READemption – A tool for the computational analysis of deep-sequencing-based transcriptome data

Konrad U. Förstner1,2, Jörg Vogel1 and Cynthia M. Sharma2∗
1Institute for Molecular Infection Biology, University of Würzburg, Würzburg
2Research Centre for Infectious Diseases (ZINF), University of Würzburg, Würzburg

ABSTRACT

Summary: RNA-Seq has become a potent and widely used method to qualitatively and quantitatively study transcriptomes. In order to draw biological conclusions based on RNA-Seq data, several steps some of which are computationally intensive, have to be taken. Our READemption pipeline takes care of these individual tasks and integrates them into an easy-to-use tool with a command line interface. To leverage the full power of modern computers, most subcommands of READemption offer parallel data processing. While READemption was mainly developed for the analysis of bacterial primary transcriptomes, we have successfully applied it to analyze RNA-Seq reads from other sample types, including whole transcriptomes, RNA immunoprecipitated with proteins, not only from bacteria, but also from eukaryotes and archaea.

Availability and Implementation: READemption is implemented in Python and is published under the ISC open source license. The tool and documentation is hosted at http://pythonhosted.org/READemption (DOI:10.6084/m9.figshare.977849).

Contact: cynthia.sharma@uni-wuerzburg.de; konrad.foerstner@uni-wuerzburg.de

1 INTRODUCTION

RNA-Seq, the examination of cDNA by massively parallel sequencing technologies, is a potent way to perform transcriptome analyses at single-nucleotide-resolution and with a high dynamic range (Wang et al., 2009). It has been successfully used to annotate transcript boundaries and to identify novel transcripts such as small regulatory RNAs in both pro- and eukaryotes (Filiatrault, 2011; Ozsolak and Milos, 2011). Most prominently, it can be applied to draw biological conclusions based on RNA-Seq data, several steps some of which are computationally intensive, have to be taken. Our READemption pipeline takes care of these individual tasks and integrates them into an easy-to-use tool with a command line interface. To leverage the full power of modern computers, most subcommands of READemption offer parallel data processing. While READemption was mainly developed for the analysis of bacterial primary transcriptomes, we have successfully applied it to analyze RNA-Seq reads from other sample types, including whole transcriptomes, RNA immunoprecipitated with proteins, not only from bacteria, but also from eukaryotes and archaea.

Availability and Implementation: READemption is implemented in Python and is published under the ISC open source license. The tool and documentation is hosted at http://pythonhosted.org/READemption (DOI:10.6084/m9.figshare.977849).

Contact: cynthia.sharma@uni-wuerzburg.de; konrad.foerstner@uni-wuerzburg.de

READemption provides the subcommands: align, coverage, gene_quant, deseq, viz_align, viz_gene_quant and viz_deseq which combine several processing substeps into comprehensible units.

Read processing and mapping: The fundamental tasks of pre-processing the input reads and aligning them to reference sequences is covered by the subcommand align. In an initial step READemption removes poly(A)-tails introduced during the library preparation and discards too short reads. For the alignment of reads to reference sequences, the short read mapper segemehl and its remapper lack (Hoffmann et al., 2009) are used. The mapping is
3 CONCLUSIONS

We present an open source pipeline for the analysis of RNA-Seq data from all domains of life. READemption generates several output files that can be examined with common office suites, graphic programs and genome browsers. Its features make it a useful tool for anybody interested in the computational analysis of RNA-Seq data with the required basic command line skills.

ACKNOWLEDGEMENT

We thank members of the Sharma and the Vogel groups especially Thorsten Bischler for testing and constructive feedback.

Funding: Work in the Sharma laboratory is supported by the ZINF Young Investigator program of the Research Center for Infectious Diseases (ZINF) at the University of Würzburg and the Bavarian BioSysNet program. CMS is supported by the Young Academy program of the Bavarian Academy of Sciences.

REFERENCES