# Supplemental Methods

## The MultiAssayExperiment class

The major design components of MultiAssayExperiment include 1) *colData*, a representation of patient-level or specimen-level characteristics as a table-like *DataFrame*, 2) *ExperimentList*, a list-like object with one element per experiment, representing any number of genomic range-based or identifier-based experimental results. Different elements do not need to have matching rows (variables/features), columns (samples) or even classes (experiments). 3) *sampleMap*, a DataFrame-class map that associates assay observations to patients, enabling replicate and missing assays. 4) a constructor function and methods for subsetting by experiments, genomic ranges, variables, samples and patients, 5) extraction methods for interfacing to statistical analysis and visualization functions in R, and 6) various helper functions for selecting patients with complete data, converting segmented genomic ranges to matrix representation, merging replicates, etc. Continuous validity requirements ensure correct matching of all patients and specimens during subsetting and re-arrangement operations. A comparison of the MultiAssayExperiment and MultiDataSet [(1)](https://paperpile.com/c/rad3b2/k3A3) approaches is made in **Supplemental Table 1**.

## The MultiAssayExperiment constructor

A *MultiAssayExperiment* class object is constructed by a call to the function of the same name, with three key arguments, each optional:

1. *experiments*: A *list* or *ExperimentList* class object providing assay data for all experiments, with one list element per experiment. Elements may be rectangular in shape (the same number of measurements per patient), or ragged in shape (e.g. segmented copy number or mutations represented by genomic coordinates).
2. *colData*: A *DataFrame* describing participant characteristics.
3. *sampleMap*: A three-column *DataFrame* with columns corresponding to assay names, patient identifiers in the *colData* table, and column identifiers in the *ExperimentList* elements.
4. metadata: A *list* of arbitrary format, describing the multi-assay experiment as a whole.

If *colData* or *sampleMap* are not provided, they are created automatically as possible from the provided experiments, in a way that guarantees that every column of the experiment list can be uniquely mapped to a single row of the patient / biological unit data (*colData*). The constructor returns an object of class *MultiAssayExperiment* (**Supplemental Table 3**).

## Subsetting

There are three main subsetting operations: 1) subset by rows 2) subset by columns 3) subset by assay. The MultiAssayExperiment can be thought of as an 3 dimensional ragged array where the first two dimensions (a table) represents one experiment and the third dimension represents the remaining experiments. The subset operations are facilitated by the *sampleMap*, which documents the relationships across experiments between patients and samples. All subset operations can be achieved via the bracket (“[“) methods with each index position corresponding to the above operations. Alternatively, functions *subsetByRow*, *subsetByColumn*, or *subsetByAssay* can be used. The double bracket (“[[“) implementation will access a single element in the ExperimentList. The dollar sign operator (“$”) returns a column of colData.

When assays have different feature types, such as gene identifiers vs. genomic ranges, the default behavior of row-based subsetting is as follows. When subsetting by gene identifiers, any genomic ranges named by those gene identifiers will be kept; other genomic ranges will be removed. Conversely, subsetting by genomic ranges will eliminate all non range-based features. Similarly, subsetting for example by microRNA identifiers will remove all rows labeled by gene identifiers. We emphasize, however, that this is default behavior and not a limitation of the MultiAssayExperiment class. In future work, we anticipate defining classes for identifiers that define what type of identifier they are and provide methods for automatically mapping them to other identifier types, such as genes to genomic ranges or to putative regulatory microRNAs. Such classes will be compatible with MultiAssayExperiment and its ability to accept a list of identifier vectors, but performing this mapping is complex and outside the scope of MultiAssayExperiment itself.

## Accessors

Whereas the extractors provide datasets in ready to analyze or visualize formats, accessors return components of the data. Accessor functions can also be used to replace data within the *MultiAssayExperiment* object. Accessors include:

1. *colData* function: returns the participant characteristics data table
2. *experiments* function: returns the list of experiments
3. The *assay* and *assays* function, which returns a *SimpleList* with one element per assay/experiment, each element containing a numeric data matrix extracted from that experiment. Range-based assays should be coerced to rectangular form beforehand. A typical use for this function would be to calculate correlations between samples or variables across different assays, or to perform unsupervised clustering. This format of *assays* output is compatible with several Bioconductor packages for multi-omics data analysis, including *omicade4*, *iClusterPlus*, *PMA, made4,* and *MCIA.*
4. *sampleMap* function: returns the map linking experimental assays to column data
5. *rownames*, *colnames*: return a list of row or column names from each assay
6. *head*: shows the result of subsetting the first five rows of each assay

## Management

A variety of functions assist with object construction and manipulation:

* *complete.cases*: returns a logical indicating, for each patient (row in *colData*), whether data are available for all assays
* *duplicated*: identify duplicate / replicate columns in experimental assays. These are identified as such because they map back to the same patient (row in *colData*)
* *mergeReplicates*:Merge multiple / duplicated observations within each experiment using a user-provided function
* *intersectRows*, *intersectColumns*: obtain a MultiAssayExperiment object where all rownames or observations by biological unit are present in all experiments
* *prepMultiAssay*: troubleshoots common problems that can occur when calling the MultiAssayExperiment constructor to create a MultiAssayExperiment object, for example, mis-matched sample identifiers due to capitalization and name-munging by R, and returns suggested “cleaned” objects for construction.

## Reshaping

These functions integrate across assays and with colData to interface for plotting and analysis functions.

* *longFormat*, *wideFormat:* integrate data from multiple assays and optionally from *colData* into either a long and skinny format or a wide format for data analysis and visualization. A typical use for these functions would be for regression analysis or plotting, and their output is compatible with many such functions in R and Bioconductor. The optional *colDataCols* argument adds data on the biological units to the assay data.

## Combining

The current API allows for the input of additional experiments via the “c” function provided that the user creates a map for the new samples to be added to the MultiAssayExperiement instance. A convenience argument is available for mapping of samples similarly to an experiment already present in the data.

## Validity

Validity checks ensure data integrity in the MultiAssayExperiment. Aside from internal helper functions that help the user create a valid MultiAssayExperiment, validity checks ensure that any further changes to the MultiAssayExperiment do not disturb the documented relationships in the object.

## TCGA MultiAssayExperiment objects

We downloaded unrestricted access data using the *RTCGAToolbox* Bioconductor package, added molecular subtype information curated from TCGA primary publications, and collected and merged an expansive set of clinical and pathological data. These objects (one per cancer type) are distributed through an Amazon S3 bucket and linked to at <http://tinyurl.com/maeourls>. In the future, they will be distributed through a new Bioconductor package using ExperimentHub*.* This data acquisition and object construction is reproducible using code at https://github.com/waldronlab/curatedTCGAData.

### Curated subtypes

The Cancer Genome Atlas (TCGA) project has produced a great amount of information via sequencing of tumor pathologies. Several studies on many of the 33 types of cancer have yielded subtype classifications that can be included in the MultiAssayExperiment data structure. However, this information was published as supplements of publications without standardized formats or centralized availability. We attempted to gather all available molecular subtype information in these publications. In particular, each primary publication was reviewed and subtype classifications were determined from the text. We then reviewed supplementary information to obtain sample subtype classification for each patient. Lastly, this information was added to the *colData* slot in the corresponding MultiAssayExperiment object. In the event that sample subtype classification information was unavailable, we contacted the corresponding authors and asked them to provide the supplementary material.

### Curated clinical data

The TCGA level 4 merged clinical data provided by the Broad Institute’s GDAC Firehose pipeline only includes about 60 to 80 select variables from the original XML format distribution [(2)](https://paperpile.com/c/rad3b2/rlKJJ). We started from complete level 1 clinical and pathological datasets, which contain hundreds more variables and in some cases have less missingness than the level 4 merged dataset. We included only variables that are not all missing, consolidated follow-up observations, and merged equivalent columns found in both clinical and pathological dataset, and ensured that patient identifiers are consistent across clinical datasets.

# Example Analyses

Analyses here are reproducible using the complete code in the vignette/extras folder at <https://github.com/waldronlab/multiassayexperiment>; only key selections of code are shown here. Reproduction of these examples currently requires the development version of MultiAssayExperiment, available in the Bioconductor development branch 3.6 and on GitHub (<https://github.com/waldronlab/multiassayexperiment>).

## Copy number and mutation burden in colon cancer and breast cancer

Datasets were downloaded from<http://tinyurl.com/MAEOurls>. Mutation load was calculated using all somatic variants and then normalized by assuming an average exome capture size of 50Mb. The somatic copy number alteration (SCNA) score was defined as the average GISTIC gene-level copy number group (0: normal, 1: gain/shallow loss, 2: amplification/deep loss). The following code snippet for colon cancer loads the complete TCGA dataset into R, selects the mutation and GISTIC copy number assays, selects only cases where both assays are available, then adds columns to colData for the number of mutations per Mb and the mean absolute copy number. The following code demonstrates loading the colon adenoma TCGA dataset, restricting to somatic mutation and copy number assays, selection of cases with complete data, and addition of copy number load to the clinical data:

mae <- readRDS("coadMAEO.rds")

mae <- mae[, , c("Mutations", "gistict")]

mae <- intersectColumns(mae)

mae$cnload <- colMeans(abs(assay(mae[["gistict"]])))

## Correlation between copy number, RNA, and protein in the NCI-60 cell lines

The NCI-60 cell line collection compiled by the U.S. National Cancer Institute (NCI) is a very widely used panel for the study of cellular mechanisms of cancer. It is a model system for the tissue types and genetic diversity of human cancers and has been extensively molecularly characterized by multiple genomic assays such as DNA copy number, DNA methylation, gene transcript, exome sequencing, miRNA profiling as well as drug activity levels profiling. We acquired processed data for DNA copy number, transcript and protein levels in tabular format through the CellMiner web application [(3)](https://paperpile.com/c/rad3b2/4cr9x). The DNA copy numbers and transcript levels provided by CellMiner were computed as a combined value from multiple platforms. Protein levels were based on reverse-phase protein lysate arrays and processed as described by Shankavaram *et al.* [(4)](https://paperpile.com/c/rad3b2/9gwOI). We assembled those multi-assay processed data into the MultiAssayExperiment object, and we showed the ease of this representation to compare the DNA copy number, transcript and protein levels of the Enabled Homolog (Drosophila) (ENAH) gene as an example gene. Pearson’s correlation coefficient was calculated between mRNA, protein, and DNA copy number, across all samples, and plotted as a 3x3 heatmap. The following code loads the full NCI-60 dataset then combines the copy number, gene expression, and protein abundance assays for a single protein coding gene (ENAH) into a data frame suitable for calculating correlations between profiles:

ncimae <- readRDS("nciMAEO.rds")

protrna.wide <- wideFormat(ncimae[“ENAH”, ,

c(“DNA Agilent", "RNA Agilent", "Protein Agilent")],

key = “assay”)

cor(as.data.frame(protrna.wide[, c(“DNA Agilent”, “Protein Agilent”, “RNA Agilent”)]))

### Methylation Quantitative Trait Loci

High-density SNP genotyping has been fruitful in elucidating regulatory and environmental mediators of cancer risk [(5,6)](https://paperpile.com/c/rad3b2/TrXQm+X1YUP). Collections of variant calls can be coupled with information about variant context and accuracy in the Variant Call Format (VCF; <https://samtools.github.io/hts-specs/>). VCF representations of genetic variation in epidemiologic cohorts are interrogated piecemeal after compression and indexing. VCF collections are incorporated into MultiAssayExperiment instances as references. To accommodate the typical approach of separating variant calls by chromosome, a data structure called “VcfStack” can be used to collect all the references constituting a genome-wide SNP scan. Infrastructure for MultiAssayExperiment and allied classes facilitates scalable combination of genomic assays and genetic variants, so that simple programming can be used to extract genotypes with genomic addresses near a genomic feature (or any given address), automatically ensuring that assay results are linked to the samples on which they are measured.

A publicly accessible archive of VCFs for the 1000 Genomes project is available at <http://1000genomes.s3.amazonaws.com/release/20130502/>. To create panel (d) of Figure 1, we assembled a MultiAssayExperiment instance with the Illumina Infinium HumanMethylation450 data published for the Yoruban HapMap population by Banovich *et al.* [(7)](https://paperpile.com/c/rad3b2/4mju8), and a reference to the Amazon S3 archive of the complete set of 1000 Genomes genotypes (produced by the stack1kg() function of Bioconductor ldblock.) We selected by name a probe (cg04793911) in the vicinity of gene HLA-DQB1, because Figure 3 of Banovich et al. gives a very detailed multi-omic display of expression and chromatin modification processes nearby. We then used the *cisAssoc* function from the *gQTLstats* package to search for SNP in the vicinity of the probe. Panel (d) of Figure 1 is created by interrogating the MultiAssayExperiment instance only, and genotype values per sample are acquired from the Amazon cloud in real time as the figure is produced. The following code loads a methylation SummarizedExperiment object from the *yriMulti* package, creates a connection to connection to the remote SNP VCF files, creates a MultiAssayExperiment instance containing these two objects, then calculates cis-correlations of SNPs the methylation site cg04793911:

data(banovichSE)

st <- stack1kg()

multiban <- MultiAssayExperiment(list(meth = banovichSE, snp = st),

colData = colData(banovichSE))

multibanfocus <- multiban[rowRanges(banovichSE)[“cg04793911”], , ]

assoc <- cisAssoc(multibanfocus[[“meth”]],

TabixFile(files(multibanfocus[[“snp”]])))

**Integrated Single Sample Gene Set Enrichment Analysis**

Gene expression (RNASeq2GeneNorm), proteomics (RPPAArray) and copy number (gistica) data of TCGA Ovarian Serous Cystadenocarcinoma (n=591) were downloaded as a MultiAssayExperiment object from Amazon <http://s3.amazonaws.com/multiassayexperiments/ovMAEO.rds> These data had been prepared using the curatedTCGAData pipeline described above and are described in **Supplemental Table 2**. Data were projected onto the common scale using multiple factor analysis [(8)](https://paperpile.com/c/rad3b2/q3ado) using the mogsa R/Bioconductor function *mcia*() and the first 15 principal components of the joint data decomposition were retained. Multiple factor analysis extracts principal component than are maximally covariant between datasets (for a detailed introduction see review [(9)](https://paperpile.com/c/rad3b2/WfWcB)). Each principal component is orthogonal and captures different information. Component are ranked such that the first component captures the most variant correlations between and within datasets. Each component can be represented in feature space as a vector of all RNAseq, protein, copy number features, the length of the vector will equal the union of all features. Thus features from diverse multi-omics data are transformed onto the same scale, allowing one to easily compute gene set scores over all features. We have described this single sample gene set analysis (ssGSA) aprpoach called multi-omics GSA (moGSA), and in experiments using simulated data moGSA outperforms ssGSEA and GSVA [(10)](https://paperpile.com/c/rad3b2/q0qxf).

We computed moGSA single sample genset scores for the 50 cancer hallmark gene sets [(11)](https://paperpile.com/c/rad3b2/UVxPQ). Ovarian cancer molecular subtypes have been described and were extracted from the supplementary data provided by Verhaak *et al.* [(12)](https://paperpile.com/c/rad3b2/4J4Xb). Germline BRCA1 and BRCA2 status of TCGA ovarian tumor were extracted from the supplementary pdf file provided with the TCGA publication [(13)](https://paperpile.com/c/rad3b2/XuY6o).

Hierarchical cluster analysis of the Pearson correlation between cancer hallmark signature scores of the serous ovarian tumors identified 4 major clusters, which were broadly consistent with the published molecular subtypes in serous ovarian cancer (**Supplemental Figure 1**), but did not distinguish ovarian tumors by BRCA1 or BRCA2 mutation status (**Supplemental Figure 2**), which is consistent with published reports [(12)](https://paperpile.com/c/rad3b2/4J4Xb). The following code loads the ovarian cancer MultiAssayExperiment object, selects the needed assays and cases with complete data for those assays, then generates a list of matrices compatible as input to the moa function from the mogsa package:

ovMAEO <- readRDS(ovMAEO.rds)

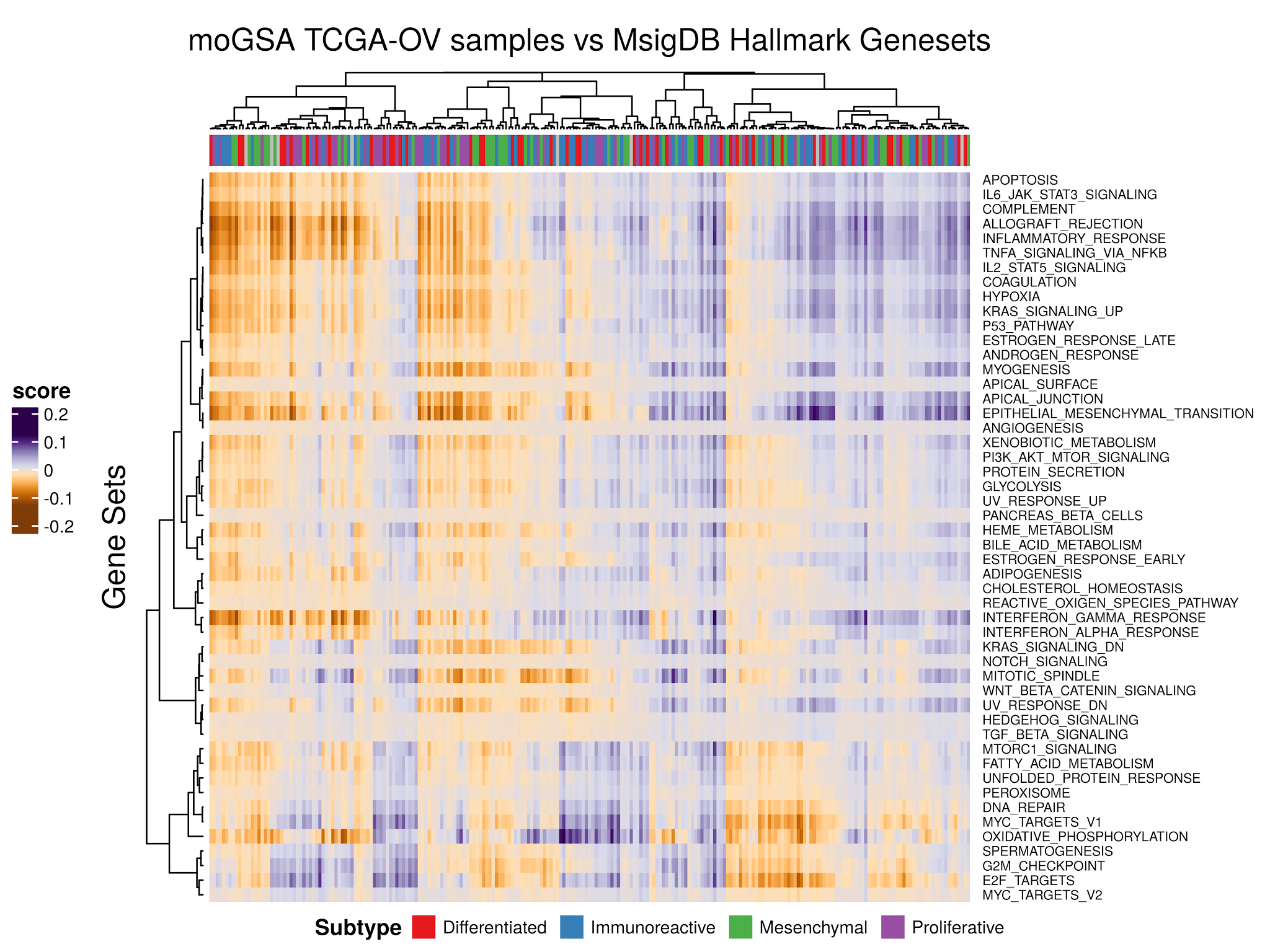
mogsaMAEO <- ovMAEO[, , c(“RNASeq2GeneNorm", "RPPAArray", "gistica") ]

mogsaMAEO <- intersectColumns(mogsaMAEO)

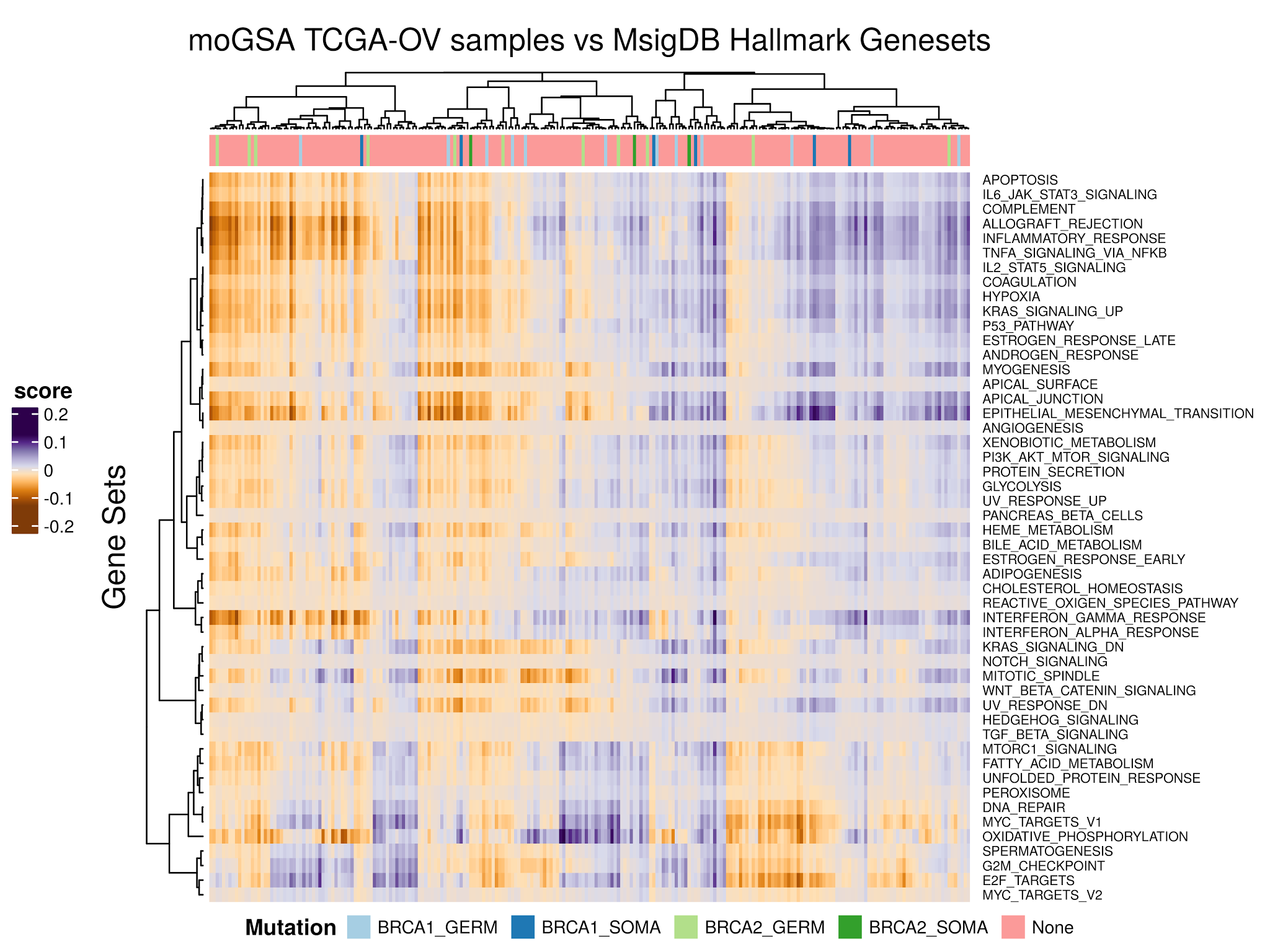
assayMats <- assays(mogsaMAEO)

ov.moa <- moa(assayMats, proc.row = “center\_ssq1”, w.data = “inertia”)

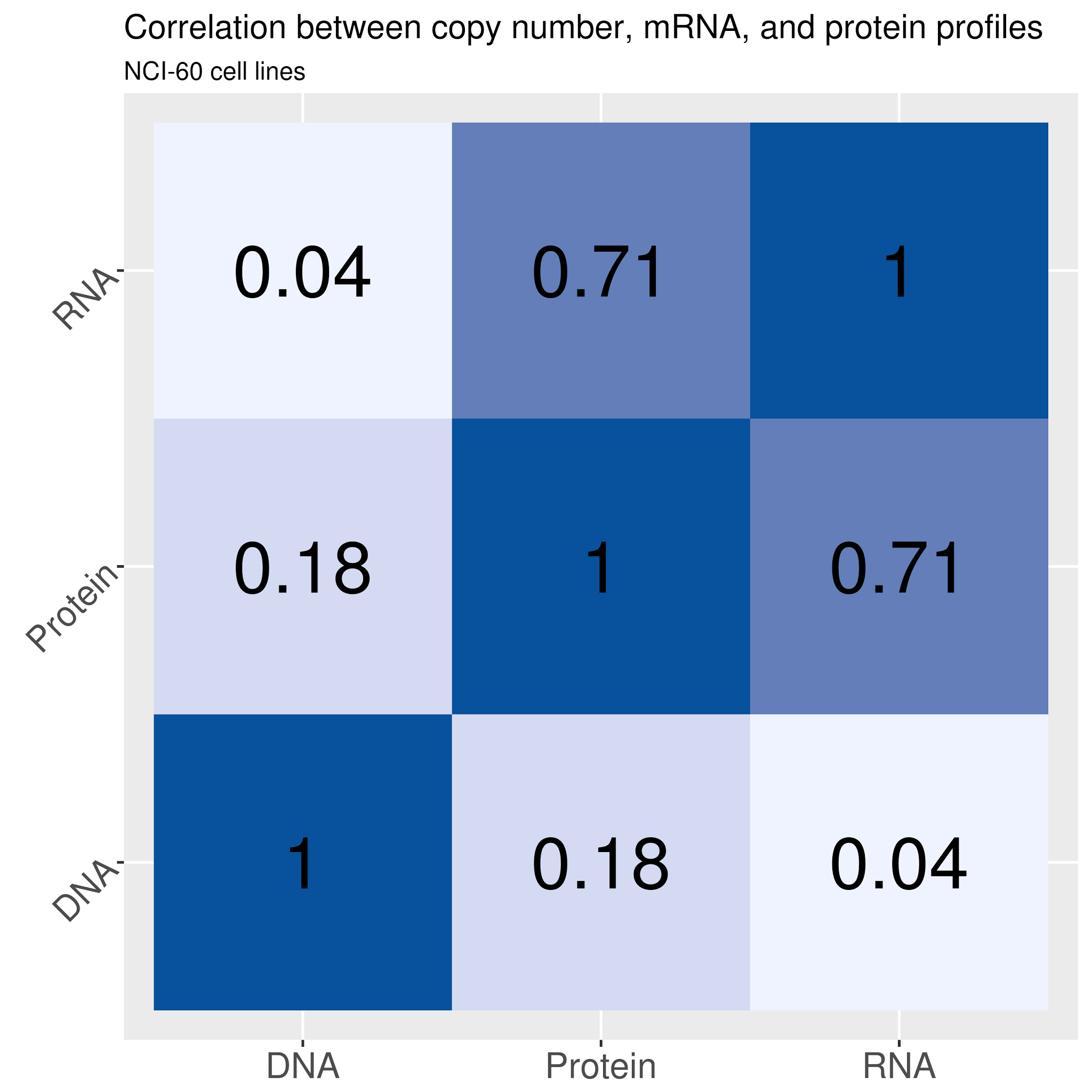
**Supplemental Figure 1. Mogsa enrichment scores for Hallmark gene sets**, calculated from platforms in the TCGA ovarian cancer cohort. The top color bar is coded by subtype.



**Supplemental Figure 2. Mogsa enrichment scores for Hallmark gene sets**, calculated from platforms in the TCGA ovarian cancer cohort. The top color bar is coded by BRCA mutation status.



**Supplemental Figure 3. DNA copy number, mRNA, and protein** levels of Enabled Homolog (Drosophila) (*ENAH*) gene in the NCI-60 cell lines show much higher correlation between mRNA and protein abundance (r=0.71), than between mRNA or protein with copy number.



|  |  |  |  |
| --- | --- | --- | --- |
| **Supplemental Table 1**. Comparison between MultiAssayExperiment and MultiDataSet | | | |
|  |  | **Package** | |
| **Feature** | **Description** | **MultiAssayExperiment** | **MultiDataSet** |
| Integrative representation | Allows the analysis of multiple data types | ✔ | ✔ |
| Coordinated subsetting | Subset rows/columns across multiple assays | ✔ | ✔ |
| Concatenation | Add experiments to existing object | ✔ | ✔ |
| Package-specific wrappers | Provides direct wrappers for other Bioconductor packages |  | ✔ |
| Feature annotation slot | Provides a slot for annotating features and ranges |  | ✔ |
| Supports open-ended set of classes | Supports additional classes without extra coding | ✔ |  |
| Out-of-memory | Supports data classes providing on-disk and remote representation of big datasets | ✔ |  |
| Consistent API | Provides a set of endomorphic operations consistent with other Bioconductor core classes | ✔ |  |
| Reshaping | Reshapes assay and primary data to wide and long formats | ✔ |  |
| Unified subject data | A single table for patient / subject data, mapped to zero, one, or more observations in each assay | ✔ |  |
| *Note*. **Package-specific wrappers.** MultiDataSet provides wrappers for several analysis packages (e.g., w\_iclusterplus, w\_mcia); MultiAssayExperiment provides only generic extraction and reshaping functions capable of generating output suitable for these and other packages. **Feature annotation slot.** MultiAssayExperiment does not provide a slot for feature annotations, under the assumption that supported single-assay data classes can provide this; however this approach allows MultiDataSet to, for example, subset non range-based data classes like eSet using genomic ranges. **Supports open-ended set of classes.** MultiDataSet operations provides special functions (i.e., add\_snps, add\_genexp, add\_methy) and uses internal code to support certain data classes, whereas MultiAssayExperiment supports an open set of data classes by default as long as they meet minimum API requirements (i.e., dim, dimnames, "["). **Consistent API** The MultiAssayExperiment API is modeled after SummarizedExperiment wherever possible; see Table 1. The CNAmet package [(14)](https://paperpile.com/c/rad3b2/PsUt) is tailored to one specific analysis so is not comparable on the items listed here. | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Supplemental Table 2. Prepared MultiAssayExperiment objects for TCGA cohorts, with number of assays, features, and samples** | | | |
| **TCGA Cohort** | **Number of Assays** | **Number of Features** | **Number of Specimens\*** |
| Adrenocortical Carcinoma | 9 | 677,947 | 180 |
| Bladder Urothelial Carcinoma | 11 | 1,236,717 | 806 |
| Breast Invasive Carcinoma | 8 | 595,134 | 1,212 |
| Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma | 9 | 892,007 | 586 |
| Cholangiocarcinoma | 9 | 610,891 | 85 |
| Colon Adenocarcinoma | 12 | 1,245,516 | 914 |
| Lymphoid Neoplasm Diffuse Large B-Cell Lymphoma | 9 | 627,264 | 94 |
| Esophageal Carcinoma | 10 | 864,036 | 373 |
| Glioblastoma Multiforme | 8 | 111,461 | 573 |
| Head and Neck Squamous Cell Carcinoma | 11 | 1,271,473 | 1,090 |
| Kidney Chromophobe | 9 | 632,321 | 132 |
| Kidney Renal Clear Cell Carcinoma | 11 | 1,195,305 | 1,059 |
| Kidney Renal Papillary Cell Carcinoma | 11 | 958,367 | 593 |
| Acute Myeloid Leukemia | 5 | 971,290 | 392 |
| Brain Lower Grade Glioma | 11 | 1,081,610 | 1,013 |
| Liver Hepatocellular Carcinoma | 7 | 577,371 | 429 |
| Lung Adenocarcinoma | 12 | 1,684,659 | 1,095 |
| Lung Squamous Cell Carcinoma | 12 | 1,425,892 | 1,035 |
| Mesothelioma | 8 | 665,207 | 173 |
| Ovarian Serous Cystadenocarcinoma | 9 | 137,987 | 591 |
| Pancreatic Adenocarcinoma | 9 | 833,255 | 368 |
| Pheochromocytoma and Paraganglioma | 9 | 890,115 | 360 |
| Prostate Adenocarcinoma | 10 | 1,283,738 | 1,029 |
| Rectum Adenocarcinoma | 12 | 865,885 | 316 |
| Sarcoma | 8 | 1,000,984 | 516 |
| Skin Cutaneous Melanoma | 8 | 1,389,241 | 938 |
| Stomach Adenocarcinoma | 11 | 1,324,766 | 906 |
| Testicular Germ Cell Tumors | 9 | 716,300 | 271 |
| Thyroid Carcinoma | 11 | 1,035,879 | 1,013 |
| Thymoma | 8 | 678,085 | 248 |
| Uterine Corpus Endometrial Carcinoma | 11 | 1,372,635 | 1,083 |
| Uterine Carcinosarcoma | 9 | 642,449 | 111 |
| Uveal Melanoma | 10 | 647,998 | 160 |
| *Note***.** All datasets are available for download at <http://tinyurl.com/maeourls>. \*Specimen counts include tumors, normal tissues, and replicate assays. | | | |

**Supplemental Table 3: Essential elements of the MultiAssayExperiment data class.**

|  |  |
| --- | --- |
| **Structure** | **Description** |
| MultiAssayExperiment | * coordinated selection of biological units and features * coordinated extraction of data matrices * operations spanning experiments * genomic range-based and ID-based subsetting |
| colData | * A DataFrame providing information on patients or biological units; one row per patient |
| ExperimentList | * a list of ”omics” datasets that support “[“ (bracket subsetting), “dimnames”, and “dim”. May be indexed by identifiers or genomic ranges. |
| sampleMap | * A three-column edge list relating colData rows to ExperimentList columns; one row per genomic profile. Enables different sample identifiers for each experiment, replicate and missing assays |

# References

1. [Hernandez-Ferrer C, Ruiz-Arenas C, Beltran-Gomila A, González JR. MultiDataSet: an R package for encapsulating multiple data sets with application to omic data integration. BMC Bioinformatics. 2017;18:36.](http://paperpile.com/b/rad3b2/k3A3)

2. [Broad Institute MIT. Firehose Broad GDAC Download [Internet]. Available from:](http://paperpile.com/b/rad3b2/rlKJJ) <https://confluence.broadinstitute.org/display/GDAC/Download>

3. [Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. Cancer Res. 2012;72:3499–511.](http://paperpile.com/b/rad3b2/4cr9x)

4. [Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, et al. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. Mol Cancer Ther. 2007;6:820–32.](http://paperpile.com/b/rad3b2/9gwOI)

5. [Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013;45:353–61, 361e1–2.](http://paperpile.com/b/rad3b2/TrXQm)

6. [Schoeps A, Rudolph A, Seibold P, Dunning AM, Milne RL, Bojesen SE, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. Genet Epidemiol. 2014;38:84–93.](http://paperpile.com/b/rad3b2/X1YUP)

7. [Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 2014;10:e1004663.](http://paperpile.com/b/rad3b2/4mju8)

8. [Meng C, Kuster B, Culhane AC, Gholami AM. A multivariate approach to the integration of multi-omics datasets. BMC Bioinformatics. 2014;15:162.](http://paperpile.com/b/rad3b2/q3ado)

9. [Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC. Dimension reduction techniques for the integrative analysis of multi-omics data. Brief Bioinform. 2016;17:628–41.](http://paperpile.com/b/rad3b2/WfWcB)

10. [Meng C, Kuster B, Peters B, Culhane AC, Gholami AM. moGSA: integrative single sample gene-set analysis of multiple omics data [Internet]. bioRxiv. 2016 [cited 2017 Feb 1]. page 046904. Available from:](http://paperpile.com/b/rad3b2/q0qxf) <http://biorxiv.org/content/early/2016/04/03/046904>

11. [Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.](http://paperpile.com/b/rad3b2/UVxPQ)

12. [Verhaak RGW, Tamayo P, Yang J-Y, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J Clin Invest. 2012;123:517–25.](http://paperpile.com/b/rad3b2/4J4Xb)

13. [Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.](http://paperpile.com/b/rad3b2/XuY6o)

14. [Louhimo R, Hautaniemi S. CNAmet: an R package for integrating copy number, methylation and expression data. Bioinformatics. 2011;27:887–8.](http://paperpile.com/b/rad3b2/PsUt)