Electronic Supplementary Material (ESI) for Molecular Omics.

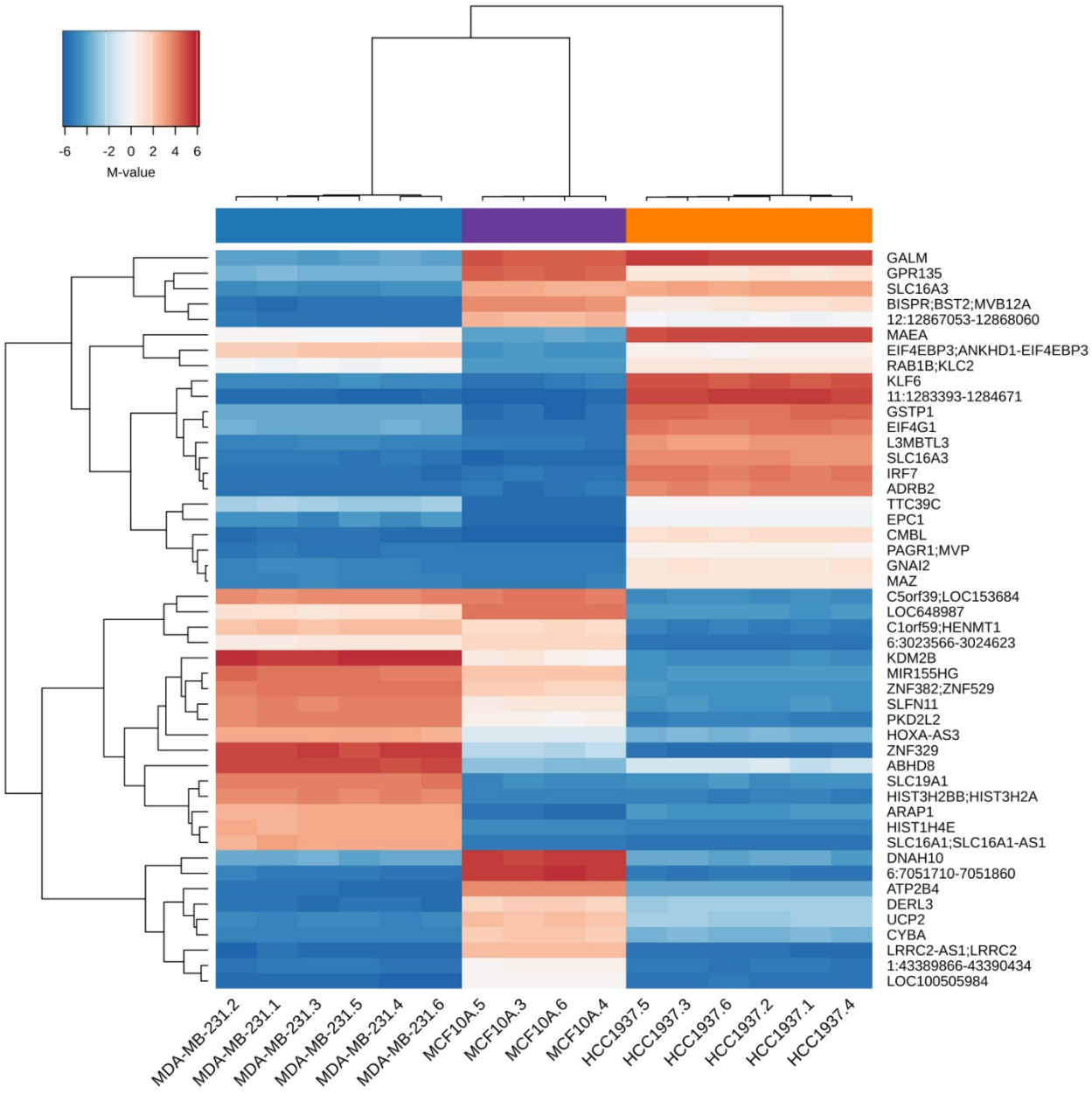
Multi-omics data integration reveals correlated regulatory features of triple negative breast cancer

Kevin Chappell[1†], Kanishka Manna[1†], Charity L. Washam[1,2], Stefan Graw[1,2,3], Duah Alkam[1], Matthew Thompson[1], Maroof K Zafar[1], Lindsey Hazeslip[1], Christopher Randolph[2], Allen Gies[1], Jordan T. Bird[1], Alicia K Byrd[1], Sayem Miah[1], Stephanie D. Byrum[1,2*]

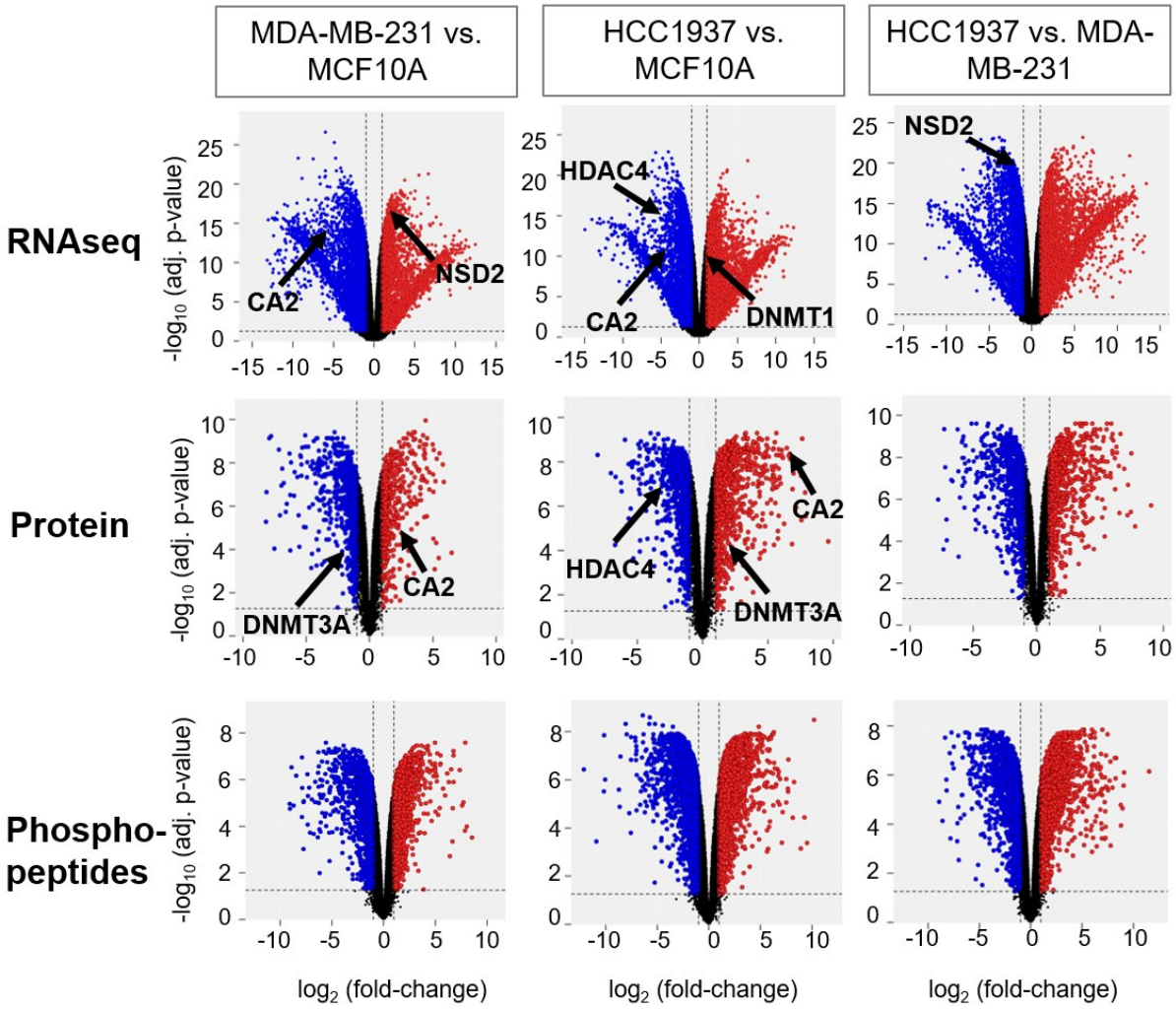
Supplemental Materials

Supplemental Table 1. Sample names for the multi-omics data integration analysis. The normalized values for each method and sample can be found in the supplemental files 1-4.

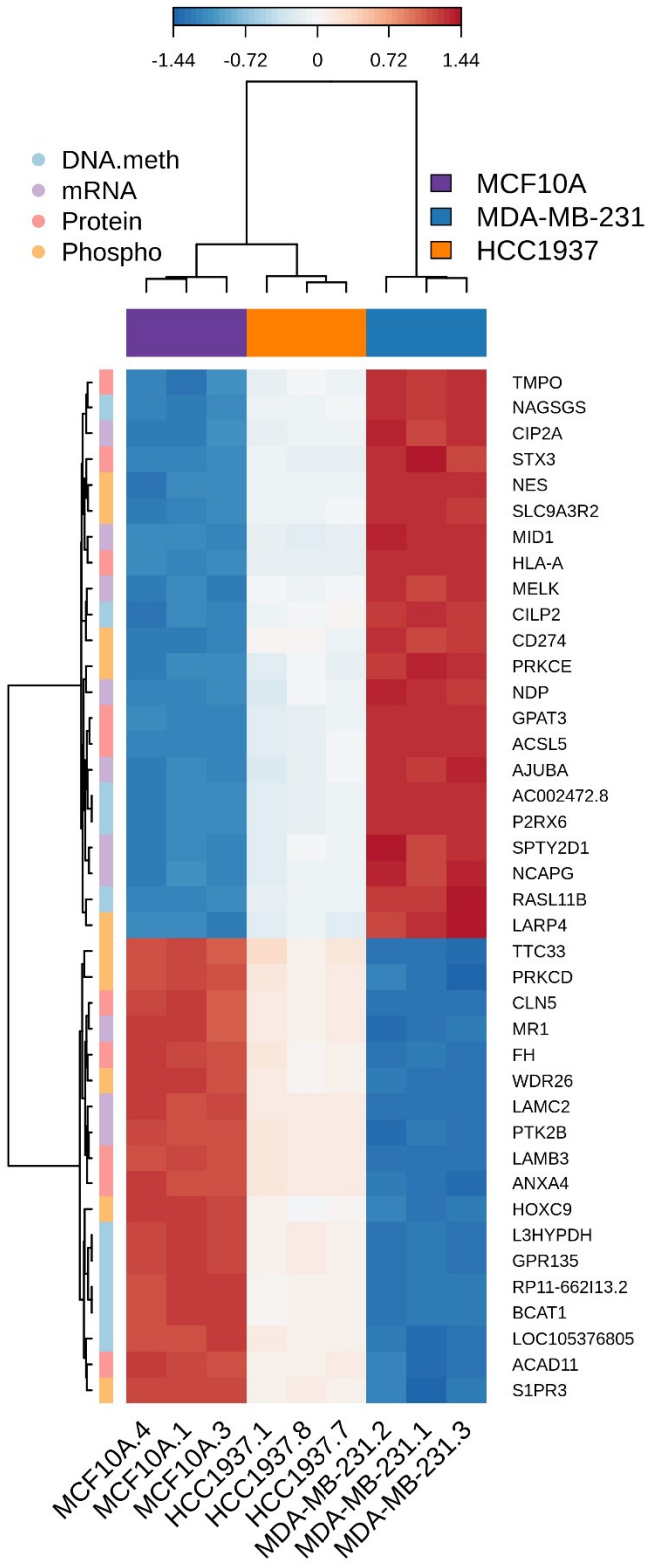
Integrated	**DNA.meth**	**RNA**	**Protein**	**Phospho**
MCF10A.1	MCF10A.5.mean	MCF10A.1.vqw	MCF10A.1	MCF10A.1
MCF10A.3	MCF10A.3.mean	MCF10A.3.vqw	MCF10A.3	MCF10A.3
MCF10A.4	MCF10A.4.mean	MCF10A.4.vqw	MCF10A.4	MCF10A.4
MDA.MB.231.1	MDA.MB.231.1.mean	MDA.MB.231.1.vqw	MDA.MB.231.1	MDA.MB.231.1
MDA.MB.231.2	MDA.MB.231.2.mean	MDA.MB.231.2.vqw	MDA.MB.231.2	MDA.MB.231.2
MDA.MB.231.3	MDA.MB.231.3.mean	MDA.MB.231.3.vqw	MDA.MB.231.3	MDA.MB.231.3
HCC1937.1	HCC1937.1.mean	HCC1937.1.vqw	HCC1937.1	HCC1937.1
HCC1937.7	HCC1937.5.mean	HCC1937.7.vqw	HCC1937.7	HCC1937.7
HCC1937.8	HCC1937.6.mean	HCC1937.8.vqw	HCC1937.8	HCC1937.8



Supplemental Figure 1: DNA methylation. A clustered heatmap of the top differentially methylated promoters regions are shown for MCF10A, MDA-MB-231, and HCC1937 cell lines. The M-value is displayed where blue indicates hypomethylation and red indicates hypermethylation of promoter regions associated with a particular gene or gene location.

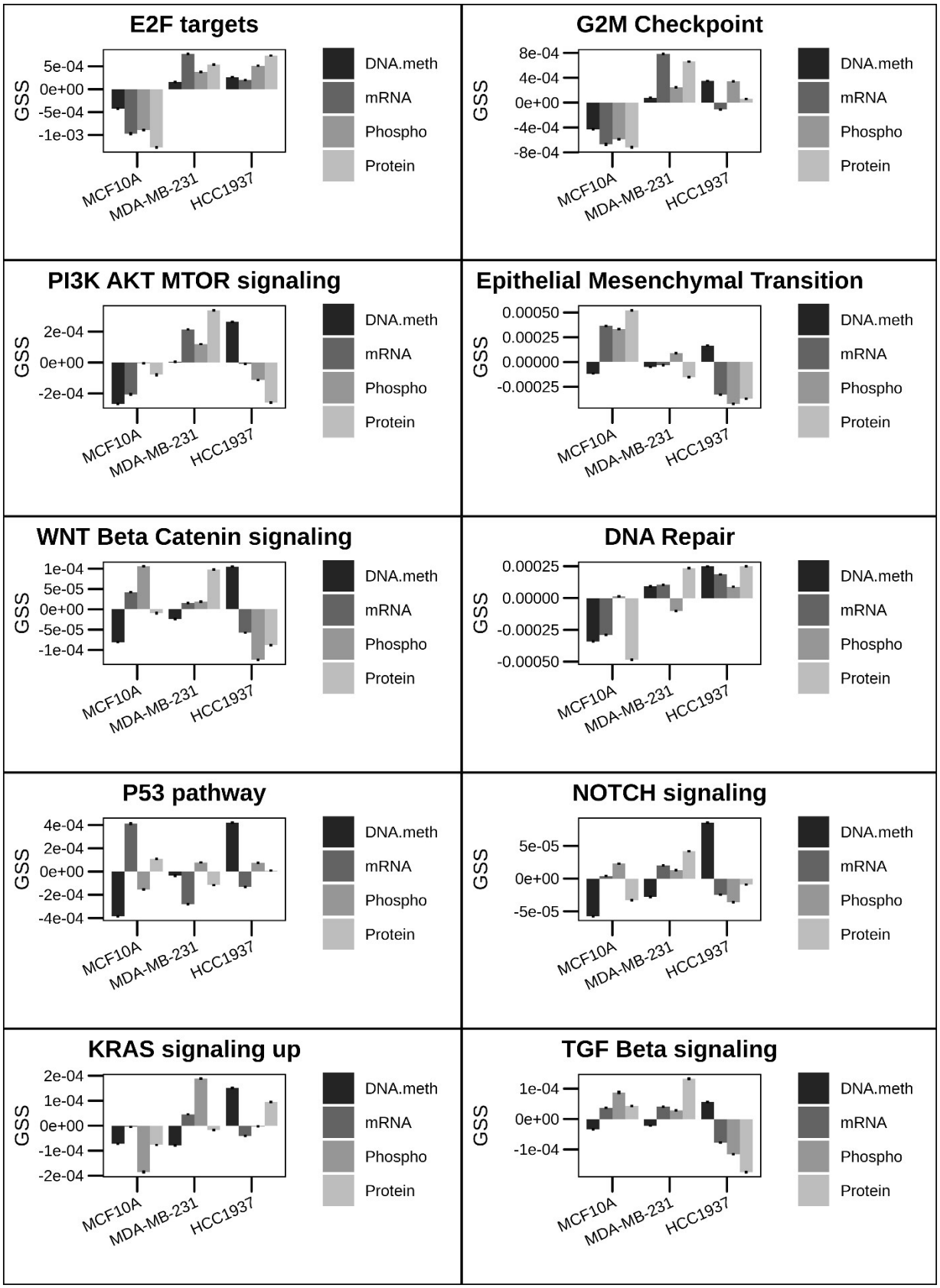


Supplemental Figure 2: Volcano plots of the differential analysis from RNAseg, protein, and phosphorylated peptides analyses. Each omics data set was normalized and analyzed using limma to identify differentially expressed features between MCF10A, MDA-MB-231, and HCC1937 cell lines. The x-axis displays the log2 fold-change and the y-axis shows the FDR adjusted p-value for each comparison. The red positive fold change values correspond to up-regulated features in the first group listed in the figure label. For example, positive features (red dots) are up-regulated in MDA-MB-231 compared to MCF10A and negative features (blue) are down in MDA-MB-231. The significance threshold was FDR-adjusted p-value < 0.05 and absolute fold change > 2.



Supplemental Figure 3: Clustered Image Map for component 2. Represents the multi-omics signature in relation with the samples. The most important features in component two that

distinguish between the three cell lines is shown as a clustered heatmap using the cimDiablo() function provided by MixOmics. Features identified from DNA methylation of promoter regions, RNAseq, protein, and phosphoproteomics data sets are included. The red and blue colors represent positive and negative correlations respectively, whereas grey represents small correlation values.


E2F targets
G2M Checkpoint
PI3K AKT MTOR signaling
Epithelial Mesenchymal Transition
WNT Beta Catenin signaling
DNA Repair
P53 pathway
NOTCH signaling
KRAS signaling up
TGF Beta signaling
GSS
DNA.meth
mRNA
Phospho
Protein
MCF10A
MDA-MB-231
HCC1937

Supplemental Figure 4: Data-wise decomposition of the gene-set score (GSS) for some of the significantly regulated gene-sets in the cell lines. The analyzed was performed using MOGSA. The influence of each multi-omics data set and its influence on the gene-set is shown.