistics* [paper] by Baggerly and Coombes