Dysregulation of the secretory pathway connects Alzheimer's disease genetics to aggregate formation

Chih-Chung Kuo[2,3], Austin WT Chiang[1], Nathan E. Lewis[1,2,3,§]

[1] Dept of Pediatrics, University of California, San Diego
[2] Novo Nordisk Foundation Center for Biosustainability at UC San Diego
[3] Dept of Bioengineering, University of California, San Diego

[§] Correspondence to: Nathan E. Lewis, nlewisres@ucsd.edu

Abstract

A hallmark of amyloid disorders, such as Alzheimer's disease, is aggregation of secreted proteins. However it is largely unclear how the hundreds of secretory pathway proteins contribute to amyloid formation. We developed a systems biology framework that integrates expression data with protein-protein interaction networks to successfully estimate a tissue's fitness for producing specific secreted proteins. Using this framework, we analyzed the fitness of the secretory pathway of various brain regions and cell types for synthesizing the Alzheimer's disease-associated amyloid-precursor protein (APP). While none of the key amyloidogenic pathway components were differentially expressed in AD brain, we found the deposition of Aβ is associated with repressed expression of the secretory pathway components proximal to APP. Concurrently, we detected systemic up-regulation of the secretory pathway components proximal to β- and γ-secretases in AD brains. Our analyses suggest that perturbations from 3 high confidence AD risk genes cascade through the secretory machinery support network for APP and into the endocytosis pathway. Thus, we present a model where amyloidogenesis is associated with dysregulation of dozens of secretory pathway components supporting APP, which could yield novel therapeutic targets for the treatment of AD.

Introduction

No mammalian cell exists alone. Indeed, each cell dedicates >1/3 of its protein-coding genes to interact directly with other cells and its environment, using hormones and receptors for communication, enzymes and other proteins to modify their extracellular matrix, transporters for exchanging metabolites, etc. The mammalian secretory pathway is tasked with the synthesis, post-translational modification (PTM), quality control, and trafficking of these secreted proteins (secPs)[1,2]. SecPs account for >25% of the total proteome mass[3,4] and are among the most tissue-specific genes in the human genome [5]. The precision and efficiency of the mammalian secretory pathway result from the concerted effort of hundreds of secretory machinery components (secMs) including chaperones, enzymes, transporters, glycosyltransferases, metabolites and lipids within the secretory pathway[6–9]. Since many secPs relay signals between cells or modify a cell's microenvironment, each cell must carefully regulate the synthesis and localization of each secP.

Perturbations to the secretory pathway result in misfolded proteins, which induce ER stress and apoptosis. In amyloid diseases, the misfolded proteins can aggregate into toxic amyloid fibrils, which ultimately lead to cell death. Aβ deposition, a major pathological hallmark of Alzheimer's disease (AD), stems from the perturbed processing of the transmembrane amyloid precursor protein (APP). In Aβ aggregation, alternative proteolytic cleavage of amyloid precursor peptide by β- (rather than α-secretase) releases the secreted form of APP, the

aggregation-prone Aβ1-42.[10–12] Furthermore, additional PTMs in the secretory pathway may affect APP cleavage,[13,14] phosphorylation,[15] glycosylation,[16,17–19] amyloidogenesis, and APP trafficking. However, protein aggregation could stem from the perturbation of diverse processes, but no systematic exploration of all processes supporting proper APP processing has been done.[20] Several large-scale GWAS also identified more than 45 AD risk loci [21–23], although for many loci, it remains unclear how they induce Aβ deposition. Furthermore, a large part of AD heritability remains unknown [21,24]. The genetic landscape of late-onset Alzheimer's (LOAD) is highly heterogeneous, with multiple complex molecular interactions contributing to the disease phenotype. Therefore, the discovery of concerted expression changes in LOAD, such as the remodeling of immune-specific modules requires systems approaches on large datasets [25].

To unravel the molecular changes leading to Aβ deposition, we focused on the roles of the secretory pathway in amyloidogenesis. The secretory pathway is responsible for the processing, quality control and trafficking of key components of the amyloidogenic pathway [26,27], such as APP and the secretases, so we investigated if there is a systemic dysregulation of the secMs supporting their production and processing. To do this, we first developed a network-based approach that leverages protein-protein interaction (PPI) and mRNA and protein abundance data to quantify a cell or tissues' "secretory machinery support". This measures the fitness of a tissue or cell for properly secreting a specific secP based on the expression of its supporting secMs. Next, we investigated if there are disruptions in the secretory machinery support for key players of the amyloidogenic pathway (i.e., APP and the secretases), leading to increased Aβ deposition in LOAD, based on data from several large-scale clinical bulk- and single-cell RNA-Seq datasets [28,29]. We found significant dysregulation of the secretory pathway proximal to APP and the secretases, and this dysregulation is a major determinant of Aβ deposition. We further demonstrated that the concerted expression changes in the secretory support modules for the APP, BACE1, and PSEN1 can be linked to known AD risk genes and their regulation targets. In terms of subcellular localization, the core perturbed network enriches for known hotspots for Aβ production such as ER, cytosol and endosomes. Moreover, we found that the AD risk loci activate endocytosis via the core support network and we identified a candidate TF binding motif that is conserved in the promoter regions of the interaction network genes. Together, our analyses suggest mechanisms underlying impaired protein secretion, which could propose novel therapeutic targets for the treatment of AD. It also proposes mechanisms by which AD genetics imbalance the secretory pathway, thus resulting in Aβ deposition, cell death, and cognitive impairment.

Results

Secreted proteins and secretory machinery show similar tissue-specific expression

The secretory pathway synthesizes and transports a variety of secreted proteins, each with different requirements for their synthesis and secretion (e.g., different physicochemical properties and post-translational modifications). With the human secretome being one of the most tissue-specific subsets of the human proteome [5,30], we hypothesized each tissue expresses just the secMs needed to synthesize and process secPs from the tissue. Supporting this, we observed that clustering tissues by secP gene expression grouped tissues similarly as when clustering by secM gene expression (Figure 1, p-value = 0.0145; Figure S1). Thus the secMs are not merely housekeeping proteins always expressed to support any proteins being secreted; rather, they express in a tissue-specific fashion to meet the demands of different tissues [7]. However, the question remains if the pairings between secMs expressed in each tissue represent those needed to specifically support the secPs they secrete.

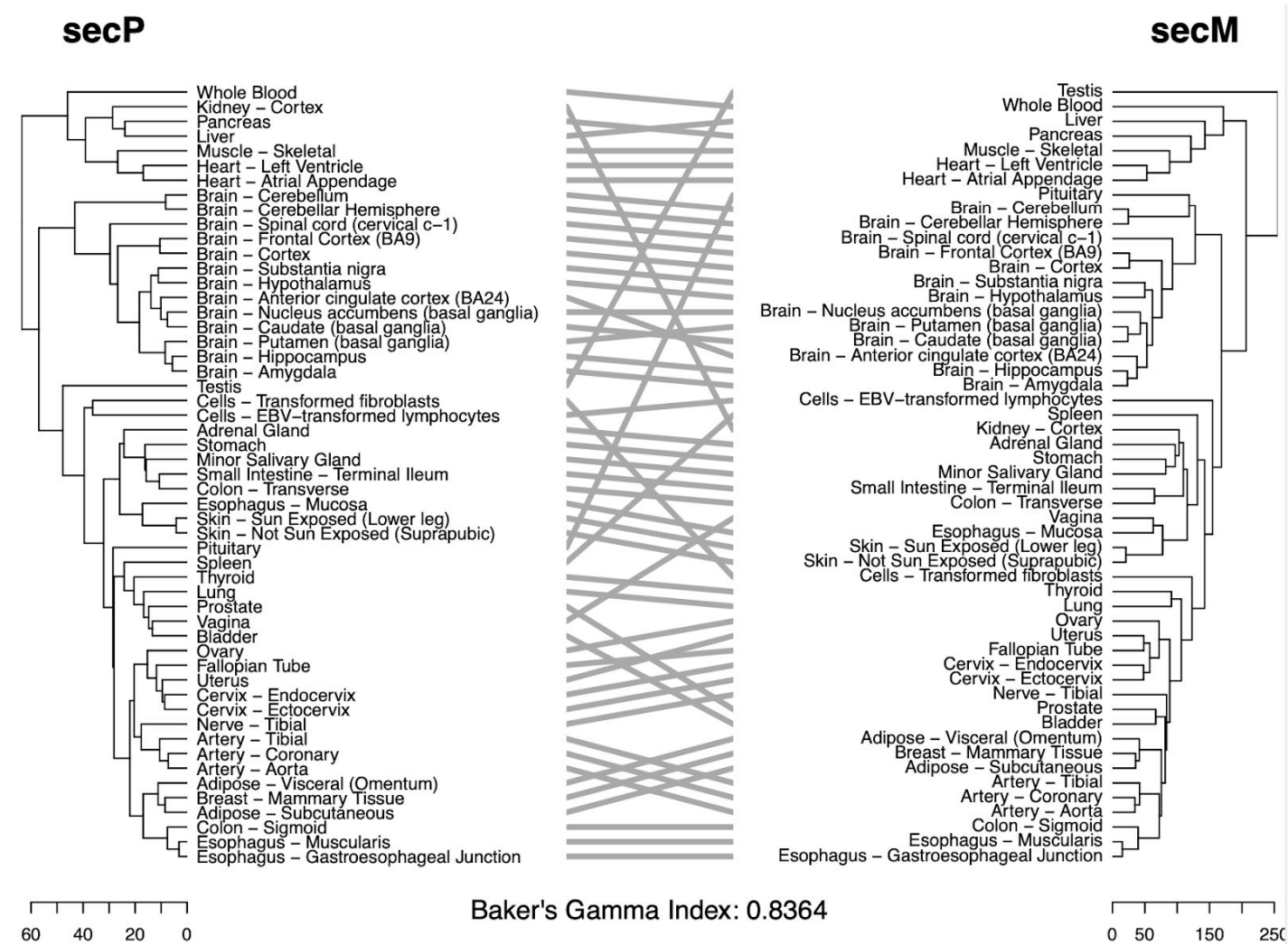
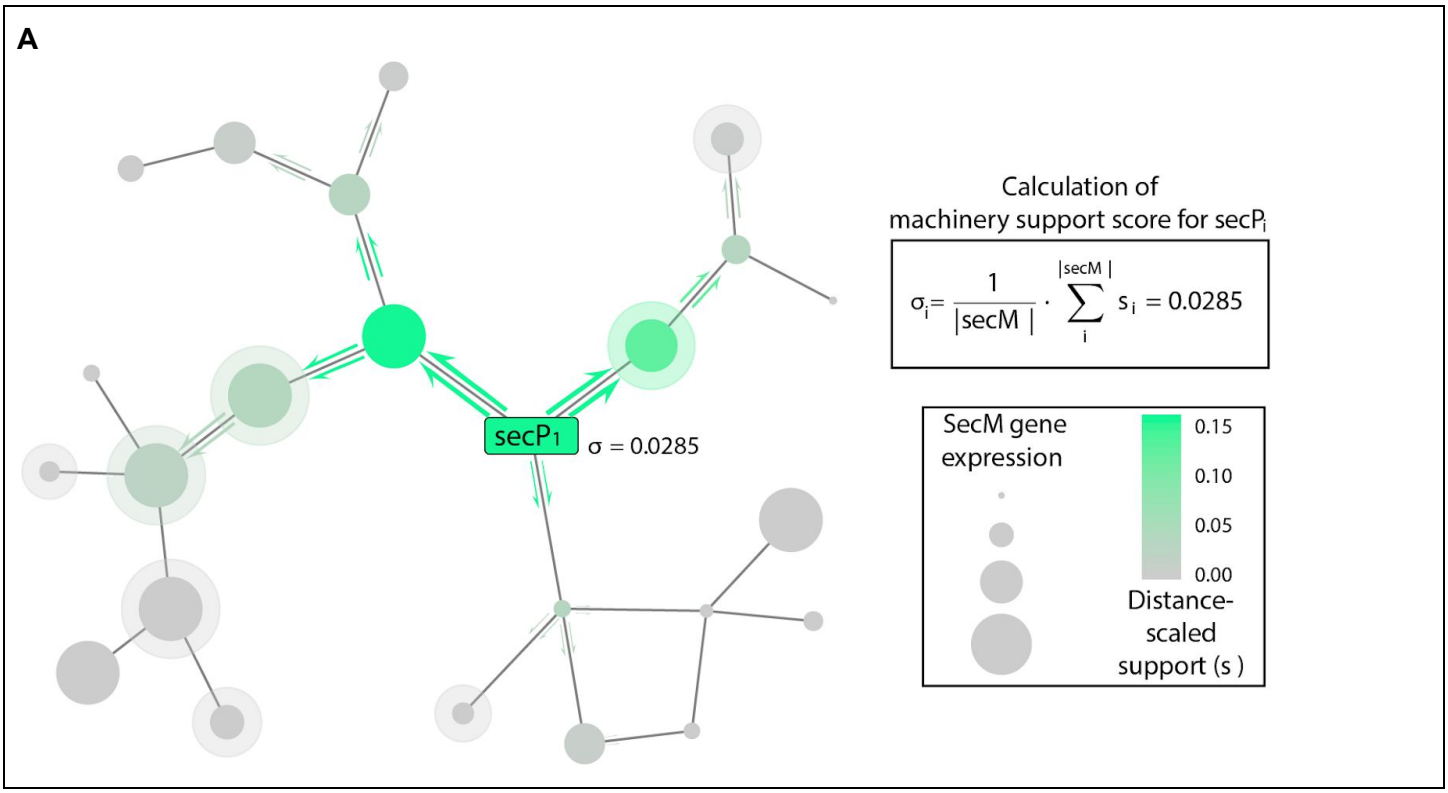


Figure 1. The secMs and the secPs show coordinated expression profiles across different human tissues. The tissue similarity structures from the perspective of the secretome and the secretory pathway expression were represented by two hierarchical clustering dendrograms, which were then compared with a tanglegram[31]. Gene expression of the secretory pathway and its clients show a high level of coordination across human tissues, with precision significantly better than expected by sampling genes from each tissue (bootstrapping p-value 0.013 and 0.045 for data from GTEx [32] and Human Protein Atlas [5] respectively).

Tissue specific expression of secMs predict expression of their client secreted proteins

To further dissect the pattern of secP-secM co-regulation seen across tissues, we incorporated two sources of information: protein-protein interactions (PPIs) and secM gene expression. We harness PPIs to identify the secMs relevant to each secP, since PPIs are one of the major modalities through which machinery proteins in the secretory pathway assist protein secretion [33–36]. Further, secMs responsible for secP post-translational modifications are well-captured by PPIs between the secPs and the secMs (Figure S2). To focus on spatially proximal interactions, we filtered the PPIs for interactions between secMs and other secretory pathway-resident proteins for each secP (see Methods), resulting in a "secM support network" consisting of 3658 genes. By overlaying secM gene expression on this network, one can quantify the secM support for secretion of each secP. To systematically quantify the fitness of the secM support network for producing each secP in a tissue, a machinery support score is calculated for each secP by a random walk algorithm that integrates secM gene expression levels proximal to each secP in its PPI network. More specifically, for each secP, we added the protein to the secM support network, and centered the network on the secP. We then performed a random walk on the secM support network starting from the secP. We adapted the transition probabilities of the random walk to incorporate gene expression of the secMs so that propagation is constrained by not only PPI network topologies but also the expression of the secM components (methods; Figure 2a). The algorithm assigns a component score to each protein in the network, representing its availability to the secP of interest. The "secretory machinery support score" (i.e., the average component scores the secretory pathway components receive from the random walk) then quantifies the overall secretory pathway support for the given secP. Using this approach, we found the secretory machinery support score for each secP increases in tissues wherein the secP is more highly expressed (Figure 2b, Figure S3). Further, the machinery support score improves the prediction of secP protein abundance from mRNA expression by >50% (Figure S4). Thus, by accounting for the mRNA/protein expression of PPIs surrounding each secP, the machinery support score quantifies a tissue's relative fitness for synthesizing and secreting the secP of interest.



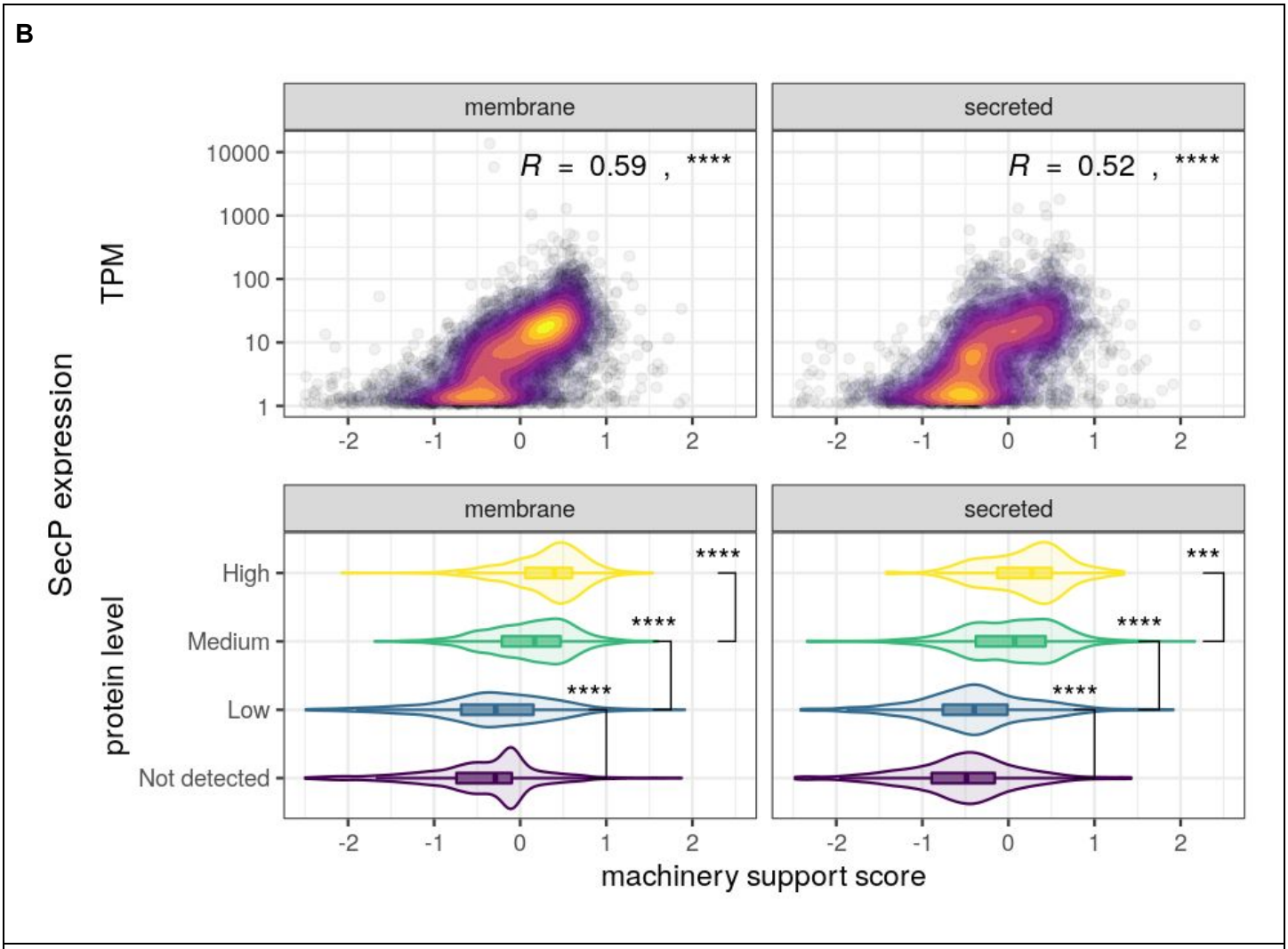


Figure 2. Expression data can quantify a tissue or cell's fitness for synthesizing a secreted or membrane protein. (a) In our systems biology approach, the mRNA or protein abundance is overlaid on a PPI network surrounding a secreted protein (secP). The secP synthesis fitness is quantified by summing the secM expression, scaled by distance from the secP (computed by a random walk), yielding a quantitative "machinery support score". The calculation of the support score also provides a sub-network of proteins contributing to secP synthesis. **(b)** We quantified the machinery support score for every secreted protein in all tissues in the Human Protein Atlas, and found a clear correlation between the secP expression and the relative machinery support score. This correlation was seen for both mRNA (top) and protein (bottom) abundance. Thus, our machinery support score allows one to quantify how fit a cell or tissue is for properly expressing and processing a secreted or membrane protein.

Aβ deposition in Alzheimer's disease is characterized by perturbed secretory support of amyloid precursor protein

The co-regulation of the secP and their cognate secMs results from millions of years of evolution. Could perturbations to such co-regulation underlie the molecular pathology in AD? Specifically, Aβ deposition is a major hallmark of AD-pathology. The precursor to Aβ, APP, is moderately to highly expressed (high transcript levels and moderate protein levels) in the cerebral cortex. While APP overexpression from APP duplication can cause early-onset (familial) AD [37,38], sporadic (non-familial) AD does not show differential transcript abundance for APP between AD and non-AD individuals despite the increase in Aβ plaques [39]. However, APP does undergo post-transcriptional processing, with pathogenic Aβ being released from APP following sequential cleavage by β- and γ-secretases, while the α-secretase promotes the correct processing of APP.

To test the relevance of secretory pathway expression to AD, we analyzed RNA-Seq data from 4 brain regions in 298 AD and age-matched control subjects (from the Mount Sinai Brain Bank[28]) and single-cell RNA-Seq from the prefrontal cortex (Brodmann area 10) of 48 individuals[29] (Figure 3). APP was not differentially expressed in AD brains at both the single-cell [29] and the tissue level [28]. This is not surprising since the amyloidogenic pathway giving rise to neurotoxic Aβ takes place post-translationally [40]. Additionally, expression of neither BACE1 nor PSEN1 correlated with plaque abundance in affected brain regions (Figure 3). However, each gene had significant changes in machinery support scores correlating with severity (Figure 3b). Specifically, the supporting machinery score for APP decreased in affected brain regions, showing suppressed scores in cells with amyloid deposition ($p<0.0066$). The largest effect was in cell types that are major producers of Aβ, including neurons [26,27,41], reactive astrocytes [42–45] (Figure 3a), and in brain regions affected early in the onset of AD (Figure 3b, Figure S5). However, BACE1 and PSEN1, which aid in amyloidogenesis, showed an opposite trend, with affected cells increasing machinery support for these secretases (Figure 3b; p-values for Brodmann area 36 (parahippocampal gyrus, BM36) and Brodmann area 44 (inferior frontal gyrus, BM44): $p<0.0038$ and $p<0.014$ for β-secretase; $p<0.13$ and $p<0.083$ for γ-secretase).

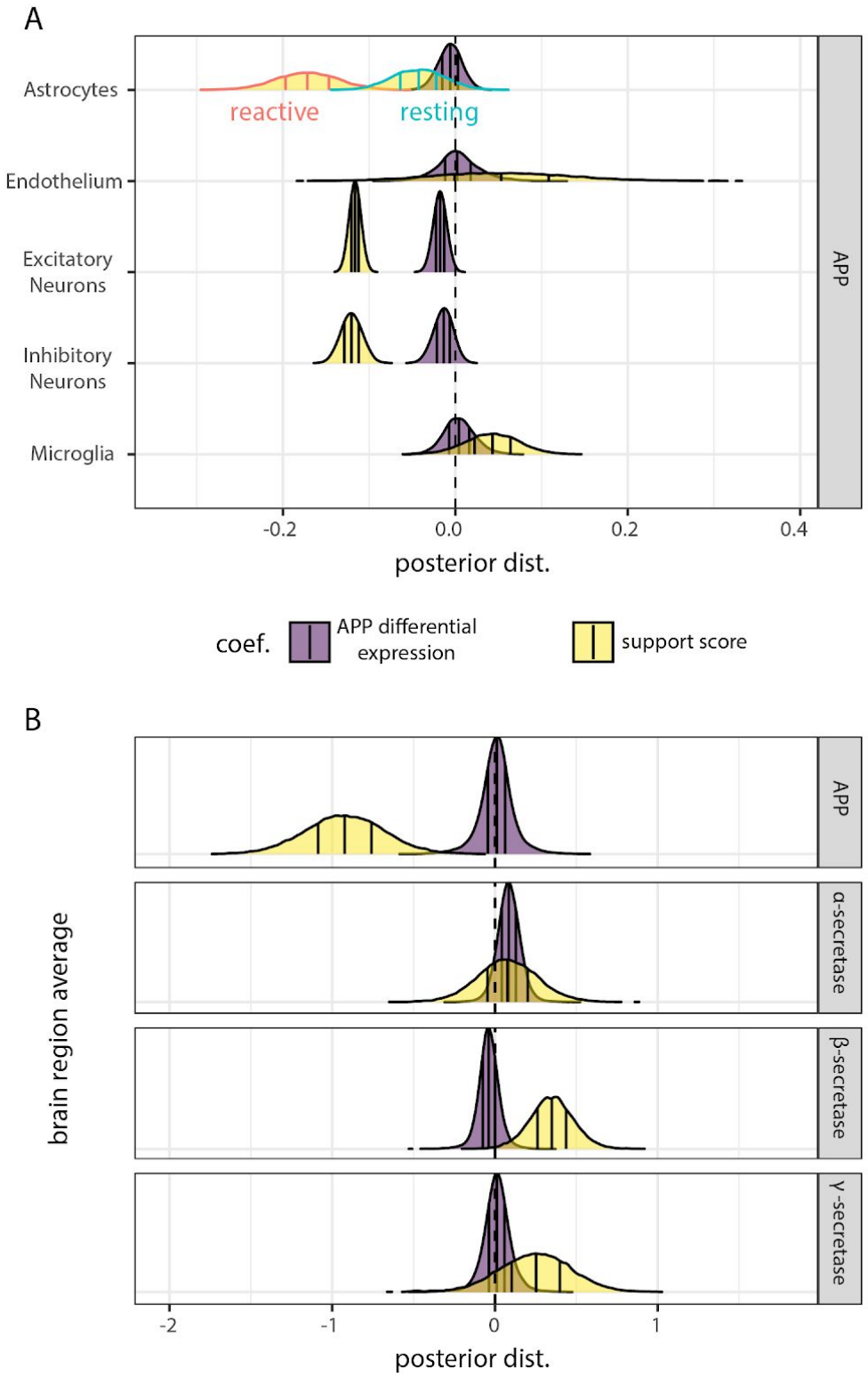


Figure 3. AD-relevant genes show perturbed secretory machinery support scores. (A) Overall, APP expression does not correlate with plaque densities across individuals in single-cell RNA-Seq. However, support scores show a negative correlation with plaque density, suggesting the secMs supporting proteostasis of APP are suppressed in AD. **(B)** Similar trends were seen in several brain regions, including the AD-affected BM36 and BM44 regions, where APP, BACE1 and PSEN1 do not correlate with plaque abundance, but the secM support score APP shows a negative correlation while the support scores for BACE1 and PSEN1 positively correlate with amyloid formation, suggesting AD pathogenesis involves dysregulation of the secretory pathway.

Secretory pathway support of APP is most strongly suppressed proximal to APP

The secretory machinery support score for APP is significantly decreased in AD. Is the decreased score a general suppression of many secMs throughout the secretory pathway, or is it specifically the secMs most proximal to APP that are suppressed? To test this we defined a core support network for APP based on network proximity and gene expression. As the abundance and proximity to APP vary across the secMs in the network, we broke down the APP support score into the individual component scores for each protein in the secM support network (Table S1). This quantifies each secM's corresponding contribution to the secretory support of APP. When the secMs are rank-ordered by their individual component scores, their contribution to the APP support score follows a pattern of exponential decay (Figure S6A), suggesting that the support score is mostly determined by a smaller number of proteins with high proximity to APP (Figure 4A). While the entire APP support network is not differentially expressed between AD and healthy brains, we wonder if this is the case with secMs that are major contributors to the support score. When we overlaid the differential expression across brain regions and cell types and considered progressively smaller subsets of the APP support network consisting of proteins with the highest component scores, we saw the strongest repression at around 20-30 secMs, suggesting that the proteins nearest to APP are the most suppressed (Figures 4A, S6B, S6C).

Changes in APP-supporting PPIs are regulated by AD risk loci

Large GWAS screens found AD risk genes impacting many pathways.[22,46] Although the secretory pathway is tasked with the synthesis and processing of APP, it is not generally implicated in LOAD pathogenesis, since secretory pathway genes are not enriched for among LOAD risk genes from large-scale GWAS studies [22]. However, our results show transcriptional perturbations of machinery support for key amyloidogenic genes; thus, there may be a concerted regulatory change for modules supporting APP production and proteostasis in the secretory pathway. Thus, we tested if regulatory AD risk loci (i.e., transcription factors) regulate the secMs interacting with APP. While the entire APP support network does not enrich for the AD risk genes nor their regulatory targets, we wondered if AD risk loci selectively target the secMs more proximal to APP in the PPI network. As we assessed enrichment for AD risk gene targets in subnetworks that were progressively closer to APP (i.e., more focused on the support network interacting most directly with APP), we found the secMs supporting APP were increasingly enriched for targets of AD risk regulatory genes (Figures 4B, S6D). The results are reproducible across multiple GWAS significance thresholds (Figure S7). Since the enrichment of AD risk gene targets steadily increased in statistical significance with increasing proximity to APP in the PPI, we focused on the top 20 proteins interacting the most closely with APP from the support network. This resulted in an APP support subnetwork where the enrichment significance for AD risk gene targets peaks. This cutoff also coincides with the subnetwork with the strongest suppression of secM expression (Figures 4A, S6B, S6C).

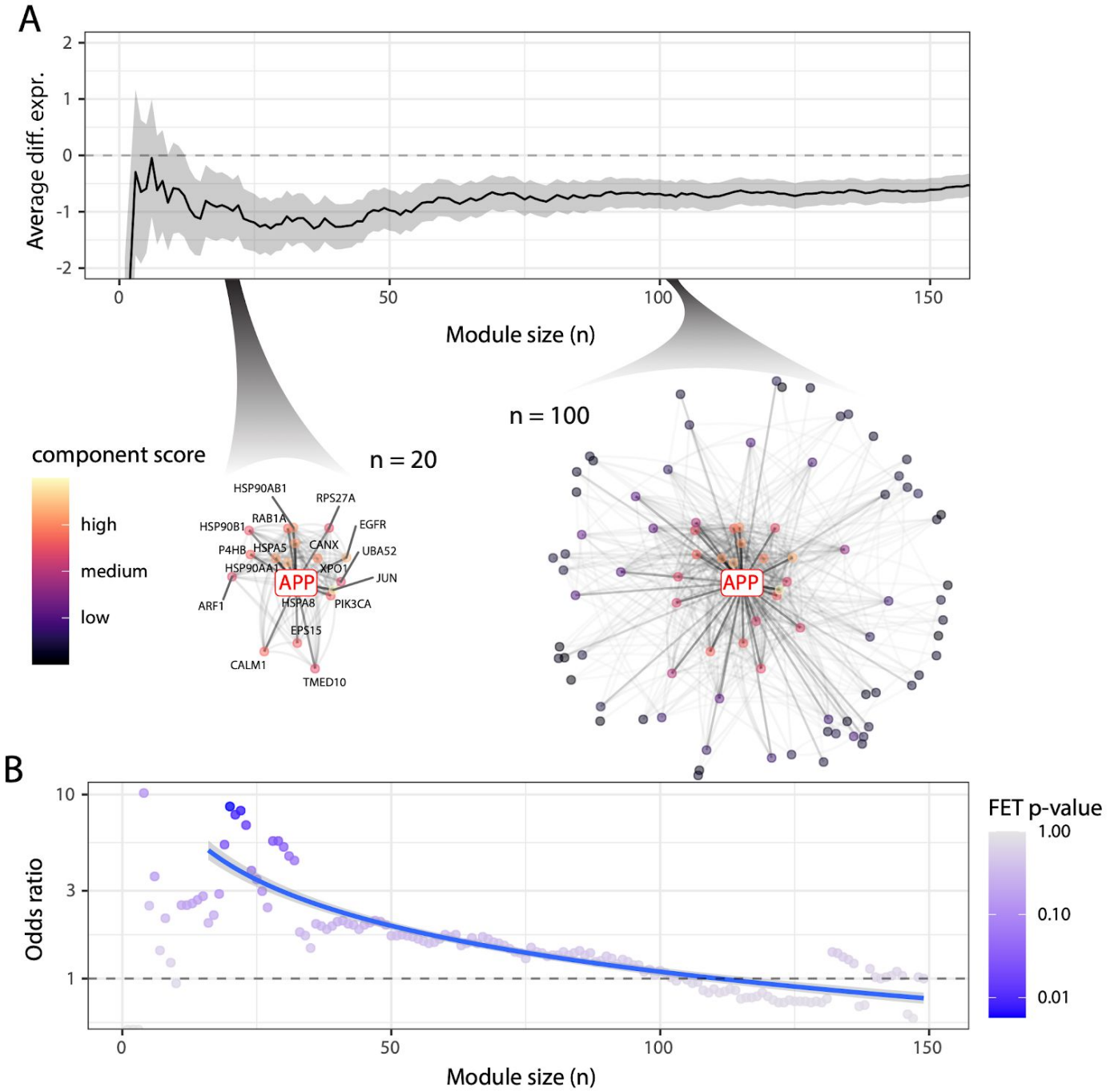


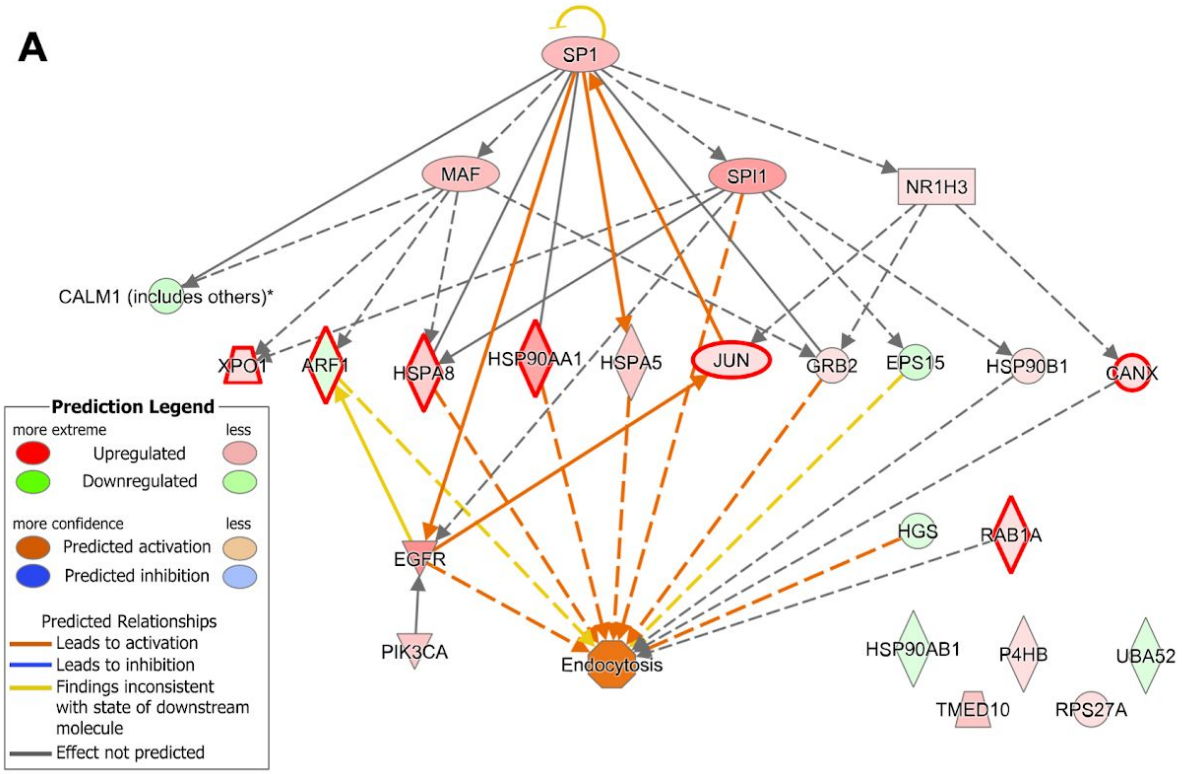
Figure 4. The APP secM proteostasis network is not enriched for AD risk genes, but is enriched for AD risk gene regulatory targets. The networks of secM proteins supporting APP production at two support component score cutoffs, representing the top *n* = 20 / *n* = 100 proteins contributing the most to the APP's support score. The top 20 proteins with the highest component scores are labeled. **(A)** Starting from n=1 where the secM with the highest component score was considered, we incrementally included secMs less proximal to APP. The sizes of the subnetworks are indicated by the x-axis. At each iteration, we calculated the average differential expression (y-axis) between AD and healthy subjects for each subnetwork. The strongest repression of the AD support network occurs at around n=15~30. **(B)** We also calculated the degree to which the subnetwork enriched for regulatory targets of known AD risk genes above the genome-wide significance threshold (y-axis). The regulatory targets of AD risk genes are generally depleted among the general non-APP-specific secMs (Figure S6 for full trend across all 3685 subnetworks), but targets of AD risk genes enrich strongly among the core secMs closest to APP.

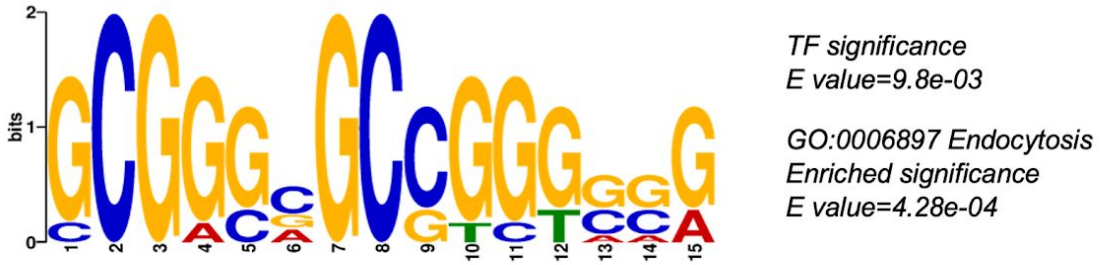
AD risk loci activate endocytosis via the core support network

We analyzed the content of the core support network, and found it is enriched for genes in the amyloidogenic pathway. Specifically, in terms of subcellular localization, we saw interactions were concentrated in the ER, endosomes, and the cytosol (Figure S8, FDR p<1.1e-3), consistent with the localization of amyloidogenesis. For example, the endosome hosts intracellular Aβ production with its β-secretase, and is enlarged in autopsies from AD[47] and stem cell models.[48] To further unravel the link between endocytosis and the core support network, we analyzed the patterns of the core network differential expression between AD and controls across multiple cohort studies using gene regulatory networks obtained from ENCODE[49] and Ingenuity Pathway Analysis (IPA)[50]. Complementing our previous observation that a significant portion of the core support network is endosome-resident, we saw significant up-regulation of genes associated with endocytosis (p-value = 8.95e-14), mediated by the core supporting machinery (Figures 5A, S9) across various brain regions. We further analyzed the APP core support network and identified transcription factors (TFs) that are most strongly associated with the perturbed secM module. The APP core support network coincides significantly with the regulatory targets of 3 genes from AD risk loci-- NR1H3, MAF and SPI1 [21–23,46]. To investigate the extent to which these AD risk genes perturb transcription of the supporting machinery in AD, we analyzed the differentially expressed genes between AD and controls across multiple cohort studies using gene regulatory networks obtained from ENCODE[49] and Ingenuity pathway analysis (IPA)[50].

In addition to predicting upstream transcriptional regulators associated with amyloidogenesis from gene signatures or differentially expressed genes, we searched for conserved TF binding sites among the top-ranked dysregulated genes of the core-network in Alzheimer's Disease. Specifically, we conducted de novo TF binding site motif discovery for the 20 genes in the core support network (see Methods). We identified one significant TF motif (E value= 9.8e-3) that is present in 7 of the genes in the core support network (Figure S10), which is associated with the SP1, SP2, and SP3 transcription factors. SP1 is important in AD[51,52] and is predicted to bind the 3 genes from AD risk loci (NR1H3, MAF, and SPI1) with high confidence via 3 elite enhancers[53] GH11J047390 (GH score=2.2), GH16J079764 (GH score=2.4), and GH11J047247 (GH score=2.1), respectively. Thus, a perturbation to multiple TFs could disrupt the APP core support network, wherein both genetic risk genes (NR1H3, MAF, and SPI1) and global regulators can contribute to the dysregulation of the core network.

Further analysis suggests the AD risk genes and the core support network genes are further co-regulated with an activation of endocytosis in the AD pathogenesis. We further characterized the functional association of conserved TF motif by scanning all promoters of genes in the genome for the motif (see Methods for details). The conserved TF motif was significantly enriched in the promoter regions of known endocytic pathway genes (p-value=4.28e-4) and several other pathways relevant to AD pathogenesis (Table S5, Figure S10). Together, these results suggest a concerted change in endosomal activities and dysregulated pathways between normal and AD brains that arises due to the differential expression of the core supporting machinery surrounding APP.





Hit	Name	Strand	Start	*p*-value	Sites
Y	XPO1	+	14	1.58e-7	ccgcaggttg GCGGGAGCCGGTGGG ggcgggccgg
Y	ARF1	+	58	2.82e-7	gcgcgtcact GCGGCCGCCGGGGCG ccgaggccag
Y	HSPA8	-	30	3.66e-7	GGGTTTGGGC GCGGGAGCCGGGCAG GAGAAGACGG
Y	JUN	+	6	1.31e-6	cggtg GCGGCCGCCGGTGGA tgacttcggg
N	P4HB	+	11	1.41e-6	aacttccggt CCGGGGGCCGGGGGG cgtgccagtt
Y	RAB1A	-	44	1.86e-6	TGCCGGGGGT GCGGGGGCGGGGAGG GGGCCGGGGT
Y	CANX	-	53	3.68e-6	AGAACCCAGC GCGAGCGCCTGGCCG GATTGGCGTG
Y	HSP90AA1	+	11	5.44e-6	ttctttggca GCGGGCGCGGCGCCA gtcaggacga

Figure 5. The regulatory relationships surrounding the core support network.
(A) Regulatory structures of the 20 proteins from the core support network were constructed using IPA. Gene expression profiles from the 4 brain regions (BM10, BM22, BM36, BM44) [28], the Mayo Study [54] and the ROSMAP study (Religious Order Study and Memory and Aging Project) [55] were averaged and overlaid on the core support network. The endocytosis pathway is strongly activated in AD brains via the core support network, which in turn is regulated by 3 genes from AD risk loci-- NR1H3, MAF and SPI1. The proteins harboring binding sites for the TF motif (shown in panel B) were outlined in red. (B) The TF motif (top panel) aligns significantly to 8 genes from the entire support network (bottom panel), of which 7 of them belong to the core support network.

Discussion

In the past four decades, the Alzheimer's research community has made huge strides towards elucidating molecular processes contributing to amyloid deposition. However, despite involving aggregation of secreted proteins, it remained unclear to what extent the core processes of protein secretion and proteostasis are involved. Here we developed a systems biology approach that analyzed the interactions between key amyloidogenic components and the secretory pathway. This approach predicted the propensity for amyloid deposition at the single-cell level. To gain systems-level insights into LOAD, we used the framework to identify a subset of the secretory pathway components on which concerted suppression converged. We further demonstrated that an increase in endocytic activities in LOAD can be attributed to key AD risk genes via the core support network.

LOAD is a complex disease. The identification of three rare mutations in APP, PSEN1 and PSEN2 that occur in early-onset familial AD and the discovery of APOE constitute our primary knowledge of the genetic landscape of LOAD. More recently, high-throughput technologies such as GWAS and whole exome sequencing have identified more than 45 genetic risk loci of LOAD. However, the additional risk loci exert only very small risk effects [56], and the link between genetic risk variants and amyloid deposition remains incompletely understood. Despite the highly heterogeneous expression of APP and other key amyloidogenic components across AD and healthy subjects, we demonstrated concerted down-regulation of secretory machinery proximal to APP in AD patients. This highlights the secretory pathway as a determinant of amyloid deposition, which had not been a major focus of AD research. Incidentally, the proteostasis network, with which the secretory pathway shares a significant overlap, has been an increasingly popular target of protein aggregation and aging studies. The human chaperone network, a major component of the proteostasis network, undergoes continual remodeling during an organism's lifespan [57–59]. However, in the aging AD brain, the directions of regulation of these chaperones are rather uncoordinated across different chaperone families and even within the same family [58]. Our observations of concerted repression of key proximal secretory pathway components show that improper expression of the secretory pathway, of which the chaperone network is a subset, is associated with the deposition of amyloid.

Even though the secretory pathway is in charge of post-translational processing and targeting of APP up until its cleavage by the secretases, its implications in AD have been insufficiently researched primarily due to the lack of AD risk genes in the pathway [60]. Our results showed that genes contributing the most to the APP support network in the secretory pathway are significantly enriched for targets of AD risk genes, suggesting a mechanistic link between AD genetic variants and secretory pathway dysregulation, complementing previous systems-level approaches to understanding LOAD with a focus on immune- and microglia-specific modules [25]. To further unravel the link, we examined the core support network consisting of secretory pathway components most proximal to APP. We noticed 3 AD risk genes that are also transcription factors showing regulatory evidence over the core support network according to both curated and *de novo* pathway analyses. More importantly, we demonstrated a regulatory cascade originating from the 3 AD risk genes, mediated by the core support network and into the endocytosis pathway. The endosome, where the β-secretase is localized to and where its acidic pH is optimal for enzymatic cleavage, is a major site of intracellular Aβ production. Our findings thus offer a mechanistic view of amyloidogenesis involving the secretory pathway and the endosomes. This is in line with observations in embryonic cortical neurons that showed increased Aβ levels as a result of increased endocytic pathway activities and reuptake in APP in aged cells[61]. We also observed significant enrichment of endosomal-localized proteins in the core support network for APP, further lending credence to the involvement of the secretory pathway in activating the endocytic pathway.

The dominant model of AD pathogenesis, the amyloid hypothesis[62–64], posits that AD pathogenesis and the rest of the disease process such as tau tangle formation [65,66] result from the accumulation of Aβ via the imbalance between Aβ production and clearance. We examined the capacity of the secretory pathway in the context of Aβ production and processing, where the secretory support of APP and β- and γ-secretases were analyzed. While the concerted dysregulation of the secretory support for these key amyloidogenic components in AD brains theoretically leads to increased Aβ production, the impact on Aβ clearance warrants further investigation. It is worth noting that detectable Aβ deposition can precede the onset of AD by more than 15 years[67,68], which likely coincides with the onset of the decline of the proteostasis network. Our findings highlight the roles of the secretory pathway in amyloidogenesis, which open new possibilities for early diagnosis and treatment research on LOAD. Furthermore, our systems approach can be further applied to other diseases in which the secretory pathway is perturbed, such as perturbed hormone secretion in endocrine disorders, changes in hepatokine secretion nonalcoholic fatty liver disease[69,70], and the secretion of diverse proteins in cancer[71].

Methods

Calculation of secretory pathway support scores

Random walk on interactome

To contextualize the secretion of a given secP, we used network propagation to quantify the influence of gene expression across neighboring genes. Let $G(V, E)$ denote an undirected interactome with vertex set V containing n proteins and an edge set E the m interactions between them. Let w_{ij} be the edge weight ($w_{ij} = 1$ if G is undirected) between edges i and j and A be the adjacency matrix of G where $A_{ij} = \{w_{ij}\ if\ \{v_i, v_j\} \in E;\ 0\ otherwise\}$. Let $x(t)$ be the location of the walk at time t. Note that given the previous walk location i at time $t-1$, we can represent the probability of the walker moving from location i to j in a single step as $W_{ij} = \Pr(x(t) = i | x(t-1) = j) = \frac{A_{ij}}{d_i}$, where $d_i = \sum_j^n A_{ij}$. Summing probabilities from all inbound locations we have $\Pr(x(t) = j | x(t-1)) = \sum_i^n W_{ij} \Pr(x(t-1) = i)$. In matrix notation, this is $\mathbf{p}(t) = \mathbf{W}\mathbf{p}(t-1)$, where W is the transition matrix where each entry W_{ij} denotes the aforementioned transition probability from i to j. In random walk with restarts [72], at each step the walk resets to the origin with probability α, and the last equation becomes $\mathbf{p_{RWR}}(t) = (1-\alpha)\mathbf{W}\mathbf{p_{RWR}}(t-1) + \alpha\mathbf{p}(0)$, where $\mathbf{p}(0) = e_k$ denotes the initial distribution if the walker starts at v_k. The restart parameter α was set to 0.1, as advised by the linear optimal model given the size of the network [73].

Expression-guided random walk with restarts

The transition matrix can be modified to incorporate gene expression into each step of the propagation. If we let $\mathbf{t} = [t_0, \ldots, t_n]^\intercal$, where t_i is the scaled expression corresponding to node v_i, $0 \le t_i \le 1$, the expression-adjusted transition matrix can thus be given by $\widehat{\mathbf{W_{adj}}} = \mathrm{diag}(\mathbf{t})\mathbf{W}$. We can normalize the adjusted transition matrix by adding in self-loop: $\mathbf{W_{adj}} = \widehat{\mathbf{W_{adj}}} + \mathbf{I}_{n\times n} - \mathrm{diag}(\mathbf{1}_{n\times 1}^\intercal \widehat{\mathbf{W_{adj}}})$, and the update rule, which we termed expression-guided random walk with restarts (eRWR), now becomes:

$$\mathbf{p_{RWR}}(t) = (1-\alpha)\mathbf{W}\mathbf{p_{RWR}}(t-1) + \alpha\mathbf{p}(0).$$

Calculation of support scores and component scores

We performed eRWR on each secP of interest for 20 iterations, and the final vector of probabilities $\mathbf{p_{RWR}}(t=20)$ represent the support component score for each gene on the network $G(V, E)$. The support score (σ) is the average of the support component scores of the secretory pathway proteins, $\sigma = \frac{1}{|\mathrm{secM}|}\sum_{i \in \mathrm{secM}} \mathbf{p_{RWR}}_i(t=20)$.

Context filtering of the secretory pathway support network

We used the composite consensus human interactome PCNet v1.3 [73] (NDEX UUID: 4de852d9-9908-11e9-bcaf-0ac135e8bacf) to be the static, context-agnostic network ($G_0(V_0, E_0)$). For each secP, we create a subnetwork containing the secP and other essential secretory pathway genes by filtering the network for human secretory pathway components [7,9] and secretory pathway-resident proteins [74] to constrain the network spatially, resulting in a vertex-induced subgraph $G(V = \{V_0 \cap (secP \cup secM \cup secResident)\}, \{uv | uv \in E_0 \text{ and } u, v \in V\})$.

Transcriptomic and proteomic data processing and support scores calculation for normal human tissues

To calculate support scores for the normal human secretome, we used two datasets, the Human Protein Atlas (HPA) [5], and the deep proteome and transcriptome abundance atlas (deep proteome) [75] for tissue-specific transcriptomes from healthy human donors in which matching proteomic data were also available. For data from the Human Protein Atlas, we downloaded the transcriptomic data-- “RNA HPA tissue gene data” and performed log- and sigmoid-transformation on the transcript abundance (TPM) data, resulting in transformed gene expression profiles in the $(0, 1)$ range. For the HPA dataset, the support score was calculated based on the tissue-median of the transformed gene expression profiles for each secP. We retained the semi-quantitative nature of the immunohistochemistry protein abundance reporting, and we calculated the support scores summary statistics for proteins belonging to each of the staining levels--"High", "Medium", "Low" & "Not detected" separately.

With the fully quantitative proteomic data from the deep proteome [75], We calculated the support scores based on the protein iBAQ values. The iBAQ abundance values were transformed in a similar fashion as the transcript abundance from the HPA dataset. Namely, they were log- and then sigmoid-transformed into the $(0, 1)$ range, before being median-summarized by tissue and subsequently used in the calculation of the support scores.

Transcriptomic and proteomic data processing and support scores calculation for AD and healthy brains

To calculate support scores for key amyloidogenic pathway components in AD and healthy brains, we used two major transcriptomic datasets from the ROSMAP project (Religious Orders Study and Memory Aging Project)-- single-cell [29,76] (Synapse ID syn18485175) and bulk RNA-seq [28] (Synapse ID syn3159438) data from individuals respectively with varying degrees of Alzheimer’s disease pathology. The single-cell transcriptomic dataset covers 80660 cells from the prefrontal cortex of 48 individuals. While annotations for major cell types were given, we further classified astrocytes into reactive and non-reactive astrocytes based on GFAP expression [45]. The bulk RNA-seq covers 4 brain regions (Brodmann areas 10, 22, 36, 44) of 364 individuals.

To transform the count data into appropriate expression inputs to the eRWR algorithm, count data from healthy tissue-specific transcriptomes and the AD single-cell RNA-seq data were first log-transformed to compress the extreme values. The values were then standardized and passed through a logistic function, where the final transformed values have a range $(0, 1)$. For the AD bulk-RNA seq data, since the counts were already normalized and transformed, they were standardized and transformed by a logistic function without first being log-transformed.

Statistical analysis

Relationship between support scores of secreted proteins and protein expression

To examine the dependencies between support scores and the transcript and protein abundances of the human secretome, we calculated the support score for each secreted protein in the human secretome [77] across various human tissues. We first calculated the spearman correlation coefficients between the tissue-median support scores and the transcript and protein abundances across all secreted proteins. To further quantify the statistical significance of transcript abundance and protein level in determining overall protein abundance, a Bayesian hierarchical model was created (model equations shown below) where the abundance of each protein across the 29 tissues is drawn from a linear combination of mRNA levels and the support scores weighted by their respective regression coefficients. We used the *rethinking* R package[78] to construct the model and sample the coefficients.

$$\mathrm{iBAQ}_{\mathrm{Protein}} \sim \mathrm{Normal}(\mu, \sigma) \quad \text{(For each tissue-protein combo, draw protein abundance)}$$

$$\mu = \mathrm{a}[i] + b_{\mathrm{genExp}}[i] \cdot \log(\mathrm{FPRM}) + b_{\mathrm{secSupport}} \cdot \mathrm{support.score} \quad \text{(Proteins } i = 1 \ldots 13002 \text{ across tissues} j = 1 \ldots 29\text{)}$$

$$\mathrm{a}[i] \sim \mathrm{Normal}(\overline{\mathrm{p}_\mu}, \overline{\mathrm{p}_\sigma}) \quad \text{(Protein}_i \text{ in tissue } j \text{ drawn from population abundance)}$$

$$\begin{bmatrix} b_{\mathrm{genExp[i]}} \\ b_{\mathrm{secSupport[i]}} \end{bmatrix} \sim \mathrm{Normal}\left(\begin{bmatrix} \overline{\mathrm{e}_\mu} \\ \overline{\mathrm{s}_\mu} \end{bmatrix}, \mathbf{S} \right) \quad \text{(each coef for geneExp and secSupport for Protein}_i \text{ in tissue } j \text{ drawn from 2D Normal)}$$

Hyper priors:

$$\overline{\mathrm{p}_\mu} \sim \mathrm{Normal}(0, 1)$$
$$\overline{\mathrm{p}_\sigma} \sim \mathrm{Exponential}(1)$$
$$\overline{\mathrm{e}_\mu} \sim \mathrm{Normal}(0, 1)$$
$$\overline{\mathrm{e}_\sigma} \sim \mathrm{Exponential}(1)$$
$$\overline{\mathrm{s}_\mu} \sim \mathrm{Normal}(0, 1)$$
$$\overline{\mathrm{s}_\sigma} \sim \mathrm{Exponential}(1)$$

Relationship between support scores of key amyloidogenic proteins and amyloid plaque densities

We built a Bayesian hierarchical model (equations shown below) to determine the extent to which the support scores for key amyloidogenic pathway components including APP and the secretases for each cell/ sample in the single-cell and bulk RNA-seq dataset affects the amount of amyloid plaque measured. We regressed the scaled amyloid plaque densities corresponding to the individual from which the single-cell/ bulk RNA-seq sample was collected against the gene expression and secretory pathway support scores of key amyloidogenic pathway components. To regularize the coefficients of interest, their Bayesian priors are all normally distributed around 0. The coefficients were sampled using the *rethinking* R package[78].

$$
\begin{aligned}
&\text{amyloid} \sim \text{Normal}(\mu, \sigma) \\
&\quad \mu = \text{a} + \sum_{j \in \textbf{Amyl}} \left(b_{\text{genExp}_j}[i] \cdot \text{genExp}_j + b_{\text{secSupport}\,j}[i] \cdot \text{support.score}_j \right) \quad (i : \text{Brain region/ cell type}; \textbf{Amyl} \subseteq \{\text{APP, secretases}\}) \\
&\quad \sigma \sim \text{Exponential}(1) \\
&\quad \text{a} \sim \text{Normal}(0, 1)
\end{aligned}
$$

Coefficients of interest:

$$
\begin{aligned}
b_{\text{genExp}_j}[i] &\sim \text{Normal}(\overline{\beta_{\text{genExp}_j}}, \overline{\sigma_{\text{genExp}_j}}) \\
b_{\text{secSupport}\,j}[i] &\sim \text{Normal}(\overline{\beta_{\text{secSupport}_j}}, \overline{\sigma_{\text{secSupport}_j}})
\end{aligned}
$$

Hyper priors:

$$
\begin{aligned}
\overline{\beta_{\text{genExp}_j}} &\sim \text{Normal}(0, 1); \quad \overline{\sigma_{\text{genExp}_j}} \sim \text{Exponential}(1) \\
\overline{\beta_{\text{secSupport}_j}} &\sim \text{Normal}(0, 1); \quad \overline{\sigma_{\text{secSupport}_j}} \sim \text{Exponential}(1)
\end{aligned}
$$

Characterizing the core support network

AD risk genes and enrichment analysis of regulatory components

We obtained 45 genome-wide significant risk loci identified by several AD GWAS studies as summarized in [21], resulting in 176 high-confidence AD risk genes. We compiled a separate set of AD risk genes from GWAS summary statistics [22,23] for loci above the genome-wide suggestive threshold, where MAGMA [79] was used to aggregate p-values for SNPs to the gene-level independently for each GWAS dataset. P-values from the two datasets for each gene were then combined using Fisher's method, resulting in 673 AD suggestive risk genes. The transcription factors and their targets were obtained from ENCODE [80] and ChEA [81] via the Enrichr portal [82]. To determine whether the core support network enriches for the regulatory targets of AD risk genes, we first calculated the level of overlap between the core support network and the targets of each transcription factor using Fisher's exact test, where significantly overlapping transcription factors were defined as those with p-values of less than 0.05. A secondary enrichment was performed to quantify the level to which the significant transcription factors overlap with known AD risk genes. As mentioned earlier, two lists of AD risk genes were used. For the 673 AD suggestive risk genes, a traditional FIsher's exact test was performed. For the risk genes originating from the 45 risk genome-wide significant risk loci, instead of calculating the direct overlap between the significant transcription factors and the 176 high-confidence risk genes, we mapped the significant transcription factors back to the 45 risk loci on which Fisher's exact test was performed. This is motivated by the fact that many risk loci contain multiple risk genes that cannot be further pinpointed due to complex linkage disequilibrium patterns, a risk locus is considered hit if at least one of its mapped risk genes appears significantly enriched as a transcription factor. We performed this two-stage enrichment analysis starting from the full static support network towards the core support network by pruning back proteins furthest from APP in each iteration.

Enrichment analysis of subcellular compartments

We compiled lists of proteins for all subcellular structures consisting of proteins known to localize to the compartment of interest within the cell [74]. We ordered the proteins in the full support network by the extent to which they deviate from their stationary support component score to control for network topology while accounting for secretory-resident proteins. To determine the degree to which the proteins from certain subcellular compartments are overrepresented in the core subnetwork, we applied Gene Set Enrichment Analysis (GSEA) [83,84] with the subcellular localization gene-sets and the ranked core support network

components as input, eliminating the need for a hard significance cut-off. Subcellular compartments significantly enriched in the core subnetwork are defined as those with an FDR p-value of 0.05 or less.

Causal gene network analysis

To robustly define the core supporting subnetwork, we iteratively constructed subnetworks from proteins most proximal to APP and progressively include more distal proteins corresponding to different significance cutoffs. To robustly select the cutoff for the core supporting subnetwork, we performed the two-stage enrichment analysis on all subnetworks as detailed above (see "AD risk genes and enrichment analysis of regulatory components"). Additionally, we calculated the average differential expression between AD and healthy individuals for each subnetwork using fold changes from bulk and single-cell RNA Seq data depending on the source expression from which the subnetwork is calculated. We selected 20 proteins most proximal to APP to include in the final core subnetwork, where the cutoff coincides with the strongest enrichment of regulatory AD risk loci and the suppression of the core subnetwork.

To determine the regulator effects, we performed two network-based analyses. We first ran the upstream regulator analysis using the curated regulator networks from IPA [50]. The algorithm took as inputs the core subnetwork and the differential expression fold changes and p-values. Batch-corrected differential gene expression profiles between AD and healthy brains from the Mount Sinai study [28], the Mayo Study [54] and the ROSMAP study (Religious Order Study and Memory and Aging Project) [55] were obtained from the AMP-AD Knowledge Portal (Synapse ID syn14237651). "Disease & functions" having considerable overlap with the core subnetwork were added, of which endocytosis is the most significant (p-value =2.34E-14).

De novo TF binding site motifs discovery and known TF binding site identification

We downloaded promoter sequences (version: GRCH38) from UCSC Genome Browser[85] for the core subnetwork. The promoter sequences are defined as sequences 1,000 bases upstream of annotated transcription start sites of RefSeq genes with annotated 5' UTRs. To conduct de novo TF binding site motifs discovery, we first ran motif discovery using the MEME suite[86] with default parameters to identify candidate TF binding site motifs within the promoter sequences by using the entire APP support network serving as background control. Then, the MEME discovered TF binding site motifs were analyzed further for matches to known TF binding sites for mammalian transcription factors in the motif databases, JASPAR Vertebrates [87], via motif comparison tool, TOMTOM[88]. We summarized all the enriched GO terms using 'Revigo'[89] (Figure S10) on the 81 GoMo identified specific enriched GO terms in the Biological Process (Table S2).

Acknowledgments

This work was supported by NIGMS (R35 GM119850, NEL) and the Ministry of Education (Taiwan). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data Collection was supported through funding by NIA grants P30AG10161, R01AG15819, RR01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute. ROSMAP data can be requested at https://www.radc.rush.edu.

References

1. Novick, P., Ferro, S. & Schekman, R. Order of events in the yeast secretory pathway. *Cell* **25**, 461–469 (1981).
2. Reynaud, E. G. & Simpson, J. C. Navigating the secretory pathway: conference on exocytosis membrane structure and dynamics. *EMBO Rep.* **3**, 828–833 (2002).
3. Hukelmann, J. L. *et al.* The cytotoxic T cell proteome and its shaping by the kinase mTOR. *Nat. Immunol.* **17**, 104–112 (2016).
4. Tan, H. *et al.* Integrative Proteomics and Phosphoproteomics Profiling Reveals Dynamic Signaling Networks and Bioenergetics Pathways Underlying T Cell Activation. *Immunity* **46**, 488–503 (2017).
5. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
6. Lund, A. M. *et al.* Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Syst. Biol.* **11**, 37 (2017).
7. Feizi, A., Gatto, F., Uhlen, M. & Nielsen, J. Human protein secretory pathway genes are expressed in a tissue-specific pattern to match processing demands of the secretome. *npj Systems Biology and Applications* **3**, 22 (2017).
8. Feizi, A., Österlund, T., Petranovic, D., Bordel, S. & Nielsen, J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS One* **8**, e63284 (2013).
9. Gutierrez, J. M. *et al.* Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. (2018) doi:10.1101/351387.
10. Vassar, R. *et al.* Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science* **286**, 735–741 (1999).
11. Lammich, S. *et al.* Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3922–3927 (1999).
12. De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387–390 (1998).

13. Thinakaran, G. & Koo, E. H. Amyloid precursor protein trafficking, processing, and function. *J. Biol. Chem.* **283**, 29615–29619 (2008).

14. Wang, X. *et al.* Modifications and Trafficking of APP in the Pathogenesis of Alzheimer's Disease. *Front. Mol. Neurosci.* **10**, 294 (2017).

15. Lee, M.-S. *et al.* APP processing is regulated by cytoplasmic phosphorylation. *J. Cell Biol.* **163**, 83–95 (2003).

16. Schedin-Weiss, S., Winblad, B. & Tjernberg, L. O. The role of protein glycosylation in Alzheimer disease. *FEBS J.* **281**, 46–62 (2014).

17. McFarlane, I., Breen, K. C., Giamberardino, L. D. & Moya, K. L. Inhibition of N-glycan processing alters axonal transport of synaptic glycoproteins in vivo. *Neuroreport* **11**, 1543–1547 (2000).

18. McFarlane, I., Georgopoulou, N., Coughlan, C. M., Gillian, A. M. & Breen, K. C. The role of the protein glycosylation state in the control of cellular transport of the amyloid β precursor protein. *Neuroscience* **90**, 15–25 (1999).

19. Joshi, G. & Wang, Y. Golgi defects enhance APP amyloidogenic processing in Alzheimer's disease. *Bioessays* **37**, 240–247 (2015).

20. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).

21. Dourlen, P., Kilinc, D., Malmanche, N., Chapuis, J. & Lambert, J.-C. The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathol.* **138**, 221–236 (2019).

22. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).

23. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).

24. Ridge, P. G. *et al.* Assessment of the genetic variance of late-onset Alzheimer's disease. *Neurobiol. Aging* **41**, 200.e13–200.e20 (2016).

25. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset

Alzheimer's disease. *Cell* **153**, 707–720 (2013).

26. Greenfield, J. P. *et al.* Endoplasmic reticulum and trans-Golgi network generate distinct populations of Alzheimer beta-amyloid peptides. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 742–747 (1999).
27. Hartmann, T. *et al.* Distinct sites of intracellular production for Alzheimer's disease A beta40/42 amyloid peptides. *Nat. Med.* **3**, 1016–1020 (1997).
28. Wang, M. *et al.* The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific Data* **5**, 180185 (2018).
29. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
30. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **5**, e1000598 (2009).
31. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
32. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
33. Pearl, L. H. & Prodromou, C. Structure and mechanism of the Hsp90 molecular chaperone machinery. *Annu. Rev. Biochem.* **75**, 271–294 (2006).
34. Anelli, T. & Sitia, R. Protein quality control in the early secretory pathway. *EMBO J.* **27**, 315–327 (2008).
35. Bonifacino, J. S. & Glick, B. S. The Mechanisms of Vesicle Budding and Fusion. *Cell* **116**, 153–166 (2004).
36. Ikawa, M. *et al.* The putative chaperone calmegin is required for sperm fertility. *Nature* **387**, 607–611 (1997).
37. Rovelet-Lecrux, A. *et al.* APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* **38**, 24–26 (2006).
38. Bushman, D. M. *et al.* Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *Elife* **4**, (2015).
39. Matsui, T. *et al.* Expression of APP pathway mRNAs and proteins in Alzheimer's disease. *Brain Res.* **1161**, 116–123 (2007).

40. De Strooper, B., Vassar, R. & Golde, T. The secretases: enzymes with therapeutic potential in Alzheimer disease. *Nat. Rev. Neurol.* **6**, 99–107 (2010).

41. Laird, F. M. *et al.* BACE1, a major determinant of selective vulnerability of the brain to amyloid-beta amyloidogenesis, is essential for cognitive, emotional, and synaptic functions. *J. Neurosci.* **25**, 11693–11709 (2005).

42. Frost, G. R. & Li, Y.-M. The role of astrocytes in amyloid production and Alzheimer's disease. *Open Biol.* **7**, (2017).

43. Phatnani, H. & Maniatis, T. Astrocytes in neurodegenerative disease. *Cold Spring Harb. Perspect. Biol.* **7**, (2015).

44. Sofroniew, M. V. & Vinters, H. V. Astrocytes: biology and pathology. *Acta Neuropathol.* **119**, 7–35 (2010).

45. Liddelow, S. A. & Barres, B. A. Reactive Astrocytes: Production, Function, and Therapeutic Potential. *Immunity* **46**, 957–967 (2017).

46. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).

47. Cataldo, A. M. *et al.* Endocytic Pathway Abnormalities Precede Amyloid β Deposition in Sporadic Alzheimer's Disease and Down Syndrome. *Am. J. Pathol.* **157**, 277–286 (2000).

48. Israel, M. A. *et al.* Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells. *Nature* **482**, 216–220 (2012).

49. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

50. Krämer, A., Green, J., Pollard, J., Jr & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).

51. Citron, B. A. *et al.* Transcription factor Sp1 inhibition, memory, and cytokines in a mouse model of Alzheimer's disease. *Am. J. Neurodegener. Dis.* **4**, 40–48 (2015).

52. Santpere, G., Nieto, M., Puig, B. & Ferrer, I. Abnormal Sp1 transcription factor expression in Alzheimer disease and tauopathies. *Neurosci. Lett.* **397**, 30–34 (2006).

53. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards.

Database **2017**, (2017).

54. Allen, M. *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data* **3**, 160089 (2016).
55. De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data* **5**, 180142 (2018).
56. Bertram, L., Lill, C. M. & Tanzi, R. E. The genetics of Alzheimer disease: back to the future. *Neuron* **68**, 270–281 (2010).
57. Walther, D. M. *et al.* Widespread Proteome Remodeling and Aggregation in Aging C. elegans. *Cell* **168**, 944 (2017).
58. Brehme, M. *et al.* A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* **9**, 1135–1150 (2014).
59. Hipp, M. S., Kasturi, P. & Hartl, F. U. The proteostasis network and its decline in ageing. *Nat. Rev. Mol. Cell Biol.* **20**, 421–435 (2019).
60. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
61. Burrinha, T., Gomes, R., Terrasso, A. P. & Almeida, C. G. Neuronal aging potentiates beta-amyloid generation via amyloid precursor protein endocytosis. *Neuroscience* 1633 (2019).
62. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
63. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608 (2016).
64. Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184–185 (1992).
65. Lewis, J. *et al.* Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP. *Science* **293**, 1487–1491 (2001).
66. Hardy, J., Duff, K., Hardy, K. G., Perez-Tur, J. & Hutton, M. Genetic dissection of Alzheimer's disease and related dementias: amyloid and its relationship to tau. *Nat. Neurosci.* **1**, 355–358 (1998).

67. Bateman, R. J. *et al.* Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* **367**, 795–804 (2012).

68. Jack, C. R., Jr *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**, 207–216 (2013).

69. Meex, R. C. *et al.* Fetuin B Is a Secreted Hepatocyte Factor Linking Steatosis to Impaired Glucose Metabolism. *Cell Metab.* **22**, 1078–1089 (2015).

70. Gorden, D. L. *et al.* Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J. Lipid Res.* **56**, 722–736 (2015).

71. Robinson, J. L., Feizi, A., Uhlén, M. & Nielsen, J. A Systematic Investigation of the Malignant Functions and Diagnostic Potential of the Cancer Secretome. *Cell Rep.* **26**, 2622–2635.e5 (2019).

72. Page, L., Brin, S., Motwani, R. & Winograd, T. The Pagerank Citation Ranking: Bringing Order to the web. *technical report* (1998).

73. Huang, J. K. *et al.* Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst* **6**, 484–495.e5 (2018).

74. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, (2017).

75. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).

76. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers. Dis.* **64**, S161–S189 (2018).

77. Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, (2019).

78. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. (CRC Press, 2020).

79. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

80. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

81. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X

experiments. *Bioinformatics* **26**, 2438–2444 (2010).

82. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).
83. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
84. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *Bioinformatics* 471 (2016).
85. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
86. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–49 (2015).
87. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–4 (2004).
88. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8**, R24 (2007).
89. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).