



Reproducible Research

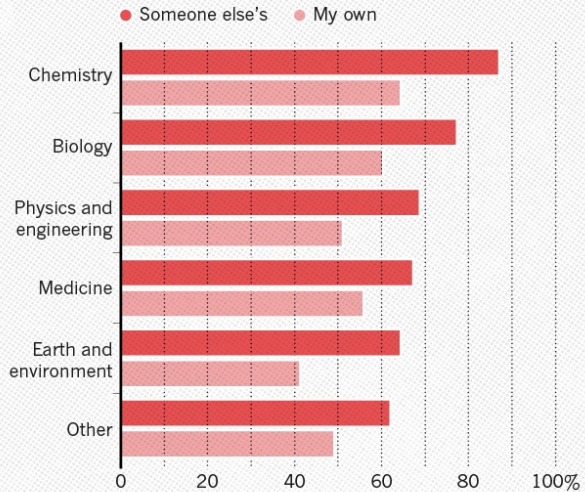
Niklaus Zemp
June 2020

Genetic Diversity Centre (GDC)
Bioinformatics
ETH Zurich

Scientific recipe-publications

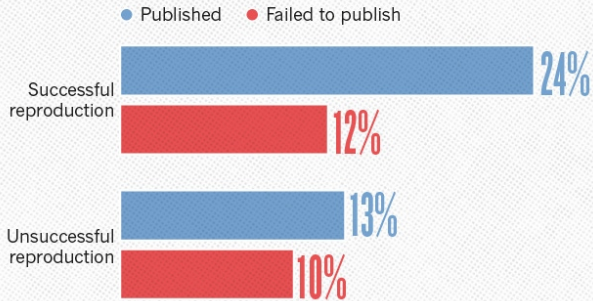
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95,
Medicine 203, Physics and engineering 236, Other 233 ©nature

IS THERE A REPRODUCIBILITY CRISIS?



©nature



Reality check on reproducibility

A survey of *Nature* readers revealed a high level of concern about the problem of irreproducible results. Researchers, funders and journals need to work together to make research more reliable.

25 May 2016



Recipe

Alles Leben strömt aus dir
 Karoline Rudolphi 1754-1811 J.-H. Tobler 1777-1838

Etwas bewegt

S/A

1. Al - les Le - ben strömt aus dir, al - les
 2. Das ich füh - le, was ich bin, das ich
 3. Dei - ner Ge - gen - wart Ge - fühl, dei - ner

T/B

1. Le - ben strömt aus dir und durch - wallt in tau - - send
 2. füh - le, was ich bin, daß ich dich, du Gro - - fier,
 3. Ge - gen - wart Ge - fühl sei mein En - gel, der mich

1. und durch
 2. daß ich
 3. sei mein

1. Bä - chen und durch - wallt in tau - - send Bä - chen al - le
 2. ken - ne, daß ich dich, du Gro - - fier, ken - ne, daß ich
 3. lei - te, sei mein En - gel, der mich lei - te, daß mein

1. wallt in tau - send Bä - chen,
 2. dich, du Gro - fier, ken - ne
 3. En - gel, der mich lei - te,

Ingredients

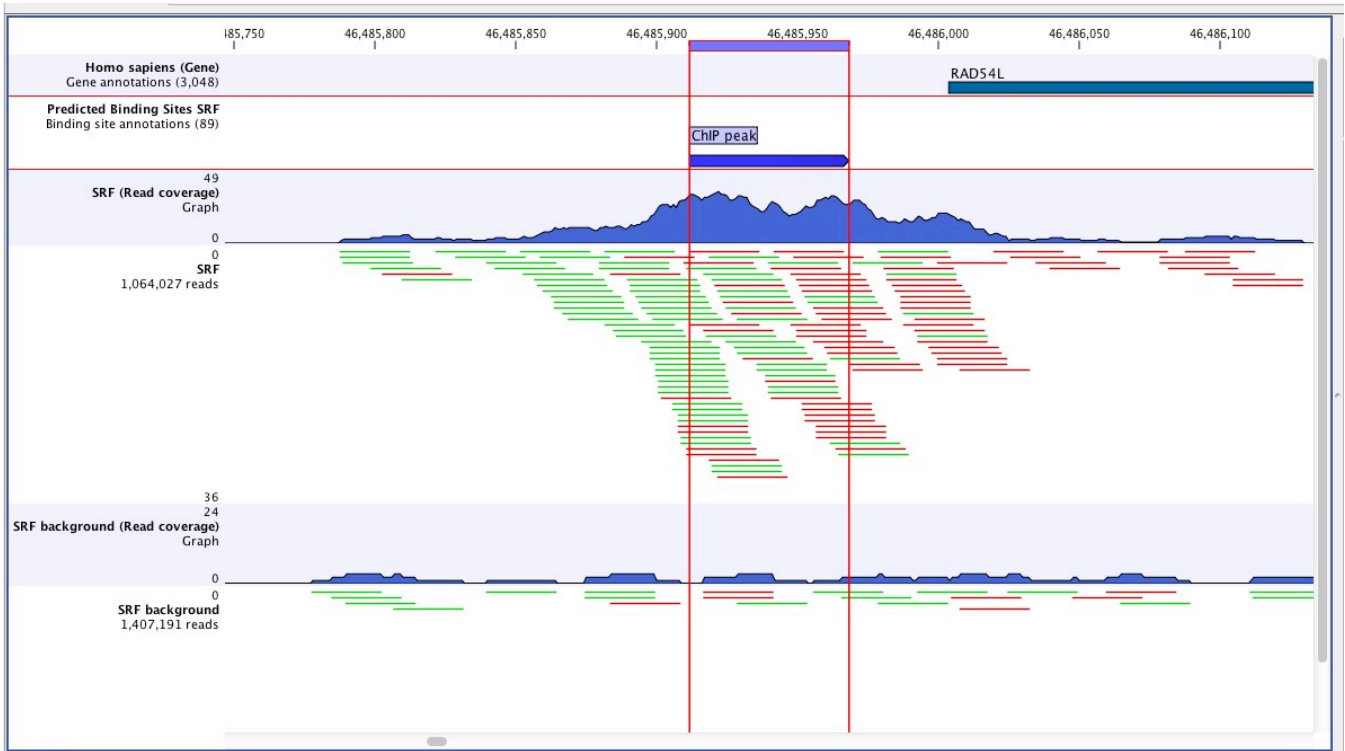
- 3 fresh red chillies
- 2 onions
- 4 cloves of garlic
- 4 large plum tomatoes
- 1 bunch of fresh coriander
- 4 large free-range chicken legs, skin on
- olive oil
- 2 teaspoons garam masala
- 1 tablespoon crumbled dried curry leaves
- 1 tablespoon mustard seeds
- 2 tablespoons white wine vinegar
- fat-free natural yoghurt

Method

1. Halve the chillies (deseed if you like), peel and finely slice the onion and the garlic. Quarter the plum tomatoes, and pick the coriander leaves.
2. Rub the chicken legs all over with a drizzle of oil and the garam masala and place in a large non-stick ovenproof pan.
3. Add another drizzle of oil and fry the chicken over a medium heat until lovely and dark golden all over. Be brave and let it get really brown to make such a difference to the end result if you get it right at this stage. Drain the fat.
4. At this point, preheat the oven to 180°C/350°F/gas 4.
5. Next, add the curry leaves, mustard seeds, chillies, onion and garlic, stirring often, for 5 minutes, then add the tomatoes and white wine vinegar.
6. Transfer the pan to the oven. Cook, uncovered, for 50 minutes, or until the chicken is cooked through and falling off the bone.
7. Pop the pan on the hob and reduce the liquid until sticky. Scatter with the coriander and serve with the cooling yoghurt. Delicious with rice or couscous.



CGI tools



Chromosome	Region	Name	p-value	Score	FDR	note
chr1	45224888..45224947	ChIP peak	0.00	0.00	4.43E-108	# forward reads : 196, # reverse reads : 222, Region containing reads : 45224467..45225
chr1	45578365..45578433	ChIP peak	3.82E-6	3.82E-6	3.63E-6	# forward reads : 43, # reverse reads : 63, Region containing reads : 45577983..4557871
chr1	46024004..46024099	ChIP peak	9.20E-6	9.20E-6	6.18E-4	# forward reads : 18, # reverse reads : 13, Region containing reads : 46023761..4602445
chr1	46485912..46485968	ChIP peak	5.02E-14	5.02E-14	1.54E-34	# forward reads : 85, # reverse reads : 80, Region containing reads : 46485517..4648639
chr1	46855203..46855264	ChIP peak	0.00	0.00	3.36E-86	# forward reads : 162, # reverse reads : 160, Region containing reads : 46854793..46855
chr1	51539515..51539583	ChIP peak	4.56E-5	4.56E-5	4.10E-22	# forward reads : 53, # reverse reads : 44, Region containing reads : 51539124..5153995



The advertisement for CLC Genomics Workbench features a blue background. At the top left is the 'g_x' logo. Below it, the text 'CLC Genomics Workbench' is written in white. Underneath that, 'GET INTO PC' is written in a stylized, green and white font. At the bottom right is the CLC bio logo. The text 'Download Free Your Desired App' is visible at the bottom left.

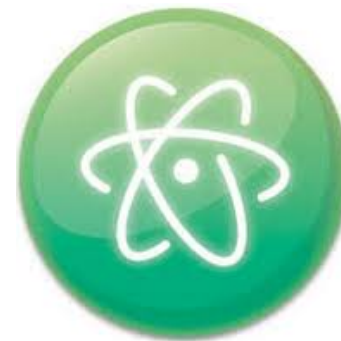


Command-line version

bwa mem Ref reads > alignment.sam



Scientific recipe



Log-file

```
#####
Nik Zemp, niklaus.zemp@env.ethz.ch, GDC, ETH Zurich
#####
####ddRAD log file, Ivo Widmer, p432
#####

#####
####Download data
#####
module load eth_proxy
bsub -n1 -W 4:00 -R "rusage[mem=1000]" \
"wget -r --no-parent --reject="index.htm*" http://gc3fstorage.u
bsub -n1 -W 4:00 -R "rusage[mem=1000]" \
"wget -r --no-parent --reject="index.htm*" http://gc3fstorage.u
bsub -n1 -W 4:00 -R "rusage[mem=1000]" \
"wget -r --no-parent --reject="index.htm*" http://gc3fstorage.u
bsub -n1 -W 4:00 -R "rusage[mem=1000]" \
```

script

```
#!/bin/bash
#BSUB -J "processradtags"
#BSUB -R "rusage[mem=10000]"
#BSUB -n 1
#BSUB -W 24:00

module load gcc/4.9.2 gdc perl/5.18.4
export PATH=$PATH:/cluster/project/gdc/shared/tools/stacks-1.48
#module load gcc/4.8.2 gdc perl/5.18.4 stacks/1.40
source /cluster/apps/gdc/perl5/etc/bashrc

mkdir samples

process_radtags -i gzfastq -f /cluster/project/gdc/people/buckleyj/Hiseq
```



Scientific recipe

- (original) author or source
- your name
- date
- version of the tool
- version of the script
- reproducible code with comments (more comments than code)
- use style guides
- syntax coloring



- keep all scripts and raw data
- use default settings or mention if not
- provide version information
- provide commands in supplementary material
- deposit scripts on github/gitlab
- Check-list

<https://www.nature.com/articles/d41586-019-03959-6>

Comment

THE 'REAPPRAISED' CHECKLIST FOR EVALUATION OF PUBLICATION INTEGRITY

Not all items will be applicable to every publication, and other questions might be relevant for individual categories.

R — Research governance

- Are the locations where the research took place specified, and is this information plausible?
- Is a funding source reported?
- Has the study been registered?
- Are details such as dates and study methods in the publication consistent with those in the registration documents?

- 'P-hacking': biased or selective analyses that promote fragile results
- Other unacknowledged multiple statistical testing
- Is there outcome switching — that is, do the analysis and discussion focus on measures other than those specified in registered analysis plans?

E — Ethics

- Is there evidence that the work has been approved by a specific, recognized committee?

I — Image manipulation

- Is there evidence of manipulation or duplication of images?





Data management

R project

- source ("function1.R")
- source ("function2.R")
- table.csv
- table2.csv
- Data.RData
- analysis1.R
- Analysis.Rmd



R with style

The tidyverse style guide

Hadley Wickham

<https://style.tidyverse.org/index.html>

Google R-style

<https://google.github.io/styleguide/Rguide.xml>



styler

The goal of styler is to provide non-invasive pretty-printing of R source code while adhering to the [tidyverse](#) formatting rules. styler can be customized to format code according to other style guides too.

Installation

You can install the package from CRAN:

```
install.packages("styler")
```

Some examples

```
# Good
day_one
day_1

# Bad
DayOne
dayone
```

```
# Good
x <- 5

# Bad
x = 5
```

```
# Good
if (y < 0) {
  stop("Y is negative")
}
```

```
# Bad
if (y < 0) stop("Y is negative")
```

```
# Good
"Text"
'Text with "quotes"'
'<a href="http://style.tidyverse.org">A link</a>'

# Bad
'Text'
'Text with "double" and \'single\' quotes'
```



Take home message

- Do reproducible research

Outlook:

- Markdown is helpful
- Remember RegEx

