SOFTvenom: an omics drug discovery approach from animal venoms

Juan Carlos Triviño[1], Miñambres Rebeca[1], Raquel Rodríguez-dePablos[1], Mayte Gil[1], Pierre Escoubas[2], Marion Verdenaud[3], Sheila Zuñiga[1], Sheila Zuñiga[1] ✉
[1]Sistemas Genomicos, Paterna, Spain
[2]Venomtech, Valbonne, France
[3]CEA, Gif sur Yvette cede, France

Motivation and Objectives

Animal venoms have been proven to be a rich source for drug development due to their efficiency and target selectivity and the subsequent reduction in side effects in a wide range of therapeutic conditions such as pain or cancer where medical needs are not properly addressed by the existing treatments. Since no reference genome is currently available for most venomous animals, research in this field has been economically restricted to small animal groups and species.

Here we present SOFTVENOM, an efficient strategy to reconstruct and characterize animal venoms. Our approach was applied to the transcriptomes analysis of three animal venoms as a pilot project using RNASeq techniques in two different NGS platforms, Illumina-HiSeq2000 and 454-GSTitanium. We functionally compare the different datasets to provide new insights into a fascinating field where little information is currently known.

Methods

Strand-specific libraries were constructed from the RNA poly(A) fraction of tarantula (*Poecilotheria regalis*), scorpion (*Parabutus transvaalicus*) and viper (*Bitis arietans*). Libraries were then sequenced with Illumina Hiseq2000 and 454-GSTitanium following a paired-end and a single-end strategy respectively. 10-20Gbs were obtained after Illumina sequencing and 0.6Gb after 454-sequencing.

Reads were trimmed, filtered and collapsed to reduce dataset complexity using a combination of Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit), Trimmomatic (Lohse *et al.,* 2012) and 'in house' scripts. Assemblies were performed with Oases (Schulz *et al.,* 2012) and Trinity (Grabherr *et al.,* 2011) and further merge with CAP3 (Huang and Madan, 1999). Functional annotation and comparison between venom assemblies was carried out using the UniProt database (http://www.uniprot.org) and custom scripts.

Results and Discussion

SOFTVENOM is an analysis framework that includes different tools to process RNA-Seq data from scratch to finally obtain a list of well-established isoforms and associated functional information.

Our results provide new insights into compounds, which may play a role in venom function. In addition, our work highlights the limitations of the applied analysis strategy and the differences found after functionally comparing both NGS platforms.

The results of this work are currently being integrated with proteomics data from mass spectrometry to obtain a complete view of the venom composition. The work done so far will serve as the fundamental basis for the study of a total of 200 animal venoms in the following three years as part of the VENOMICS project, an international effort to uncover the secrets behind venom activity and their potential use in the development of drugs to improve Human Health.

Acknowledgements

This work has been financed by the 7th Framework Program.

References

Grabherr, M. G. et al. (2011): Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**, 644-652. doi:10.1038/nbt.1883

Huang, X. & Madan, A. CAP3 (1999): A DNA sequence assembly program. *Genome Research* **9**, 868-877. doi: 10.1101/gr.9.9.868

Lohse, M. et al. (2012): RobiNA: a user-friendly, integrated software solution for RNA-seq-based transcriptomics. *Nucleic Acids Research* **40**, W622-W627. doi:10.1093/nar/gks540

Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases (2012): robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086-1092. doi: 10.1093/bioinformatics/bts094