

Global proteome profiling of CRISPR/Cas9 induced insertions and deletions

Featuring SWATH® Acquisition on a TripleTOF® 6600+ LC-MS/MS System

Michal Lubas¹, Thomas Eriksen¹, Malene Ambjørn¹, Eric Paul Bennett^{2,3}, Ignacio Ortea³, Jens-Ole Bock³, Ferran Sanchez⁴, Antonio Serna-Sanz⁴, Kerstin Pohl⁵

¹H. Lundbeck A/S, Denmark; ²Copenhagen Center for Glycomics, Department of Odontology, Faculty of Health Sciences, University of Copenhagen, Denmark; ³COBO Technologies, Denmark; ⁴SCIEX, Spain; ⁵SCIEX, US

Recent advances in gene editing technologies have revolutionized biomedical research. Some technologies enable the introduction of specific single nucleotide changes, insertions and deletion (InDels) in genomes of cells, tissues and whole organisms of any species opening the door for treatment of a variety of genetic-based diseases.¹ In this regard, CRISPR/Cas9 has been widely adopted as the first choice of gene editing modality in the field, due to its simple mode of action and great flexibility in use.² With CRISPR/Cas9 moving from a research tool into adaption as personalized medicine, safety concerns have to be addressed: an understanding of the effects on the targeted gene, but also potential off-target effects is desired. Analytical methods such as polymerase chain reaction (PCR)-based methods are widely used. However, the verification on gene level, does not reveal direct information on the protein expression. Immunoblotting techniques can give further insights, but the approaches are limited to the availability of antibodies and furthermore only provide information on the targeted proteins. Mass spectrometry in comparison allows for unbiased information on all detectable proteins addressing the need for understanding the overall impact of gene therapies. Still,

challenges such as limited sample amount, sensitivity, specificity and last but not least reproducibility for simultaneous relative quantification on all proteins in a sample need to be overcome. Here, an MS-based strategy is described addressing these additional challenges.

Key features of SWATH Acquisition for CRISPR proteome profiling

- Unbiased MS method without the need for antibody production or extensive method development
- Quantitative analysis with extremely high sensitivity and specificity at the MS/MS level for highly accurate and reproducible results
- Quantification of the target product, but also monitoring of changes in the total proteome, thus exploring the overall impact of gene editing on the protein expression, serving a need to ensure safe therapies

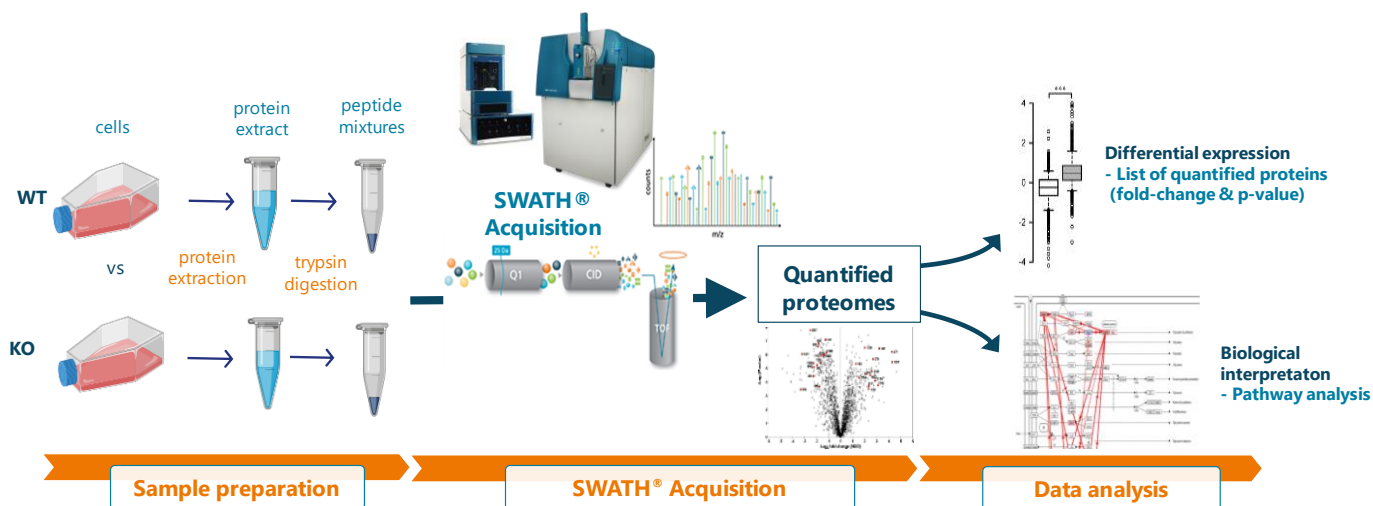


Figure 1: Overview of the proteome expression analysis platform (PIPPR). Proteins of wildtype (WT) and treated cells, e.g. knock-out (KO) cells are extracted and digested followed by LC-MS/MS analysis using SWATH Acquisition. Differential expression are evaluated based on fold-change and p-value, and affected pathways are analyzed to enable biological interpretation of results.

- Functionality-based bioinformatics post-acquisition enables interpretation of large-scale proteomics data providing biologically meaningful results

Methods

CRISPR/Cas9 targeting and validation: The human gene OTUB1 coding for ubiquitin thioesterase OTUB1 was targeted using guiding ribonucleic acid (gRNA) sequences: OTUB1-gRNA1 (5'-CGATAGAAACAGTTGCCGTC-3') or OTUB1-gRNA3 (5'-GACGGCAACTGTTTCTATC-3') (ThermoFisher, TrueGuide). Both gRNAs were functionally validated by IDAA using the CRISPR InDel Profiling Platform (CIPP)³ based on previously described protocols⁴. In brief, 2 µg Cas9 protein (Thermo Fisher Scientific) and 0.5 µg gRNAs were complexed, and electroporated into 1e⁵ HEK293 cells in a 24-well format (for Western blot analysis) or 1e⁴ cells for 96 format (for IDAA analysis). Five days post transfection, cells were harvested by trypsinization and washed with phosphate buffered saline (PBS), spun down at 1000xg, lysed in 50 µL COBO extraction buffer and analyzed by IDAA. InDel detection and profiling by IDAA is based on a PCR using three primers (tri-primer principle): two gene specific primers; in this case specific for OTUB1 A (FAM-seq-5'-TCCTTAAGTGCCCAGCTTCC-3') and B (5'-TGCAACTCCTTGCTGTCATC-3'), that span across the OTUB1 target site, and a third universal fluorescein-labeled primer (FAM-5'-AGCTGACCGGCAGCAAATTG-3') that is specific for an extension present on one of the gene specific primers (A-primer). Amplicons were size-discriminated by capillary electrophoretic (CE) fragment analysis using a standard DNA-sequencing instrument (Thermo Fisher Scientific). The InDel profiles were generated from the raw data files using ProfileIt-V2 software available from COBO Technologies.⁵

Sample preparation for MS analysis: A HEK293 knock-out cell line and a non-targeted control cell line were used for this study. Samples were prepared as follows: cells were collected in ice-cold PBS and stored as pellets at -80°C. Cell lysis was performed by incubation for 10 min in boiling lysis buffer (6 M guanidine chloride (Gua-HCl), 5 mM tris(2-carboxyethyl)phosphine, 10 mM iodoacetic acid and 100 mM tris(hydroxymethyl)aminomethane (Tris) pH. 8.5) followed by a sonication step. Samples were diluted to 2M Gua-HCl and LysC protein digest performed using 1:200 (w:w) enzyme:protein ratio for 1 h at room temperature. Samples were further diluted to 0.6M Gua-HCl with 25 mM Tris, pH 8.5 and a second digestion step was carried out overnight with trypsin (1:20 (w:w)). Protease activity was quenched by acidification with trifluoroacetic acid to a final concentration of approximately 1%, and the resulting peptide mixture was desalted using StageTips (100 µL, Pierce).

Chromatography: 2 µg of the tryptic digested sample were separated by reversed-phase chromatography using a SCIEX NanoLC™ 425 HPLC System in low microflow configuration (5 µL/min) equipped with a YMC Triart column (150x0.3mm, particle size 3 µm, pore size 12 nm). Mobile phase A consisted of water with 0.1% formic acid, while mobile phase B consisted of acetonitrile with 0.1% formic acid. A gradient from 8% to 30% mobile phase B within 30 min and 40°C column oven temperature was applied. The column was washed during 5 min at 95% B and equilibrated during another extra 5 min at 5% B before the next injection. Same separation and injection conditions were used for library generation via information-dependent acquisition (IDA) runs as well as data-independent acquisition (SWATH Acquisition).

Mass spectrometry: For library generation and sample testing, data were acquired in replicates on the TripleTOF 6600+ System using an IDA method monitoring the 100 most intense peaks within a total cycle time of 2.5 s. SWATH Acquisition data were acquired using a method with 100 variable windows and 25 ms accumulation time for each window.

MS data processing: For ion library generation, a combined search of six IDA replicates per sample was done using ProteinPilot™ Software 5.0, resulting in 3556 protein groups with an FDR cut off of 1%. SWATH Acquisition data were subsequently processed using OneOmics™ Project in SCIEX Cloud and on a local workstation using the SWATH Acquisition MicroApp 2.0 within PeakView® Software 2.2 and MarkerView™ Software 1.3 deploying the previously generated ion library.

Analytical tools for CRISPR/Cas9

All current gene editing technologies share the common principle of inducing a DNA double strand break (DSB) at a user-specified site in the genome, followed by cellular repair of the induced DSB. In mammalian cells, the non-homologous end joining repair pathway dominates the cellular repair and, as a consequence, the primary outcome after gene editing is the formation of InDels at the targeted site. When the target site is chosen to be positioned in the protein encoded gene sequence (exon), these gene editing induced InDels can result in disruption of the protein encoded reading frame (out-of-frame causing InDels), which leads to abrogation of gene function. In these so-called knock-out (KO) gene editing experiments, the primary objective is to induce a high frequency of these out-of-frame causing InDels, that can be detected by genetic analysis of cells after gene editing.

Among the available InDel detection methodologies, InDel Detection by Amplicon Analysis (IDAA) has shown great promise as a fast, robust and cost efficient method, with InDel detection

sensitivity being comparable to next generation sequencing.^{6,7} However, InDel detection and profiling will only provide genetic information of the induced gene editing events at a specific genomic site. When working with cell populations, tissues or whole organisms, it does not reveal what consequences it might inflict at the protein level for the gene of interest, or globally at the proteome level.

Alternatively, immunoblotting assays can be applied for this purpose, but the use of antibodies has several drawbacks: in many cases there are no antibodies available for the target protein. Even in the cases where antibodies are available, they are limited to given epitopes and can suffer from low specificity because of cross-reactivity, poor quality or batch-to-batch variability. As a consequence of gene editing, non-desired, off-target changes can occur, which are difficult to be detected and measured reliably, as they can be located throughout the genome. Since they are not known at the outset, targeted methods such as specific antibodies cannot be used to detect such changes.

In comparison, mass spectrometry (MS)-based techniques offer high resolution and specificity, thus, it can be used to confirm that the gene editing resulted in the desired effect at the protein level in an unbiased way. In addition to measuring the expression of protein product of the edited gene, MS experiments can provide information on the global proteome expression. This allows for detection of off-target editing events and mapping of differentially regulated proteins in one effort. However, a few challenges need to be overcome. In order to provide a full picture of the changes in a given heterogenous sample, all analytes need to be detected with high sensitivity, specificity and reproducibility. Furthermore, quantitative information on all of the analytes—from high abundance to very low abundance—in a complex matrix is desired.

Verification of CRISPR/Cas9 target by PCR using CIPP

OTUB1 has recently been reported to impact the clearance of aggregated Tau protein.⁸ Thus, there is interest to better

describe the function and cellular processes in which OTUB1 may play a role in connection with Alzheimer's disease pathology.

The target validation of OTUB1 by CRISPR/Cas9 was performed using IDAA. IDAA is based on a PCR with a tri-primer principle: two OTUB1 gene specific primers A and B that span across the OTUB1 target site, and a third universal, labeled primer that is specific for an extension present on one of the gene specific primers (A-primer). The tri-primer methodology ensures homogeneous labeling of amplicons that are subsequently size-discriminated and detected by CE analysis. For the cells targeted with gRNA1, gRNA3, the wild type (WT) and the non-transfected (NT) cells, the amplification of the target gene OTUB was proven to be successful using the tri-primer principle (Figure 2B). The subsequent CE analysis revealed InDels of various lengths in the gRNA1 and gRNA3 targeted cells (Figure 2C), while the amount of the expected length of the OTUB gene product (yellow peak in Figure 2C) was significantly decreased compared to the WT and NT cells. Since a cell population can contain a mixture of cells which were successfully treated with the CRISPR/Cas9 approach and such cells which weren't, there can be still evidence of the gene of interest (yellow peak in Figure 2C for gRNA1 and gRNA3). The WT and NT cells did not show any additional peaks, as expected. Overall, these results prove the successful CRISPR/Cas9 targeting on gene level.

InDels leading to disruption of the protein encoded reading frame (out-of-frame InDels) were indicated in blue (Figure 2C) and the total percentage compared to the wildtype was calculated (Table 1). InDels which didn't lead to a disruption of the reading frame were indicated in white (Figure 2C). The sum of both types of InDels was used for the quantification of total InDels for both targets (gRNA1 and gRNA3, Table 1): The gene inactivation efficiency in the cell pool of 5×10^6 cells was determined by CIPP with overall InDel formation efficiencies of >91% and importantly, with out-of-frame causing InDel efficiencies >80% (Table 1). Out-of-frame InDels are desired as they usually lead to the abrogation of gene function, whereas in-frame InDels might not.

Table 1: OTUB1 gene editing induced InDel profiles for samples targeted with gRNA1 and gRNA3. Percentages calculated based on peak area of CE analysis.

Sample	Total InDels* [%]	Out-of-frame InDels* [%]	InDels in size (percentage)^
<i>gRNA1</i>	91.0	80.1	1 (50.4%) -1 (8.6%) -8 (6.0%) -6 (5.6%) -10 (3.9%) -7 (2.1%) -9 (1.8%) -2 (1.5%) -3 (1.5%) -4 (1.4%)
<i>gRNA3</i>	94.8	85.8	1 (58.5%) -4 (6.9%) -12 (4.5%) -8 (3.7%) -16 (3.4%) -2 (3.2%) -10 (1.9%) -5 (1.8%) -9 (1.7%) -1 (1.4%)

*calculated including wildtype and unmodified alleles

^10 most abundant InDels with % detected peak area indicated in parenthesis

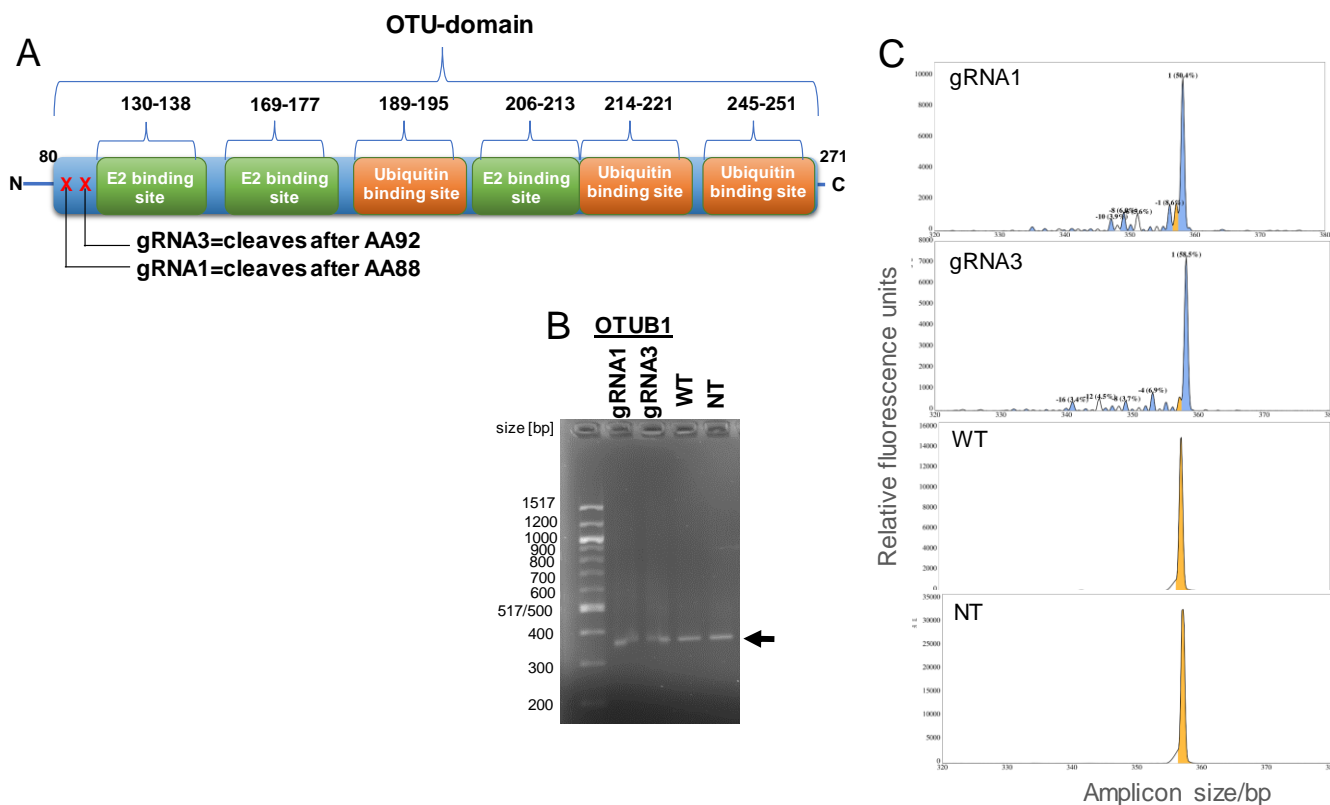


Figure 2. CIPP analysis after CRISPR/Cas9 OTUB1 gene targeting of gRNA1 and gRNA3 in human HEK293 cells. A: Illustration of OTU domain structure with indication of the regions targeted (gRNA1 and 3, indicated with red crosses). B: Gel electrophoresis results of tri-primer amplification of OTUB1 target locus after Cas9/OTUB1-gRNA1 or -3 editing showing a single specific fluorophore-labelled amplicon in HEK293 edited samples (indicated by arrow); gRNA1, gRNA2, wild type (WT) and non-transfected (NT) cells. C: CIPP functional validation and profiling of gRNA1 and -3 InDel formation potential, showing >90% efficiency and no InDel formation in WT nor NT cells. Peak profiles were generated by IDAA fragment analysis and ProfileIt-V2 profiling displaying unmodified alleles in yellow, out-of-frame InDel possessing alleles in blue and in-frame InDel alleles in white. Insertions and deletions are displayed to the right and left of the unmodified yellow peak, respectively. Top five observed InDels are indicated above peaks and calculated total frequencies shown in Table 1.

Evaluation of expressional changes

CRISPR technology was used to create an OTUB1 gene KO HEK293 cell line and SWATH Acquisition was applied to assess the impact of gene-inactivation on the overall changes in the cellular proteome in a quantitative manner.

The high CRISPR/Cas9 OTUB1 knock-out efficiency determined by CIPP was confirmed at the protein level by western blotting and MS (Figure 3 and 4): both protein detection techniques were able to detect OTUB1 protein in unmodified cells and confirmed its knock-out in CRISPR/Cas9-modified cells, which demonstrates the applicability of SWATH Acquisition is an alternative protein detection methodology.

In particular, in experimental settings where antibodies specific towards the protein of interest are not available, alternative methodologies are needed. Additionally, MS analysis is unbiased and provides access to the entire proteome, unlike antibodies targeting specific epitopes. Still, the quantification of analytes via

MS poses challenges especially in complex matrices such as cells or tissues since the MS information can suffer from interferences leading to wrong results. Hence, quantification on MS/MS level is favorable as it provides additional specificity. However, for understanding the overall impact of CRISPR/Cas9 technology on the proteome, the target analytes cannot be predefined which limits the usage of standard quantification methods such as multiple reaction monitoring (MRM). Since the data independent approach of SWATH Acquisition provides MS and MS/MS data of the entire mass range in each cycle in a manner compatible with any LC time scale, fragment information can be used for quantification of any detectable analyte without the need to define the target analyte upfront. The spectral, quantification-ready information is stored as a digital fingerprint and proteins of interest for quantification can be defined after acquisition.

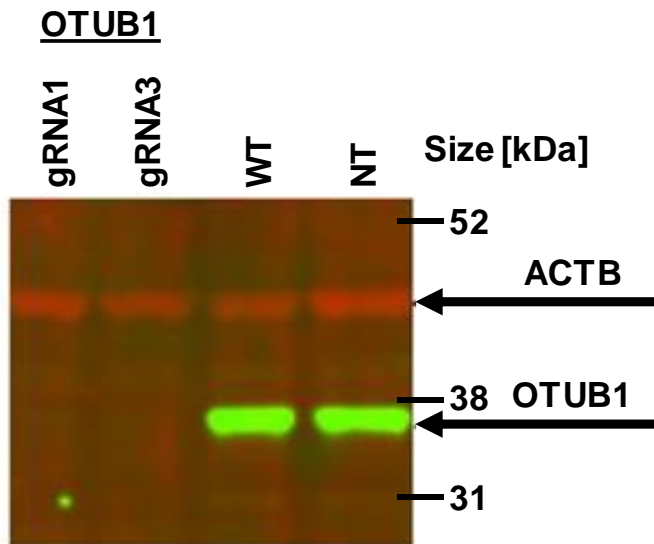


Figure 3: Western blot analysis. Actin (ACTB) serves as positive control, being present in the gRNA1 and gRNA3 targeted cells, the wild type (WT) and the non-transfected (NT) cells. The OTUB1 protein was only detected in the WT and NT cells, indicating a successful knockout in the gRNA1 and gRNA targeted cells.

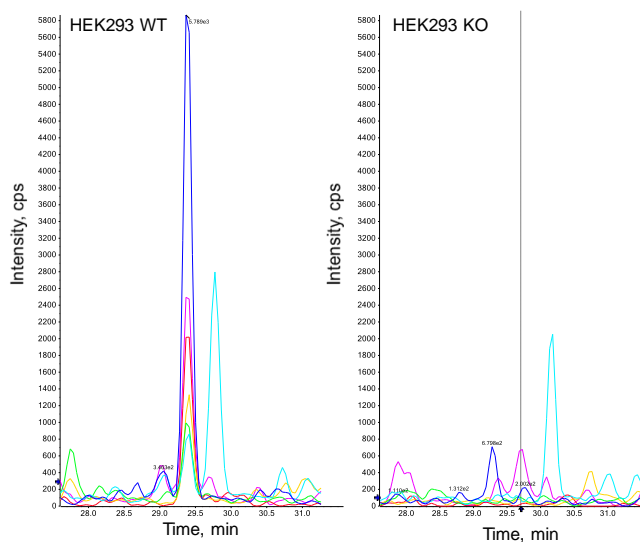


Figure 4: Extracted ion chromatograms (XIC) for fragments of peptide derived from the OTUB1 protein. Example showing the XICs of fragments of a peptide (AFGFSLHLEALLDDSK) in the wildtype HEK293 cells (WT) and the knockout HEK293 cells (KO). The KO cell line does not show clear evidence of the specific fragments, indicating a successful knockout compared to the WT.

The data-independent SWATH Acquisition was performed in triplicate resulting in 16550 identified peptides with a CV value of 20% (cut-off applied for quantification, Figure 5). A total of 2819 different protein groups could be quantified based on these peptides (Figure 5).

Principal component analysis (PCA) in MarkerView™ Software based on all MS signals of replicates of control and the OTUB1 targeted KO cells revealed that first two components explained 85% of the total variance (Figure 6). An excellent separation of

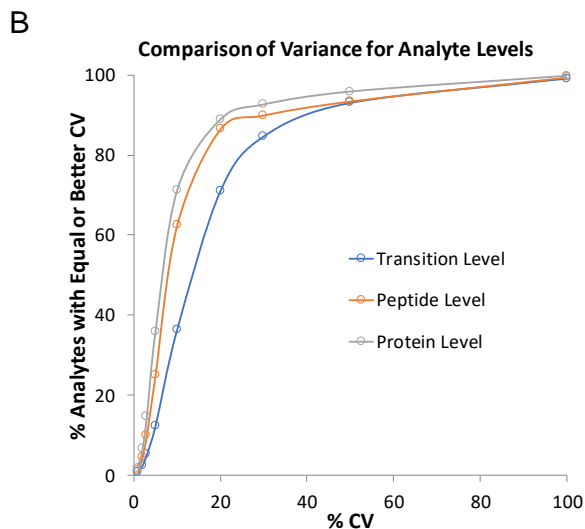
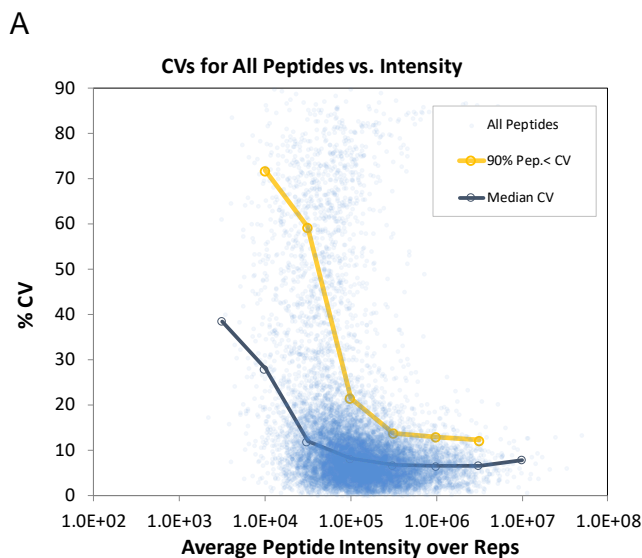


Figure 5: Results of SWATH Acquisition. A: %CV dependence on peptide intensity ($n = 3$). Peptides below yellow curve with a %CV of 20% or less represent 90% of all peptides. These peptides will be taken into account for quantification. B: Percentage of analytes versus their %CV for fragment (transition) level, peptide and protein level ($n = 3$). More than 88% of all detected proteins showed a %CV lower or equal to 20% and were used for quantification.

Scores for PC1 (77.8 %) versus PC2 (7.2%), Pareto

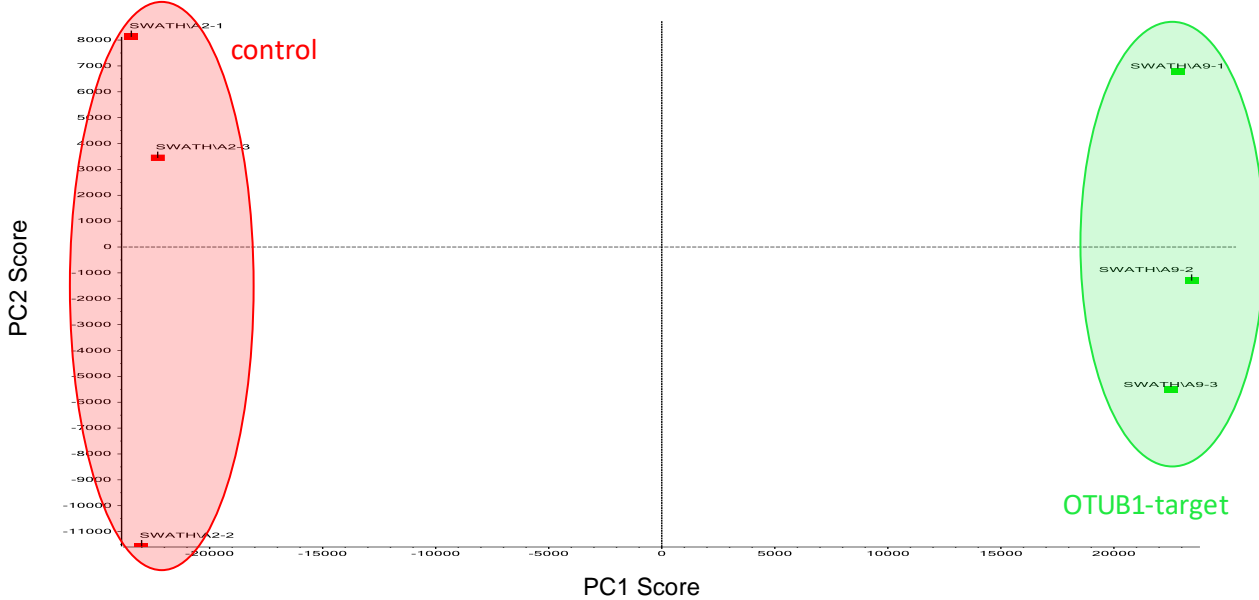


Figure 6: Principle component analysis in MarkerView Software. Scores plot showing 85% of the total data variance (77.8% for PC1 and 7.2% for PC2). Separation based on PC1 was observed for the control and OTUB1 targeted KO cells.

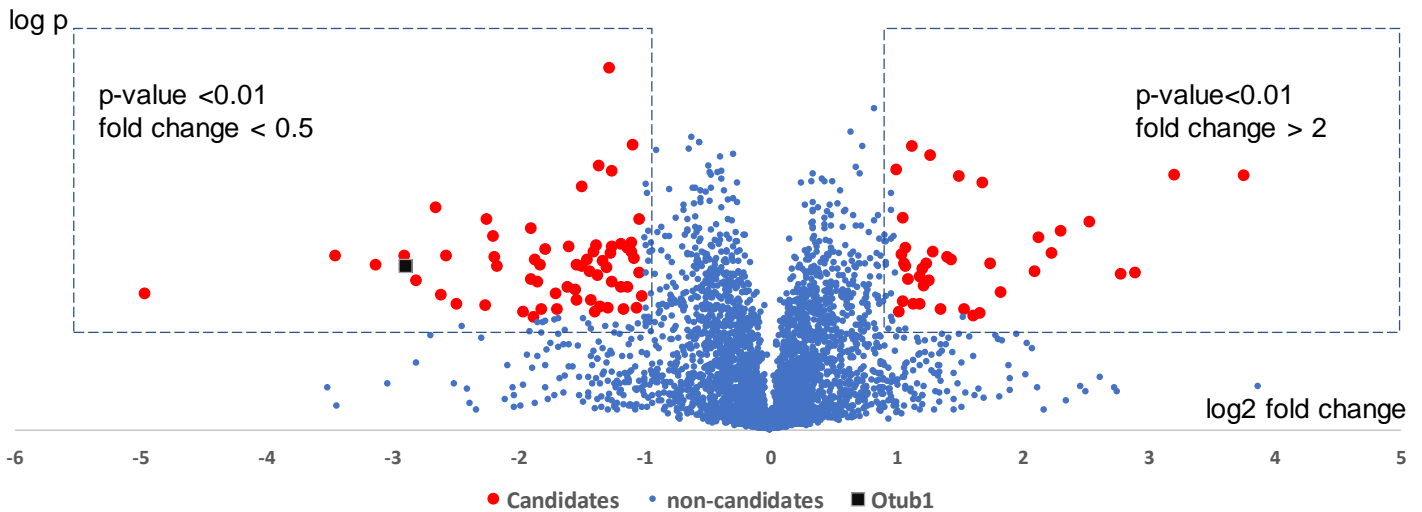


Figure 7: Impact of CRISPR OTUB1-knock-out on proteome level. Volcano plot: representation of binary logarithm of fold changes against decimal log of p-values. Candidates (red) were selected based on lower than 0.5 or higher than 2 fold change ($\log_2 < -1$ or $\log_2 > 1$) with p values better than 0.01. A total of 100 candidates passed this selection filter. Otub1 protein indicated as black square showing a down-regulation.

the control and the KO cells for the first component was observed (Figure 6).

In total, 100 proteins showed changes when being compared to controls: 62 were down-regulated and 38 up-regulated (Figure 7) when applying the following criteria: a fold change lower than 0.5 or higher than 2 and a p-value lower or equal than 0.01. The subsequent pathway analysis of the deregulated proteins, using the Pathway Browser Reactome (access via OneOmics Project),

revealed that affected pathways were mainly related to gene expression, cell cycle progression and DNA repair (Figure 8). This information can be used for in depth analysis of each pathway increasing the understanding of OTUB1's cellular functions.

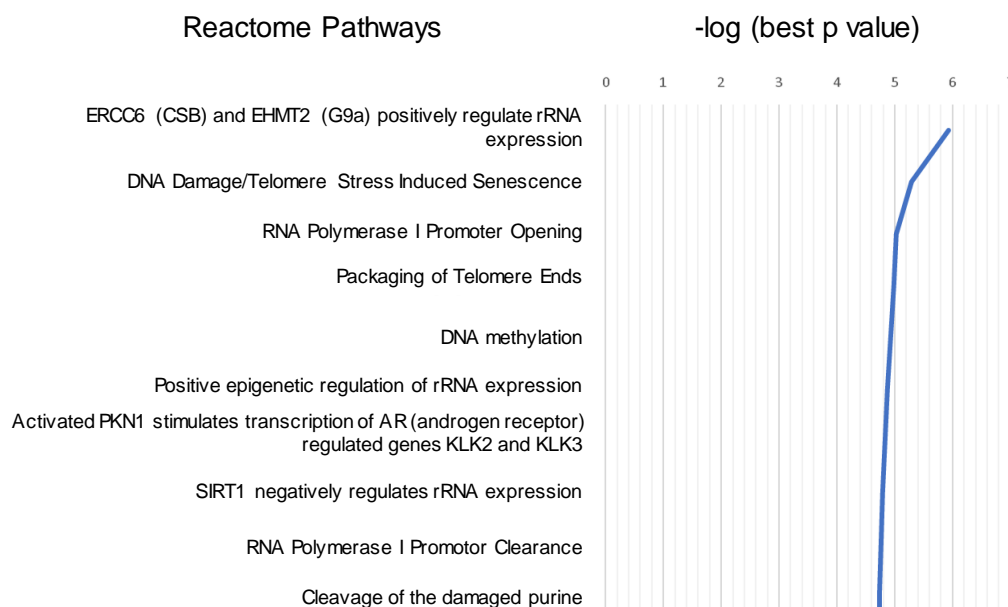


Figure 8: Impact of CRISPR OTUB1-knock-out on Reactome pathways. Ten most affected pathways based on candidate selection as indicated in the Pathway Browser Reactome. Data was filtered using a p-value equal or better than 0.001.

Conclusion

- The successful knockout of the OTUB1 gene in HEK293 cells was confirmed by SWATH Acquisition, in alignment with other protein detection techniques (Western blotting) and genomic-based techniques (PCR-based IDAA)
- The accurate, rapid and robust method presented is not limited to measuring protein expression of a target gene, but can also detect non-desired, off-target events, which may occur across the entire genome, within the same single method
- The functionality-based post-analysis of protein expression data can identify changes in the signal transduction and metabolic pathways allowing for identification of responses at the functional/molecular mechanisms level derived from gene editing
- CRISPR/Cas 9 and SWATH Acquisition are excellent partners.⁹ They allow monitoring of the target protein in parallel with thousands of other proteins across sample sets, ensuring reproducibility and accuracy by increasing specificity via MS/MS-based quantification.

References

1. Chandrasegaran S and Carroll D (2016) Origins of programmable nucleases for genome engineering. [J. Mol. Biol.](#), **428**, 963–989.
2. Knott GJ, Doudna, JA (2018) CRISPR-Cas guides the future of genetic engineering. [Science](#), **31(6405)**, 866-869.
3. CRISPR InDel Profiling Platform (CIPP) powered by IDAA: <https://cobotechnologies.com/services/Indel-profiling/>
4. Lonowski LA *et al.* (2017) Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. [Nat Protoc](#) **12(3)**, 581-603.
5. ProfileIt-V2 software available from COBO Technologies: <https://cobotechnologies.com/software/InDel-analysis-software/>.
6. Yang Z *et al.* (2015) Fast and sensitive detection of indels induced by precise gene targeting. [Nucleic Acids Res](#) **43(9)**, e59.
7. König S *et al.* (2019) Fast and quantitative identification of *in vivo* precise genome targeting-induced indel events by IDAA. [Methods Mol Biol](#) **1961**, 45-66.
8. Wang P *et al.* (2017) Tau interactome mapping based identification of Otub1 as Tau deubiquitinase involved in accumulation of pathological Tau forms *in vitro* and *in vivo*. [Acta Neuropathol](#) **133(5)**, 731-749.
9. Confirming gene mutation by CRISPR-Cas9 at the protein level and identifying proteome-wide changes. [SCIEX technical note RUO-MKT-02-9247-A](#).

Legal notice

A patent application covering the IDAA method is pending.

CIPP and ProfileIt-V2 are trademarks of COBO Technologies.

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to <https://sciex.com/diagnostics>. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries.

© 2020 DH Tech. Dev. Pte. Ltd. RUO-MKT-02-12095-A. AB SCIEX™ is being used under license.



Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at sciex.com/offices