TITLE: Enhanced JBrowse plugins for epigenomics data visualization

AUTHORS

Brigitte T. Hofmeister[1] and Robert J. Schmitz[2]

[1] Institute of Bioinformatics, University of Georgia, Athens, GA 30602, U.S.A.

[2] Department of Genetics, University of Georgia, Athens, GA 30602, U.S.A.

CORRESPONDING AUTHOR: Brigitte T. Hofmeister, bth29393@uga.edu and Robert J. Schmitz, schmitz@uga.edu

KEYWORDS: Epigenomics, genomics, genome browser, visualization

ABSTRACT

New sequencing techniques require new visualization strategies, as is the case for epigenomics data such as DNA base modifications, small non-coding RNAs, and histone modifications. We present a set of plugins for the genome browser JBrowse that are targeted for epigenomics visualizations. Specifically, we have focused on visualizing DNA base modifications, small non-coding RNAs, stranded read coverage, and sequence motif density. Additionally, we present several plugins for improved user

experience such as configurable, high-quality screenshots. In visualizing epigenomics with traditional genomics data, we see these plugins improving scientific communication and leading to discoveries within the field of epigenomics.

BACKGROUND

As next-generation sequencing techniques for detecting and quantifying DNA nucleotide variants, histone modifications and RNA transcripts become widely implemented, it is imperative that graphical tools such as genome browsers are able to properly visualize these specialized data sets. Current genome browsers such as UCSC genome browser [1], AnnoJ [2], IGV [3], IGB [4], and JBrowse [5], have limited capability to visualize these data sets effectively, hindering the visualization and potential discoveries with new sequencing technologies. JBrowse is used by numerous scientific resources, such as Phytozome [6], CoGe [7], WormBase [8], and Araport [9], because it is highly customizable and adaptable with modular plugins [5].

Epigenomics is an emerging area of research that generates a significant amount of specialized sequencing data which cannot be efficiently visualized using standard genome browsers. New sequencing technologies such as whole-genome bisulfite sequencing (WGBS) [2, 10], Tet-assisted bisulfite sequencing (TAB-seq) [11], single-molecule real-time sequencing (SMRT) [12], chromatin immunoprecipitation sequencing (ChIP-seq) [13], assay for transposase-accessible chromatin sequencing (ATAC-seq)

[14], RNA-seq [15–17], and small RNA-seq [18] have been instrumental in advancing the field of epigenomics. Epigenomic data sets generated from these techniques typically include: DNA base modifications, mRNAs, small RNAs, histone modifications and variants, chromatin accessibility, and DNA sequence motifs. These techniques have allowed researchers to map the epigenomic landscape at high resolution, greatly advancing our understanding of gene regulation. DNA methylation (4-methylcytosine, 4mC; 5-methylcytosine, 5mC; 5- hydroxylmethylcytosine, 5hmC; and 6-methyladenine, 6mA) and small non-coding RNAs (smRNAs) are modifications often found in epigenomic data sets, and function to regulate DNA repair and transcription by localizing additional chromatin marks or inducing post-transcriptional gene regulation [19–21].

We have developed several JBrowse plugins to address the current limitations of visualizing epigenomics data, which include visualizing base modifications and small RNAs as well as stranded-coverage tracks and sequence motif density. Additionally, we have developed several plugins that add features for improved user experience with JBrowse, including high-resolution browser screenshots. These plugins are freely available and can be used together or independently as needed. In visualizing epigenomics with traditional genomics data, we see these plugins improving scientific communication and leading to discoveries within the field of epigenomics.

IMPLEMENTATION

Plugins are implemented to work with JBrowse's modular plugin system. Client-side logic, such as visualization, fetching data, and interaction, are written in Javascript relying on the Dojo library [22]. This includes JavaScript classes for viewing data and storing data. Raw data files are standard in genomics, including BAM files for next-generation sequencing reads [23] and BigWig files for quantitative coverage tracks [24]. Python scripts are included to convert output from analysis pipelines to BigWig files needed by JBrowse. Additional styling for each plugin is provided using CSS. Wherever possible, colorblind safe colors were used to improve accessibility.

RESULTS

Base modifications. We have developed a plugin to visualize the quantity of 4mC, 5mC, 5hmC, and 6mA at single base-pair resolution. When studying 5mC, the modification is split into two (CG and CH; where H is any nucleotide expect G) sequence contexts for animals or three (CG, CHG, and CHH) sequence contexts for plants, as each context is established and/or maintained by different pathways with different functional roles [20]. Our plugin visualizes the quantity of methylation at each cytosine or adenine using a bar plot (Fig. 1), where values are positive or negative to signify the DNA strand. In most genome browsers, each sequence context must be shown as a different track (Fig. 1a). This is cumbersome when viewing multiple samples and makes it more difficult to determine overlap between context or samples. Our plugin

89 is advantageous because, we color-code 4mC, 5mC, 5hmC, and 6mA sequence
90 contexts and display them on a single track (Fig. 1b, Additional file 1: Fig. S1). However,
91 focusing on a single context or modification can be important, thus our plugin offers
92 several filtering options including by sequence context and base modification.
93
94 **Small RNAs.** Currently, JBrowse represents each sequenced RNA as a single read and
95 is colored by sequenced strand (Fig. 2a). When analyzing smRNAs, strand alone does
96 not always provide sufficient information; the size (nucleotides [nt]) of smRNA and
97 strandedness indicate potential function [19]. For example, in plants, 21nt microRNAs
98 can be aligned to single strand and 24nt small interfering RNAs can be aligned to both
99 strands [25]. Products of RNA degradation, however, have varying sizes and align to
100 one strand. To improve smRNA visualization, we color-code reads by smRNA size and
101 retain strand information by placement of smRNAs within the track relative to the y-axis
102 (Fig. 2b). This plugin also includes the ability to filter the reads in a track or multiple
103 tracks by size, strand, and read quality.
104
105 **Stranded read coverage.** Quantitative coverage tracks are necessary for any
106 worthwhile genome browser. It is important for visualizing DNA-protein interactions via
107 ChIP-seq and chromatin accessibility via ATAC-seq where coverage is computed in a
108 strand-independent manner. However, for strand-dependent data types, such as 5mC,
109 small RNAs, and mRNAs, read coverage can greatly vary for opposite strands. The
110 default coverage tracks are unable to handle this, thus we developed a plugin which

111 shows stranded read coverage. For example, WGBS can have uneven coverage on
112 both strands which can make only one strand seem methylated (Fig. 3a).
113
114 **Motif density**. Sequence motifs not only have important roles for protein binding, i.e.
115 binding motifs, but can also impact chromatin formation [26] and recombination hotspots
116 [27]. When correlating the frequency of a sequence motif with another characteristic, i.e.
117 5mC or histone modification localization, it is preferred to visualize motif density over
118 larger regions compared to single base-pair resolution. To address this, we developed a
119 plugin which visualizes sequence motif density across the genome as a heatmap (Fig.
120 3b). Users can input multiple motifs in a single track and IUPAC degenerate nucleotides
121 are supported. We also include several options for heatmap coloring and density
122 computation configuration options.
123
124 **Exporting browser images**. One of the most difficult tasks working with any genome
125 browser is obtaining high-quality screenshots for presentations or publications. We have
126 developed a plugin for JBrowse, which allows the user to take high quality and highly
127 configurable screenshots without installing additional software. A dialog window allows
128 users to set general, output, and track-specific configuration options (Fig. 4).
129 Additionally, our plugin is able to create the screenshot with vector graphic objects,
130 which is preferred for publication-quality screenshots, without needing to change the
131 underlying track configuration parameters.
132

Customization. To improve user experience, we have developed several additional JBrowse plugins. These plugins include: (i) Selecting or deselecting all tracks in a category from a hierarchical track list; (ii) An easily customizable y-scale range and location; and (iii) An option to force a track to stay in “feature” view or “histogram” view regardless of the zoom.

CONCLUSIONS

With these plugins, we aim to improve epigenomics visualization using JBrowse, a user-friendly genome browser familiar to the research community. All the plugins described can be used together or independently as needed. All plugins are freely available for download and additional customization.

AVAILABILITY AND REQUIREMENTS

Project name: Epigenomics in JBrowse

Project home page: http://github.com/bhofmei/bhofmei-jbplugins

Operating systems(s): Platform independent

Programming language: Javascript, Python

Other requirements: JBrowse 1.11.6+

License: Apache License, Version 2.0

Any restrictions to use by non-academics: none

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

See Additional file 1 for availability and description of data processing for samples used in the figures.

Competing interests

Not applicable

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (T32GM007103) to BTH, the National Science Foundation (IOS-1546867) to RJS., and the Office of Research at the University of Georgia to BTH and RJS.

Authors' contributions

Conceptualization and design: BTH and RJS; Implementation and testing: BTH; Writing: BTH; Review and editing: RJS. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Adam Bewick, Lexiang Ji, William Jordan, and Melissa Shockey for comments and discussions. We would like to thank Eric Lyons and Colin Diesh for open-source software code that influenced these plugins early in development. We would like to thank all members of the Schmitz lab for using the plugins during development and suggesting additional features. Additionally, we would like to thank Scott Cain and Mathew Lewsey for being early adopters.

FIGURES

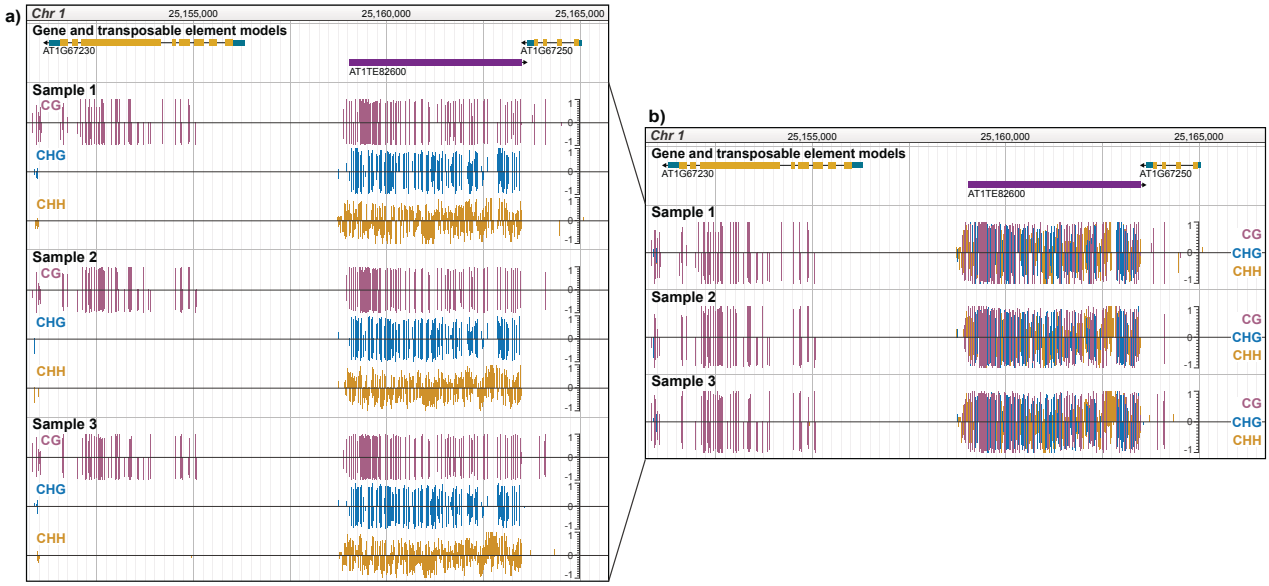

Fig 1. Visualizing DNA base modifications. Top track shows gene models in gold and transposable element models in purple. a) Viewing 5mC in three *A. thaliana* samples without the plugin. b) Viewing 5mC in the same samples with the plugin. For all tracks, height and direction of bar indicates methylation level and strand, respectively. Bars are colored by 5mC sequence context.

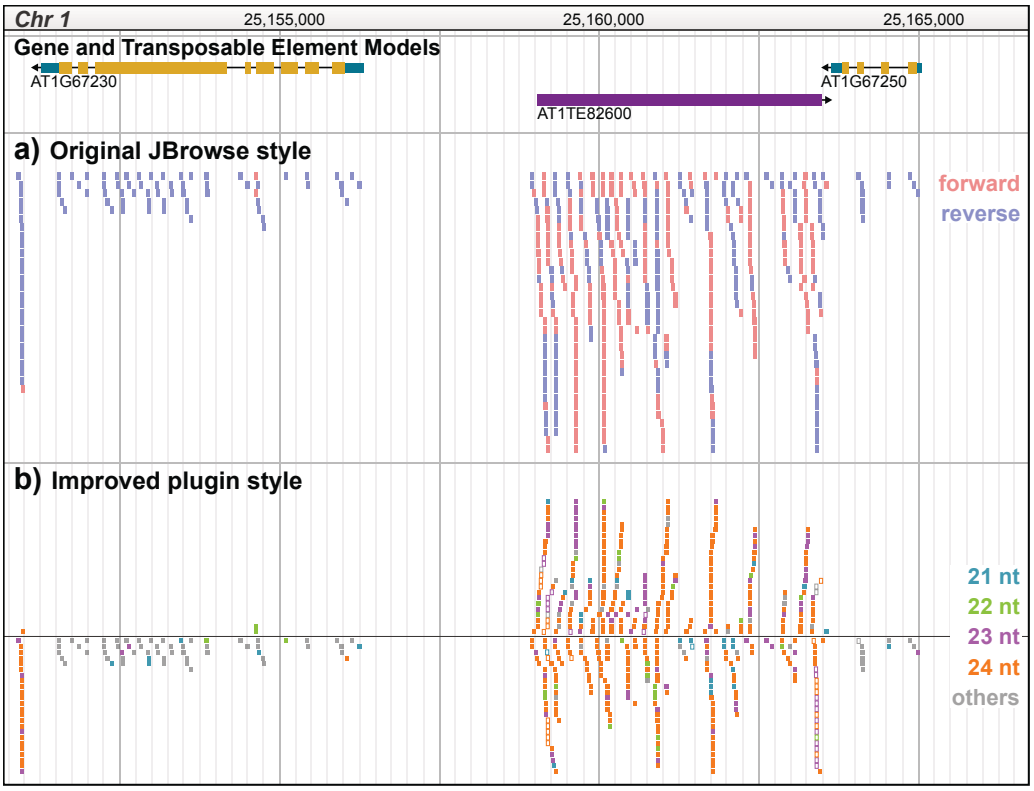

Fig 2. Visualizing small RNAs. Top track shows gene models in gold and transposable element models in purple. a) Viewing smRNA reads, 18 nt - 30 nt, in an *A. thaliana* sample using the general JBrowse alignments track. Color indicates strand; red, forward; blue, reverse. b) Viewing the same smRNA reads using the smRNA alignments track provided by the plugin. Color indicates read length. Position above and below the y-axis origin indicates forward and reverse strand, respectively. Unfilled reads map to multiple genomic locations and filled reads map uniquely.

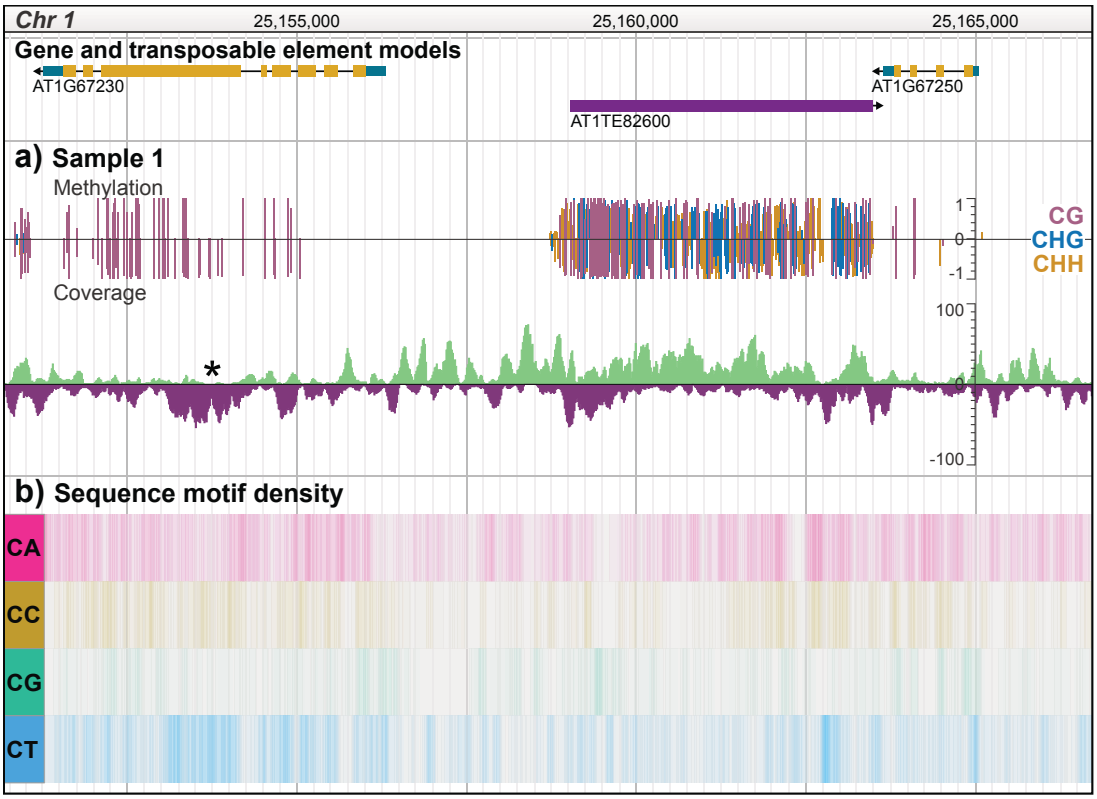

Fig 3. Visualizing stranded coverage and sequence motif density. Top track shows gene models in gold and transposable element models in purple. a) Stranded read coverage for sample used in the methylation track. Asterisk (*) indicates uneven strand coverage which affects the perceived methylation level. b) Dinucleotide sequence motif density in *A. thaliana*. Darker color indicates higher density.

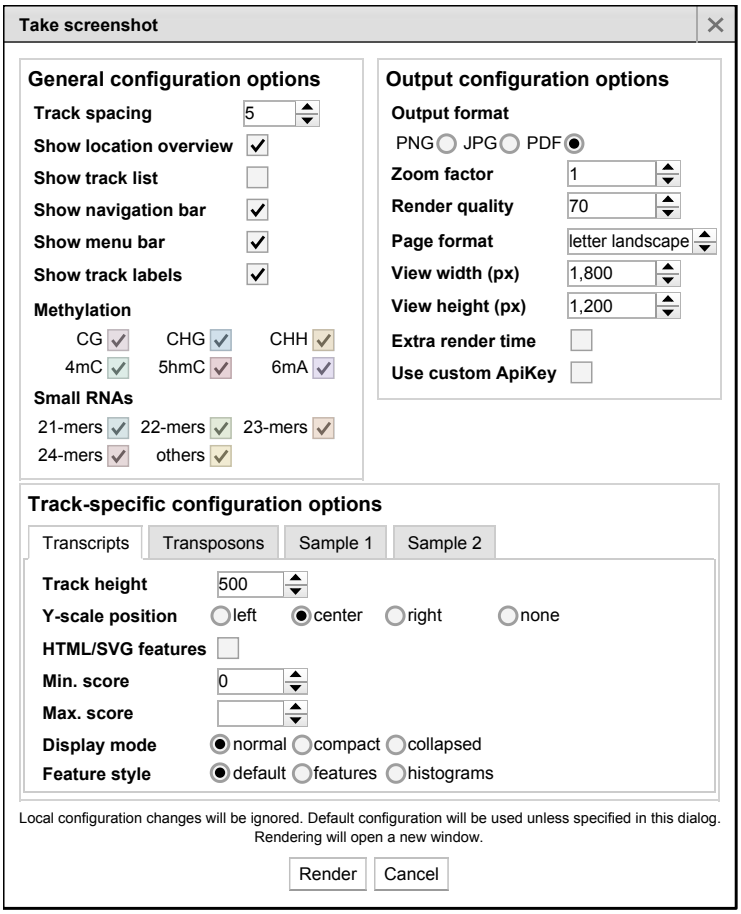

Fig. 4. Screenshot dialog window. The dialog window that opens when taking screenshots with our plugin. There are numerous configuration options for general visualization, image output, and track-specific settings. This includes exporting each track using vector objects.

SUPPLEMENTARY FILES

Additional file 1. Supplementary figure S1 and supplementary methods. PDF. Additional-file.pdf (103 KB)

234

235 **REFERENCES**

236

237 1. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM et al. The
238 human genome browser at UCSC. Genome research. 2002; 12: 996-1006.
239 2. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH et al.
240 Highly integrated single-base resolution maps of the epigenome in Arabidopsis.
241 Cell. 2008 May 2; doi:10.1016/j.cell.2008.03.029.
242 3. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G et al.
243 Integrative genomics viewer. Nature biotechnology. 2011; 29: 24-26.
244 4. Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics
245 platform for genomics. Bioinformatics. 2016 Jul 15;
246 doi:10.1093/bioinformatics/btw069.
247 5. Buels R, Yao E, Diesh CM, Hayes RD, Munoz-Torres M, Helt G et al. JBrowse: a
248 dynamic web platform for genome visualization and analysis. Genome Biol. 2016;
249 doi:10.1186/s13059-016-0924-1.
250 6. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J et al.
251 Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res.
252 2012; 40: D1178-D1186.
253 7. Lyons E, Freeling M. How to usefully compare homologous plant genes and
254 chromosomes as DNA sequences. Plant J. 2008 Feb; doi:10.1111/j.1365-
255 313X.2007.03326.x.

256 8. Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P et al. WormBase 2016:
257 expanding to enable helminth genomic research. Nucleic Acids Res. 2016 Jan 04;
258 doi:10.1093/nar/gkv1217.
259 9. Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M et
260 al. Araport: the Arabidopsis information portal. Nucleic Acids Res. 2015 Jan;
261 doi:10.1093/nar/gku1200.
262 10. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD et al. Shotgun
263 bisulphite sequencing of the Arabidopsis genome reveals DNA methylation
264 patterning. Nature. 2008; 452: 215-219.
265 11. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A et al. Base-resolution
266 analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012 Jun 08;
267 doi:10.1016/j.cell.2012.04.027.
268 12. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA et al. Direct
269 detection of DNA methylation during single-molecule, real-time sequencing. Nat
270 Methods. 2010 Jun; doi:10.1038/nmeth.1459.
271 13. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo
272 protein-DNA interactions. Science. 2007; 316: 1497-1502.
273 14. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of
274 native chromatin for fast and sensitive epigenomic profiling of open chromatin,
275 DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;
276 doi:10.1038/nmeth.2688.
277 15. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M et al. The

278 transcriptional landscape of the yeast genome defined by RNA sequencing.
279 Science. 2008 Jun 06; doi:10.1126/science.1158441.
280 16. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK et al. Stem
281 cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods.
282 2008 Jul; doi:10.1038/nmeth.1223.
283 17. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
284 mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;
285 doi:10.1038/nmeth.1226.
286 18. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL et al.
287 Application of massively parallel sequencing to microRNA profiling and discovery
288 in human embryonic stem cells. Genome Res. 2008 Apr; doi:10.1101/gr.7179508.
289 19. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev
290 Genet. 2009 Feb; doi:10.1038/nrg2504.
291 20. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation
292 patterns in plants and animals. Nat Rev Genet. 2010 Mar; doi:10.1038/nrg2719.
293 21. Marinus MG, Løbner-Olesen A. DNA Methylation. EcoSal Plus. 2014 May;
294 doi:10.1128/ecosalplus.ESP-0003-2013.
295 22. Dojo Toolkit: Reference Guide. https://dojotoolkit.org/reference-guide/1.10/.
296 Accessed 15 July 2017.
297 23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. The sequence
298 alignment/map format and SAMtools. Bioinformatics. 2009;
299 doi:10.1093/bioinformatics/btp352.

300 24. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed:
301 enabling browsing of large distributed datasets. Bioinformatics. 2010 Sep 1;
302 doi:10.1093/bioinformatics/btq351.
303 25. Finnegan EJ, Matzke MA. The small RNA world. J Cell Sci. 2003 Dec 01;
304 doi:10.1242/jcs.00838.
305 26. Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, Moore IK et al. A
306 genomic code for nucleosome positioning. Nature. 2006 Aug 17;
307 doi:10.1038/nature04979.
308 27. Myers S, Freeman C, Auton A, Donnelly P, McVean G. A common sequence motif
309 associated with recombination hot spots and genome instability in humans. Nat
310 Genet. 2008 Sep; doi:10.1038/ng.213.
311