Protein homeostasis of a metastable subproteome associated with Alzheimer’s disease

Rishika Kundra, Prajwal Ciryam, Richard I. Morimoto, Christopher M. Dobson, and Michele Vendruscolo

Alzheimer’s disease is the most common cause of dementia. A hallmark of this disease is the presence of aberrant deposits containing the Aβ peptide (amyloid plaques) and the tau protein (neurofibrillary tangles) in the brains of affected individuals. Increasing evidence suggests that the formation of these deposits is closely associated with the age-related dysregulation of a large set of highly expressed and aggregation-prone proteins, which make up a metastable subproteome. To understand in more detail the origins of such dysregulation, we identify specific components of the protein homeostasis system associated with these metastable proteins by using a gene coexpression analysis. Our results reveal the particular importance of the protein trafficking and clearance mechanisms, including specific branches of the endosomal-lysosomal and ubiquitin-proteasome systems, in maintaining the homeostasis of the metastable subproteome associated with Alzheimer’s disease.

Significance

Alzheimer’s disease is a neurodegenerative disorder whose molecular origins have been associated with the dysregulation of a set of metastable proteins prone to aggregation. Under conditions of cellular and organismal health, the protein homeostasis system prevents effectively the misfolding and aggregation of these metastable proteins. Although it is well established that such regulatory mechanisms become progressively impaired with aging, resulting in an accumulation of protein deposits, the specific nature of such impairment has remained incompletely characterized. Through a gene coexpression analysis, here we identify the endosomal-lysosomal and ubiquitin-proteasome systems, and more generally the protein trafficking and clearance mechanisms, as key components of the protein homeostasis system that maintains the metastable proteins in their functional states.

Author contributions: R.K., P.C., R.I.M., C.M.D., and M.V. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Freely available online through the PNAS open access option.

1To whom correspondence should be addressed. Email: mv245@cam.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1618417114/-/DCSupplemental.
of a set of aggregation-prone proteins in the proteome (29). Many of the proteins involved in this metastable subproteome are components of the mitochondrial respiratory chain, an observation consistent with the well-characterized mitochondrial disruption associated with neurodegenerative disorders and specifically with AD (39). We refer to the proteins expressed by this subset of genes as the “AD metastable subproteome.” The primary aim of the present study is to understand the different ways in which these metastable proteins are controlled, as illustrated schematically in Fig. 1. Our goal is thus to identify the specific components of the protein homeostasis system that are most closely involved in the regulation of the AD metastable subproteome.

Coexpression Analysis of the AD Metastable Subproteome and Its Associated Protein Homeostasis Components. Because the proteins in the AD metastable subproteome are intrinsically aggregation-prone, we searched for the specific protein homeostasis components that maintain the solubility and folding of these proteins. We therefore set out to identify an “AD metastable network” as a network of genes that encode for the AD metastable subproteome and its associated protein homeostasis components and that are correlated with the disease status (Materials and Methods).

To find the AD metastable network, we first carried out a weighted gene correlation network analysis (WGCNA) (33, 34) of the set of metastable proteins that we previously identified (29) and of the known components of the overall protein homeostasis system (23) (Materials and Methods). WGCNA is a robust method of performing gene coexpression analysis that has been shown to be particularly effective when large transcriptional datasets are available (33). As our aim was to study how metastable proteins are regulated across health and disease, we pooled together extensive microarray data obtained from postmortem brain tissues of patients diagnosed with late-onset AD (LOAD) and of matched controls (40) (Dataset S1, Table S1 and Materials and Methods).

WGCNA uses the Pearson’s coefficient of correlation between each pair of genes and their “topological overlap,” which is a measure of their connectivity based on their shared neighbors, to identify biologically meaningful groups of coexpressed genes; these groups are called “modules” and labeled by different colors (34) (Dataset S1, Tables S2 and S3 and Materials and Methods).

As WGCNA captures the underlying network structure in large-scale gene expression studies, it has been used to study the global changes associated with a range of disease states, with the preservation of groups of coexpressed genes across species, and with the identification of hub genes associated with particular traits (28, 38, 41–44). We observed that the genes encoding for metastable proteins and for certain components of the protein homeostasis network are organized into well-defined modules (Fig. 2A, with the genes encoding for proteins in the metastable subproteome shown in pink in each module), where each module consists of tightly coexpressed genes. We found that the majority of metastable proteins belong to four specific modules, and we refer to them as “modules enriched in metastable proteins” (MEMPs)—MEMP-1 (blue), MEMP-2 (turquoise), MEMP-3 (green-yellow), and MEMP-4 (black) (Fig. 2A)—which consist of 659, 688, 35, and 74 genes, with 220, 91, 10, and 10 of these genes corresponding to metastable proteins, respectively (Dataset S1, Table S3).

Identification of an AD Metastable Network. As the next step to identify an AD metastable network, we performed an analysis to identify a module eigengene (ME) for each module, which is the first principal component (PC) of the expression values across genes in each module (Materials and Methods). The ME therefore provides a representative value for the expression of a group of genes in a particular module (34). This approach offers a significant advantage in correlating pairs of modules, as it eliminates the problem of multiple testing and noise by reducing the number of comparisons to just one instead of several hundreds. The higher the value of the Pearson’s coefficient of correlation between two MEs, the more closely the two modules are related (Materials and Methods).

We therefore identified which of these modules showed the most significant relationship to the disease status by looking at the correlation between the MEs and disease status (Materials and Methods). We found six modules (Fig. 2B) to be significantly correlated with disease status, of which three modules (magenta, pink, and red) had very few metastable genes (4, 1, and 1, respectively). We excluded these modules from further analysis, as they mostly contained components of the protein homeostasis machinery whose expression levels do not correlate well with the metastable genes. The other three modules (MEMP-1, MEMP-2, and MEMP-3) were those most enriched for metastable genes that we described in the previous section (Fig. 2B).

As these three MEMPs were the only modules both significantly correlated with disease status and significantly enriched in the genes encoding for metastable proteins (Fig. 2B), we chose them for further analysis. They are also, in fact, closely related to each other based on the correlation of their MEs with a Pearson’s correlation coefficient of 0.78 between MEMP-1 and MEMP-2 and 0.68 between MEMP-2 and MEMP-3 (Dataset S1, Table S4). To control for possible biases of the modules because of the use of a particular dataset, we cross-validated the results of module detection with a hippocampal gene expression dataset, as the hippocampus is among the regions typically most affected in AD (35). We observed that the MEMPs, along with most of the other modules, were well preserved between the two datasets (Fig. S1).

We then performed a gene ontology enrichment analysis to characterize these modules, finding that protein ubiquitination was the most enriched GO term for MEMP-1, MEMP-2, and MEMP-3 (Fig. S2). We also asked if the genes contained in these
modules are overrepresented in any biochemical pathway. To this end, by analyzing the KEGG biochemical pathways (45), we found that they are strongly overrepresented in the pathways associated with the AD metastable subproteome and the ubiquitin-proteasome and endosomal–lysosomal systems (Fig. S3).

Based on these results, we identified the AD metastable network as the set of genes in MEMP-1, MEMP-2, and MEMP-3 (Dataset S1, Table S2).

Identification of the Hub Genes and of Their Roles in the AD Metastable Network. Because any given module is comprised of a large number of genes, it is helpful to identify the most highly connected genes within a particular module. These “hub” genes are likely to reveal the main processes carried out by the AD metastable network, even if individually they may not be crucial for the disease itself. To achieve this goal, we defined the “module membership” (MM) score by using the intramodular connectivity (kME; Materials and Methods), which is a measure of how strongly connected—that is, coexpressed—a given gene is to all of the other genes in a module (34). Hub genes were defined as those genes having an absolute kME value greater than 0.8.

The hub genes in the AD metastable network were found to be highly enriched in the KEGG biochemical pathways of cellular degradation (proteasome and ubiquitin-mediated proteolysis) and trafficking in addition to those previously associated with metastable proteins such as oxidative phosphorylation, AD, PD, and HD (Fig. 3A). These results are fully consistent with those reported above for the full list of genes in the AD metastable network (Fig. S3). In any given module, a high mean MM value indicates how tightly coexpressed the genes are within that module. We observed that the genes encoding for the AD metastable subproteome in the AD metastable network have a significantly high mean MM value that is significantly greater than that of other genes in that module (Fig. 3B). In addition, more than two thirds of the genes encoding for metastable proteins in these modules are hub genes, indicating their central importance in their respective modules.

Test of Module Generality Using a Consensus Network Analysis with a Visual Cortex Dataset. We next sought to determine whether the modules that we identified are general or instead specific to the dataset or brain region that we analyzed. To check the robustness of the modules identified in this study, we constructed a consensus network (Materials and Methods) using WGCNA on another dataset from the visual cortex (VC) of AD patients and healthy controls (40), along with the dataset for the dorsolateral prefrontal cortex (PFC) used previously (Materials and Methods), to examine whether or not our network is preserved. To assess the level of preservation, we used the “consensus network” construction, which identifies groups of genes that are tightly coexpressed across multiple studies (46). The consensus MEs (consMEs) represent modules in each of the two sets (46) (Materials and Methods). Each gene is assigned to a single consensus module, but there are two sets of consMEs for each module as a given module can have a different expression profile in the two datasets. We found that all of the modules identified in our study have a consensus counterpart in the VC dataset, indicating that the module structure in the two datasets is similar (Fig. S4).

We then constructed the two sets of eigengene dendrograms and eigengene heatmaps based on the consMEs (one for each study), and the results indicate that the overall modular structure in the two sets is quite similar. The preservation heatmap shows the preservation network, defined as one minus the absolute difference of the eigengene networks in the two datasets (Fig. S5A). The overall degree of preservation between the two networks is 0.87, and the mean preservation of relationships for each eigengene is consistently high for all of the modules except the “red” one, as shown by the preservation heatmap and bar plot (Fig. S5B), thus indicating that the modules identified in the analysis detailed in the study are highly robust.

These results suggest that the difference between a healthy state and a disease state does not involve a reorganization of the modules but rather a variation in the expression levels of specific genes within the modules. In the following, we therefore carried out further investigations to identify such genes.

Protein Homeostasis of the AD Metastable Subproteome. We next asked the central question of this work—How is the AD metastable subproteome regulated? To answer this question, we analyzed which components of the protein homeostasis system are coexpressed with the AD metastable subproteome in the AD metastable network, as we expect that the knowledge of such components could offer insight into the regulation of these metastable proteins (Fig. 4 and Fig. S6). To this end, we identified the most important hub genes, among those described above, by visualizing them within the AD metastable network; we used the Cytoscape software for
Although the present analysis of PTK2B to BIN1 (65) to the regulation of KSTRU et al.0.05. (Identification of KEGG biochemical pathways enriched in hub genes > detail. thesis; Fig. 4 components (autophagy, metabolism, signaling, and protein synthesis) close link with AD(48 degradation are components of the protein homeostasis system consistent with experimental evidence that protein trafficking and as the most connected hub genes (Table 1). These results are further assess the significance of our analysis, we compared our results to the one that we have used here to associate genes with metastability to aggregation of their products. These two approaches this finding, it has been suggested that this protein is involved in a regulatory mechanism that responds to increased protein accumulation (55). In addition, overexpression of RAB1, another small GTPase closely related to RAB6A, has been shown to alleviate ER stress in yeast models of PD (56). Hence, RAB6A, which is a central gene in the AD metastable network described in this work, could play an important role in the regulation of the metastable proteins by directing them toward the endosomal-lysosomal degradation machinery, thereby preventing their accumulation in the cytoplasm. Another two genes in the group that we found are ATP6V1H, which encodes a protein subunit of a vacuolar ATPase involved in clathrin-mediated endocytosis (57, 58) and whose role in regulating lysosomal pH has been recently been linked to neurodegeneration (59), and ATTL1, which is involved in ER trafficking (60, 61). In fact, all 10 genes that we found to be related to trafficking are part of the endosomal-lysosomal system. Specifically, SH3GL2, SLC9A6, and CLTA are localized in the endocytic vesicle membrane (62, 63), and NSF is involved in vesicle-mediated transport and acts as a fusion protein through the SNARE proteins (64). Our results, therefore, indicate the importance of the endosomal–lysosomal system in controlling the metastable subproteome. These findings extend the well-known role of this system in the processing of Aβ(65) to the regulation of a broader range of aggregation-prone proteins.

**Ubiquitin–Proteasome System.** Among the genes associated with the ubiquitin–proteasome system, we found ENCB, which is an actin binding protein that has been reported to modulate the aggregation of mutant huntingtin in yeast under ER stress (66). MYCBP2, FBXL2, and RNF128 are E3 ubiquitin ligases and are essential components of the ubiquitin-dependent degradation of proteins (67–69). These results indicate that metastable proteins are likely to be regulated upstream of the proteasomes at the ubiquitin ligase stage.

**Molecular Chaperones.** We also found a number of components of molecular chaperone networks coexpressed with the AD metastable subproteome (Dataset S1, Table S5). Such components include co-Hsp70/Hsp90 species, which are known to assist the Hsp70/Hsp90 system to degrade protein aggregates (70, 71). Among such molecular chaperones, we found DNAJC6, a J-domain cochaperone with a role in HSC70-mediated uncoating of the clathrin-coated vesicles in neurons by recruiting HSC70. Also seen as hub genes were TOR1A, with chaperone activity and a member of the AAA family of ATPases, and ERLEC1, which has a role in ER quality control (72, 73).

Taken together, these results indicate that the components of the AD metastable subproteome, which consists of proteins inherently at risk of aggregation, tend to be highly coexpressed with multiple components of the protein homeostasis system. These findings illustrate how during the course of AD, when a dysregulation and collapse of these systems is increasingly likely to occur, these metastable proteins are likely to represent an enhanced risk due to the dysfunction of the regulatory mechanisms associated with their folding, transport, and degradation.

**Relationship with Genome-Wide Association Studies (GWAS).** To further assess the significance of our analysis, we compared our results with genetic loci identified by GWAS. These studies have reported that several genes associated with the trafficking and degradation systems are closely associated with AD (74, 75). In particular, seven GWAS genes (PICALM, SORL1, CD33, BIN1, CD2AP, ABCA7, and RIN3) are associated with the endosomal-lysosomal system, and two GWAS genes (CLU and PTK2B) are associated with the ubiquitin–proteasome pathway (74, 75). These results are highly consistent with the conclusions of the present study, as 17 GWAS genes (among the 28 that we considered) are present in the AD metastable network identified in this work (Fig. 5). This consistency is remarkable, as the GWAS strategy, where genes are typically associated with disease on the basis of single nucleotide polymorphism (SNP) statistics, is independent from the one that we have used here to associate genes with disease through the combination of their coexpression and the metastability to aggregation of their products. These two approaches...
are therefore complementary, as a coexpression analysis can identify a large number of genes and therefore reveal the biochemical pathways involved in the disease and help rationalize the specific roles of the GWAS genes but may not capture important relationships, such as in the present case the role of ADAM10, PSEN1, and PSEN2 in the processing and regulation of APP (Fig. 5).

Consensus Network Analysis of AD, PD, and HD. As noted above, the phenomenon of protein misfolding and aggregation is a common feature of many neurodegenerative disorders, including AD, PD, HD, and ALS. Although these diseases are characterized by a variety of different clinical manifestations and features, there is increasing interest in understanding the extent to which they share common molecular origins (1–9). To address this question in the present context, we investigated whether or not the regulation of the metastable proteins, in terms of their coexpression with specific protein homeostasis components, is similar across AD, PD, and HD.

Because oxidative phosphorylation is the most significantly enriched pathway among the metastable genes, we analyzed the coexpression of genes involved in this specific pathway and in the protein homeostasis components. We built a consensus network for gene expression data (Dataset S1, Table S6) from hippocampal tissue (35), substantia nigra (76), and PFC (73), obtained postmortem from patients diagnosed with AD, PD, and HD, respectively, and from age-matched controls. The network heatmaps indicate the correlation of various eigengenes within the AD, PD, and HD networks (Fig. S7) and the preservation heatmaps (Fig. 6) reveal that the overall preservation of the three networks is highly significant (shown in red). The mean preservation of the three networks exceeds 0.7 in all three cases (Fig. S7), indicating that the global structures of the coexpression

![Diagram](image_url)

**Table 1. List of hub genes used to identify the components of the protein homeostasis system associated with the AD metastable subproteome**

<table>
<thead>
<tr>
<th>Hub genes</th>
<th>Proteins</th>
<th>Known functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP6V1H</td>
<td>V-type proton ATPase subunit H</td>
<td>Clathrin coated endocytosis, formation of endosomes</td>
</tr>
<tr>
<td>SH3GL2</td>
<td>Endophilin-A1</td>
<td>Synaptic vesicle endocytosis</td>
</tr>
<tr>
<td>SLC9A6</td>
<td>Sodium/hydrogen exchanger 6</td>
<td>Exchange of protons across the membrane of early and recycling endosome</td>
</tr>
<tr>
<td>RAB6A</td>
<td>Ras-related protein Rab-6A</td>
<td>Retrograde transport from Golgi to ER, transport from endosome to plasma membrane</td>
</tr>
<tr>
<td>CDH13</td>
<td>Cadherin-13</td>
<td>Regulation of endocytosis</td>
</tr>
<tr>
<td>RBFX1</td>
<td>RNA binding protein fox-1 homolog 1</td>
<td>RNA binding protein, regulation of alternative splicing events</td>
</tr>
<tr>
<td>CLTA</td>
<td>Clathrin light chain A</td>
<td>Major protein of the polyhedral coat of coated pits and vesicles</td>
</tr>
<tr>
<td>NSF</td>
<td>Vesicle-fusing ATPase</td>
<td>SNARE binding, regulation of exocytosis</td>
</tr>
<tr>
<td>CNK2</td>
<td>Connecter enhancer of kinase suppressor of ras 2</td>
<td>Adaptor protein, regulation of signal transduction</td>
</tr>
<tr>
<td>ATL1</td>
<td>Atlastin-1</td>
<td>ER to Golgi vesicle transfer</td>
</tr>
<tr>
<td>ENC1</td>
<td>Ectoderm-neural cortex protein 1</td>
<td>Proteosomal ubiquitin-independent protein catabolic process</td>
</tr>
<tr>
<td>MYCBP2</td>
<td>E3 ubiquitin-protein ligase MYCBP2</td>
<td>Ubiquitin ligase, protein ubiquitination</td>
</tr>
<tr>
<td>FBXL2</td>
<td>F-box/LRR-repeat protein 2</td>
<td>Ubiquitin ligase, protein ubiquitination</td>
</tr>
<tr>
<td>RFPL1</td>
<td>Ret finger protein-like 1</td>
<td>Zinc ion binding</td>
</tr>
<tr>
<td>RNF128</td>
<td>E3 ubiquitin-protein ligase RNF128</td>
<td>Ubiquitin ligase, ubiquitin-dependent protein catabolic process</td>
</tr>
<tr>
<td>TUSC3</td>
<td>Tumor suppressor candidate 3</td>
<td>Magnesium transporter</td>
</tr>
</tbody>
</table>

These hub genes are shown in Fig. 4 and are reported here together with their corresponding proteins and their known functions. The list of hub genes corresponding to the AD metastable subproteome is reported in Dataset S1, Table S7.
Fig. 5. The majority of GWAS genes are found in the AD metastable network. Shown are the number of genes identified by GWAS (74, 75) that are present in the AD metastable network or in the other modules described in this work. Seventeen out of 28 genes identified by GWAS are present in the AD metastable network. The names of the genes are shown in their respective modules. The genes shown in red belong to the ubiquitin–proteasome system, and those in green belong to the trafficking system.

networks are similar for the three diseases. These results thus suggest that the differences between these diseases may be found in the dysregulation of specific genes within the consensus network (Fig. 6 and Fig. S7).

Discussion

Specific Components of the Protein Homeostasis System That Regulate Protein Aggregation. In this work, we have taken the view that a major hallmark of aging and neurodegeneration is the progressive impairment of the balance between protein aggregation and its control by the protein homeostasis system, which leads to the characteristic accumulation of aberrant protein aggregates (1–17, 21–29) (Fig. 1). In this context, we have previously reported that large numbers of proteins are inherently metastable to aggregation because of their elevated expression levels relative to their solubilities (16, 17). We have also observed a specific transcriptional down-regulation of genes encoding these proteins in AD (29) as well as a tissue-specific vulnerability to AD caused by an imbalance between aggregation-prone proteins and their protein homeostasis regulators (22).

To identify the protein homeostasis mechanisms that control the metastable proteins associated with AD, in this study we have analyzed together a set of proteins inherently prone to aggregation (29) and a set of proteins that make up the overall protein homeostasis system (23). Our analysis started from a metastable subproteome corresponding to the overlap between genes encoding for proteins that are supersaturated and transcriptionally down-regulated in AD but not in aging (29). We then constructed an AD metastable network composed of genes encoding this set of metastable proteins together with the corresponding components of the protein homeostasis system. We have found that this specific AD metastable network consists of well-defined modules of coexpressed genes (Fig. 2), enabling us to identify key players of the ubiquitin–proteasome and endosomal–lysosomal systems, along with some specific molecular chaperones (Fig. 4).

The systems-level approach that we have adopted in this work provides an understanding of the regulation of the AD metastable subproteome as a whole, as opposed to the regulation of individual proteins by specific components of the protein homeostasis system. Our results show that, from a list of about 2,000 components of the protein homeostasis system (23), just a relatively small number of specific proteins in the degradation and trafficking machinery along with specific molecular chaperones are primarily responsible for handling the metastable proteins with a high propensity to misfold and aggregate (Fig. 7).

Endosomal–Lysosomal and Ubiquitin–Proteasome Regulation of the Proteins Involved in Oxidative Phosphorylation. The expression levels of most of the components of the protein homeostasis system identified in this study have been previously seen to decrease with aging (29). Hence, during aging and disease, with the suppression of the protein homeostasis system, these proteins could become particularly vulnerable to aggregation because of their inherent metastability. Because these proteins perform fundamental functions, including in particular energy metabolism through oxidative phosphorylation (29), their aggregation could result in triggering a cascade of events contributing to disease pathology and ultimately to neuronal death. In addition, such dysregulation poses a pronounced threat to neurons due to their postmitotic state and increased dependence on mitochondria for energy production. Indeed, there is a substantial overlap in the genes involved in the pathways associated with oxidative phosphorylation and AD, HD, and PD (Dataset S1, Table S6), indicating again that the proteins encoded by these genes are highly metastable and hence are significant in the context of disease pathology. Mitochondria play a central role in aging and in regulating cell death (39, 77) as well as in the overall maintenance of cellular health. Whether mitochondrial dysfunction is the cause or effect of the disease pathology is still, however, unclear. Mitochondria have been shown to interact with aggregation-prone proteins, including α-synuclein and Aβ. More specifically, Aβ has been shown to be localized on the mitochondrial membrane in a transmembrane arrested form, possibly disrupting protein import into the mitochondria (78, 79).

Our results also point to a possible dependence of the proteins in the respiratory chain complex on the endosomal–lysosomal system, which we identified using the hub genes RAB6A, ATP6V1H,
ATL1, SH3GL2, SLC9A6, and CLTA, and on the ubiquitin–proteasome system, where the hub genes are the E3 ubiquitin ligases MYCBP2, FBXL2, and RNF128 (Fig. 4). These indications are consistent with the observation in yeast that accumulation of mitochondrial proteins in the cytoplasm leads to activation of the unfolded protein response (80). Furthermore, recent studies have reported the presence of polyubiquitinated mitochondrial proteins, suggesting that they are substrates of the ubiquitin proteasome system (81, 82) and that in yeast the expression of the proteasome is up-regulated upon cytoplasmic accumulation of mitochondrial proteins (80, 83). If mitochondrial import is disrupted and these metastable proteins therefore accumulate in the cytoplasm, the cell responds by clearing them through degradation. When, however, this disruption happens in an environment where protein homeostasis is compromised, these proteins would be particularly at danger of aggregation. We observed, in addition, a similar pattern of coexpression of genes encoding for mitochondrial membrane proteins across AD, PD, and HD, indicating that even though the initial cause of dysregulation might be different, these diseases are likely to share common molecular mechanisms at a later stage of progression, with regulation of mitochondrial membrane proteins playing an important role.

The finding that a metastable subproteome that is specifically associated with AD is primarily regulated by the protein trafficking and degradation systems provides important insights into the control of protein misfolding in this disease. These results suggest that in a setting of compromised protein folding, the maintenance of proteins in their soluble states may move away from regulating conformations and toward regulating concentrations.

Overall, the results that we have reported suggest an extension of the view that AD is associated with an age-related protein homeostasis failure that results in the aggregation of Aβ and tau—thus, neuronal dysfunction appears to be a consequence of a loss of ability of major branches of the protein homeostasis system to regulate a much wider group of aggregation-prone proteins making up a metastable subproteome.

Conclusions
We have described specific components of the protein homeostasis system that regulates a metastable subproteome associated with AD. This analysis has revealed the central roles of the ubiquitin–proteasome and endosomal–lysosomal degradation pathways, whose relevance to AD is well known (48–54), in the maintenance of a pool of metastable proteins prone to aggregation. By identifying a series of regulatory pathways associated with AD, these findings also help to rationalize the roles in the disease of the individual genes resulting from GWASs. We anticipate that an increasingly detailed understanding of the mechanisms of regulation of the metastable subproteome will contribute to the development of therapeutic strategies against neurodegenerative diseases aimed at promoting the maintenance of aggregation-prone proteins in their soluble states.

Materials and Methods
Dataset Acquisition. Microarray data for brain tissues from postmortem AD patients and healthy controls were downloaded from the Gene Expression
Omnibus (GEO) database (84). The following datasets were used for analysis (Dataset S1, Table S1): GSE44770, containing tissues derived through autopsy from the substantia nigra of PD patients and from healthy controls; GSE1297, containing hippocampal gene expression data from LOAD patients and from healthy controls; GSE33000, containing dorsolateral PFC tissue from HD patients and from healthy controls obtained from the Harvard Brain Tissue Resource Center (HBTRC); and GSE20292, containing postmortem brain tissue from the substantia nigra of PD patients and from healthy controls. Using the GEOQuery package, data were downloaded into R and checked for missing values (33).

Sample Clustering. Samples in each dataset were hierarchically clustered within GEOquery to detect outliers. One sample from GSE44771 (GSM1090949) and one sample from GSE20292 (GSM508732) were found to be outliers and hence removed from further analysis.

Generation of a “Weighted Gene Correlation Network.” A distance measure commonly used for coexpression analysis is based on the Pearson’s coefficient of correlation; in this approach, gene pairs with a coefficient of correlation below a given cutoff value (e.g., 0.8) are considered as not correlated. However, this kind of “hard thresholding” may be insensitive to subtle and yet important expression patterns (85). We therefore used the WGCNA method (33, 34), which uses a “soft thresholding” and the concept of topological overlap or shared neighbors to identify clusters of coexpressed genes. The soft thresholding method assigns a weight to each pair of interacting genes and uses such weights, along with the topological overlap, to identify modules of coexpressed genes in the expression data (33, 34).

The construction of a Weighted Gene Correlation Network was performed using the R package for WGCNA (33). Absolute values of Pearson’s coefficient of correlation were calculated for the expression values of each gene pair across all microarray samples. WGCNA uses a power function to transform the coexpression similarities (given by a similarity matrix $S = [s_{ij}]$) into connection strengths (given by an adjacency matrix $A = [a_{ij}]$):

$$a_{ij} = |s_{ij}|^\beta,$$

where $\beta$ is the soft thresholding power. In unweighted networks, the entries $a_{ij}$ of the adjacency matrix are either 1 or 0, indicating whether or not a pair of nodes is connected. In weighted networks, the values are real numbers ranging from 0 to 1. Due to the noise in microarray data and the limited number of samples, we weighted the Pearson’s coefficients of correlation by taking their absolute values and raising them to the power $\beta$. To choose the value of $\beta$, we observed that many biological networks, especially gene expression networks, have been found to exhibit approximate scale-free topology (86)—that is, the connectivity distribution $p(k)$ for each node $k$ follows a power law, $p(k) \sim k^{-\gamma}$, with exponent $\gamma$. This “scale-free” relationship indicates that there are a few nodes that are highly connected, whereas others have much fewer connections. Through these considerations, we chose $\beta = 3$ (34). This procedure results in a weighted network in which the continuous nature of the gene-expression values is preserved (as opposed to unweighted networks); the results are robust with respect to the choice of $\beta$, as opposed to the high sensitivity to the cutoff value of unweighted networks.

Identification of Modules in the Weighted Gene Correlation Network. Modules were defined as groups of genes having high correlation and high topological overlap (34). The topological overlap of two nodes refers to their interconnectedness, which is measured as the number of shared neighbors between two nodes. It provides a similarity measure that has been shown to be very useful in biological networks (87) and was used here as the basis for average linking hierarchical clustering to identify modules of coexpressed genes.

MEs. The ME, which is defined as the first PC of a given module, can be considered as a representative of the gene expression profiles in a module (33). The connectivity of a gene $v$ with a module $k$ ($MM_v$) is defined as the Pearson’s coefficient of correlation of the expression value of that gene with the ME of the module. It is a measure of MM for a particular gene. Specifically:

$$MM_v(k) = \text{cor}(\text{e}(v), e_k),$$

where $MM_v(k)$ is a measure of MM for gene $v$ with respect to module $k$, $e(v)$ is the expression profile of gene $v$, and $e_k$ is the eigenvector of module $k$. The intra-modular connectivity (IMC) is defined as the connectivity of a gene with its own module. The ME is also used to calculate the Pearson’s coefficient of correlation and the associated student $P$ value of each module with disease status; the disease status is encoded as binary information for disease or healthy.

Module Preservation and Consensus Analysis. WGCNA provides various measures of module preservation statistics, which assess whether or not the interconnections among the genes within a module and connectivity patterns of individual modules (for example, intramodular hub gene status) are preserved between two datasets. To assess the preservation of our disease-associated modules found in the PFC dataset (the network that we analyzed) and in a hippocampal gene expression dataset (test network), we used the modulePreservation function in the WGCNA R package (46). In brief, this function provides an average measure of several preservation statistics generated through many permutations of the data, the $Z_{\text{summary}}$ value. In general, modules with $Z_{\text{summary}}$ scores $> 10$ are interpreted as strongly preserved (that is, densely connected, distinct, and reproducible modules), $Z_{\text{summary}}$ scores between 2 and 10 are weakly to moderately preserved, and $Z_{\text{summary}}$ scores $< 2$ are not preserved (46). Another way to look at module preservation is to rank the modules by their overall preservation in the test set, which gives a relative measure of module preservation. Median rank is a measure that relies on observed preservation statistics rather than the permutation $Z$ statistics (46). It is calculated as described previously (46).

Consensus analysis is a way to identify modules present in several independent datasets. Consensus modules group together genes densely connected in all conditions and are defined from the clustering of consensus similarity:

$$\text{Module}_{\text{consensus}} = \min(\text{Network 1, Network 2}).$$

Consensus modules are by construction present (i.e., preserved) in all input datasets. If a module identified in a reference dataset is strongly preserved in test datasets, it would also be a consensus module among the reference and test datasets. Each consensus module has one eigengene per dataset. Eigengene correlation helps to visualize the overall network structure and also to compare a given network between different datasets. An eigengene network ($A_{\text{e}}$) is defined as a signed network with a soft thresholding power of 1. A preservation network (Pres) measures the correlation of eigengene correlation among different networks (88):

$$\text{Pres}_{(1,2)} = 1 - \left| \max(A_1, A_2) \cdot \cdots \cdot \min(A_1, A_2) \right|,$$

where $\text{Pres}_{(1,2)}$ is the preservation network for any networks 1 and 2. The overall mean preservation of eigengene networks is given by (88):

$$D_{(1,2)} = \text{mean}_{\text{pres}}(\text{Pres}_{(1,2)}).$$


Kundra et al.