**Unbiased spatial proteomics analysis of SWIPP1019R mutant mouse brain reveals significant disruption of endo-lysosomal pathways.**  pathogenicConceptually, the goal of spatial proteomics is to predict the subcellular localization of proteins. In practice, this is done by fractionating biological samples to resolve various membrane and non-membrane bound compartments followed by quantitative protein mass spectrometry analysis. Subsequent analysis is performed to identify groups of proteins that covary together across subcellular space. These modules (aka clusters) of proteins represent hypothesized subcellular organelles. Given a map defining protein subcellular localization, we then explore how the organization of the proteome may be altered. Here we analyze modules for differential abundance in order to identify modules that are perturbed by the SWIP mutation. In addition to evaluating differential abundance at the level of individual proteins, we assessed overall differences between WT and SWIPP1019R conditions at the level of protein modules. Using spatial proteomics, we extend inference from individual proteins to the level of protein-groups to explore the cell biological consequences of the SWIPP1019R mutation.

We analyzed the brains of control and SWIPP1019R mutant mice by subcellular fractionation and 16-plex TMT MS profiling. We isolated seven biological fractions (BioFractions) by following the subcellular fractionation protocol established by Geladaki *et al*., with modifications for homogenization of brain tissue (Geladaki et al., 2019; Hallett et al., 2008). Our experimental design is summarized in Supplemental Figure X. Proteins were quantified by analysis with MSstatsTMT, a statistical tool for normalization and statistical testing in TMT mass spectrometry experiments (REF:Huang2020). The normalized protein data were then used to construct a spatial proteomics network in which edges between nodes are the pairwise Pearson correlation between scaled protein profiles. Proteins with similar subcellular distribution are highly correlated. To resolve the structure of the spatial proteomics network, Geledaki *et al.* utilizea support vector machine learning algorithm that relies upon a predefined set of known, high-confidence subcellular markers. Here, we take an unbiased, data-driven approach, partitioning the spatial proteomics network by optimization of a quality statistic which describes the overall quality of a network partitioned into communities.

Brains from 10-month-old mice were gently homogenized to release intact organelles, followed by successive centrifugation steps to enrich subcellular compartments into different fractions based on their density (Figure 2C) (Geladaki et al., 2019). Seven WT and seven MUT fractions (each prepared from one brain, 14 samples total) were labeled with unique isobaric tandem-mass tags and concatenated. We also included two sample pooled quality controls (SPQCs), which allowed us to assess experimental variability and perform normalization between experiments. By performing this experiment in triplicate, deep coverage of the mouse brain proteome was obtained—across all 48 samples we quantified 86,551 peptides, corresponding to 7,488 proteins. After data pre-processing, normalization and filtering we retained 6,919 reproducibly quantified proteins in the final dataset (Table S2).

MSstatsTMT utilizes a linear mixed-model (LMM) framework. Following data normalization and protein summarization, each protein in the dataset is fixed with a LMM expressing the major sources of variation in the experimental design. Model-based comparisons are then made between pairs of treatment conditions. Using MSstatsTMT we assessed two types of protein-level contrasts: intra-BioFraction comparisons between WT and MUT genotypes for each of the seven biological fractions, and the overall comparison across all BioFractions between WT and MUT groups.

In the first analysis, there were 65 proteins with significantly altered abundance in at least one of the 7 subcellular fractions (Benjamini-Hochberg FDR < 0.05, Table S2 and Figure 2-figure supplement 2). Five proteins were differentially abundant between WT and MUT in all 7 fractions, including four WASH proteins and RAB21A—a known WASH interactor that functions in early endosomal trafficking (WASHC1, WASHC2, WASHC4, WASHC5, Figure 2E) (Del Olmo et al., 2019; Simpson et al., 2004).

Next, to evaluate global differences between WT and MUT brain, we analyzed the average effect of genotype on protein abundance across all BioFractions. At this level, there were 728 differentially abundant proteins between WT and MUT brain (Benjamini-Hochberg FDR < 0.05) (Table S2). All five WASH complex proteins exhibited a significant reduction for the overall Mutant-Control comparison. These data affirm that the SWIPP1019R mutation destabilizes the WASH complex, resulting in an ~50% reduction WASH complex proteins.

We then aimed to place these differentially abundant proteins into a more meaningful biological context using a systems-based approach. For network-based analyses, we clustered the protein covariation network defined by pairwise correlations between all 6,919 proteins. Our data-driven, quality-based approach used Network Enhancement (Wang et al., 2018) to remove biological noise from the covariation network and employed the Leiden algorithm (Traag et al., 2019) to identify optimal partitions of the graph. Clustering of the protein covariation graph identified 51 modules of proteins that covaried together. We analyzed these modules for enrichment of proteins predicted to reside in the 10 subcellular compartments defined by Geledaki et al. All 10 subcellular compartments are represented by unique modules in our spatial proteomics network.

Endoplasmic reticulum (ER), Lysosome, Peroxisome, Mitochondria, Nucle

The strength of linear mixed-models lies in their flexibility. To test for module-level differences between WT and MUT brain, we fit linear mixed-models to each module-level subset of the data. The data for each protein were sum normalized, to scale protein measurements. In a mixed-model the response variable is taken to be a function of both fixed- and random-effects. If the set of possible levels of a covariate is fixed and reproducible, then the factor is modeled as a fixed-effect parameter. In contrast, if the levels of an observation reflect a sampling of the set of all possible levels, then the covariate is modeled as a random-effect. Random or mixed-effects represent categorical variables that reflect experimental or observational units within the dataset. As such, mixed-effect parameters account for the variation occurring among lower levels of an upper level unit in the data (Bates et al., 2015). We extended the LMM framework developed by MSstatsTMT to perform inference at the level of protein groups. Given a map partitioning the proteome into modules of covarying proteins, we wish to assess the module-level difference between control and SWIPP1019R conditions. We fit the data for each module in the dataset with a LMM. We represent the proteins within each module as the mixed-effect term Protein, capturing variation among a module’s constituent proteins.

3 of the 51 modules exhibited an overall significant difference between WT versus MUT brain (Bonferroni P-Adjust < 0.05, percent change > ± 5.0 % Control; Table S3).

M23 (n = 134)

M21 (n = 146) Lysosome

M33 (n = 62) enriched for several protein complexes CORVET HOPS VPS

~~Across all fractions, the abundance of M23 was significantly lower in MUT brain compared to WT, providing evidence that the SWIP~~~~P1019R~~ ~~mutation reduces the stability of this protein subnetwork and impairs its function (Figure 2F-G).~~

~~In contrast to the decreased abundance of the WASH complex/endosome module, M19, we observed three modules (M2, M159, and M213) which were enriched for lysosomal protein components (Geladaki et al., 2019), and exhibited increased abundance in MUT brain (Figure 3). M159 (Figure 3B) contained the lysosomal protease Cathepsin A (CTSA), while M213 (Figure 3D) contained Cathepsin B (CTSB), as well as two key lysosomal hydrolases GLB1 and MAN2B2, and M2 (Figure 3C) contained two Cathepsins (CTSS and CTSL) and several lysosomal hydrolases (e.g. GNS, GLA, and MAN2B1) (Eng and Desnick, 1994; Mayor et al., 1993; Mok et al., 2003; Moon et al., 2016; Patel et al., 2018; Regier and Tifft, 1993; Rosenbaum et al., 2014). Notably, M2 also contained the lysosomal glycoprotein progranulin (GRN), which is integral to proper lysosome function and whose loss is widely linked with neurodegenerative pathologies (Baker et al., 2006; Pottier et al., 2016; Tanaka et al., 2017; Zhou et al., 2018). In addition, M2 contained the hydrolase IDS, whose loss causes a lysosomal storage disorder that can present with neurological symptoms (Hopwood et al., 1993; Schröder et al., 1994). The overall increase in abundance of modules M2, M159, and M213, and these key lysosomal proteins (Figure 3E-G), may therefore reflect an increase in flux through degradative lysosomal pathways in SWIP~~~~P1019R~~ ~~brain.~~

~~Furthermore, Module 2 (Figure 3C) included multiple membrane proteins and extracellular proteins, such as ITGA5 (an integrin shown to be upregulated and redistributed upon loss of WASH1), ATP13A2 (a cation transporter whose loss causes a Parkinsonian syndrome), and MMP17 (an extracellular metalloprotease), suggesting a link between these proteins and lysosomal enzymatic function (English et al., 2000; Ramirez et al., 2006; Zech et al., 2011). Increased abundance of these M2 proteins in MUT brain may indicate that WASH complex disruption alters their cellular localization. Taken together, these changes appear to reflect a pathological condition characterized by distorted lysosomal metabolism and altered cellular trafficking.~~

~~In addition to these endo-lysosomal changes, network alterations were evident for an endoplasmic reticulum (ER) module (M83), supporting a shift in the proteostasis of mutant neurons (Figure 2-figure supplement 3B). Notably, within the ER module, M83, there was increased abundance of chaperones (e.g. HSPA5, PDIA3, PDIA4, PDIA6, and DNAJC3) that are commonly engaged in presence of misfolded proteins (Bartels et al., 2019; Kim et al., 2020; Montibeller and de Belleroche, 2018; Synofzik et al., 2014; Wang et al., 2016). This elevation of ER stress modulators can be indicative of neurodegenerative states, in which the unfolded protein response (UPR) is activated to resolve misfolded species (Garcia-Huerta et al., 2016; Hetz and Saxena, 2017). These data demonstrate that loss of WASH function not only alters endo-lysosomal trafficking, but also causes increased stress on cellular homeostasis.~~

~~Finally, besides these endo-lysosomal and homeostatic changes, we also observed two synaptic modules (M35 and M248) that were reduced in MUT brain (Figure 2-figure supplement 3C-D). These included mostly excitatory post-synaptic proteins such as HOMER2 and DLG4 (also identified in WASH1-BioID, Figure 1), consistent with endosomal WASH influencing synaptic regulation. Decreased abundance of these modules indicates that loss of the WASH complex may result in failure of these proteins to be properly trafficked to the synapse. In line with these findings, we observed less excitatory synapses in adult MUT brain compared to WT (Figure 2-figure supplement 4), validating that these module-level differences correlate with cellular alterations~~ *~~in vivo~~*~~.~~