

TECHNICAL NOTE

Open Access

iReport: a generalised Galaxy solution for integrated experimental reporting

Saskia Hiltemann^{1,2*}, Youri Hoogstrate^{1,2}, Peter van der Spek¹, Guido Jenster² and Andrew Stubbs¹

Abstract

Background: Galaxy offers a number of visualisation options with components, such as Trackster, Circster and Galaxy Charts, but currently lacks the ability to easily combine outputs from different tools into a single view or report. A number of tools produce HTML reports as output in order to combine the various output files from a single tool; however, this requires programming and knowledge of HTML, and the reports must be custom-made for each new tool.

Findings: We have developed a generic and flexible reporting tool for Galaxy, iReport, that allows users to create interactive HTML reports directly from the Galaxy UI, with the ability to combine an arbitrary number of outputs from any number of different tools. Content can be organised into different tabs, and interactivity can be added to components. To demonstrate the capability of iReport we provide two publically available examples, the first is an iReport explaining about iReports, created for, and using content from the recent Galaxy Community Conference 2014. The second is a genetic report based on a trio analysis to determine candidate pathogenic variants which uses our previously developed Galaxy toolset for whole-genome NGS analysis, CGtag. These reports may be adapted for outputs from any sequencing platform and any results, such as *omics* data, non-high throughput results and clinical variables.

Conclusions: iReport provides a secure, collaborative, and flexible web-based reporting system that is compatible with Galaxy (and non-Galaxy) generated content. We demonstrate its value with a real-life example of reporting genetic trio-analysis.

Keywords: Galaxy, Generic reporting application, Next generation sequencing, Genetic variation, Pathogenic gene selection

Findings

Structured reporting and documentation of experimental outcome is required for the successful transfer of knowledge from the research scientist to their peers and to the broader academic community.

Galaxy is a platform that aims to provide complex bioinformatics services and tools in an easy-to-use web-based graphical user interface [1-3]. The output from these tools can be displayed using built-in Galaxy visualisation applications [4], via specialised visuals implemented as a component in the workflow deployed in Galaxy [5]

or by downloading the results and visualising the output with applications external to Galaxy (e.g., Excel, TIBCO spotfire, R, spreadsheet programs, etc).

Galaxy has the capacity to track the provenance of the source data, the workflow, as well as the workflow components used to analyse the data. Currently users can share their workflow and results within Galaxy, but do not have access to a simple method to summarise results from multiple tools and/or workflows in an integrated report. To address this issue we have developed iReport, an integrated reporting application that provides users with a flexible means to produce dynamic HTML reports which can be shared with other Galaxy users or downloaded to disk.

Systems used by end-users to deliver graphical output range from open-source applications, such as *Ad Hoc*

*Correspondence: s.hiltemann@erasmusmc.nl

¹Department of Bioinformatics, Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

²Department of Urology, Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

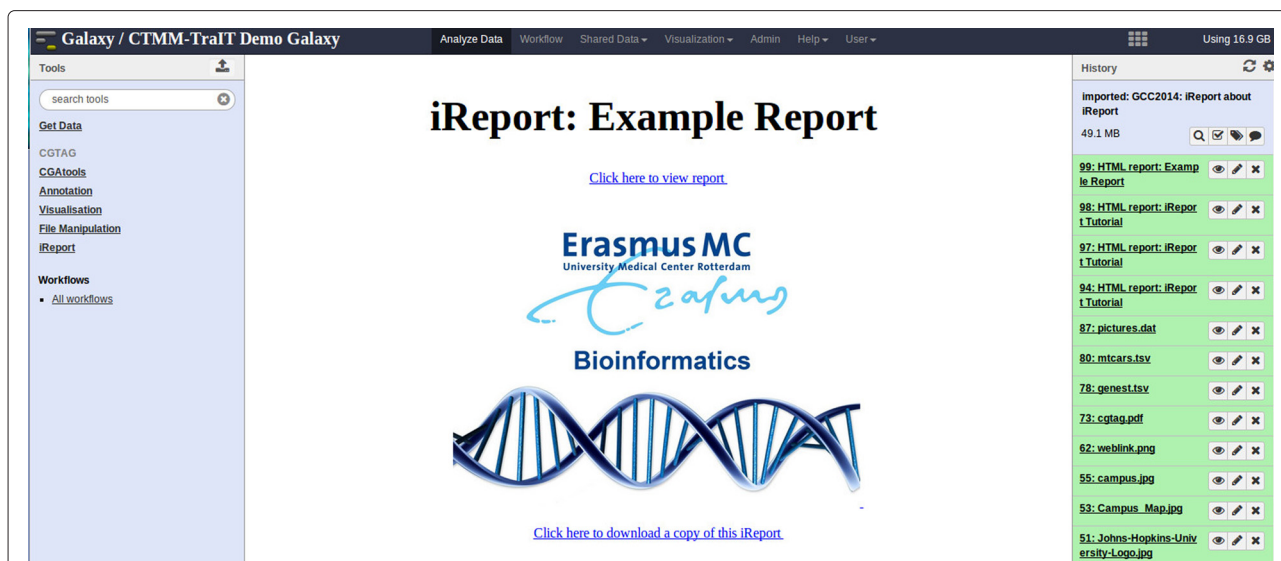


Figure 1 Example cover page. Example of a cover page with title *Example Report* and the default cover image. A link to download the entire iReport web page is also provided.

reports [6], Google charts (and docs) [7] and OpenOffice [8], to commercial applications such as Microsoft Office. Indeed scientific reporting applications both open-source (Bioconductor [9], Circos [10,11]) and commercial software (e.g., Omniviz [12], Partek [13]) include a multitude of visualisation capabilities with a focus on

data reporting and presentation of data in the context of the experimental design and with associated meta-data. There are some applications, like TIBCO spotfire [14], which are capable of integrating results from multiple sources including associated text and meta-data and other applications which serve as an electronic lab note

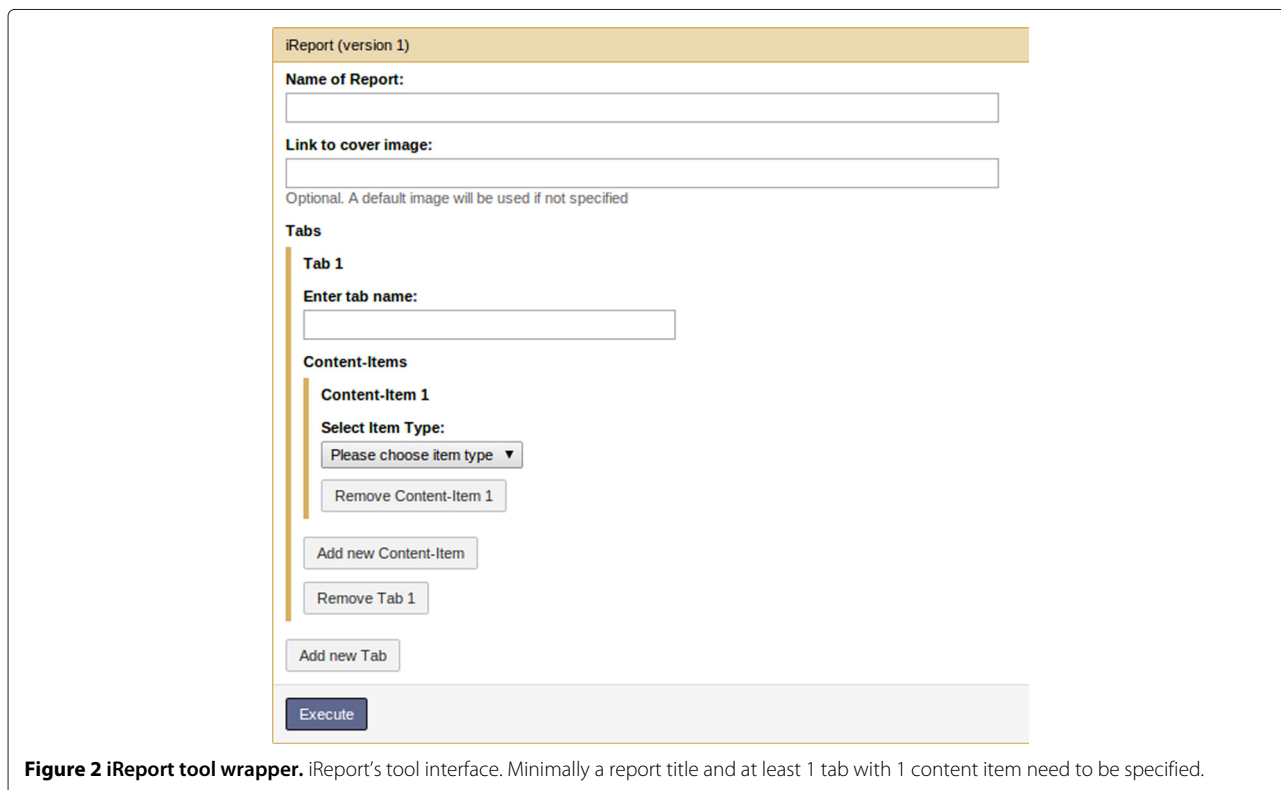


Figure 2 iReport tool wrapper. iReport's tool interface. Minimally a report title and at least 1 tab with 1 content item need to be specified.

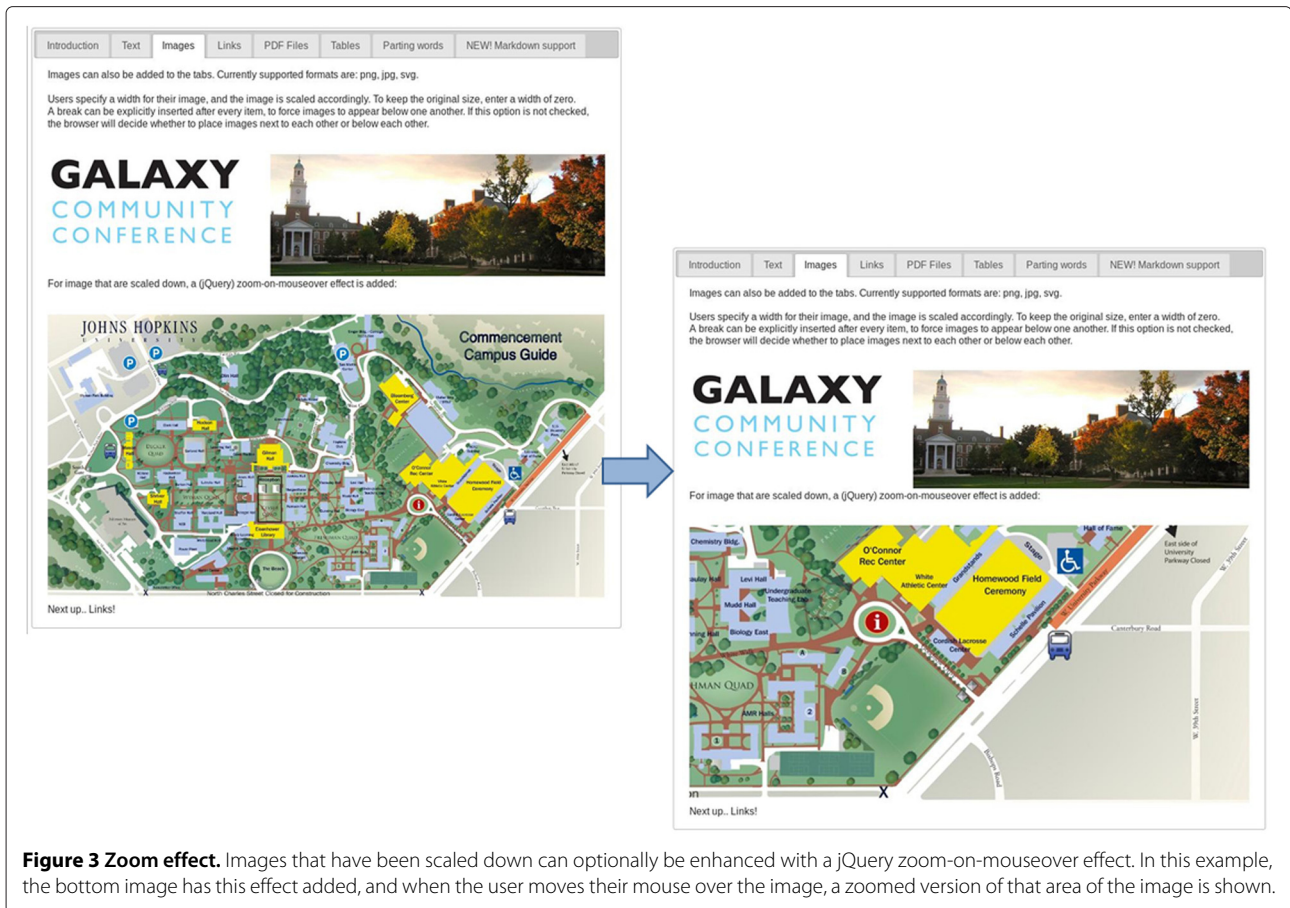


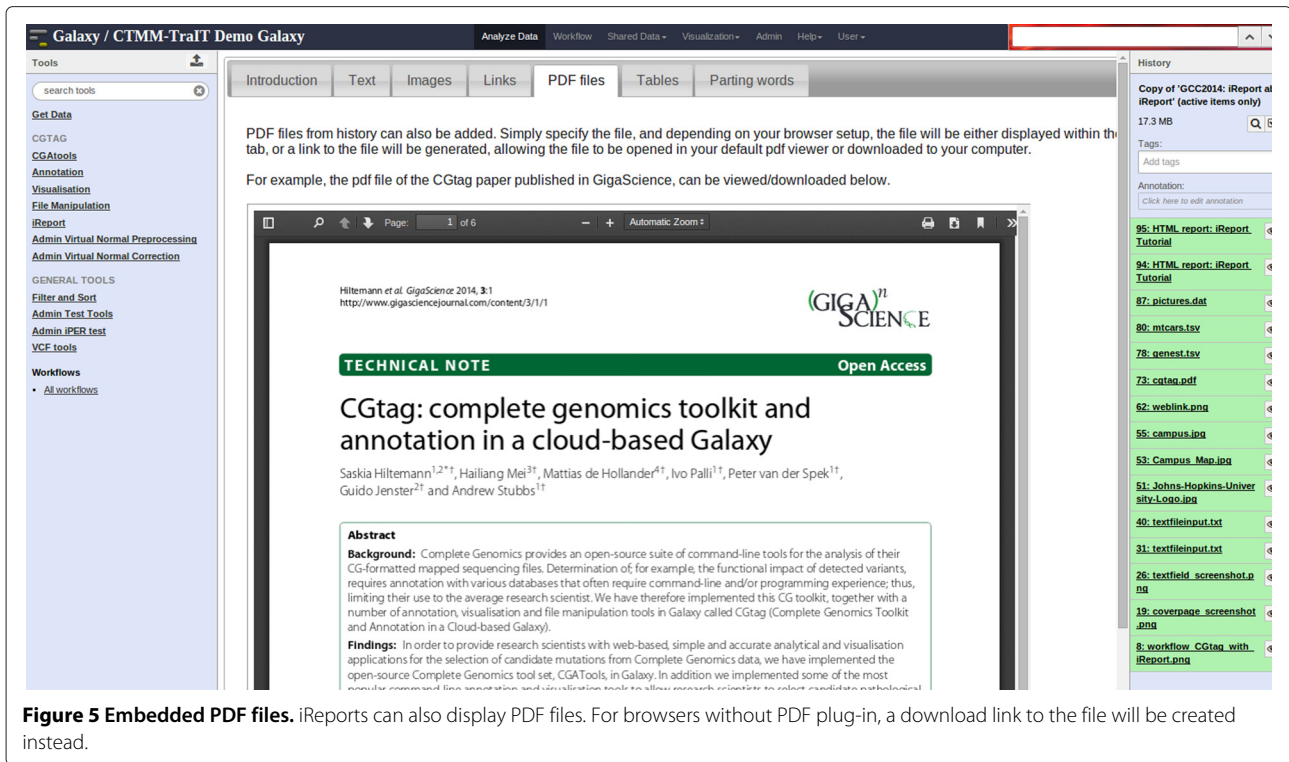
Figure 3 Zoom effect. Images that have been scaled down can optionally be enhanced with a jQuery zoom-on-mouseover effect. In this example, the bottom image has this effect added, and when the user moves their mouse over the image, a zoomed version of that area of the image is shown.

book (e.g., IDBS [15]). Additionally there have been many products developed to address the selection and reporting of variants for pathogenic variant selection including the workflow to identify those variants (e.g., Gensight [16], Cartagenia [17], Clinical Genomicist [18]). For data generated in R, dynamic reporting packages such as KnitR [19], Sweave [20] and R-Markdown [21], allow for the integration of data-generating code within the report specification itself. Similar systems exist for other programming languages, for example Tangle [22] (JavaScript), Active Markdown [23] (CoffeeScript) or IPython Notebooks [24] (Python). These are very versatile tools, but require programming knowledge to use effectively. iReport offers an open-source application for both Galaxy and non-Galaxy produced results allowing for the generation of customized integrated reports for any type of project or workflow. The advantage to Galaxy users is that the output from any application can be included into any report, and that a report template can be reused for other projects. Also, the report may be securely shared with one or many users of that Galaxy instance or made publicly available. iReports can be completely configured from the tool web interface, and requires no programming or knowledge of the underlying system.

We demonstrate iReport's utility through an example where a genetic report is generated from outputs of an existing Galaxy next-generation sequencing (NGS) toolkit, CGtag [5]. iReport can also be used as an electronic lab notebook by creating an iReport which links out to various other iReports containing different analysis reports from various samples. It can be also be coupled to output from other Galaxy instances, for example output generated by specialized Galaxy instances such as Conifero [25], ORIONE [26], and Galaxy-P [27].

Functionality

iReport dynamically generates HTML, and employs JavaScript and jQuery to create interactive components, such as searchable, sortable, paginated tables and zoomable images. iReport is ideally suited to use as the final step in a workflow; the pipeline developer configures the report once and end-users are then presented with a templated report each time users run the workflow, while only needing to provide the input files for the pipeline [28]. iReport can also be used directly by end-users as a means of easily sharing their results with other Galaxy users, or the public via Galaxy's native sharing capabilities.



(Figure 1). By clicking on the image, or the link above it, the user can access the main report page. There is also a link to download the entire iReport webpage, including all dependency files, for storing or viewing on different systems.

Main report page

An arbitrary number of tabs may be added via a repeat parameter. Each tab can be labelled with a name specified by the user. An arbitrary number of *content items* may then be added to each tab in a repeat parameter. A type must be specified for each content item (e.g., text, image, table *etc.*), as well as several other parameters depending on the type chosen (Figure 2). Layout is mostly left up to the browser, but users can explicitly add a line-break after each item to force items to appear underneath each other.

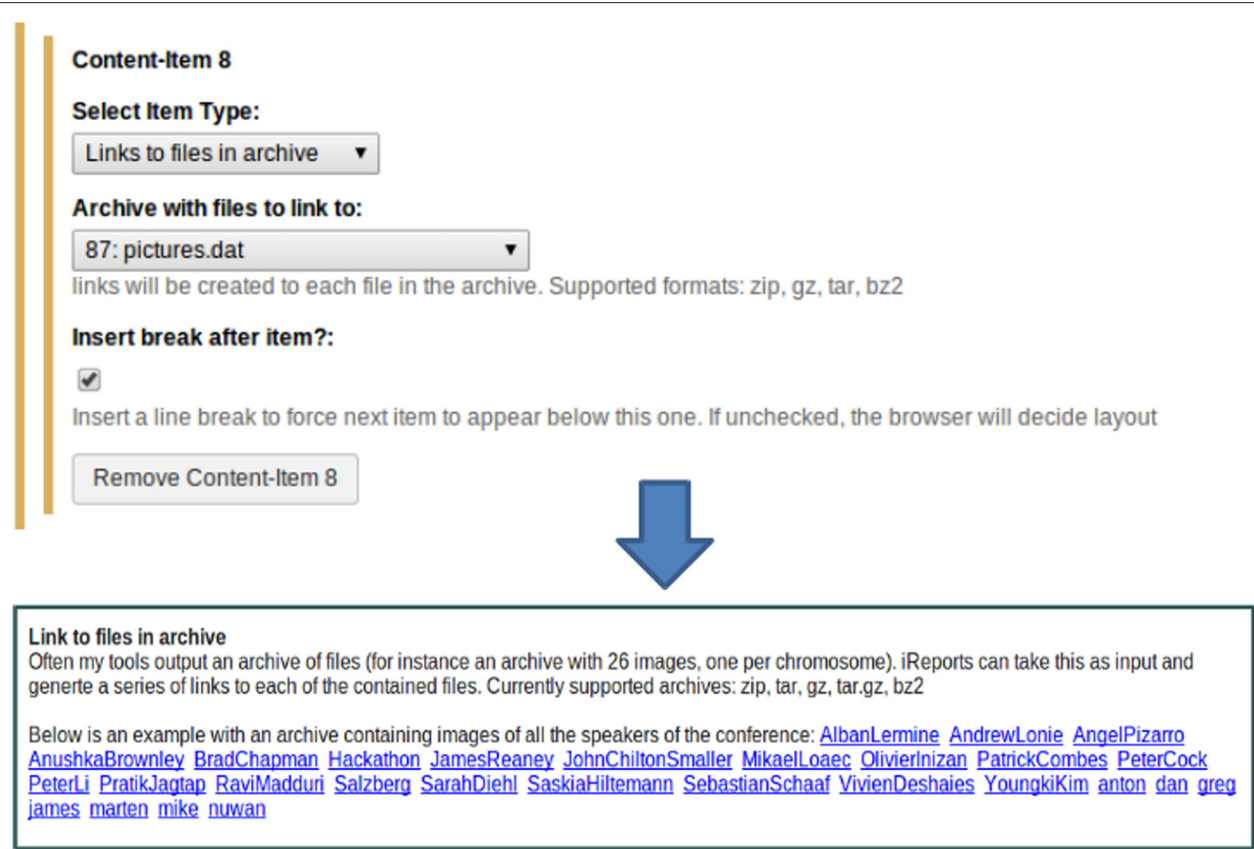
Content item: text field Text can be entered in a text field in the tool interface, for example to create an introduction paragraph and to give a description of the items on the page. Text is printed verbatim, although a small number of HTML tags are allowed in order to give the user some control over formatting (e.g., *b*, *i*, *em*, *strong*, *h1-h6* tags). Text files can also be specified, and the contents of the file will be printed to the screen verbatim.

Content item: images Many tools produce images as output, which can also be displayed by iReport. Users

specify the image file from their Galaxy history, and the desired image size. For images that have been scaled down, an optional jQuery zoom-on-mouseover effect may be added (Figure 3) [31]. Currently supported image formats are JPG, PNG and SVG.

Content item: tables iReport can also display tables. The input must be a tab-delimited file from the users' Galaxy history, and the first nonempty line not starting with a hash symbol (#) is assumed to contain the column headers. The jQuery library *DataTables* [32] is used to create tables which can be searchable, sortable and paginated, if requested by the user. There is an option to create hyperlinks within the columns of a table by providing a column number, a URL prefix and a URL suffix. This is illustrated in Figure 4, where the first column contains gene names and by including the GeneCards [33,34] URL prefix "http://www.genecards.org/cgi-bin/carddisp.pl?gene=". This generates a hyperlink to the corresponding GeneCards entry for every item in the column in the table.

Content item: PDF files This is one of the simplest content items. The user provides a PDF file from the Galaxy history, which will be embedded in the page. If the browser does not have the necessary plug-ins installed, a download link for the file will be generated instead (Figure 5).



Content-Item 8

Select Item Type:
Links to files in archive ▼

Archive with files to link to:
87: pictures.dat ▼
links will be created to each file in the archive. Supported formats: zip, gz, tar, bz2

Insert break after item?:

Insert a line break to force next item to appear below this one. If unchecked, the browser will decide layout

Remove Content-Item 8

Link to files in archive
Often my tools output an archive of files (for instance an archive with 26 images, one per chromosome). iReports can take this as input and generate a series of links to each of the contained files. Currently supported archives: zip, tar, gz, tar.gz, bz2

Below is an example with an archive containing images of all the speakers of the conference: [AlbanLermine](#) [AndrewLonie](#) [AngelPizarro](#) [AnushkaBrownley](#) [BradChapman](#) [Hackathon](#) [JamesReaney](#) [JohnChiltonSmaller](#) [MikaelLoaec](#) [OlivierInizan](#) [PatrickCombes](#) [PeterCock](#) [PeterLi](#) [PratikJagtap](#) [RaviMadduri](#) [Salzberg](#) [SarahDiehl](#) [SaskiaHiltemann](#) [SebastianSchaaf](#) [VivienDeshaies](#) [YoungkiKim](#) [anton](#) [dan](#) [greg](#) [james](#) [marten](#) [mike](#) [nuwan](#)

Figure 6 Links to all files in an archive. Given an archive of files, iReport can create a series of links to all files contained in the archive. Link texts are the filenames (without file extension).

Content item: links Users can create links to web locations by specifying a URL and a link text. Links to datasets in the history can also be created here by specifying a dataset and a link text. Several tools create archives of files as output (for example a zip file containing the plots for each chromosome). Links to all files contained in an archive can also be created, and will be named with the file names (excluding file extension). Currently the supported archive formats are zip, bz2, tar, gz and tar.gz. An example can be seen in Figure 6, where an archive with images was used as input and a series of links to each contained file was created. An option to create a link to an iReport is also present. This allows users to create a kind of electronic lab notebook, by creating an overview of all their samples and linking to one or more iReports for each sample.

Genetic report for a trio of HapMap individuals

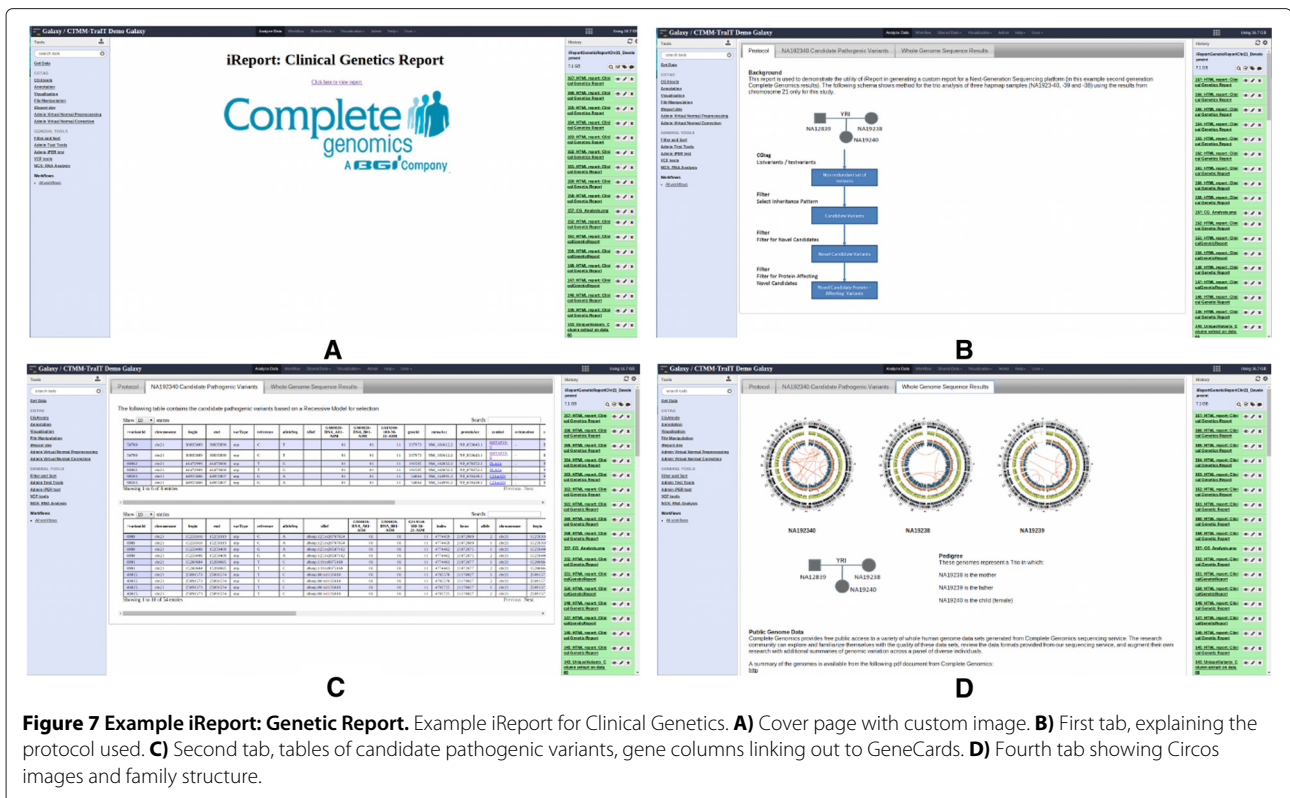
Accurate, reproducible and traceable reporting is an essential requirement to the evaluation of the genetic outcome from any assay [35], including those variations predicted from NGS analysis. Since iReport is capable of including many formats, we have used the outcome from

a trio analysis generated from the Complete Genomics [36] NGS platform to demonstrate its utility in representing these data in a user-defined format, which contains the provenance of the underlying analysis. In this example we use a trio of individuals sequenced in the International HapMap Project [37,38], to demonstrate how to select protein affecting candidate variants based on a recessive genetic model. All data in this example is freely available for download from the Complete Genomics website [39].

This example iReport has one tab devoted to explaining the protocol used (Figure 7B), one tab with circos plots and an explanation of the family structure (Figure 7D), and one tab with tables containing the candidate pathogenic variants determined by the protocol based on a recessive model for selection. This iReport is also available as a published history on the TraIT-CTMM public Galaxy [40].

Conclusions

iReport is a easy-to-use, flexible tool for generating traceable, standardized reports which are easily shared



between users within and across platforms. We have demonstrated that iReport is capable of creating a customised genetics report from results generated within Galaxy and may be shared with collaborators on the same platform, or with the public. Additionally, data or results generated externally can be uploaded into Galaxy and can also be used by iReport. These reports are generated as web pages and may be downloaded in their entirety to be easily shared across systems.

The genetics report presented here represents the bare minimal reporting that is required to summarise the output for a genetic variation analysis. Whilst we used a trio of individuals to demonstrate how to select protein-affecting candidate variants based on a recessive model, any number of model outcomes and other assay results may be included in an iReport.

We developed iReport to simplify reporting and sharing the output from *omics* and non-high throughput assays analysed both in and external to Galaxy. We have also utilised iReport for more complex analysis workflows, such as summarising translational research and diagnostic applications for cancer and immunological research and diagnostics.

Availability and requirements

Project name: iReport

Project home page: <https://github.com/shiltemann/iReport>

CTMM-TraIT public Galaxy instance: galaxy.ctmm-trait.nl

iReport tool shed repository: <https://toolshed.g2.bx.psu.edu/view/saskia-hiltemann/ireport>

Operating system(s): Unix-based Operating Systems

Programming languages: Bash, Perl, Python

Other Requirements: Galaxy

License: GNU GPL

Any restrictions to use by non-academics: none

Examples:

iReport about iReport published history: <http://galaxy.ctmm-trait.nl/u/saskia-hiltemann/h/gcc2014-ireport-about-ireport,ortinyurl.com/llrzz9w>

Clinical Genetics iReport published history: <http://galaxy.ctmm-trait.nl/u/andrew-stubbs/h/ireportgeneticreportchr21>

Availability and supporting data

The iReport tool, user manual (published page), and example data and histories are available at the CTMM-TraIT Galaxy server [40].

Abbreviations

CGtag: Complete genomics toolkit and annotation in a cloud-based galaxy;
 CTMM-TraIT: Center for Translational Molecular Medicine - Translational IT;
 NGS: Next generation sequencing; URL: Uniform resource locator.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SH, GJ, and AS contributed to the design and coordination of iReport and manuscript preparation. SH and AS contributed to implementing iReport. SH, GJ, YH, PvdS and AS contributed to testing of iReport and all authors read and approved the final manuscript.

Acknowledgements

This study was performed within the framework of the Center for Translational Molecular Medicine (CTMM). TraIT project (grant 05T-401). This work was sponsored by the BiG Grid project for the use of the computing and storage facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO).

Received: 30 July 2014 Accepted: 29 September 2014

Published: 13 October 2014

References

1. Goecks J, Nekrutenko A, Taylor J, The Galaxy Team: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**(8):R86.
2. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J: **Galaxy: a web-based genome analysis tool for experimentalists.** *Curr Protoc in Mol Biol* 2010. Chap. 19.10.1–21. [<https://wiki.galaxyproject.org/CitingGalaxy>]
3. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A: **Galaxy: a platform for interactive large-scale genome analysis.** *Genome Res* 2005, **15**(10):1451–1455.
4. Goecks J, Eberhard C, Too T, The Galaxy Team, Nekrutenko A, Taylor J: **Web-based visual analysis for high-throughput genomics.** *BMC Genomics* 2013, **14**:397.
5. Hiltemann S, Mei H, de Hollander M, Palli I, van der Spek P, Jenster G, Stubbs A: **CGtag: Complete Genomics toolkit and annotation in a cloud-based Galaxy.** *GigaScience* 2014, **3**(1):1.
6. **Ad Hoc Reporting** [<http://reporting.inetsoftware.de/public/remote/adhoc>]
7. **Google Charts** [<https://developers.google.com/chart/>]
8. **Apache OpenOffice - The Free and Open Productivity Suite** [<https://www.openoffice.org/>]
9. **Bioconductor: Open Source Software for Bioinformatics** [<http://www.bioconductor.org/>]
10. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**(9):1639–1645.
11. **Circos Circular Visualisation** [<http://circos.ca>]
12. **OmniViz** [<http://www.instem.com/solutions/omniviz.html>]
13. **Partek: NGS and Microarray Data Analysis Software** [www.partek.com]
14. **TIBCO Spotfire: Business Intelligence Analytics Software and Data Visualization** [[spotfire.tibco.com](http://www.tibco.com)]
15. **IDBS E-Workbook** [<http://www.idbs.com/en/platform-products/e-workbook/inforsense-for-e-workbook/>]
16. **GenSight: Enterprise Portfolio Management Solutions** [www.gensight.com]
17. **Cartagenia. Confidently Interpret, Report and Share Genomic Variants.** [www.cartagenia.com]
18. Sharma MK, Philips J, Agarwal S, Wiggins WS, Shrivastava S, Koul SB, Bhattacharjee M, Houchins CD, Kalakota RR, George B, Meyer RR, Spencer DH, Lockwood CM, Nguyen TT, Duncavage EJ, Al-Kateb H, Cottrell CE, Godala S, Lokineni R, Sawant SM, Chatti V, Surampudi S, Sunkishala RR, Darbha R, Macharla S, Milbrandt JD, Virgin HW, Mitra RD, Head RD, Kulkarni S, *et al*: **Clinical genomics workstation.** *AMIA Jt Summits Transl Sci Proc* 2013, **19**(9):156–157.
19. **Knitr: A General-purpose Package for Dynamic Report Generation in R** [<http://cran.r-project.org/web/packages/knitr/>]
20. Leisch F: **Sweave: Dynamic generation of statistical reports using literate data analysis.** *Proc Comp Stat* 2002:575–580.
21. **R-Markdown** [<http://shiny.rstudio.com/articles/rmarkdown.html>]
22. **Tangle** [<http://worrydream.com/Tangle/>]
23. **Active Markdown** [<http://activemarkdown.org/>]
24. **IPython Notebooks** [<http://ipython.org/notebook.html>]
25. Hermida L, Poussin C, Stadler MB, Gubian S, Sewer A, Gaidatzis D, Hotz H-R, Martin F, Belcastro V, Cano S, Peitsch MC, Hoeng J: **Confero: an integrated contrast data and gene set platform for computational analysis and biological interpretation of omics data.** *BMC Genomics* 2013, **14**:514.
26. Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, Uva P, Zanetti G, Fotia G: **Orione, a web-based framework for NGS analysis in microbiology.** *Bioinformatics* 2014, **30**(10):1928–1929.
27. **Galaxy-P** [<https://usegalaxy.org/>]
28. **CGtag Pipeline with iReport as Final Step** [<http://galaxy.ctmm-trait.nl/u/saskia-hiltemann/p/cgtag>]
29. **iReport Example: Tutorial GCC2014.** [<http://galaxy.ctmm-trait.nl/u/saskia-hiltemann/h/gcc2014-ireport-about-ireport>]
30. **iReport Example: Genetic Report** [<http://galaxy-demo.trait-ctmm.cloudlet.sara.nl/u/andrew-stubbs/h/ireportgeneticreportchr21>]
31. **jQuery Zoom Library** [<http://www.jacklmoore.com/zoom/>]
32. **DataTables | Table Plug-in for jQuery** [<https://datatables.net>]
33. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D: **GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support.** *Bioinformatics* 1998, **14**(8):656–664.
34. **GeneCards - The Human Gene Compendium** [<http://www.genecards.org/>]
35. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ: **Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.** *Bioinformatics* 2010, **26**(7):966–968.
36. Drmanac R, Sparks A, Callow M, Halpern A, Burns N, Kermani B, Carnevali P, Nazarenko I, GB Nilsen G, Yeung G, Dahl F, Fernandez A, Staker B, Pant K, Baccash J, Borcharding A, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert J, Hacker C, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, *et al*: **Human genome sequencing using unchained base reads on self-assembling dna nanoarrays.** *Science* 2010, **327**(5961):78–81.
37. The International HapMap Consortium: **The international hapmap consortium. the international hapmap project.** *Nature* 2003, **426**:789–796.
38. The International HapMap Consortium: **Integrating common and rare genetic variation in diverse human populations.** *Nature* 2010, **467**:52–58.
39. **Complete Genomics Public Datasets** [<http://www.completegenomics.com/public-data/>]
40. **CTMM-TraIT Public Galaxy Instance** [<http://galaxy.ctmm-trait.nl>]

doi:10.1186/2047-217X-3-19

Cite this article as: Hiltemann *et al.*: iReport: a generalised Galaxy solution for integrated experimental reporting. *GigaScience* 2014 **3**:19.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

