multiSLIDE: a query-driven heatmap visualization tool for multi-omics data

Soumita Ghosh[1,2,3], Abhik Datta[4,5], Hyungwon Choi[1,3,*]

1. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
2. Saw Swee Hock School of Public Health, National University of Singapore, Singapore
3. Institute of Molecular and Cell Biology, Agency for Science, Technology, Research, Singapore
4. Centre for BioImaging Sciences, National University of Singapore, Singapore
5. Department of Biological Sciences, National University of Singapore, Singapore

*Corresponding author: Hyungwon Choi (hwchoi@nus.edu.sg)

Running title: multiSLIDE: a web-based multi-omics visualization tool

Abbreviations

ANOVA: analysis of variance

API: application programming interface

CPDB: Consensus Path DB

CPTAC: Clinical Proteomic Tumor Analysis Consortium

CSS: Cascading Style Sheet

DTT: dithiothreitol

EGFR: epidermal growth factor receptor

ER: endoplasmic reticulum

ERBB2: erb-b2 receptor tyrosine kinase 2

ESR1: estrogen receptor alpha

GO: Gene Ontology

HGNC: HUGO Gene Nomenclature Committee

JSON: JavaScript Object Notation

PDI: protein disulfide isomerase

PPI: protein-protein interaction

RefSeq: Reference Sequence

SVG: scalable vector graphics

TCGA: The Cancer Genome Atlas

TF: transcription factor

TS: Type Script

UPR: unfolded protein response

Abstract

We present multiSLIDE, an open-source tool for query-driven visualization of quantitative single- or multi-omics data. Using pathways and networks as the basis for data linkage, multiSLIDE provides an interactive platform for querying the multi-omics data by genes, pathways, and intermolecular relationships. Representing individual -omics levels as separate heatmaps, multiSLIDE visualizes quantitative data for selected genes at all omics levels in a single snapshot. The tool also provides functionalities to arrange data both ways, by their phenotypic characteristics and by their molecular interactions or co-membership to common pathways. Built-in statistical tests and clustering methods provide display subsets of interesting genes or rearrange the genes based on expression patterns. All visualization panels are fully customizable, and both the graphics and the analysis workspace can be saved and shared between collaborating parties. We demonstrate the utility of multiSLIDE through two example studies. First, with a time-course data of HeLa cells, subjected to dithiothreitol induced endoplasmic reticulum stress, we visualized different stages of unfolded protein response and identified temporal patterns of gene expression response at the mRNA and protein levels. Second, through joint visualization of mRNA and protein expression of TCGA/CPTAC Invasive Breast Carcinoma data, we explored the estrogen receptor α regulon and prioritized clusters of genes associated with PAM50 basal-like subtype.

Keywords: Visualization, multi-omics

Introduction

Visualization is an important tool for understanding any data. Large-scale data sets from high-throughput -omics platforms, such as massively parallel sequencing or mass spectrometry, often require significant reduction of data using statistical filters or abstraction via a projection of data into a lower dimensional space to facilitate interpretation. Although data reduction is unavoidable for effective presentation of data, our dependence on these filters blinds us from other intrinsic features that fail to pass the filter. Therefore, there is need for visualization tools that enable detailed exploration of full data set prior to data filtering. Easy-to-use, versatile visualization tools are all the more in increasing demand, especially for emerging multi-omics data sets. With the leading -omics technologies maturing, we have learned that different omics-level measurements often lead to concordant and discordant observations from the same biological samples (1,2), although ultimately the most relevant data for biological phenomenon is the expression of the final gene products, i.e. proteins (3).

Indeed, there already exists a wealth of bioinformatics tools for multi-omics data visualization. **Supplementary Table S1** gives a summary of existing tools in four categories: open-source data portals, networks-based tools, pathway-based tools, and heatmap-based omics integration tools. Open-source, data-rich web resources such as cBioPortal (4), UCSC Xena (5) , and LinkedOmics (6) have made it possible for systematic exploration and visualization of public cancer-omics datasets. However, most of these tools, apart from UCSC Xena, do not visualize user's own data. Moreover, visualization in these data portals are predominantly single gene-based or key cancer driver gene-based exploration modules.

In pathway-based visualizations (7–11), pathway diagrams are augmented with quantitative data to enable users to gain more meaningful insights for phenotypic variations at the level of gene groups. However, it is challenging to visualize interconnections between omics layers within the pathway-based visualization modules. Network-based visualization approaches (12–21) are popular alternatives for visualization of complex interconnectivities between biomolecules, yet visualized networks often run into the "hairball" problem even with a few hundred nodes (22,23).

83 Integrated heatmap-based visualizations (24,25) can help disentangle the crowding of network
84 diagrams without overwhelming the graphical interface. While heatmaps have not been the primary mode
85 of representation in multi-omics data visualization tools, many tools often include heatmaps as an additional
86 visualization. For instance, PaintOmics3 supplements their pathway diagrams with additional heatmaps
87 displaying omics expression profiles, where the interactivity and data handling capacity of these heatmaps
88 are usually restricted even for moderately sized data sets, limiting its utility.

89 Motivated by the difficulties presented in the existing tools, we developed multiSLIDE, an
90 interactive data visualization tool for easy exploration of multi-omics data. In multiSLIDE, quantitative
91 multi-omics data are visualized for genes in specific pathways or gene ontology (GO) categories preselected
92 by the user, simultaneously at different molecular levels. The tool displays the abundance measurements of
93 DNA copy number, mRNA transcripts, and proteins in functional groups (e.g. pathways and GO terms)
94 selected by the user in aligned panels of heatmaps, synchronizing them by gene identifiers and sample
95 names.

96 multiSLIDE also integrates biological networks in queries, including transcription factor (TF)
97 regulatory networks and protein-protein interaction networks, which capture the dependencies across and
98 within molecular levels, respectively. These networks can be quickly queried for genes of interest, which
99 allows the user to add other genes interacting with them to the current heatmap visualization in real time.
100 At the same time, sample specific phenotype labels are visualized as side bars in the heatmaps. This
101 integrated visualization enables exploring molecular profiles linked with phenotypes at various omics levels
102 concurrently. The visualizations in multiSLIDE are fully customizable, with multiple ordering options for
103 samples and genes, such as hierarchical clustering, multi-level phenotypic sorting, and statistical filtering.

104 We demonstrate multiSLIDE using two publicly available multi-omics datasets. In the first case
105 study, we visualize the time-course mRNA and protein expression data in HeLa cells subjected to
106 endoplasmic reticulum (ER) stress (26). The joint analysis reveals time-dependent patterns of unfolded
107 protein response (UPR), distinctly regulated at the mRNA and protein levels. In the second case study, we
108 visualized mRNA and protein data for 73 tumors from TCGA/CPTAC invasive ductal breast carcinoma

109 (TCGA-BRCA) cohort, aiming to visualize key hormone receptor ESR1 and growth factor EGFR proteins

110 in the four intrinsic molecular subtypes (27).

111

Experimental Procedures

The Visualization Workflow

Figure 1A shows the web-based visualization interface of multiSLIDE. In multiSLIDE, data analysis begins with the user's selection of pathways, GO terms, or individual genes. multiSLIDE provides an intuitive keyword-based search syntax for searching multiple pathways, GO terms, and genes. The user selects relevant genes and gene groups from the search results and those genes are visualized with clicks on the group names, as shown in **Figure 1B**. In addition, the user can choose to add network neighbors of a target gene on protein-protein interaction (PPI) and TF regulatory networks via a network neighborhood search, all enabled by a simple right-click on the gene of interest. Selected network neighbors are added to the visualization in real time (**Figure 1C**). Alternatively, the user can visualize genes in GO terms or pathways, with the side tracks indicating the membership of genes to the functional groups (side bars on the left side of heatmaps, **Figure 1A**).

The scales and dynamic ranges of detection and quantification techniques can be inconsistent for different -omics data. Therefore, we made the graphical parameters customizable in each omics data separately. The user can set a suitable binning range, select different color schemes differently in individual omics panels (see heatmap settings panel, **Figure 1D**). Settings applicable to all heatmaps, such as zoom (or resolution) and orientation of heatmaps, are applied to all heatmaps simultaneously using the global settings panels (**Figures 1E** and **1F**). The heatmaps in different panels (different omics data) are anchored by the genes and samples so as to enable simultaneous scrolling at any instant, using the scrolling panel (**Figure 1G**).

multiSLIDE has no restrictions on the amount of data that can be viewed in a single snapshot, although the screen size of the user's computer has natural limits to which the visualization can be effective. As different systems and browsers have variable computing capabilities, this choice is left to the user. Using the layout panel (**Figure 1E)**, the size of a single snapshot can be optimized, depending on the data transfer rate between the multiSLIDE server and the browser, and the browser's latency in rendering the data.

multiSLIDE has multiple sorting, clustering, and filtering methods built in to help users discover patterns in the data. Interesting genes can be difficult to discern when they are incoherently mixed with other genes, particularly when visualizing large pathways or networks. With the appropriate ordering of genes and samples, however, previously undetectable structures in the data can become apparent. In multiSLIDE, features can be sorted by gene groups, based on significance level in differential expression analysis, or based on hierarchical clustering. Samples can be ordered by a combination of phenotypes or based on hierarchical clustering. The hierarchical clustering can be further modified by selecting different linkage functions, distance metrics, and leaf ordering schemes, although the dendrograms are not visualized.

The user can also remove statistically non-significant genes from the panels using differential expression analysis. In large pathways and networks, a substantial number of genes are not differentially expressed between sample groups, and removal of these genes may improve the visualization clarity. The statistical test for differential expression analysis depends on the type of the selected phenotype. For binary, categorical, and continuous phenotypes, two-sample t-test, analysis of variance (ANOVA), and linear least squares regression are performed, respectively. multiSLIDE automatically classifies phenotypes into one of the three variable types.

The changes in the customizations are reflected in the visualizations in real time. The user can create curated feature lists by adding individual genes or gene groups from various heatmap panels. The "User List" panel in the far-right side of **Figure 1A** shows an example gene list curated by the user. These gene subsets can be easily re-visualized within the analysis with a single click. Feature lists can also be created by uploading a list of gene identifiers as a delimited text file. The user can also maintain multiple save points in multiSLIDE, to which they can go back to and restart the analysis. The visualizations can be saved as high-resolution PDF files, and the analysis workspaces with customized heatmaps can be saved as ".mslide" files. The feature lists can be also exported as text files. The saved analysis workspaces can later be loaded back into multiSLIDE for continued analysis, or simply to revisit the analysis steps leading to the current visualizations. This functionality makes multiSLIDE a useful tool to share analysis results in collaborative projects.

163

164 ***Software Design***

165 *Databases: Pathways, GO terms and Molecular Networks.* Comprehensive genome-wide annotations and
166 gene ontology databases for mouse and human genes were extracted using libraries in R Bioconductor
167 (28,29). Data is highly structured in these R packages and are routinely used by bioinformaticians to analyze
168 their data. multiSLIDE recognizes Entrez, HUGO Gene Nomenclature Committee (HGNC) Gene Symbols,
169 ENSEMBL identifiers, NCBI Reference Sequence (RefSeq) identifiers and UniProt identifiers. Internally
170 within multiSLIDE, the gene identifiers are standardized by conversion to Entrez.

171 Comprehensive biological pathways were obtained from ConsensusPathDB (CPDB)
172 (http://cpdb.molgen.mpg.de/) (30,31). Validated miRNA - target interactions on pathways and GO were
173 extracted from miRWalk2.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) (32).

174 Various networks indicating relationships between molecules within the same molecular level such
175 as PPI network (within proteins), as well as networks indicating relationships between molecules at
176 different levels such as TF regulatory networks, have been integrated in multiSLIDE. The networks come
177 from diverse sources and the interactions are either experimentally validated or computationally predicted.
178 multiSLIDE also integrates Human Transcription Factor (TF) - targets network information from additional
179 databases: TRED (33), ITFP (34), ENCODE (35), Neph2012 (36), TRRUST (37), Marbach2016 (38).
180 Mouse Transcription Factor (TF) - targets network information was obtained directly from TRRUST.
181 Physical interactions between proteins was sourced from iRefIndex
182 (http://irefindex.org/wiki/index.php?title=iRefIndex) (39), which indexes protein-protein interaction
183 networks from a number of databases. multiSLIDE also includes miRNA-mediated gene regulation
184 information from TargetScan (http://www.targetscan.org) (40), a database housing mostly predicted targets
185 of miRNAs.

186

187 *Software Implementation.* multiSLIDE is built on a three-tier, client-server architecture separating the core
188 computational logic, user interface and data storage tiers. A multi-tier architecture has the flexibility of

189 developing and extending each tier independently without affecting the other, provided that the tiers
190 communicate with each other using predefined application programming interfaces (APIs). Moreover, such
191 architectures facilitate highly parallel communication. Similar to SLIDE, a tool we had developed for on-
192 line visualization of single -omics data (41), multiSLIDE is available online and can also be used as a
193 standalone software. Due to its modular design, multiSLIDE can also easily scale to distributed multi-node
194 environments.

195 **Supplementary Figure S1** shows a schematic view of multiSLIDE's software architecture. The
196 server side of multiSLIDE consists of an HTTP server and a database server. The client can be any modern
197 web browser. The HTTP server hosts the analytics module, which is the main computation engine, as well
198 as the graphics server. The analytics module carries out the bulk of the computation. The HTTP server and
199 the browser communicate via highly optimized JavaScript Object Notation (JSON) objects. The data tier,
200 implemented using MongoDB, manages the physical storage of all curated gene annotation, regulatory
201 networks, biological pathways and Gene Ontology (GO) tables. In multiSLIDE, individual components
202 within each tier are also highly compartmentalized. At the client, the data and presentation layers are also
203 decoupled. Layouts and graphics can therefore be altered without the need for fetching the data again from
204 the server. User interactions and styles were implemented using TypeScript (TS) and Cascading Style
205 Sheets (CSS). The visualizations are rendered using resolution independent Scalable Vector Graphics
206 (SVG).

207 A key design philosophy in multiSLIDE is to visualize only user queried genes (molecules). As a
208 result of the extensive user interactions available in multiLSIDE, there is frequent communication between
209 client and server. At the same time, owing to multiple -omics datasets, the server has to manage a much
210 larger amount of data. In multiSLIDE, lazy execution combined with memoization is extensively used to
211 balance the server-side memory footprint and response times. The computation intensive modules are
212 developed by leveraging the advantages of Java and Python. Aggressive caching in the browser and using
213 delta loads for data transfer gives multiSLIDE a desktop-like user experience.

214

215 *Input Data Format.* In multiSLIDE, the user creates an analysis by uploading a set of required input files.
216 Inputs to multiSLIDE should be delimited ASCII text files (one for each -omics dataset), containing
217 quantitative measurements across samples. These files can be created and edited using any text editor and
218 there is no restriction on the number of input files that can be uploaded for visualization. Measurements can
219 be counts, categorical data or continuous data that have already undergone standard pre-processing and
220 transformations. The databases integrated into multiSLIDE contain functionally annotated genes assigned
221 to pathways and GO terms. To fully utilize multiSLIDE's querying capabilities, individual -omics input
222 files should have at least one column with standard gene identifiers available in multiSLIDE. Also, high
223 throughput -omics data, such as DNA methylation, are molecular features at specific genomic coordinates.
224 Since multiSLIDE's search capabilities are gene-based, for omics data with no designated gene identifiers
225 (e.g. sequence variants, CpG islands in DNA methylation), genomic coordinates of the molecular features
226 have to be mapped to the nearest gene and labeled as such in the current version of the software.
227 In addition, a separate file containing sample attributes (e.g. clinical data) is also required,
228 formatted as a separate delimited ASCII text file. The information in this file should map the samples in -
229 omics data files to their corresponding phenotype information. The attributes file may include optional
230 sample information such as descriptive sample names, replicate names, and time points.
231
232 ***Data preprocessing for TCGA/CPTAC breast cancer data***
233 Genes that had fewer than 80% samples expressed in Luminal A, Luminal B, and basal-like subtypes in
234 both -omics level were removed. For the HER2-enriched subtype, genes with fewer than ten samples
235 expressed at either -omics levels were removed from the corresponding -omics level. For the remainder of
236 genes, weighted k-nearest neighbor method, where $k = 10$, (42) was used to impute the missing data prior
237 to importing into multiSLIDE. The mRNA expression values were log-transformed (base 2) and each gene
238 was centered by the mean of their expression levels in multiSLIDE visualization. For the protein level data,
239 the iTRAQ ratios were also median centered for visualization.

Results

We present two case studies using previously published datasets to illustrate multiSLIDE's features. The first case study explores multiple pathways and GO terms simultaneously in a time-course multi-omics dataset. The second case study illustrates exploration of the TF regulatory networks and statistical filtering-based gene prioritization functionalities on a paired transcriptomic and proteomic TCGA/CPTAC breast cancer dataset. In addition, we also demonstrate the exploration of the PPI network around a receptor tyrosine kinase HER2/ERBB2 on the same dataset in **Supplementary Figure S2**.

UPR reveals independent gene expression regulation at the mRNA and protein level

Using multiSLIDE, we first visualize the time-course mRNA and protein expression data of HeLa cells in response to dithiothreitol (DTT) treatment, which induces ER stress (26). In the study, cells were sampled at eight time points (0, 0.5. 1, 2, 8, 16, 24, and 30h) and their transcriptome and proteome were measured. Both -omics data were normalized by dividing the measurements at post-treatment time points by their respective measurement at 0h. During the ER stress, we expect to observe complex signaling cascades involved in UPR, translation attenuation, ER-associated protein degradation, and cellular apoptosis (43–46).

Before we explore this multi-omics data, we first visualize the entire data sets separately using the SLIDE tool, another related tool that we had previously developed for full-scale single -omics data visualization (41). We applied hierarchical clustering with Euclidean distance and complete linkage to the whole transcriptome data comprising of 16,704 genes in SLIDE (**Supplementary Figure S3A**). The global view of the clustering profiles in **Supplementary Figure S3A** shows three phases of the ER stress response characterized by Cheng *et al*: early phase (< 2 h), intermediate phase (2 - 8 h), and late phase (> 8 h). The whole mRNA expression regulation suggested a spike-like pattern in the transition from the early phase to the intermediate phase, peaking in the intermediate phase before returning to original levels in the late phase.

265 One of the direct consequences of ER stress is the aggregation of misfolded, unassembled proteins
266 in the organelle. As a survival mechanism to avert the loss of homeostasis, the ER responds by increasing
267 the protein folding capacity. This activated pro-survival cellular mechanism, otherwise known as the UPR,
268 is involved in extensive reprogramming of the transcriptional and translational regulation (43,47,48). First,
269 to identify action patterns of UPR related genes, the global view in **Supplementary Figure S3A** was tagged
270 to highlight genes belonging to the GO term 'endoplasmic reticulum unfolded protein response' (green bars
271 to the right of the heatmap). As an adaptive response, a hallmark of UPR is to reduce ER stress and restore
272 homeostasis by the coordinated transcriptional upregulation of ER chaperones and protein folding enzymes.
273 The initial visual inspection of the whole mRNA data confirmed the upward expression trends of the ER
274 chaperones, heat shock protein family A (Hsp70) member 5 (HSP5A/GRP78/BiP) and folding chaperones
275 such as protein disulfide isomerase (PDI) family genes (**Supplementary Figure S3B**).
276 Cheng *et al*. also analyzed the dynamics of 1,237 mRNA/protein pairs that passed the filtering for
277 missing data and noise. **Supplementary Figure S3C** shows the protein abundances of these 1,237 proteins
278 in SLIDE. Most UPR related proteins expressed upward trends in the proteomics data, similar to the mRNA
279 expression regulation, but with a delayed temporal response. The global heatmap clearly shows that this
280 up-regulation persisted even at the late phase of the stress response.
281 To investigate these patterns in the transcriptomic and proteomic data simultaneously, we loaded
282 the data sets into multiSLIDE. Hierarchical clustering of UPR-related genes at the mRNA level in
283 multiSLIDE reveals a cluster of upregulated genes. The master sensor of misfolded proteins in the ER,
284 HSPA5 (also known as GRP78 or BiP), is a member of the heat shock protein 70 family, which is
285 upregulated at both mRNA and protein level. In **Figure 2**, the mRNA level data shows early up-regulation
286 of HSPA5, at 0.5 h of stress induction. HSPA5 regulates the activation of ER stress transducers, including
287 PERK (protein kinase R-like ER kinase, also known as EIF2AK3/PEK), inositol-requiring 1 (IRE1) and
288 activating transcription factor 6 (ATF6/ACHM7). Under basal conditions, HSPA5 keeps the three ER stress
289 transducers inactive (47). The downstream effectors of these three key UPR regulators converge on
290 promoting ER chaperone synthesis, ER-associated protein degradation and ER membrane biogenesis (49).

291 Observing the dynamics of transcriptome regulation in multiSLIDE in **Figure 2**, we also notice the
292 early phase activation of ER chaperones, heat shock protein 90 beta family member 1 (HSP90B1),
293 DNAJC3/P58IPK (member of the HSP40 chaperone family) and protein disulfide isomerases (PDIs),
294 PDIA4, PDIA6 and PDIA3, that remains activated in the intermediate and late phases. PDIs are known to
295 be responsible for the oxidation (formation), reduction (break down) and isomerization (rearrangement) of
296 protein disulfide bonds via disulfide interchange activity. The other major role of PDIs is in general
297 chaperone activity and recent studies have also identified PDI for its role as PERK activator (50–52). In
298 **Figure 2**, it is interesting to observe that mRNA-level upregulation is countered by protein-level down-
299 regulation of DnaJ heat shock protein family (Hsp40) member C3 (DNAJC3), which is a known inhibitor
300 of PERK (53). This suggests that an active feedback control defense mechanism is in place to mitigate ER
301 stress, allowing the PERK pathway to remain uninhibited (54,55).
302 The reversal of expression levels in the ER resident proteins SELK and MANF to their pre-
303 treatment levels at the very late phase can also be attributed to the ER stress attenuation mechanism of UPR.
304 Up-regulation of selenoprotein K (SELK) and mesencephalic astrocyte-derived neurotrophic factor
305 (MANF) are a protective mechanism to avoid ER stress mediated cell death.
306 In summary, jointly visualizing the mRNA- and protein-level expression data in multiSLIDE helps
307 the user to uncouple the distinct expression regulation patterns at different response phases of UPR.
308 Clustering genes at the mRNA level and applying the same ordering at the protein level helped visualize
309 whether the clusters propagate across -omics levels. An activated UPR initiates an adaptive stress response
310 to regulate downstream effectors and further through a feedback control switches on/off transcriptional
311 regulation and protein synthesis to restore ER homeostasis. A failure to attain homeostasis leads to
312 programmed cell death (56,57).
313
314 **Exploring ESR1 regulon in basal-like breast cancer tumors**
315 In recent years, genome-scale characterization of breast cancer subtypes has helped us understand its tissue
316 heterogeneity and identify additional therapeutic targets such as human epidermal growth factor receptor 2

317 (HER2/ERBB2). The four major intrinsic subtypes of breast cancer are defined as luminal A, luminal B,
318 HER2–enriched and basal-like (27).

319 However, immunohistochemistry (IHC)-based surrogate subtyping, which are still routinely used
320 for classification in pathological settings, is based on three key hormone receptors: estrogen receptor
321 (ESR1), progesterone receptor (PGR), and HER2 (58). Basal-like tumors tend to have the worst prognosis,
322 with all three receptors negative, making them unresponsive to endocrine therapy. Specifically, the
323 expression of ESR1 is widely linked to better survival and considered a major regulator of the phenotypic
324 properties of these breast cancers. ESR1 is a hormone-regulated transcription factor, a member of a
325 superfamily of nuclear receptors, and plays a key role in cell proliferation. Typically, cell growth in breast
326 cancer is stimulated either by the hormone estrogen (17β-estradiol) or growth factors such as EGF.
327 Interestingly, transcription of EGF receptor is regulated by ESR1. Therefore, exploring the behavior of TF
328 targets of ESR1, particularly in the basal-like subtype, can reveal genes that bypass ESR1 negativity.

329 We visualized proteomics data from CPTAC (59) and transcriptomics data (60) from TCGA
330 invasive ductal breast carcinoma (TCGA-BRCA) in multiSLIDE. To understand the role of ESR1 as a
331 master regulator of pathways, we first query and visualize all its transcription target genes at the mRNA
332 and protein levels in multiSLIDE. A query for the TF targets of ESR1 in multiSLIDE's databases resulted
333 in a total of 3,787 targets, of which the 2,312 unique targets present in the datasets were included in the
334 analysis.

335 Next, using the PAM50 subtype classification as phenotype, we performed differential expression
336 analysis at the protein level in multiSLIDE. With (unadjusted) p-value < 0.001 from analysis of variance
337 as the statistical filter, we were able to select 172 differentially expressed genes. We applied hierarchical
338 clustering to these genes with Euclidean distance and complete linkage, resulting in four distinct clusters
339 consistent at mRNA and protein level. For visual clarity, we visualized the four clusters in two parts, as
340 shown in **Figures 3A and 3B** (the side track on top of each heatmap indicating the subtypes). The genes in
341 **Figure 3A** show two distinct clusters of genes (top-half and bottom-half of the figure) where the ESR-
342 tumors (basal-like and HER2-enriched subtypes) are predominantly down-regulated (blue) while the other

343 subtypes are mostly up-regulated (red), across both -omics levels. These genes had signatures that were
344 strongly coherent with that of ESR1. The clusters shown in **Figure 3B** had expression patterns that were
345 opposite to that of ESR1 across subtypes. The latter of the two clusters (bottom-half of **Figure 3B**) had
346 pronounced up-regulation only in the basal-like subtype but no distinct signature in the other subtypes.

347 In **Figure 3A**, the two clusters included key tumorigenic genes down-regulated in basal-like tumor
348 samples and some HER2-enriched samples, such as progesterone receptor (PGR), carbonic anhydrase 12
349 (CA12), androgen receptor (AR), growth regulating estrogen receptor binding 1 (GREB1), and GATA
350 binding protein 3 (GATA3) that control cell proliferation. The under-expression of ESR1 and these targets
351 in the basal-like subtype suggests the role of alternative pathways for growth and proliferation of basal-like
352 tumors, bypassing ESR1-dependent stimulation. One such pathway is epidermal growth factor (EGF)-
353 induced tumor growth. In **Figure 3B**, EGFR is over-expressed in the basal-like subtype at both mRNA and
354 protein levels. An overexpression of EGF receptor (EGFR) drives an aggressive form of cell proliferation,
355 which is often associated with poor survival (61–63). In addition, **Figure 3B** also shows that the FAT
356 atypical cadherin 1 (FAT1) gene, a member of the cadherin superfamily, widely considered as a tumor
357 suppressor, is overexpressed in the basal-like subtype at both mRNA and protein levels. This is indeed
358 consistent with recent literature indicating a more multifaceted role for FAT1, suggesting it as either a tumor
359 suppressor or tumor promoter in a context dependent manner (64,65). Most of the other genes in **Figure**
360 **3B**, that are upregulated in the basal-like subtype are known for their role in cell motility, growth and
361 positive regulation of GTPase activity and has to do with tumor survival, migration and metastasis.

362 Furthermore, using a network-based integration method for multi-omics data, Koh *et al*. identified
363 overexpressed TFs such as CCAAT enhancer binding protein beta (CEBPB), nuclear factor I X (NFIX),
364 WW domain containing transcription regulator 1 (WWTR1), and WD repeat domain 74 (WDR74), as
365 unique drivers of basal-like subtypes (66). Among the targets of ESR1 in **Figure 3B**, we also observed
366 CEBPB and NFIX were overexpressed at both -omics levels. Given that these are TFs that in turn affect
367 expression of their target genes, their overexpression in the basal-like subtype is suggestive that they play
368 a role as key regulators of cell proliferation and aggressive tumor development in the subtype.

369 **Supplementary Figures S4A and S4B** visualizes the targets of these four transcription factors: CEBPB,
370 NFIX, WWTR1, WDR74. Statistical filtering and hierarchical clustering of this data revealed a cluster
371 downregulated in the ESR- basal-like subtype (**Supplementary Figures S4A**) as well as a cluster
372 upregulated in the ESR- basal-like subtypes (**Supplementary Figures S4B**). The TF search keys are
373 highlighted by green arrows in **Supplementary Figures S4B**.

374 In summary, concurrent visualization of the curated ESR1 regulon both at the protein and mRNA
375 level reveals target genes that are under-expressed in basal-like subtype due to ESR1 negativity, consistent
376 at both molecular levels. At the same time, we were able to zoom into a set of specific target genes that
377 bypass ESR1 negativity and are activated in the basal-like subtype. The same type of query-driven
378 inspection of multi-omics data can be generically extended to other applications.

379

380 # Discussion

381 In this paper, we described multiSLIDE, a new web-based tool for query-driven interactive heatmap-based
382 visualization of multi-omics data. With steady growth in multi-omics experiments across many domains of
383 biomedical research, it has become increasingly important to develop open-source analysis and
384 visualization tools. This was the primary motivation behind the development of multiSLIDE. We
385 demonstrated how multiSLIDE enables targeted exploration of large multi-omics datasets within a
386 biological context. From a practical point of view, multiSLIDE was designed to treat both the data analysis
387 and visualization as resources to be shared and disseminated for collaborative research, a feature that is
388 often missing in current tools.

389 Methods in multi-omics integration are often devised making many assumptions. For instance, in
390 correlation-based integration methods, genes with correlated expression profiles across -omics levels are
391 considered closely associated. However, there could be numerous other factors such as time delay in
392 response between -omics levels or proteo-static regulation of protein expression that violates this
393 assumption. At the same time, anti-correlated patterns may be of interest as well. In the ER stress time-
394 course multi-omics data, we identified the example of DNAJC3 gene, which was up-regulated at the
395 transcriptome level but down-regulated at the protein level. Such patterns, if not known beforehand, may
396 not be accounted for in the assumptions and thus remain undiscovered. Visualizing pathways in their
397 entirety is a useful alternative for exploring the complex response patterns.

398 In multiSLIDE, the ordering of genes across all -omics datasets can be determined by clustering
399 the expression profiles of any one of them. This sorting functionality, imposed on one of the datasets and
400 propagated to the others simultaneously, can reveal clusters with concordant or discordant gene expression
401 regulation across different -omics layers. In the breast cancer data, for instance, the PAM50 classifier used
402 for identifying breast cancer subtypes was created using mRNA data (27). In **Figures 3A and 3B**, upon
403 hierarchical clustering in multiSLIDE, distinct clusters corresponding to PAM50 classifications were found

404 at the transcriptome level, as expected, but the same clusters were also visible at the proteome level just by
405 enforcing the same ordering as that of the transcriptome level.

406 Currently, multiSLIDE automatically links different datasets using standard gene identifiers.
407 Therefore, visualizing molecular features that are not gene products is not possible in the current
408 implementation, and we are currently exploring a new version of the tool to address this limitation. For
409 visualizing proteomics data and metabolomics data concurrently, for example, the map between genes and
410 metabolites has not been completely charted by experimental means. Hence an ideal architecture for a future
411 version of multiSLIDE may need to infer networks between genes and other types of molecules from the
412 data directly, in conjunction with known connections.

413 Other data types that may clash with gene-based summaries are found in existing -omics data such
414 as DNA methylation data, ChIP/ATAC-seq, and histone modification profiles to name a few, which cover
415 more vaguely defined units over the genome. For example, DNA Methylation at a gene promoter or
416 enhancer region acts to repress gene transcription. When multiple regions close to the gene are methylated
417 or in regions with dense population of genes, care has to be taken to denote them as gene promoter or gene
418 body. Such one to many mappings will require further modification to the software architecture and will be
419 included in future iterations of multiSLIDE.

420

421 Acknowledgments

422 This work was supported in part by grants from Singapore Ministry of Education (MOE2016-T2-1-001 and
423 MOE2018-T2-2-058 to H.C.), National Medical Research Council of Singapore (NMRC-CG-M009 to
424 H.C.) , and the support by the Institute of Molecular and Cellular Biology, Agency for Science, Technology
425 and Research.

426

427 Data and Software Availability

The source code, installation instructions and a demo online version of multiSLIDE are available at https://github.com/soumitag/multiSLIDE. The tool is accessible by any modern web-browser with the server deployed either in a distributed environment or installed locally.

References

1. McManus J, Cheng Z, Vogel C. Next-generation analysis of gene expression regulation–comparing the roles of synthesis and degradation. Mol Biosyst. 2015;11(10):2680–9.
2. Liu Y, Aebersold R. The interdependence of transcript and protein abundance: new data–new complexities. Mol Syst Biol. 2016 Jan 1;12(1):856.
3. Ebhardt HA, Root A, Sander C, Aebersold R. Applications of targeted proteomics in systems biology and translational medicine. Proteomics. 2015 Sep 1;15(18):3193–208.
4. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr 2;6(269):pl1–pl1.
5. Goldman M, Craft B, Kamath A, Brooks A, Zhu J, Haussler D. The UCSC Xena Platform for cancer genomics data visualization and interpretation. bioRxiv. 2018 Jan 1;326470.
6. Vasaikar S V, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. 2017/11/09. 2018 Jan 4;46(D1):D956–63.
7. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLoS Comput Biol. 2015 Aug 1;11(8).
8. Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, et al. KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol. 2005 Jul;138(3):1289–300.
9. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, et al. mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004 Mar 1;37(6):914–39.
10. Hernández-de-Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas Jr GJ, et al. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. Nucleic Acids Res. 2018/05/25. 2018 Jul 2;46(W1):W503–9.
11. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, et al. PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput Biol. 2015;11(2):e1004085.
12. Wang Q, Tang B, Song L, Ren B, Liang Q, Xie F, et al. 3DScapeCS: application of three dimensional, parallel, dynamic network visualization in Cytoscape. BMC Bioinformatics. 2013;14(1):322.
13. Jang Y, Yu N, Seo J, Kim S, Lee S. MONGKIE: an integrated tool for network analysis and visualization for multi-omics data. Biol Direct. 2016 Mar 18;11(1):10.
14. Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, et al. NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. Bioinformatics. 2009/10/16. 2009 Dec 15;25(24):3327–9.
15. Junker BH, Klukas C, Schreiber F. VANTED: A system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics. 2006;7(1):109.
16. Hu Z, Hung J-H, Wang Y, Chang Y-C, Huang C-L, Huyck M, et al. VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. Nucleic Acids Res. 2009 May 21;37(suppl_2):W115–21.
17. Xia J, Gill EE, Hancock REW. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat Protoc. 2015;10(6):823.
18. Theocharidis A, Van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout Express 3D. Nat Protoc. 2009;4(10):1535.
19. Zhou G, Xia J. OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. Nucleic Acids Res. 2018/06/07. 2018 Jul 2;46(W1):W514–22.
20. Pavlopoulos GA, O'Donoghue SI, Satagopam VP, Soldatos TG, Pafilis E, Schneider R. Arena3D: visualization of biological networks in 3D. BMC Syst Biol. 2008;2(1):104.
21. Shi Z, Wang J, Zhang B. NetGestalt: integrating multidimensional omics data over biological networks. Nat Methods. 2013 Jul;10(7):597–8.
22. Schulz H-J, Hurter C. Grooming the hairball-how to tidy up network visualizations? In 2013.
23. Nocaj A, Ortmann M, Brandes U. Untangling hairballs. In: International Symposium on Graph Drawing. Springer; 2014. p. 101–12.
24. Lex A, Streit M, Schulz H, Partl C, Schmalstieg D, Park PJ, et al. StratomeX: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. In: Computer graphics forum. Wiley Online Library; 2012. p. 1175–84.

25. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PLoS One. 2011 May 13;6(5):e19541–e19541.
26. Cheng Z, Teo G, Krueger S, Rock TM, Koh HW, Choi H, et al. Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. Mol Syst Biol. 2016;12(1):855–855.
27. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27(8):1160.
28. Carlson M. org. Hs. eg. db: Genome Wide Annotation for Human. R package version 3.8.2. 2019.
29. Carlson M. org. Mm. eg. db: Genome wide annotation for Mouse. R package version 3.8.2. Bioconductor. 2019.
30. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. Nucleic Acids Res. 2010/11/11. 2011 Jan;39(Database issue):D712–7.
31. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2012/11/10. 2013 Jan;41(Database issue):D793–800.
32. Dweep H, Gretz N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat Methods. 2015 Jul 30;12:697.
33. Jiang C, Xuan Z, Zhao F, Zhang MQ. TRED: a transcriptional regulatory element database, new entries and other development. Nucleic Acids Res. 2007 Jan;35(Database issue):D137–40.
34. Zheng G, Tu K, Yang Q, Xiong Y, Wei C, Xie L, et al. ITFP: an integrated platform of mammalian transcription factors. Bioinformatics. 2008 Aug 19;24(20):2416–7.
35. Consortium TEP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 5;489:57.
36. Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. Circuitry and dynamics of human transcription factor regulatory networks. Cell. 2012/09/05. 2012 Sep 14;150(6):1274–86.
37. Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, et al. TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep. 2015 Jun 12;5:11432.
38. Marbach D, Lamparter D, Quon G, Kellis M, Kutalik Z, Bergmann S. Tissue-specific regulatory circuits reveal variable modular perturbations across complex diseases. Nat Methods. 2016 Mar 7;13:366.
39. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics. 2008 Sep 30;9:405.
40. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Izaurralde E, editor. Elife. 2015;4:e05005.
41. Ghosh S, Datta A, Tan K, Choi H. SLIDE – a web-based tool for interactive visualization of large-scale –omics data. Bioinformatics. 2018 Jun 28;35(2):346–8.
42. Schwender H. Imputing Missing Genotypes with Weighted k Nearest Neighbors. J Toxicol Environ Heal Part A. 2012 Apr 15;75(8–10):438–46.
43. Ron D, Walter P. Signal integration in the endoplasmic reticulum unfolded protein response. Nat Rev Mol cell Biol. 2007;8(7):519.
44. Kim I, Xu W, Reed JC. Cell death and endoplasmic reticulum stress: disease relevance and therapeutic opportunities. Nat Rev Drug Discov. 2008;7(12):1013.
45. Oslowski CM, Urano F. Measuring ER stress and the unfolded protein response using mammalian tissue culture system. In: Methods in enzymology. Elsevier; 2011. p. 71–92.
46. Tabas I, Ron D. Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. Nat Cell Biol. 2011;13(3):184.
47. Rutkowski DT, Arnold SM, Miller CN, Wu J, Li J, Gunnison KM, et al. Adaptation to ER stress is mediated by differential stabilities of pro-survival and pro-apoptotic mRNAs and proteins. PLoS Biol. 2006;4(11):e374.
48. Walter P, Ron D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. Science (80-). 2011 Nov 25;334(6059):1081 LP – 1086.
49. Clarke R, Cook KL, Hu R, Facey COB, Tavassoly I, Schwartz JL, et al. Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. Cancer Res. 2012 Mar 15;72(6):1321–31.
50. Ferrari DM, SÖLING H-D. The protein disulphide-isomerase family: unravelling a string of folds. Biochem J. 1999;339(1):1–10.
51. Perri ER, Thomas CJ, Parakh S, Spencer DM, Atkin JD. The Unfolded Protein Response and the Role of Protein Disulfide Isomerase in Neurodegeneration . Vol. 3, Frontiers in Cell and Developmental Biology .

541 2016. p. 80.
542 52. Kranz P, Neumann F, Wolf A, Classen F, Pompsch M, Ocklenburg T, et al. PDI is an essential redox-
543 sensitive activator of PERK during the unfolded protein response (UPR). Cell Death &Amp; Dis. 2017 Aug
544 10;8:e2986.
545 53. Yan W, Frank CL, Korth MJ, Sopher BL, Novoa I, Ron D, et al. Control of PERK eIF2α kinase activity by
546 the endoplasmic reticulum stress-induced molecular chaperone P58IPK. Proc Natl Acad Sci.
547 2002;99(25):15920–5.
548 54. Rainbolt TK, Saunders JM, Wiseman RL. Stress-responsive regulation of mitochondria through the ER
549 unfolded protein response. Trends Endocrinol Metab. 2014;25(10):528–37.
550 55. Lebeau J, Saunders JM, Moraes VWR, Madhavan A, Madrazo N, Anthony MC, et al. The PERK Arm of the
551 Unfolded Protein Response Regulates Mitochondrial Morphology during Acute Endoplasmic Reticulum
552 Stress. Cell Rep. 2018 Mar 13;22(11):2827–36.
553 56. Schröder M, Kaufman RJ. ER stress and the unfolded protein response. Mutat Res Mol Mech Mutagen.
554 2005;569(1–2):29–63.
555 57. Sano R, Reed JC. ER stress-induced cell death mechanisms. Biochim Biophys Acta (BBA)-Molecular Cell
556 Res. 2013;1833(12):3460–70.
557 58. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical
558 characterization of the basal-like subtype of invasive breast carcinoma. Clin cancer Res. 2004;10(16):5367–
559 74.
560 59. Mertins P, Mani DR, Ruggles K V, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects
561 somatic mutations to signalling in breast cancer. Nature. 2016;534(7605):55.
562 60. Network TCGA, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, et al. Comprehensive
563 molecular portraits of human breast tumours. Nature. 2012 Sep 23;490:61.
564 61. Dickson RB, McManaway ME, Lippman ME. Estrogen-induced factors of breast cancer cells partially
565 replace estrogen to promote tumor growth. Science (80-). 1986;232(4757):1540–3.
566 62. Russo J, Fernandez S V, Russo PA, Fernbaugh R, Sheriff FS, Lareef HM, et al. 17-Beta-estradiol induces
567 transformation and tumorigenesis in human breast epithelial cells. FASEB J. 2006 Aug 1;20(10):1622–34.
568 63. Reis-Filho JS, Tutt ANJ. Triple negative tumours: a critical review. Histopathology. 2008;52(1):108–18.
569 64. Morris LGT, Kaufman AM, Gong Y, Ramaswami D, Walsh LA, Turcan Ş, et al. Recurrent somatic
570 mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. Nat Genet. 2013 Jan
571 27;45:253.
572 65. van Roy F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer.
573 2014 Jan 20;14:121.
574 66. Koh HWL, Fermin D, Vogel C, Choi KP, Ewing RM, Choi H. iOmicsPASS: network-based integration of
575 multiomics data for predictive subnetwork discovery. NPJ Syst Biol Appl. 2019;5(1):22.
576
577

578 # Figure Legend

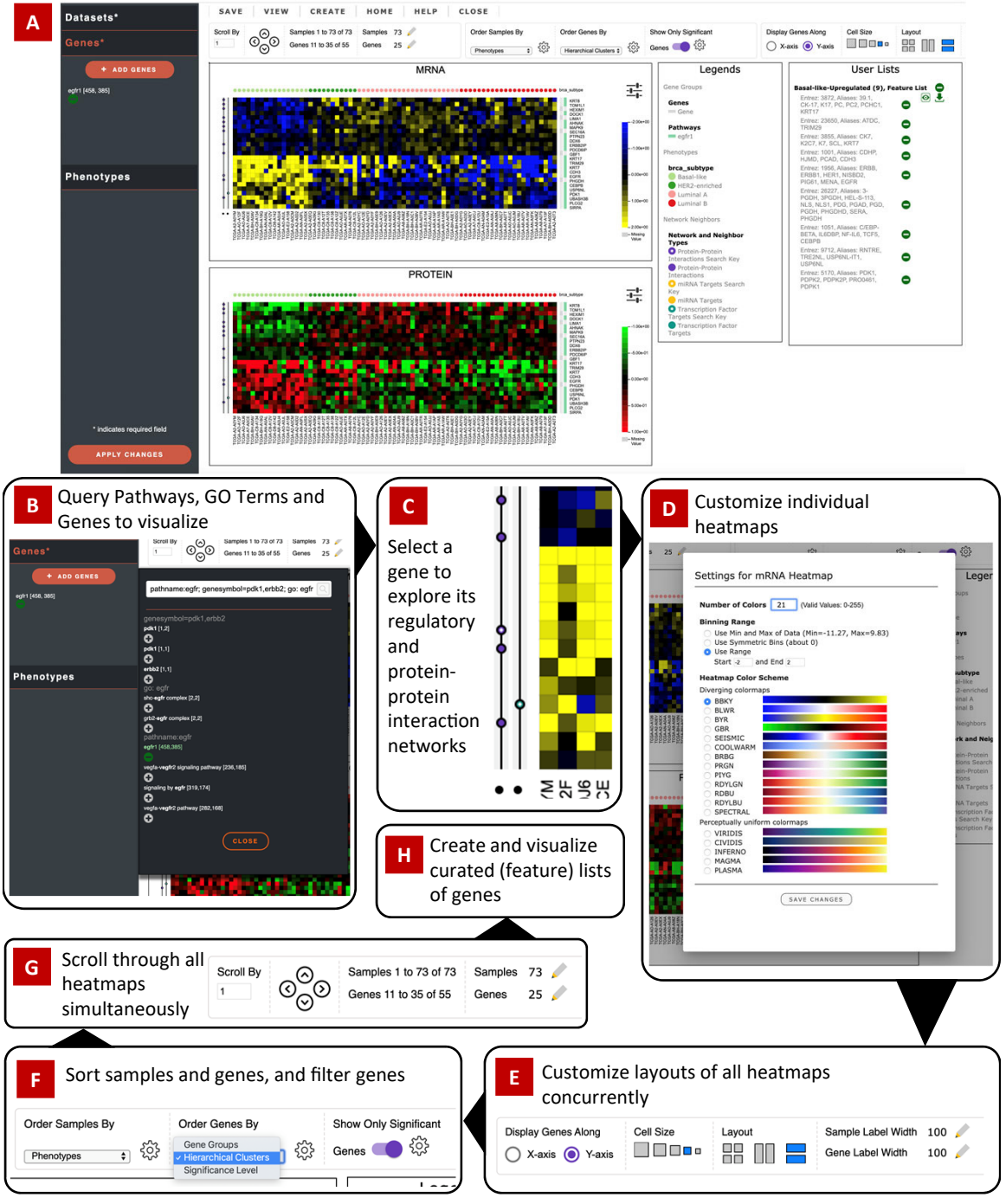
579

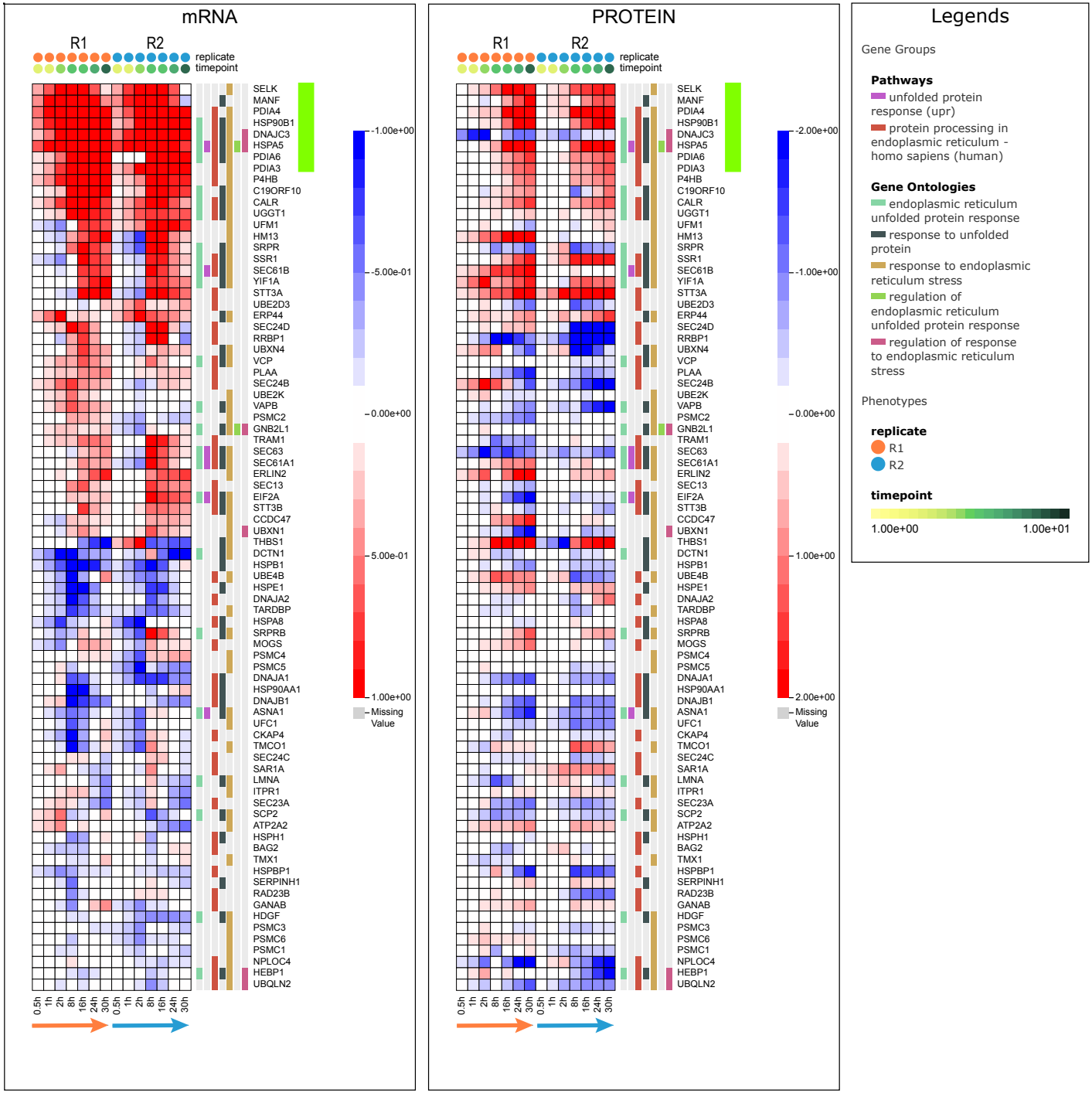
580 **Figure 1: Visualization workflow of multiSLIDE. (A)** The web-based interface of multiSLIDE,
581 visualizing expression patterns of 25 significant genes in the 'EGFR1' pathway across mRNA and protein
582 levels. In multiSLIDE, sample phenotypes, gene groups, and user-selected molecular interactions are
583 visualized alongside the expression profiles. Here, phenotypes, molecular interactions, and gene groups are
584 shown above, to the left of, and to the right of the heatmap, respectively. **(B)** Visualization in multiSLIDE
585 starts with the user selecting datasets, gene groups (relevant pathways, GO terms, or genes), and phenotypes
586 to visualize using the Selection Panel. Multiple gene groups can be simultaneously searched and selected
587 using an intuitive search syntax. **(C)** Putative transcription factor targets or physically interacting proteins
588 of a selected gene can be queried with a single right-click on the gene name. Interesting regulons or their
589 interacting genes can be selected from the list and added to the heatmaps, with the molecular interactions
590 visualized as tracks alongside the heatmaps. **(D)** For each -omics level heatmap, the data range, the number
591 of color bins, and color scheme can be adjusted individually. Given the variability between -omics data,
592 this customization is essential for a meaningful comparative analysis. **(E)** Settings common to all -omics
593 levels (global settings), such as heatmap cell size and orientation, can be applied immediately to all
594 heatmaps using the Layout Panel. **(F)** Using the Sorting and Filtering Panel, genes and samples can be
595 ordered in multiple different ways, while non-significant genes can be filtered out. Appropriate ordering
596 and filtering are essential for revealing structure in the data. **(G)** All the heatmaps are synchronized by
597 samples and genes and can be scrolled simultaneously using the Scrolling Panel. H. Multiple curated lists
598 of genes (feature lists) can be created and maintained within multiSLIDE. These lists can be re-visualized
599 in multiSLIDE or downloaded as text files with a single click.

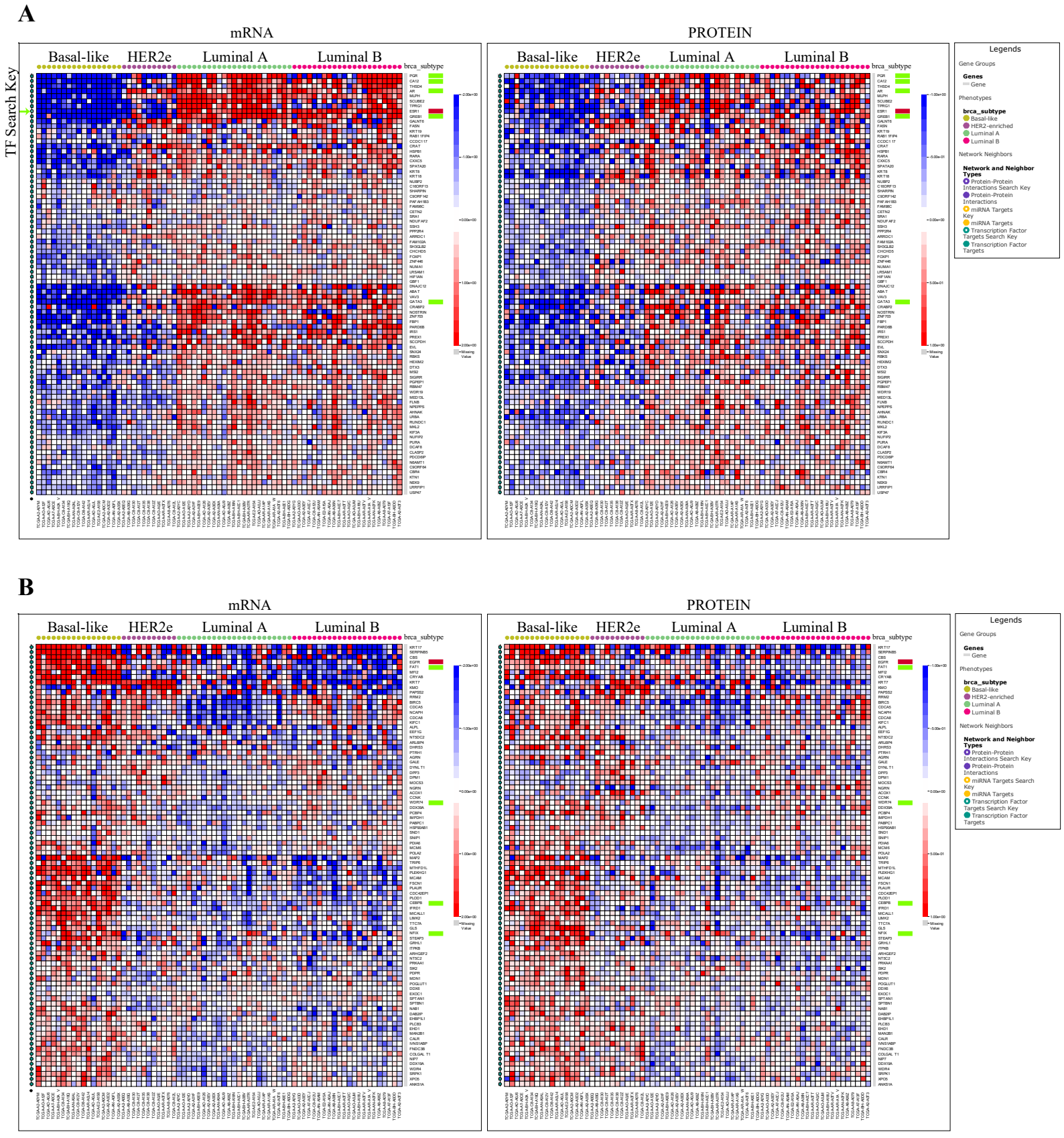
600

601 **Figure 2: Visualization of unfolded protein response in mammalian cells responding to stress**. mRNA
602 and protein-level expression across eight time-points and two replicates, as visualized in multiSLIDE. Gene

603 expression profiles belonging to the GO terms and pathways shown in the "Legends" panel are visualized
604 in the heatmaps. Association between genes and GO terms/pathways are indicated by colored tags in
605 vertical tracks alongside the heatmap. Sample information, such as replicate number and timepoint, is
606 visualized above the heatmaps. Genes highlighted with a green band show early (< 2 h) spikes in response
607 to stress at the mRNA level. Among these, except DNAJC3, all others have a delayed upregulation at the
608 protein level, beginning in the intermediate (2 - 8 h) phase and persisting through the late phase (> 8 h).
609
610 **Figure 3: Visualization of significant transcription regulation targets of ESR1.** The TF targets of ESR1
611 were queried and visualized simultaneously across the mRNA and protein level in multiSLIDE.
612 Significance based filtering (one-way ANOVA, p-value <0.001) in multiSLIDE left 172 targets, which
613 were clustered using Euclidean distance and complete linkage. Two of these clusters are shown in **(A)** and
614 the other two clusters are shown in **(B)**.


A
Datasets*
Genes*
+ ADD GENES
egfr1 [458, 385]
Phenotypes
* indicates required field
APPLY CHANGES
SAVE
VIEW
CREATE
HOME
HELP
CLOSE
Scroll By
Samples 1 to 73 of 73
Genes 11 to 35 of 55
Samples 73
Genes 25
Order Samples By
Phenotypes
Order Genes By
Hierarchical Clusters
Show Only Significant
Genes
Display Genes Along
X-axis
Y-axis
Cell Size
Layout
MRNA
PROTEIN
Legends
Gene Groups
Genes
Gene
Pathways
egfr1
Phenotypes
brca_subtype
Basal-like
HER2-enriched
Luminal A
Luminal B
Network Neighbors
Network and Neighbor Types
Protein-Protein Interactions Search Key
Protein-Protein Interactions
miRNA Targets Search Key
miRNA Targets
Transcription Factor Targets Search Key
Transcription Factor Targets
User Lists
Basal-like-Upregulated (9), Feature List
B
Query Pathways, GO Terms and Genes to visualize
pathname:egfr; genesymbol=pdk1,erbb2; go: egfr
C
Select a gene to explore its regulatory and protein-protein interaction networks
D
Customize individual heatmaps
Settings for mRNA Heatmap
Number of Colors 21 (Valid Values: 0-255)
Binning Range
Use Min and Max of Data (Min=-11.27, Max=9.83)
Use Symmetric Bins (about 0)
Use Range
Start -2 and End 2
Heatmap Color Scheme
Diverging colormaps
BBKY
BLWR
BYR
GBR
SEISMIC
COOLWARM
BRBG
PRGN
PIYG
RDYLGN
RDBU
RDYLBU
SPECTRAL
Perceptually uniform colormaps
VIRIDIS
CIVIDIS
INFERNO
MAGMA
PLASMA
SAVE CHANGES
H
Create and visualize curated (feature) lists of genes
G
Scroll through all heatmaps simultaneously
F
Sort samples and genes, and filter genes
Order Samples By
Phenotypes
Order Genes By
Gene Groups
Hierarchical Clusters
Significance Level
Show Only Significant
Genes
E
Customize layouts of all heatmaps concurrently
Display Genes Along
X-axis
Y-axis
Cell Size
Layout
Sample Label Width 100
Gene Label Width 100


mRNA
R1
R2
replicate
timepoint
SELK
MANF
PDIA4
HSP90B1
DNAJC3
HSPA5
PDIA6
PDIA3
P4HB
C19ORF10
CALR
UGGT1
UFM1
HM13
SRPR
SSR1
SEC61B
YIF1A
STT3A
UBE2D3
ERP44
SEC24D
RRBP1
UBXN4
VCP
PLAA
SEC24B
UBE2K
VAPB
PSMC2
GNB2L1
TRAM1
SEC63
SEC61A1
ERLIN2
SEC13
EIF2A
STT3B
CCDC47
UBXN1
THBS1
DCTN1
HSPB1
UBE4B
HSPE1
DNAJA2
TARDBP
HSPA8
SRPRB
MOGS
PSMC4
PSMC5
DNAJA1
HSP90AA1
DNAJB1
ASNA1
UFC1
CKAP4
TMCO1
SEC24C
SAR1A
LMNA
ITPR1
SEC23A
SCP2
ATP2A2
HSPH1
BAG2
TMX1
HSPBP1
SERPINH1
RAD23B
GANAB
HDGF
PSMC3
PSMC6
PSMC1
NPLOC4
HEBP1
UBQLN2
0.5h
1h
2h
8h
16h
24h
30h
-1.00e+00
-5.00e-01
0.00e+00
5.00e-01
1.00e+00
Missing Value
PROTEIN
R1
R2
replicate
timepoint
-2.00e+00
-1.00e+00
0.00e+00
1.00e+00
2.00e+00
Missing Value
Legends
Gene Groups
Pathways
unfolded protein response (upr)
protein processing in endoplasmic reticulum - homo sapiens (human)
Gene Ontologies
endoplasmic reticulum unfolded protein response
response to unfolded protein
response to endoplasmic reticulum stress
regulation of endoplasmic reticulum unfolded protein response
regulation of response to endoplasmic reticulum stress
Phenotypes
replicate
R1
R2
timepoint
1.00e+00
1.00e+01


A
mRNA
PROTEIN
Basal-like
HER2e
Luminal A
Luminal B
brca_subtype
TF Search Key
Legends
Gene Groups
Genes
Gene
Phenotypes
brca_subtype
Basal-like
HER2-enriched
Luminal A
Luminal B
Network Neighbors
Network and Neighbor Types
Protein-Protein Interactions Search Key
Protein-Protein Interactions
miRNA Targets Key
miRNA Targets
Transcription Factor Targets Search Key
Transcription Factor Targets
B
mRNA
PROTEIN
Basal-like
HER2e
Luminal A
Luminal B
brca_subtype
Legends
Gene Groups
Genes
Gene
Phenotypes
brca_subtype
Basal-like
HER2-enriched
Luminal A
Luminal B
Network Neighbors
Network and Neighbor Types
Protein-Protein Interactions Search Key
Protein-Protein Interactions
miRNA Targets Search Key
miRNA Targets
Transcription Factor Targets Search Key
Transcription Factor Targets