

Modulation of the host cell RNA splicing program by the gastric pathogen *Helicobacter Pylori*

Frithjof Glowinski, Fernando Garcia-Alcalde, Konstantin Okonechnikov, Thomas F Meyer✉

Department of Molecular Biology, Berlin, Germany

Motivation and Objectives

Helicobacter pylori is a Gram-negative bacterial pathogen colonizing the human stomach. Infection with *H. pylori* causes chronic inflammation of the gastric mucosa and may lead to peptic ulceration and/or gastric cancer. Using a quantitative phosphoproteomic approach several splicing factors were found to be differentially phosphorylated upon infection (Holland *et al.*, 2011).

Serine arginine rich (SR) proteins, in particular, which are involved in the regulation and control of alternative splicing, are affected by the differential phosphorylation. SR proteins are regulated in their subcellular location dependent on their different phosphorylation states. To investigate the functional consequences of such alterations in phosphorylation we analysed the changes in splicing of a small set of known targets of SR-protein dependent alternative splicing. Within this set, two genes, BRCA1 and BMF, were confirmed to be differentially spliced after infection. For a more comprehensive picture of changes in splicing the host cell mRNA was sequenced using next generation sequencing, RNA-seq.

Methods

The mRNA of infected host cells was analyzed using by RNA-seq and further characterized for changes in gene expression and alternative splicing. Sequencing samples were obtained from *in vitro* infections. Human gastric AGS cells were infected with *Helicobacter pylori* strain P12. Total RNA was collected and the polyA enriched fraction was used for strand-specific library preparation. Sequencing was performed on the Illumina HiSeq2000 platform with 90 bp strand paired-end reads and a fragment size of 300 bp. To estimate variability 3 biological replicates were performed, each with a non-infected vs infected sample. Reads were mapped with TopHat (Trapnell *et al.*, 2009) and defaults parameters.

TopHat mapping was supplemented with splice junctions previously identified by PASSion (Zhang *et al.*, 2012), and mapping quality assessed using QualiMap (García-Alcalde *et al.*, 2012). All further analyses were based on this combined TopHat-PASSion mapping. Differential alternative splicing was analyzed using either Cufflinks (Trapnell *et al.* 2010) or DEXSeq (Anders *et al.*, 2012). Both pipelines were compared to an experimental in-house pipeline. This in-house tool uses statistical testing on the obtained isoform expression levels for detection of differentially spliced transcripts.

Results and Discussion

All six samples were sequenced with approx. 40 Mio paired-end reads. Reads were mapped and further analysed for differential expression and splicing.

Differential exon usage was analysed using the DEXSeq package. 15 genes were found to contain exons, which are significantly changed in their abundance after infection. Differential expression and alternative splicing was further analysed using the Cufflinks pipeline. Using the workflow described in Trapnell *et al.* (2012) 892 genes were found to be differentially expressed between the non-infected and infected cells. Accordingly, 178 genes were identified to be differentially spliced between the two conditions.

Notably both tools identified a considerably different number of genes to be affected by regulation of cellular splicing with just minimal overlap. To further investigate this contradicting results we developed a new statistical approach to detect genes significantly differing in their isoform distribution between conditions.

The results of this new method and a detailed comparison with the tools mentioned above will be presented.

Acknowledgements

This work was supported by the International-Max-Planck-Research-School IMPRS-IDI (FG).

References

- Anders, S., Reyes, A. & Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. *Genome research* **22**(10), 2008–2017.
- García-Alcalde, F. *et al.* (2012) Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**(20), 2678–2679.
- Holland, C. *et al.* (2011) Quantitative phosphoproteomics reveals link between *Helicobacter pylori* infection and RNA splicing modulation in host cells. *Proteomics* **11**(14), 2798–2811.
- Trapnell, C. *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**(3), 562–578.
- Trapnell, C. *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**(5), 511–515.
- Trapnell, C., Pachter, L., Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9), 1105–1111.
- Zhang, Y. *et al.* (2012) PASSion: a pattern growth algorithm-based pipeline for splice junction detection in paired-end RNA-Seq data. *Bioinformatics* **28**(4), 479–486.