

Multi-Omics Analysis of Human Embryonic Stem Cell Neural Differentiation

Processing SWATH® Acquisition Data in the Cloud with the OneOmics™ Project

Hunter CL¹; Williams KE^{2,3}; Chen H²; Robinson JF²

¹SCIEX, USA; ²Department of Obstetrics, Gynecology & Reproductive Sciences, UCSF, USA; ³Sandler Moore Mass Spectrometry Core Facility, UCSF, USA

A quantitative proteomics study was performed and analyzed using the OneOmics™ project suite of applications. Undifferentiated human embryonic stem cells (hESCs) were compared to their neuronal derivatives, *i.e.*, neuronal progenitor cells (NPCs), to examine common and unique differences at the transcriptional and protein level. hESCs were differentiated into NPCs using a previously developed method^{1,2}, which involves cells growing in suspension and the addition of neural promoting factors. RNA and protein fractions from both cell populations were collected and analyzed.

Quantitative proteomics was performed using SWATH® Acquisition and data were processed using the suite of tools in the OneOmics™ Project in the SCIEX Cloud. Transcriptomic data was analyzed using standard procedures, then both the protein and RNA data was loaded into iPathwayGuide (Adviata) for comparison. For the proteins/genes identified to significantly different between NPCs and hESCs, good correlation in differential expression was observed, especially for the

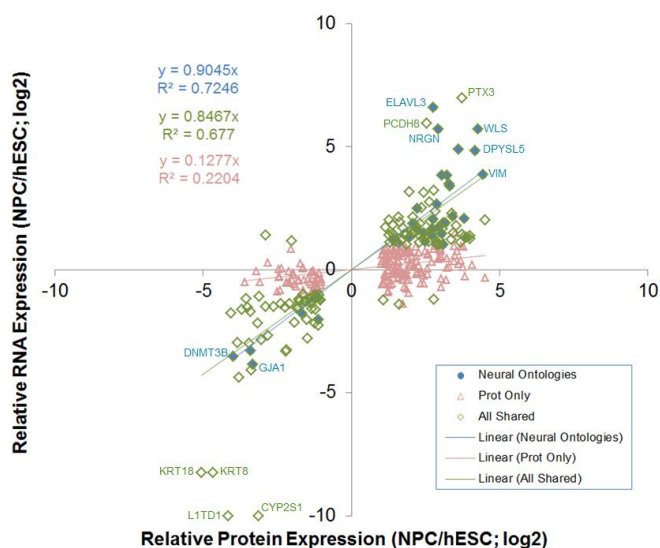


Figure 1. Comparing Protein and RNA Expression Changes between hESCs and NPCs.

After alignment of the proteins and genes, proteins with significant differential expression were plotted vs the observed RNA expression differences. Green diamonds show the protein set (144) where differential RNA expression was also observed (blue fill indicates proteins (37) involved in a neuronal biological process). Good correlation was observed in terms of the magnitude and direction of differential expression. Orange triangles represent the differentially expressed proteins (197) with no significant RNA expression difference, highlighting the importance of studying both protein and RNA expression.

molecules involved in neuronal development-related biological processes (Figure 1).

Key Advances of the OneOmics™ Project

- Comprehensive SWATH® Acquisition datasets can be generated on proteomics samples for protein expression
- Improved depth of coverage obtained using Variable Q1 Window Acquisition³
- Large datasets can be processed quickly in the cloud using the SWATH Acquisition Proteomics Toolkit in SCIEX cloud
- Powerful visuals for assessing MS data quality and understanding protein expression differences are automatically generated
- Ability to compare across or between proteomic and transcriptomic datasets, to identify common and uniquely differentially expressed proteins/RNAs
- Drill into the biology with powerful tools such as iPathwayGuide (Adviata)

Methods

Cell Preparation and RNA/Protein Isolation: Undifferentiated hESCs (UCSF4 line) were cultured in mTeSR Medium (StemCell Technologies). NPCs were derived using a previously established method (Swistowski et al., 2009; Robinson et al., 2016), and grown in Neurobasal Medium (Gibco), containing NEAA, L-Glutamine, penicillin/ streptomycin, B27, FGF2, and LIF. Cells were cultured at 37°C in 5% CO₂ and 8% O₂. Total RNA was obtained using the Qiagen RNeasy Micro Plus RNA Isolation Kit. RNA quality was examined using the Agilent RNA 6000 Nano LabChip Kit and Bioanalyzer 2100 system (RIN > 9). For protein, cells were collected in 1% SDS and 50mM ammonium bicarbonate and quantified using the Pierce BCA Kit. All samples were stored at -80°C before further processing.

Transcriptomic analysis: We isolated RNA from hESC and NPCs and evaluated relative expression using the Affymetrix Human Gene 2.0 ST microarray platform. Hybridization and array scanning was performed at the UCSF Gladstone (NHLBI) Genomics Core Facility. The signal intensity fluorescent images produced during Affymetrix GeneChip hybridizations were read using the Affymetrix Model 3000 Scanner and converted into GeneChip probe results files (CEL) using Command and Expression Console software (Affymetrix). CEL files were RMA normalized and differential gene expression between the two cell populations was determined.

Protein Sample Preparation: Protein isolated from the hESC and NPC cells were reduced in 5 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP), alkylated with 10 mM iodoacetic acid (IAA), trypsinized (Promega, Sequencing Grade), and processed using Pierce Detergent Removal Spin Columns according to manufacturer's specifications. Collections were vacuum-dried and re-suspended at a concentration of 1 µg/µl in 98% HPLC-grade water, 2% HPLC-grade acetonitrile, and 0.1% formic acid for MS analyses. For library generation a pool of all the digested samples was fractionated into 8 fractions using SCX Spin Tips (Protea) according to manufacturer's protocol.

Chromatography: Separations of the tryptic peptides from digested samples and sample fractions were performed on a NanoLC™ 425 System (SCIEX) in trap elute mode, using a 75 µm x 150 mm column and a 0.35 x 0.5 mm trap (both ChromXP™ C18CL, 5 µm, 120 Å phase - SCIEX). A linear gradient of 5-35% over 90 min with a flow rate of 300 nL/min was used and the column was maintained at 35 °C. Mobile phase A was 100% water with 0.1% formic acid. Mobile phase B was 100% acetonitrile with 0.1% formic acid.

Mass Spectrometry: MS analyses were performed using either data dependent acquisition (DDA) or SWATH® Acquisition on a TripleTOF® 6600 System equipped with a NanoSpray® Source (SCIEX). Variable Q1 window SWATH Acquisition methods (100 windows)³ were built in high sensitivity MS/MS mode with Analyst® TF Software 1.7.1.

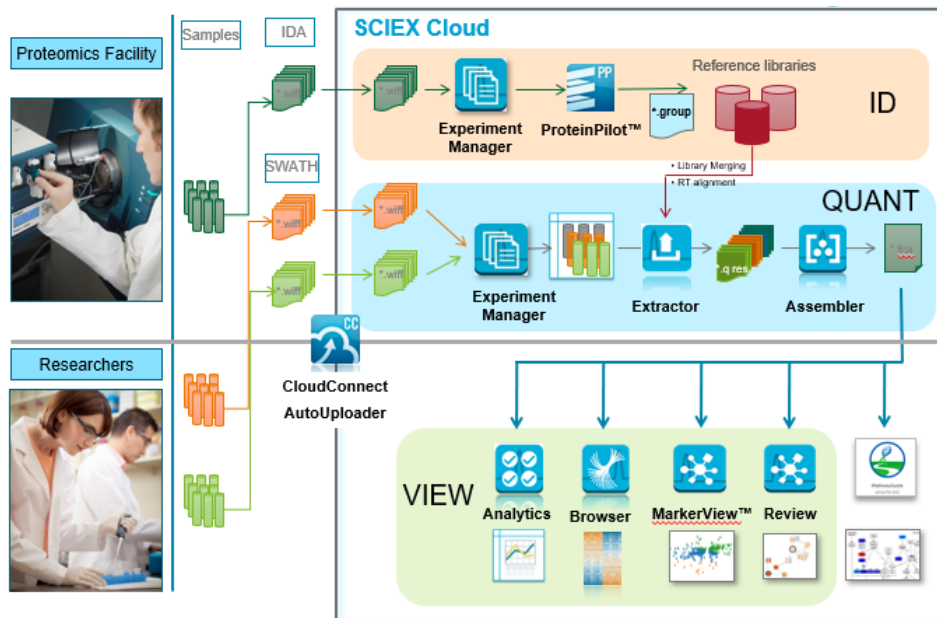


Figure 2. Overview of the Processing and Visualization Pipeline. Samples analyzed by SWATH® acquisition are uploaded into SCIEX cloud then the metadata describing the study is specified in Experiment manager. The Extractor uses a spectral ion library to extract the peptide information from the data, followed by peak group scoring and FDR analysis. Next, data normalization and protein fold changes are computed in the Assembler application. Then data can be visualized with a variety of applications to assess MS data quality (Analytics), view protein expression differences (Browser), find trends in the data (MarkerView™ software), and assess the biological significance of the results (Review and iPathwayGuide (Advaita)).

Data Processing: DDA data was processed with ProteinPilot™ Software and the group file was used as the spectral ion library. Library and SWATH acquisition data were uploaded to the SCIEX cloud and data were processed using OneOmics™ project tools (Figure 2). SWATH acquisition data extraction was followed by most likely ratio (MLR) normalization and fold change (FC) calculations². Protein quantitation data was compared to transcriptomic data using comparison tools in OneOmics and iPathwayGuide (Advaita).

Quality Assessment of SWATH® Acquisition Data using Analytics Application

Shown in Figure 3 are a few of the many data tools provided for assessing MS data quality provided in the Analytics application in Workspaces in the OneOmics™ Project. Good discrimination was observed in the false discovery rate (FDR) analysis, shown by plotting the peak group score distributions of the forward and decoy peptides. The reproducibility between the technical/biological replicates in this experiment centered around 10% CV as seen in the %CV vs frequency plots. The MLR normalization is performed during the Assembler data processing, and the results samples showed good alignment after normalization, as seen by viewing the alignment of the ratio distributions before and after normalization. Many more figures are provided to allow the MS operator to confirm the quality of the SWATH® Acquisition data.

Visualizing the Protein Expression Changes

Using a library generated from a pool of NPC and hESC cells, 2278 proteins were reliably quantified using SWATH® Acquisition across the 2 sample types (3 biological replicates of each). The Browser application in Workspaces can be used to explore the protein expression data; after filtering, 280 proteins were differentially expressed (2 or more peptides per protein, protein fold change confidence > 70%). Figure 4 (right) shows the heat map for these 280 proteins. Example data for an up-regulated protein (Dihydropyrimidinase-related protein 5, DPYL5) in the NPC cells is shown (top left); 8 peptides were measured for this protein. DNA methyltransferase 3 beta (DNMT3B) is down

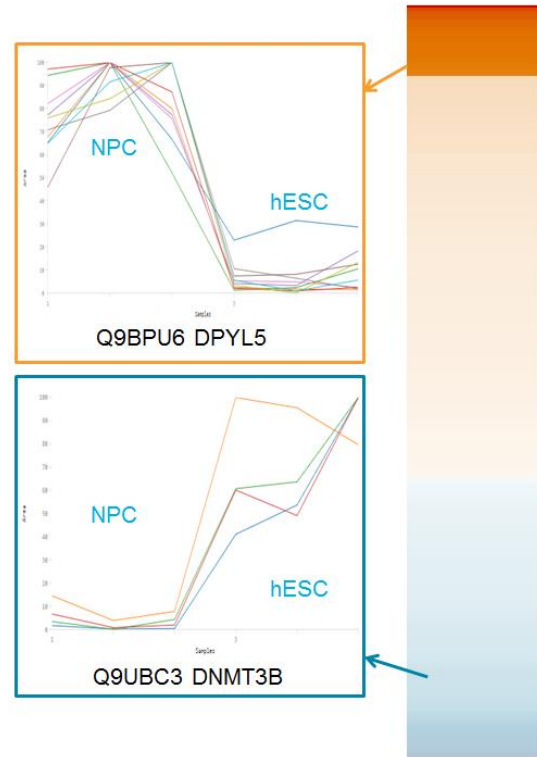


Figure 4. Protein Expression Differences in Browser. In the browser application, one can view the protein expression data as a heat map, here comparing the protein expression in NPC vs the hESC cells. 280 proteins are significantly differential expression with 2 or more peptides (filtered at 70% fold change confidence). An example of an up-regulated protein (orange) is shown for Dihydropyrimidinase-related protein 5 (8 peptides quantified). An example of a down-regulated protein (blue) is DNA methyltransferase 3 beta.

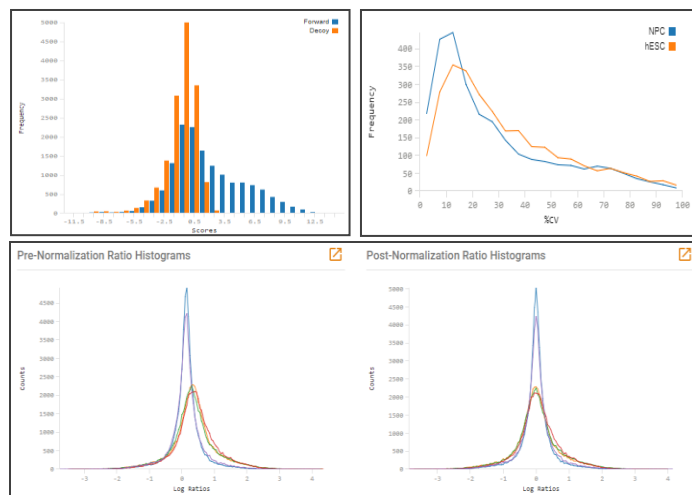


Figure 3. Quality Assessment of SWATH® Acquisition Data using Analytics. Top left shows the output from the FDR analysis, plot of the score distributions of the forward and decoy peptides, showing good differentiation between the two. Top right shows the CV vs frequency plot, between the 3 biological replicates of each experimental group, hESC (orange) and NPC (blue). Bottom shows the improvement in sample alignment after MLR normalization in the ratio histogram plots.

regulated, with 4 peptides quantified. Both proteins are believed to be involved in neuronal development.

Once proteins of interest are identified, more information can be easily obtained on each. Information from UniProt is pulled into the Browser session, including protein sequence and ontology information. As an example, another up-regulated protein (Cadherin 2) was found to be upregulated in NPCs in both the protein and RNA data. This protein is involved in neurogenesis, playing a role in development of the nervous system and formation of cartilage and bone. It was found that 3 peptides were quantified for this protein and they spanned both the cytoplasmic and extracellular domains (Figure 5). When available, information on potential post-translational modification sites and many other sequence features are displayed in this view.

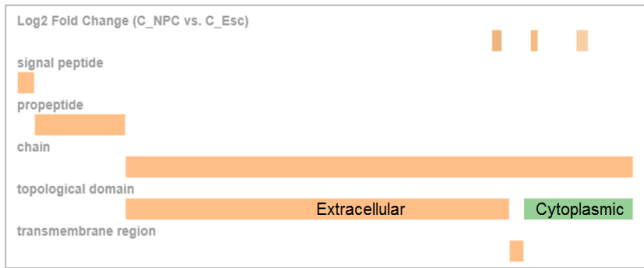


Figure 5. Aligning Peptide Information with Protein Sequence. Cadherin 2 was found to be upregulated in NPCs in both the protein and RNA data. Three peptides were quantified for this protein, bridging both the cytoplasmic and extracellular portion of this transmembrane protein. Protein sequence information is pulled from UniProt.

Finding Trends in the Protein Expression Data

The protein expression data can also be mined using multivariate statistics using the MarkerView™ Software in Workspaces. Here, sets of proteins showing large or small, up- or down-regulation were found, as seen on the Loadings plot (Figure 6, top) from the PCA-PCVG analysis. The orange PCVG group contains a set of proteins showing larger up-regulation in the NPC cells vs the hESC cells. Protein area data for six proteins from this group is shown (Figure 6, bottom left). Conversely, in the brown group, there is a set of proteins showing a small decrease in expression in NPC vs hESC, and 5 proteins from this group are shown (bottom right).

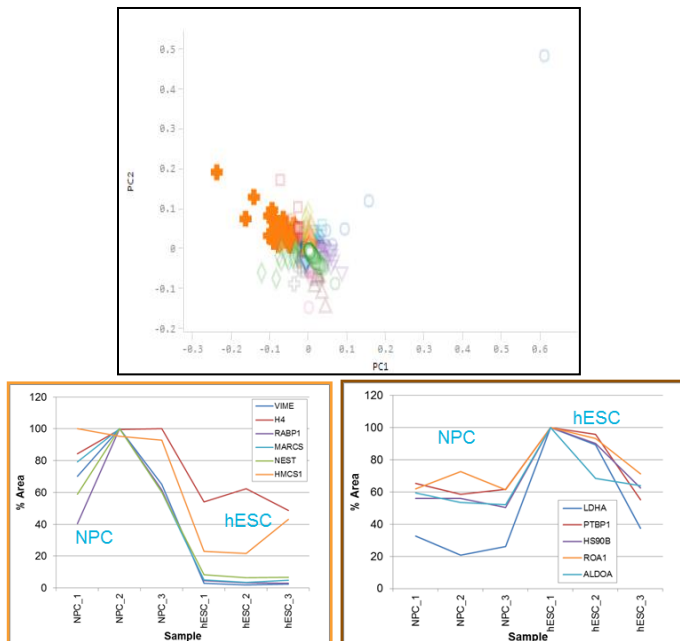


Figure 6. Trend Analysis with PCA. Multivariate analysis can also be used in the MarkerView™ Software App to find groups of proteins that show the same expression patterns, using PCA-PCVG and K-means clustering. Top is the loadings plot with a variety of colored groups from PCA-PCVG analysis. The orange group shows a set of proteins that are higher in expression in NPC. The brown group contains a set of proteins that show a very small decrease in expression in NPC.

Understanding the Biology

To begin to dig into the underlying biology, the perturbed ontologies can be visualized using the ontology wheels in the Browser application (Figure 7). The ontologies are first retrieved for each protein from UniProt. Data is filtered using protein fold change confidence and the number of proteins per ontology. In comparing the differentially expressed proteins between hESC and NPC, it was observed that the nervous system development ontology was strongly perturbed in this contrast between NPC vs hESC. There were 13 proteins measured in this particular ontology and 12 of these proteins were upregulated in NPC cells. This observation is expected as these cells are actively differentiated into neuronal cells from stem cells.

Comparing Protein and RNA Expression Data

The comparison of the protein expression results to the previously obtained RNA express data was performed using iPathwayGuide (Advaita). The protein and RNA data showing significant expression changes were loaded and aligned. After alignment, it was found that 144 proteins/RNA were significant in both datasets (Figure 8, top). Of these 144, there were 37 that mapped to biological processes involving neuronal development (Figure 8, bottom). Very good correlation of expression was seen

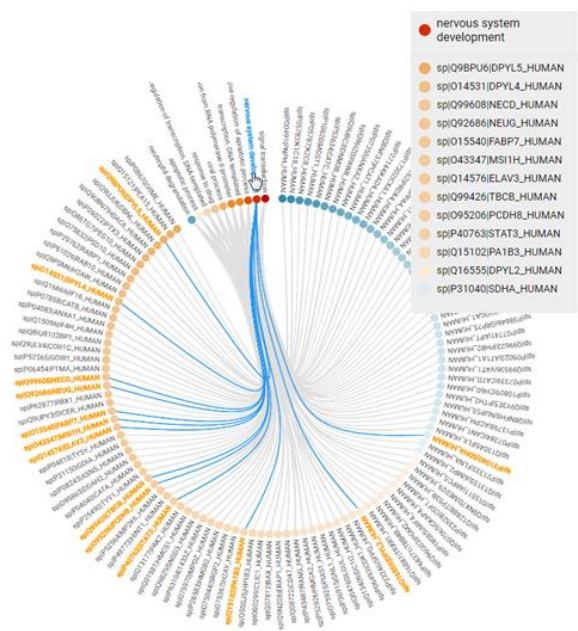


Figure 7. Perturbed Ontologies Viewer in Browser. The differences between 2 experimental groups are analyzed to determine which ontologies are perturbed. When assessing NPC vs hESC cells, the nervous system development ontology was strongly enriched as expected. There were 13 proteins measured in this ontology and 12 of these proteins were upregulated in NPC cells relative to hESC cells.

between these proteins and genes.

There were an additional 197 proteins that had significant differential expression at the protein level but not at the RNA level; 26 of these also mapped to neuronal processes (Figure 8 bottom). This data is also plotted in Figure 1, which further highlights the good correlation observed between the protein and RNA data.

Conclusions

A project was performed to evaluate the changes observed at the RNA and protein level, between human embryonic stem cells and the differentiated neural progenitor cells. The cloud based tools for multi-omics analysis in the OneOmics™ Project were used to explore the dataset.

- Identified significant differential expression of 280 proteins with 2 or more peptides from SWATH® Acquisition data
 - Filtered at 70% fold change confidence
- Significant correlation between changes in protein and RNA levels was found between NPCs and hESCs
 - Many of these involved in neurogenesis
- 26 proteins found to be differentially expressed in the protein data that also mapped to neuronal development processes that were not found to be differentially expressed in the RNA data, highlighting the importance of measuring changes at the protein level.

References

1. Swistowski *et al.*, (2009) *PLoS One* **4(7)**, e6233.
2. Robinson *et al.*, (2016) *Reprod Toxicology* **60**, 1-10.
3. Improved Data Quality Using Variable Q1 Window Widths in SWATH® Acquisition, SCIEX Technical Note, RUO-MKT-02-2879-B.

AB Sciex is doing business as SCIEX.

© 2019 AB Sciex. For Research Use Only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Document number: RUO-MKT-02-7115-B

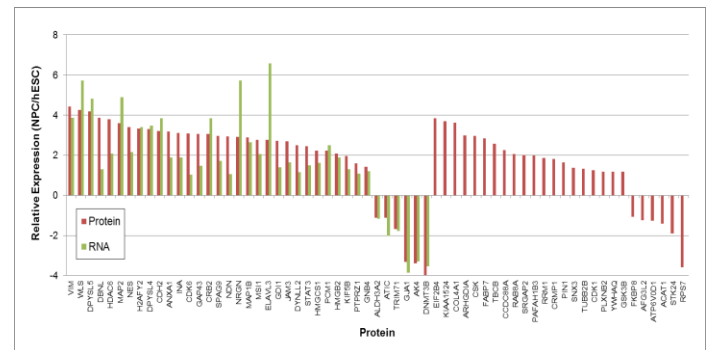
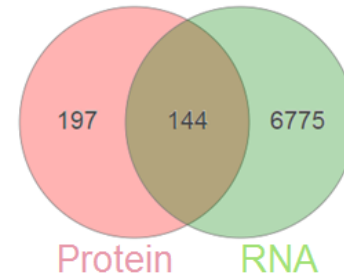


Figure 8. Correlation between the RNA and Protein Expression Data. The protein and RNA data was uploaded for analysis in iPathwayGuide (Advaita), then proteins and genes are aligned. Data was pre-filtered using a fold change of >2 for both protein and RNA, then a p-value of <0.0001 was used for the RNA data and a p-value of <0.01 was used for the protein level. (Top) As shown in the Venn diagram, 144 proteins/genes were found to be differential in both the RNA and protein data. There were 197 proteins that were found to be differential in the proteomics but not in the RNA data, highlighting the importance of using both techniques. (Bottom) Of these differential proteins, a subset of 63 proteins which were related to the enriched neuronal biological processes were plotted. Very good correlation was observed between the protein and the RNA data (37 proteins where there was matched RNA data).