

# Why and when I develop R packages?

Alejandro Reyes  
CBDS workshop (2019)

@areyesq  
[www.alejandre Reyes.org](http://www.alejandre Reyes.org)

# An R package is simply a very well organized directory

```
├── DESCRIPTION
├── PackageHow.Rproj
├── NAMESPACE
├── R
│   └── hello.R
└── man
    └── hello.Rd
```

# Forces you to organize your code into functions

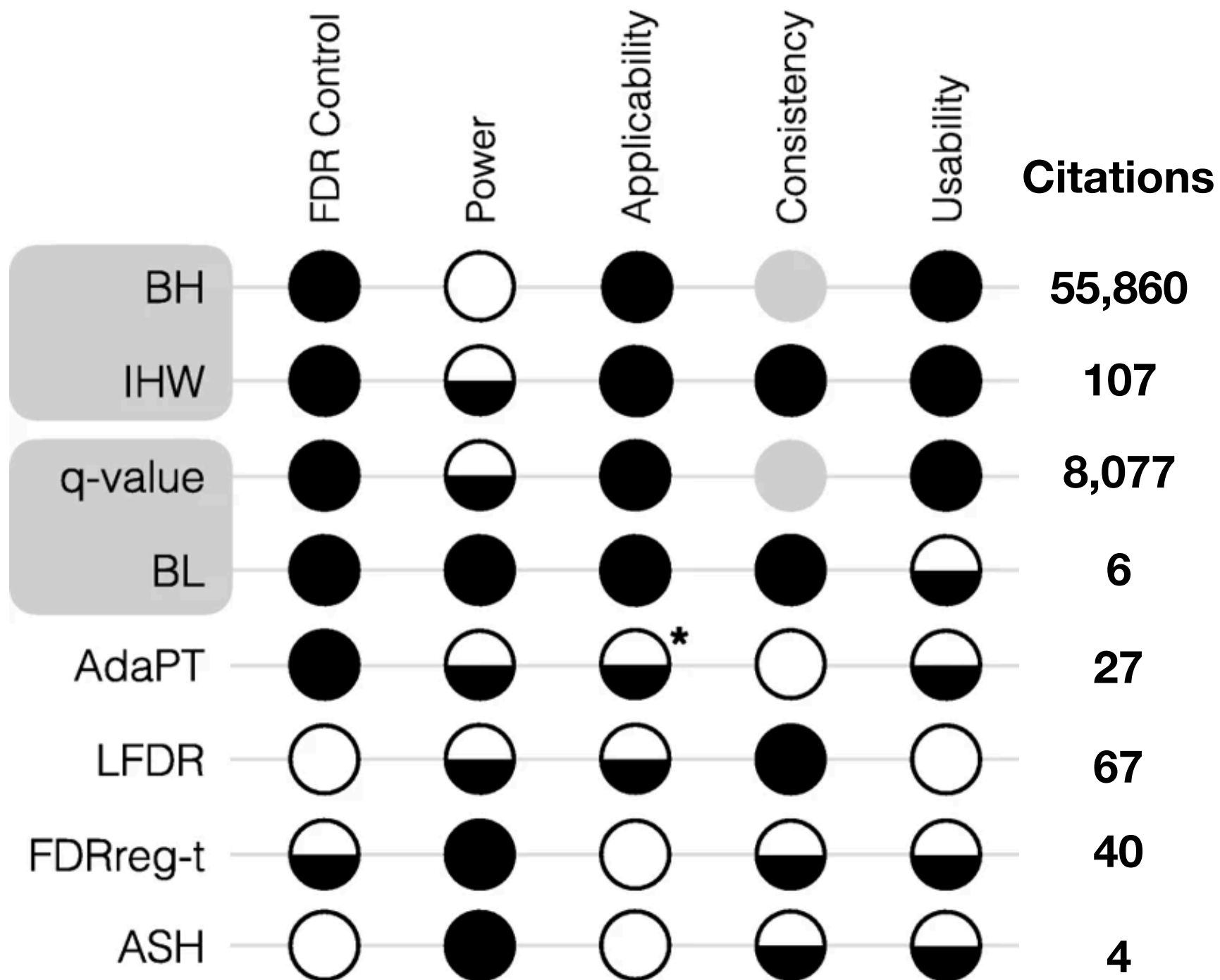
```
d <- dist( t( assay(rld) ) )
mdsResult <- cmdscale( d )
mdsResult <- as.data.frame( mdsResult )
colnames( mdsResult ) <- sprintf( "coord%d", seq_len(ncol(mdsResult)))
mdsResult$condition <- colData(rld)$condition
pl <- ggplot( mdsResult, aes(coord1, coord2, col=condition)) +
  geom_point() + labs(x="Coordinate 1", y="Coordinate 2")
pl
```

```
plotMDS <- function( rld ){
  stopifnot( is( rld, "DESeqTransform" ) )
  d <- dist( t( assay(rld) ) )
  mdsResult <- cmdscale( d )
  mdsResult <- as.data.frame( mdsResult )
  colnames( mdsResult ) <- sprintf( "coord%d", seq_len(ncol(mdsResult)))
  mdsResult$condition <- colData(rld)$condition
  pl <- ggplot( mdsResult, aes(coord1, coord2, col=condition)) +
    geom_point() + labs(x="Coordinate 1", y="Coordinate 2")
  pl
}

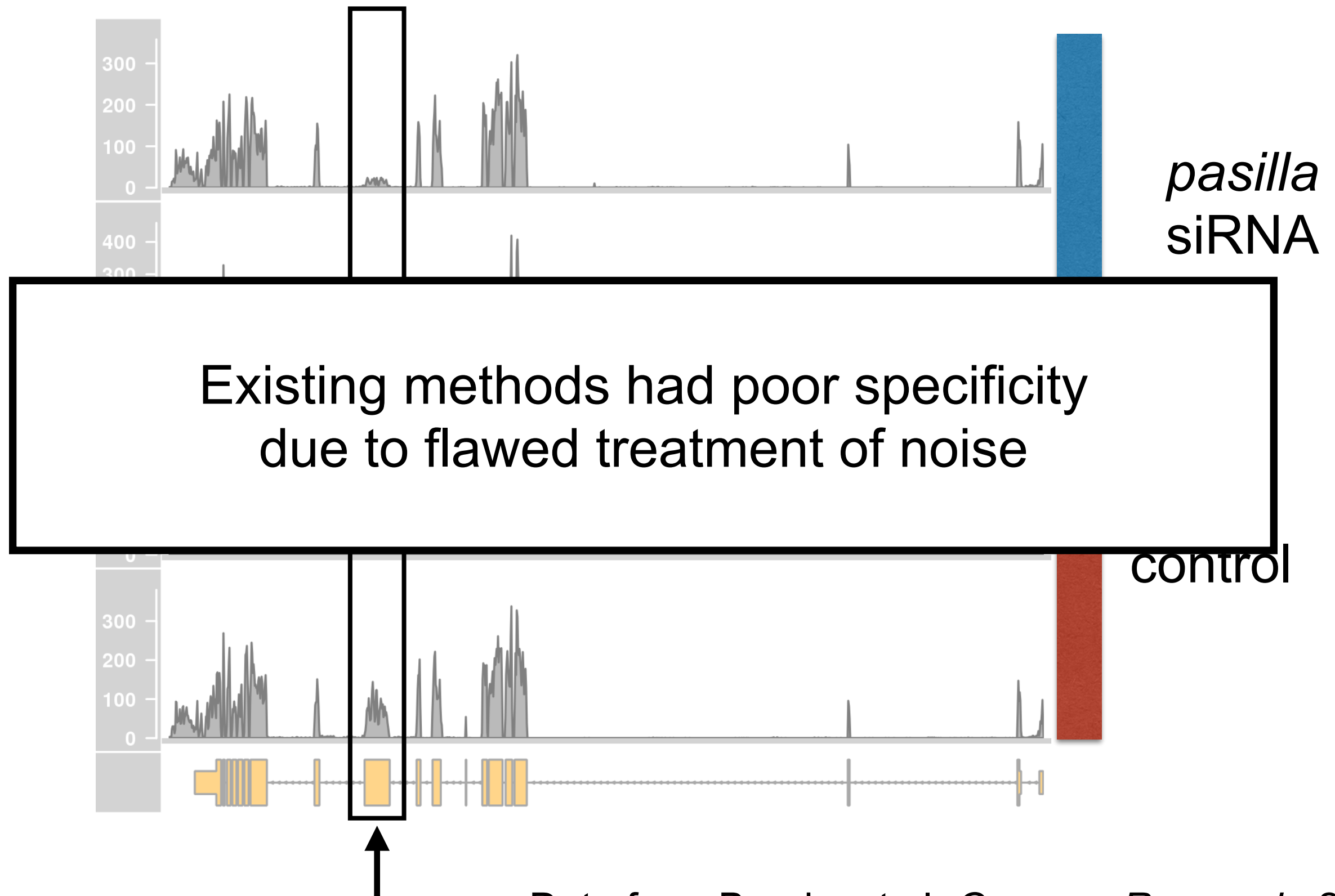
plotMDS(rld)
```

## **Case A: translational methods**

***No matter how awesome your method is, if you don't provide a good software implementation, no one will use it***



# High-throughput RNA sequencing allows an unbiased characterisation of isoform expression

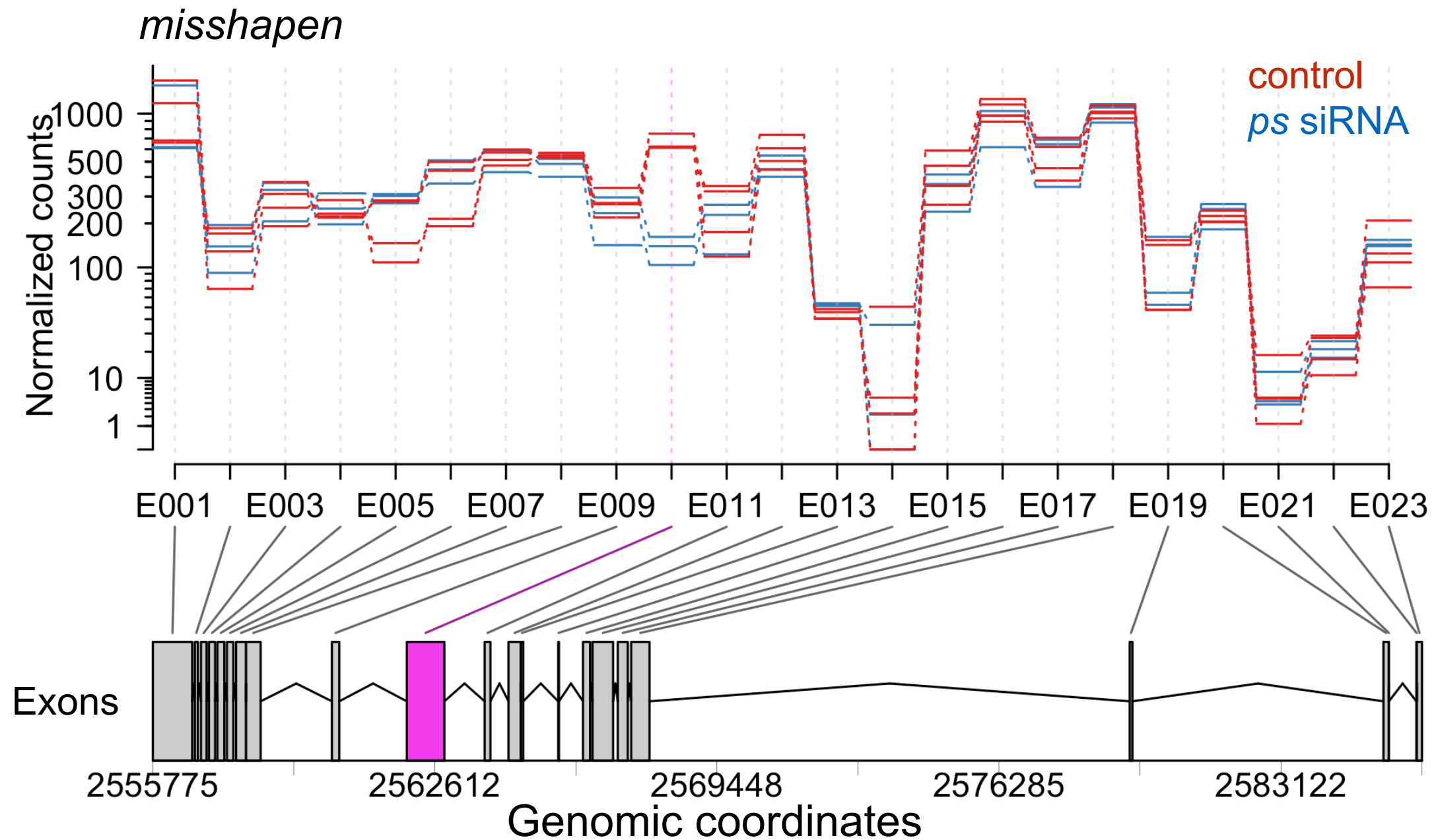


# DEXSeq: inference of differential exon usage

	treated1fb	treated2fb	treated3fb	untreated1fb	untreated2fb	untreated3fb	untreated4fb
E001	1997	494	562	1150	2514	570	547
E002	122	112	180	69	203	156	142
E003	276	293	305	190	398	312	259
E004	420	200	182	230	446	183	185
E005	416	217	279	146	170	237	231
E006	486	357	471	190	337	418	364
E007	574	465	536	469	805	480	496
E008	536	417	447	541	832	475	472
E009	191	237	216	217	427	286	222
E010	188	130	96	617	1177	520	508
E011	165	212	210	118	275	294	269
E012	536	437	414	441	792	619	504
E013	72	41	49	40	76	34	38
E014	3	0	33	5	0	2	42

$$\text{exon usage} = \frac{\text{\# of transcripts including an exon}}{\text{\# total transcripts}}$$

# DEXSeq: inference of differential exon usage





## **Case B: reproducible research**

*Full transparency and reproducibility of published results*

# Most published science is not reproducible

Analysis | Published: 28 January 2009

## Repeatability of published microarray gene expression analyses

John P A Ioannidis✉, David B Allison, Catherine A Ball, Issa Coulibaly, Xiangqin Cui, Aedín C Culhane, Mario Falchi, Cesare Furlanello, Laurence Game, Giuseppe Jurman, Jon Mangion, Tapan Mehta, Michael Nitzberg, Grier P Page, Enrico Petretto & Vera van Noort

*Nature Genetics* **41**, 149–155 (2009) | [Download Citation](#) ⚡

“... Here we evaluated the replication of data analyses in 18 articles ... We reproduced two analyses in principle and six partially or with some discrepancies; **ten could not be reproduced ...**”

# Publication of *R* packages containing data and code to support publications

## Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues

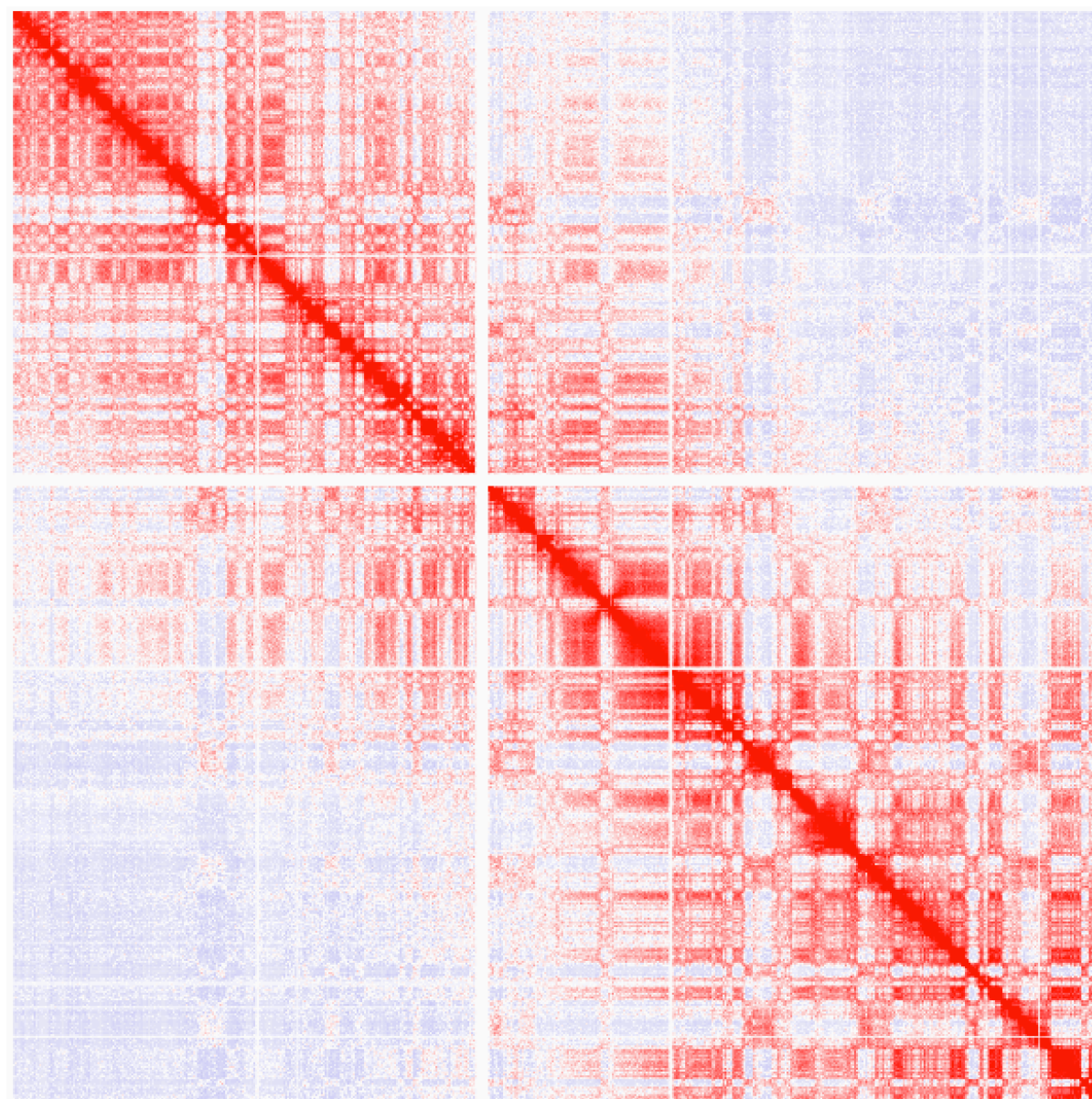
[Alejandro Reyes](#) , [Wolfgang Huber](#) 

*Nucleic Acids Research*, Volume 46, Issue 2, 25 January 2018, Pages 582–592,

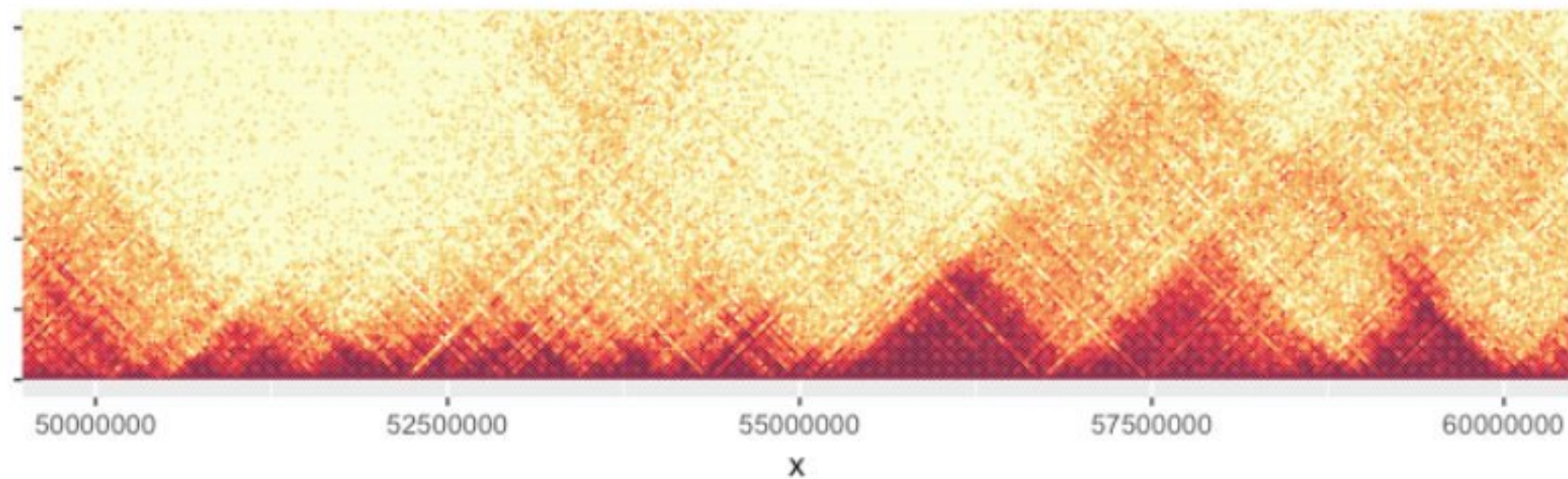
<https://doi.org/10.1093/nar/gkx1165>

**Published:** 30 November 2017    **Article history** ▼

**Case C: organize code into functions, avoid copy/pasting, and make code available to everyone**







**Reuse:**

**Copy/paste**



**Wrote a function**

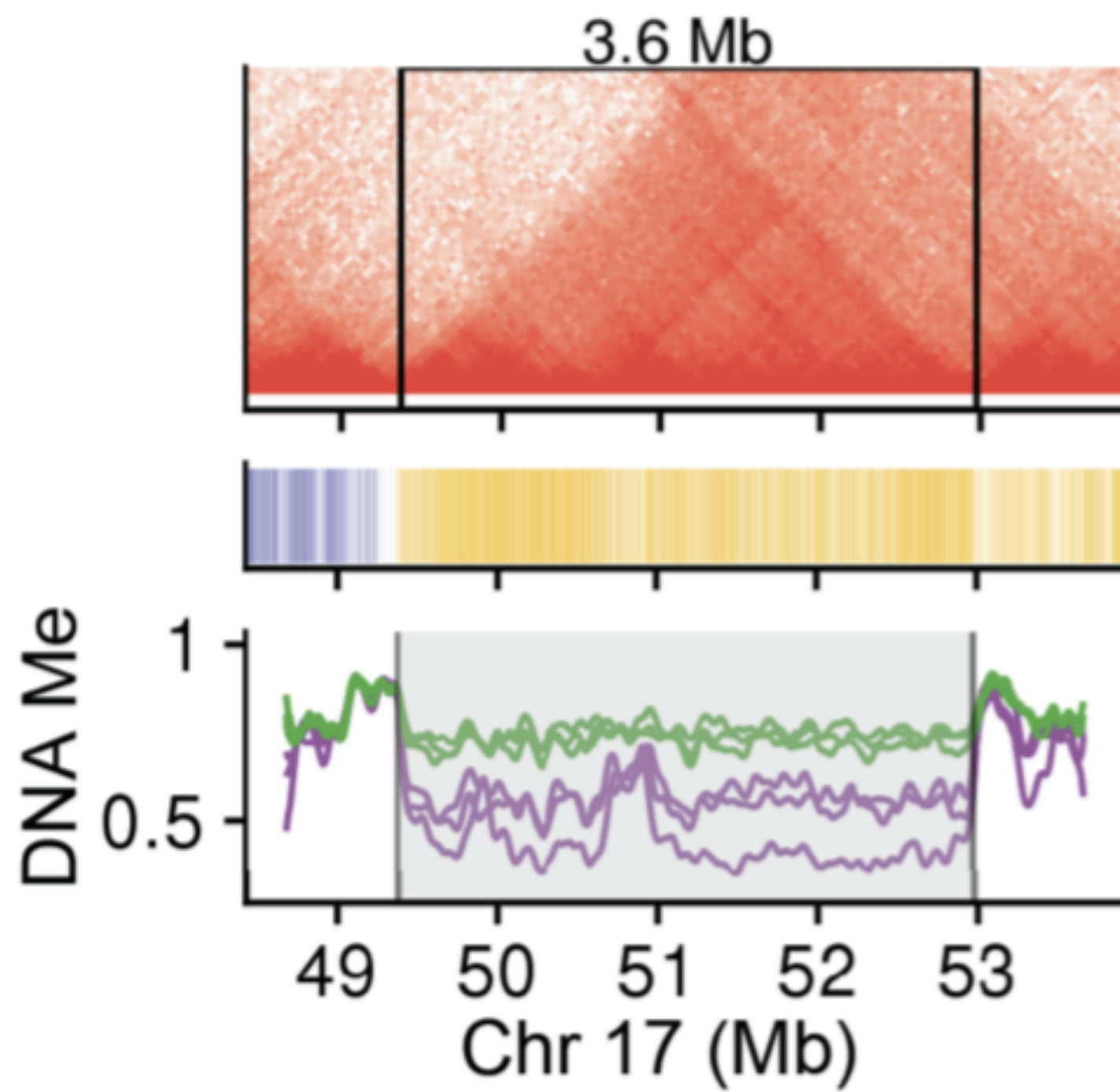
**Share:**

**Share code  
by e-mail**



**Wrote a  
well-documented  
package**

 [areyesq89 / GenomeMatrix](#)



# Enhance the impact of your work!

## Case A: translational methods

☐ [Detecting differential usage of exons from RNA-seq data](#)  
S Anders\*, A Reyes\*, W Huber  
Genome Research

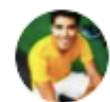
788

2012

## Case B: reproducible research

~2 monthly with questions about the package

## Case C: sharing code



Alejandro Reyes  
@areyesq

Wrote a one-function [#rstats](#) package to rotate  $-45^\circ$  a matrix and plot the upper half using [#ggplot2](#). Useful for plotting HiC data along with other information, such as chromatin marks or gene annotations.

[github.com/areyesq89/Geno...](https://github.com/areyesq89/Geno...)

39 Retweets 140 Likes



Replying to [@areyesq](#)  
Eres un crack!

Jun 22, 2018

1



1





# Discussion