

HPC for AspPic: a distributed web resource for alternative splicing prediction and transcript isoform characterization

ID - 137

Castrignano Tiziana¹, D'antonio Mattia¹, D'onorio De Meo Paolo¹, Sanna Nico¹, Rizzi Raffaella², Mignone Flavio³, Bonizzoni Paola², Pesole Graziano⁴

¹Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca, CASPUR, Rome, Italy

²DISCo, University of Milan Bicocca, via Bicocca degli Arcimboldi, 8, Milan, 20135, Italy

³Dipartimento di Scienze Biomolecolari e Biotecnologie, via Celoria 26, 20133 Milano, Italy

⁴Dipartimento di Biochimica e Biologia Molecolare, University of Bari, via Orabona, 4, Bari 70126, Italy

Motivation

Alternative splicing (AS) is the process by which one pre-mRNA yields different mRNAs leading to the production of protein isoforms, which can have diverse functions. AS affects the great majority of intron-containing transcripts and thus is a major mechanism in the expansion of transcript and protein complexity in eukaryotes. Recent descriptions of the functional implications of AS in tissue-specificity, different biological processes and tumor development has generated an explosion of interest and activity in this field.

In order to analyse a) transcriptome and proteome complexity of multicellular organisms, b) gene classes involved in human health and disease, a software platform for high-performance large-scale alternative splicing analysis and transcript isoform characterization could provide increasingly important results in many areas of basic and applied biomedical research.

Methods

The developed platform consists of a multi threaded Java application for data preprocessing and a distributed system of Java applications for alternative splicing prediction and transcript isoforms characterization. The software architecture is organized in two different phases, time independent. The preprocessing phase produces inputs for the next (computational) phase from a list of genes or chromosomal ranges, utilizing two customized web services, querying on a MySQL DataBase for the genome annotation. Computational phase consists of a scalable set of distributed remote servers and a centralized management Java application. A web interface allows the final user to submit input lists and shows obtained results for both preprocessing and computation phases.

Results

We have developed a platform (HPC for ASPic) independent, flexible and scalable Java environment for highperformance large-scale alternative splicing analysis and transcript isoform characterization, which integrates some web services, databases and computational intensive algorithms. It has been optimised programming multi threaded powerful Java client for data preprocessing and several distributed application servers for intensive computation. The system is fault-tolerant. The web resource is available free of charge for academic and non-profit institutions.

Page A.3/272

Availability: <http://www.caspur.it/HPCforASPic>

Email: tiziana.castrignano@caspur.it