

## **Data integration: overcoming challenges and creating opportunities**

*Jana Sponarova, Stefan Bleuler and Philip Zimmermann; NEBION AG, Hohlstrasse 515, 8048 Zurich, Switzerland*

The advent of genome-wide measurements has led to the publication of thousands of studies and corresponding data files. The availability of these data to the wider research community should facilitate re-analysis and meta-analysis, generating novel insight that goes beyond the primary purpose of these studies. Two major challenges in integrating such public studies are i) the heterogeneity in annotations and technology, and ii) the unreliable quality levels. The curation and integration of transcriptomic studies on a large scale require the use of controlled vocabularies for several biological dimensions such as tissues, genotypes, diseases, treatments, etc. To add further value to the data, a strict quality control ensuring to only integrate high-quality samples and experiments is mandatory, to avoid bias in the results. Integration of such deeply curated data results in a database suitable for meta-analyses, comprising extensive and flexible drill-down possibilities. In addition, incorporating data-measurements together with detailed curated sample information, enables on-the-fly calculations. Smart integrated systems with high quality curated data empower biologists to perform compendium-wide searches and analyses, advancing their research and accelerating novel scientific discoveries. We discuss how the innovative expression database platform GENEVESTIGATOR is overcoming the described challenges. Its unique approach enables researchers to globally explore public and proprietary expression data for research and creates opportunities for clinical and translational applications.