Prototype meta-analysis demonstration for ImmuneSpaceR, using designated S4 objects

Dror Berel, Raphael Gottardo

INTRODUCTION

ImmuneSpace is a powerful management and analysis engine web-portal for integrative modeling of human immunological data. Currently, it contain data from 66 human immunology studies, covering a total of 4,084 participants. Each study is comprised of multiple data types including microarray, flow cytometry, hemagglutination inhibition assay, among others. Some of these assays contain thousands of biological measures (e.g. mRNA gene transcripts, FACS analytes / markers). The data is standardized and annotated for both explanatory clinical outcomes and biological ontologies. In addition to the web-portal, the entire ImmuneSpace databases are also accessible for direct download via **ImmuneSpaceR**, an API R/Bioconductor package.

GOAL

Aggregating such comprehensive data across all studies, all subjects and all variables (both dependent and independent) is an exhaustive task. Here we demonstrate an R analysis pipeline to accomplish this task, and demonstrate a meta-analysis using a specific hypothesis of interest.

METHODS

The R/Bioconductor **MultiAssayExperiment** (MAE) package is a designated S4 class for integrative omic data from multiple assays. All data from a study is converted and stored in a single MAE object, which is a non-atomic R object. A tibble R class is used to systematically access multiple non-atomic objects in a fashion reminiscent of the canonical R data frames. 10 out of the 66 ImmuneSpace studies include complete data for both microarray (gene markers) and clinical outcome that is derived from the Hemagglutination assay (HAI). For each study, association between the clinical outcome and each of the genes (separately) is modeled via a **logistic regression**, and summarized as an odds-ratio (OR) estimation. For each gene, a '**meta**' OR across all single-studies is calculated by taking into account the relative weight for each of the study's effect sizes using the **Mantel-Haenszel** method. A **forest plot** summarize all single-studies and 'meta' ORs.

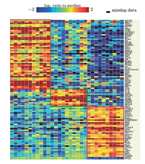
CONCLUSION

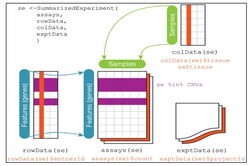
The combination of such well-annotated standardized data, and designated tools for accessing it, enables modification and extension of this prototype analysis into broader pipelines of meta-analysis hypotheses testing.

References

- Consortium, P. (2014). Computational resources for high-dimensional immune analysis from the Human Immunology Project Consortium. Nature Biotechnology *32*, 146–148.
- SIG M (2017). MultiAssayExperiment: Software for the integration of multi-omics experiments in Bioconductor. R package version 1.3.20
- Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods *12*, 115–121.

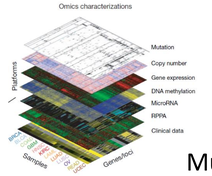
Single Assay (S4)

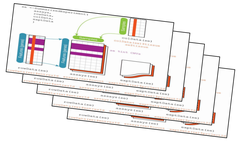




SummarizedExperiment[features (genes), subjects]

Single Multi-Assay (S4)

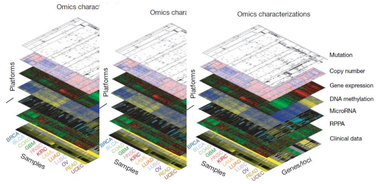




MultiAssayExperiment[features (genes), subjects , Assays]

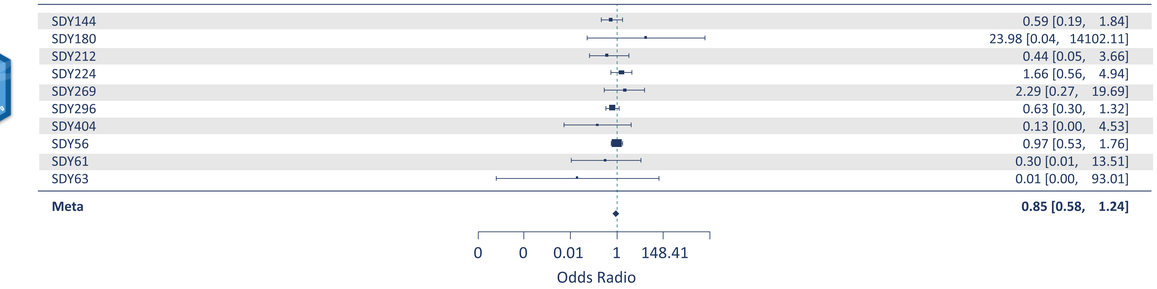
Multiple Multi-Assays

LIST($S4_1$, $S4_2$, ..)



Study	MAE S4 objects	expressionSet S4 objects	Nested data frames
Study 1	MAE $S4_1$	ExpressionSet $S4_1$	
Study 2	MAE $S4_2$	ExpressionSet $S4_2$	
Study 3	MAE $S4_3$	ExpressionSet $S4_3$	

HAI association with NUAK1 gene



FRED HUTCHINSON CANCER RESEARCH CENTER

dberel@fredhutch.org