**Capseome Project Team 2**

1. **Project Background**

Nephronophthisis (NPHP) is a genetic disorder characterized by cystic kidney disease and progressive fibrosis, frequently linked to defects in primary cilia. It affects approximately 1 in 50,000 children and young adults, and is often linked to syndromic variants of Polycystic Kidney Disease, such as Joubert, Bardet–Biedl, and Meckel–Gruber syndromes. (Gupta et al., 2021). While multiple genes have been associated with NPHP, the molecular mechanisms driving renal fibrosis remain poorly understood. FBXW7 is an E3 ubiquitin ligase known to regulate the degradation of various transcription factors, including c-Myc and SOX9, which are implicated in cell proliferation and differentiation (Shimizu et al., 2018). Previous studies suggested that loss of FBXW7 may disrupt protein homeostasis and contribute to tissue remodeling, but its role in kidney-specific fibrotic processes was not fully elucidated (Petsouki et al., 2021). This study aimed to investigate the functional consequences of FBXW7 deletion in renal tissue, particularly in the context of ciliary maintenance, inflammation, and fibrotic progression. Using a mouse model with targeted deletion of FBXW7 in renal epithelial cells, combined with quantitative mass spectrometry, the researchers sought to uncover downstream effectors and signaling pathways that mediate fibrosis, with the goal of identifying new therapeutic targets for NPHP and related ciliopathies.

1. **Summary of the Original Study**

This report details the processing and analysis of proteomics data from the PRIDE dataset PXD061542, which is part of the project *"SOX9-dependent fibrosis drives renal function in nephronophthisis"(Maulin Mukeshchandra Patel, 2025)*.

The study (Patel et al., 2025) investigates the role of FBXW7 in the development of renal fibrosis associated with NPHP. Using a conditional knockout mouse model (Fbxw7-null mIMCD-3 cell lines isolated from the kidney of an adult mouse), the researchers conducted quantitative bottom-up proteomic analysis using data-independent acquisition (DIA), where they processed the total protein from each sample. Briefly, the total protein suspension was reduced, alkylated and purified, digested with trypsin and separated by reverse phase. Eluted peptides were ionised by electrospray followed by mass spectrometric analysis in an Orbitrap Exploris 480 mass spectrometer. The resulting data were deposited in the PRIDE repository, a component of the ProteomeXchange consortium that facilitates the sharing of proteomics data. To uncover downstream molecular changes. Mass spectrometry was used to generate high-resolution proteomic data, resulting in the identification of 893,271 peptide-spectrum matches (PSMs), 38,737 unique peptides, and 4,923 unique proteins.

A list of text on a white background

AI-generated content may be incorrect.

Figure 1: Overview of Identified Spectra, Peptides, and Proteins.

One of the major findings was the downregulation of TMEM237, a protein critical for ciliary function, upon FBXW7 deletion (Fig.2A-C). This downregulation was associated with ciliary structural defects (Fig.2D-F). TMEM237 is a known component of the NPHP module, a group of proteins genetically associated with nephronophthisis (Gana et al., 2022). Its downregulation provides a mechanistic link between FBXW7 loss and ciliary dysfunction, establishing that the fibrotic phenotype observed in these mice may arise not only from transcriptional reprogramming (e.g., SOX9 upregulation) but also from direct structural defects in cilia due to impaired expression of transition zone components.

A close-up of a graph

AI-generated content may be incorrect.

Figure 2: knockout of FBW7 shows TMEM237 downregulation with ciliary defects.

**A**) Volcano plot showing differentially expressed proteins from Fbxw7-null cells (Fbxw7 KO#2 and Fbxw7 KO#7) versus empty vector (EV) pool mIMCD3 cells. The red dot represents the TMEM237 protein that is significantly downregulated in Fbxw7-null cells compared to EV pool mIMCD3 cells. n = 3 independent experiments. The data was normalized using cyclic loess, and statistical analysis was performed using linear models for microarray data (limma) with empirical Bayes (eBayes) smoothing to the standard errors. Proteins with an FDR-adjusted P value <0.05 and a fold change >2 were considered significant. (**B**) Validation of quantitative proteomics using immunoblotting and (**C**) quantification of TMEM237 from n = 3 independent experiments. Statistical analysis was performed using one-way ANOVA followed by Šídák’s multiple comparisons test and is presented as the mean ± SEM. (**D**, **F**) Each data point represents the mean ciliary length per field of view from >20 images or percent ciliation in cystic cells from n ≥ 3 animals. Statistical analysis was performed using the Mann–Whitney test and is presented as the mean ± SEM.

The proteomic data also revealed upregulation of transcription factors SOX9 and c-Myc. This was consistent with their known roles as FBXW7 degradation targets (Suryo Rahmanto et al., 2016). While both were elevated at the protein level, functional experiments revealed a clear difference in their contributions to disease. c-Myc upregulation had little impact on fibrosis severity or inflammation (Fig.3A), suggesting it was not a key driver in this context. In contrast, SOX9 played a central role in mediating the fibrotic phenotype. Its expression correlated with fibrotic regions, and deleting SOX9 in FBXW7-deficient mice significantly reduced fibrosis, inflammation, and tissue damage. The study also linked SOX9 to upregulation of WNT4 (Aggarwal et al., 2024), a profibrotic signaling molecule, indicating that SOX9 may activate a downstream transcriptional program promoting fibrosis (Fig.3B-G). These findings demonstrate that although both factors accumulate when FBXW7 is lost, only SOX9 is essential for disease progression. Additionally, the study identified WNT4 signaling dysregulation, likely acting downstream of SOX9, linking transcriptional reprogramming to fibrosis. These findings highlight a SOX9-dependent fibrotic program triggered by FBXW7 deletion, with implications for future therapeutic targeting.

A close-up of a graph

AI-generated content may be incorrect.

Figure 3: Effect of c-myc and SOX9 on disease progression.

(A) Representative images of whole kidney section scan, each data point represents one animal. Statistical analysis was performed using one-way ANOVA followed by Šídák's multiple comparisons test and is presented as the mean ± SEM. Scale bar: 400 µm.. (**B**) an average of WNT4 (n ≥ 5) MFI per animal. Statistical analysis was performed using one-way ANOVA followed by Šídák’s multiple comparisons test and is presented as the mean ± SEM. (**C**) qPCR of Wnt4 mRNA from EV pool and Fbxw7-null (Fbxw7 KO#2) mIMCD3 cells from n = 6 experiments. (**D**) Serum BUN (n ≥ 10), (**E**) Creatinine (n ≥ 10), (**F**) Cytastin C (n ≥ 9), and (**G**) 2KW/BW (n ≥ 10) from 3-month-old Fbxw7f/f, Cdh16Cre;Fbxw7f/f, and Cdh16Cre;Fbxw7f/f;Sox9+/f mice. Each data point represents one animal. Statistical analysis was performed using one-way ANOVA followed by Šídák’s multiple comparisons test and is presented as the mean ± SEM. Source data are available online for this figure

1. **Team Findings and Comparative Analysis**

The dataset was accessed using the rpx R package. While the standard workflow specified the use of .mzID files, none were provided in the PRIDE repository for this project. Instead, the available .mzML raw files were downloaded and processed locally using **SearchGUI**, a widely used search engine frontend for proteomics data analysis.

Out of twelve .mzML files, **two were found to be corrupt and could not be analyzed**. The remaining ten were successfully processed in SearchGUI, and search results were exported as .mzID files.

Due to the large size of the .mzID files (approximately 1 GB each), direct processing in R was impractical. To address this, the .mzID files were converted to .csv format using **Python in an Anaconda environment**, enabling lightweight and flexible downstream analysis in R.

Additionally, to allow for a direct comparison with the publication results and to compare differences between two peptide and protein identification approaches, we used EncyclopeDIA. EncyclopeDIA is a library search engine for peptide identification that counts with several algorithms for DIA data analysis [2]. Briefly, these kinds of tools search the spectra of peaks against a FASTA database containing protein sequences to match peaks and peptides and/or proteins. EncyclopeDIA allows for the use of various workflows. In this project, we generated an empiricarlly-corrected DIA chromatogram library [3](.elib format) from the raw .mzML files available; together with a Prosit-generated library for *Mus musculus* that is used as scaffold for the analysis, and the corresponding protein FASTA file. The software first creates the chromatogram library, which is then used for peptide and protein identification.

The software outputs several graphs that can be used to assess the quality of the analysis. Here are some examples from one of the runs (we first run each sample independently):



Figure 4 Delta RT from library. The histogram shows the difference in retention time (RT) between the observed peptides in the sample and the expected RTs from the spectral library. The x-axis represents the D(RT) in minutes, while the y-axis shows the frequency.

Figure 5 & 2: MS1 & MS2 Mass error (PPM): mass accuracy of the fragment ions matched to the spectral library. In the x-axis, the mass error in parts per million.

The y-axis displays the relative frequency of the mass errors.



Due to the nature of the experiment, the quantification of peptides and proteins is more challenging. Briefly, in a DIA experiment the quantification engine selects the best ion to quantify. Hence, it does not make sense to perform quantitation per sample because the engine might select a different ion in each of the samples. Meaning, the quantification results are only generated by integrating the previous analysis and generating a “shared” quantitation report that could then be used quantitatively. Unfortunately, because of the heavy nature of the analysis, we did not have time to generate the integrated quantification results.

1. **PSM Object Creation and Preprocessing**

Searchgui output:

The ten .csv files were read and merged into a unified PSM dataset using R. These files had already been filtered by SearchGUI to include only confident matches (non-decoy, rank 1). The combined dataset was converted into a tibble (idtbl) for downstream manipulation.

An additional filtering step was initially applied to remove ambiguous PSMs (spectrum IDs matching multiple peptides). However, this step removed nearly all data and was therefore excluded, as ambiguity resolution had likely been handled during the initial SearchGUI processing.

EncyclopeDIA output:

EncyclopeDIA provides the raw match-level data between the DIA spectra and the library in the features.txt files – which is basically already akin to the PSM object we have seen in the course. There are some differences because of the way EncyclopeDIA searches the library and produces the output, and because our data comes from a DIA experiment:

* The decoy hits are provided in a separate file.
* The spectra are not automatically assigned a rank (because of the way that the matching is done when using a DIA). To circumvallate this, we manually calculate a rank by grouping the results by scan and ranking them according to the HyperScore value.
* The identified peptides and proteins are already corrected for FDR < 0.01 (EncyclopeDIA outputs a separate result file where the q-score (or FDR) can be checked.

The combined feature file for all samples was then filtered for rank = 1 (in other words, selecting the peptide with higher score, we used HyperScore forthis), and the decoy hits were removed from the data frame.

1. **Identification Summary**

|  |  |  |
| --- | --- | --- |
|  | **Searchgui** | **EncyclopeDIA** |
| **Total PSMs (after filtering)** | 1,505,627 | 105,600 |
| **Unique peptide sequences** | 18,687 | 48,672 |
| **Unique protein accessions** | 11,860 | 12803 |
| **Unique razor proteins** |  | 11790 |

These results indicate a high-confidence and deep coverage of the proteome in the analysed dataset.

1. **Comparison with previous results**

The authors provide a summary of their analysis in an .mztab file.

**References**

Aggarwal, S., Wang, Z., Rincon Fernandez Pacheco, D., Rinaldi, A., Rajewski, A., Callemeyn, J., Van Loon, E., Lamarthée, B., Covarrubias, A.E., Hou, J., Yamashita, M., Akiyama, H., Karumanchi, S.A., Svendsen, C.N., Noble, P.W., Jordan, S.C., Breunig, J.J., Naesens, M., Cippà, P.E., Kumar, S., 2024. SOX9 switch links regeneration to fibrosis at the single-cell level in mammalian kidneys. Science 383, eadd6371.

Gana, S., Serpieri, V., Valente, E.M., 2022. Genotype–phenotype correlates in Joubert syndrome: A review. American Journal of Medical Genetics Part C: Seminars in Medical Genetics 190, 72-88.

Gupta, S., Ozimek-Kulik, J.E., Phillips, J.K., 2021. Nephronophthisis-Pathobiology and Molecular Pathogenesis of a Rare Kidney Genetic Disease. Genes 12, 1762.

Maulin Mukeshchandra Patel, V.G., Bryan Lettenmaier, Eleni Petsouki, Kurt A Zimmerman, John A Sayer, and Leonidas Tsiokas, 2025. SOX9-dependent fibrosis drives renal function in nephronophthisis.

Patel, M.M., Gerakopoulos, V., Lettenmaier, B., Petsouki, E., Zimmerman, K.A., Sayer, J.A., Tsiokas, L., 2025. SOX9-dependent fibrosis drives renal function in nephronophthisis. EMBO Molecular Medicine, 1-21-21.

Petsouki, E., Gerakopoulos, V., Szeto, N., Chang, W., Humphrey, M.B., Tsiokas, L., 2021. FBW7 couples structural integrity with functional output of primary cilia. Communications Biology 4, 1066.

Shimizu, K., Nihira, N.T., Inuzuka, H., Wei, W., 2018. Physiological functions of FBW7 in cancer and metabolism. Cellular Signalling 46, 15-22.

Suryo Rahmanto, A., Savov, V., Brunner, A., Bolin, S., Weishaupt, H., Malyukova, A., Rosén, G., Čančer, M., Hutter, S., Sundström, A., Kawauchi, D., Jones, D.T.W., Spruck, C., Taylor, M.D., Cho, Y.J., Pfister, S.M., Kool, M., Korshunov, A., Swartling, F.J., Sangfelt, O., 2016. FBW7 suppression leads to SOX9 stabilization and increased malignancy in medulloblastoma. The EMBO Journal 35, 2192-2212-2212.