Science
AAAS

Supplementary Materials for

GPNMB confers risk for Parkinson's Disease through interaction with alpha-synuclein

Maria E. Diaz-Ortiz, Yunji Seo, et al.

Correspondence to: chenplot@pennmedicine.upenn.edu

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods:

RNA isolation and library preparation from human brain samples:

Human postmortem brain samples from neurologically normal controls (NC, n = 2) and PD (n = 4) individuals were obtained from the University of Pennsylvania Center for Neurodegenerative Disease Research (CNDR) Brain Bank and dissected as previously described (*29*). Demographics are summarized in **Table S1**. The regions analyzed included the caudate nucleus, cingulate gyrus, and cerebellum. Samples comprised predominantly gray matter. Genotypes for the brain samples were obtained as previously described (*29, 30*).

Total RNA was isolated from postmortem brain samples using the Qiagen RNeasy kit as previously described (*29*). After RNA isolation, RNA purity and integrity were assessed by spectrophotometric measurement of 260/280 nm OD ratios and by capillary electrophoresis on an Agilent 2100 Bioanalyzer.

Library preparation was performed with the KAPA RNA HyperPrep Kit (KR1350, Illumina Platforms, Roche, WI). Briefly, RNA was fragmented using heat and magnesium for 8 minutes at 94°C to obtain 100-200bp fragments. The first cDNA strand was then synthesized using a thermocycler, followed by the second strand synthesis combined with A-tailing. Unique adapters were then ligated onto the library insert fragments and amplified using high-fidelity, low-bias PCR. The products then underwent a bead-based cleanup for purification of the cDNA libraries.

SeqCap RNA ChoiceTM probe pool design:

The SeqCap RNA probe design pool was designed by Roche Sequencing Solutions Custom Design (Roche, WI). The pool was designed to contain probes about 60bp in length with no more than 20 “close matches” in the genome, as determined by the Sequence Search and

Alignment by Hashing Algorithm (SSAHA), for the purposes of providing sufficient coverage of transcripts of interest while minimizing potential off-target effects. A "close match" was defined as any genomic sequence that differed from one of the probe sequences by five or fewer single-base insertions, deletions, or substitutions. The majority of included probes had no off-target matches, with an exceedingly small percentage of probes displaying very few off-target matches. Only exonic probes covering the transcripts of interest were included and, to minimize SNP-mediated capture bias, probes that overlapped any SNPs in linkage disequilibrium ($r^2 > 0.2$) with the sentinel PD GWAS SNP were excluded.

Target cDNA enrichment and sequencing:

Target cDNA enrichment and sequencing was performed as previously described (*14*) using the SeqCap EZ Accessory kit v2 (07145594001, Roche, WI). Briefly, equimolar amounts of cDNA libraries were combined for a total mass of 1ug. Each of these libraries underwent multiplexed PCR with unique index oligonucleotides. The libraries were then hybridized with capture probes using the SeqCap EZ Hybridization and Wash Kit (05634261001, Roche, WI). In brief, libraries were dried with heat in a vacuum and resuspended with hybridization reagents. Pooled capture probes for each region were added to the resuspended libraries and incubated for 20 hrs in a thermocycler at 47°C, with the lid temperature at 57°C. The captured multiplex cDNA samples then underwent stringent washing steps and were amplified using ligation-mediated PCR (LM-PCR). These post-capture PCR amplified libraries were pooled and sequenced on two lanes of an Illumina HiSeq 2500 with 150bp paired-end reads, yielding ~150 million read pairs per lane.

Read mapping and allele specific expression analyses:

To assess RNA-seq reads quality we employed FastQC (*31*), while for reads quality filtering and trimming we used Trimmomatic (Version 032) (*32*). We ran Trimmomatic to remove low quality fragments in a 4 base wide sliding window (average window quality below PHRED 20), and low quality leading and trailing bases (below PHRED 10). We also dropped all the reads with average PHRED quality below 25, as well as reads shorter than 75 bases. Depending on the sample, 65–80% of reads passed this trimming and filtering step, resulting in 6.5–24 million read pairs per sample for mapping.

To perform unbiased allele specific read mapping to the reference human genome (hg19) we applied WASP–STAR pipeline (**Figure S7**). First, we mapped reads with STAR (*33*) applying 2 step alignment and filtered them for mapping bias using WASP (*34*). Before proceeding with variant calling, we removed duplicate reads using rmdup_pe.py script incorporated into WASP pipeline. To call and filter SNV we used GATK tools; HaplotypeCaller, SelectVariants and VariantFiltration. We obtained allele specific read counts by GATK – ASEReadCounter. In order to filter out intergenic variants, we functionally annotated SNVs using VariantAnnotation (*35*) and TxDb.Hsapiens.UCSC.hg19.knownGene (*36*) R packages. To test for allele specific expression (ASE) at the gene level, we first selected proxy SNP's that were highly linked ($r^2 > 0.6$) with rs199347 and located within a coding region for the gene of interest in order to assign allele of origin. For *GPNMB*, we assigned the allele of origin for each transcript read based on genotype at rs199355, and for *KLHL7*, we assigned the allele of origin based on genotype at rs2072368. We then tested for allelic imbalance with a beta-binomial model with overdispersion using the MBASED R package (*13*). P-values were adjusted for false discovery rate using the Benjamini-Hochberg method (*37*).

Colocalization analysis:

Colocalization analysis was performed as previously described (*11*) using the 'coloc' package in R (version 4.0-2). Summary statistics from the latest PD GWAS (*2*) excluding the 23andMe cohort were downloaded from the article's supplementary information. The data includes 17.5M SNPs for 33.6K cases and 44.9K controls. All SNPs within 1Mb up or downstream of the top GWAS SNP (rs199351) were included in the analysis. Full summary eQTL data for whole blood, caudate nucleus, and substantia nigra were downloaded from the GTEx portal (Version 7) (*10*). GTEx contains paired-end RNA-seq (Illumina TruSeq) and whole genome sequencing data from normal individuals. The prior probabilities of p1, p2, and p12 were set to 1e-4, 1e-4, and 1e-5 respectively. Significant colocalization was determined as having a PPH4 > 75%. Sensitivity analysis was performed to determine whether PPH4 is robust over plausible values of p12 (1e-5 to 1e-6) (*38*). Association plots were generated using LocusComparer package (version 1.0.0) (*39*) in R.

Immortalized cell line transfection and collection:

For co-immunoprecipitation (co-IP) experiments, HEK293 or HeLa cells were plated at 700,000 or 1 million cells per well, respectively, and 18 hours after plating, transfection was performed with 4ug of each DNA construct and Lipofectamine 2000 (Thermo Scientific) in serum-free DMEM. Cells were transfected with either GPNMB-myc-DDK pCMV6-Entry (Origene) and 5' UTR-Syn pcDNA3.1+ (obtained from Dr. Kelvin Luk) for the GPNMB-flag IP or GPNMB untagged clone pCMV6-XL4 (Origene) and Syn-Myc pcDNA3.1+ (obtained from Dr. Kelvin Luk) for the Syn-Myc IP. 4 hours post-transfection, media was changed to DMEM with 10% FBS 1% L-Glut and 1% Pen-Strep. 20 hours post-transfection, cells were washed in dPBS and lysed in CHAPS buffer (25mM Tris, 150mM NaCl, 1mM EDTA 1% CHAPS, 5%

glycerol, pH to 7.4). Lysates were collected and spun down at 4ºC for 30 minutes at 21380 xg. BCA assays (Thermo Scientific) were used to determine protein concentrations.

For immunofluorescence experiments, HEK293 or HeLa cells were plated at either 100,000 or 200,000 cells per well on either PDL-coated or untreated 12mm glass coverslips in a 12-well format. 18 hours after plating, the cells were transfected with 1.6ug of each DNA construct (GPNMB untagged clone pCMV6-XL4 and 5' UTR-Syn pcDNA3.1+) and Lipofectamine 2000 (Thermo Scientific) in serum-free DMEM. Cells were fixed, stained, and imaged 20 hours post-transfection as described in the immunofluorescence section.

Co-immunoprecipitation:

To immunoprecipitate (IP) GPNMB, 300 µL anti-Flag-conjugated beads (Sigma A2220) or mouse IgG-conjugated beads (Sigma A0919, used as a negative control) were used to IP from 1,000 µg of lysate from GPNMB-myc-DKK and 5'UTR-syn double-transfected cells in CHAPS buffer overnight at 4ºC. To IP α-synuclein, 300 µL anti-c-Myc-conjugated beads (Sigma A7470) or rabbit IgG-conjugated beads (Sigma A2909, used as a negative control) were used to IP from 1,000 µg of lysate from GPNMB untagged clone and Syn-Myc double-transfected cells in CHAPS buffer overnight at 4ºC. After 24 hours, the protein-conjugated beads were washed 3x with CHAPS buffer and the bound protein was competitively eluted from the beads using either 250 µM 3x flag peptide (Sigma) or 250 mM myc peptide (Sigma) for 1 hour at 4ºC. The input, eluates, and flow-throughs were run on a mini-protean 4-20% polyacrylamide gradient gel at an input:IP:flow-through ratio of 1:10:1 or 1:100:1 ratio, as indicated in the text.

CRISPR-Cas9 knock-out of human iPSC's:

Human induced pluripotent stem cells (iPSC) with *Neurog1* and *Neurog2* in a bicistronic doxycycline-inducible expression cassette (*22*) were used for generation of GPNMB knock-out

(KO) iPSC and rapid induction of iPSC-derived neurons (iPSC-N). The protocol for generating CRISPR edited iPSC's is summarized in **Figure S3**. iPSCs were dissociated with Accutase (Innovative Cell Technologies), spun down at 200xg for 4 minutes, and plated at 240k cells/well on Matrigel (CORNING)-coated 12-well plates in mTeSR1 media (STEMCELL Technologies) with 2uM Thiazovivin (Cayman Chemical). The next day, GPNMB sgRNA (Synthego) and TrueCut Cas9 (Invitrogen) were mixed at a 0.75:1 molar ratio in OptiMEM. An equal volume of OptiMEM + 1:10 Lipofectamine STEM reagent (Thermo Fisher) was added to the sgRNA/Cas9 mixture and allowed to incubate at room temperature for 15 minutes. The media was replaced with serum-free OptiMEM + 2uM thiazovivin and transfected with the sgRNA/Cas9/Lipofectamine mixture for 3-4 hours, after which, an equal amount of mTeSR1 was added. 18h after transfection, the media was replaced with 100% mTeSR1 and 48h later, the cells were dissociated with Accutase and plated as single cells in a low-density manner (2,000-4,000 cells per 10cm plate).

Colonies resulting from individual clones were isolated; a small fraction of the colony was processed for genomic DNA using QuickExtractTM DNA extraction solution (Lucigen) and the rest was plated for expansion. QuickExtract DNA from individual clones was screened for indels using T7 Endonuclease 1 (T7E1) assay. Briefly, a region spanning the cut site in an asymmetric manner was amplified by PCR using the following primers:

Fwd: 5'-TACGTGTCGCTTGGAATCTTGA-3'

Rev: 5'-ATATTGCCAAGTCCAGGTGC-3'

The PCR product was incubated with T7E1 for 30 minutes at 37ºC, after which the reaction was stopped with EDTA. The reaction product was run on a 1.5% agarose gel and clones with 2 bands of different sizes (rather than a single band of larger size) were prioritized for further

validation. PCR products were purified and Sanger sequenced. The sequencing results were analyzed with Synthego's ICE tool (https://ice.synthego.com/#/), which allowed us to identify 5 clones of interest (showing two bands on T7E1 assay and predicted indels on one or both DNA strands) that, along with a wild-type (WT) clone, were expanded and validated by TOPO cloning.

Briefly, high quality genomic DNA was isolated using QIAamp DNA mini kit (QIAGEN) and the region of interest was PCR amplified with the primers listed above. PCR products were purified and incubated with pCR™4-TOPO® vector (Invitrogen) for 5 minutes at RT. One Shot® chemically competent *E. coli* were transformed with the TOPO reaction products, streaked on LB + Kanamycin selective plates, and incubated at 37ºC overnight. The next morning, 30 colonies per clone were picked and miniprepped, and the resulting DNA was sent for sanger sequencing. Sequences were aligned using Benchling's alignment tool and protein sequences were predicted using the Expasy protein translator tool (https://web.expasy.org/translate/). Cells populations were determined to be clonal if only 2 alleles were present at a ~50:50 ratio (tested by binomial likelihood ratio test).

GPNMB KO in iPSC's was validated by qPCR using either the SYBR-green or TaqMan master mixes and the following primers:

GPNMB (fwd): 5'- CTTCTGCTTACATGAGGGAGC-3'

GPNMB (rev): 5'- CTCCCTTCCAGGAGTTTTTCC-3'

ACTB: 4352935E, Applied Biosystems

qPCR data was analyzed using the ΔΔCt method.

Culture and differentiation of iPSC-Neurons:

Prior to differentiation, iPSC cultures were maintained on Matrigel-coated 6cm tissue culture plates with mTeSR1 media and mechanically passaged every 4 days with StemMACS passaging solution XF. The neural induction protocol is described in detail in **Figure S4** (*40*). To initiate neuronal induction, iPSCs were dissociated with Accutase, spun down at 200xg for 4 minutes, and plated at 700k cells/well on Matrigel-coated 6-well plates in N2 pre-differentiation media (KnockOut DMEM/F12 (Gibco) with 1x N-2 supplement (Gibco), 1x MEM Non-essential Amino Acid (MEM NEAA) solution (Gibco), 10ng/ml BDNF (Gibco), 10ng/ml NT-3 (Gibco, Preprotech), 0.2ug/ml Laminin (Gibco), and 2ug/ml doxycycline hyclate (Sigma)) with 10ug/ml Y-27632 (Tocris). Full media changes were performed at 24 and 48hrs with and without Y-27632 respectively. 12mm round glass coverslips were placed in a 12-well plate, and coverslips or 6-well plates were coated with 10ug/mL poly-D-lysine (PDL, Sigma) in borate buffer (pH = 8.4) for 1hr at room temperature. The PDL-coated coverslips or plates were then rinsed three times with tissue culture grade water (Lonza) and coated with 1:100 Matrigel in DMEM/F12 (Gibco) for 1hr at room temperature prior to being stored at 37ºC overnight. On day 3, the cells were re-plated at 200,000 cells/well on PDL/Matrigel coated coverslips (for immunofluorescence) or at 1 million cells/well in PDL/Matrigel coated 6-well plates (for immunoblotting or RNA isolation). Pre-differentiated iPSC-N were dissociated with Accutase, spun down at 200xg for 4 minutes, and resuspended in classic neuronal media (50% DMEM/F12, 50% Neurobasal-A media (Gibco), 1x MEM NEAA, 0.5x N-2 supplement, 0.5x GlutaMAX supplement (Gibco), 0.5x B27 supplement (Gibco), 10ng/ml BDNF, 10ng/ml NT-3, 1ug/ml Laminin, and 2ug/ml doxycycline) for plating. Half media changes were performed every 7 days with neuronal media without doxycycline.

Immunofluorescence:

The media was aspirated from coverslip-containing wells, and the cells were fixed in 2% paraformaldehyde in dPBS for 15 minutes at room temperature. After fixing, the coverslips were rinsed 5x with dPBS and blocked/permeabilized with blocking buffer (3% bovine serum albumin + 0.05% saponin in dPBS) for 1hr prior to incubating in primary antibody overnight. The coverslips were then rinsed 4x with blocking buffer and incubated with secondary antibody for 1hr at room temperature. Finally, the coverslips were rinsed 4x with blocking buffer and either incubated in DRAQ5 for 30 minutes or mounted onto microscope slides without DRAQ5 using a small amount of ProLong Gold (ThermoFisher). Slides were sealed with clear nail polish and stored at 4C. Image stacks of 1um thickness were acquired by confocal microscopy (Leica SP5) using a 40x oil immersion objective with 2x Zoom. Antibody concentrations are summarized in the Key Resources Table.

Image processing and quantification:

Image processing to quantify synaptic proteins was performed using a Cell Profiler pipeline based on Danielson et. al.'s previously published work (*41*). A single slice was chosen from each stack to focus our analysis on the plane with the most abundant synapsin-1 staining. TUBB3 images underwent image enhancement for neurite-like features and were used to calculate TUBB3+ area for normalization and expanded to generate a mask. Synapsin-1 and α-synuclein images underwent enhancement for speckle-like features, followed by object identification and characterization of object size/intensity. The total integrated intensity (i.e. sum for all particles) was used for analysis. The pipeline used has been made publicly available as supplementary material.

aSyn pathology in iPSC-N was also quantified using CellProfiler. Maximum intensity projections of z-stack images were created using FIJI. Desired 81a+ objects were identified with

intensity thresholding and the toal area of 81a+ objects were normalized with total area of DRAQ5+ objects. To quantify MJFR1+ staining in HEK293 overexpressing GPNMB, nuclei artifact staining was removed by generating DRAQ5+ nuclei mask and excluding MJFR1+ speckles within the mask. Total area of MJFR1+ objects was normalized by total area of DRAQ5+ objects.

A researcher blinded to sample identity was provided with randomly ordered maximum intensity projection of z-stack images of iPSC-N treated with AF594-PFF. Scores ranging from 0 to 3 were given for each criterion: intensity, length and frequency of internalized PFF. The three scores were added, and were normalized to the average score of WT iPSC-N for each individual experiments.

Synaptosome extraction:

Synaptosomes were extracted using Syn-PER™ Synaptic Protein Extraction Reagent (Thermo-Fisher) per the user instructions. Briefly, 14-day and 21-day iPSC-N's cultured in 6-well plates were rinsed twice with ice cold PBS, then scraped into 200ul SYN-PER reagent plus protease inhibitors (25mg Pepstatin in 3mL EtOH, 25mg Leupeptin, 25mg TPCK, 25mg TLCK, 25mg Trypsin Inhibitor from soybean, 5mL of 0.5M EDTA and add H_2O to a final volume of 25mL, to use at 1:1000) and phenylmethylsulfonyl fluoride (VWR) per well. The samples were spun down for 10 minutes at 1200 xg and 4°C and the supernatant was transferred to a new tube. The supernatant was then spun down for 20 minutes at 15,000 xg and 4°C, after which the supernatant (cytosolic fraction) was transferred to a new tube and the residual pellet (synaptosomal fraction) was resuspended in 20ul of Syn-PER reagent per well.

Immunoblotting:

Samples were diluted in 5X concentrated sample buffer (10g Sucrose, 1.85mL 0.5M Tris, pH 6.8, 1.0mL 0.1M EDTA, 1.0mL of 0.1% Bromophenol Blue, 1.0mL of 0.05% Pyronine Yellow, 0.615g of DTT, 10mL of 10% SDS, and adjusted to a final volume of 20mL with Milli-Q Water) and boiled at either 100ºC for 10 minutes or 70ºC for 15 minutes for heat-sensitive proteins (GPNMB). Samples were run on 4-20% polyacrylamide TGX gels (Bio-Rad Laboratories) and transferred onto 0.2 nitrocellulose membrane (Bio-Rad Laboratories). When blotting for α-synuclein, membranes were fixed in 0.4% PFA for 30 minutes (*42*). Membranes were blocked in 5% milk in TBS for 1 hour and blotted overnight at 4ºC with specific antibodies. The membranes were incubated for 2 hours in HRP-conjugated secondary antibodies and developed using Western Bright ECL and Sirius HRP substrates (Advansta). Antibody concentrations are summarized in the Key Resources Table. Densitometry was performed using the Bio-Rad Image Lab Software.

RNA sequencing of iPSC-N:

RNA was extracted from iPSC-N on days 0 and 14 after doxycycline induction using QIAGEN RNAeasy ® mini kit. RNA integrity was measured using RNA nano chips (Agilent) on an Agilent 2100 Bioanalyzer. All samples had RIN > 9.4. Library preparation was performed by the Penn Next Generation Sequencing (NGS) Core staff with the TruSeq Stranded mRNA Library Prep kit (Illumina cat #20020595) using 225ng total RNA and following manufacturer instructions. High sensitivity DNA chips (Agilent) were used to balance libraries prior to sequencing. In total 30 RNA samples (3 genotypes (WT, Het and KO) x 2 differentiation time (days 0 and 14) x 5 biological replicates) were sequenced. All 30 samples were sequenced together on a single S2 flow-cell generating >1.1 x 10^9 of 100 bp single end reads in total (on

average 3.69 x 10^7 reads per sample, ranging from 2.9 x 10^7 to 5.3 x 10^7). A summary of the Illumina NovaSeq run is provided in **Table S12**.

Read quality control and filtering:

A quality check of the raw reads was assessed using FastQC (*31*) and summarized with MultiQC *(43)*. Next, the adapters were removed and filtering of the low quality reads was conducted employing Trimmomatic, Version 0.39 (*32*) with the following parameters: *ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:5:20 MINLEN:50 AVGQUAL:30 HEADCROP:10*. The adapters that matched the sequences provided by TruSeq3-SE.fa were removed. The low-quality bases were cut off the start and off the end of reads if their PHRED score was below 5. The sliding window of 5 bp was applied to trim the bases if the window PHRED score dropped below 20. The 10 bases from the start of the reads were also cropped. Finally, the reads were discarded if the average PHRED score of the read was below 30.

Read alignment and quantification:

The sequence alignment BAM files generated by Bowtie 2 were used as the input to RSEM software tool, in order to quantify gene expression levels (*44*). On average, 3.61 x 10^7 (97.8%) of reads per sample passed the filtering steps and were mapped versus reference genome GRCh38 using Bowtie 2 the with the following parameters: *bowtie2 -q --phred33 --sensitive --dpad 0 --gbar 99999999 --mp 1,1 --np 1 --score-min L,0,-0.1 --nofw -p 32 -k 200 -x*, as suggested by RSEM manual. The read alignment rate was on average >89% (**Table S13**). To obtain the raw read counts at the gene level, an expectation maximization algorithm (RNA-seq by Expectation Maximization) was run by the following RSEM command: *rsem-calculate-expression --bowtite2 --forwardprob 0.*

Differential gene expression analysis:

To detect significant differences in iPSC-N gene expression data between Day 0 and Day 14 and between the genotypes (WT, Het, and KO), DESeq function form DESeq2 Bioconductor's R package (*45*) was used. RNA-seq data was modeled using the negative binomial distribution that accounts for overdispersion. In particular, DESeq function calculates the library size factors for each sample using the *median of ratios method*. Then, the gene-wise dispersion was assessed using the parametric argument of the *estimateDispersions* function, which estimates the dispersions using the negative binomial distribution. To conduct the pairwise comparisons (Day 14 vs. Day 0 in WT, KO vs. WT, Het vs. WT, and KO vs Het at Day 14) for each gene the *nbinomWaldTest* statistic was used. The Wald test P-values were calculated by scaling the coefficients by their standard errors and then compared to a standard Normal distribution. After DGE analysis, the lfcShrink function from R package apeglm *(46)* was applied to shrink the log2FC. We corrected for FDR using the Benjamini & Hochberg method (*37*). Genes with a BH corrected P-value < 0.01 and a |log2 fold-change| > 1 were considered to be significantly differentially expressed. To visualize the results of DGE analyses we used EnhanceVolcano R package version 1.10.0 (*47*).

Modular co-expression network analysis:

Modular co-expression network analysis of iPSC-Ns was performed using R package *CEMiTool* (Co-Expression Molecules identification Tool) version 1.14.1 (*48*). The rlog normalized expression data (15981 genes in 15 samples) was used as an input to *cemitool* function. As a first step, cemitool conducted an unsupervised filtering of expression data using the inverse gamma distribution to model the variance of genes. Out of 15981 genes, 1100 genes with the P-value < 0.1 survived this filtering were used for downstream analyses. The genes were

then separated into modules using the dissimilarities measures and the *Dynamic Tree Cut* package (*49*). The minimum number of genes per module was set to 20, and similar modules were merged together based on correlation.

To determine biological functions associated with the co-expression modules, the Gene Ontology gene sets *c5.all.v7.4*, which contain GO resources (BP = biological process, CC = cellular component and MF = molecular function) and human phenotype ontology (HPO) from MSigDB database (*50*), were used in overrepresentation analysis (ORA). The module gene set enrichment analysis (GSEA) was performed using the *fgsea* R package (*51*) within the CEMitool pipeline. First, a z-score gene normalization on all input genes was performed following by calculation of the mean for each sample class (GPNMB WT, Het, and KO). Next, a pre-ranked GSEA was performed independently for each GPNMB genotype. The module activity was visualized by normalized enrichment score (NES), which represents enrichment score (ES) for a module in each iPSC-N GPNMB genotype normalized by the number of genes in the module. To construct the gene interactions network in the co-expression modules, the combined human protein-protein interaction data, downloaded from GeneMANIA (*24*), were added to the CEMiTool object using the function *interactions_data*. The interactions between the genes were visualized using the *plot_interactions* function.

<u>aSyn pre-formed fibril internalization and immunofluorescence experiments:</u>

Human wild-type pre-formed fibrils (PFF) and fluorescent fibrils including Alexa Fluor 594 (AF594) conjugated aSyn were provided by Dr. Kelvin Luk at the Penn CNDR. All fibrils were kept at -80°C until use. After the stock of 5mg/mL was diluted to a working concentration of 0.1mg/mL in dPBS, fibrils were sonicated (Diagenode Biorupter® Plus) on high, sonicating for 30 sec followed by 30 sec rest for total of 10 minutes. Sonicated PFFs were then added to

classic neuronal media according to appropriate concentrations. Each PFF transduction experiment was repeated and quantified at least three times from independently differentiated neurons.

For AF594-PFF internalization experiments using iPSC-N, 10ug of PFF were added per well of a 12-well plate. iPSC-N were incubated in 4°C for 30 min, followed by an additional 30 min incubation at 4°C for the negative control (to inhibit endocytosis) or 1 hour 30 min in 37°C for experimental wells. All cells were immediately fixed and stained afterwards.

To demonstrate aSyn pathology following PFF transduction, we replaced regular media of iPSC-N at day 14 with classical neuronal media containing total of 1ug of human wild-type PFF per well. iPSC-N were fixed and stained after 14 days (at day 28 of neuronal induction). To extract soluble proteins, coverslips were fixed with 2% paraformaldehyde and 1% Triton X-100 in dPBS. 3% bovine serum albumin without saponin in dPBS was used as blocking buffer.

For wild-type PFF internalization experiments using HEK293, cells were plated at 200,000 cells per well in PDL-coated 12mm glass coverslips in a 12-well format. 18 hours after plating, the cells were transfected with 1ug of DNA (GPNMB untagged clone pCMV6-XL4 or TMEM106B-Flag) and Lipofectamine 2000 (Thermo Scientific) in serum-free DMEM. At 20 hours post-transfection, media was replaced with DMEM containing 1ug of PFF per well. The plates were incubated for 30 min at 4°C followed by an additional 30 min incubation at 4°C for negative control (to inhibit endocytosis) or 1 hr 30 min in 37°C for experimental wells. All cells were immediately fixed and stained for immunofluorescence afterwards.

Patient samples and genotypes:

Plasma and CSF samples of PD patients and neurologically normal controls were obtained as part of the Penn CNDR Neuropathology, Biomarker, and Genetics Biobank and the

Parkinson's Disease Molecular Integration in Neurological Diagnosis Initiative (MIND) studies. PD patients had a clinical diagnosis of PD made by a movement disorders specialist at the Parkinson's Disease and Movement Disorder Clinic (PDMDC) at the University of Pennsylvania, while controls had no known neurological disorder. Data were stored in the Penn Integrated Neurodegenerative Disease Database (INDD) (*30*). These studies were approved by the UPenn Institutional Review Board (IORG0000029). Informed consent was obtained at study enrollment. Participant demographics are reported in **Tables S4 – S7**. Individuals with known *GBA1* mutations (N370S, E365K, L444P) were excluded from the analysis, since GPNMB elevation has been reported in patients with Gaucher disease, a lysosomal storage disorder caused by homozygous mutations in GBA1 (*52, 53*). Genotyping of SNP rs199347 was performed by Infinium Global Screening Assay (Illumina), NeuroX genotyping platform (*30*), PANDoRA (Sequenom) panel (*30*), or MIND panel based on allele-specific PCR performed using FlexSix Dynamic Array integrated fluidic circuits (Fluidigm) and genotyping using BioMark HD system (Fluidigm) (*54*).

A subset of PD individuals had measures of disease severity as assessed by the motor subscale of the Unified Parkinson's Disease Rating Scale (UPDRS- III) (*25*). These individuals are part of the University of Pennsylvania U19 Cohort (formerly Udall Cohort), which has been previously described (*55*).

<u>Enzyme-Linked Immunosorbent Assay (ELISA):</u>

GPNMB protein levels within human plasma and CSF samples were measured with ELISA kits (R&D systems) according to manufacturer's instructions. CSF and plasma samples were diluted by factors of 1 in 2 and 1 in 30 respectively to obtain optical density measurements within the standard range. All samples were run in duplicates and absorbance at 450nm was

394 determined by a microplate reader (Berthold Technologies, Tristar LB 941). Only duplicate
395 samples with a coefficient of variation (CV) <25% were retained for analysis, and the average
396 CV across all samples used was 3.3%. Moreover, replicate samples assayed by ELISA on
397 different days, by different operators, across multiple freeze-thaw cycles, demonstrated excellent
398 reproducibility (Pearson r=0.97).
399 The log_{10}GPNMB concentration (pg/mL) for each sample was interpolated from a
400 sigmoidal (4 parameters logistic) standard curve using GraphPad Prism 9.
401 Multiple Linear Regression Analyses:
402 Multiple linear regression was used to determine factors that are significant predictors of
403 GPNMB concentration in plasma and CSF. Age, sex, PD status, and rs199347 genotype were
404 included as independent variables. A codominant genetic model, which considers each allele
405 combination (GG, AG, AA) as a separate factor (*56*), was used to model the effect of rs199347.
406 Regression coefficients were calculated with GG as the reference.
407 The relationship between GPNMB levels and disease severity was established in two
408 different PD cohorts. GPNMB values for the discovery cohort were generated by ELISA (R&D
409 systems), whereas GPNMB values for the replication cohort were generated using an aptamer-
410 based platform (*26*). UPDRS-PIII values were downloaded from the Penn INDD.
411 Statistical analyses:
412 Statistical analyses were performed with either PRISM or R. Data was tested for
413 normality with a Shapiro-Walk test. Data with more than two categorical groups was analyzed
414 by either 1-way ANOVA (followed by post-hoc Tukey or Dunnett tests) or Kruskal-Wallis test
415 (followed by post-hoc Dunn's test) depending on the data's distribution. Data with only two
416 groups was analyzed by either a Welch's two-sample t-test or Mann-Whitney-U test depending

on the data distribution. Correlations were determined using Spearman's rank correlation. Multiple linear regression (glm) was used to determine associations between two variables adjusting for potentially confounding covariates. Outlier determination was performed using ROUT method, Q (FDR)=1%.

Immunohistochemistry

Formalin-fixed, paraffin-embedded cingulate and temporal cortex samples were obtained from the Penn CNDR. Patient demographics are reported in **Table S1**. 6μm sections were cleared in xylenes and a descending EtOH series. Endogenous peroxidases were quenched in 30% H_2O_2 and 70% MeOH solution for 30 minutes. Slides were microwaved in citric acid Antigen Unmasking Solution (Vector Laboratories). After cooling, slides were rinsed in TBS-T (0.1 M Tris Buffer/0.05% TWEEN) and blocked (TBS-T/2%FBS/3%BSA). Sections were incubated overnight at 4°C in the primary antibody (see Key Resources table for antibody conditions). Once washed with TBS-T, sections were incubated for 1 hour at room temperature in the secondary antibody (see Key Resources table for antibody conditions). VECTASTAIN ABC Standard (Vector Laboratories) was applied for 1 hour at room temperature followed by ImmPACT DAB (Vector Laboratories). Sections were counterstained with Harris Hematoxylin (Thermo Scientific) for 40 seconds. Slides were dehydrated in an ascending EtOH series and xylenes then coverslipped with Cytoseal (Thermo Scientific).

Dephosphorylation and deglycosylation

For deglycosylation of cell and brain lysates, 5.0uL of PNGase F was added to 125ug of brain lysates or 60ug of cell lysates and incubated at 37°C for 30 minutes. For dephosphorylation experiments, 1uL of Lambda protein phosphatase was added to 125ug of brain lysates or 50ug of cell lysates and incubated at 30°C for 30 minutes.

Key Resources Tables:

Antibodies		
Name	Product ID	Dilution
Anti-Alpha-synuclein antibody (MJFR1)	Abcam (ab138501)	IF: 1:300 – 1:1000* WB: 1:1000
Anti-Synapsin-1	SySy (106-011)	IF: 1:500 WB: 1:1000
Anti-α-synuclein Antibody (syn211)	Santa Cruz (sc-12767)	WB: 1:200
Anti-α-synuclein (phospho S129) Antibody (81a)	Provided by Center for Neurodegenerative Disease Research (CNDR) at University of Pennsylvania	IF: 1:5000
Anti-Human Osteoactivin/ GPNMB Antibody	RnD Systems (AF2550)	IF: 1:500 IHC: 1:200 WB: 1:1000
Anti-GPNMB (E1Y7J) Rabbit mAb	Cell Signaling Technology (#13251)	WB: 1:1000
Anti-GPNMB Antibody (D-9)	Santa Cruz (sc-271415)	IF: 1:500 WB: 1:250
Anti-beta Actin antibody (AC-15)	Abcam (ab6276)	WB: 1:5000
Anti-beta-III Tubulin Antibody	Novus Biologicals (NB100-1612)	IF: 1:500
Anti-Flag M2 Mouse Monoclonal Antibody	Sigma Aldrich (F1804)	IF: 1:1000
Anti-GAPDH Antibody	Advanced Immunochemical Anti-GAPDH	WB: 1:3000
Anti-LAMP1 Antibody (H4A3)	University of Iowa Hybridoma Bank	IF: 1ug/mL
Goat Anti-Chicken Alexa Fluor 647	Invitrogen (A-32933)	IF: 1:1000
Goat Anti-Rabbit Alexa Fluor 488	Invitrogen (A-21206)	IF: 1:1000
Donkey Anti-Mouse Alexa Fluor 594	Invitrogen (A-21203)	IF: 1:1000
DRAQ5	Thermo-Fischer (62251)	IF: 1:5000
Goat Anti-Mouse HRP	Jackson Immuno Research (115-035-062)	WB: 1:3000

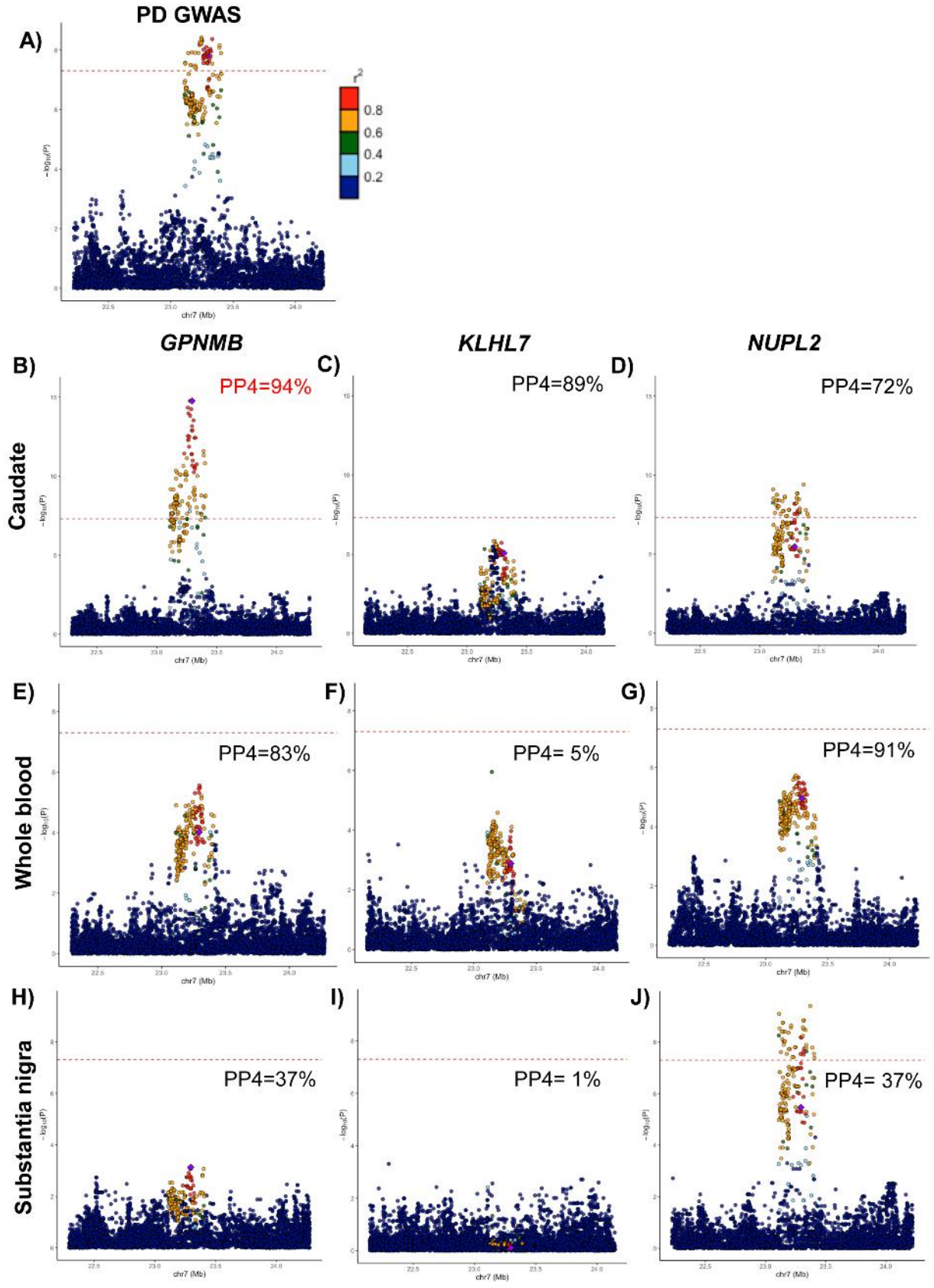
Goat Anti-Rabbit HRP	Jackson Immuno Research (111-035-144)	WB: 1:3000
Biotinylated Rabbit Anti-Goat	Vector Laboratories (BA-5000-1.5)	IHC: 1:5000

IF: Immunofluorescence, WB: Western blot, IHC: Immunohistochemistry

Cell Lines			
Name	Species	Obtained from	Citations
HeLa	Homo sapiens	Michael S. Marks	
QBI293	Homo sapiens	Penn Center for Neurodegenerative Disease Research (CNDR)	
iPSC-N	Homo sapiens	George Church	(*22)*

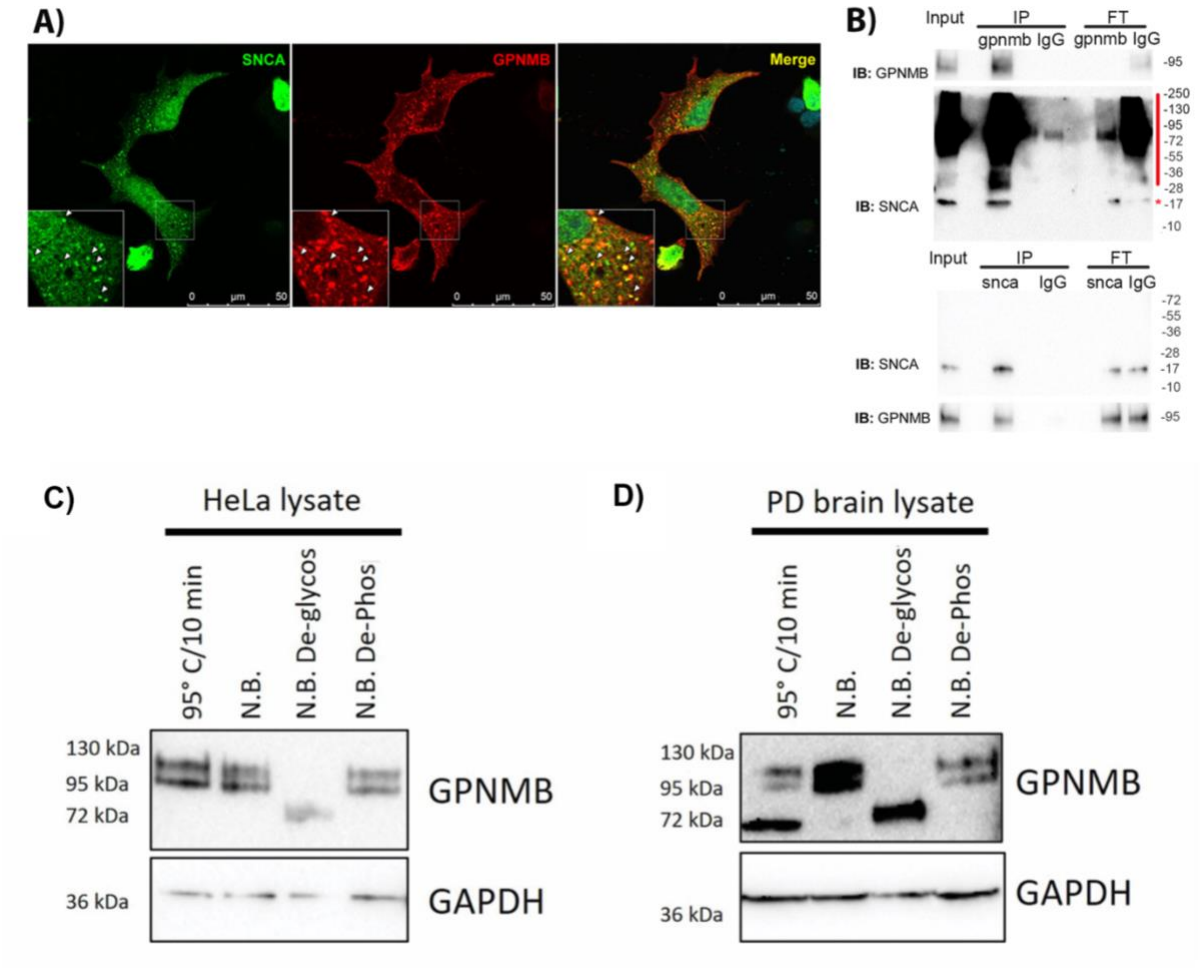
Deglycosylation and dephosphorylation		
Name	Product ID	Protocol
PNGase F (Deglycosylation)	New England Biolabs Inc., Catalog # P0704L	As per product instructions
Lambda Protein Phosphatase (Dephosphorylation)	New England Biolabs Inc., Catalog # P0753S	As per product instructions

*MJFR1 was used at 1:300 to stain endogenous α-synuclein in iPSC-N and 1:1000 to stain overexpressed α-synuclein in HEK293 and HeLa cells.


A)
PD GWAS
r²
0.8
0.6
0.4
0.2
chr7 (Mb)
GPNMB
KLHL7
NUPL2
B)
C)
D)
Caudate
PP4=94%
PP4=89%
PP4=72%
E)
F)
G)
Whole blood
PP4=83%
PP4= 5%
PP4=91%
H)
I)
J)
Substantia nigra
PP4=37%
PP4= 1%
PP4= 37%

450

Fig. S1. Colocalization analyses suggest *GPNMB* as the target gene of chromosome 7 PD GWAS risk locus. (A) Locus plot of PD-risk GWAS p-values *(2)* for chromosome 7 locus. **(B-J)** Locus plots showing p-values for GTEx *(10)* tissue-specific eQTL's in the caudate nucleus **(B-D)**, whole blood **(E-G)**, and substantia nigra **(H-J)** for potential target genes *GPNMB* **(B,E,H)**, *KLHL7* **(C,F,I)**, and *NUPL2* **(D,G,J)**. Posterior probability of colocalization (PP4) for each tissue-specific eQTL and PD-risk GWAS is specified on the upper right of each panel. Significant colocalization is defined as PP4 > 75%. Red font indicates significant colocalization of GWAS and genome-wide significant (p < 5E-8) eQTL signal. The purple diamond corresponds to the top GWAS SNP, rs199347. All other SNPs are colored according to their r^2 measure of linkage disequilibrium with rs199347. Dashed line indicates genomewide significance threshold of 5E-8.



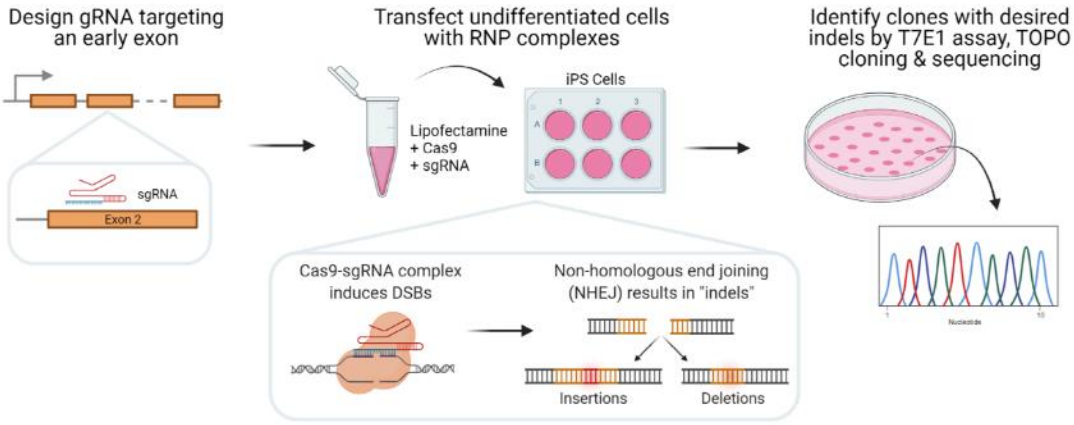
463

464 **Fig. S2: Characterization of GPNMB in HEK293 cells and human brain lysate.**

465 HEK293 cells over-expressing GPNMB and alpha-synuclein (SNCA) show **(A)** co-localization
466 of GPNMB (red, D9 antibody) and SNCA (green, MJFR1 antibody) on immunofluorescence
467 confocal microscopy, **(B)** co-IP of SNCA when GPNMB is IP'd (top, SNCA detected with
468 MJFR1 antibody), and co-IP of GPNMB when SNCA is IP'd (bottom, GPNMB detected with
469 E1Y7J antibody). For all IP's, a negative control IgG IP is shown, along with input (10% of IP),
470 and flow-through (FT) lanes. When GPNMB is IP'd, bands corresponding to higher-molecular
471 weight multimers or aggregates of SNCA are observed occasionally (red bar) in HEK293 cells,

whereas no other bands appear on the GPNMB immunoblots. In both HeLa cells **(C)** and human brain lysates **(D)**, GPNMB appears as 1-2 bands >95kD, collapsing to a single 72kD band with deglycosylation. For **(C)** and **(D)**, the order of the bands is as follows: sample heated to 95°C for 10 minutes, sample kept on ice (not boiled or N.B.), sample deglycosylated with PNGase F (N.B. De-glycos), and sample dephosphorylated with Lambda protein phosphatase (N.B. De-Phos).

A) CRIPSR-Cas9 KO Workflow:



B)

Cell Line	Indel	Result
1C (WT)	0	NA
6D (KO1)	-1	Early stop, truncated at AA **36**
	-7	Early stop, truncated at AA **41**
6H (KO2)	+1	Early stop, truncated at AA **49**
	-1	Early stop, truncated at AA **36**
2A (Het)	0	NA
	-18	Deletion of 6 AA's (PSAYMR) at position **34 - 39**

C)

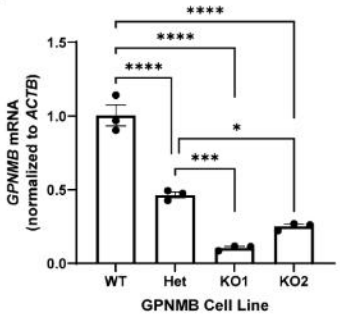
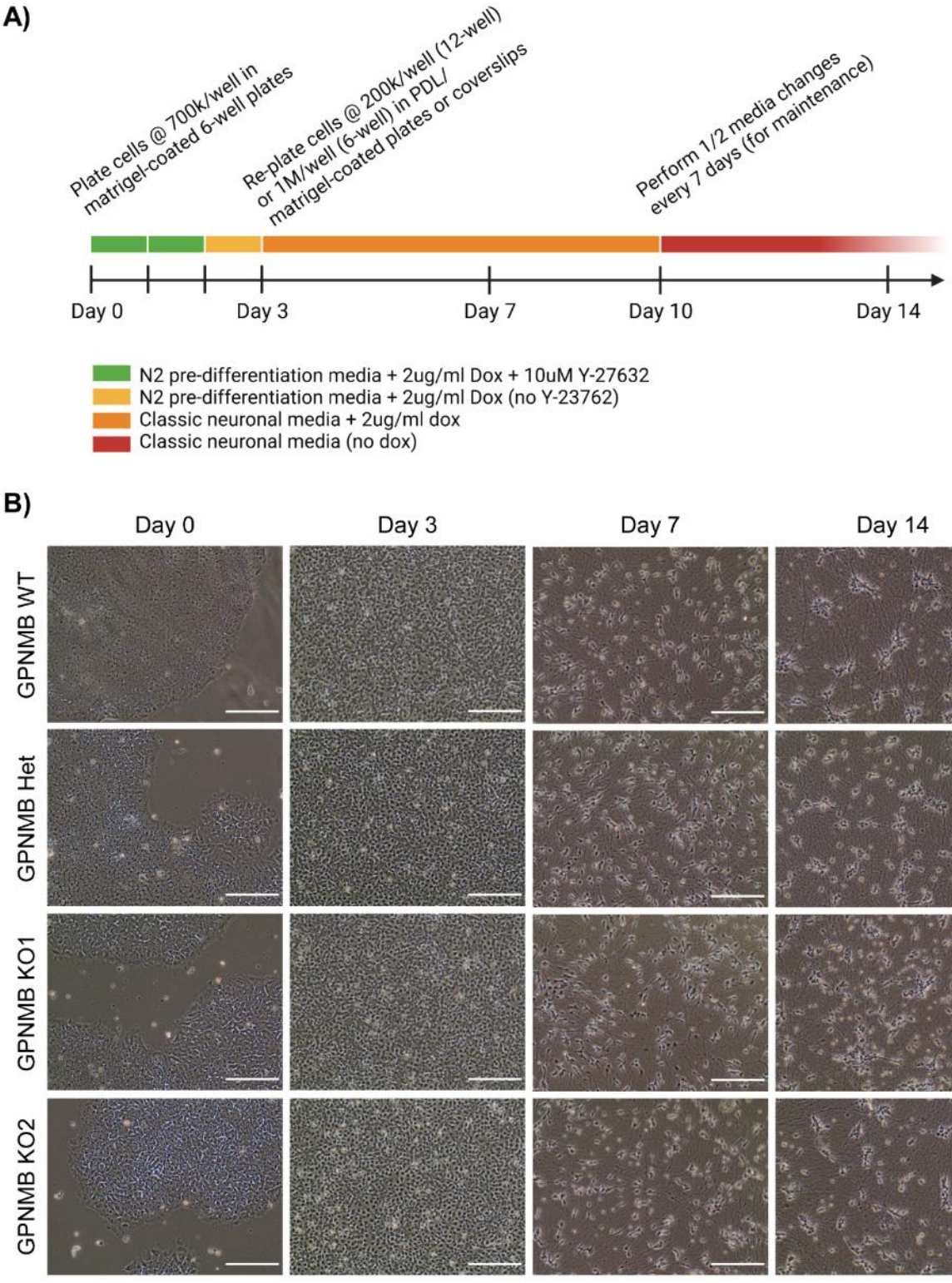


Fig. S3. CRISPR-Cas9 generation of GPNMB WT, Het, and KO iPSC lines.

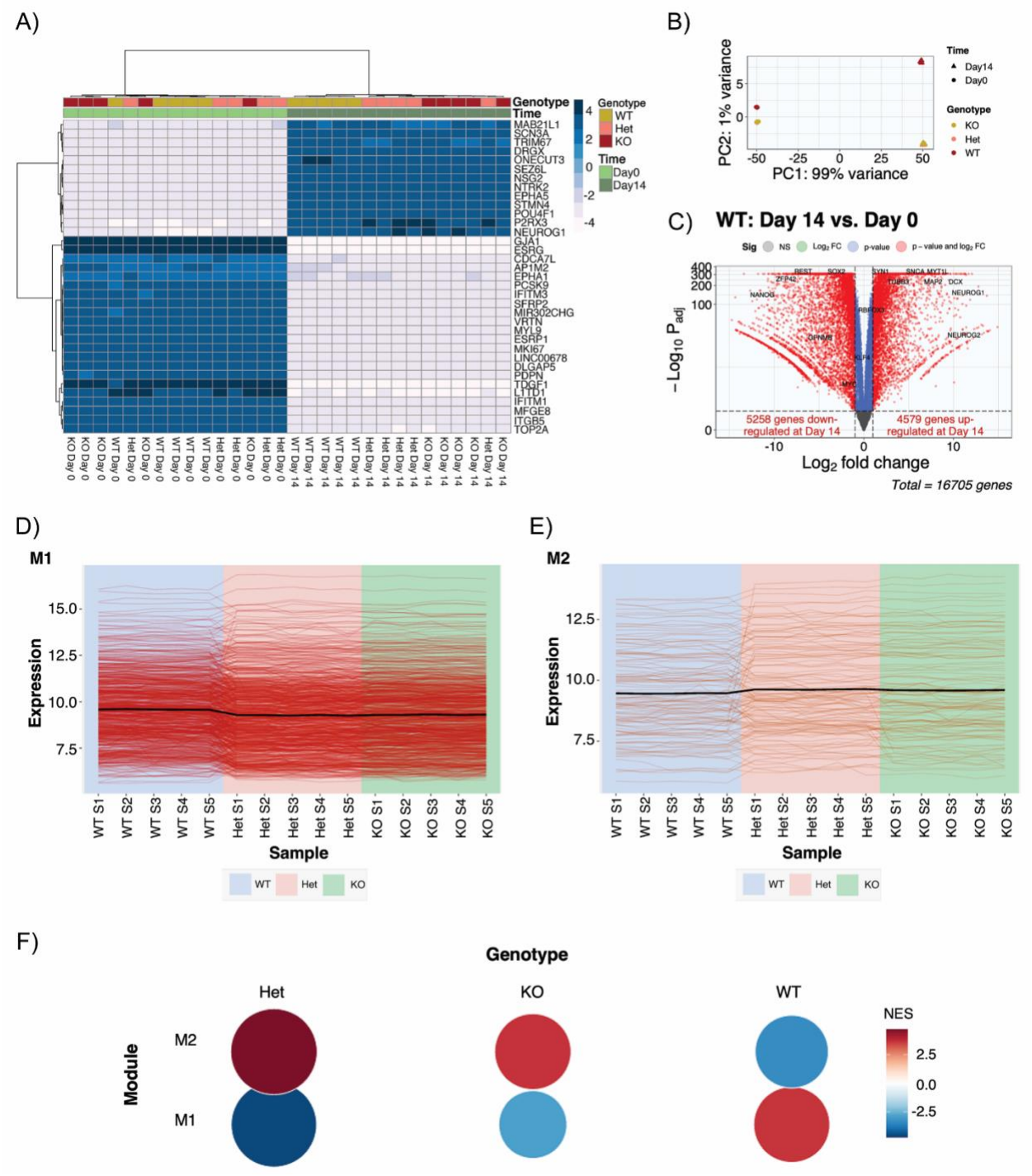
A) Schematic outlining experimental design. sgRNA targeting *GPNMB* exon 2 was designed using Synthego software. Cells were transfected with Cas9 RNP using lipofectamine STEM reagent. Desired clones were identified by T7E1 assay, TOPO cloning, and sequencing. Created with BioRender.com. **B)** Indel length and predicted protein translation products based on sequencing of individual strands for each cell line. **C)** qPCR validation of *GPNMB* knock-out in iPSC. n = 3 wells from independent differentiations per line. Statistical test used was one-way ANOVA, followed by post-hoc Tukey tests. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$.


A)
Plate cells @ 700k/well in matrigel-coated 6-well plates
Re-plate cells @ 200K/well (12-well) or 1M/well (6-well) in PDL/ matrigel-coated plates or coverslips
Perform 1/2 media changes every 7 days (for maintenance)
Day 0
Day 3
Day 7
Day 10
Day 14
N2 pre-differentiation media + 2ug/ml Dox + 10uM Y-27632
N2 pre-differentiation media + 2ug/ml Dox (no Y-23762)
Classic neuronal media + 2ug/ml dox
Classic neuronal media (no dox)
B)
Day 0
Day 3
Day 7
Day 14
GPNMB WT
GPNMB Het
GPNMB KO1
GPNMB KO2

493

Fig. S4. Differentiation of GPNMB-edited iPSC lines into neurons (iPSC-N).

A) Schematic outlining differentiation protocol for iPSC-N. Created with BioRender.com. **B)** Brightfield images of GPNMB WT, Het, KO1, and KO2 lines on days 0, 3, 7, and 14 after doxycycline induction. Scale bar = 200μm.



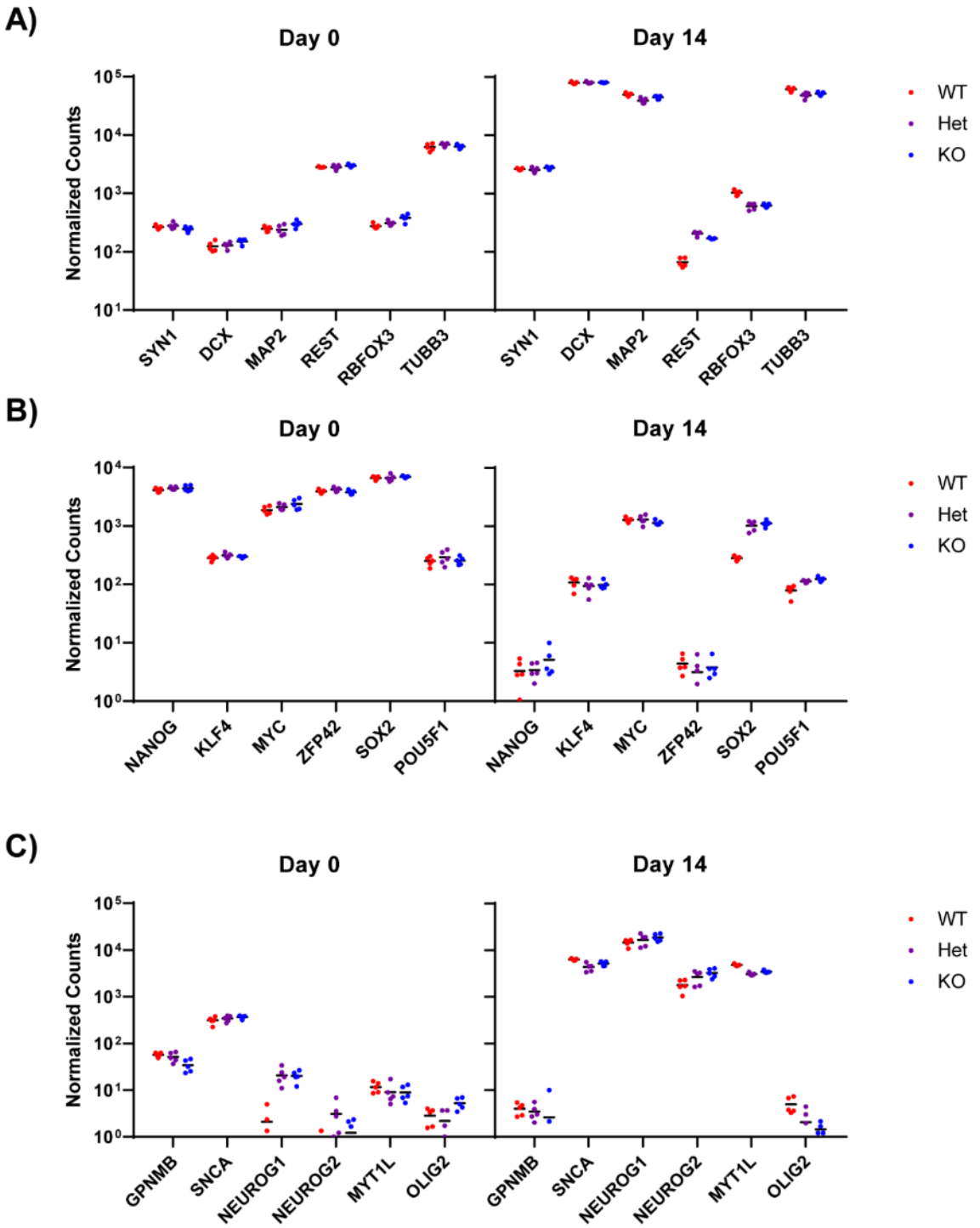
499

500 **Fig. S5. Additional RNAseq Analysis of Day 0 and 14 iPSC-N.**

501 **(A)** Heatmap showing clustering of days 0 and 14 iPSC-N using the top 35 most variable genes.

502 The first bar along the top row shows clustering of iPSC-N cells by GPNMB status within

timepoints, whereas the second bar shows stronger clustering by timepoint (day 0 or 14 after neuronal indeuction). Samples (columns) and genes (rows) were clustered by hierarchical clustering (hclust) using the complete linkage method and Euclidian distances. **(B)** Principal component analysis (PCA) using all expressed genes shows, as expected, highly differential gene expression in undifferentiated (day 0) vs. neuronally-differentiated (day 14) samples (separation along PC1). Additionally, while undifferentiated samples showed some separation by GPNMB KO status, this separation widened substantially with neuronal differentiation (day 14, separation along PC2). **(C)** Volcano plot showing genes with differing expression in day 14 (neuronal) cells compared to day 0 (undifferentiated) cells, independent of GPNMB status. The horizontal axis indicates the $\log_2$ fold-change ($\log_2$FC) in gene expression. The vertical axis indicates the $-\log_{10}$ of Benjamini-Hochberg adjusted p-value (P_{adj}). The horizontal dotted line indicates the significance threshold ($P_{adj} = 0.01$), with genes above this line showing significant differential expression (FDR < 0.01). The genes are red colored if they passed both thresholds ($P_{adj} < 0.01$ and $|\log_2 FC| > 1$), blue if they met only the P_{adj} criteria, green if they passed only the $\log_2$FC criteria, and dark gray if they met neither criterion. **(D-E)** Module gene expression profiles for module M1 **(D)** and M2 **(E)** show the individual gene expression profiles within modules. The colors represent GPNMB status (WT, Het, or KO). The median expression of all genes within the module is indicated by the black line. **(F)** Module activity for each GPNMB subtype (WT, Het, or KO). The enrichment score (ES) of a module is normalized (NES) by the number of genes in the module. Red color indicates higher activity and blue represents lower activity. The circle size and color intensity are proportional to the NES values.


A)
Day 0
Day 14
Normalized Counts
SYN1
DCX
MAP2
REST
RBFOX3
TUBB3
WT
Het
KO
B)
Day 0
Day 14
Normalized Counts
NANOG
KLF4
MYC
ZFP42
SOX2
POU5F1
WT
Het
KO
C)
Day 0
Day 14
Normalized Counts
GPNMB
SNCA
NEUROG1
NEUROG2
MYT1L
OLIG2
WT
Het
KO

525

526

Fig. S6. RNAseq expression of key genes over time.

(A-C) Normalized expression of neuronal differentiation markers (**A**), pluripotency markers (**B**), and other genes of interest (**C**) in GPNMB WT, Het, and KO iPSC-N at days 0 and 14 after dox induction. n = 5 samples from 3 independent differentiations per cell line per timepoint.

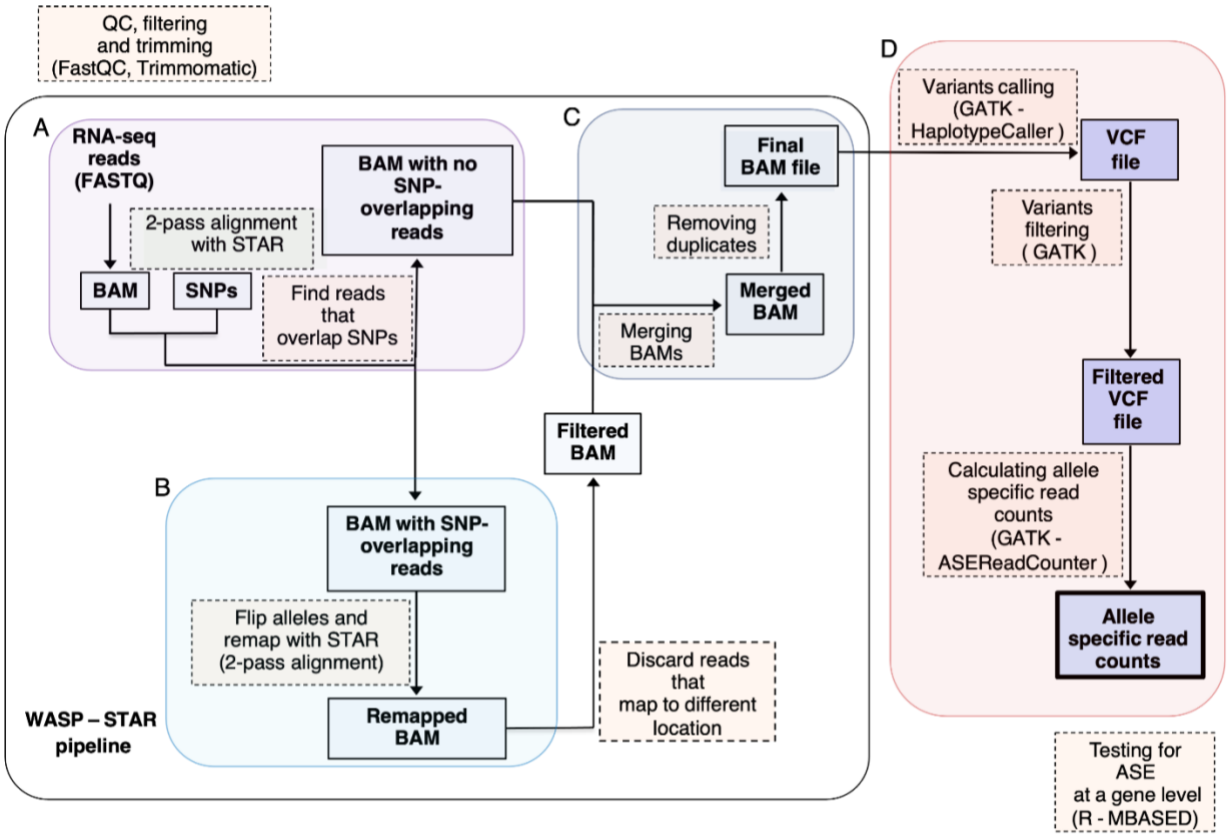


Fig. S7. Allele specific expression (ASE) workflow.

(A) RNA-seq data were first QC filtered and then mapped to the human reference genome (GRChg37) with STAR aligner using a 2-step approach. To remove allelic bias the aligned reads were split into reads that did and did not overlap with SNVs and into separate BAM files. **(B)** . The SNV genotypes were flipped in all reads that overlapped a SNV (in each read the genotype was swapped with that of the other allele) and re-mapped to reference genome again. Reads that did not re-mapped to the original location were discarded. (**C)** Filtered BAM files were then merged with BAM files that didn't overlap with SNVs. Duplicate reads were removed using *rmdup_pe.py* script incorporated into the WASP pipeline. **(D)** SNV were called and filtered using GATK tools. The allele specific counts were obtained by *ASEReadCounter* GATK tools. MBASED algorithm was used to test for ASE at the gene level.

Table S1.

Summary of demographics and brain regions used for human postmortem brain experiments. ASE=allele-specific expression, WB=Western blot, IHC=immunohistochemistry. The representative cases depicted in Figure 1's IHC are Patient ID 119534 and Patient ID 118273, but all 6 cases were stained and evaluated.

Assay	Patient ID	Clinical Diagnosis	Age at Death (yrs)	Sex	Disease Duration (yrs)	Caud-ate	Cingu-late	Cere-bellum	Tem-poral Cortex
ASE	100671	PD with Dementia	88	F	17	x	x	x	
	104173	PD with Dementia	73	M	14	x	x	x	
	105484	PD with Dementia	67	M	6	x	x	x	
	114691	PD with Dementia	72	M	8	x	x	x	
	101060	Normal	72	M		x		x	
	118624	Normal	70	M		x			
WB	118624	Normal	70	M		x			
	113756	Depression	43	M	6	x			
	112090	Normal	83	F		x			
	101060	Normal	72	M		x			
	116519	Normal	67	F		x			
	114653	PD with Dementia	80	M	14	x			
	104717	PD with Dementia	87	M	28	x			
	108649	PD with Dementia	72	M	23	x			
	116280	PD with MCI	82	F	23	x			
	109853	PD with Dementia	77	M	16	x			
IHC	121111	Normal	68	M			x		x
	117925	Normal	78	M			x		x
	119534	Normal	72	F			x		x
	125654	Parkinsonism	56	M	5		x		x
	118273	PD with Dementia	83	M	10		x		x

	121672	DLB	74	M	6		x		x

551
552

Table S4.

Cohort demographics for comparison of GPNMB plasma levels between PD and NC.

	Diagnosis		Statistics		
Variable	PD	NC	Test	P-value	Signif.
N	731	59	NA		
Sex (N)	Female: 247 Male: 484	Female: 35 Male: 24	Fisher's Exact	0.0002	***
Age (Yrs) Mean ± SD	67.64 ± 8.96	71.53 ± 7.84	T-test	0.0005	***
Disease Duration (Yrs) Mean ± SD	7.16 ± 5.94	NA			

n.s. = $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S5.

Cohort demographics for GPNMB plasma pQTL analysis.

	Diagnosis		Statistics		
Variable	PD	NC	Test	P-value	Signif.
N	724	32			
Sex (N)	Female: 246 Male: 478	Female: 21 Male: 11	Fisher's Exact	0.0005	***
Age (Yrs) Mean ± SD	67.69 ± 8.98	69.06 ± 6.65	T-test	0.2694	n.s.
Disease Duration (Yrs) Mean ± SD	7.15 ± 5.93	NA			

n.s. = $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S6.

Cohort demographics for comparison of GPNMB CSF levels between PD and NC.

	Diagnosis		Statistics		
Variable	PD	NC	Test	P-value	Signif.
N	72	40			
Sex (N)	Female: 37 Male: 35	Female: 20 Male: 20	Fisher's Exact	1	n.s.
Age (Yrs) Mean ± SD	66.46 ± 6.61	69.2 ± 6.70	T-test	0.0456	*
Disease Duration (Yrs) Mean ± SD	6.38 ± 4.42	NA			

n.s. = $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S7.

Cohort demographics for GPNMB CSF pQTL analysis.

	Diagnosis		Statistics		
Variable	PD	NC	Test	P-value	Signif.
N	70	31			
Sex (N)	Female: 36 Male: 34	Female: 18 Male: 13	Fisher's Exact	0.666	n.s.
Age (Yrs) Mean ± SD	66.44 ± 6.67	67.61 ± 6.18	T-test	0.395	n.s.
Disease Duration (Yrs) Mean ± SD	7.22 ± 4.51	NA			

n.s. = $p \geq 0.05$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S8.

Results for linear regression of plasma GPNMB ~ PD diagnosis + rs199347 genotype (codominant model) + age + sex.

Variable	Estimate (β)	Std. Error	t value	P-value	Signif.
(Intercept)	2.735117	0.132445	20.65103	2.21E-75	****
Sex (Male vs. Female)	0.042786	0.028966	1.477142	0.140057	
Age (Yrs)	0.000796	0.00155	0.51368	0.607627	
Cohort (PD vs. NC)	0.141329	0.068611	2.059848	0.039757	*
rs199347 (AA vs. GG)	0.044773	0.038803	1.153834	0.248935	
rs199347 (GA vs. GG)	0.061689	0.037284	1.654552	0.098433	.

. = p < 0.1, * = p < 0.05, ** = p < 0.01, *** = p < 0.001, **** = p < 0.0001

Table S9.

Results for linear regression of CSF GPNMB ~ PD diagnosis + rs199347 genotype (codominant model) + age + sex.

Variable	Estimate (β)	Std. Error	t value	P-value	Signif.
(Intercept)	1.82835	0.168931	10.82303	2.92E-18	****
Sex (Male vs. Female)	0.04961	0.029236	1.696886	0.092992	.
Age (Yrs)	0.010435	0.002301	4.534395	1.68E-05	****
Cohort (PD vs. NC)	0.06314	0.031755	1.988362	0.049649	*
rs199347 (AA vs. GG)	0.199285	0.043897	4.539812	1.65E-05	****
rs199347 (GA vs. GG)	0.123633	0.040998	3.015597	0.00329	**

. = p < 0.1, * = p < 0.05, ** = p < 0.01, *** = p < 0.001, **** = p < 0.0001

Table S10.

Results for linear regression of plasma GPNMB ~ UPDRS-PIII score + age + sex in the Discovery cohort.

Variable	Estimate (β)	Std. Error	t value	P-value	Signif.
(Intercept)	2.611333	0.310946	8.398015	5.32E-13	****
UPDRS-PIII	0.007614	0.00268	2.841321	0.005532	**
Age (Yrs)	0.001364	0.004803	0.283907	0.777119	
Sex (Male vs. Female)	0.075098	0.078812	0.952885	0.343145	

. = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S11.

Results for linear regression of plasma GPNMB ~ UPDRS-PIII score + age + sex in the Replication cohort.

Variable	Estimate (β)	Std. Error	t value	P-value	Signif.
(Intercept)	3.953842	0.090524	43.67715	7.65E-63	****
UPDRS-PIII	0.001392	0.000774	1.797537	0.075567	.
Age (Yrs)	-0.00042	0.001328	-0.31824	0.751034	
Sex (Male vs. Female)	-0.01343	0.018397	-0.73021	0.46714	

. = $p < 0.1$, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$

Table S2.

Module 1 genes in top 10 categories from overrepresentation analysis.

Table S3.

Module 2 genes in top 3 categories from overrepresentation analysis.

Table S12.

Summary of Illumina NovaSeq RNAseq run.

Data S1.

Cell profiler pipeline used for image processing.