Abstract

Genetic studies in late-onset Alzheimer disease (LOAD) are aimed at identifying core disease mechanisms and providing potential biomarkers and drug candidates to improve clinical care for AD. However, due to the complexity of LOAD, including pathological heterogeneity and disease polygenicity, extracting actionable guidance from LOAD genetics has been challenging. Past attempts to summarize the effects of LOAD-associated genetic variants have used pathway analysis and collections of small-scale experiments to hypothesize functional convergence across several variants. In this review, we discuss how the study of molecular, cellular and brain networks provides additional information on the effect of LOAD-associated genetic variants. We then discuss emerging combinations of omic data types in multiscale models, which provide a more comprehensive representation of the effect of LOAD-associated genetic variants at multiple biophysical scales. Further, we highlight the clinical potential of mechanistically coupling genetic variants and disease phenotypes with multiscale brain models.

Introduction

Alzheimer disease (AD) is a common, progressive, lethal neurodegenerative disorder. No preventative or curative treatment exists for this disease with high emotional and economic costs. In the USA, the prevalence of AD is expected to triple by 2050, but the global ramifications are even larger, as at that time, the majority of cases will occur in
developing countries. Genetic factors explain most of the variation in the risk of AD, especially in familial AD (a rare form of the disease with early onset), in which most genetic variants relate to amyloid-β (Aβ) processing. However, no single genetic, lifestyle or environmental factor is sufficient to predict late-onset AD (LOAD) with clinically relevant certainty. Variants have been detected in more than 20 genes, and are involved in a wide range of functions including metabolism, inflammation, synaptic activity and intracellular trafficking. Functional effects of these genetic variants have been observed across multiple cell types and on processes such as intracellular signalling, cell morphology and regional brain connectivity, which span several physiological scales. However, the identification of LOAD variants has not yet led to preventative treatments or provided clinical guidance for carriers of specific AD-associated genetic variants.

Previous work on AD genetics has demonstrated the utility of grouping disease-associated variants into canonical pathways. However, manually curated pathways and scientific literature are incomplete, can lack disease specificity, and can reflect historical biases. Network-based approaches (Box 1), comprise complementary methods to identify the function of genetic variants and the basis of AD pathology, by mapping them onto the interactions between the components of various biological systems. This framework facilitates the use of recent omics datasets and primary data from cohorts of individuals with AD to explore the effects of AD genetic variants.

In this review, we focus on how molecular networks provide a functional context for genetic risk variants in AD, and their potential for personalized diagnosis and disease stratification. We first consider ongoing efforts to identify genetic variants and current assessments of their role in AD pathology. Then, we examine network approaches to understand the function of AD variants, both at the molecular and whole-brain neuroimaging levels. Finally, we consider how multiscale models might bridge the gap between the effects of AD-associated variants at the molecular level and their function at the level of the brain, which could enhance the clinical relevance of genetic variants.

**Early genetic findings in AD**

**The Aβ hypothesis and APOE variants**

Early genetic findings reported large effect sizes of genetic variants occurring in a group of genes that encode interacting proteins involved in Aβ processing. The resulting amyloid hypothesis profoundly influenced the direction of studies designed to prevent AD or modify the course of the disease. Although these findings have been applied to understanding LOAD, they were obtained in the context of familial Alzheimer disease (which is typically early-onset), with the discovery of mutations in the genes APP, PSEN1 and PSEN2, and of relatively common AD-associated alleles of APOE, which encodes a protein that physically interacts with Aβ and tau, and can substantially increase risk of LOAD. The amyloid hypothesis has dominated the field for decades, and continues to be the major influence on clinical trials.

Several trials of Aβ modulators are ongoing. For example, the Alzheimer's Prevention Initiative launched two clinical trials to evaluate the effects of early anti-amyloid treatment...
in populations at risk of AD — a large familial group in Columbia with early-onset AD driven by \textit{PSEN1} mutations\textsuperscript{28}, and in individuals with homozygous \textit{APOE} ε4 alleles. Similarly, the Dominantly Inherited Alzheimer’s Network (DIAN) study selected participants from approximately 500 families carrying dominant mutations in \textit{APP}, \textit{PSEN1} or \textit{PSEN2} and is testing the effect of Aβ modulators\textsuperscript{29}. Pioglitazone, a PPAR-γ agonist that is already approved and in widespread use for the treatment of type 2 diabetes, is also being tested for efficacy in preventing AD in people with specific \textit{APOE–TOMM40} haplotypes\textsuperscript{30}. These trials, and many others, represent ongoing high levels of investment in the amyloid hypothesis.

Since the US government began tracking AD clinical trials consistently in 2002, more than 400 clinical trials have tested more than 200 compounds\textsuperscript{27}, of which only Memantine showed a mild reduction in symptoms\textsuperscript{31,32}. Half of these trials involved Aβ-related therapies and failed to improve cognition in patients with AD, even when they significantly reduced Aβ levels\textsuperscript{33-36}. These results could have been expected given that after the onset of overt cognitive impairment, increasing Aβ levels are only weakly correlated with further impairment in cognitive function\textsuperscript{37-39}. Thus, therapeutic intervention in the early stages of LOAD might be required for Aβ-targeted therapies to show efficacy, a concept being tested in the ongoing trials in individuals with mild cognitive impairment (MCI)\textsuperscript{35} or high-risk asymptomatic participants\textsuperscript{40}. It should be noted, however, that long-term administration of expensive agents, with potentially severe adverse effects, warrants careful ethical consideration, as AD-associated brain changes begin a decade or longer before AD-related cognitive decline\textsuperscript{29,41,42}, and many individuals with moderate AD neuropathology do not have cognitive impairment\textsuperscript{37,43}.

Regardless of the actual source, the continued failure of trials that evaluate AD treatments is disconcerting, in light of the large and increasing public health burden of the disease\textsuperscript{7}. While these trials testify to the strength of the early findings on genetic perturbations of the \textit{APP} processing in AD, this approach is still strongly tied to scientific advances made three decades ago, as opposed to incorporating more-recent genetic findings. It might therefore be helpful to use existing genetics findings and novel ‘omics’ technologies to uncover the disease mechanisms of AD and identify novel therapeutic targets.

**Genome-wide association studies**

After the early work on variants in early-onset AD, genetic research shifted towards the identification of AD-associated variants that have weak effects or are very rare. To date, GWAS have identified more than 20 genetic variants that influence the risk of AD\textsuperscript{12,44,45}. Although these studies have provided insights into possible pathophysiological mechanisms of AD\textsuperscript{14,15,22}, AD-associated variants that are rare or that have a weak effect are challenging to use in prognosis\textsuperscript{46,47}, as their associated disease risk is much less than that of \textit{APOE} variants, even when the variants are considered in aggregate\textsuperscript{48}. These findings raise the question of how the identification of a wide range of variants can contribute to a coherent theory of AD pathology. In the past two years, whole exome or targeted sequencing studies\textsuperscript{49-51} have discovered rare variants that increase the risk of AD more strongly than do...
common variants (certain *TREM2* variants are on par with the odds ratio of *APOE* e4 variants).

Ongoing whole-genome sequencing projects, such as the Alzheimer’s Disease Sequencing Project will continue to identify susceptibility loci in AD; however, the overall scientific returns of hunting for genetic variants are diminishing relative to the increasing size of the projects. Gaining insight into the multigenic and multifactorial disease mechanisms in AD requires methods that go well beyond simple genotype–phenotype models. We must now start to scientifically consider and understand many genes, many phenotypes, and the interactions between them. This nascent third stage of AD genetics, the main focus of this review, uses systems biology methods that emphasize how variants interact and how their effects propagate through networks of molecular and brain circuit interactions, which span biophysical scales.

**Molecular interactions in AD**

**Epistatic regulation of AD**

The effect of the interaction between genetic variants might help explain the gap between the high heritability of AD and the weak effect of genetic variants. Epistasis, which can consist of synergistic or dysynergistic gene–gene interactions, is the simplest form of multigene interaction and is related to phenotypes or disease status. As described below, the combinatorial nature of epistasis may be useful in accounting for the personalized molecular basis for AD risk in a given patient. Incorporating epistatic interactions to understand AD heritability provides a direct connection between traditional additive genetics (GWAS) and the large-scale molecular-interaction networks described in the following sections.

Early attempts to find epistasis in AD were largely hypothesis-driven, and only a minority of the proposed epistatic interactions were replicated. Given the large burden associated to multiple testing and the large sample size necessary to detect genetic interactions, most studies have been focused on detecting interactions between marginally associated variants, and primarily found effects between such variants and *APOE* haplotypes. However, certain gene–gene interactions can be associated with AD even if neither locus is independently associated with the disease. Considering all possible combinations of variants produces a high multiple testing burden, which is magnified when examining multiple phenotypes, such as the volume of multiple brain regions. Therefore, some studies have opted for a middle ground: they limited epistasis testing to pairs of genes in the same functional pathways, and found epistasis for hippocampal atrophy and temporal lobe volume.

**Edgetic effects**

Genetic variation that alters the affinity of specific protein–protein interactions (the edges between proteins in a network diagram), as opposed to all the interactions of a given protein, are referred to as ‘edgetic’ effects. Edgetic interactions are based on protein binding information, in contrast to the epistatic effects mentioned above, which are primarily statistical interactions, whose physical basis may be unclear. Edgetic alterations resulting
from genetic variants can be identified experimentally or predicted by high-resolution, three-dimensional models of protein structure and protein–protein interactions (FIG. 1a). This concept is helpful to associate specific alterations in protein interactions with disease states. For example, pleiotropic effects can sometimes be explained by alterations of distinct protein interfaces of a single protein: each interface affects distinct partners and triggers distinct phenotypes. In the context of Mendelian disorders, disease-associated genetic variants are likely to result in edgetic effects. Whether this edgetic framework can be applied to study the effects of AD genetic variants is unclear.

Targeted sequencing or exome sequencing efforts carried out in the past few years have identified coding variants in ABCA7, ADAM10, BIN1, CD2AP, CLU, CR1, EPHA1, MS4A4A/MS4A6A, PICALM, PLD3, SORL1 and TREM2 that are associated with late-onset AD, in addition to the familial AD variants in PSEN1/2 and APP. Thus it is likely that, as whole-genome sequencing becomes more common, person-specific coding variants might also be mapped edgetically to generate detailed individual molecular networks—a process that is already possible for several thousand proteins with known structures (FIG. 1a).

Network-based approaches to AD genetics

Aggregating a collection of epistatic or edgetic interactions to identify specific disease mechanisms is still challenging, owing to missing information about how they affect specific cellular systems or neuronal activity. Therefore, alternative network-based approaches (Box 1) leverage large-scale molecular networks to help identify coherent biological functions. The structure of these molecular networks contains information that is useful to determine the function of biological systems. For instance, molecules involved in a particular biological function tend to be densely and mutually connected. Network approaches are different from pathway analysis: they can provide tissue-specific and disease-specific information from the latest omic technologies, and are not limited by the state of knowledge about canonical pathways. In some cases, network-based approaches intersect with the study of AD-associated genetic variants, but we also review select cases outside of AD, which involve principles or practical approaches that could be useful in characterizing the function of AD variants.

Coexpression networks

Coexpression-based networks (FIG. 1a, b, Box 2) are a common type of data-driven network, in which links represent the strength of gene–gene correlations. These links between genes can be generated by many mechanisms such as microRNAs, cell-type specificity, chromosome conformation, epigenetics and cell type proportions, that regulate gene expression. Correlation links between genes are therefore useful in the context of disease: clusters of coexpressed genes can be a proxy for alterations in gene expression regulatory mechanisms, and the level of coexpressed genes can be matched to disease phenotypes to prioritize certain molecular systems for follow-up experiments. Advantages of the coexpression-analysis approach include coherence with genetic findings, potential tissue-specificity and greater robustness than univariate or single–gene approaches.
**Genetic integration**—Genetic variants that interact in protein–protein interaction (PPI) networks (Box 2) and tend to be coexpressed have been identified in AD, autism and schizophrenia. In autism, the coherence between coexpression and PPIs has been used to filter for molecular systems enriched in *de novo* mutations, in autistic individuals.

**Tissue specificity**—Obtaining tissue-specificity and building coexpression networks in tissues relevant to disease can be crucial for the identification of relevant and accurate coexpression networks in neurological disorders. The Genotype-Tissue Expression project (GTEx) consortium was created to make tissue-specific data more accessible by generating transcriptomic profiles of a large number of human tissues, with plans to examine the genetic effects of the tissue context.

**Robustness**—Patterns of gene coexpression in human CA1 and CA3 were compared to disease progression and pathology to prioritize disease-associated molecular systems. These coexpression clusters were identified without using genetic information per se; however, they showed coherence with subsequent coexpression, genetic and experimental studies, in terms of the TYROBP–TREM2 signalling cascade.

**Network-based stratification**

Network-based patient stratification, an approach utilized in oncology, provides a clear example of how molecular interaction networks can aggregate diverse genetic effects to uncover disease mechanisms that are relevant to personalized treatment. Briefly, genes carrying disease-linked variants were located in publically-available PPI networks (Box 2). Variants carried by patients with similar prognoses were mostly found in genes that were located in certain areas of the PPI network. Thus, the protein-protein interaction network can be used to stratify patients and to highlight candidate disease mechanisms that involve molecules in the affected area of the network. Drug–target interactions can also be included in these networks to enhance their clinical utility: such networks would enable the identification of drugs that target the specific regions of a network that are affected in a particular person. Although in AD, the effects of individual genetic variants are relatively small, network-based stratification is an example of how genetic variation can be aggregated in a clinically relevant manner. Applications of this tool to large cancer cohorts has provided a roadmap for how molecular networks could be helpful in identifying personalized mechanisms of AD, especially as whole-genome sequencing becomes commonplace. Similar approaches have been utilized in schizophrenia and autism to find network regions enriched in disease-associated mutations, though such approaches have yet to be applied to AD.

**Analysis of directed networks**

To estimate the effects of disease-associated variants, the studies cited thus far utilized undirected networks, such as coexpression networks or PPI’s, in which elements are connected by links without specific direction that would imply causality. Combining the genetic analysis of pathological variants with gene expression data provides a causal basis for expression changes, which can be used to infer directed networks. Previously, the use
of directed networks required large computing clusters, but the CINDERella has enabled researchers to infer disease-specific directed networks using standard hardware\(^1\).

A previous generation of the directed/causal network approach was applied to microarray and single nucleotide polymorphism (SNP) data obtained from tissue samples of the prefrontal cortices of 549 healthy individuals and individuals with AD\(^2\). This approach, which included the analysis of expression quantitative trait loci (eQTLs), implicated TYROBP (which encodes part of a microglial membrane receptor complex) as an upstream node in an AD-associated network of several hundred genes related to multiple immune subsystems, including previous AD susceptibility genes identified from GWAS CD33, MS4A4A, MS4A6A and TREM2.

**Disease state-specific networks**

Many network studies were carried out with the assumption that networks generated from datasets in control/healthy systems are sufficiently similar to networks found in the disease state; however, pathological processes or responses to disease can alter network structure. For example, certain gene–gene correlations might only be observed in disease-state data (differential coexpression)\(^3,4\). Traditional differential expression analysis might not detect cases of differential coexpression because gene–gene correlations can be altered without changes in the average expression level of the two genes\(^1\). Differential coexpression has been observed between several molecular systems when comparing control to AD gene expression patterns\(^2,5\) and between PSEN1 and groups of genes highly expressed in oligodendrocyte and microglia, when comparing mouse and human coexpression patterns\(^6\). However, it is challenging to collect samples of ‘true’ control-state or disease-state networks, as the phenotypic variability in AD is a continuum and AD pathology can be present long before the onset of clinical signs and symptoms\(^7\). A hybrid approach, which requires the presence of both clinical and pathological manifestations of the disease to establish an AD diagnosis, might lead to the loss of information on the unique molecular basis of the different—but related—phenotypic aspects of AD\(^8\).

**Neuropathological studies of AD variants**

A fundamental assumption of AD case–control GWAS is that persons can be meaningfully classified according to observable clinical behaviour and accepted definitions of the disease. However, many individuals who are classified as ‘cognitively normal’ harbour AD, infarcts and Lewy bodies\(^9\), and many individuals who are clinically diagnosed with AD are affected by other pathologies\(^10\) (Box 3). This inaccuracy in disease status and the many comorbid (yet typically unmeasured) conditions present in older populations reduce statistical power and places practical limits on the effectiveness of case–control studies of AD. By conducting GWAS on AD endophenotypes (such as neuropathological and neuroimaging features), it is possible to reduce the effect of confounding factors associated with subclinical disease and comorbid conditions, and to increase statistical power with quantitative outcomes\(^11\). The ultimate goal of this approach is to compile these partially overlapping genetic signatures into a more robust description of AD.
Collecting neuropathological data is onerous, but such efforts are leading to the identification of novel loci associated with different types of AD and their related neuropathological features\(^98,102\). GWAS on common age-related neurodegenerative disorders, such as hippocampal sclerosis and cerebral amyloid angiopathy, have identified loci involved in each pathology, half of which are also found via case–control AD GWAS\(^98\). Combining genetic data with detailed neuropathological phenotypes has the advantage of placing the genetic variants in the context of a particular pathophysiological process. Moreover, this approach has identified loci that are not found by traditional GWAS, and that contribute to pathologies associated with AD dementia.

**Insights from AD networks neuroimaging**

Genetics shape brain structure, connectivity and function\(^103\). Alterations in various brain regions have been linked to AD\(^104\), with canonical sites of early AD-associated effects in the hippocampal region\(^105,106\). To understand how AD affects macroscopic brain structures and brain dynamics on a global scale that is relevant to cognitive functions, some neuroimaging efforts have adopted a network–based approach\(^107-109\). This work is based on the observation that the networks of the healthy brain must simultaneously achieve two objectives, which can appear opposed. On one hand, brain networks must support functional specialization, which requires some isolated modules or clusters; on the other hand, brain networks must support coordination and information flow between diverse systems, which requires the existence of short-cut paths between modules\(^110,111\). Healthy brain networks generally exhibit a balance of high modularity (locality) and short path lengths (integration) to support perceptual and cognitive processing\(^107,109,112,113\). In the field of network neuroimaging, AD pathogenesis has been characterized as an imbalance between modular and integrative processes (described below).

**Balancing local and global functions**

When a network simultaneously supports modular function (local clusters) as well as integration (a short average path length between nodes), it is said to have a ‘small world’ (SW) organization\(^114\)(FIG. 1c). The extent of SW organization in brain networks is heritable\(^115,116\), and predictive of AD status\(^117,118\) and progression\(^119\). Moreover, changes to the SW balance can be detected prior to neurodegeneration or cognitive decline in people with elevated brain amyloid levels\(^120\).

In AD, changes in SW organization can arise from changes in the shortest average path between nodes and/or changes to the clustering coefficient\(^121,122\)(FIG. 1c). These changes, which may be due to amyloid accumulation in the hub regions of the brain\(^123\), can be visualized with several imaging methods, including functional MRI (fMRI)\(^121,124\), magnetoencephalography\(^125\), EEG\(^124\) and structural imaging\(^122,126\). Importantly, although the presence of short paths in networks has been associated with cognitive function in healthy cohorts\(^112,127\), specific cognitive deficits in AD have not been conclusively linked to this metric. Cohort studies have shown that the alteration of brain networks in MCI is generally less severe than in AD\(^118\); however, some individuals with MCI can show
selectively increased connectivity as a potential compensatory mechanism,\textsuperscript{124,128,129} (FIG. 1c).

**Critical nodes: hubs and rich clubs**

Brain region connectivity has implications for the location and propagation of AD pathology. For instance, ‘hubs’ are regions with many connections to other regions, and changes in hubs are implicated in multiple brain disorders.\textsuperscript{130} Hubs are preferentially affected in AD, potentially owing to unique metabolic demands or cellular processes in these regions.\textsuperscript{131,132} A large study comprising patients with MCI and AD showed decreased global integration and decreased interconnections mediated by hubs within the default mode network (DMN).\textsuperscript{133} (FIG. 1c). Brain hubs that are densely coupled to each other are termed a ‘rich club’.\textsuperscript{134} Damage to rich club nodes could be especially disruptive to brain function, as they are the intersection point for many paths that link distal regions.\textsuperscript{135} Consistent with this hypothesis, the connectivity of rich club nodes is affected in APOE ε4 carriers, and alterations of hub connectivity are correlated with cognitive performance in patients with AD.\textsuperscript{137} However, connectivity of rich club brain regions could be more strongly altered in early-onset AD and frontotemporal dementia than in LOAD.\textsuperscript{138,139}

**Network changes linked to AD genetics**

Global and local brain network changes are associated with AD diagnosis, but few studies have examined the effects of AD-associated genomic variants on brain networks, independently of diagnosis. Most of the studies focusing on APOE ε4 carriers versus non-carriers have found alterations of the DMN,\textsuperscript{140-142} and of the SW integration-modularity balance,\textsuperscript{143} with studies of all APOE alleles pointing towards additional subnetworks, the activation of which is influenced by genetics.\textsuperscript{144} These changes take place years before the onset of AD,\textsuperscript{140} are stable through mid-life,\textsuperscript{145} and might be associated with decreased connectivity and metabolism in the DMN in AD.\textsuperscript{146-148} Similar effects on the DMN are seen in the form of familial AD that results from mutations in genes involved in Aβ processing.\textsuperscript{149} Although an AD-associated CLU variant has also been shown to correlate with brain structural path lengths in healthy individuals,\textsuperscript{150} most genetic studies on brain networks in AD have focused on APOE, rather than other recent AD GWAS hits. It is unclear if this is due to historical effects or reflects truly larger effects of APOE alleles on brain network structure.

**Challenges and next steps**

As mentioned above, an imbalance of modularity and integration is regularly reported in individuals with AD. However, the features that drive the imbalance varies across studies and methods. Many (but not all) imaging studies have found increased path length in AD brain networks,\textsuperscript{121,122,154-157} in particular, the studies that measured structural connectivity networks. By contrast the other component of SW networks – clustering coefficient – is decreased in AD in some studies,\textsuperscript{125,151,152,158,159} and, increased in others.\textsuperscript{126,157,160} Differences in imaging modalities, network connectivity normalization in the context of neurodegeneration, or intrinsic limitations of the resting-state paradigm might be responsible for this range of results.
Although inferred functional brain networks are constrained by underlying structural connections\textsuperscript{161}; the correspondence of structural and functional networks is not straightforward\textsuperscript{162}: and all networks should not be expected to show identical effects in AD. This confounding factor is exacerbated by variations in cohort characteristics, and by the effects of haemodynamics and arousal, which are difficult to control in clinical neuroimaging settings\textsuperscript{163,164}. Guidelines are emerging for optimal data preprocessing\textsuperscript{165}, which should improve the consistency and replicability of network characterizations in AD. Neuroimaging studies sometimes treat brain networks to be static; however, dynamic functional connectivity studies suggest that resting-state networks are a blurred representation of transiently activated regional subnetworks\textsuperscript{107,166-168}. Considering these network fluctuations in disease states\textsuperscript{169-171} may resolve discrepancies among studies that have assumed a constant network architecture\textsuperscript{171}.

To truly characterize the multiscale processes that are affected in AD, it is crucial to develop network signatures that go beyond the measurement of the SW balance\textsuperscript{172}. This can be done by identifying specific network systems (e.g. configurations of DMN interactions) that are associated with functions (e.g. spatial episodic memory) that can be behaviourally assayed in individuals at risk of AD. This way, the ‘function’ of a particular functional network can be connected more directly to a behavioural phenotype of AD. Finally, as we discuss in the next section, it is crucial that the next generation of network signatures are derived from biologically constrained multiscale models of AD pathology.

**Multiscale models of AD**

To date, studies have typically examined the effects of AD genetics on molecular networks (FIG. 1a) or on large-scale brain networks (FIG. 1c). Those two perspectives are missing a description of how the molecular effects of AD-associated variants translate into structural brain changes, which in turn can represent cognitive decline in AD. Describing how molecular and brain networks are mechanistically coupled together may help to reduce false positives in drug development, as the molecular assays to test drugs could be linked to brain activity measures that are closer to clinical-level effects. Similarly, brain connectivity and neuroimaging findings are generally described in a manner that is uncoupled from molecular activity. Coupling brain connectivity to molecular properties provides an experimentally tractable basis to address changes in disease.

Multiscale models help to fill the gap between the effects of genetic variants on molecular networks and brain networks\textsuperscript{173,174}. These models go beyond typical imaging-genetics approaches, and can be useful in identifying which molecular and cellular features have an effect on a given tissue or have clinical-level properties, such as patterns of fMRI connectivity that are associated with cognitive function. A concrete example of a neuronal property that could be well-represented by multiscale models is long-term potentiation (LTP) of synapses, which is triggered by the coincident presynaptic input with large postsynaptic depolarization. While initiation of LTP is rapid, the gene expression, translation, and cytoskeletal rearrangement it entails are relatively slow. In the context of AD, a multiscale model could include the experimentally observed effects of AD-associated variants on cellular processes (e.g. the effects of amyloid on LTP), which are then ‘scaled
up’ into larger and more realistic brain networks, through computational simulations (Fig. 2).\textsuperscript{175-178}

**Potential to complement AD research**

Multiscale models are particularly important for AD because the effects of the disease on executive control and memory can be realistically represented by whole-brain models\textsuperscript{109,179}, whereas therapies are generally developed at the molecular level. These models have the potential to rapidly identify specific molecular properties that exert the strongest effects on the clinical readout\textsuperscript{180,181,182}; thus, they are valuable in silico tools to understand drug mechanisms of action and to evaluate the molecular systems affected by an individual’s genetics. The potential use of multiscale models should not be perceived as a replacement for experimental work; it is rather a way of aggregating and extracting the implications of experimental results. Such models are necessary given the diverse systems involved in AD pathology, its decades-long prodrome, and the lack of sufficiently predictive animal models\textsuperscript{183-185}.

Presently, network models of AD typically operate on a single physical scale, and ignore other spatial scales and the temporal component entirely.\textsuperscript{186} Whereas multiscale models are rapidly developing in other areas of neuroscience\textsuperscript{180,187-190}, they are only starting to be utilized in AD.\textsuperscript{191} Therefore, we examine below how multiscale modelling in other diseases and biological systems can enable realistic and rapid examination of the effect of AD genetic variants, in a way that is useful for clinical progress. In particular, two bridges between scales are likely to be essential for multiscale models of AD to become useful in a preclinical setting: gene– electrophysiology coupling, and neuron–tissue coupling.

**Gene expression and neuronal activity**

Both gene expression and electrophysiology studies have a strong influence on AD research. Integrating these approaches in a multiscale model would have conceptual and practical benefits for the development of AD therapeutics, as cell membranes are very accessible to drugs and can be mathematically modeled in great detail and are essential to brain function. Few studies describe the feedback between gene expression and electrophysiology in a manner that enables predictive multiscale models of AD. Nonetheless, a study of the suprachiasmatic nucleus (SCN)\textsuperscript{192} exemplified the potential of combining gene expression and neuronal activity into a unified multiscale model, which can be used to simulate neuronal activity under many different physiological conditions. The cellular aspects included in the simulations span a range of temporal and physical scales (from smallest to largest): intracellular calcium levels, circadian genes expression, ion channels, intracellular neuropeptides and synaptic connectivity. All of these components of the model were mathematically coupled with interacting differential equations to represent the relationships observed experimentally. This simulation was helpful for reconciling long-standing experimental differences observed between results obtained in various laboratories on the effect of \textgreek{g}amma-aminobutyric acid (GABA) on SCN activity. This example illustrates the complete lifecycle of multiscale models: the association of gene expression to systems-level phenomena, such as neurotransmission and circadian activity, resulting in testable hypotheses that are examined experimentally.\textsuperscript{193} Increasing the use of CRISPR, optogenetics
and single-cell RNAseq or ATACseq will facilitate experiments to simultaneously measure relationships between gene expression and electrophysiological activity in closely controlled systems. Such experiments are prime sources for the components of multiscale gene–electrophysiology models applicable to AD.

**Neuron-tissue coupling**

In the process of scaling the effect of genetic variants to the level of whole-brain activity, a second major bridge across physiological scales is between single neuron properties and large-scale networks, that model the activity of millions of neurons\(^\text{194-196}\). Simulations of large-scale networks can include different excitatory and inhibitory cell types, with their respective ion channel conductances, connected with realistic columnar and inter-regional patterns (FIG. 2b). Far more information is available to generate this aspect of the multiscale model, than is available to model the effects of genetic variants. A key challenge is therefore to determine the useful level of biological detail in simulations\(^\text{196}\). Simulated neuronal activity can be linked to brain imaging studies by pooling the activity of groups of neurons to represent different brain regions (FIG. 2c)\(^\text{194}\). Such whole-brain networks show dynamic response patterns that are associated with perceptual and memory processes, and are helpful to understand the interplay between chemical, structural and functional aspects of neurological diseases, as demonstrated in recent multiscale models of schizophrenia\(^\text{197}\). Although these models cannot yet be used to model the effect of genetic variants, they already contain downstream components of genetic effects, such as intracellular components, detailed neuronal morphology and connectivity, which, together, reproduce fMRI features associated with memory\(^\text{198}\) and cognition\(^\text{179}\). As such multiscale models begin to incorporate disease-specific effects, including genetics, they will become useful *in silico* screening tools for AD cellular interventions.

**Roadmap to multiscale models of AD**

Here, we specify the components of one of the possible multiscale models of AD and we provide one possible roadmap (FIG. 2) for scaling the effects of AD genetic variants up to the level of whole-brain models\(^\text{195,199-201}\). First, the effect of AD-associated variants on the electrophysiological and morphological features of neurons (either directly or via other cell types) can be assessed either by changes in ion channel populations or by fitting the observed electrophysiological recordings to mathematical models of neuronal activity\(^\text{202}\) (FIG. 2a). This strategy is helpful for bypassing some details of the intracellular signaling that transduces the effects of variants to membrane properties. In the second stage of multiscale modelling, the spiking and network activity of models with and without genetic variants (FIG. 2b) can be connected to each other with the desired level of detail, mainly limited by computational power. In sufficiently large numbers or with approximations of the average activity of large numbers of neurons, these computational models can generate a reasonable approximation of the whole-brain dynamics observed in resting-state fMRI studies\(^\text{194}\) (FIG. 2c).

**Practical challenges**

Challenges in the implementation of multiscale models that incorporate the effects of AD-associated genetic variants include identifying the exact location of the relevant variant,
which can often fall in noncoding regions, and the fact that some AD-associated genetic variants, particularly those that affect non-neuronal cell types, may not have known synaptic effects. The analysis of epigenetic information can narrow the size of the relevant locus. The epigenetic resources to localize AD-associated variants are limited, but they are rapidly expanding and becoming more relevant to brain cell types. The effects of non-coding AD-associated variants can be assessed by creating cell lines each with a mutation in different subregions of the locus, and evaluating the effects of those mutations in electrophysiological and cell imaging experiments. Effects observed in these experiments can be included in neuronal models of disease and compared with output of control-state models, or with the effects of other variants. Regarding variants that affect non-neuronal cell types, the effects of such cell types on neurons can be included in neuronal simulations. For example, certain TREM2 variants are associated with AD and activation of TREM2 signaling in microglia can lead to decreased neurite length. Such effects can be included in neuronal models, which are, in turn, components of whole-brain models.

Conceptual challenges

Ambitious modelling projects, such as multiscale modelling of the effects of disease-associated variants (FIG. 2), are sometimes dismissed because genes, cells and brain networks have some important properties that are not yet measured, and that can lead to variable results. Indeed, the absence of constraining data is a challenge for modeling efforts in many domains of biology, and has no easy solution. However large-scale and multiscale models have yielded useful insights in several areas of biology, including whole-cell modeling, cancer, cardiology, immunology and neuroscience. Moreover, the data required to constrain multiscale models is rapidly accumulating as part of the data-heavy, multicontinental collaboration initiatives led by private and public funders, such as the large-scale coordination of open science for target discovery in the Accelerating Medicines Partnership for Alzheimer Disease (AMP-AD). A particular advantage of multiscale models is that they remain constrained by a large set of results (FIG. 2): for example, fMRI data can be used to constrain the model parameters of cellular membranes by comparing the output of the model in ‘healthy’ and ‘disease’ states to actual healthy fMRI, EEG or magnetoencephalography results. Finally, we emphasize that these models do not attempt to copy the brain in all of its complexity. The goal is to simplify the real system, to distill some of its core functions with sufficient accuracy to predict the behavior of the real system in some limited setting. In this light, multiscale modelling is simply a more quantitative and systematic approach to the general scientific endeavour — one which explicitly couples different experimental programs, rather than leaving their interaction to chance.

Conclusions

The past decade of AD genetic discoveries has been marked by the search for coherence among disease-associated variants with weak effects and functional diversity. Concurrently, omic technologies and neuroimaging have produced detailed descriptions of molecular and brain networks. These trends of diverse genetic findings and biological networks are converging in studies of large AD cohorts, which shed light on the functional roles of AD-
associated variants and point towards convergent functions of such variants. At the molecular level, several types of molecular interactions including epistasis, protein–protein interactions and gene coexpression, define the intricate relationships between variants and genome-wide molecular systems. Measurements of differential coexpression and edgetic changes allow the identification of networks that only exist in the disease state, and these novel interaction structures may be crucial for understanding pathogenesis. The effects of individual and combinations of AD-associated genetic variants can also be observed on structural and functional brain connectivity patterns. These show that some variants affect the integration–segregation balance of brain networks, which is critical for perceptual and cognitive function. However, identifying how variants alter brain connectivity patterns entails understanding their effects on cell morphology and electrophysiology, and creating integrated multiscale models to capture their full effects on brain microcircuits and regional connectivity.

While molecular and brain networks are the most complete description of the biological processes altered in patients with AD to date, they are still incomplete, potentially biased, and represent a static picture of the disease. Advancing to the point where network tools can generate a dynamic, multiscale description of Alzheimer's that is sufficiently accurate to provide personalized diagnosis or screen potential therapeutic targets involves challenges in computational infrastructure, omics data acquisition and the social organization of science. Specifically, building multiscale models requires openness and novel collaborations among groups of investigators, breaking out from traditional academic boundaries, to become more aligned with patterns of molecular interactions. While daunting, such coordination between researchers is possible, as demonstrated by the centralized efforts to annotate genome-wide metabolic networks.

**Acknowledgements**

This work was supported by NIH grants U01AG46152, R01AG36042, P30AG10161, RF1AG15819, R01AG17917, R01AG36836.

**Biography**

Chris Gaiteri is a computational neuroscientist who studies the molecular and neuronal basis of psychiatric and neurodegenerative diseases. To help identify the origins of these diseases, he examines how biological functions can be represented by molecular and cellular interaction networks. He has applied this network perspective to omics research at Sage Bionetworks, in Seattle, Washington, USA, the Allen Institute for Brain Science, in Seattle, Washington, USA, and presently at Rush University, in Chicago, Illinois, USA, where he is an Assistant Professor.

Sara Mostafavi is a computational biologist who develops models and algorithms to understand the genetic and genomic basis of complex neurological and psychiatric disease. She is an Assistant Professor at the Department of Statistics and the Department of Medical Genetics, at University of British Columbia, Vancouver, Canada, and is a Canada Research Chair in Computational Biology. She obtained her PhD in Computer Science from the...
University of Toronto, Canada, and performed her postdoctoral fellowship at Stanford University, California, USA.

Christopher Honey is a cognitive neuroscientist who studies large-scale neural dynamics, asking how brain regions communicate and how memory arises in hierarchical brain processes. Born and raised in southern Africa, he studied mathematics and literature at the University of Cape Town, South Africa, obtained his PhD in psychology and cognitive science from Indiana University, Bloomington, Indiana, USA, completed postdoctoral training at Princeton University, New Jersey, USA and is an Assistant Professor at the University of Toronto, Canada.

Philip De Jager is an Associate Professor of Neurology at Harvard Medical School, Boston, Massachusetts, USA and Director of the Program in Translational NeuroPsychiatric Genomics within the Ann Romney Center for Neurologic Diseases in the Department of Neurology at Brigham and Women's Hospital, in Boston, Massachusetts, USA. He is the first incumbent of the Steven R. and Kathleen P. Haley Distinguished Chair for the Neurosciences. He is a practicing clinical neuroimmunologist. The goal of his work as a clinician-scientist is to apply modern methods of neuroimmunology, statistical genetics and computational biology to first delineate and then intervene in the sequence of events leading from health to neurodegenerative diseases.

David A. Bennett is the Director of the Rush Alzheimer's Disease Center, Chicago, Illinois, USA and the Robert C. Borwell Professor of Neurological Sciences at Rush University Medical Center. He studies the causes of common chronic neurologic conditions of the aging nervous system and novel drug target nomination and validation. He is principal investigator of several studies funded by the National Institutes of Health including the Religious Orders Study and Rush Memory and Aging Project, both community-based clinical-pathologic cohort studies that incorporate deep omics, biomarkers, neuroimaging, biomedical devices, and brain, spinal cord, nerve and muscle pathology. He has more than 600 peer-reviewed publications, 60,000 citations, and an h-index of 118.

Glossary

**Amyloid precursor protein (APP)**
Gene that encodes the amyloid-β precursor protein, which is cleaved to form amyloid-β peptides; presenilin (*PSEN1, PSEN2*) mutations promote the cleavage of APP into plaque-forming peptide type.

**APOE**
Dosage of the ε4 allele *APOE*, which encodes the apolipoprotein E, is the strongest genetic risk factor for late-onset Alzheimer disease; APOE interacts with amyloid-β to influence plaque aggregation.

**Amyloid hypothesis**
Proposal according to which the root cause of Alzheimer disease is the accumulation of amyloid-β, with nuances around sufficiency and form of amyloid-β responsible for pathogenesis.
Clustering coefficient
Measure of modularity around a network node. Number of connections among neighbors of a node, divided by the maximum possible number of connections among those neighbors

Cognitive reserve
Tolerance and adaptation to neuropathology, in part attributed to genetic and life-style associated factors such as education and social activity, and their neural correlates

Directed networks (directed graphs)
Networks where elements are connected by links with a specific direction that represents an asymmetric temporal or transfer function

Genetic variants
Insertion, deletion or alternative coding of DNA

Late-onset (sporadic) Alzheimer Disease (LOAD)
In contrast to early-onset Alzheimer disease, this is the most common form of the neurodegenerative disease, typically diagnosed clinically after age 65 and definitively diagnosed post-mortem. LOAD is associated with functionally diverse, weak genetic variants

Multiscale modelling
Mathematical or conceptual models that couple processes that occur on a varying range of physical or temporal scales, which are typically studied in isolation

Peroxisome proliferator-activated receptor gamma (PPAR-γ)
Component of a nuclear receptor complex, which includes the retinoid X receptor. This complex is activated endogenously by fatty acids and leads to transcription of APOE, among other genes

Pleiotropy
When a single gene or variants of that gene can contribute to two or more ‘non-related’ phenotypes

Small-world network
Networks in which most nodes are connected by a small number of hops, yet form relatively isolated (modular) clusters

References


Key points

- Genetic findings in late-onset Alzheimer disease (AD) have not yet resulted in strategies to prevent or treat AD
- Examining the position of genes carrying disease-associated variants in large-scale molecular networks can help identify coherent disease mechanisms
- Not all network approaches are equal: recent approaches involve networks that are directed (with causal links), and specific to the tissue and disease-state
- AD and AD-associated genetic variants decrease efficiency of information transfer in the brain connectome, which can be quantified by measuring structural and functional patterns in brain networks
- Building multiscale models of the effects of AD-associated genetic variants with large neuronal simulations is now feasible and will be useful to understand the effects of such variants and to screen therapeutics in silico
Box 1 | Network-analysis approaches

Over the past decade, the term “network analysis” has been used to refer to a variety of analytical approaches to analyze genomics data. By considering many relationships among biological entities simultaneously, network analyses shift the assumptions from a one-to-one genotype–phenotype mapping to a many-to-many mapping, which better accommodates the complexity of both the genetic basis and the phenotypic expression of chronic diseases. In this approach, networks (known as weighted graphs in Computer Science and Mathematics) provide a generic framework to represent relationships or links between a set of entities which, in the context of genomic analysis can be molecules, cells, tissues, or organisms. Network-based approaches utilize one or more large knowledge sources to describe the functional relationships between units. These networks are most commonly utilized in the ‘interpretation’ phase of a genome-wide analysis, to look for functional enrichment among the top hits. Increasingly, molecular networks have also been used during the discovery phase of genomic experiments (as with the NetWas approach) to identify novel genes or mechanisms associated with diseases. Two main sources of data can be used to construct gene networks: literature-curated interactions from low-throughput experiments, and high-throughput datasets that measure gene expression (and/or activity) across a large set of individuals or conditions. These two sources vary in false positive rates, coverage, prior knowledge, and context-specificity. The meaning of a link in a molecular network can vary depending on the type of network: links can be physical links between molecules, represent signal transduction such as phosphorylation, or consist of statistical inferences derived from primary data, without specifying a particular mechanism of interaction, or involve unmeasured intermediaries. The connectivity in several types of networks has been used to provide context for Alzheimer disease (AD) variants (guilt by association) or to identify molecules and mechanisms associated with AD.
Box 2 | Common types of biological networks

Protein–protein interaction networks

- Protein interaction networks compile experimentally tested or predicted physical binding affinities between proteins
- Tissue-specific protein interactions can be obtained experimentally via yeast two-hybrid or affinity purification–mass spectroscopy approaches, or identified in non-tissue-specific databases, using tissue-specific gene expression signatures
- Disease-associated genetic variants can lead to edgetic changes that alter specific protein–protein physical binding interactions
- The use of databases such as HINT and H2-II-14 could help avoiding the historical publication bias that can be linked to protein interactions generated in small-scale experiments

Coexpression networks

- Coexpression networks represent the links, or gene–gene correlations, between genes that have similar expression patterns which can reflect common regulatory mechanisms
- The biological bases of coexpression links are diverse and include coregulation via chromatin conformation, transcription factors, epigenetics, noncoding RNAs and cell-type variation
- Analysis of samples from multiple tissues can uncover differential gene expression between cell types and brain regions
- Coexpression relationships can be utilized even with small samples sizes, with databases such as COEXPRESdb and GeneMANIA, to find genes coexpressed with gene sets of interest
- Coexpression relationships can be specific to the tissue or the disease state

Causal networks — directed networks that predict signal propagation

- Inference of causal networks that contain directed links typically require hundreds of samples and/or multilevel ‘omics’ data
- Inference of directionality is statistically difficult: accuracy decreases as the number of network nodes increases
- Ideal data sources for these networks include microarray or RNAseq time-series experiments and systematic perturbation data, which are rarely available for brain tissue
- Expression quantitative trait loci (eQTLs) are useful to generate directed networks, especially in inbred populations in which the single nucleotide polymorphism–to–gene coupling is strong
Toolboxes to extract directed networks from gene expression data, or expression and genetics data, are now publicly available\textsuperscript{91}
Box 3 | Cognitive ability and reserve

Baseline cognitive function and decline in cognitive function in non-demented individuals are affected by genetics\textsuperscript{235,236}, several neuropathologies\textsuperscript{97}, and lifestyle factors, such as social\textsuperscript{237} and physical\textsuperscript{238} activity and life-long cognitive activity\textsuperscript{239,240}. Similarly, in the population with Alzheimer disease (AD), the level of cognitive function is influenced by lifestyle factors (including education and occupation\textsuperscript{241,242}, and disease-related decline\textsuperscript{243}). However, classic AD pathology explains less than a third of the variance of cognitive decline\textsuperscript{38}. Furthermore, many individuals presenting with AD neuropathology still maintain adequate cognitive function\textsuperscript{37,99}. The concept of cognitive reserve has been proposed to help explain this difference between predicted and actual cognitive decline\textsuperscript{242,244}. This disconnection between pathological indices and cognitive function is a strong motivation to study the genetic basis of cognitive function and cognitive decline seen in AD.

High cognitive function in midlife can be considered a component of cognitive reserve, as it can delay AD diagnosis\textsuperscript{245,246}. Twin studies indicate that cognitive function has a substantial genetic component\textsuperscript{247}. For example, out of 13 single nucleotide polymorphisms associated with general cognitive function, four (\textit{TOMM40}, \textit{APOE}, \textit{MEF2C} and \textit{ABCG1}) have been consistently implicated in AD\textsuperscript{248}, and others were associated with the rate of cognitive decline (\textit{CR1}\textsuperscript{249}, episodic memory (\textit{PICALM})\textsuperscript{250} and working memory performance tasks (\textit{BIN1})\textsuperscript{251}. The most consistent finding, by far, was the association of \textit{APOE} with various cognitive phenotypes and the rate of cognitive decline with age\textsuperscript{252,253}. However, some of this overlap in the genetic bases of cognitive function and that of AD-associated dementia might be due to the fact that decline in cognitive ability can occur naturally with time or with AD.
Figure 1. Effects of Alzheimer disease genetic variants on molecular networks and global brain topology

A. The phenotypic effects of Alzheimer disease (AD)-associated genetic variants are exerted through several types of molecular networks (here extracted from GeneMANIA), or even alter their structure. For example, some AD-associated SORL1 mutations (blue dots) are predicted to affect specific interaction interfaces (edgetic effects). The effects of genetic variants on molecular networks, in turn, influence processes at higher physiological scales (cellular morphology, cell–cell interactions and tissue morphology) that ultimately affect

Natl Rev Neurol. Author manuscript; available in PMC 2017 January 01.
global brain structure. B. Each type of molecular network is prone to different types of biases: in canonical pathways, most links are centred around APP, PSEN1/2 and APOE, potentially reflecting historical interest in those genes; in coexpression analyses and other high-throughput networks, links between genes carrying AD-associated variants are more evenly distributed. C. The functional and structural connectivity of brain networks seen in patients with AD or with mild cognitive impairment (MCI) shows stereotypical changes, in part attributable to the collective effect of AD-associated variants. The circular networks are conceptual representations of small-world brain networks (green zone), which can gain links (pink links in MCI) or lose links (dashed links in AD and MCI), according to disease status. Generally, AD brains are characterized by less efficient and less globally integrated networks, which are seen as a parallel for cognitive dysfunction.
Figure 2. Components of a multiscale model of the effects of genetic variants

A. The direct and indirect effects of Alzheimer disease (AD)-associated genetic variants on neurons can be assessed experimentally in gene editing-based and/or stem cell-based disease models. The resulting patterns of activity can be fitted to mathematical neuronal models — the building blocks of larger microcircuits and brain region models.

B. Connectivity between and within brain regions, such as different cortical layers (L1, L2/3, L4) and other structures, including components of the thalamus can be extracted from human and other primate brain tissue (top left). This information can be combined with cellular parameters, such as cell...
morphology and connectivity, that can be measured by electron microscopy (EM), and inter-regional parameters that can be measured by diffusion spectrum imaging (DSI), to generate detailed realistic connectivity matrices (top right). Such matrices establish a synaptic connection among dynamic neuron models (a), from which spontaneous or evoked activity can be simulated (spike raster plots at the bottom). The effects of AD variants on network activity can be observed by comparing neuronal network models that do or do not incorporate the effects of AD variants. C. The model activity in various brain regions, with or without AD variants, can be temporally smoothed to provide an output analogous to functional MRI (fMRI) data, which can then be compared with published studies. The actions of various drugs or molecular interventions can also quickly be examined at this level — a level much closer to a cognitive phenotype — which may be helpful in screening potential therapies.