SUPPLEMENTARY DATA:

Analysis of the genomic response of human prostate cancer cells to histone deacetylase inhibitors

Madeleine S.Q. Kortenhorst, Michel D. Wissing, Ronald Rodriguez, Sushant Kachhap, Judith J.M. Jans, Petra Van der Groep, Henk M.W. Verheul, Anuj Gupta, Paul O. Aiyetan, Elsken van der Wall, Michael A. Carducci, Paul J. Van Diest, Luigi Marchionni

May 31, 2013

The present "pdf" document contains cross-referencing hyperlinks to each section listed in the Table of Contents available on the next page.

Complete Supplementary Figures and Tables for the Analysis of Functional Annotation linked to external genomic annotation databases are available at:

- http://luigimarchionni.org/HDACIs.html

Gene expression data compliant with MIAME standards analyzed in this study is available from the NCBI Gene Expression Omnibus (GEO) database[1], and the Connectivity Map database[2,3]:

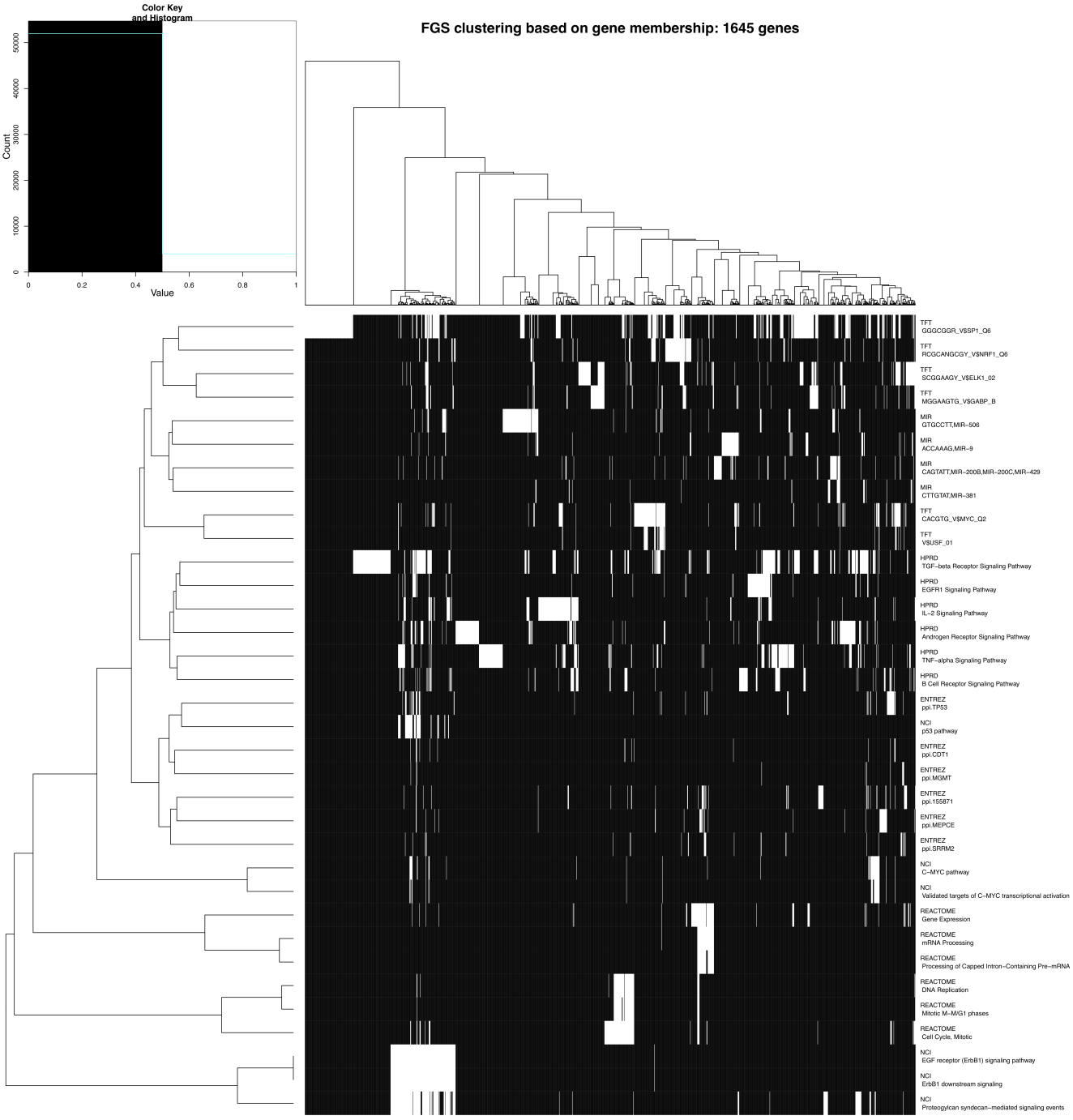
- http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE34452
- http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8645
- http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31620
- http://www.broad.mit.edu/cmap

Contents

1 Supplementary Results

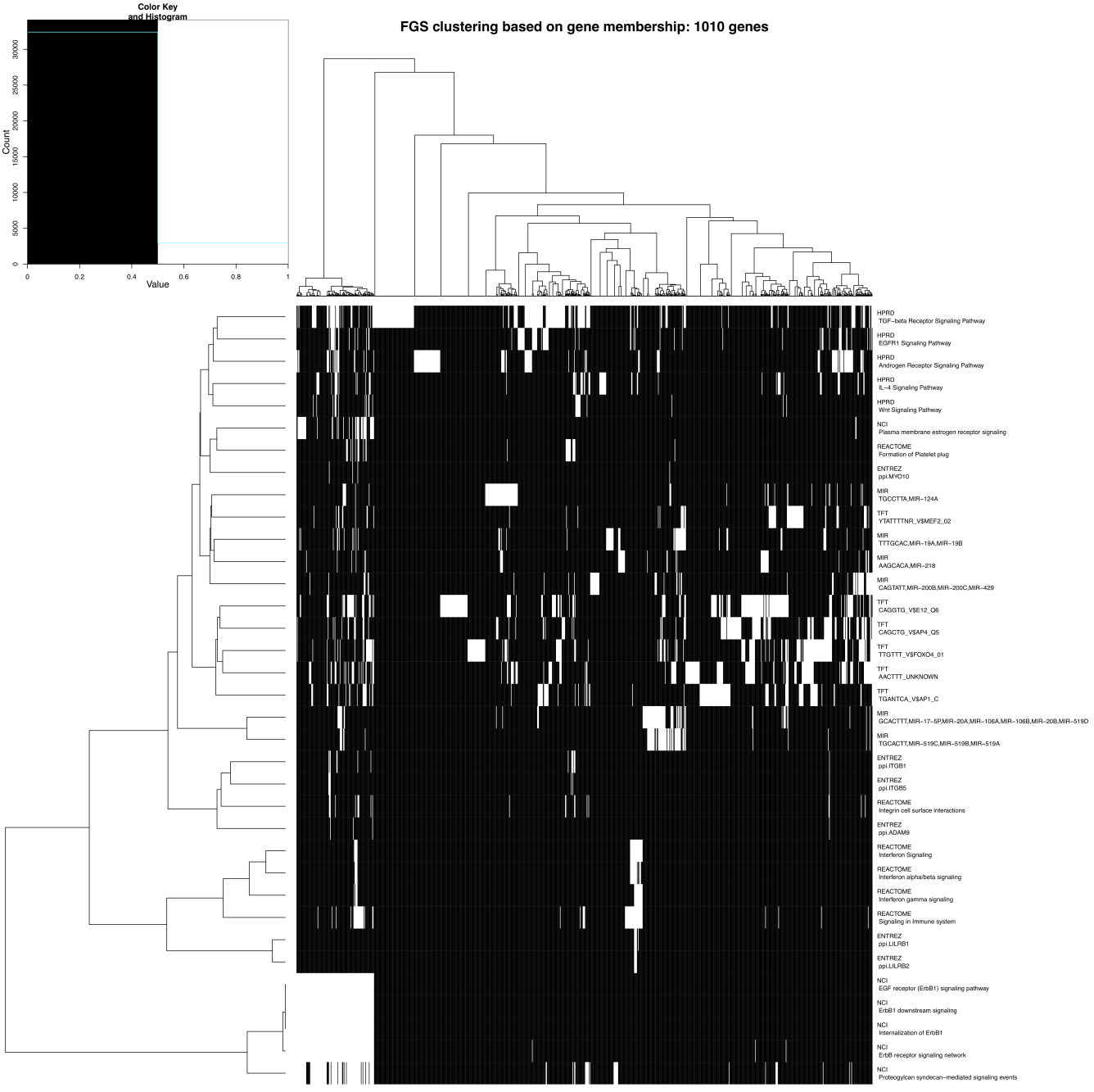
1.1 Supplementary Figure 1

Supplementary Figure S1: Gene membership for differentially expressed FGS after HDACI-treatment. Heatmap visualizing gene membership for the differentially expressed Functional Gene Sets (FGS) enriched in DU-145 and PC3 cells upon HDACI treatment (GSE34452). Rows in the heatmap corresponds to the FGS from Figure 1A in the main paper, while the columns represent the most differentially expressed genes (FDR $< 0.1\%$) annotated to such FGS. In the heatmap when a particular gene is annotated to a specific FGS it is highlighted in white, while black is used when the gene does not belong to the gene list. Hierarchical clustering of rows and columns was obtained using the binary distance and the Ward clustering method. A number of distinct FGS clusters are evident, based on distinct subset of differentially expressed genes in common among the FGS. A high resolution version of this figure can be downloaded here.



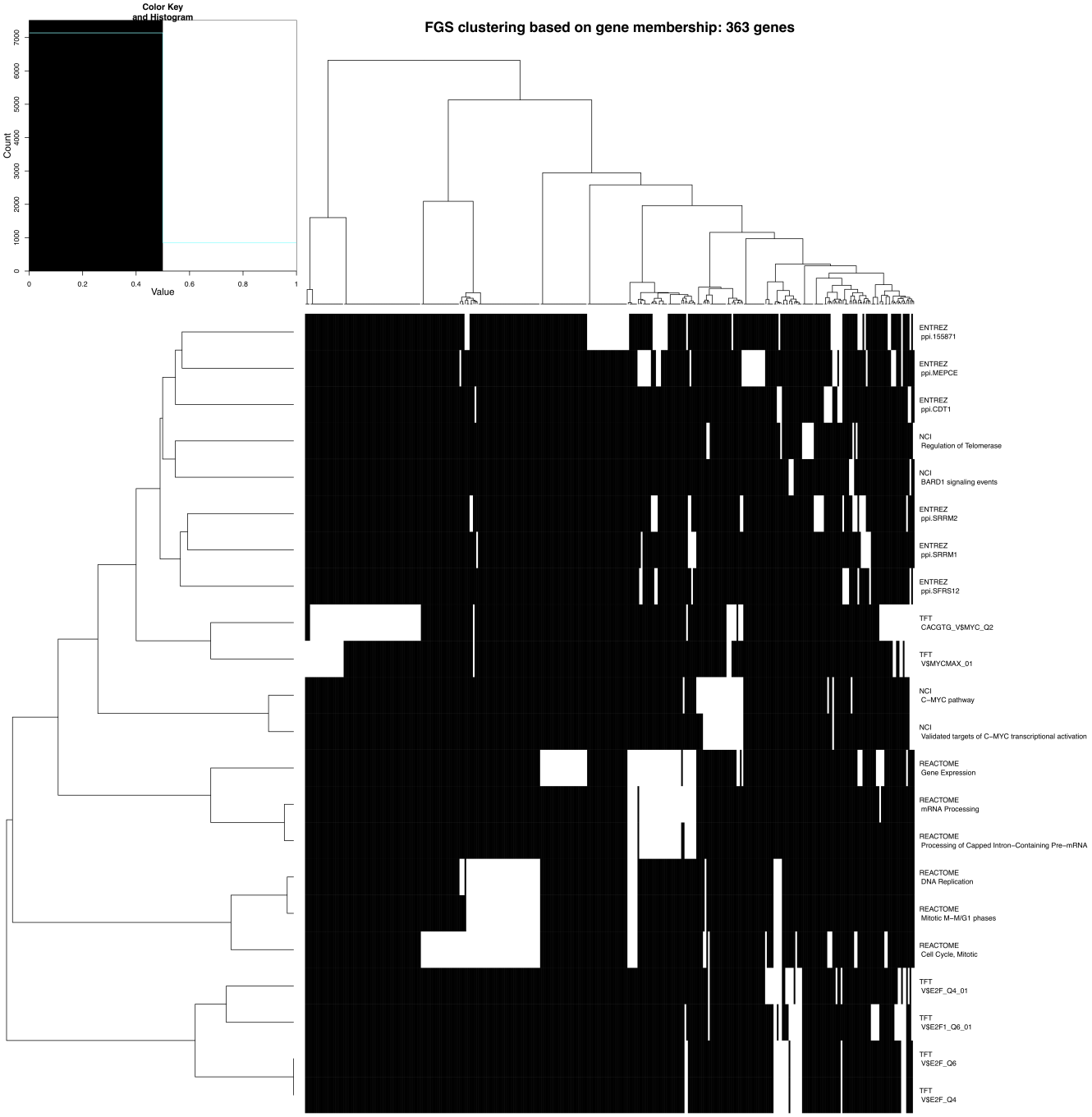
1.2 Supplementary Figure 2

Supplementary Figure S2: Gene membership for up-regulated FGS after HDACI-treatment. Heatmap visualizing gene membership for the up-regulated Functional Gene Sets (FGS) in DU-145 and PC3 cells upon HDACI treatment (GSE34452). Rows in the heatmap corresponds to the FGS from Figure 1B in the main paper, while the columns represent the most up-regulated genes (FDR $< 0.1\%$) annotated to such FGS. In the heatmap when a particular gene is annotated to a specific FGS it is highlighted in white, while black is used when the gene does not belong to the gene list. Hierarchical clustering of rows and columns was obtained using the binary distance and the Ward clustering method. A number of distinct FGS clusters are evident, based on distinct subset of up-regulated genes in common among the FGS. A high resolution version of this figure can be downloaded here.



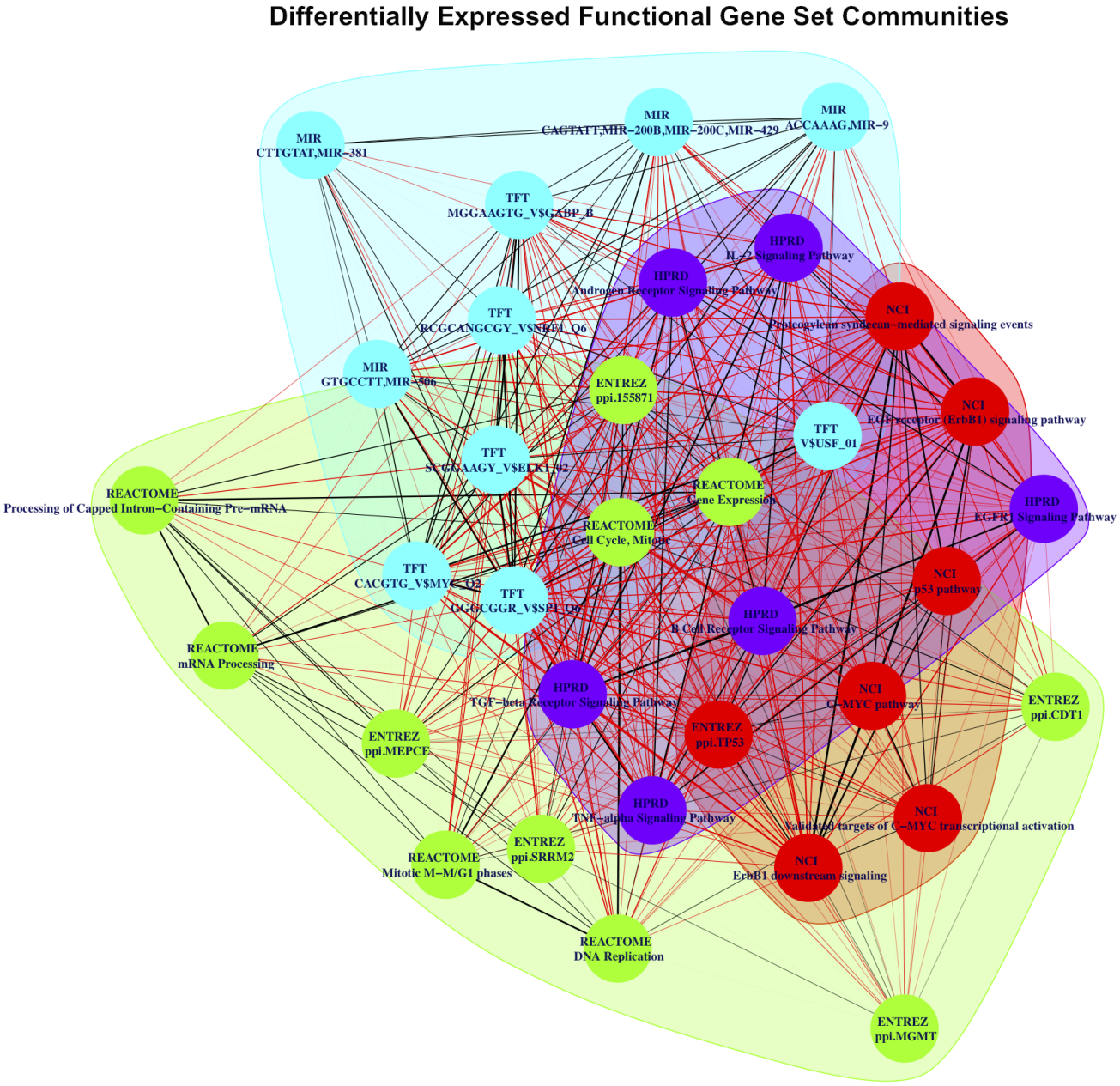
1.3 Supplementary Figure 3

Supplementary Figure S3: Gene membership for down-regulated FGS after HDACI-treatment. Heatmap visualizing gene membership for the down-regulated Functional Gene Sets (FGS) in DU-145 and PC3 cells upon HDACI treatment (GSE34452). Rows in the heatmap corresponds to the FGS from Figure 1C in the main paper, while the columns represent the most up-regulated genes (FDR $< 0.1\%$) annotated to such FGS. In the heatmap when a particular gene is annotated to a specific FGS it is highlighted in white, while black is used when the gene does not belong to the gene list. Hierarchical clustering of rows and columns was obtained using the binary distance and the Ward clustering method. A number of distinct FGS clusters are evident, based on distinct subset of down-regulated genes in common among the FGS. A high resolution version of this figure can be downloaded here.



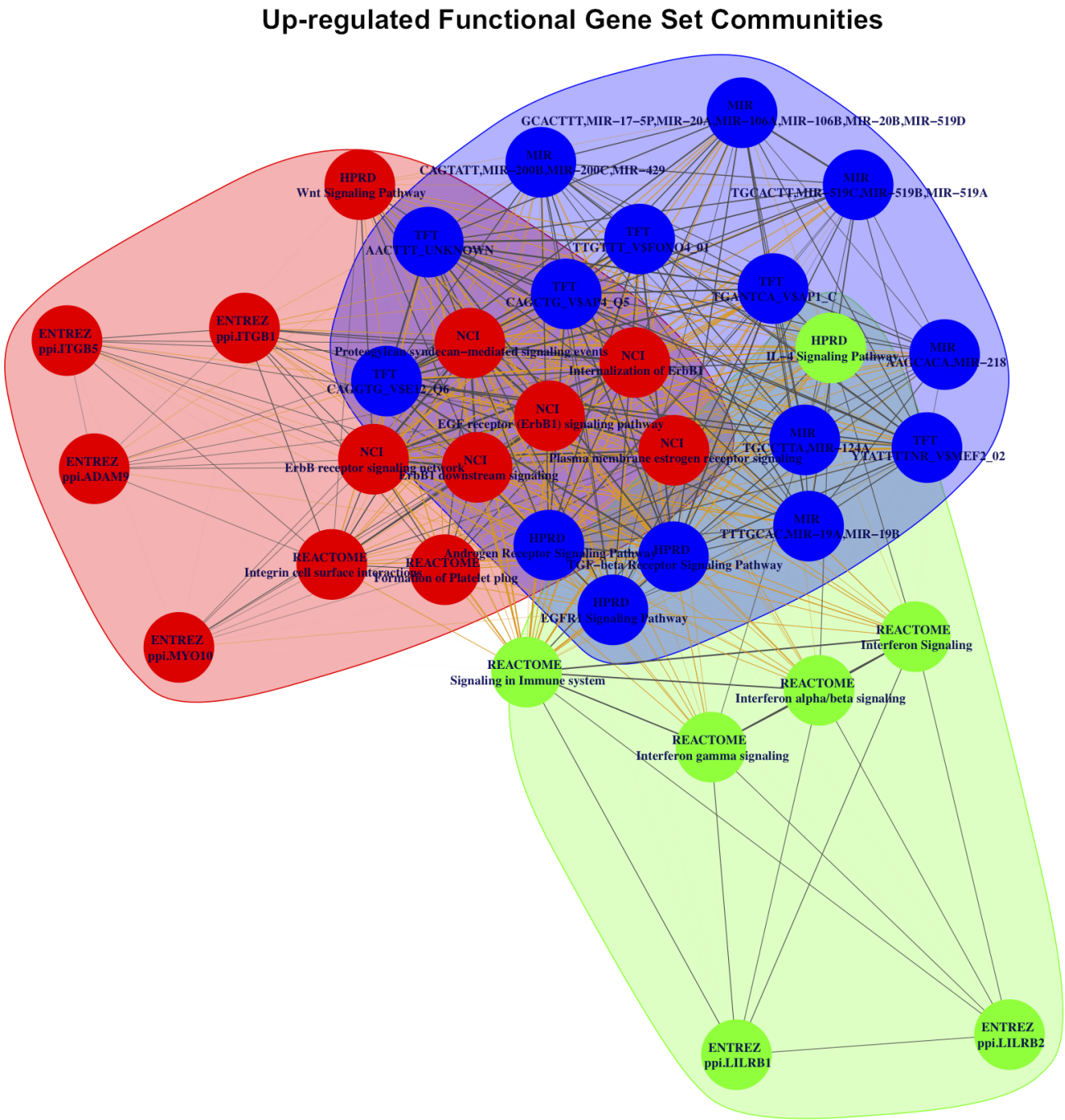
1.4 Supplementary Figure 4

Supplementary Figure S4: Social network analysis of FGS differentially expressed after HDACI-treatment. The figure depicts the weighted undirected network based on the differentially expressed genes in common among the enriched FGS from Figure 1A in the main paper (see also Supplementary Figure **S11**). In the network vertexes represent the specific FGS, while the edges (and their weights) are based on the number of differentially expressed genes in common among the FGS. A number of distinct FGS "communities" (i.e. subgraphs of FGS sharing common subset of genes) were identified using the fast greedy modularity optimization algorithm described by Clauset and colleagues [4], and are shown in the figure with distinct colors. A high resolution version of this figure can be downloaded here.



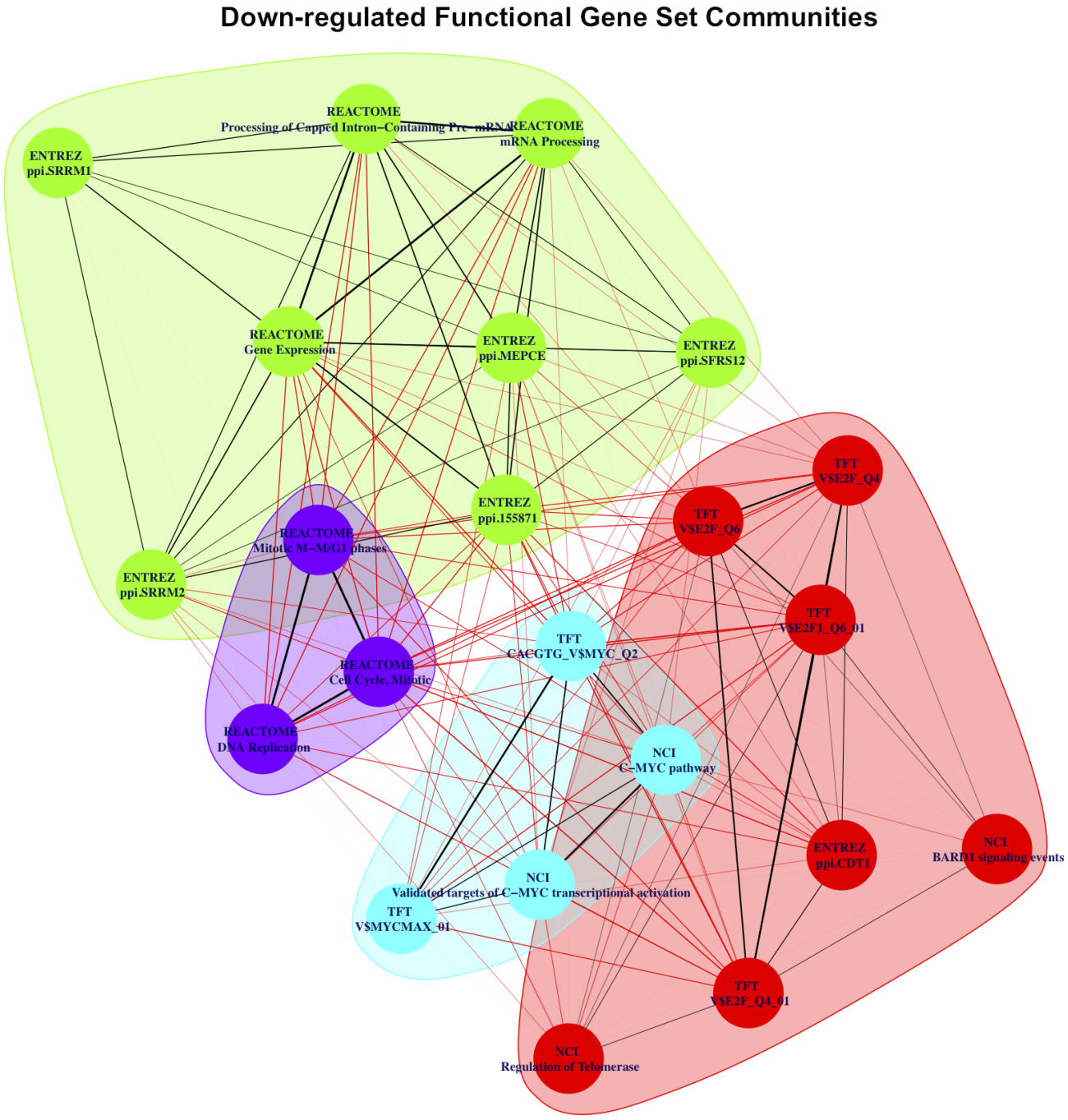
1.5 Supplementary Figure 5

Supplementary Figure S5: Social network analysis of FGS up-regulated upon HDACI-treatment. The figure depicts the weighted undirected network based on the up-regulated genes in common among the up-regulated FGS from Figure 1B in the main paper (see also Supplementary Figure **S12**). In the network vertexes represent the specific FGS, while the edges (and their weights) are based on the number of up-regulated genes in common among the FGS. A number of distinct FGS "communities" (i.e. subgraphs of FGS sharing common subset of genes) were identified using the fast greedy modularity optimization algorithm described by Clauset and colleagues[4], and are shown in the figure with distinct colors. A high resolution version of this figure can be downloaded here.



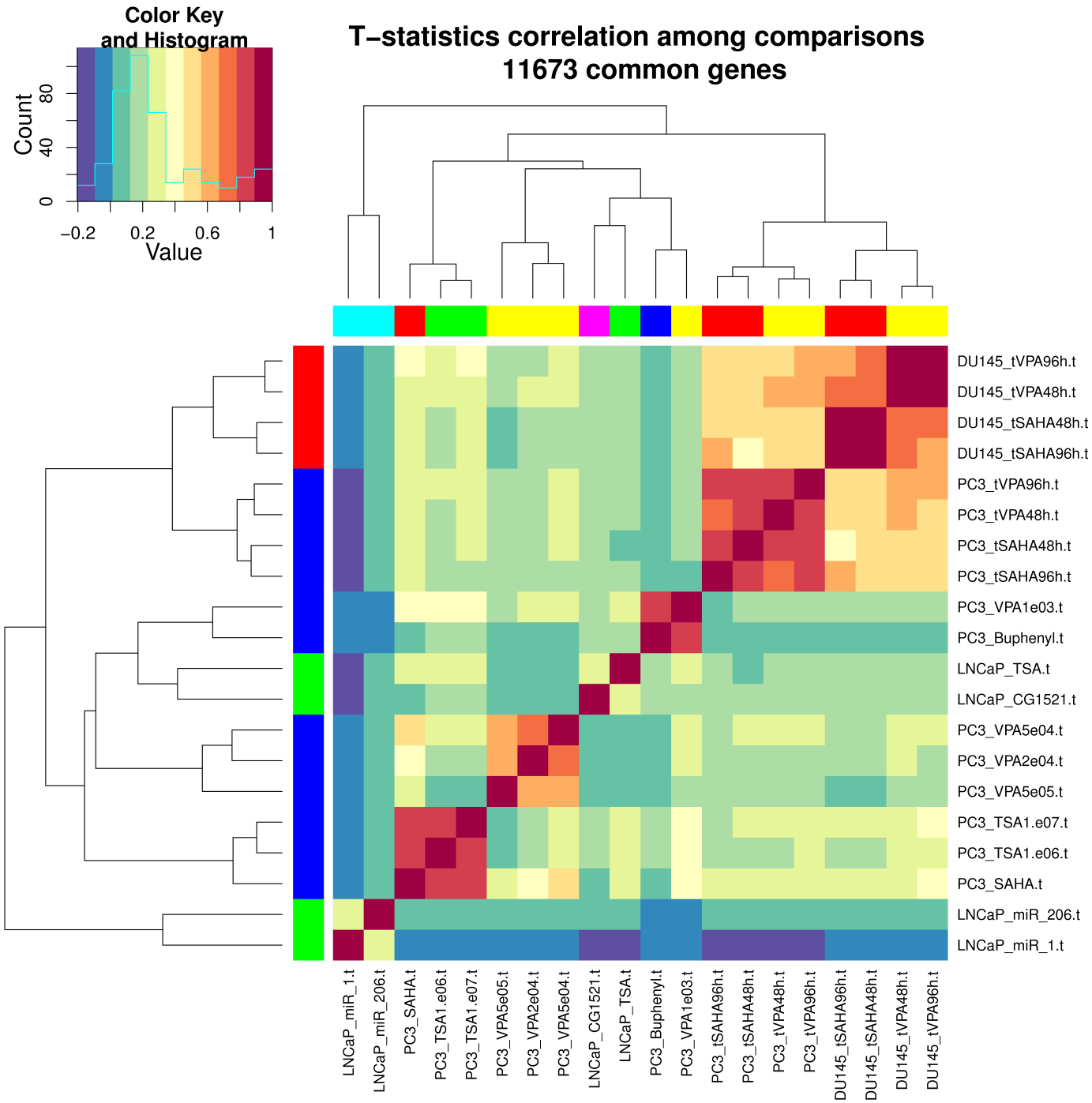
1.6 Supplementary Figure 6

Supplementary Figure S6: Social network analysis of FGS down-regulated upon HDACI-treatment. The figure depicts the weighted undirected network based on the down-regulated genes in common among the down-regulated FGS from Figure 1C in the main paper (see also Supplementary Figure **S13**). In the network vertexes represent the specific FGS, while the edges (and their weights) are based on the number of down-regulated genes in common among the FGS. A number of distinct FGS "communities" (i.e. subgraphs of FGS sharing common subset of genes) were identified using the fast greedy modularity optimization algorithm described by Clauset and colleagues [4], and are shown in the figure with distinct colors. A high resolution version of this figure can be downloaded here.



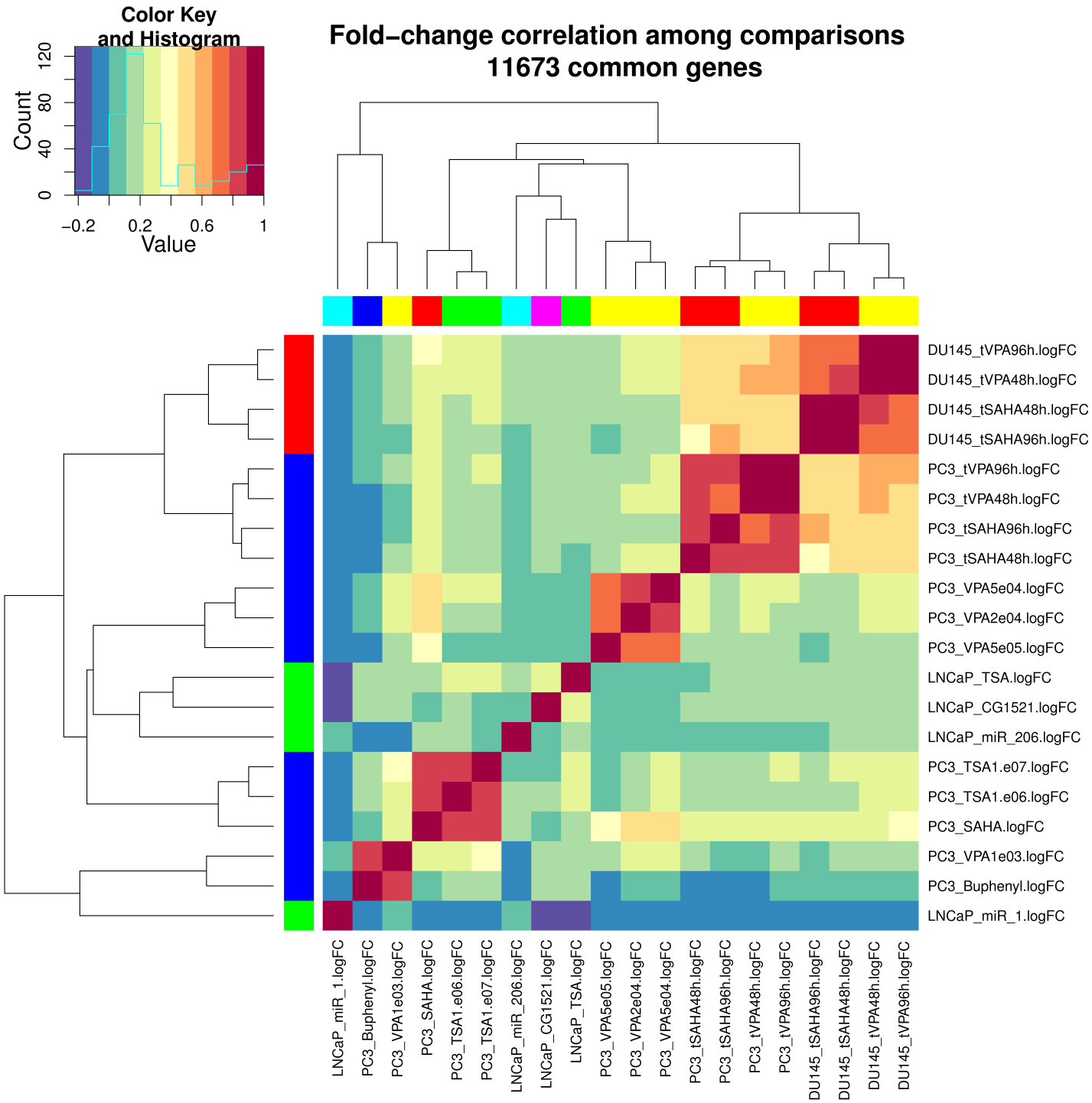
1.7 Supplementary Figure 7

Supplementary Figure S7: Heatmap showing the correlation across distinct gene expression experiments comparing HDACI treatment versus control. This dataset accounts for 375 microarray experiments, it was obtained from four distinct datasets (GSE34452, GSE8645, GSE31620, and Connectivity Map), and encompasses different cell lines, inhibitors, dosages, and time points. The color code used for the columns highlights the different inhibitors that were used: VPA in yellow, SAHA in red, Buphenyl in blue, TSA in green, CG1521 in purple, and miR transfection in cyan. The color code used for the rows highlights the different cell lines that were analyzed: DU-145 in red, PC3 in blue, and LNCaP in green. Correlations were computed between moderated t-statistics expressing the degree of differential gene expression between treatment and control. As expected the highest degree of correlations are observed within study. A high resolution version of this figure can be downloaded here.



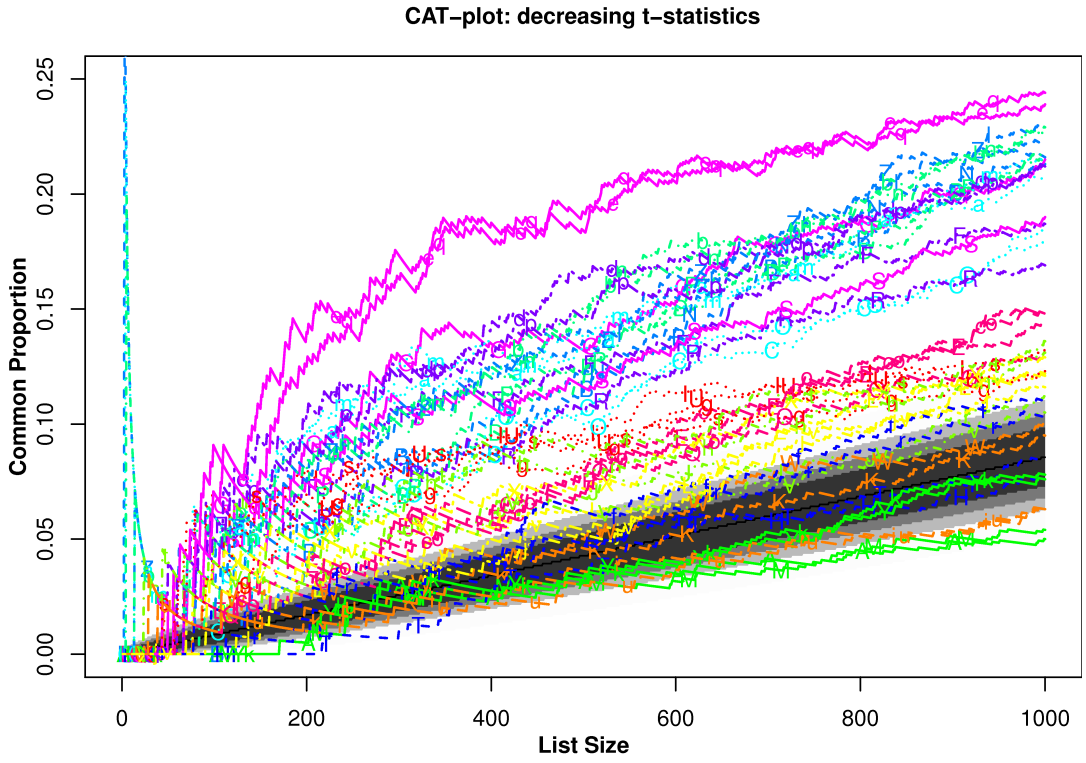
1.8 Supplementary Figure 8

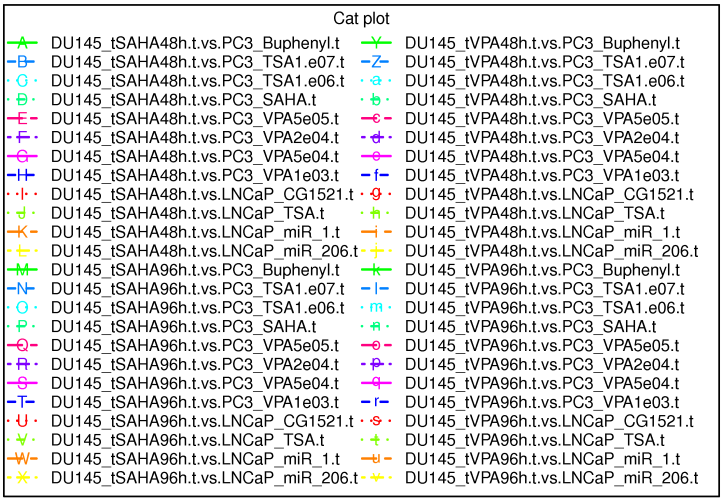
Supplementary Figure S8: Heatmap showing the correlation across distinct gene expression experiments comparing HDACI treatment versus control. This dataset accounts for 375 microarray experiments, it was obtained from four distinct datasets (GSE34452, GSE8645, GSE31620, and Connectivity Map), and encompasses different cell lines, inhibitors, dosages, and time points. The color code used for the columns highlights the different inhibitors that were used: VPA in yellow, SAHA in red, Buphenyl in blue, TSA in green, CG1521 in purple, and miR transfection in cyan. The color code used for the rows highlights the different cell lines that were analyzed: DU-145 in red, PC3 in blue, and LNCaP in green. Correlations were computed between log2 fold-change expressing the degree of differential gene expression between treatment and control. As expected the highest degree of correlations are observed within study. A high resolution version of this figure can be downloaded here.



1.9 Supplementary Figure 9

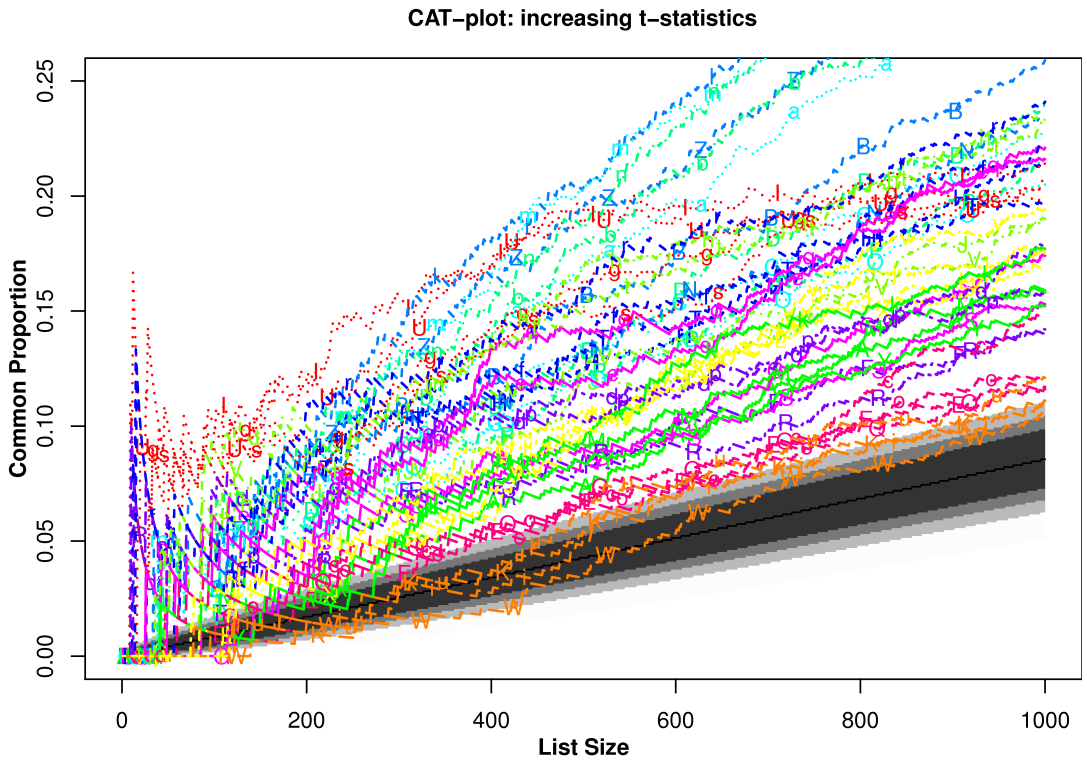
Supplementary Figure S9: Correspondence at the top curves for the 1000 most up-regulated genes upon HDACi inhibition in the DU145 cell line. Genes were ranked in decreasing order by the moderated t-statistics obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion of differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the DU145 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

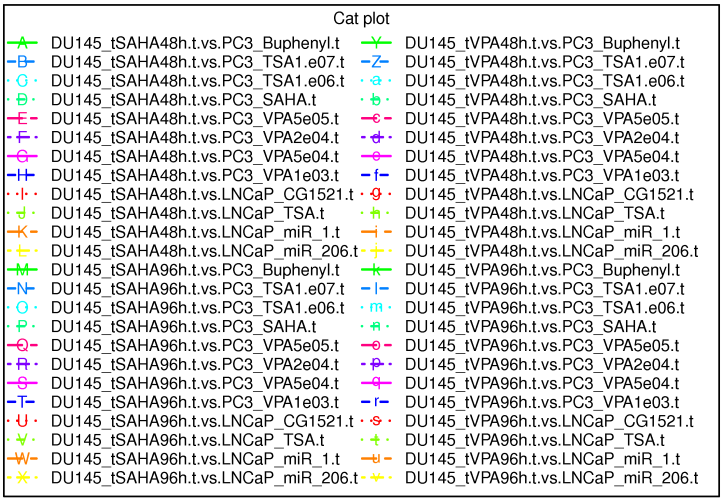




1.10 Supplementary Figure 10

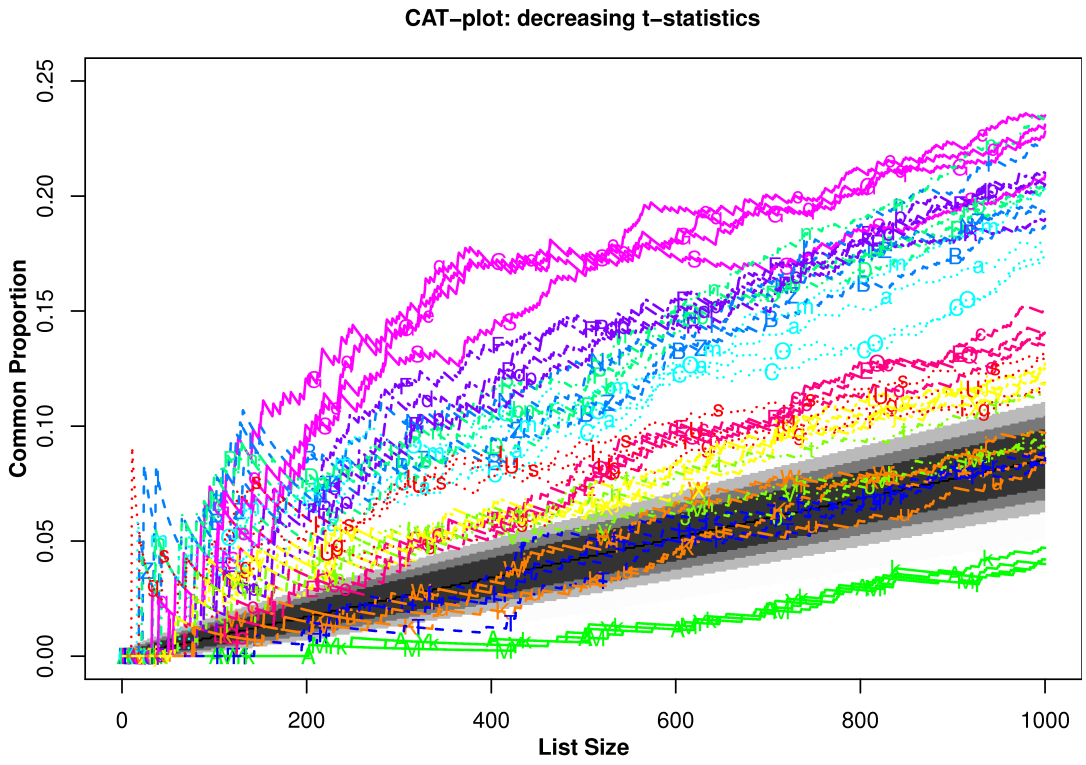
Supplementary Figure S10: Correspondence at the top curves for the 1000 most downregulated-regulated genes upon HDACi inhibition in the DU145 cell line. Genes were ranked in decreasing order by the moderated t-statistics obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the DU145 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

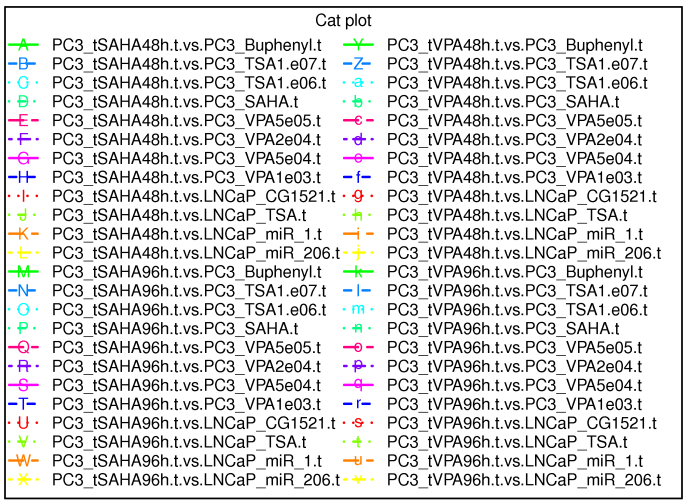




1.11 Supplementary Figure 11

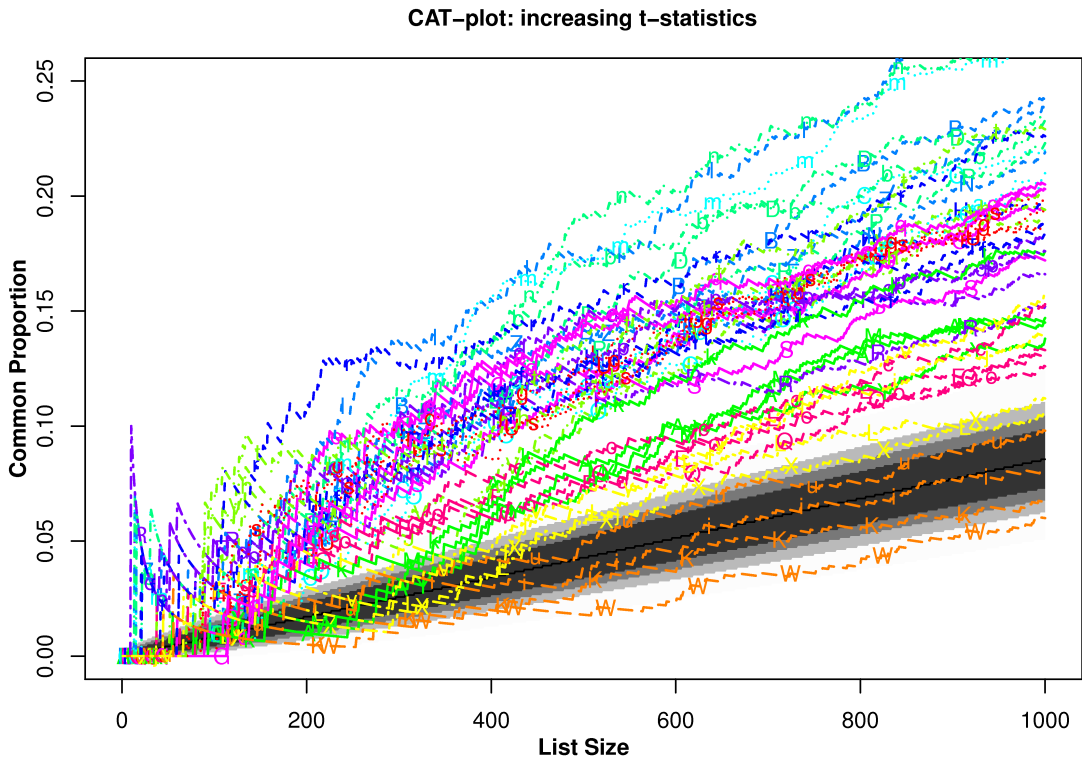
Supplementary Figure S11: Correspondence at the top curves for the 1000 most up-regulated genes upon HDACi inhibition in the PC3 cell line. Genes were ranked in decreasing order by the moderated t-statistics obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion of differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the PC3 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

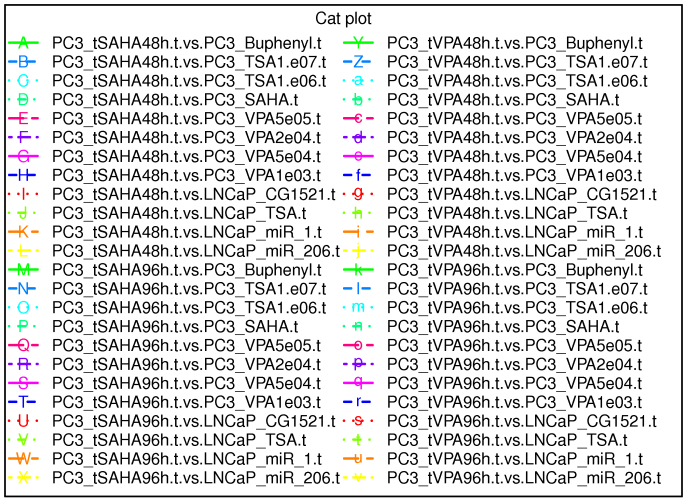




1.12 Supplementary Figure 12

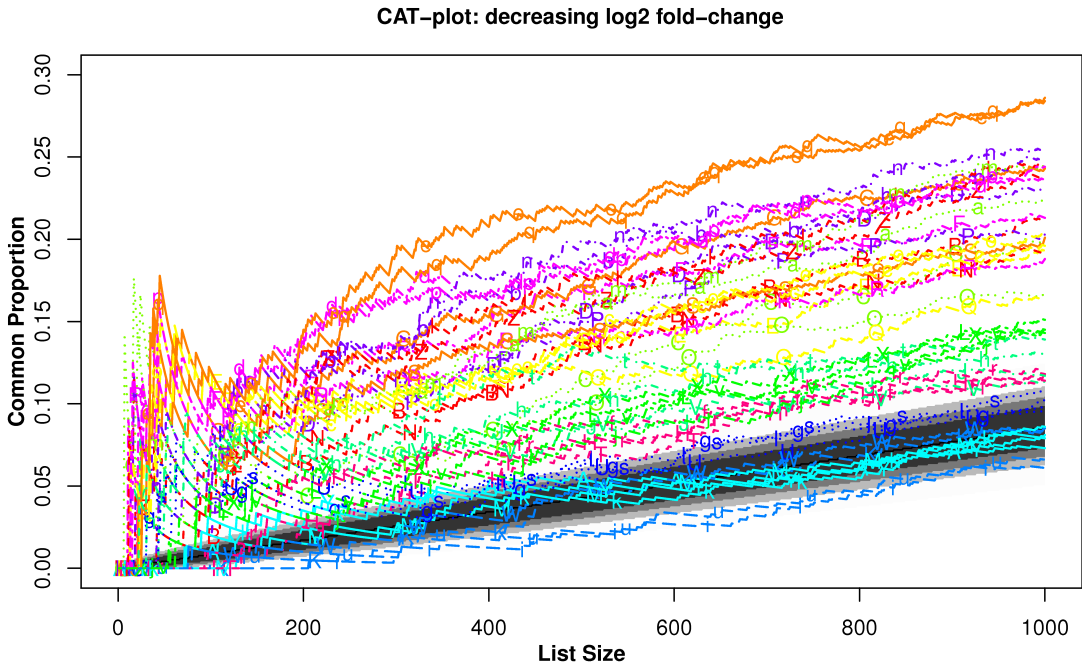
Supplementary Figure S12: Correspondence at the top curves for the 1000 most downregulated-regulated genes upon HDACi inhibition in the PC3 cell line. Genes were ranked in decreasing order by the moderated t-statistics obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the PC3 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

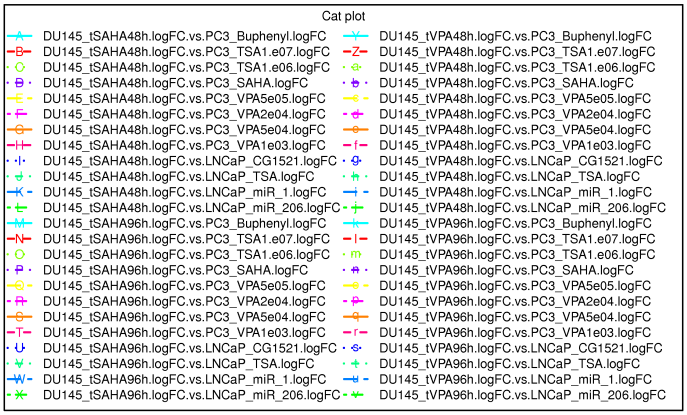




1.13 Supplementary Figure 13

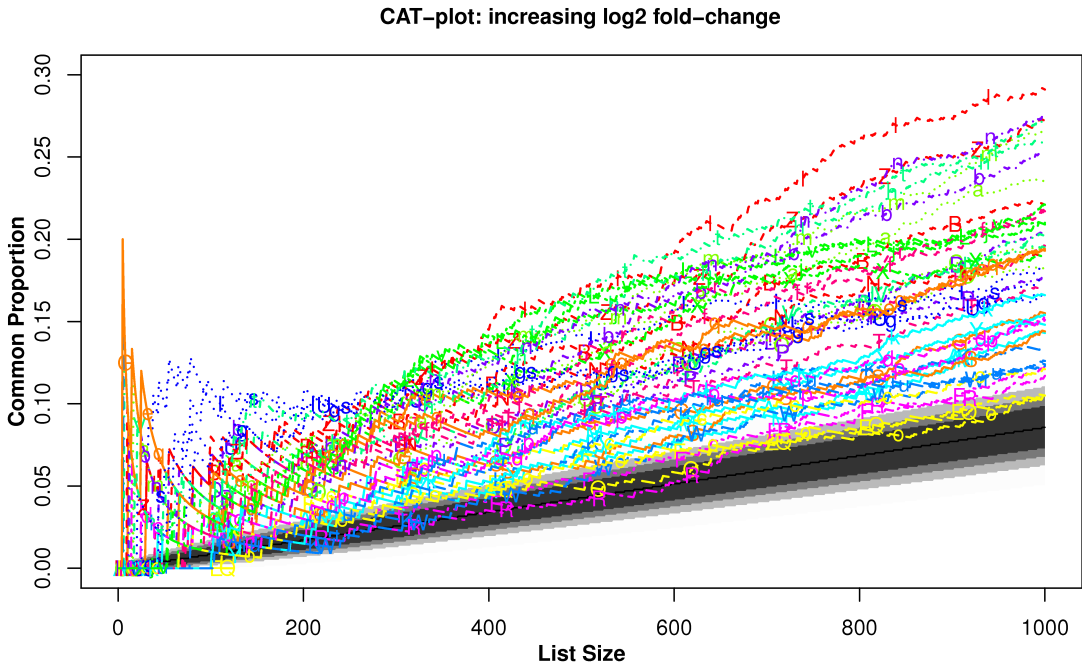
Supplementary Figure S13: Correspondence at the top curves for the 1000 most up-regulated genes upon HDACi inhibition in the DU145 cell line. Genes were ranked in decreasing order by the log2 fold-change obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion of differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the DU145 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

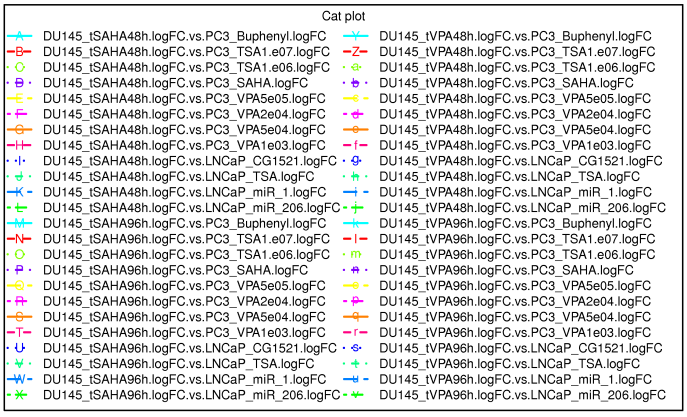




1.14 Supplementary Figure 14

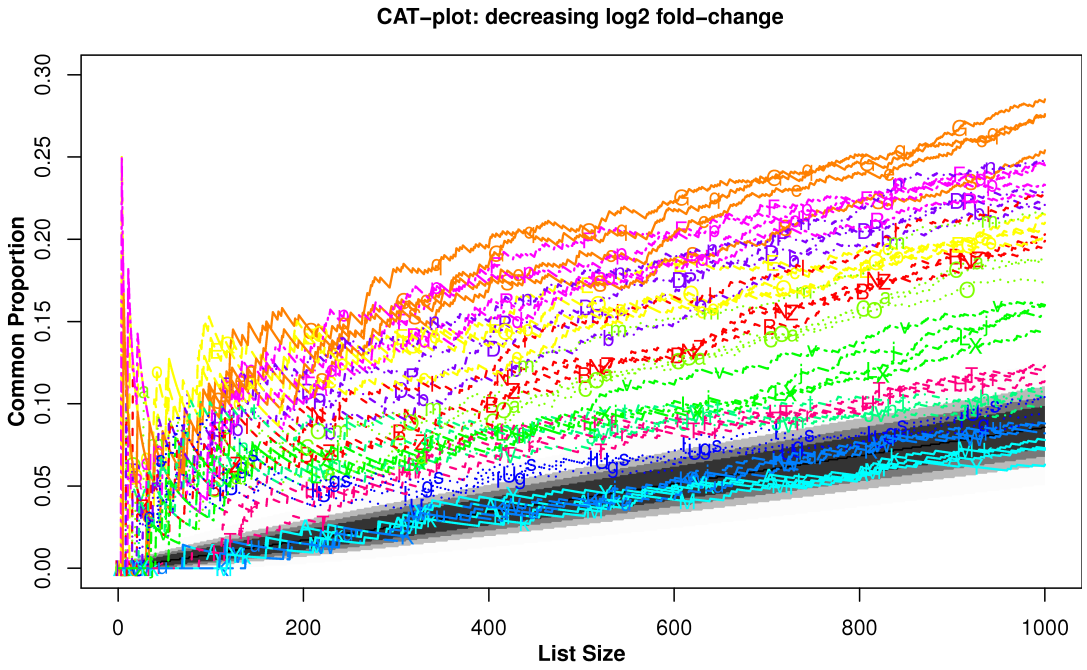
Supplementary Figure S14: Correspondence at the top curves for the 1000 most downregulated-regulated genes upon HDACi inhibition in the DU145 cell line. Genes were ranked in decreasing order by the log2 fold-change obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the DU145 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

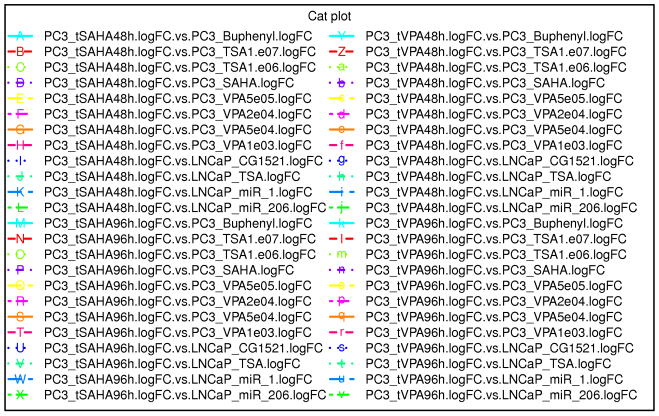




1.15 Supplementary Figure 15

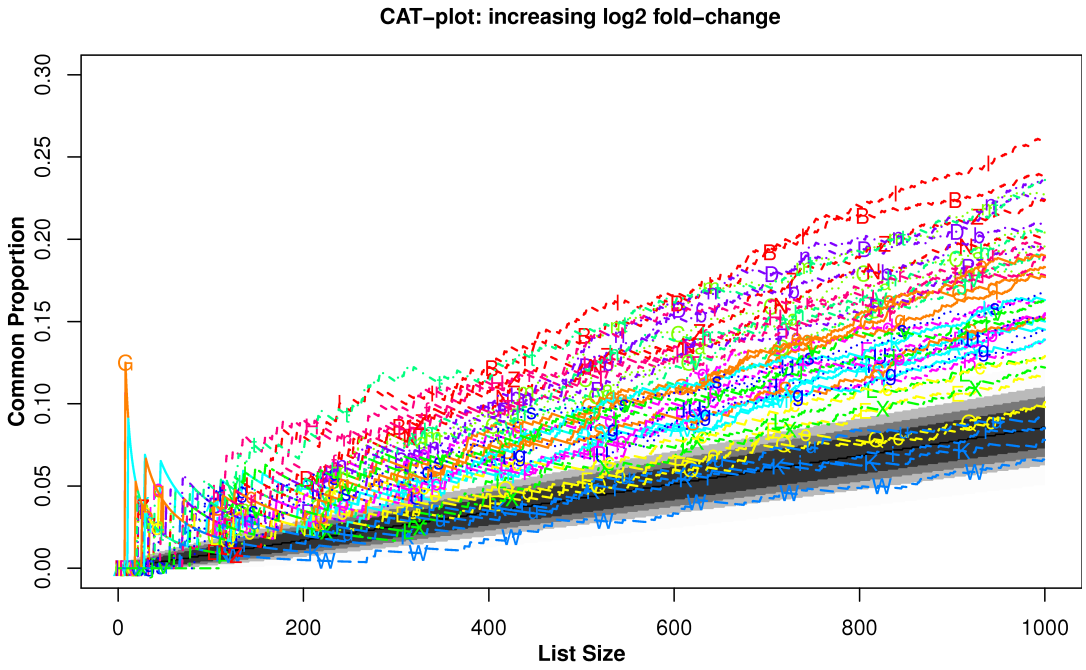
Supplementary Figure S15: Correspondence at the top curves for the 1000 most up-regulated genes upon HDACi inhibition in the PC3 cell line. Genes were ranked in decreasing order by the log2 fold-change obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion of differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the PC3 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.

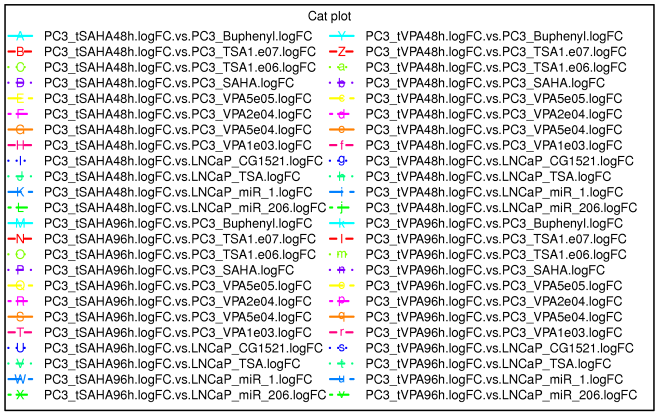




1.16 Supplementary Figure 16

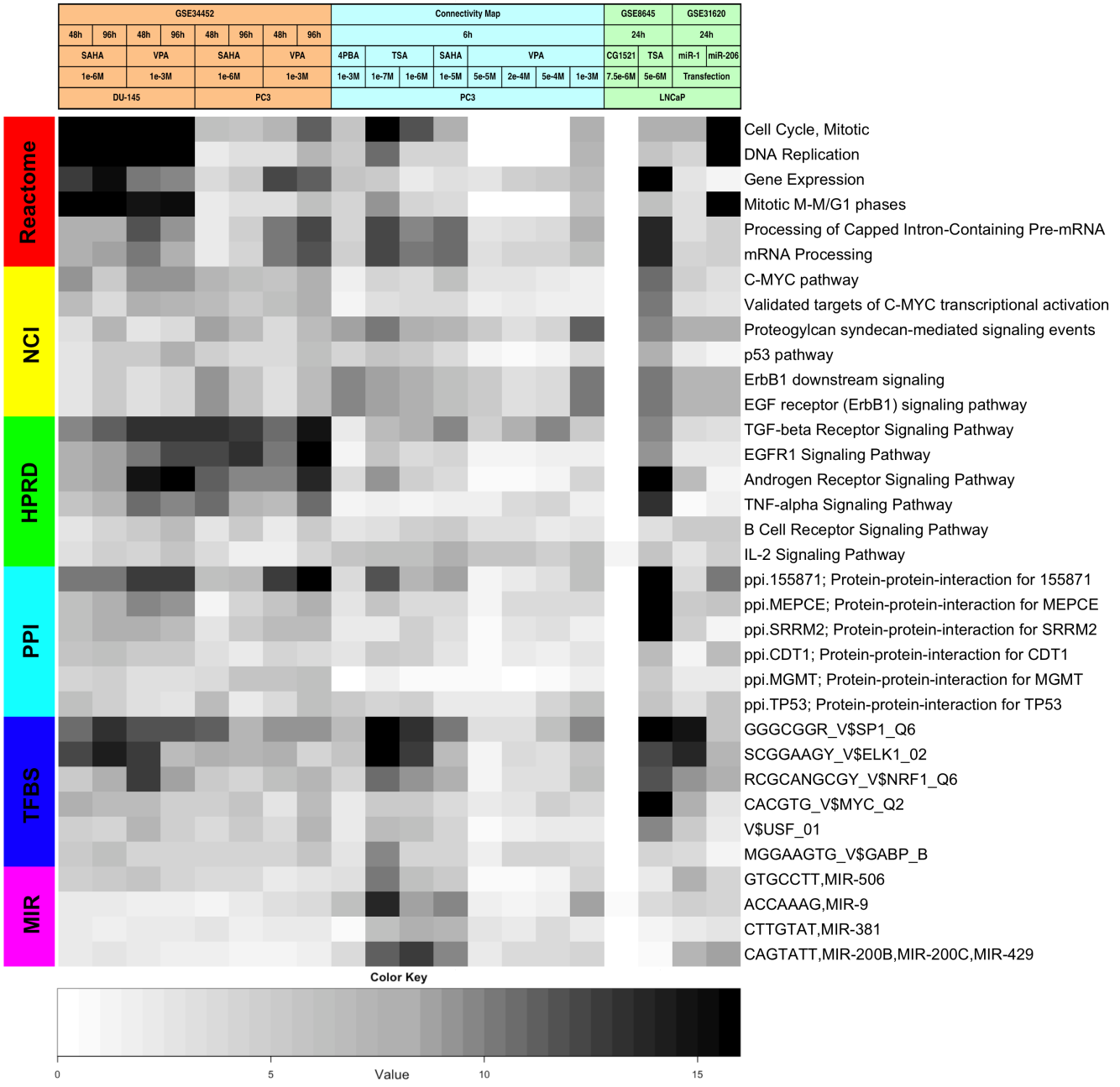
Supplementary Figure S16: Correspondence at the top curves for the 1000 most downregulated-regulated genes upon HDACi inhibition in the PC3 cell line. Genes were ranked in decreasing order by the log2 fold-change obtained from the comparison between treated and untreated cells. Each CAT curve represents the proportion differentially expressed genes that are in common between two comparisons. All time points and HDACI treatments in the PC3 cell line from our study (GSE34452) were compared to those obtained from the remainder studies (GSE8645, GSE31620, and Connectivity Map). CAT curves in the white area above the gray shading indicate agreement, while the curves below indicate disagreement between experiments. The grey shading represents the 99% probability intervals of agreement by chance, therefore CAT curves in the white represent agreement beyond what it would be expected by chance alone. Overall the we observed good agreement across studies, apart from comparisons involving miR and Buphenyl. A high resolution version of this figure can be downloaded here.





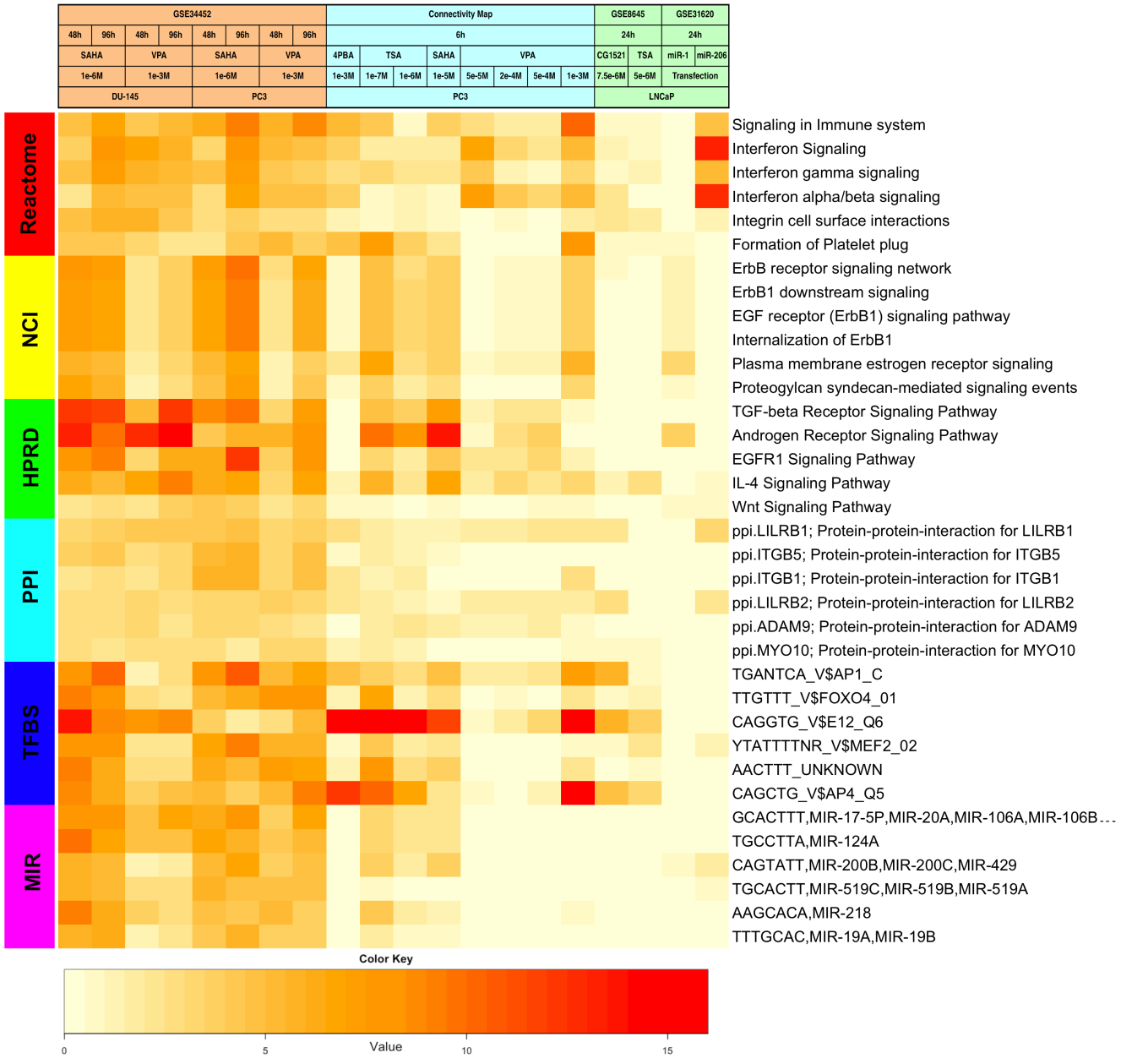
1.17 Supplementary Figure 17

Supplementary Figure S17: AFA results displaying enriched FGS after HDACI-treatment. Heatmap visualizing Functional Gene Sets (FGS) enriched in DU-145 and PC3 cells upon HDACI treatment (GSE34452), along with the level of enrichment for similar experiments in other studies (GSE8645, GSE31620, and Connectivity Map). This heatmap corresponds to Figure 1A in the main paper, in which enrichment was assessed after ordering the genes based on absolute moderated t-statistics, thus irrespective of the direction of gene expression modulation upon HDAC-inhibition. Each row represents a distinct FGS, while each column represents a distinct coefficient from our previous linear model analysis. The most enriched FGS across all the comparisons performed are shown in the figure (top 5 FGS showing an adjusted p-value $\leq 5\%$, or more in case of ties). Color scales representing the enrichment correspond to the absolute adjusted p-values obtained from our analysis after base 10 logarithmic transformations (i.e. the number on the color scale increases with decreasing adjusted p-values). Enriched FGS were selected from different collections in order to encompass distinct biological concepts, as shown by the color bar on the left of each heat map. Cell signaling FGS are highlighted in red and yellow (Pathway Commons Reactome and NCI pathways, respectively), signaling pathway target gene sets in green (Human Protein Reference Database, HPRD), protein-protein-interaction networks in cyan (PPI, as compiled in the NCBI Entrez Gene database), FGS for shared transcriptional factor binding sites (TFBS) in blue, and microRNA (MIR) targets gene sets in pink (both from the Broad Institute Molecular Signature Database collections). A high resolution version of this figure can be downloaded here.



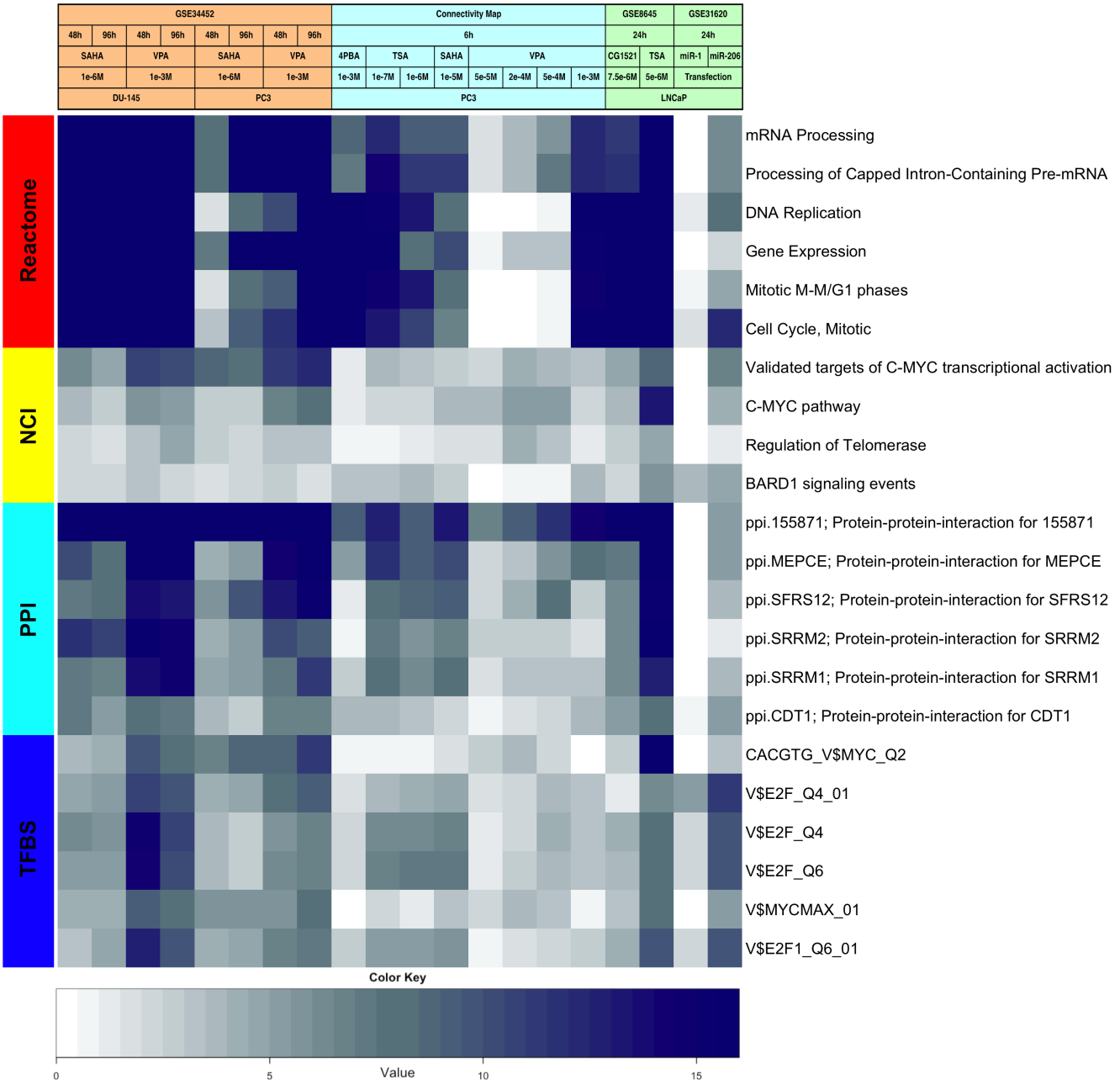
1.18 Supplementary Figure 18

Supplementary Figure S18: AFA results displaying down-regulated FGS after HDACI-treatment. Heatmap visualizing Functional Gene Sets (FGS) enriched in DU-145 and PC3 cells upon HDACI treatment (GSE34452), along with the level of enrichment for similar experiments in other studies (GSE8645, GSE31620, and Connectivity Map). This heatmap corresponds to Figure 1C in the main paper, in which the enrichment was assessed after increasing ordering the genes based by signed t-statistics. Each row represents a distinct FGS, while each column represents a distinct coefficient from our previous linear model analysis. The most enriched FGS across all the comparisons performed are shown in the figure (top 5 FGS showing an adjusted p-value $\leq 5\%$, or more in case of ties). Color scales representing the enrichment correspond to the absolute adjusted p-values obtained from our analysis after base 10 logarithmic transformations (i.e. the number on the color scale increases with decreasing adjusted p-values). Enriched FGS were selected from different collections in order to encompass distinct biological concepts, as shown by the color bar on the left of each heat map. Cell signaling FGS are highlighted in red and yellow (Pathway Commons Reactome and NCI pathways, respectively), signaling pathway target gene sets in green (Human Protein Reference Database, HPRD), protein-protein-interaction networks in cyan (PPI, as compiled in the NCBI Entrez Gene database), FGS for shared transcriptional factor binding sites (TFBS) in blue, and microRNA (MIR) targets gene sets in pink (both from the Broad Institute Molecular Signature Database collections). A high resolution version of this figure can be downloaded here.



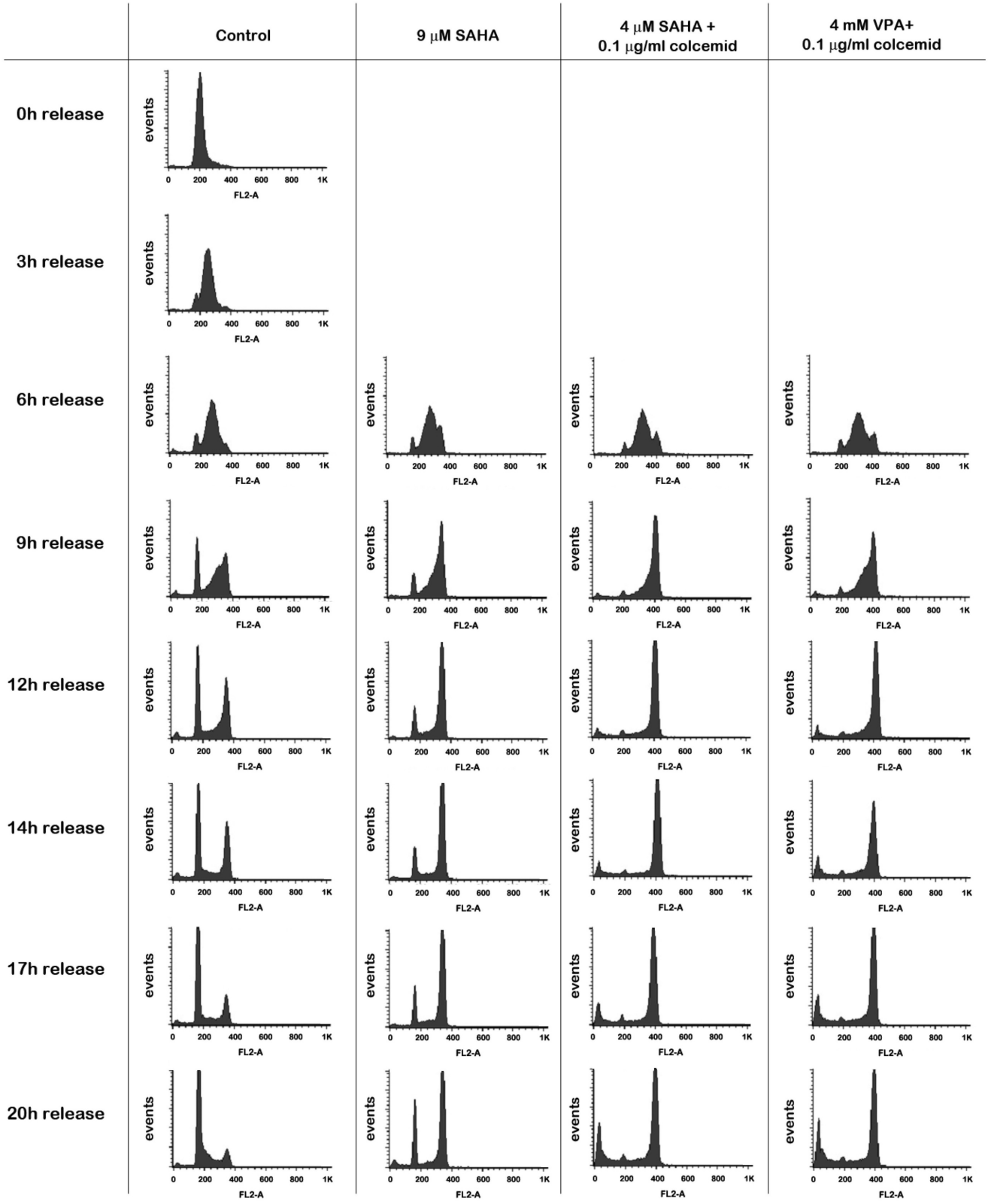
1.19 Supplementary Figure 19

Supplementary Figure S19: AFA results displaying up-regulated FGS after HDACI-treatment. Heatmap visualizing Functional Gene Sets (FGS) enriched in DU-145 and PC3 cells upon HDACI treatment (GSE34452), along with the level of enrichment for similar experiments in other studies (GSE8645, GSE31620, and Connectivity Map). This heatmap corresponds to Figure 1B in the main paper, in which enrichment was assessed after decreasing ordering the genes based by signed t-statistics. Each row represents a distinct FGS, while each column represents a distinct coefficient from our previous linear model analysis. The most enriched FGS across all the comparisons performed are shown in the figure (top 5 FGS showing an adjusted p-value $\leq 5\%$, or more in case of ties). Color scales representing the enrichment correspond to the absolute adjusted p-values obtained from our analysis after base 10 logarithmic transformations (i.e. the number on the color scale increases with decreasing adjusted p-values). Enriched FGS were selected from different collections in order to encompass distinct biological concepts, as shown by the color bar on the left of each heat map. Cell signaling FGS are highlighted in red and yellow (Pathway Commons Reactome and NCI pathways, respectively), signaling pathway target gene sets in green (Human Protein Reference Database, HPRD), protein-protein-interaction networks in cyan (PPI, as compiled in the NCBI Entrez Gene database), FGS for shared transcriptional factor binding sites (TFBS) in blue, and microRNA (MIR) targets gene sets in pink (both from the Broad Institute Molecular Signature Database collections). A high resolution version of this figure can be downloaded here.



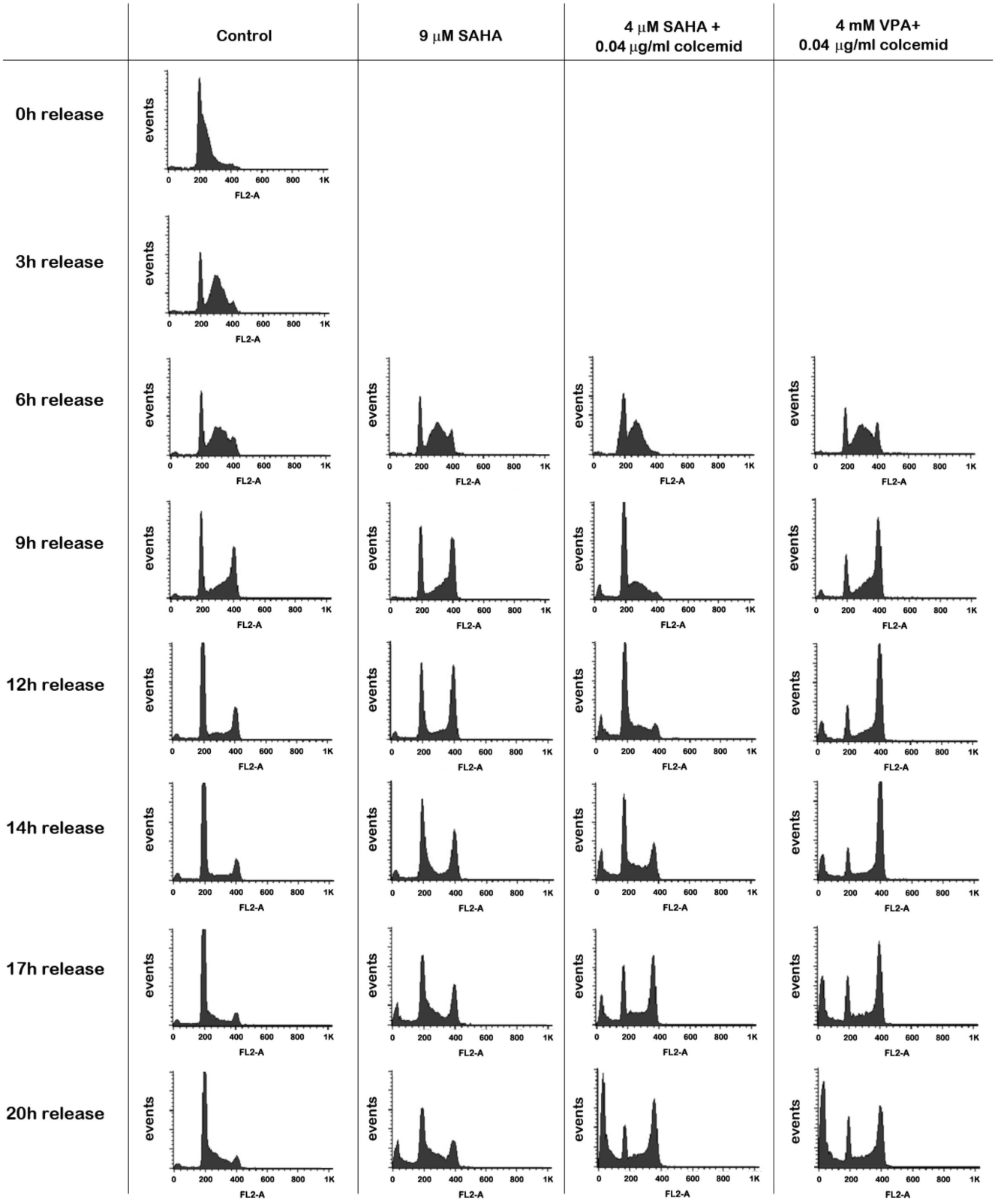
1.20 Supplementary Figure 20

Supplementary Figure S20: Flow cytometry graphs after treatment of PC3 cells with HDACIs, alone or with combinations of Colcemid with HDACIs, containing all time points measured. PC3 cells were treated for variable periods with either 9 μM SAHA alone, or a combination of 0.1 μg/ml Colcemid with 4 μM SAHA or 4 mM VPA. Every 2-3 hours cells were harvested and stained with propidium iodide and flow cytometry analyses were performed. Combining HDAC-inhibitors with Colcemid resulted in mitotic accumulation of PC3 cells. A high resolution version of this figure can be downloaded here.



1.21 Supplementary Figure 21

Supplementary Figure S21: Flow cytometry graphs after treatment of DU-145 cells with HDACIs alone or with combinations of Colcemid with HDACIs, containing all time points measured. DU-145 cells were treated for variable periods with either 9 μM SAHA alone or a combination of 0.04 μg/ml Colcemid with 4 μM SAHA or 4 mM VPA. Every 2-3 hours cells were harvested and stained with propidium iodide and flow cytometry analyses were performed. Combining HDAC-inhibitors with Colcemid resulted in a time-dependent increase of a sub-G0 population in DU-145 cells. A high resolution version of this figure can be downloaded here.



2 Study Synopsis

Analysis of Functional Annotation (AFA)[5,6] is conceptually similar to Gene Set Enrichment Analysis[7–9], and it is used to mine microarray gene expression data. Common functional themes used are Gene Ontology (GO) terms[10], and pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG)[11,12], as well as gene lists from other sources (i.e. the Molecular Signature data base[8,9]).

We applied AFA using several Functional Gene Set (FGS) collections, including protein-protein interaction (PPI) and transcription factor binding sites (TFBS), to extract biological meaning from differential gene expression as measured by microarray analysis upon HDAC inhibitors (HDACIs) treatment in prostate cancer cell lines. To this end we compared differential gene set enrichment across different prostate cancer cell lines (DU-145, PC3, and LNCaP), distinct HDACIs (SAHA, VPA, TSA, Buphenyl, CG-1521), and different treatment times (48 and 96 hours), using four independent datasets (GSE34452, GSE8645, GSE31620, and Connectivity Map).

3 Supplementary Materials and Methods

3.1 Microarray Pre-processing and Differential Gene Expression Analysis

3.1.1 GSE34452 data

Expression data for our original HDAC-inhibition study in PC3 and DU-145 cell (GSE34452) was processed using the R-Bioconductor[13,14] library limma[15–18]. A detailed explanation of all procedures and methods used for microarray data pre-processing, and differential gene expression analysis and detection was described previously[19].

3.1.2 GSE8645 data

LNCaP Prostate Cancer cells were treated for a period of 24h with either CG-1521 (7.5uM) or TSA (5uM) following a 24h seeding period. At the selected time point, total RNA was harvested from the cells for hybridization and analysis by Nimblgen Systems Inc using the homo sapiens gene expression array.

We downloaded the raw data files from GEO[20]. These could be found at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8645. The "Rquantile" normalization method[17,21] was applied to standardize log2 Cy5/Cy3 ratio (the so called M-value) distributions across different arrays. In order to assess the quality of the data, several diagnostic plots such as boxplot, MAplot, heatmap etc. were made.

Differential gene expression was investigated for the following contrasts:

- CG-1521 *versus* Control, using Control as the denominator in the log-ratio;
- TSA *versus* Control, using Control as the denominator in the log-ratio;
- Both (CG-1521+TSA) *versus* Control, using Control as the denominator in the log-ratio;

The data had biological replicates for which a correlation coefficient was computed between replicates and the associated consensus correlation was added to the model[16]. Finally, for each analyzed feature moderated t-statistics, log-odds ratios of differential expression (B-statistics), raw and adjusted p-values (FDR control by the Benjamini and Hochberg method[22]) were obtained. Gene annotation was based on the R/Bioconductor package `org.Hs.eg.db`. Gene-set enrichment analysis was performed to identify the biological concepts associated with the phenotypes and/or comparisons of interest.

3.1.3 GSE31620 data

Hudson and colleagues[23] monitored global miRNA expression changes in prostate cancer LNCaP cells treated with the epigenetic compounds 5-Azacytidine (5-AzaC) and/or trichostatin A (TSA). Cells were treated with epigenetic drugs for 36 hours and total RNA was isolated for hybridization to miRNA microarrays. 5 independent experiments were performed. The candidate prostate tumor suppressor miRNAs, miR-1, miR-206, and miR-27 were up-regulated in LNCaP cells for Affymetrix microarray analysis. LNCaP cells were transfected with pre-miR oligos and 24 hr post-transfection total RNA was collected for microarray analysis; total of three independent experiments.

We downloaded the raw data files from GEO[20]. These could be found at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31620. For miRNA data within-array dye effects were corrected by the "loess" normalization method[24]. The "Rquantile" normalization method[17,21] was applied to standardize log2 Cy5/Cy3 ratio (the so called M-value) distributions across different arrays. No background subtraction was performed prior to normalization[25,26]. The expression data was normalized using rma normalization method[27–29]. In order to assess the quality of the data, several diagnostic plots such as boxplot, MAplot, heatmap etc. were made.

For miRNA data, differential gene expression was investigated for the following contrasts:

- AzaC *versus* Control, using Control as the denominator in the log-ratio;
- TSA *versus* Control, using Control as the denominator in the log-ratio;
- Both (AzaC+TSA) *versus* Control, using Control as the denominator in the log-ratio;

For expression data, differential gene expression was investigated for the following contrasts:

- miR-206 *versus* Control, using Control as the denominator in the log-ratio;
- miR-1 *versus* Control, using Control as the denominator in the log-ratio;
- miR-27 *versus* Control, using Control as the denominator in the log-ratio;

Finally, for each analyzed feature moderated t-statistics, log-odds ratios of differential expression (B-statistics), raw and adjusted p-values (FDR control by the Benjamini and Hochberg method[22]) were obtained. Gene annotation was based on the R/Bioconductor package `org.Hs.eg.db`. Gene-set enrichment analysis is performed to identify the biological concepts associated with the phenotypes and/or comparisons of interest.

3.1.4 Connectivity Map data

Lamb and colleagues[2,3] monitored global miRNA expression changes, using the Affymetrix Human Genome U133A Array platform, in 4 distinct human cancer cell lines upon treatment with various doses of a wide range of FDA approved drugs (*"perturbagens"* in their definition). The cell lines utilized in this experiment included:

- the MCF7 breast cancer cell line;
- the HL60 leukemia cell line;
- the SKMEL5 melanoma cell line; and
- the PC3 prostate cancer cell line.

Overall in this experiment more than 1300 different compounds were used to treat these 4 cell lines, measuring gene expression before and after treatment for a total of over 7500 experiments. In Connectivity Map the PC3 cells were treated with various doses, as follows:

- Sodium phenylbutyrate (Buphenyl), 0.001M, 1 experiment;

- TSA 1^{-7}M, 39 experiments;
- TSA 1^{-6}M, 16 experiments;
- SAHA 1^{-5}M, 2 experiments;
- VPA 5^{-5}M, 2 experiments;
- VPA 5^{-4}M, 2 experiments;
- VPA 2^{-4}M, 2 experiments;
- VPA 1^{-3}M, 4 experiments;
- Buphenyl 1^{-3}M, 1 experiments;
- Untreated control, 237 experiments;

All cells were treated the drugs for 6 hours and total RNA was isolated for hybridization to the Affymetrix arrays.

We downloaded the complete set of raw data in the form of 7500 CEL files, along with phenotypic information about treatment regimens, from the Connectivity Map data base (http://www.broad.mit.edu/cmap). We used the 'frozen-RMA' normalization method[30,31] implemented by McCall and Irizarry in order to achieve a better control of batch effect. We further normalized across DNA-chips by quantile normalization[21]. In order to assess the quality of the data, several diagnostic plots such as boxplot, MAplot, heatmaps, etc. were made.

After gene expression data pre-processing using the entire dataset, we further analyzed the PC3 cell lines treated with HDACIs along with the corresponding untreated samples. Differential gene expression was investigated separately for each drug/dose combination mentioned above, controlling for platform and experimental batch, using a generalized linear model approach as implemented in the `limma` R/Bioconductor package. Finally, for each analyzed feature moderated t-statistics, log-odds ratios of differential expression (B-statistics), raw and adjusted p-values (FDR control by the Benjamini and Hochberg method[22]) were obtained. Gene annotation was based on the R/Bioconductor package `hgu133a.db`. Gene-set enrichment analysis is performed to identify the biological concepts associated with the phenotypes and/or comparisons of interest.

3.2 Correspondence-at-the-top and correlation analysis

The "Correspondence-at-the-top" curves (CAT-plot)[32] was implemented and successfully used to evaluate the agreement among distinct microarray studies, as previously described[33,34] This technique allows for comparing the agreement (*i.e* "correspondence") between two ranked vectors of features starting from the top, as follows:

1. The two vectors are ordered based on a ranking statistics (*e.g.* t-statistics for differential gene expression, p-values for significance, ...);
2. The proportion of overlapping elements is then computed starting from the top, considering more and more features until all used;
3. Finally the proportion of common elements are plotted against the increasing size of the vectors being compared creating a CAT-curve.

Since CAT-curves focus on the agreement between vectors at the top, they are particularly useful to compare vectors ordered based on differential gene expression studies, since only a small fraction of genes is expected to be different over the large total number of analyzed genes. We used cat-plots to evaluate the agreement among the different contrasts investigating differential gene expression upon HDAC-inhibition in prostate cancer cell lines (see Figures **S9** ,**S10**, **S11**, **S12**, **S13**, **S14**, **S15**, **S16**). To this end we ranked the genes based on the moderate t-statistics as obtained from our linear model analysis and performed the CAT-curve analysis as follows:

- By increasing ordering using the signed moderate t-statistics to investigate the down-regulated genes;

- By decreasing ordering using the signed moderate t-statistics, to investigate the up-regulated genes.

We also computed the Pearson's correlation between all pairs of moderated t-statistics and log2 fold-change obtained from our linear model analysis and displayed them using heatmaps (see Figures **S7**, **S8**). In this representation clustering based on this pari-wise correlation was achieved using the euclidian distance and the average clustering method.

Summary tables reporting the list of the differentially expressed genes identified for each study can be accessed for each analyzed contrast and for each study in html format:

- GSE34452 study, differentially expressed genes:
  - Comparisons to controls, on-line tables;
  - Comparisons by drug or cell line, on-line tables;
- GSE8645 study, differentially expressed genes, on-line tables;
- GSE31620 study, differentially expressed mRNA, on-line tables;
- GSE31620 study, differentially expressed miR, on-line tables;
- Connectivity Map study, differentially expressed genes, on-line tables;

3.3 Analysis of Functional Annotation

Enrichment analysis of functional themes was performed to capture biological processes affected by HDAC inhibition in the studied cellular models. The Wilcoxon rank sum test, as implemented in the `geneSetTest` function of the `limma` R-Bioconductor package[15,35], was applied to test whether each FGS was differentially expressed, up-regulated, or down-regulated across all the 8 investigated contrasts, using the moderated t-statistics from the linear model analysis to order the genes (see Kortenhorst et al for details[19]) . This function computes a p-value to test the hypothesis that a given genes set, defined by any functional annotation of interest, tends to be more highly ranked on a given statistic.

In the present study, individual, non redundant probes on the microarray were ranked by their absolute moderated t-statistics, and the enrichment p-values were computed by one-sided Wilcoxon rank sum tests. This approach enabled the identification of biological concepts enriched by differential gene expression irrespective to up- or down-regulation. In addition, genes were also ranked by their signed moderated t-statistics, performing separate tests on each tail, therefore enabling the identification of FGS enriched either by gene up-regulation upon HDACIs treatment or gene down-regulation.

The enrichment analysis was performed using all non-redundant genes present on the microarray, according to the NCBI Entrez Gene database annotation[36] (see details below). Filtering of redundant microarray features (i.e. probes mapping to the same NCBI Entrez Gene identifier) was achieved by retaining only the probes with the largest absolute t-statistics for further analysis.

Correction for multiple hypothesis testing was obtained separately for each FGS collection, by applying the Benjamini and Hochberg method[37] as implemented in the `multtest` R/Bioconductor package. Overall, our approach is analogous to Gene Set Enrichment Analysis (GSEA) like procedures[8,9]), and has already been successfully applied in other studies[5,6].

3.4 Microarray and Functional Gene Set Annotation

In the present study mappings between each FGS considered and the individual probes of the Agilent microarray were based on NCBI Entrez Gene identifiers[36], as obtained from the `hgug4110b.db`, `hgu133a.db`, and `org.Hs.eg.db` R/Bioconductor packages (see details below).

3.4.1 The hgug4110b.db metadata package

The hgug4110b.db metadata package was obtained from R/Bioconducor[13,14]. This package contains annotation information for the human Agilent array hgug4110b, as detailed below:

```
> hgug4110b()

Quality control information for hgug4110b:


This package has the following mappings:

hgug4110bACCNUM has 20173 mapped keys (of 20173 keys)
hgug4110bALIAS2PROBE has 65694 mapped keys (of 110538 keys)
hgug4110bCHR has 18252 mapped keys (of 20173 keys)
hgug4110bCHRLENGTHS has 93 mapped keys (of 93 keys)
hgug4110bCHRLOC has 18110 mapped keys (of 20173 keys)
hgug4110bCHRLOCEND has 18110 mapped keys (of 20173 keys)
hgug4110bENSEMBL has 17773 mapped keys (of 20173 keys)
hgug4110bENSEMBL2PROBE has 16597 mapped keys (of 19887 keys)
hgug4110bENTREZID has 18253 mapped keys (of 20173 keys)
hgug4110bENZYME has 2274 mapped keys (of 20173 keys)
hgug4110bENZYME2PROBE has 919 mapped keys (of 936 keys)
hgug4110bGENENAME has 18253 mapped keys (of 20173 keys)
hgug4110bGO has 16963 mapped keys (of 20173 keys)
hgug4110bGO2ALLPROBES has 13163 mapped keys (of 13360 keys)
hgug4110bGO2PROBE has 9949 mapped keys (of 10161 keys)
hgug4110bMAP has 18199 mapped keys (of 20173 keys)
hgug4110bOMIM has 13394 mapped keys (of 20173 keys)
hgug4110bPATH has 5758 mapped keys (of 20173 keys)
hgug4110bPATH2PROBE has 214 mapped keys (of 214 keys)
hgug4110bPFAM has 18196 mapped keys (of 20173 keys)
hgug4110bPMID has 18216 mapped keys (of 20173 keys)
hgug4110bPMID2PROBE has 272631 mapped keys (of 283543 keys)
hgug4110bPROSITE has 18196 mapped keys (of 20173 keys)
hgug4110bREFSEQ has 18212 mapped keys (of 20173 keys)
hgug4110bSYMBOL has 18253 mapped keys (of 20173 keys)
hgug4110bUNIGENE has 18223 mapped keys (of 20173 keys)
hgug4110bUNIPROT has 17718 mapped keys (of 20173 keys)


Additional Information about this package:

DB schema: HUMANCHIP_DB
DB schema version: 2.1
Organism: Homo sapiens
Date for NCBI data: 2010-Sep7
Date for GO data: 20100904
Date for KEGG data: 2010-Sep7
Date for Golden Path data: 2010-Mar22
Date for IPI data: 2010-Aug19
Date for Ensembl data: 2010-Aug5
```

3.4.2 The `org.Hs.eg.db` metadata package

The `org.Hs.eg.db` metadata package was obtained from R/Bioconducor[13,14]. This package contains annotation information for the human genome, as detailed below:

```
> org.Hs.eg()

Quality control information for org.Hs.eg:


This package has the following mappings:

org.Hs.egACCNUM has 30045 mapped keys (of 44811 keys)
org.Hs.egACCNUM2EG has 656242 mapped keys (of 656242 keys)
org.Hs.egALIAS2EG has 110538 mapped keys (of 110538 keys)
org.Hs.egCHR has 44424 mapped keys (of 44811 keys)
org.Hs.egCHRLENGTHS has 93 mapped keys (of 93 keys)
org.Hs.egCHRLOC has 22107 mapped keys (of 44811 keys)
org.Hs.egCHRLOCEND has 22107 mapped keys (of 44811 keys)
org.Hs.egENSEMBL has 19496 mapped keys (of 44811 keys)
org.Hs.egENSEMBL2EG has 19887 mapped keys (of 19887 keys)
org.Hs.egENSEMBLPROT has 19461 mapped keys (of 44811 keys)
org.Hs.egENSEMBLPROT2EG has 75463 mapped keys (of 75463 keys)
org.Hs.egENSEMBLTRANS has 19494 mapped keys (of 44811 keys)
org.Hs.egENSEMBLTRANS2EG has 109368 mapped keys (of 109368 keys)
org.Hs.egENZYME has 2142 mapped keys (of 44811 keys)
org.Hs.egENZYME2EG has 936 mapped keys (of 936 keys)
org.Hs.egGENENAME has 44811 mapped keys (of 44811 keys)
org.Hs.egGO has 17794 mapped keys (of 44811 keys)
org.Hs.egGO2ALLEGS has 13360 mapped keys (of 13360 keys)
org.Hs.egGO2EG has 10161 mapped keys (of 10161 keys)
org.Hs.egMAP has 37845 mapped keys (of 44811 keys)
org.Hs.egMAP2EG has 2601 mapped keys (of 2601 keys)
org.Hs.egOMIM has 14704 mapped keys (of 44811 keys)
org.Hs.egOMIM2EG has 17368 mapped keys (of 17368 keys)
org.Hs.egPATH has 5501 mapped keys (of 44811 keys)
org.Hs.egPATH2EG has 214 mapped keys (of 214 keys)
org.Hs.egPFAM has 24976 mapped keys (of 44811 keys)
org.Hs.egPMID has 30298 mapped keys (of 44811 keys)
org.Hs.egPMID2EG has 283543 mapped keys (of 283543 keys)
org.Hs.egPROSITE has 24976 mapped keys (of 44811 keys)
org.Hs.egREFSEQ has 28641 mapped keys (of 44811 keys)
org.Hs.egREFSEQ2EG has 91755 mapped keys (of 91755 keys)
org.Hs.egSYMBOL has 44811 mapped keys (of 44811 keys)
org.Hs.egSYMBOL2EG has 44796 mapped keys (of 44796 keys)
org.Hs.egUCSCKG has 20528 mapped keys (of 44811 keys)
org.Hs.egUNIGENE has 25212 mapped keys (of 44811 keys)
org.Hs.egUNIGENE2EG has 25814 mapped keys (of 25814 keys)
org.Hs.egUNIPROT has 18990 mapped keys (of 44811 keys)


Additional Information about this package:

DB schema: HUMAN_DB
DB schema version: 2.1
Organism: Homo sapiens
Date for NCBI data: 2010-Sep7
Date for GO data: 20100904
Date for KEGG data: 2010-Sep7
Date for Golden Path data: 2010-Mar22
Date for IPI data: 2010-Aug19
Date for Ensembl data: 2010-Aug5
```

3.4.3 The hgu133a.db metadata package

The **hgu133a.db** metadata package was obtained from R/Bioconducor[13,14]. This package contains annotation information for the Affymetrix hgu133a platform, as detailed below:

```
> hgu133a()

Quality control information for hgu133a:


This package has the following mappings:

hgu133aACCNUM has 22283 mapped keys (of 22283 keys)
hgu133aALIAS2PROBE has 53200 mapped keys (of 110538 keys)
hgu133aCHR has 20267 mapped keys (of 22283 keys)
hgu133aCHRLENGTHS has 93 mapped keys (of 93 keys)
hgu133aCHRLOC has 20066 mapped keys (of 22283 keys)
hgu133aCHRLOCEND has 20066 mapped keys (of 22283 keys)
hgu133aENSEMBL has 19742 mapped keys (of 22283 keys)
hgu133aENSEMBL2PROBE has 12921 mapped keys (of 19887 keys)
hgu133aENTREZID has 20273 mapped keys (of 22283 keys)
hgu133aENZYME has 3002 mapped keys (of 22283 keys)
hgu133aENZYME2PROBE has 869 mapped keys (of 936 keys)
hgu133aGENENAME has 20273 mapped keys (of 22283 keys)
hgu133aGO has 19270 mapped keys (of 22283 keys)
hgu133aGO2ALLPROBES has 12901 mapped keys (of 13360 keys)
hgu133aGO2PROBE has 9648 mapped keys (of 10161 keys)
hgu133aMAP has 20229 mapped keys (of 22283 keys)
hgu133aOMIM has 16682 mapped keys (of 22283 keys)
hgu133aPATH has 7585 mapped keys (of 22283 keys)
hgu133aPATH2PROBE has 214 mapped keys (of 214 keys)
hgu133aPFAM has 20157 mapped keys (of 22283 keys)
hgu133aPMID has 20197 mapped keys (of 22283 keys)
hgu133aPMID2PROBE has 266196 mapped keys (of 283543 keys)
hgu133aPROSITE has 20157 mapped keys (of 22283 keys)
hgu133aREFSEQ has 20182 mapped keys (of 22283 keys)
hgu133aSYMBOL has 20273 mapped keys (of 22283 keys)
hgu133aUNIGENE has 20232 mapped keys (of 22283 keys)
hgu133aUNIPROT has 19681 mapped keys (of 22283 keys)


Additional Information about this package:

DB schema: HUMANCHIP_DB
DB schema version: 2.1
Organism: Homo sapiens
Date for NCBI data: 2010-Sep7
Date for GO data: 20100904
Date for KEGG data: 2010-Sep7
Date for Golden Path data: 2010-Mar22
Date for IPI data: 2010-Aug19
Date for Ensembl data: 2010-Aug5
```

3.4.4 Functional Gene Set Collections

Overall we analyzed 43 FGS collections, which were obtained from various databases, encompassing distinct biological and molecular concepts (see Table **T1** below for details) including:

1. Cytogenetic bands and chromosomes;
2. Gene Ontology Terms (GO)[10,38],
3. Signaling pathways from KEGG[11,12,39] and other databases;
4. Functional themes from the Molecular Signature Database (MSigDb)[8,9], (see: MSigDb);
5. Protein-Protein-Interaction (PII) networks obtained from Biogrid[40], the Biomolecular Interaction Network Databas (BIND)[41], and the Human Protein Reference Database (HPRD)[42] databases;
6. Genes sharing conserved Transcription Factor Binding Site (TFBS), as defined in the University of California at Santa Cruz (UCSC) GoldenPath data base[43,44] (see Table **T1** below for details);
7. MicroRNA targets according to a number of different databases and prediction algorithms, including miRGen data base[45], PicTar[46], TargetScanS[47], tarbase[48,49], miRBase[50], mirtarget2[51], miRanda[52], and DIANA-microT[53,54]) (see Table **T1** below for details);
8. Genes co-cited in published manuscripts as recorded in PubMed;
9. Genes co-cited in the Online Mendelian Inheritance (OMIM) database;
10. Genes sharing similar protein domain, as defined in the Prosite database;
11. Genes annotated to the same Enzyme Commission number (EC number);

Table **T1** reports the description and source of each Functional Gene Sets collection used in the presents study to perform the the enrichment analysis.

Supplementary Table T1: Description and source of each Functional Gene Set (FGS) collections used in the present study

FGS Collection	Description	Source
GO	Gene Ontology; FGS defined based on Gene Ontology annotation, as obtained from the R/Bioconductor org.Hs.eg.db metadata package. In this GO collection each gene is associated to all parents GO terms.	R/Bioconductor
KEGG	KEGG; FGS defined based on the KEGG pathways data base. Data obtained from the R/Bioconductor org.Hs.eg.db metadata package	R/Bioconductor
Broad.c1.CYTOBAND	Broad Institute MSigDB Molecular Signature Database (MSigDB) Gene sets corresponding to each human chromosome and each cytogenetic band harboring at least one gene	MsigDB
Broad.c2.CGP	Broad Institute MSigDB CGP gene sets: chemical and genetic perturbations. These FGS represent gene expression signatures of genetic and chemical perturbations	MsigDB
Broad.c2.CP	Broad Institute MSigDB CP gene sets: canonical pathways. These FGS are canonical representations of a biological process compiled by domain experts	MsigDB
Broad.c2.CP.BIOCARTA	Broad Institute MSigDB canonical pathways BioCarta gene sets. These FGS are derived from the BioCarta pathway database	MsigDB
Broad.c2.CP.KEGG	Broad Institute MSigDB canonical pathways KEGG gene sets. These FGS are derived from the KEGG pathway database	MsigDB
Broad.c2.CP.REACTOME	Broad Institute MSigDB canonical pathways Reactome gene sets. These FGS are derived from the REACTOME pathway database	MsigDB

Continued on next page

Supplementary Table T1 – *Continued from previous page*

FGS Collection	Description	Source
Broad.c3.MIR	Broad Institute MSigDB regulatory motifs gene sets. FGS that contain genes that share a cis-regulatory motif that is conserved across the human, mouse, rat, and dog genomes. The motifs are catalogued in Xie, et al (2005, Nature), and represent known or likely regulatory elements. This collection contains 3'-UTR microRNA binding motifs.	MsigDB
Broad.c3.TFT	Broad Institute MSigDB regulatory motifs gene sets. FGS that contain genes that share a cis-regulatory motif that is conserved across the human, mouse, rat, and dog genomes. The motifs are catalogued in Xie, et al (2005, Nature), and represent known or likely regulatory elements. This collection contains FGS accounting for the genes that share a specific transcription factor binding site as defined in the TRANSFAC database (version 7.4)	MsigDB
Broad.c4.CGN	Broad Institute MSigDB computational gene sets: CGN, cancer gene neighborhood. FGS computationally derived from large collections of cancer-oriented microarray data. This collection accounts for expression neighborhoods centered on 380 cancer-associated genes, as defined in Brentani, et al (2003, PNAS)	MsigDB
Broad.c4.CM	Broad Institute MSigDB computational gene sets: CM, cancer modules. FGS computationally derived from large collections of cancer-oriented microarray data. This collection accounts for gene expression modules defined by Segal et al (Nature Genetics, 2004). Briefly, the authors compiled gene sets ('modules') from a variety of resources such as KEGG, GO, and others. By mining a large compendium of cancer-related microarray data, they identified 456 such modules as significantly changed in a variety of cancer conditions.	MsigDB
Broad.c5.BP	Broad Institute MSigDB Gene Ontology gene sets: BP FGS are derived from the Biological Process Gene Ontology (see guidelines)	MsigDB
Broad.c5.CC	Broad Institute MSigDB Gene Ontology gene sets: CC FGS are derived from the Cellular Component Gene Ontology (see guidelines)	MsigDB
Broad.c5.MF	Broad Institute MSigDB Gene Ontology gene sets: MF FGS are derived from the Molecular Function Gene Ontology (see guidelines)	MsigDB
PMID	PubMed; FGS defined based on the PubMed identifiers obtained from the org.Hs.eg.db R/Bioconductor metadata package	R/Bioconductor
OMIM	OMIM; FGS defined based on the OMIM identifiers obtained from the org.Hs.eg.db R/Bioconductor metadata package	R/Bioconductor
ChromosomalTiles5Mb	This collection accounts for FGS containing all the genes located on consecutive five Mb chromosomal tiles, as obtained from Stanford Microarray Database; a description is available at Synthetic genes page on SMD	Tibshirani's webpage
Prosite	Prosite; FGS defined based on proteins domains, families, and functional sites, as defined in the Prosite data base. Data obtained from the org.Hs.eg.db R/Bioconductor metadata package	R/Bioconductor
Enzyme	Enzyme Commission number; FGS defined based on the Enzyme Commission number. Data obtained from the org.Hs.eg.db R/Bioconductor metadata package	R/Bioconductor
ppi.BIND	Protein-protein-interaction data from the BIND database, as listed in the Entrez Gene data base	NCBI Entrez Gene
ppi.BioGRID	Protein-protein-interaction data from the BioGRID database, as listed in the Entrez Gene data base	NCBI Entrez Gene
ppi.HPRD	Protein-protein-interaction data from the HPRD database, as listed in the Entrez Gene database	NCBI Entrez Gene
ppi.anyDB	Protein-protein-interaction data from of any of the data bases listed above, as listed in the Entrez Gene data base	NCBI Entrez Gene
pathwayCommons.cell-map	FGS corresponding to pathways defined in the Pathway Commons data base. This collection accounts for gene lists from the Cancer Cell Map collection (Memorial Sloan-Kettering Cancer Center)	Pathway Commons

Continued on next page

Supplementary Table T1 – *Continued from previous page*

FGS Collection	Description	Source
pathwayCommons.humancyc	FGS corresponding to pathways defined in the Pathway Commons data base. This collection accounts for gene lists from HumanCyc collection (the Encyclopedia of Human Genes and Metabolism)	Pathway Commons
pathwayCommons.nci-nature	FGS corresponding to pathways defined in the Pathway Commons data base. This collection accounts for gene lists from the NCI/Nature Pathway Interaction Database. The Pathway Interaction Database is a collaborative project between the US National Cancer Institute (NCI) and Nature Publishing Group (NPG)	Pathway Commons
pathwayCommons.reactome	FGS corresponding to pathways defined in the Pathway Commons database. This collection accounts for gene lists from the Reactome collection, which is a knowledgebase of biological processes	Pathway Commons
miranda.targets	miRNA targets as predicted by the miranda algorithm. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
mirbase.targets	miRNA targets as obtained from the miRBasedatabase. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
mirtarget2.targets	miRNA targets as obtained from the mirtarget2 algorithm. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
pictar.targets	miRNA targets as predicted by the PicTar algorithm. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
tarbase.targets	miRNA targets as obtained from the TabBasedatabase. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
targetscan.targets	miRNA targets as predicted by the targetscanalgorithm. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
miRNAtargetIntersection	FGS resulting from the combinations of target predictions resulting from different algorithms and databases. In particular FGS in this collection account for predicted target **intersections of any three** of the following: targetscan, tarbase, pictar, mirtarget2, mirbase, or miranda. Such intersection lists might provide more specific results. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
miRNAtargetUnion	FGS resulting from the combinations of target predictions resulting from different algorithms and databases. In particular FGS in this collection account for predicted target **unions of any three** of the following: targetscan, tarbase, pictar, mirtarget2, mirbase, or miranda. Such intersection lists might provide more specific results. Data obtained from the RmiR.Hs.miRNA R/Bioconductor metadata package	R/Bioconductor
hprdBatch.UP	Genes up-regulated by pathway activation; the genes contained in this collection are the up-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (batch download)	NetPath
hprdBatch.DOWN	Genes down-regulated by pathway activation; the genes contained in this collection are the down-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (batch download)	NetPath
hprdBatch.DIFFERENT	Genes up- and down-regulated by pathway activation; the genes contained in this collection are the up- and down-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (batch download)	NetPath
hprdManual.UP	Genes up-regulated by pathway activation; the genes contained in this collection are the up-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (manual download)	NetPath
hprdManual.DOWN	Genes down-regulated by pathway activation; the genes contained in this collection are the down-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (manual download)	NetPath

Continued on next page

Supplementary Table T1 – *Continued from previous page*

FGS Collection	Description	Source
hprdManual.DIFFERENT	Genes up- and down-regulated by pathway activation; the genes contained in this collection are the up- and down-regulated targets induced by the activation of the signaling pathway; the gene lists were manually curated, result from the evaluation of evidence available from the literature, and are available from the HPRD data base (manual download)	NetPath
tfbsK3Z3	TRANSFAC Transcription Factor Binding Site (TFBS); the genes contained in this collection have a TFBS in the genomic region around their transcription starting site (TSS); The genomic window considered spans from 3kb before the TSS to 3kb after the TSS, with a Z-score for conservation of 3.0, corresponding to a False Discovery Rate of less than 1%. Details are available from the UCSC Genome Browser	GoldenPath

3.5 AFA results exploration

We selected and considered significantly enriched FGS that showed an adjusted p-values $< 5\%$ after correction for multiple testing. Summary tables, reporting the top differentially expressed FGS (adjusted p-values of 5% or less), can be accessed for each endpoint and for each study in html format:

- GSE34452 study, AFA results on-line:
  - Comparisons to control;
  - Comparisons by cell line or drug;
- GSE8645 study, AFA results on-line;
- GSE31620 study, AFA results on-line;
- Connectivity Map study, AFA results on-line;

Heatmaps were also used to display and explore significant AFA results for all studies, and can be accessed on line. When we considered the GSE34452 study alone, we selected FGS, if any, that were significantly enriched (adjusted p-values $< 5\%$) across **all** considered comparisons. When we considered all the studies together (GSE34452, GSE8645, GSE31620, and Connectivity Map) we selected FGS, if any, that were significantly enriched (adjusted p-values $< 5\%$) in a least **one** comparison. In all theses heatmaps for GSE34452 we used only the comparisons to control cells.

- GSE34452 study alone:
  - FGS enriched (FDR $< 5\%$ in all comparisons) by differential gene expression;
  - FGS enriched (FDR $< 5\%$ in all comparisons) by gene up-regulation;
  - FGS enriched (FDR $< 5\%$ in all comparisons) by gene down-regulation;
- All studies together (GSE34452, GSE8645, GSE31620, and Connectivity Map):
  - FGS enriched (FDR $< 5\%$ in at least one comparison) by differential gene expression;
  - FGS enriched (FDR $< 5\%$ in at least one comparison) by gene up-regulation;
  - FGS enriched (FDR $< 5\%$ in at least one comparison) by gene down-regulation;

We used the color scale to represent the absolute negative base 10 logarithm obtained from of the adjusted Wilcoxon rank sum test P values. FGS were represented in the as rows, and were clustered using the Euclidian distance and the average clustering method, while the different HDACIs treatments were represented by column and were not reordered. This exploratory approach intuitively allowed us to detect commonalities and peculiarities among the different biological contrasts we evaluated.

3.6 FGS communities

We used social network analysis to explore the relationship among the enriched FGS and to assess whether the enrichment was driven by common or distinct gene modules, as follows:

- First we assembled "gene by FGS" matrices indicating the membership of each gene to the enriched FGS from Figures 1A, 1B, and 1C in the main paper;
- We then performed hierarchical clustering to group the enriched FGS based on common genes, using the binary distance and the Ward clustering method (see Figures **S1**, **S1**, and **S1**);
- We then represented the gene-FGS membership data as adjacency matrices and use this information to reconstruct the corresponding networks using weighted undirected graphs;
- We subsequently performed social network analysis to identify distinct FGS communities using the fast greedy modularity optimization algorithm described by Clauset and colleagues[4] (see Figures **S4**, **S5**, and **S6**);
- We also repeated all analytical steps above filtering the genes based on increasing significance in differential gene expression analysis (FDR $< 10\%$, $< 5\%$, $< 1\%$, and $< 0.1\%$) with similar results.

Similarly to what we have done for AFA, we applied social network analysis separately for differentially expressed, up-regulated, and down-regulated FGS and genes. Overall this analysis revealed distinct FGS communities for which the enrichment is driven by distinct sets of differentially expressed genes. Even most interestingly these FGS communities represent distinct and complementary biological concepts. For instance the FGS communities for the upstream signaling pathways proved to be distinct from those for the downstream target genes. These findings may suggest that HDACIs treatment not only can modulate cell signaling pathways, but also that such modulation may result in the reactivation of the downstream responses depending on such pathways.

For instance, among the FGS that were differentially expressed upon HDACIs treatment in PC3 and DU-145 cells four main subgraphs were identified based on the most significant differentially expressed genes (FDR $< 0.1\%$, see Figure **S4**):

1. FGS related to modulation of cell-cycle and gene expression (highlighted in green in Figure **S4**):
   - REACTOME gene lists for "Gene Expression", "mRNA Processing", and "Processing of Capped Intron-Containing Pre-mRNA";
   - ENTREZ protein-protein-interaction networks for MEPCE, MGMT CDT1, and 155871.
2. FGS related to cell signaling (highlighted in red in Figure **S4**):
   - NCI Pathways for C-MYC, p53 and EGF receptor signaling;
   - ENTREZ protein-protein-interaction networks for TP53.
3. FGS related to cell signaling transcriptional responses (highlighted in purple in Figure **S4**):
   - HPRD gene lists for downstream targets of "IL-2 Signaling Pathway", "Androgen Receptor Signaling Pathway", "EGFR1 Signaling Pathway", "B Cell Receptor Signaling Pathway", "TGF-beta Receptor Signaling Pathway", and "TNF-alpha Signaling Pathway".
4. Regulatory networks associated with specific TFBS and miRs (highlighted in cyan in Figure **S4**):
   - TFBS target gene lists for "V$MYC_Q2", "V$NRF1_Q6", "V$GABP_B", "V$ELK1_02", "V$MYC_Q2", "V$USF_01", and "V$SP1_Q6"
   - miR target gene lists for "MIR-381", "MIR-9", "MIR-506", and "MIR-200B,MIR-200C,MIR-429".

Among the FGS that were up-regulated upon HDACIs treatment in PC3 and DU-145 cells three main subgraphs were identified based on the most significant up-regulated genes (FDR $< 0.1\%$, see Figure **S5**):

1. Immune system related FGS (highlighted in green in Figure **S5**):
   - REACTOME gene lists for "Signaling in Immune system", "Interferon Signaling", "Interferon alpha/beta signaling", and "Interferon gamma signaling";
   - ENTREZ protein-protein-interaction networks for LILRB1, and LILRB2.
2. EGF signaling related gene lists (highlighted in red in Figure **S5**):
   - NCI Pathways for "ErbB receptor signaling network", "EGF receptor (ErbB1) signaling pathway", "Internalization of ErbB1", "ErbB1 downstream signaling", "Proteogylcan syndecan-mediated signaling events", "Plasma membrane estrogen receptor signaling", and "Integrin cell surface interactions";
   - ENTREZ protein-protein-interaction networks for ITGB1, ITGB5, ADAM9, and MYO10;
   - HPRD gene lists for downstream targets of "Wnt Signaling Pathway".
3. Transcriptional responses and regulatory networks associated with specific TFBS and miR (highlighted in blue in Figure **S5**):
   - HPRD gene lists for downstream targets of "EGFR1 Signaling Pathway"; "HPRD Androgen Receptor Signaling Pathway", "TGF-beta Receptor Signaling Pathway";
   - TFBS targets potentially downstream immune system signaling pathways: "V$E12_Q6" and "V$E12_Q6";
   - TFBS targets potentially downstream the EGFR signaling pathway: "V$AP1_C" and "V$FOXO4_01";
   - miR target lists for "MIR-17-5P,MIR-20A,MIR-106A,MIR-106B,MIR-20B,MIR-519D", "MIR-200B,MIR-200C,MIR-429", "MIR-218", "MIR-519C,MIR-519B,MIR-519A", "MIR-124A", and "MIR-19A,MIR-19B".

Finally, among the FGS that were down-regulated upon HDACIs treatment in PC3 and DU-145 cells four main subgraphs were identified based on the most significant down-regulatedgenes (FDR $< 0.1\%$, see Figure **S6**):

1. FGS related to gene expression (highlighted in green in Figure **S6**):
   - REACTOME gene lists for "Gene Expression", "mRNA Processing", and "Processing of Capped Intron-Containing Pre-mRNA";
   - ENTREZ protein-protein-interaction networks for MEPCE, SRRM1, SRRM2, SFRS12, and 155871.
2. FGS related to cell-cycle progression (highlighted in purple in Figure **S6**):
   - REACTOME gene lists for "DNA Replication", "Cell Cycle, Mitotic", and "Mitotic M?M/G1 phases".
3. Transcriptional responses and regulatory networks associated with cell cycle progression and regulation (highlighted in red in Figure **S6**):
   - TFBS targets potentially potentially regulated by the the E2F family of transcription factors: "V$E2F_Q4" and "V$E2F_Q6", "V$E2F1_Q6_01", and "V$E2F_Q4_01" ;
   - NCI Pathways for "BARD1 signaling events" and "Regulation of Telomerase".
4. Transcriptional responses and regulatory networks associated with the c-Myc signaling pathway (highlighted in cyan in Figure **S6**):
   - TFBS targets potentially potentially regulated by the the c-Myc: "V$MYC_Q2" and "V$MYCMAX_01";
   - NCI Pathways for "C?MYC pathway" and "Validated targets of C?MYC transcriptional activation".

Similar networks and communities were obtained by applying less stringent criteria to filter the underlying genes (FDR $< 10\%$, $< 5\%$, $< 1\%$, data not shown).

3.7 Software

All analyses were performed using analytical packages from the R/Bioconductor project [13,14], including `limma` [16], `affy` [55], `RTopper` [56], `matchBox`, `igraph`, and `multtest`. Additional functions and methods were developed by Dr. Marchionni and implemented in additional packages available from http://luigimarchionni.org/software.html.

4 Literature Cited

References

[1] David L Wheeler, Tanya Barrett, Dennis A Benson, Stephen H Bryant, Kathi Canese, Vyacheslav Chetvernin, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res, Jan 2007, 35(Database issue):D5–12.

[2] Justin Lamb. The Connectivity Map: a new tool for biomedical research. Nature Reviews Cancer, 2007, 7(1):54–60.

[3] Justin Lamb, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science (New York, NY), September 2006, 313(5795):1929–1935.

[4] Aaron Clauset, M Newman, and Cristopher Moore. Finding community structure in very large networks. Physical Review E, December 2004, 70(6):066111.

[5] E. M. Schaeffer, L. Marchionni, Z. Huang, B. Simons, A. Blackman, W. Yu, et al. Androgen-induced programs for prostate epithelial growth and invasion arise in embryogenesis and are reactivated in cancer. Oncogene, Dec 2008, 27(57):7180–7191.

[6] Vincent C Daniel, Luigi Marchionni, Jared S Hierman, Jonathan T Rhodes, Wendy L Devereux, Charles M Rudin, et al. A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. Cancer Res, Apr 2009, 69(8):3364–3373.

[7] Seon-Young Kim and David J Volsky. Page: parametric analysis of gene set enrichment. BMC Bioinformatics, 2005, 6:144.

[8] V. K. Mootha, P. Lepage, K. Miller, J. Bunkenborg, M. Reich, M. Hjerrild, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc Natl Acad Sci U S A, 2003, 100(2):605–10. 0027-8424 (Print) Journal Article.

[9] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, Oct 2005, 102(43):15545–15550.

[10] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet, 2000, 25(1):25–9. 1061-4036 (Print) Journal Article.

[11] H. Ogata, S. Goto, W. Fujibuchi, and M. Kanehisa. Computation with the KEGG pathway database. Biosystems, 1998, 47(1-2):119–28. 0303-2647 (Print) Journal Article.

[12] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 1999, 27(1):29–34. 0305-1048 (Print) Journal Article.

[13] R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics, 1996, 5:299–314.

[14] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol, 2004, 5(10):R80. 1465-6914 (Electronic) Journal Article.

[15] G. K. Smyth. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 2004, 3(Article 3).

[16] G. K. Smyth, J. Michaud, and H. S. Scott. Use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics, 2005, 21(9):2067–75. 1367-4803 (Print) Evaluation Studies Journal Article Validation Studies.

[17] G. K. Smyth and T. Speed. Normalization of cDNA microarray data. Methods, 2003, 31(4):265–73. 1046-2023 (Print) Journal Article.

[18] G. K. Smyth, Y. H. Yang, and T. Speed. Statistical issues in cDNA microarray data analysis. Methods Mol Biol, 2003, 224:111–36. 1064-3745 (Print) Journal Article.

[19] Madeleine S Q Kortenhorst, Marianna Zahurak, Shabana Shabbeer, Sushant Kachhap, Nathan Galloway, Giovanni Parmigiani, et al. A multiple-loop, double-cube microarray design applied to prostate cancer cell lines with variable sensitivity to histone deacetylase inhibitors. Clin Cancer Res, Nov 2008, 14(21):6886–6894.

[20] D. L. Wheeler, T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, D. M. Church, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res, 2005, 33(Database issue):D39–45. 1362-4962 (Electronic) Journal Article.

[21] Y. H. Yang and N. Thorne. Normalization for two-color cDNA microarray data. In D. R. Goldstein, editor, *Science and Statistics: A Festschrift for Terry Speed*, volume 40, pages 403–41. 2003.

[22] Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, and I. Golani. Controlling the false discovery rate in behavior genetics research. Behav Brain Res, 2001, 125(1-2):279–84. 0166-4328 (Print) Comment Journal Article.

[23] Robert S. Hudson, Ming Yi, Dominic Esposito, Stephanie K. Watkins, Arthur A. Hurwitz, Harris G. Yfantis, et al. Microrna-1 is a candidate tumor suppressor and prognostic marker in human prostate cancer. Nucleic Acids Res, 2012, 40.

[24] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res, 2002, 30(4):e15. 1362-4962 (Electronic) Journal Article.

[25] Robert B Scharpf, Christine A Iacobuzio-Donahue, Julie B Sneddon, and Giovanni Parmigiani. When should one subtract background fluorescence in 2-color microarrays? Biostatistics, Oct 2007, 8(4):695–707.

[26] M. Zahurak, G. Parmigiani, W. Yu, R. B. Scharpf, D. Berman, E. Schaeffer, et al. Pre-processing Agilent microarray data. BMC Bioinformatics, 2007, 8:142. 1471-2105 (Electronic) Journal Article Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S.

[27] Rafael A. Irizarry, Benjamin M. Bolstad, Francois Collin, Leslie M. Cope, Bridget Hobbs, and Terence P. Speed. Summaries of affymetrix genechip probe level data. Nucleic Acids Research, 2003, 31(4):e15.

[28] B.M. Bolstad, R.A Irizarry, M. ÃĚstrand, and T.P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and biasmyth. Bioinformatics, 2003, 19(2):185–193.

[29] Rafael A. Irizarry, Bridget Hobbs, Francois Collin, Yasmin D. Beazer Barclay, Kristen J. Antonellis, Uwe Scherf, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 2003, 4(2):249–264.

[30] Matthew N McCall, Harris A Jaffee, and Rafael A Irizarry. fRMA ST: Frozen robust multiarray analysis for Affymetrix Exon and Gene ST arrays. Bioinformatics (Oxford, England), October 2012.

[31] Matthew N McCall and Rafael A Irizarry. Thawing Frozen Robust Multi-array Analysis (fRMA). BMC Bioinformatics, 2011, 12:369.

[32] Rafael A Irizarry, Daniel Warren, Forrest Spencer, Irene F Kim, Shyam Biswal, Bryan C Frank, et al. Multiple-laboratory comparison of microarray platforms. Nat Methods, May 2005, 2(5):345–350.

[33] B Benassi, R Flavin, L Marchionni, S Zanata, Y Pan, D Chowdhury, et al. c-Myc is activated via USP2a-mediated modulation of microRNAs in prostate cancer. Cancer Discovery, 2012.

[34] Ashley E Ross, Luigi Marchionni, Milena Vuica-Ross, Chris Cheadle, Jinshui Fan, David M Berman, et al. Gene expression pathways of high grade localized prostate cancer. The Prostate, February 2011.

[35] G. K. Smyth. Limma: linear models for microarray data. In R. Gentleman, R. V. Carey, S. Dudoit, R. Irizarry, and W. Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420. Springer, New York, 2005.

[36] Lewis Y Geer, Aron Marchler-Bauer, Renata C Geer, Lianyi Han, Jane He, Siqian He, et al. The ncbi biosystems database. Nucleic Acids Res, Jan 2010, 38(Database issue):D492–D496.

[37] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 1995, 57:289–300.

[38] M. A. Harris, J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res, 2004, 32(Database issue):D258–61. 1362-4962 (Electronic) Journal Article.

[39] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. The KEGG resource for deciphering the genome. Nucleic Acids Res, 2004, 32(Database issue):D277–80. 1362-4962 (Electronic) Journal Article.

[40] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. BioGRID: a general repository for interaction datasets. Nucleic Acids Res, Jan 2006, 34(Database issue):D535–D539.

[41] G. D. Bader, I. Donaldson, C. Wolting, B. F. Ouellette, T. Pawson, and C. W. Hogue. BIND–The Biomolecular Interaction Network Database. Nucleic Acids Res, Jan 2001, 29(1):242–245.

[42] Suraj Peri, J. Daniel Navarro, Troels Z Kristiansen, Ramars Amanchy, Vineeth Surendranath, Babylakshmi Muthusamy, et al. Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res, Jan 2004, 32(Database issue):D497–D501.

[43] R. M. Kuhn, D. Karolchik, A. S. Zweig, T. Wang, K. E. Smith, K. R. Rosenbloom, et al. The ucsc genome browser database: update 2009. Nucleic Acids Res, Jan 2009, 37(Database issue):D755–D761.

[44] W. James Kent, Charles W Sugnet, Terrence S Furey, Krishna M Roskin, Tom H Pringle, Alan M Zahler, et al. The human genome browser at ucsc. Genome Res, Jun 2002, 12(6):996–1006.

[45] Molly Megraw, Praveen Sethupathy, Benoit Corda, and Artemis G Hatzigeorgiou. mirgen: a database for the study of animal microrna genomic organization and function. Nucleic Acids Res, Jan 2007, 35(Database issue):D149–D155.

[46] Azra Krek, Dominic Grün, Matthew N Poy, Rachel Wolf, Lauren Rosenberg, Eric J Epstein, et al. Combinatorial microrna target predictions. Nat Genet, May 2005, 37(5):495–500.

[47] Benjamin P Lewis, Christopher B Burge, and David P Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. Cell, Jan 2005, 120(1):15–20.

[48] Praveen Sethupathy, Benoit Corda, and Artemis G Hatzigeorgiou. Tarbase: A comprehensive database of experimentally supported animal microrna targets. RNA, Feb 2006, 12(2):192–197.

[49] Giorgos L Papadopoulos, Martin Reczko, Victor A Simossis, Praveen Sethupathy, and Artemis G Hatzigeorgiou. The database of experimentally supported targets: a functional update of tarbase. Nucleic Acids Res, Jan 2009, 37(Database issue):D155–D158.

[50] Sam Griffiths-Jones, Russell J Grocock, Stijn van Dongen, Alex Bateman, and Anton J Enright. mirbase:

microrna sequences, targets and gene nomenclature. Nucleic Acids Res, Jan 2006, 34(Database issue):D140–D144.

[51] Xiaowei Wang and Issam M El Naqa. Prediction of both conserved and nonconserved microrna targets in animals. Bioinformatics, Feb 2008, 24(3):325–332.

[52] Bino John, Anton J Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, and Debora S Marks. Human microrna targets. PLoS Biol, Nov 2004, 2(11):e363.

[53] M. Maragkakis, M. Reczko, V. A. Simossis, P. Alexiou, G. L. Papadopoulos, T. Dalamagas, et al. Diana-microt web server: elucidating microrna functions through target prediction. Nucleic Acids Res, Jul 2009, 37(Web Server issue):W273–W276.

[54] Marianthi Kiriakidou, Peter T Nelson, Andrei Kouranov, Petko Fitziev, Costas Bouyioukos, Zissimos Mourelatos, et al. A combined computational-experimental approach predicts human microrna targets. Genes Dev, May 2004, 18(10):1165–1178.

[55] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics, 2004, 20(3):307–15. 1367-4803 (Print) Evaluation Studies Journal Article.

[56] Svitlana Tyekucheva, Luigi Marchionni, Rachel Karchin, and Giovanni Parmigiani. Integrating diverse genomic data using gene sets. Genome biology, October 2011, 12(10):R105.