



SOFTWARE TOOL ARTICLE

REVISED regionReport: Interactive reports for region-level and feature-level genomic analyses [version 2; referees: 2 approved, 1 approved with reservations]

Previously titled: regionReport: Interactive reports for region-based analyses

Leonardo Collado-Torres¹⁻³, Andrew E. Jaffe¹⁻⁴, Jeffrey T. Leek^{1,3}

¹Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA

²Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, Maryland, 21205, USA

³Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, 21205, USA

⁴Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, 21205, USA

v2 First published: 01 May 2015, 4:105 (doi: [10.12688/f1000research.6379.1](https://doi.org/10.12688/f1000research.6379.1))
 Latest published: 29 Jun 2016, 4:105 (doi: [10.12688/f1000research.6379.2](https://doi.org/10.12688/f1000research.6379.2))

Abstract

regionReport is an R package for generating detailed interactive reports from region-level genomic analyses as well as feature-level RNA-seq. The report includes quality-control checks, an overview of the results, an interactive table of the genomic regions or features of interest and reproducibility information. regionReport provides specialised reports for exploring DESeq2, edgeR, or derfinder differential expression analyses results. regionReport is also flexible and can easily be expanded with report templates for other analysis pipelines.



This article is included in the **RPackage** channel.



This article is included in the **Bioconductor** channel.

Open Peer Review

Referee Status:

	Invited Referees		
	1	2	3
REVISED			
version 2		report	
published		↑	
29 Jun 2016			
version 1			
published	report	report	report
01 May 2015			

- 1 Timothy J. Triche Jr**, Keck School of Medicine of the University of Southern California USA
- 2 Karthik Ram**, University of California Berkeley USA
- 3 David Robinson**, Princeton University USA

Discuss this article

Comments (0)

Corresponding author: Jeffrey T. Leek (jtleek@gmail.com)

How to cite this article: Collado-Torres L, Jaffe AE and Leek JT. **regionReport: Interactive reports for region-level and feature-level genomic analyses [version 2; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2016, 4:105 (doi: [10.12688/f1000research.6379.2](https://doi.org/10.12688/f1000research.6379.2))

Copyright: © 2016 Collado-Torres L *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: J.T.L. was partially supported by NIH Grant 1R01GM105705, L.C-T. was supported by Consejo Nacional de Ciencia y Tecnología Mexico 351535, AEJ was partially supported by 1R21MH109956.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 01 May 2015, 4:105 (doi: [10.12688/f1000research.6379.1](https://doi.org/10.12688/f1000research.6379.1))

REVISED Amendments from Version 1

We would like to thank the reviewers for their excellent reviews and valuable feedback. We have taken into consideration all of the reviews and comments, and have improved the software, updated the software vignettes, and added more figures to the manuscript. In particular we

- merged David Robinson's pull request, added the `customCode` argument, provide code for making histograms instead of density plots which is used in one of the example use cases, and requested that quotation marks are typeset in a way that makes them easy to copy-paste into R;
- made it easier to use and install dependencies via a new function `load_install()` based on Karthik Ram's feedback, added a link to the bibliography file;
- modified the manuscript and added more example use cases to motivate users to use `regionReport`, particularly with `DESeq2` or `edgeR` results which expands the software's usefulness to a broader user base.

This manuscript matches the software release from Bioconductor version 3.3 which uses by default `rmarkdown`'s version 0.9.5 capabilities. We considered using `git2r` and have not added support for so far given that (A) we mostly use `regionReport` in git-controlled directories that are not on the `gh-pages` branch and (B) that some of the reports can be large in size. However, we might add a function for uploading the resulting report to GitHub via `git2r` in the future.

See referee reports

Introduction

Many analyses of genomic data result in regions along the genome that associate with a covariate of interest. These genomic regions can result from identifying differentially bound peaks from ChIP-seq data¹, identifying differentially methylated regions (DMRs) from DNA methylation data², or performing base-resolution differential expression analyses using RNA sequencing data^{3,4}, among other analysis pipelines. The genomic regions themselves are commonly stored in a `GRanges` object from `GenomicRanges`⁵ when working with R or the BED file format on the [UCSC Genome Browser](#)⁶. Other information on these regions, for example summary statistics on the magnitude of effects and statistical significance, also provide useful information and can be stored as metadata in `GRanges` objects. The usage of R in genomics is increasingly common due to the usefulness and popularity of the Bioconductor project⁷, and in the latest version (3.3), 300 unique packages use `GenomicRanges` for many workflows, demonstrating the widespread utility of identifying and summarizing characteristics of genomic regions.

Bioconductor is particularly strong for differential expression analyses, with 206 packages using the [Differential Expression](#) `BiocView`. RNA-seq data is commonly used to perform feature-level analyses at either the transcript, gene or exon levels with Bioconductor packages `DESeq2`⁸ and `edgeR`⁹⁻¹¹, among others. The features can also be expressed regions identified in an annotation-agnostic procedure by `derfinder`³. In an exploratory data analysis of `DESeq2` or `edgeR` results it is common to create a set of plots in order to identify potentially problematic samples or features. For example, in such an exploratory analysis it is common to use a

dimension reduction technique such as principal component analysis to determine if samples are clustering by group or another variable of interest. This type of plot is useful for detecting artifacts, such as mislabeling of samples.

Here we introduce `regionReport` which allows users to explore genomic regions of interest, `derfinder`, `DESeq2`, and `edgeR` results through interactive stand-alone HTML reports that can be shared with collaborators. These reports are flexible enough to display plots and quality control checks within a given experiment, but can easily be expanded to include custom visualizations or text describing the main conclusions of the exploratory analysis. The resulting HTML report emphasizes reproducibility of analyses¹² by including all the R code without obstructing the resulting plots and tables. Alternatively, static PDF reports can be generated and easily shared among collaborators. We envision `regionReport` will provide a useful tool for exploring and sharing genomic region-based, `DESeq2`, and `edgeR` results from high throughput genomics experiments.

Methods

Implementation

The package includes R Markdown templates which are processed using `rmarkdown`¹³ and `knitr`¹⁴ to produce HTML or PDF reports. HTML reports can be styled using `knitrBootstrap`¹⁵ or with `rmarkdown` templates that include interactive features. The `regionReport` package generates a report that includes a series of plots for checking the quality of the results and an interactive table with the best regions or features. Each element of the report has a brief explanation, although actual interpretation of the results is dataset- and workflow-dependent. To facilitate navigation a menu is included, which is useful for users interested in a particular section of the report. [Figure 1A](#) shows the menu of the general report for a set of regions with associated p-values. The code for each plot or table is hidden by default and can be shown by clicking on the "code" button as shown in [Figure 2](#). Further customization of the reports can be done by providing custom code, changing the default plots, or by modifying the R Markdown templates included in `regionReport`.

General region report

Quality checks

This section of the report includes a variety of quality control steps which help the user determine whether the results are sensible. The quality control steps explore:

- P-values, Q-values, and FWER adjusted p-values
- Region width
- Region area: sum of single-base level statistics (if available)
- Mean coverage or other score variables (if available)

A combination of density plots and numerical summaries are used in these quality checks. If there are statistically significant regions, the distributions are compared between all regions and the significant ones. For example, the distribution region widths might have a high density of small values for the global results, but shifted towards higher values for the subset of significant regions as shown in [Figure 3](#).

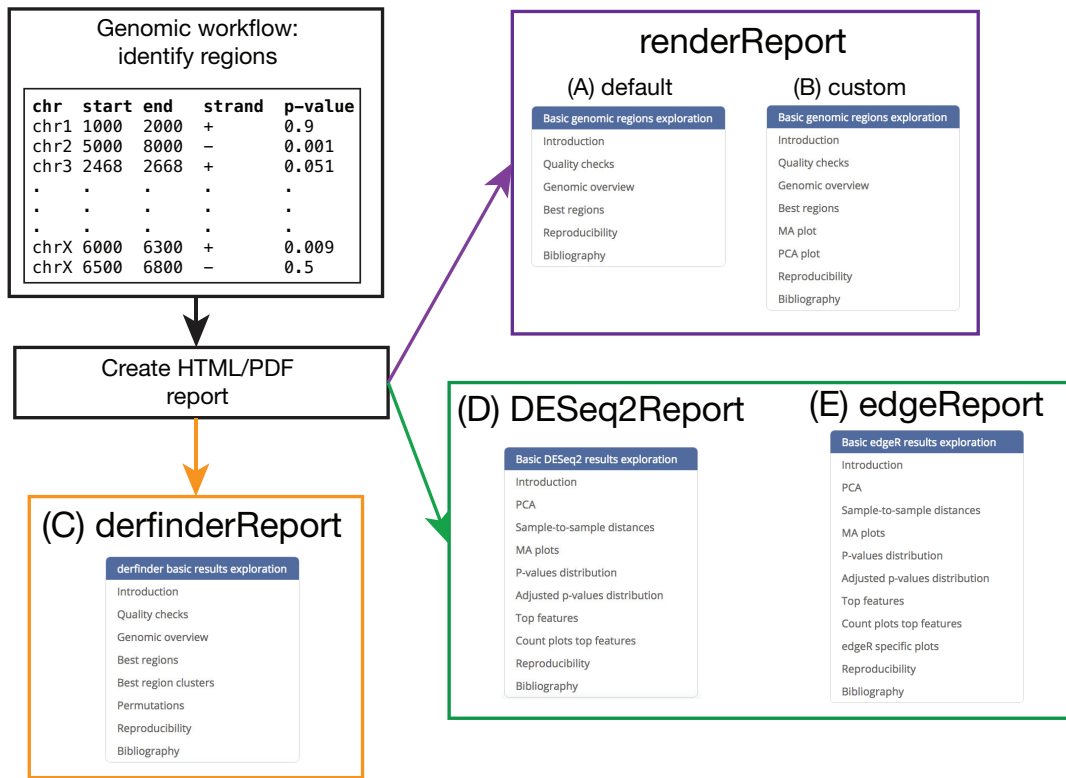


Figure 1. regionReport overview. Example region input, the appropriate regionReport function to use, and menu of the resulting report for: (A) the general use case, (B) a customised report, (C) derfinder results, (D) DESeq2 results and (E) edgeR results.

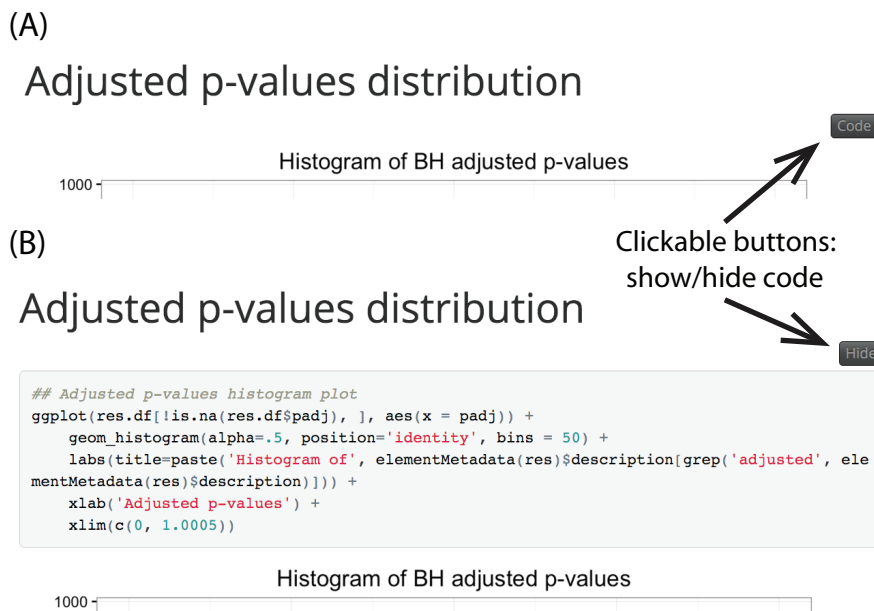


Figure 2. Interactively display the code for each table/figure in the report. (A) View by default and (B) after clicking on the “code” toggle for a section in the report and the HTML reports include a toggle to hide/show all the R code.

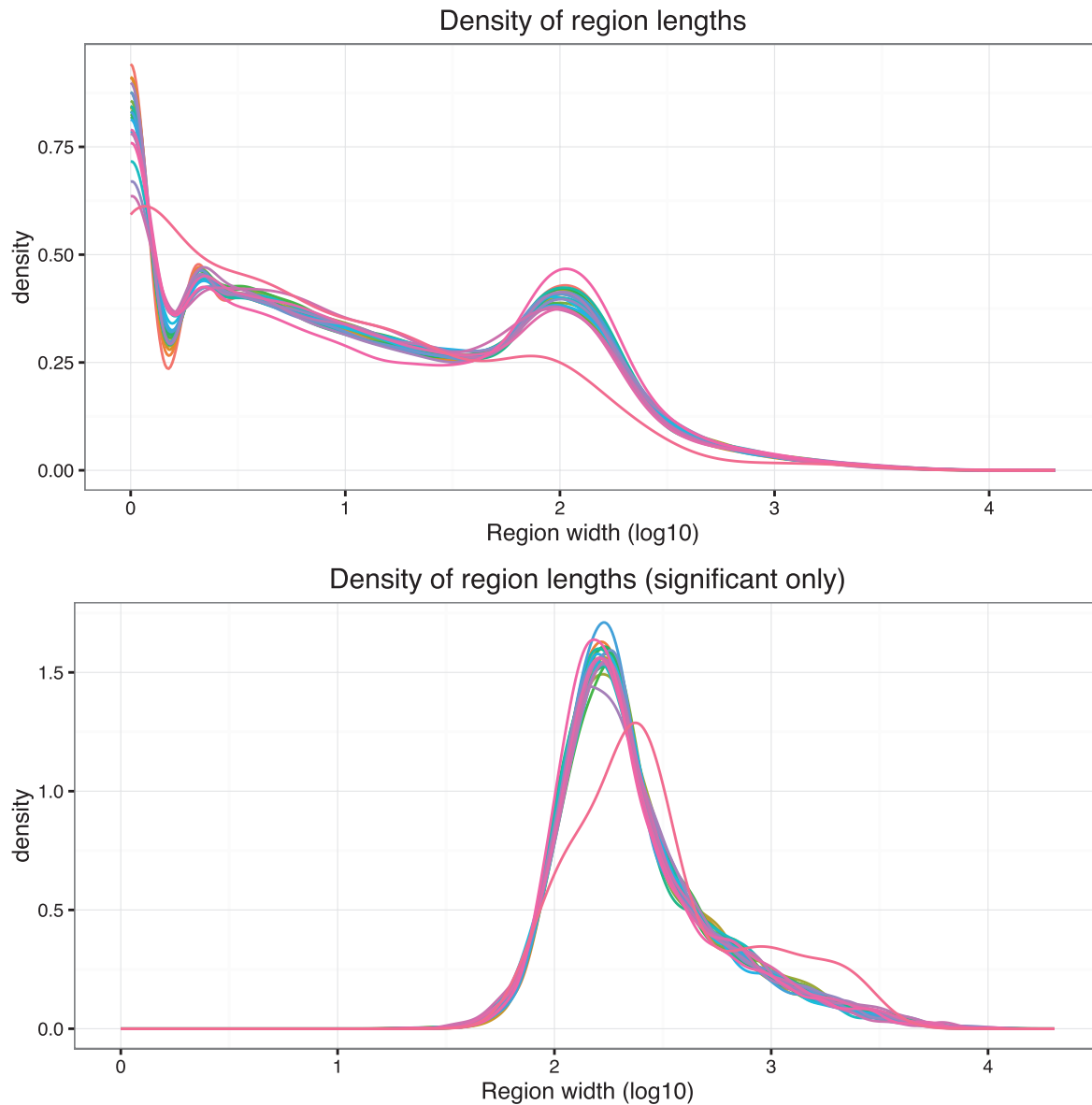


Figure 3. Distribution of region widths for all regions in the *derfinder* use case example with the *BrainSpan* dataset. The top figure shows the region width distribution for all regions while the bottom one shows it only for the significant regions. One line is shown per chromosome in each of the plots.

Genomic overview

The report includes plots to visualize the location of all the regions as well as the significant ones. Differences between them can reveal location biases. The nearest known annotation feature for each region is summarized and visually inspected in the report. This type of plot can be useful to quickly check whether significant regions are concentrated in a chromosome or in an annotation type. For example, [Figure 4](#) shows the annotation information for the

significant regions with most regions contained inside genes, which is expected with RNA-seq data.

Best regions

An interactive table with the top regions (500 by default) is included in this section as shown in [Figure 5A](#). This allows the user to sort the region information according to their preferred ranking option. For example, lowest p-value, longest width, chromosome, nearest

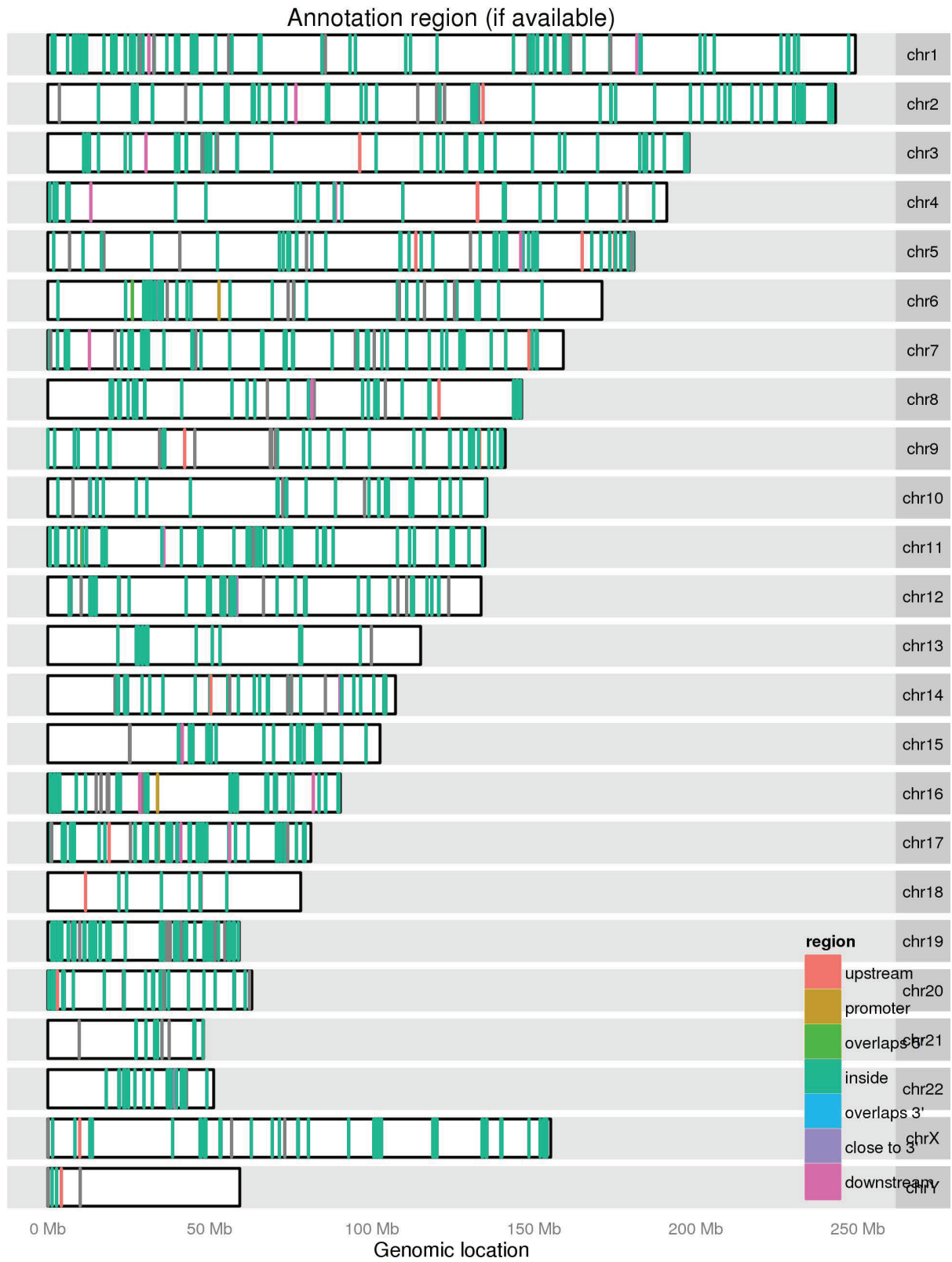


Figure 4. Genomic overview of the annotation type for the significant regions in the *dexfinder* use case example with the *Hippo* dataset.

(A) Show entries Search:

seqnames	start	end	width	strand	area	value	cluster	L	clusterL	nar
chr1	2316	2631	316	*	-1.58	15.81	2	10	29.00	DDX1
chr2	451	551	101	*	1.59	4.77	3	3	20.00	FAM1
chr1	456	526	71	*	1.07	3.20	1	3	29.00	DDX1
chr1	2176	2211	36	*	0.78	1.57	2	2	29.00	DDX1
chr1	2841	2841	1	*	1.20	1.20	2	1	29.00	DDX1

(B) Show entries Search:

value	cluster	L	clusterL	name	annotation	description	region	distance	s
15.81	2	10	29.00	DDX11L1	NR_046018	upstream	upstream	9243	
3.20	1	3	29.00	DDX11L1	NR_046018	upstream	upstream	11348	
1.57	2	2	29.00	DDX11L1	NR_046018	upstream	upstream	9663	
1.20	2	1	29.00	DDX11L1	NR_046018	upstream	upstream	9033	
0.78	1	1	29.00	DDX11L1	NR_046018	upstream	upstream	11103	
0.77	2	1	29.00	DDX11L1	NR_046018	upstream	upstream	9628	
0.62	1	1	29.00	DDX11L1	NR_046018	upstream	upstream	11243	
0.61	1	1	29.00	DDX11L1	NR_046018	upstream	upstream	11873	
0.58	2	1	29.00	DDX11L1	NR_046018	upstream	upstream	8963	
0.52	2	1	29.00	DDX11L1	NR_046018	upstream	upstream	9068	

Showing 1 to 10 of 10 entries (filtered from 15 total entries) First Previous Next Last

Figure 5. Interactive table with results for the top regions in the general use case example using bumpHunter results. The interactive table can (A) show all the top regions of (B) a subset of the results by using the search box. The table can also be sorted by each of the different columns.

annotation feature, etc. The table also allows the user to search and subset it interactively as shown in [Figure 5B](#). A common use case is when the user wants to check if any of the regions are near a known gene of their interest.

Reproducibility

At the end of the report, detailed information is provided on how the analysis was performed. This includes the actual function call to generate the report, the path where the report was generated, time spent, and the detailed R session information including package versions of all the dependencies. An example is shown in [Figure 6](#) with the R package information truncated.

The R code for generating the plots and tables in the report is included in the report itself, thus allowing users to manually reproduce any section of the report, customize them, or simply change the graphical parameters to their liking.

Customization

`regionReport` allows users to customize the reports to their liking. This can be done in different ways depending on the amount of customization the user is looking for. Several plots are made with `ggplot2` and the user might want to change the default theme, for example to a black and white theme as shown in the function call in [Figure 6](#). Another user might be interested in adding

Reproducibility

The input for this report was generated with DESeq2 (Love, Huber, and Anders, 2014) using version 1.11.42 and the resulting features were called significantly differentially expressed if their BH adjusted p-values were less than `alpha = 0.1`. This report was generated in path `/Users/lcollado/Dropbox/JHSPH/Code/regionReportSupp` using the following call to `DESeq2Report()`:

```
## DESeq2Report(dds = dds, project = "DESeq2 HTML report", intgroup = c("condition",
##   "type"), outdir = "DESeq2-example", output = "index", theme = theme_bw())
```

Date the report was generated.

```
## [1] "2016-04-12 07:31:25 EDT"
```

Wallclock time spent generating the report.

```
## Time difference of 22.351 secs
```

R session information.

```
## Session info -----
## -----
```

```
## setting value
## version R version 3.3.0 alpha (2016-03-23 r70368)
## system x86_64, darwin13.4.0
## ui X11
## language (EN)
## collate en_US.UTF-8
## tz America/New_York
## date 2016-04-12
```

•••

```
## rtracklayer 1.31.10 2016-04-07 Bioconductor
## S4Vectors * 0.9.46 2016-04-07 Bioconductor
## scales 0.4.0 2016-02-26 CRAN (R 3.3.0)
## stringi 1.0-1 2015-10-22 CRAN (R 3.3.0)
## stringr 1.0.0 2015-04-30 CRAN (R 3.3.0)
## SummarizedExperiment * 1.1.23 2016-04-06 Bioconductor
## survival 2.38-3 2015-07-02 CRAN (R 3.3.0)
## VariantAnnotation 1.17.23 2016-04-07 Bioconductor
## XML 3.98-1.4 2016-03-01 CRAN (R 3.3.0)
## xtable 1.8-2 2016-02-05 CRAN (R 3.3.0)
## XVector 0.11.8 2016-04-06 Bioconductor
## yaml 2.1.13 2014-06-12 CRAN (R 3.3.0)
## zlibbioc 1.17.1 2016-03-19 Bioconductor
```

Pandoc version used: 1.17.0.3.

Figure 6. Reproducibility section for a report using DESeq2 results. The reproducibility information includes the actual function call used to generate the report, the path where the report was generated, the time it took to create the report, details about the R session information, and the pandoc version used for rendering the HTML report. For reports based on DESeq2 results, the version used to perform the differential expression analysis and cutoff used are also displayed. Note that DESeq2 version used for the analysis and for the report might differ.

code that creates more plots than the ones included by default in the report. For example, the user might be interested in adding a MA and a PCA plot to the default report. This can be done via the `customCode` argument which results in new sections added to the menu as shown in [Figure 1B](#) compared to [Figure 1A](#). Further customization can be achieved by modifying the templates included in `regionReport` and using the `template` argument.

derfinder report

When exploring `derfinder` results from the single base-level approach, for each of the best 100 (default) DERs a plot showing the coverage per sample is included in the report. These plots allow the user to visualize the differences identified by `derfinder` along known exons, introns and isoforms. The plots are created using `derfinderPlot`¹⁶. Due to the intrinsic variability in RNA-seq coverage data or mapping artifacts, in situations where there are two candidate DERs that are relatively close there might be reasons to consider them a single candidate DER and its important to visualize them. This tailored report groups candidate DERs into clusters based on a distance cutoff. After ranking them by their area, for the top 20 (default) clusters it plots tracks with the coverage by sample, the mean coverage by group, the identified candidate DERs colored by whether they are statistically significant,

and known alternative transcripts as shown in [Figure 7](#). [Figure 1C](#) shows the main categories of the report generated from a richer region data set than in the general case.

DESeq2 and edgeR reports

Feature-level differential expression analyses result in a set of features (genes, exons) with a p-value for each feature. To perform such analyses, some phenotype information about the samples is usually available. With this information, you can explore the raw data to identify potentially problematic samples using principal component analysis and sample distance plots. You can also explore the results and check the features marked as differentially expressed with MA plots and a histogram of the p-values distribution. `regionReport` provides a template that allows you to create all these plots easily for DESeq2 results ([Figure 1D](#)). It has similar components to the region-level reports such as an interactive table for the top features as shown in [Figure 8](#), but also highlights specific exploratory plots for this type of results. `regionReport` can also be used for edgeR results ([Figure 1E](#)) resulting in very similar reports given the internal implementation. The only difference is that reports for edgeR results include sections for visualizing the biological coefficient of variation and the multidimensional scaling plot of distances between feature

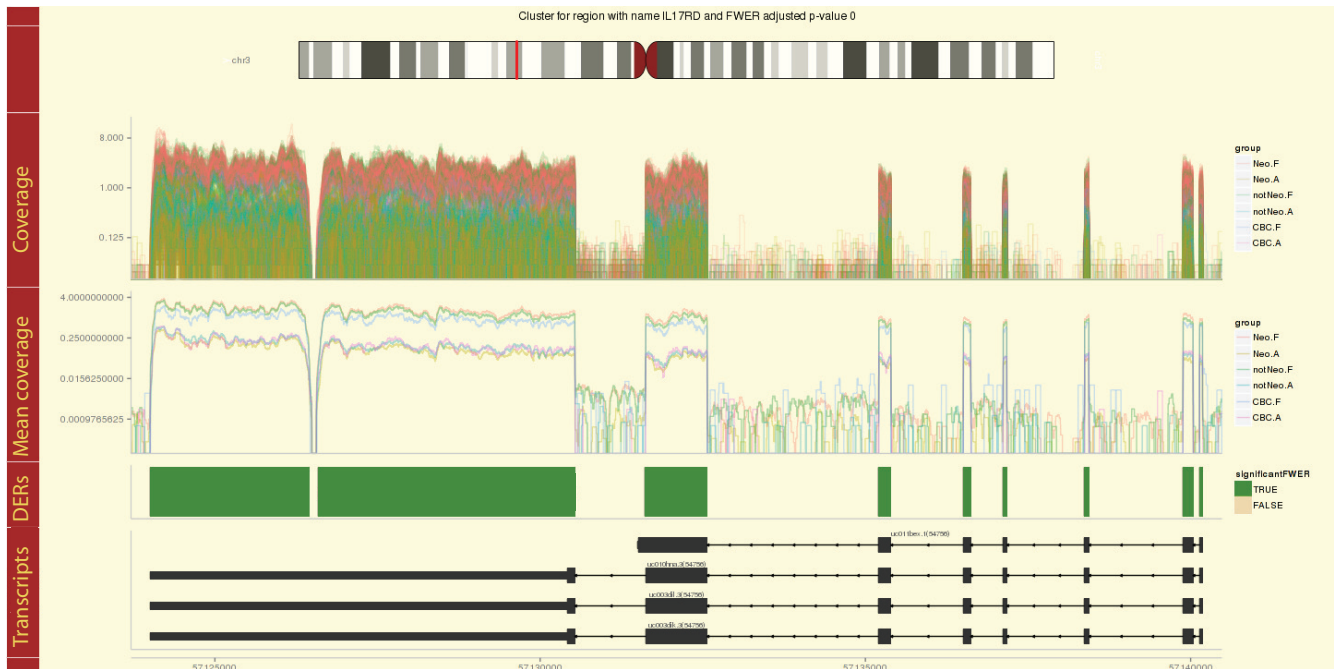


Figure 7. Example region cluster plot for the `derfinder` use case example with the *BrainSpan* dataset. Coverage curves are shown for each sample colored by their group membership. Mean coverage curves by group, differentially expressed regions (DERs) and known transcripts are shown in the remaining tracks.

Top features

This interactive table shows the top 500 features ordered by their BH adjusted p-values. Use the search function to find your feature of interest or sort by one of the columns.

[Code](#)

Show entries Search:

Feature	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
FBgn0039155	453.28	3.71	0.16	23.21	4.013291e-119	3.380797e-115
FBgn0029167	2165.04	2.08	0.10	20.10	6.684454e-90	2.815492e-86
FBgn0035085	366.83	2.23	0.14	16.26	1.888618e-59	5.303239e-56
FBgn0029896	257.90	2.21	0.16	13.91	5.854593e-44	1.232977e-40
FBgn0034736	118.41	2.57	0.18	13.88	8.067448e-44	1.359204e-40
FBgn0040091	610.60	1.43	0.12	11.91	1.114552e-32	1.564831e-29
FBgn0000071	180.01	-2.14	0.18	-11.81	3.498482e-32	4.210174e-29
FBgn0011260	140.36	-1.96	0.17	-11.61	3.723734e-31	3.921092e-28
FBgn0034434	76.91	2.39	0.21	11.26	2.045205e-29	1.914312e-26
FBgn0001226	686.72	-1.53	0.14	-11.10	1.310312e-28	1.103807e-25

Showing 1 to 10 of 500 entries First Previous 1 2 3 4 5 ... 50 Next Last

Figure 8. Interactive table for top features from the DESeq2 use case example.

expression profiles. See the use cases for example reports from DESeq2 and edgeR results.

Operation

Installation. `regionReport` and required dependencies can be easily installed from Bioconductor with the following commands:

```
source("http://bioconductor.org/biocLite.R")
biocLite("regionReport")
```

Input. To generate the report, the user first has to identify the regions of interest according to their analysis workflow. For example, by performing bump hunting to identify DMRs with `bumphunter`. The report is then created using `renderReport()` which is the main function in this package as shown in [Figure 1A,B](#).

For the `derfinder` use case, the `derfinderReport()` function creates the recommended report that includes visualizations of the coverage information for the best regions and clusters of regions. Similarly `DESeq2Report()` and `edgeReport()` create reports for DESeq2 and edgeR results, respectively.

Output. A small example can be generated using:

```
example("renderReport", "regionReport", ask=FALSE)
```

The resulting HTML file will open in the users default browser when using R in an interactive session. Note that alternative output formats such as PDF files can also be generated, although they are not as dynamic and interactive as the HTML format.

Use cases

The supplementary website contains reports using `DiffBind`, `bumphunter`, `derfinder`, `DESeq2`, and `edgeR` results. The `derfinder` use case is illustrated with data sets previously described³ with a moderately sized data set (25 samples), and a large data set with 484 samples. We encourage you to explore the following example reports:

- general **HTML** report example using `bumphunter` results,
- customized general **HTML** report using `DiffBind` results with histograms instead of density plots,

- DESeq2 [HTML](#) and [PDF](#) reports,
- edgeR [HTML](#) and [PDF](#) reports using the custom `ggplot2` theme `theme_linedraw()`,
- edgeR-robust [HTML](#) report,
- [HTML](#) report using `derfinder` results with the *BrainSpan* dataset (484 samples) and styled with `knitrBootstrap`,
- [HTML](#) report using `derfinder` results with the *Hippo* dataset (25 samples) and styled with `knitrBootstrap`.

Summary

`regionReport` creates interactive reports from a set of regions and can be used in a wide range of genomic analyses. Reports generated with `regionReport` can easily be extended to include further quality checks and interpretation of the results specific to the data set under study. These shareable documents are very powerful when exploring different parameter values of an analysis workflow or applying the same method to a wide variety of data sets. The reports allow users to visually check the quality of the results, explore the properties of the genomic regions under study, and inspect the best regions and interactively explore them.

Furthermore, `regionReport` promotes reproducibility of data exploration and analysis. Each report provides R code that can be used as the starting point for other analyses within a dataset. `regionReport` provides a flexible output for exploring and sharing results from high throughput genomics experiments.

Software availability

Software access

`regionReport` is freely available via Bioconductor at Bioconductor.org/packages/regionReport. The supplementary website <http://leekgroup.github.io/regionReportSup/> hosts the code and

output for generating all the use cases described. Versions of all software used are included in the reports.

Latest source code

The latest source code is available at github.com/leekgroup/regionReport. However, we highly recommend users to install `regionReport` directly from Bioconductor at bioconductor.org/packages/regionReport.

Archived source code as at the time of publication

Archived source code available at dx.doi.org/10.5281/zenodo.55274

License

Artistic-2.0.

Author contributions

L.C.-T. conceived and developed the `regionReport` package, supervised by A.E.J. and J.T.L. All authors wrote and approved the final manuscript.

Competing interests

No competing interests were disclosed.

Grant information

J.T.L. was partially supported by NIH Grant 1R01GM105705, L.C.-T. was supported by Consejo Nacional de Ciencia y Tecnología México 351535, AEJ was partially supported by 1R21MH109956.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to acknowledge Michael I. Love for his feedback and input in creating the report specific to DESeq2 results.

References

1. Stark R, Brown G: *DiffBind: differential binding analysis of ChIP-Seq peak data*. 2011. [Reference Source](#)
2. Jaffe AE, Murakami P, Lee H, *et al.*: *Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies*. *Int J Epidemiol*. 2012; 41(1): 200–209. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Torres LC, Frazee AC, Love MI, *et al.*: *derfinder: Software for annotation-agnostic RNA-seq differential expression analysis*. *bioRxiv*. 2015; 015370. [Publisher Full Text](#)
4. Frazee AC, Sabuncyan S, Hansen KD, *et al.*: *Differential expression analysis of RNA-seq data at single-base resolution*. *Biostatistics*. 2014; 15(3): 413–26. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Lawrence M, Huber W, Pagès H, *et al.*: *Software for computing and annotating genomic ranges*. *PLoS Comput Biol*. 2013; 9(8): e1003118. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Rosenbloom KR, Armstrong J, Barber GP, *et al.*: *The UCSC Genome Browser database: 2015 update*. *Nucleic Acids Res*. 2015; 43(Database issue): D670–D681. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

7. Huber W, Carey VJ, Gentleman R, *et al.*: **Orchestrating high-throughput genomic analysis with Bioconductor**. *Nat Methods*. 2015; **12**(2): 115–121.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biol*. 2014; **15**(12): 550.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics*. 2010; **26**(1): 139–140.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. McCarthy DJ, Chen Y, Smyth GK: **Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation**. *Nucleic Acids Res*. 2012; **40**(10): 4288–4297.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Zhou X, Lindsay H, Robinson MD: **Robustly detecting differential expression in RNA sequencing data using observation weights**. *Nucleic Acids Res*. 2014; **42**(11): e91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Sandve GK, Nekrutenko A, Taylor J, *et al.*: **Ten simple rules for reproducible computational research**. *PLoS Comput Biol*. 2013; **9**(10): e1003285.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Xie Y: **Dynamic Documents with R and knitr**. CRC Press, 2013; 216.
[Reference Source](#)
14. Rstudio: **RMarkdown: Dynamic Documents for R**. 2014.
[Reference Source](#)
15. Hester J: **Knitr Bootstrap framework, R package**. 2015.
[Reference Source](#)
16. Collado-Torres L, Jaffe AE, Leek JT: **derfinderPlot: Plotting functions for derfinder**. 2015.
[Reference Source](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 30 June 2016

doi:[10.5256/f1000research.9717.r14687](https://doi.org/10.5256/f1000research.9717.r14687)



Timothy J. Triche Jr

Jane Anne Nohl Division of Hematology, USC/Norris Comprehensive Cancer Center, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA

The authors have gone far beyond my expectations in addressing the lack of visual aids and examples in the original manuscript; the (new?) DiffBind example, as well as the excerpted figures from the BrainScan, clearly show strengths that derFinder and regionReport bring to bear when compared to existing tools.

All of my previous reservations (and then some) have been thoroughly addressed and I believe the resulting manuscript is a strong argument for the authors' toolchain.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 22 June 2015

doi:[10.5256/f1000research.6840.r8559](https://doi.org/10.5256/f1000research.6840.r8559)



David Robinson

Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA

The authors present regionReport, an R package to produce interactive HTML reports from a genomic-region based analysis, such as those produced by derfinder, bumphunter or DiffBind. The report shows quality control summaries, interactive tables of the most significant regions, and information on how regions lie in exons, introns, and intergenic regions. Both novice and expert genomicists are sure to gain much from this, and the paper clearly and concisely explains the software and its use, while providing useful examples and instructions.

The idea of producing common reports for a Bioconductor object is ingenious, and will hopefully inspire packages for other types of biological data. One of the great strengths of the package is the reproducibility practices it follows. For example, the section at the end of the produced report that shows reproducibility

information, such as the original command, the session info, and the amount of time the report took to generate, is a great idea. (Indeed, the option to add sessionInfo() and timers could probably be baked into rmarkdown, or a thin wrapper thereof). Another strength is the use of modern knitr templates, such as expandable tables. Scientists who want to develop automated reports should use this package as a guide.

Overall my concerns are minor, and mostly concern the package rather than the paper, some of which I attempt to address in a [GitHub pull request](#).

In pull request

- If the renderReport function leaves early (for example, if it is interrupted by the user hitting Stop) it strands the user's R session in a working directory. Using the on.exit function, as described [here](#), lets R return to the original directory instead.
- The options for customization of the report are limited, by the customCode argument, to chunks between the main text and the reproducibility section. Genomicists may wish to take advantage of these reports while customizing some of their outputs. (For example, the authors of region-finding packages may wish to wrap renderReport with a customized template for their own objects). I've added a template argument in my pull request, and go over another suggestion below.

Not in pull request

- The `template` argument is a start towards greater customization, but a further improvement would be to allow the user to provide a list of customized internal chunks (for example, density-pvalue). As it is now, these are constructed in the renderReport function and cannot be altered without rewriting the entire function. This suggests finding a way to abstract them, such as bringing them in from a separate file, would be useful.
- As one example of an important customization I'd make: the reports show density plots of p-values and q-values, but in my experience genomicists are more accustomed to histograms (especially since bumps in density plots may be misleading, while histograms can get a better sense of which bumps are meaningful). I understand if the authors wish to keep it as a density plot, but if so I would appreciate a way to change it for my own use.

Minor issues

- The use of "smart quotes" in code within the PDF, such as source ("http://bioconductor.org/biocLite.R"), make it inconvenient to copy and paste them into an R terminal. If there's any way this could be remedied by the author or editors, it should.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 11 June 2015

doi:[10.5256/f1000research.6840.r8558](https://doi.org/10.5256/f1000research.6840.r8558)



Karthik Ram

Berkeley Institute for Data Science, University of California Berkeley, Berkeley, USA

Review of regionReport: Interactive reports for region based analysis.

This short software tool article describes a new R package, `regionReport`, available from Bioconductor that generates HTML reports which allows users to explore genomic regions and quickly scan quality control information. The reports also provide provenance of code used in the analysis, including detailed session information to facilitate as much reproducibility as possible.

The paper/report clearly describes functionality of the tool, potential use cases, and details on installation and operation.

Suggestions for improvement

1. Given that you are describing an HTML application, it would be really helpful to include more screenshots/figures rather than just the one, and the workflow diagram. Most readers are unlikely to install the package immediately (see further comments on 3) and so it would help to make the value proposition clear. I would also suggest annotating these figures highlighting the key parts. Happy to approve the narrative itself after this revision.
2. Given that the package primarily generates html based reports, it would be of great value to have these files in the `gh-pages` branch of a GitHub repo, such that reports could be automatically made available under `https://USERNAME.github.io/repo/file.html` much like the page that describes the supplementary material. One way to easily enable this would be to use the functionality in the `git2r` package (disclosure: I am a coauthor on the package) to programmatically create a new branch (if it doesn't already exist), generate the report, then add those files and push to GitHub (assuming the same folder in under git revision control). Obviously I am not expecting the authors to add this suggestion to the current version of the package, but as something to consider for future versions.
3. Reduce the number of dependencies. `locfdr` is no longer on CRAN. On a slightly slower than normal connection (currently on travel) it took a fairly long time to track down and install all the dependencies. I'd recommend moving non-essential dependencies to suggests and using something like this to selectively install packages as needed using `requireNamespace(pkg, quietly = TRUE)`. It was disappointing to go down a rabbit hole of dependencies and still not be able to install and run examples. However, I found the report examples posted online (here: <http://leekgroup.github.io/regionReportSup/bumphunter-example/index.html> and here: <http://leekgroup.github.io/regionReportSup/DiffBind-example/index.html>) extremely useful. The use of Twitter bootstrap also adds a layer of a familiarity that I found extremely useful.
4. It would be nice to have the package generate a direct link to the bib file under the bibliography.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: In my suggestions to improve the software, I've recommended adding one dependency tool on which I am a coauthor. The software itself is free and I don't benefit from any citations (there is no paper associated with that software). In this case it would really help make it easier for researchers to publish their reports generated by the software described in this paper. It did not affect my review and I have offered that addition only as a suggestion (not a requirement to acceptance).

Referee Report 18 May 2015

doi:10.5256/f1000research.6840.r8554



Timothy J. Triche Jr

Jane Anne Nohl Division of Hematology, USC/Norris Comprehensive Cancer Center, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA

Needs more figures to demonstrate why a user would choose this tool. For example <http://leekgroup.github.io/regionReportSupp/bumphunter-example/index.html> (but even better would be to show an example, e.g. ITGB2 exon inclusion/exclusion or multiscale DMRs, where in our hands at least, nothing else short of IGV really does the job, and IGV doesn't do it that well.) The software is a firm foundation but the writeup needs work if it is to be compelling and thus influence readers to try out an unfamiliar tools.

My apologies for being harsh, but without figures, an applied paper simply will not be read. I would be less harsh if the underlying work were not compelling enough to command broader interest. A poor writeup will doom the work to obscurity.

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 18 May 2015

Jeffrey Leek, Johns Hopkins Bloomberg School of Public Health, USA

Thanks, we will update with more figures and expand the description. This is meant to be a short description of the software but we certainly appreciate the feedback on how to increase users. We will update the draft and respond shortly.

Competing Interests: No competing interests were disclosed.