

The STATegra project: new statistical tools for analysis and integration of diverse omics data



Ana Conesa (on behalf of the STATegra consortium)

Scientific coordinator of the FP7 STATegra project, Genomics of Gene expression Lab, Prince Felipe Research Center, Valencia, Spain

Conesa A (2014) *EMBnet.journal* **20**(Suppl A), e768. <http://dx.doi.org/10.14806/ej.20.A.768>

Next generation sequencing has speed up genome analysis and brought omics research closer to many organisms and biological scenarios. Today an increasing number of research projects propose the combined use of different omics platforms to investigate diverse aspects of genome functioning. These proposals ideally seek to provide complementary sources of molecular information that eventually can be put together to obtain systems biology models of biological processes. Hence, it is not rare anymore to find experimental designs involving the collection of genome, transcriptome, epigenome and even metabolome data on a particular system. However, standard methodologies for the integration of diverse omics data types are not yet ready and researchers frequently face post-experiment question on how to combine data of different nature, variability, and significance into an analysis routine that sheds more light than the

analysis of individual datasets separately. The [STATegra project](#)¹ has been conceived to address these problems and provide the genomics community with user-friendly tools for the integration of different omics data types. STATegra targets several sequencing based functional genomics methods, proteomics and metabolomics. In this presentation I will report about current results of the project that include the STATegraEMS, an experiment management system for storage and annotation of complex omics experiments, novel data integration visualisation tools, statistical approaches to integrate RNA-seq data with different regulators of gene expression, transcriptomics measurements combined with downstream features such as proteomics and metabolomics, and data mining strategies to leverage public domain datasets in the integrative effort. I will also present the STATegRa, a new Bioconductor R package for integrative omics data analysis.

¹ stategra.eu