Relating Global and Local Connectome Changes to Dementia and Targeted Gene Expression in Alzheimer's Disease

Samar S. M. Elsheikh[1], Emile R. Chimusa[2], Alzheimer's Disease Neuroimaging Initiative*, Nicola J. Mulder[1,†], and Alessandro Crimi[3,4]

[1]Computational Biology Division, Department of Integrative Biomedical Sciences, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, 7925, South Africa
[2]Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, 7925, South Africa
[3]University Hospital of Zürich, Zürich, 8091, Switzerland
[4]African Institute for Mathematical Sciences, Ghana
[†]nicola.mulder@uct.ac.za
*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

ABSTRACT

Networks are present in many aspects of our lives, and networks in neuroscience have recently gained much attention leading to novel representations of brain connectivity. The integration of neuroimaging and genetics allows a better understanding of the effects of the genetic variations on brain structural and functional connections. The current work uses whole-brain tractography in longitudinal case-control setting and measures the brain structural connectivity changes to study the neurodegeneration of Alzheimer's. This is accomplished by examining the effect of targeted genetic risk factors on the most common local and global brain connectivity measures. Furthermore, we examined the extent to which Clinical Dementia Rating is affected by brain connections longitudinally and genetic variation. Here, we show that the expression of *PLAU* and *HFE* genes increases the change in betweenness centrality related to the fusiform gyrus and clustering coefficient of cingulum bundle over time, respectively. We also show that betweenness centrality has a high contribution to dementia in distinct brain regions. Our findings provide insights into the complex longitudinal interplay between genetics and brain characteristics and highlight the role of Alzheimer's genetic risk factors in the estimation of regional brain connection alterations.

Introduction

There are many factors which may affect susceptibility to Alzheimer's disease (AD) and various ways to measure the disease status. However, there is no single factor which can be used to predict the disease risk sufficiently[1]. Genetics is believed to be the most common risk factor in AD development[2]. Towards studying the etiology of the disease, a number of genetic variants located in about 20 genes have been reported to affect the disease through many cell-type specific biological functions[3]. Those efforts resulted from omic studies such as Genome-Wide Associations Studies (GWAS), which highlighted dozens of multi-scale genetic variations associated with AD risk[4–6].

From the early stages of studying the disease, the well known genetic risk factors of AD were found to lie within the coding genes of proteins involved in amyloid-$\beta(A\beta)$ processing. These include the well-known Apolipoprotein E gene (*ApoE*) that increases the risk of developing AD[7], the Amyloid precursor protein (APP)[8], presenilin-1 (PSEN1) and presenilin-2 (PSEN2)[9,10]. The advancement in technologies and the integration of genetic and neuroimaging datasets have taken Alzheimer's research steps further, and produced detailed descriptions of molecular and brain aspects[11]. Studies have utilised the *connectome*[12] to study different brain diseases through associating genetic variants to brain connectivity[13]. A *structural connectome* is a representation of the brain as a network of distinct brain regions (nodes) and their structural connections (edges), calculated as the number of anatomical tracts. Those anatomical tracts are generally obtained by diffusion-weighted imaging (DWI)[14]. DWI is the most commonly used method for mapping and characterizing the diffusion of water molecules in three-dimensions, as a function of the location in order to construct a contrast in the magnetic resonance images. This representation highlighted a network based organization of the brain with separated subnetworks (*network segregation*) which are connected by few nodes (*network integration*)[15]. Given such a "small-world" representation of the brain, it is also possible to represent each

individual brain as single scalar metrics which summarize peculiar properties of segregation and integration[16]. Alternatively, those global metrics can also be used to quantify local properties of specific nodes/areas.

Early work demonstrated that *ApoE*-4 carriers have an accelerated age-related loss of global brain inter-connectivity in AD subjects[17], and topological alterations of both structural and functional brain networks are present even in healthy subjects carrying the *ApoE* gene[18]. A more recent study has shown the association between *ApoE* expression and brain segregation changes[6]. Going beyond the *ApoE* gene, Jahanshad et al.[19] used a data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to carry out a GWAS of brain connectivity measures and found an associated variant in F-spondin (*SPON1*), previously known to be associated with dementia severity. A meta-analysis study also showed the impact of APOE, phosphatidylinositol binding clathrin assembly protein (PICALM), clusterin (CLU), and bridging integrator 1 (BIN1) gene expression on resting state functional connectivity in AD patients[20]. In this work we utilised the most commonly used metrics of network segregation and integration. The path length, for example, was used in a plethora of studies as biomarker to study the complex brain connectivity in schizophrenia[21], while the local efficiency and characteristic path lengths were used to study the structural organisation of the brain network in autism[22]. For a detailed review on work using such metrics we refer the reader to[23]. For the local area features, we used a measure of segregation, one of integration and one of centrality. These metrics are described in detail in the Material and Methods section. Briefly, *Louvain modularity* is a community (cluster) detection method, which iteratively transforms the network into a set of communities. *Transitivity* also quantifies the segregation of a network, as the normalized counting of the fraction of triangles around an individual node. *Characteristic path length* is the average shortest path, and the *global efficiency* is its inverse. *Betweenness centrality* is fraction of all shortest paths in the network that pass through a given node. Those metrics represent large-scale organization which in turn might be used to represent well functioning cognitive functions[24]. Measures of centrality (e.g. degree and betweeness centrality) can be related each other as they measure similar things, and indeed it has been shown that many of them can be simply led back to the node degree measurement at the point to be considered its "sprangles"[25]. Therefore, along with using them we computed their inter-correlations as well as their correlation with the nodal degree metric which is the simplest feature we can extract from a graph node, to investigate this aspect.

Focusing on AD, we cannot fail to mention that it is a common dementia-related illness; in the elderly, AD represents the most progressive and common form of dementia. Accordingly, incorporating and assessing dementia severity when studying AD provides more insights into the disease progression from a clinical point of view. A reliable global rating of dementia severity is the Clinical Dementia Rating (CDR)[26], which represents a series of evaluations specific for memory, orientation, judgment and problem solving, community affairs, home and hobbies (intellectual interests maintained at home), and personal care assessed at the time of the visit.

Summarizing this paper focuses on the following questions:

- Are there brain connectivity metrics that discriminate changes longitudinally in AD patients compared to healthy control subjects?
- Is there a most representative metric, or a redundancy in the chosen metrics for the case in examination?
- Is there a correlation between the metrics used and known genes expression?
- Is there a correlation between connectivity metrics and clinical ratings?

To address the above questions, we initially study the longitudinal connectivity changes and CDR score, then we regress the changes overtime in the connectivity metrics on gene expression, lastly we perform a ridge regression between the difference in CDR scores and brain connectivity. The relationships are investigated both for global and local brain connectivity metrics, as a different impact could be possible at global level overall or specific to regions.

This paper uses a dataset from ADNI (http://adni.loni.usc.edu/) and presents an integrated association study of specific AD risk genes, dementia scores and structural connectome characteristics. We adapted a longitudinal case-control study design to mainly examine the association of known AD risk gene expression with local and global connectivity metrics. We also aim at testing the longitudinal effect of brain connectivity on different CDR scores, and carry out a multivariate analysis to study the longitudinal effect of gene expression and connectome changes on CDR. Our underlying hypothesis can be summarized in the simplistic representation in Figure 1, where specific genes affect decreases in connectivity comparing baseline and follow-up and this ultimately affects intellectual abilities and CDR scores.

Results

Longitudinal Connectivity Changes and CDR

Initially, we used descriptive statistics plots to visualize the data for the two populations; the AD and matched control subjects. To facilitate the integrated analysis, we looked into the different sets of data individually to have a better understanding of the underlying statistical distribution of each, and chose the best analysis methods accordingly. Firstly, we plotted the global and local connectivity metrics in a way that illustrates the longitudinal change. Those longitudinal changes are measured after 1 year followup from baseline screening. The global connectivity metric box plots show the baseline and follow-up distributions for both AD and controls for transitivity, Louvain modularity, characteristic path length and global

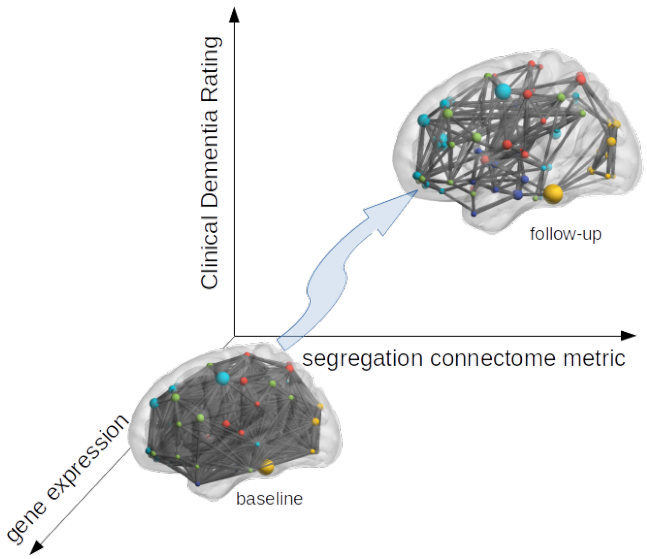

Figure 1. Simplistic representation of our approach which relates connectome metrics of segregation (disconnection), cognitive decline and gene expression.

efficiency (Figure 2). Figure 2 shows that the longitudinal changes in connectome metrics are statistically significant among the AD subjects and not mere artifacts, but not within the control population which has non significant changes. In fact, comparing the two groups (AD and controls) in terms of the change in global connectivity metrics overtime, the only significant differences found between baseline and follow-up were within the AD group. These were found in the characteristic path length (p-value 0.0057), global efficiency (p-value 0.0033), and Louvain modularity (p-value 0.0086). The test used here was the Wilcoxon signed rank test for paired samples. The threshold used here was $\frac{0.05}{4} = 0.0125$, where the division by 4 is due to the multiple hypothesis correction.

Supplementary Figures S1, S2 and S3 show the distribution of the local efficiency, clustering coefficient and betweenness centrality connectivity metrics, respectively as their absolute differences at all atlas brain regions. These figures convey that these local connectivity metrics have different values across the Automated Anatomical Labeling (AAL) regions, and that there is a difference between the baseline and follow-up. A list of the brain atlas region names and abbreviations are available in Supplementary Table S1. Moreover, we show, in Supplementary Figure S4, the scatter and violin plots of the six CDR scores, at the baseline and follow-up, illustrated in the Materials and Methods (and also in Supplementary Figure S4). Both global and local connectivity features show non-symmetric distribution in the baseline, follow-up and absolute change between them. Therefore, we use non-parametric models and statistical tests in the following analysis.

Analyzing the relationship between connectivity metrics and their possible redundancy, as shown in Figure 3, there is a relative correlation between nodal degree and betweenness centrality. However, the correlation between degree and other metrics measuring segregation and integration is weak. Comparing the features to each other by using Spearman correlation, as shown in Figure 4, there is a high redundancy between local efficiency and cluster coefficients, assuming they are computed at the same time point.

Gene Expression

We derived a list of 17 AD risk factor genes from BioMart, and retrieved 56 related probes sets from ADNI data. We performed a Mann-Whitney U test which aims at testing whether a specific probe set's expression is different between AD and controls. For each gene, we chose the probe set that has the lowest p-value. Table 1 reports the selected probe set with the smallest p-value, at each gene. It is worth mentioning that we are using probe sets, rather than single probes. These probe sets represent different parts of the transcripts rather than representing single alternative splicing of the gene. The probe sets for each gene were normalised across probes. Expression values were normalised using the robust multi-chip average method in the Affymetrix U219 array, which consists of a total number of 49,293 probe sets and 530,467 probes, as explained in the Genetic Data Acquisition section.

After estimating the expression of the 17 genes, as explained in the Materials and Methods, we plotted a heatmap of the related gene expression profiles shown in Supplementary Figure S5. Here, some of the genes appear to be highly expressed in the profiles (e.g. *SORL1* and *PSEN1*), while others show very low expression (e.g. *HFE* and *ACE*).

Table 1. Mann-Whitney U test top results for the difference between AD and controls in probe-set expression

Gene	Top results		
	Chromosome	Probe set id	p-value
APBB2	4	11734823_a_at	0.02575
MPO	17	11727442_at	0.38631
APP	21	11762804_x_at	0.01396
ACE	17	11752871_a_at	0.24478
PLAU	10	11717154_a_at	0.01396
PAXIP1	7	11755176_a_at	0.45499
HFE	6	11736346_a_at	0.11881
SORL1	11	11743129_at	0.10912
A2M	12	11715363_a_at	0.28592
NOS3	7	11725467_a_at	0.04261
BLMH	17	11757556_s_at	0.09356
ADAM10	15	11751180_a_at	0.14278
PLD3	19	11715382_x_at	0.17304
ApoE	19	11744068_x_at	0.05962
PSEN1	14	11718678_a_at	0.29453
PSEN2	1	11723674_x_at	0.04862
ABCA7	19	11755091_a_at	0.45499

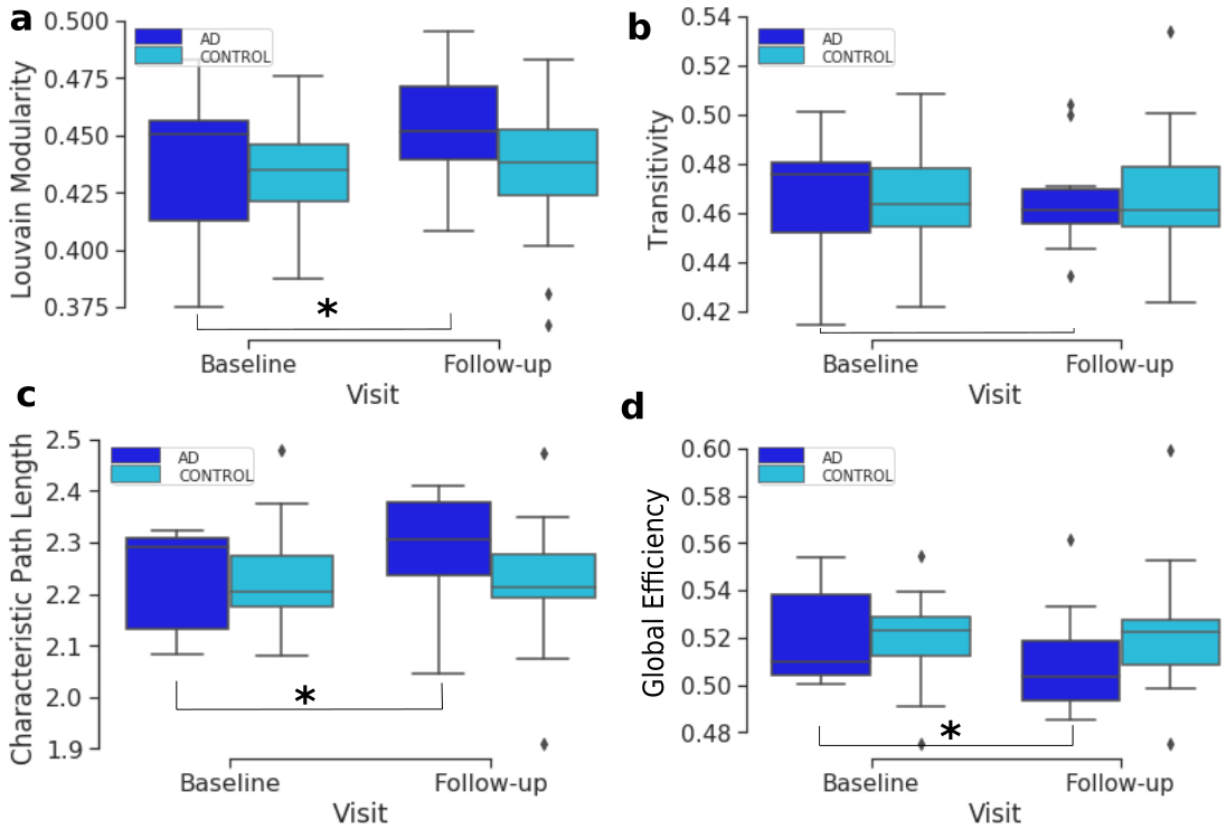

Figure 2. Box plots of the distribution of brain segregation and integration global connectivity metrics comparing the two time points. The plots compare the baseline and follow-up distributions for AD and controls for Louvain modularity (a), transitivity (b), characteristic path length (c) and global efficiency (d). The asterisk denotes that there is a significant change from baseline to the follow-up visit.

Association Analysis

We studied the undirected associations of the 17 gene expression values with the longitudinal change in global and local brain connectivity, as well as the associations with longitudinal CDR and connectivity changes. The total sample size after integrating all the datasets was 47 participants. Firstly, we performed an association analysis of gene expression with the connectivity changes locally, at each AAL brain region. In Table 2 we show the top results reported along with the Spearman correlation co-efficient. The *APP* gene (ρ =-0.58, p-value=1.9e-05) and *BLMH* (ρ =0.57, p-value=2.8e-05) are the top and only significant genes in the list, and associate with the change in local efficiency at the right middle temporal gyrus (Temporal_Mid_R AAL region) and clustering coefficient at the left Heschl gyrus (Heschl_L), respectively. Supplementary Figure S6 shows the scatter plots related to these significant associations.

In Table 2, there is a similar pattern observed in association results between the clustering coefficient and local efficiency, e.g. both metrics are associated with *BLMH* at the left Heschl gyrus (Heschl_L), *APP* at the right middle temporal gyrus (Temporal_Mid_R) and *PLAU* at the right angular gyrus (Angular_R). We interpret this by the strong correlation that exists between the local efficiency and clustering coefficient at the baseline, follow-up and also, the absolute change (see Figure 4). On the other hand, Table 3 reports the top results of the

Table 2. Top results of Spearman associations between AD gene expression and local connectivity metrics.

Gene	**Sorted by P-value. Dashed line: threshold**$=\frac{0.05}{17\times 90}=3.27e-05$				
	Region	Region id	Metric	ρ	P-value
APP	Temporal_Mid_R	Region86	local_eff	-0.5805	1.9e-05
BLMH	Heschl_L	Region79	cluster_coef	0.5708	2.8e-05
PSEN1	Occipital_Mid_R	Region52	b_centrality	-0.5598	4.3e-05
BLMH	Heschl_L	Region79	local_eff	0.5591	4.4e-05
APP	Temporal_Mid_R	Region86	cluster_coef	-0.5197	0.000182
PAXIP1	Amygdala_L	Region41	cluster_coef	0.5064	0.000281
PLAU	Angular_R	Region66	cluster_coef	0.484	0.000567
PLAU	Angular_R	Region66	local_eff	0.4838	0.00057
ACE	Postcentral_L	Region57	b_centrality	0.4648	0.000998
ADAM10	Postcentral_L	Region57	local_eff	-0.4602	0.001133
PAXIP1	Parietal_Sup_L	Region59	b_centrality	-0.4585	0.00119
PLAU	Fusiform_L	Region55	b_centrality	0.4564	0.001262
SORL1	Putamen_R	Region74	local_eff	-0.4528	0.001395
PSEN2	Frontal_Inf_Oper_R	Region12	local_eff	0.4457	0.001693
PLAU	Frontal_Inf_Oper_L	Region11	b_centrality	0.4454	0.001704
ABCA7	Temporal_Inf_L	Region89	local_eff	0.442	0.001866

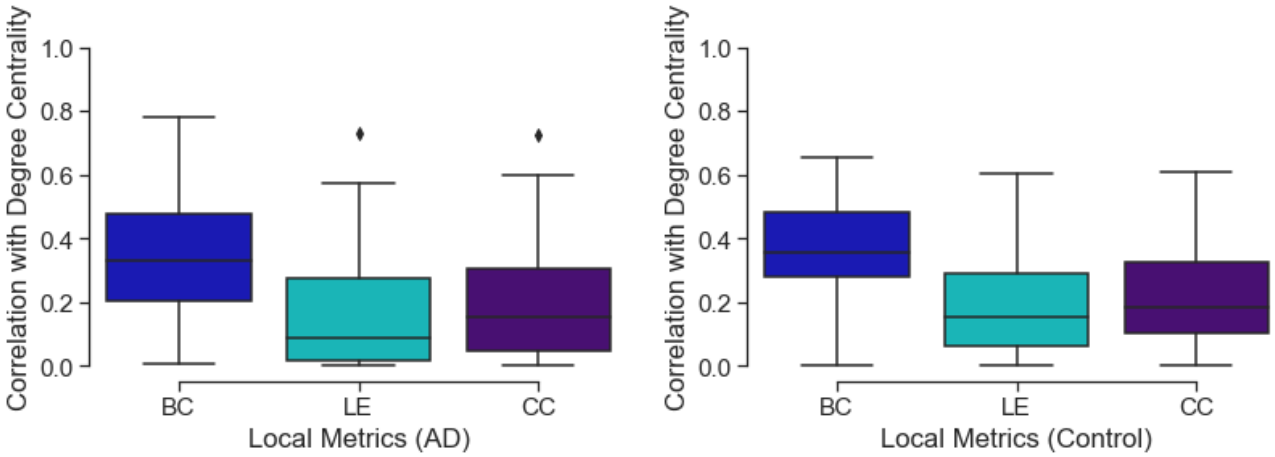

Figure 3. The figure shows three box plots respectively (a) for the AD and (b) control subjects. Each boxplot represents the correlation values between the degree value of a node and the related local efficiency (LE), betweenness centrality (BC) and clustering coefficient (CC). Values close to 0 represent low correlation while those close to 1 show a high correlation.

association between gene expression and the change in brain global connectivity. In this case, all observed associations were not statistically significant after correcting for multiple hypotheses (threshold is stated in Table 3).

Table 3. Top 20 Spearman association results of the change in global network metrics with targeted AD gene expressions with threshold $= \frac{0.5}{17} = 0.0029$

Gene	Results are sorted according to p-value		
	ρ	P-value	Global Feature
PAXIP1	0.3889	0.0069	transitivity
PLAU	-0.3824	0.008	global_eff
ACE	-0.3696	0.0106	transitivity
PLAU	-0.3523	0.0151	char_path_len
ABCA7	-0.3492	0.0161	char_path_len
PSEN1	-0.299	0.0412	transitivity
APP	-0.2602	0.0774	char_path_len
PLAU	-0.2542	0.0847	Louvain
ApoE	-0.2506	0.0893	char_path_len
ADAM10	-0.2365	0.1095	Louvain
ACE	0.2291	0.1213	Louvain
NOS3	-0.2207	0.1359	char_path_len
NOS3	-0.2202	0.1369	global_eff
ABCA7	-0.2164	0.1441	global_eff
HFE	-0.2012	0.1752	char_path_len

Regressing Change in Local and Global Brain Connectivity on Gene Expression

We analyzed the directed association through regressing the change in local connectivity (as a dependant variable), at each AAL region, on gene expression using (as an independent variable or predictor) a quantile regression model. Table 4 reports the top results, along with the regression coefficient, p-values and t-test statistic. *PLAU* was the most significant gene affecting the absolute change in betweenness centrality at left Fusiform gyrus (Fusiform_L) with an increase of 487.13 at each unit increase in *PLAU* expression (p-value$= 3e-06$). This was followed by the expression of *HFE* with an effect size of 0.1277 on the change in local efficiency at the right anterior cingulate and paracingulate gyri (Cingulum_Ant_R). Those observed associations are illustrated in Figure 5. Moreover, we report the protein-protein interaction[27] of the aforementioned genes in Figure 6. Supplementary Figures S7, S8 and S9 show the Manhattan plots for the -log10 of the p-values corresponding to the quantile regression models of the change in local efficiency, clustering coefficient and betweenness centrality, respectively.

Similarly, we regressed the absolute change of global connectivity measures on gene expression values and the top results are shown in Table 5. All the results have p-values less than the threshold we set ($\frac{0.05}{17} = 0.0029$).

Additive Genetic Effect on Brain Regions

To visualize the overall contribution of AD gene risk factors used in this work on distinct brain areas, we added up the -log10 p-values for the gene expression coefficients at each of the 90 AAL regions. The p-values were obtained from the quantile regression analysis between the gene expression values and each of the three connectivity metrics - those are the absolute difference between baseline and follow-up of local efficiency, clustering coefficient and betweenness. Figure 7 summarizes this by 1) representing the brain connectome without edges for each of the connectivity metrics, 2) each node represents a distinct AAL region and is annotated with the name of the region, 3) the size of each node is the sum -log10 of the regression coefficient associated p-vales for all the genes. It is clear from Figure 7 that local efficiency and clustering coefficient show more similar patterns of association with genes compared to betweenness centrality. This

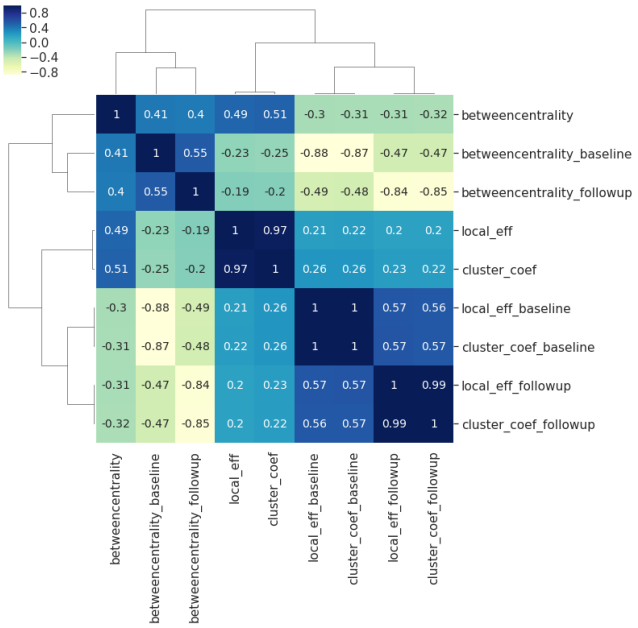

Figure 4. Spearman correlations between the three local connectivity metrics; local efficiency, clustering coefficient and betweenness centrality, at baseline (suffix: _baseline), follow-up (suffix:_followup) and the absolute difference between them (no suffix). The calculation of Spearman's coefficient combines all 90 brain regions, and both AD and controls. The plot illustrates the very strong relationship between the clustering coefficient and local efficiency at baseline, follow-up and the absolute difference between the two visits.

means that the gene expression has stronger association with local structure of the brain when using clustering measures (e.g. clustering coefficients), while the pattern of association tends to be weaker when using measures expressing the state of a region being between two others (e.g. betweenness centrality).

The colors are assigned automatically by the BrainNet Viewer. Overall, although the gene contributions to the absolute change in local efficiency have a similar pattern to that of clustering coefficient, the contribution to betweenness centrality change is relatively small.

Regressing the difference in CDR on the difference in Global and Local Connectivity

To asses the directed and undirected association of the longitudinal measures of global connectivity and CDR scores, we calculated the difference between baseline and follow-up visits for both CDR and global connectivity metrics , i.e. $CDR_{baseline} - CDR_{follow-up}$ and $metric_{baseline} - metric_{follow-up}$, respectively. The Spearman and quantile regression results are both shown in Table 6. We observed a correlation between the increase of the transitivity score (global brain segregation) and the CDR memory score over time ($\beta = -6.14e-06$, p-value$= 0.0034$). On the other hand, there is a positive correlation between global efficiency (global brain integration) and the CDR "home and hobbies" score.

Similarly, in Supplementary Table S2 we looked at the monotonic effect of local connectivity metrics on the seven CDR scores, both represented as the subtraction of the follow-up visit from the baseline visit. The increase in betweenness centrality was shown to have different effects on the CDR score over the one-year time period. For example, as the betweenness centrality decreases over time, the judgement and problem solving increases in severity by 1.06e-08 over time (p-value=1.32e-17), in the frontal lobe (Frontal_Inf_Oper_L).

Table 4. Top 50 quantile regression results of the change in local network metrics (y) and targeted AD gene expression (x)

Region	**Sorted by p-value. Dashed line: threshold**$= \frac{0.05}{17\times 90} = 3.27e-05$					
	Region	R. id	Beta	Statistic	P-value	Metric
PLAU	Fusiform_L	55	487.1319	5.3836	3e-06	b_centrality
HFE	Cingulum_Ant_R	32	0.1277	4.8139	1.7e-05	local_eff
PAXIP1	Parietal_Sup_L	59	-147.3175	-4.5608	3.9e-05	b_centrality
HFE	Cingulum_Ant_R	32	0.1662	3.9835	0.000246	cluster_coef
APP	Amygdala_R	42	-0.1349	-3.8548	0.000365	local_eff
PLAU	Hippocampus_L	37	0.1073	3.4801	0.001125	local_eff
ADAM10	Postcentral_L	r57	-0.0871	-3.4376	0.001275	cluster_coef
ApoE	Frontal_Inf_Orb_L	15	153.3117	3.3627	0.001584	b_centrality
APBB2	Amygdala_L	41	0.2054	3.3517	0.001635	cluster_coef
ApoE	Frontal_Sup_Medial_L	23	0.1912	3.2788	0.002015	cluster_coef
MPO	Cingulum_Mid_L	33	0.0293	3.2465	0.00221	cluster_coef
MPO	Cingulum_Mid_L	33	0.0281	3.2143	0.00242	local_eff
PLAU	Cingulum_Ant_R	32	0.1428	3.1969	0.002543	local_eff
ADAM10	Postcentral_L	57	-0.0517	-3.1541	0.002867	local_eff
ApoE	Postcentral_L	57	0.0806	3.0931	0.003398	local_eff
PLD3	Olfactory_R	22	-0.1268	-3.0463	0.003867	cluster_coef
ABCA7	Frontal_Inf_Orb_R	16	34.9538	2.9489	0.005043	b_centrality
A2M	Putamen_R	74	-0.0543	-2.9171	0.005495	local_eff
PLAU	Hippocampus_L	37	0.1472	2.9023	0.005717	cluster_coef
HFE	Frontal_Inf_Tri_R	14	0.1852	2.8813	0.006047	cluster_coef
HFE	Frontal_Inf_Tri_R	14	0.0926	2.8594	0.006411	local_eff
APP	Amygdala_R	42	-0.1753	-2.8288	0.006953	cluster_coef
ApoE	Occipital_Mid_R	52	44.0624	2.7995	0.007512	b_centrality
HFE	Calcarine_R	44	0.1403	2.7916	0.007669	cluster_coef
APP	Temporal_Mid_R	86	-0.0692	-2.7396	0.008787	cluster_coef
APP	Temporal_Mid_R	86	-0.0386	-2.7297	0.009016	local_eff
PLD3	Olfactory_R	22	-0.0987	-2.713	0.009413	local_eff
A2M	Olfactory_R	22	-30.9342	-2.6919	0.009941	b_centrality
APP	Cuneus_R	46	0.1443	2.6845	0.010131	cluster_coef
PSEN1	Frontal_Inf_Tri_L	13	-0.1344	-2.6492	0.01109	local_eff
PSEN2	Temporal_Mid_L	85	0.0528	2.6465	0.011168	local_eff
PSEN1	Frontal_Inf_Tri_L	13	-0.2431	-2.6432	0.011262	cluster_coef
PLAU	Frontal_Mid_R	8	0.0854	2.6384	0.011401	local_eff
ApoE	Putamen_L	73	372.8291	2.638	0.011411	b_centrality
A2M	Occipital_Mid_R	52	0.0535	2.6213	0.011907	cluster_coef
APP	Cuneus_R	46	0.0798	2.6189	0.011979	local_eff
ADAM10	Temporal_Sup_L	81	-0.1123	-2.6134	0.012147	cluster_coef
ApoE	SupraMarginal_L	63	0.1158	2.5663	0.013677	local_eff
HFE	Calcarine_R	44	0.0755	2.533	0.014866	local_eff
PLAU	Occipital_Mid_L	51	0.1268	2.5193	0.01538	cluster_coef
MPO	Pallidum_R	76	0.024	2.5101	0.015732	local_eff
ABCA7	Temporal_Inf_L	89	0.0249	2.5083	0.015804	local_eff
A2M	Occipital_Mid_R	52	0.0268	2.5023	0.01604	local_eff
PLAU	Hippocampus_R	38	182.0756	2.5007	0.016105	b_centrality
APP	Frontal_Med_Orb_L	25	0.1167	2.4968	0.01626	cluster_coef
HFE	Cingulum_Post_R	36	28.3207	2.4708	0.017334	b_centrality
ACE	Occipital_Mid_R	52	0.0433	2.4632	0.017659	local_eff
NOS3	Olfactory_L	21	-0.1615	-2.4596	0.017815	cluster_coef
ABCA7	Temporal_Inf_L	89	0.0429	2.4374	0.018808	cluster_coef
SORL1	Paracentral_Lobule_L	69	91.7443	2.4364	0.018855	b_centrality

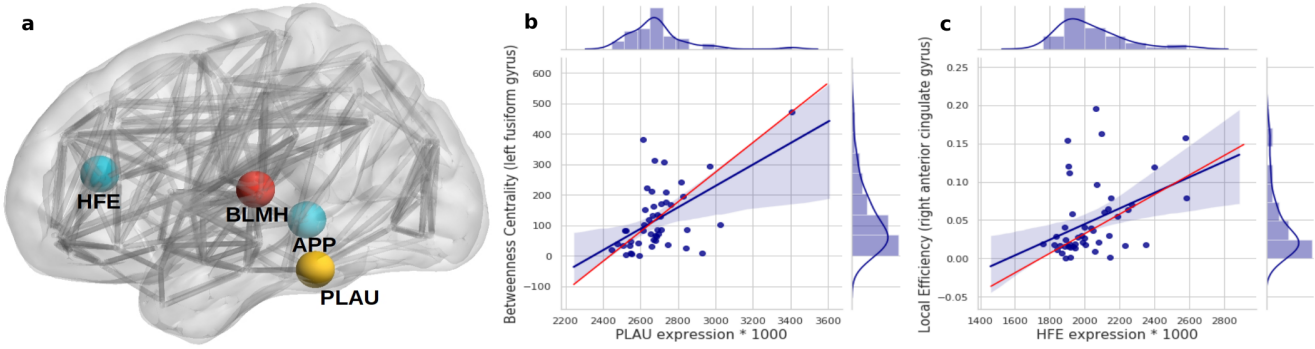

Figure 5. Subfigure (a) higlights regions in the brain where significant associations - between gene expression and longitudinal change in local connectivity metrics were found, using quantile regression (*HFE* and *PLAU*) and Spearman associations (*APP* and *BLMH*). Each gene is plotted at the AAL brain region where the association was significant; *APP* at Temporal_Mid_R, *BLMH* at Heschl_L, *PLAU* at Fusiform_L and *HFE* at Cingulum_Ant_R. (b) and (c) are scatter plots to visualize the association between *PLAU* gene expression and betweenness centrality in the left fusiform gyrus (a), and between the expression of *HFE* gene with local efficiency in right anterior cingulate gyrus (b). The red line on the plots represents the median (quantile) regression line, while the blue line represents the ordinary least square line.

Table 5. Top quantile regression results of the change in global network metrics and targeted AD gene expressions. Threshold= $\frac{0.5}{17} = 0.0029$.

Gene	Results are sorted according to p-value			
	Beta	Statistic	P-value	Metric
PAXIP1	0.0155	2.1179	0.039738	transitivity
PSEN1	-0.0366	-2.0821	0.04305	transitivity
A2M	0.0178	1.9728	0.054679	Louvain
PLAU	-0.0157	-1.9288	0.06008	global_eff
APBB2	0.0166	1.7579	0.085573	Louvain
ABCA7	0.0077	1.5185	0.135881	transitivity
BLMH	0.0101	1.3476	0.184529	transitivity
ACE	-0.0162	-1.2618	0.213524	transitivity
ADAM10	-0.0147	-1.2538	0.21638	Louvain
ApoE	-0.0555	-1.1673	0.249246	char_path_len
PLD3	-0.0096	-1.1433	0.258978	transitivity
ABCA7	-0.0156	-1.1367	0.261684	char_path_len
SORL1	0.0147	1.1278	0.265372	transitivity
ApoE	-0.0099	-1.1048	0.275101	global_eff
HFE	-0.0161	-1.0436	0.302239	Louvain

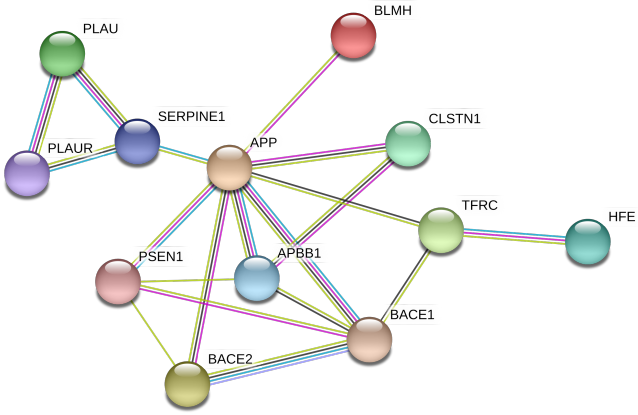

Figure 6. Protein–protein interactions between the genes with significant correlation to brain connectivity metrics. It is evident that the genes interact either directly with APP or via one intermediate node/gene. In this sub-network extracted from STRING[27], the different color lines represents different types of interaction: Cyan edges are interactions from curated databases, purple are experimentally determined, yellow are from text mining, and black are from co-expression data.

Multivariate Analysis: Ridge Regression

We regressed the difference in CDR visits (response variable; Y), one score at a time, on both the difference in global brain connectivity (predictor; X1), one connectivity metric at a time and all gene expression values (predictor; X2), using the ridge regression model. Supplementary Table S3 reports the mean squared error (the score column) and shows the top hits in the multiple ridge regression. It shows that the α (alpha column) could not converge, using the cross-validation, when the response variables were judgment or personal care. However, the CDR score results show that genes and connectivity metrics have a small effect (β) on the response variables (the change in CDR scores over time), and the larger effects were observed when using the total CDR score (CDR_diff) as a response variable. The connectivity metrics and expression of genes have both negative and positive effects on CDR change. The expression of *ApoE*, for example, has a negative effect (β) of -0.24 on the change in memory score, i.e. the memory rating decreases by 0.24 as the *ApoE* expression increases. While when the expression of *ApoE* increases one unit, the home and hobbies score increases, over time, by 0.12.

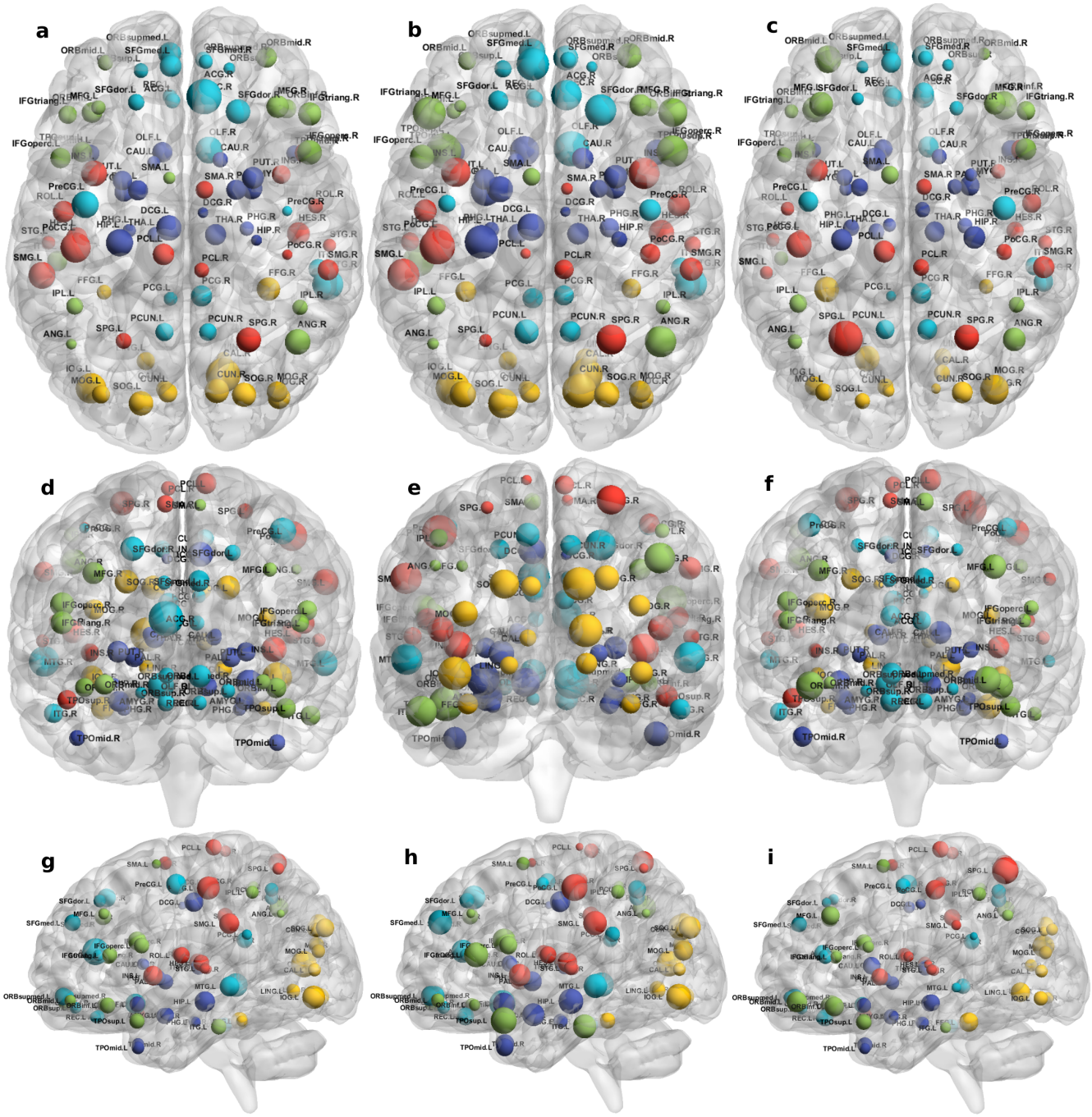

Figure 7. Connectome representations showing the metric additive genetic effect at each AAL node. The subfigures show the axial (top; (a), (b) and (c)), coronal (middle; (d), (e) and (f)), and sagittal (bottom; (g), (h) and (i)) planes of the brain, the node size represents the local efficiency (left; (a), (d) and (g)), clustering coefficient (middle; (b), (e) and (h)) and betweenness centrality (right; (c), (f) and (i)). Colors of the nodes are automatically assigned by the BrainNet Viewer. The acronyms of the brain regions are explained in Supplementary Table S1.

Discussion

In this paper we aim to analyse the longitudinal change of local and global brain connectivity metrics for AD and controls, using a dataset from ADNI. We evaluate the associations between the most commonly used connectivity metrics and the expression of known AD risk genes. Finally, we study the multivariate association between connectivity metrics, gene expression and dementia clinical ratings.

Table 6. Quantile regression results of the difference in CDR (y) with the difference in global connectivity (x)

Global Metric	Threshold= $\frac{0.05}{6} = 0.00833$. * represents significant p-value.					
	CDMEMORY	CDORIENT	CDJUDGE	CDCOMMUN	CDHOME	CDCARE
Q. Regression:	β **(p-value)**					
transitivity	-6.14e-06 (0.0034*)	-1.8e-06 (nan)	-3.2e-07 (nan)	8.4e-07 (0.9249)	4.13e-06 (0.7324)	-3.1e-07 (nan)
global_eff	1.3e-06 (0.8572)	3.36e-06 (0.9944)	3.5e-07 (0.9826)	3.09e-06 (0.3131)	9.5e-06 (0.1613)	-2.5e-07 (nan)
Louvain	-2.64e-06 (0.1683)	-6.84e-06 (0.0352)	1.21e-06 (0.0012*)	-1.01e-06 (0.8125)	-2.16e-06 (0.9424)	-6.3e-07 (0.1787)
char_path_len	-5.8e-07 (0.8562)	-8.3e-07 (0.8791)	-8e-08 (0.8637)	-8.3e-07 (0.0361)	-2.14e-06 (0.0442)	-2e-08 (nan)
Spearman:	ρ **(p-value)**					
transitivity	-0.3395 (0.021)	-0.0763 (0.6142)	-0.0618 (0.6835)	0.0661 (0.6623)	0.161 (0.2851)	-0.0081 (0.9574)
global_eff	0.0483 (0.7497)	0.0056 (0.9707)	0.0505 (0.7388)	0.2685 (0.0712)	0.4145 (0.0042*)	-0.0263 (0.8625)
Louvain	-0.1955 (0.1928)	-0.3119 (0.0349)	0.2077 (0.166)	-0.0968 (0.522)	-0.11 (0.4666)	-0.0628 (0.6784)
char_path_len	-0.1183 (0.4337)	-0.0516 (0.7333)	-0.0281 (0.8531)	-0.2909 (0.0498)	-0.3811 (0.009)	-0.0065 (0.9658)

Our results show that Alzheimer's risk genes can manipulate the amount of change observed in the structural connectome, measured as the absolute difference of longitudinal connectivity metrics. Here, we show that longitudinal regional connectivity metrics, global brain segregation and integration have effects on the CDR scores. More specifically, we observe a consistent decrease, over time, in the local efficiency (a network connectivity metric defined as the length of the shortest path between a node j and h, that contains only the neighbors of a node i[16]) in response to the increase in *APP* expression, at the right middle temporal gyrus (Temporal_Mid_R; see Table 2). The same connectivity metric increases over time as the expression of *HFE* increases, at the right anterior cingulate and paracingulate gyri (see Table 4). Furthermore, as the disease progresses, we observe a correlation between brain segregation and cognitive decline, the latter is measured as CDR memory scores. Moreover, we noticed a correlation between the increase in the global efficiency and the increase in the home and hobbies CDR scores (see Table 6).

Prescott et al.[28] have investigated the differences in the structural connectome in three clinical stages of AD, using a cross-sectional study design, and targeted regional brain areas that are known to have increased amyloid plaque. Their work suggested that connectome damage might occur at an earlier preclinical stage towards developing AD. Here, we further adapted a longitudinal study design and incorporated known AD risk genes. We showed how the damage in the connectome is affected by gene expression, and that the change in connectome affects dementia, globally and locally at distinct brain regions. Aside from our previous work[6], which examined the *ApoE* associations with longitudinal global connectivity in AD using longitudinal global connectivity metrics, this study, to our knowledge, is the first of its type to include gene expression data with global and local brain connectivity. However, similar work has been done in schizophrenia structural brain connectivity, where longitudinal magnetic resonance imaging features, derived from the DWI, were associated with higher genetic risk for schizophrenia[29].

The results obtained here align with findings in the literature of genetics and neuroimaging. Specifically, Robson et al.[30] studied the interaction of the *C282Y* allele *HFE* - the common basis of hemochromatosis - and found that carriers of *ApoE*-4, the C2 variant in TF and C282Y are at higher risk of developing AD. Moreover, the *HFE* gene is known for regulating iron absorption, which results in recessive genetic disorders, such as hereditary haemochromatosis[31]. According to Pujol et al.[32], the association between the harm avoidance trait and right anterior cingulate gyrus volume was statistically significant. In their study, they examined the association between the morphology of cingulate gyrus and personality in 100 healthy participants. Personality was assessed using the Temperament and Character Inventory questionnaire. Higher levels of harm avoidance were shown to increase the risk of developing AD[33]. We show here that *HFE* expression affects the local efficiency at the right anterior cingulate gyrus (see

Table 4 and Figure 5). This might indicate a possible effect of *HFE* expression on the personality of AD patients or those at risk of developing the disease.

Moreover, in this study we found that the Plasminogen activator, urokinase (*PLAU*) expression affects the betweenness centrality (a measure of the region's (or node) contribution to the flow of information in a network[16]) in the left fusiform gyrus, over time (see Table 4 and Figure 5). Although the functionality of this region is not fully understood, its relationship with cognition and semantic memory was previously reported[34]. *PLAU*, on the other hand, was shown to be a risk factor in the development of late-onset AD as a result of its involvement in the conversion of plasminogen to plasmin - a contributor to the processing of APP by the urokinase-type plasminogen activator (uPA)[35].

When examining the linear associations between gene expression and local connectivity (see Table 2 and Supplementary Figure S6), we found that the right middle temporal gyrus, known for its involvement in cognitive processes including comprehension of language, negatively associates with *APP* expression. Additionally, the left Heschl gyrus positively correlates with bleomycin hydrolase (BLMH) expression. In the human brain, the BLMH protein is found in the neocortical neurons and senile plaques[36], microscopic decaying nerve terminals around the amyloid occurring in the brain of AD patients. Some studies[37,38] have found that a variant in the *BLMH* gene, which leads to the Ile443→Val in the BLMH protein, increases the risk of AD; this was strongly marked in *ApoE*-4 carriers. The BLMH protein can process the $A\beta$ protein precursor and is involved in the production of $A\beta$ peptide[39].

The spatial location of expression highlighted in regions in Figure 5 are relatively in line with previous studies (e.g.[40]). Indeed, the regional expression of the *APP* gene has been shown to be positively correlated with the severity of regional amyloid deposition observed in PET studies. In[40] the temporal medial region and the fusiform gyrus of the brain are among the most affected by the amyloid deposition and also have high levels of *APP* expression. Recent findings from tau-sensitive positron emission tomography data also confirm the spatial correspondence between accumulation of tau pathology and neurodegeneration in AD patients in the same regions, though only correlations to the *ApoE* genes were investigated[41]. The BLMH protein alters the processing of APP and significantly increases the release of its proteolytic fragment. It has been previously reported to be expressed and have an impact on the hippocampal tissues, but not investigated in other brain regions[42]. To our knowledge, apart from the general expression in the brain parenchyma reported in the Allen brain atlas[43], no study has shown spatial expression of the *PLAU* and *HFE* genes among AD patients. Nevertheless, we can hypothesize that the significance of expression in specific nodes of the brain connectome is related to their interaction with the *APP* gene which is particularly expressed in the nodes highlighted in Figure 5, and this is supported by the protein-protein interaction shown in Figure 6.

Even though none of the AD risk genes showed a significant effect on the longitudinal change in global connectivity (see Tables 3 and 5), the genes showed significant effects on local connectivity changes at regional brain areas (see Table 4 and Table 2). The global connectivity metrics of the brain, on the other hand, have shown promising results in affecting the change observed in CDR scores, including memory, judgement and problem solving, as well as home and hobbies, as shown in Table 6. Previous work studied the association between genome-wide variants and global connectivity of Alzheimer's brains, represented as brain integration and segregation, and found that some genes affect the amount of change observed in global connectivity[6]. This suggests that a generalisation of the current study at a gene-wide level might warrant further analysis.

Considering the possible redundancy of brain connectivity metrics, we looked for correlations with the nodal degrees and other features and observed only a relative correlation to betweenness centrality. Comparing all features to each other, we found only a correlation between local efficiency and cluster coefficient metrics. In summary, we hypothesize that the degree is a simpler representation than betweennes centrality, and it could be used as a substitute, but it is not showing redundancy with the other metrics, and therefore either cluster coefficient or efficiency should also be investigated in similar studies. More generally, when correcting our threshold, we did not consider the number of metrics, because of the high co-linearity. Therefore, reporting all, or only one, did not affect the results.

Our work provides new insights, though replication on a larger sample size is required. Indeed, one limitation here was the small sample size available. We needed to narrow down our selection of participants to those with both baseline and follow-up visits, and that have CDR scores, genetic and imaging information available. Another limitation is given by the use of only two time points, the baseline and the first follow-up visit. This does not allow capturing the effects of connectivity changes over a longer-term or studying the survival probabilities in AD. Extending to more time points would have been useful, but it would have further reduced the dataset. We foresee future work in using a more complex unified multi-scale model to facilitate studying the multivariate effect of clinical and genetic factors on the connectome, besides considering the complex interplay of genetic factors. Nevertheless, previous studies with similar sample sizes have been able to provide relevant insights for the gene and brain interaction networks[44].

In this work, we conducted an association analysis of targeted gene expression with various longitudinal brain connectivity features in AD. Aiming at estimating the neurodegeneration of the connectome, we obtained local and global connectivity metrics at two visits, baseline and follow-up, after 12 months. We calculated the change between the two visits and carried out an association analysis, using quantile

and ridge regression models to study the relationship between gene expression and disease progression globally and regionally at distinct areas of the brain. We tested the effect of the change in connectivity on the longitudinal CDR scores through quantile regression. Furthermore, using a ridge regression model, we controlled for the genetic effects in the previous settings to study the effect of connectivity changes on the CDR change.

The present analysis was conducted in AD using a longitudinal study design and highlighted the role of *PLAU*, *HFE*, *APP* and *BLMH* in affecting how the pattern information is propagated in particular regions of the brain, which might have a direct effect on a person's recognition and cognitive abilities. Furthermore, the results illustrated the effect of brain structural connections on memory and cognitive process of reaching a decision or drawing conclusions. The findings presented here might have implications for better understanding and diagnosis of the cognitive deficits in AD and dementia *in vivo* estimation of a regional disease progression pattern is of high interest for the neuropathology community working on Alzheimers. Indeed the Braak staging hypothesis is still controversial, and other *in-vivo* studies have shown regional effects from positron emission tomography (PET) imaging (temporal lobe, the anterior cingulate gyrus, and the parietal operculum)[40], and impact on the default mode network connectivity[45]. Therefore further investigations of regional patterns is relevant. Very recent results on treatments in mice showed that drug based modulated neuronal activity can reduce amyloid plaques in specific locations and circuitry[46]. In view of future treatments based on specific spatial location and genetic influences, our study provides some initial insights into connectivity outcomes, or at least enhances our understanding of the regions/circuits that show amyloid aggregation or neurodegeneration.

Materials and Methods

Data Description

We used two sets of data from ADNI, which are available at adni.loni.usc.edu. The experiments have been conducted on the publicly available datasets described below, for which ethical approval has already been granted, and data acquisition has been conducted according to the Helsinki II regulations. To use this data for our analysis, we also received an ethical approval from the Faculty of Health Sciences Human Research Ethics Committee.

To fulfil our objectives, unless otherwise specified, we merged neuroimaging, genetic and CDR datasets for all the participants with those three types of data at two-time points available. We considered follow-up imaging and CDR acquisition one year later than the baseline visit. Given those constraints, we ended up with a total of 47 participants. We adopted a case-control study design; 11 of the participants are AD patients, while 36 are controls. The data were matched by age, and the distribution of age in AD ranges between 76.5±7.4 for cases, and 77.0±5.1 years in controls.

Imaging Data

For the imaging, we obtained the DWI volumes at two time points, the baseline and follow-up visits, with one year in between. Along with the DWI, we used the T1-weighted images which were acquired using a GE Signa scanner 3T (General Electric, Milwaukee, WI, USA). The T1-weighted scans were obtained with voxel size $= 1.2 \times 1.0 \times 1.0 mm^3 TR = 6.984 ms$; TE = 2.848 ms; flip angle= 11°), while DWI was obtained with voxel size $= 1.4 \times 1.4 \times 2.7 mm^3$, scan time = 9 min, and 46 volumes (5 T2-weighted images with no diffusion sensitization b0 and 41 diffusion-weighted images b= $1000 s/mm^2$).

Pre-processing of Imaging Data

Each DWI and T1 volume have been pre-processed performing Eddy current correction and skull stripping. Given the fact that DWI and T1 volumes were already co-registered, the AAL atlas[47], and the T1 reference volume are linearly registered according to 12 degrees of freedom.

Moreover, we used the same T1 reference to get the information needed to compute the partial volume effect from the tissue segmentation by using FSL.

Genetic Data Acquisition

We used the Affymetrix Human Genome U219 Array profiled expression dataset from ADNI. The RNA was obtained from blood samples and normalised before hybridization to the array plates. Partek Genomic Suite 6.6 and Affymetrix Expression Console were used to check the quality of expression and hybridization[48]. The expression values were normalised using the Robust Multi-chip Average[49], after which the probe sets were mapped according to the human genome (hg19). Further quality control steps were performed by checking the gender using specific gene expression, and predicting the single nucleotide plymorphisms from the expression data[50,51]

In this work, we targeted specific genes which have been reported to affect the susceptibility of AD. We used the BioMart software from Ensembl to choose those genes by specifying the phenotype as AD[52]. We obtained a total of 17 unique gene names and retrieved a total of 56 probe sets from the genetic dataset.

Clinical Dementia Rating

The Clinical Dementia Rating, or CDR score is an ordinal scale used to rate the condition of dementia symptoms. It range from 0 to 3, and is defined by four values: 0, 0.5, 1, 2 and 3, ordered by severity, which stand for none, very mild, mild and severe, respectively. The scores evaluate the cognitive state and functionality of participants. Here, we used the main six scores of CDR; memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care. Besides the latter, we used a global score, calculated as the sum of the six scores. We obtained the CDR scores at two time points in accordance with the connectivity metrics time points.

Connectome Construction

We generated tractographies by processing DWI data with Dipy[53], a Python library. More specifically, we used the constant solid angle model[54], and Euler Delta Crossings[53] deterministic algorithm. We stemmed from 2,000,000 seed-points, and as stopping condition we used the anatomically-constrained tractography approach based on partial volume effect[55]. We also discarded all tracts with sharp angle (larger than 75°) or those with length $< 30mm$.

To construct the connectome, we assigned a binary representation in the form of a matrix whenever more than three connections were present between two regions of the AAL, for any pair of regions. Tracts shorter than 30 mm were discarded. The FA threshold was chosen in a such a way that allows reasonable values of characteristic path length for the given atlas. Though the AAL atlas has been criticized for functional connectivity studies[56], it has been useful in providing insights in neuroscience and physiology, and is believed to be sufficient for our case study[56].

Global and Local Connectivity Metrics

To quantify the overall efficiency and integrity of the brain, we extracted global measures of connectivity from the connectome, represented here in four values of network integration and segregation. Specifically, we used two network integration metrics 1) the global efficiency (E; Equation 1), and 2) the weighted characteristic path length (L; Equation 2). Both are used to measure the efficiency of which information is circulated in a network. On the other hand, we used; 1) Louvain modularity (Q; Equation 3), and 2) transitivity (T; Equation 4) to measure the segregation of the brain, that is, the capability of the network to shape sub-communities which are loosely connected to one another while forming a densely connected sub-network within communities[15,16].

Suppose that n is the number of nodes in the network, N is the set of all nodes, the link (i, j) connects node i with node j and a_{ij} define the connection status between node i and j, such that $a_{ij} = 1$ if the link (i, j) exist, and $a_{ij} = 0$ otherwise. We define the global connectivity metrics as;

$$E = \frac{1}{n(n-1)} \sum_{i \in N} \sum_{j \in N, j \neq i} d_{ij}^{-1}, \tag{1}$$

where, $d_{ij} = \sum_{a_{uv} \in g_{i \leftrightarrow j}} a_{uv}$, is the shortest path length between node i and j, and $g_{i \leftrightarrow j}$ is the geodesic between i and j.

$$L = \frac{1}{n(n-1)} \sum_{i \in N} \sum_{j \in N, j \neq i} d_{ij}. \tag{2}$$

$$Q = \frac{1}{l} \sum_{ij \in N} \left[a_{ij} - \frac{k_i k_j}{l} \right] \delta(c_i, c_j), \tag{3}$$

where $l = \sum_{i,j \in N} a_{ij}$, m_i and m_j are the modules containing node i and j, respectively, and $\delta(c_i, c_j) = 1$ if $c_i = c_j$ and 0 otherwise.

$$T = \frac{\sum_{i \in N} 2t_i}{\sum_{i \in N} k_i(k_i - 1)}, \tag{4}$$

where $t_i = \frac{1}{2} \sum_{j,h \in N} (a_{ij} a_{ih} a_{jh})$ is the number of triangles around node i.

Using the AAL atlas, we constructed the following local brain network metrics at each region or node. We used the local efficiency ($E_{loc,i}$; Equation 5), clustering coefficient (C_i; Equation 6) and betweenness centrality (b_i; Equation 7) at each node to quantify the local connectivity. Both local efficiency and clustering coefficient measure the presence of well-connected clusters around the node, and they are highly correlated to each other. The betweenness centrality is the number of shortest paths which pass through the node, and measures the effect of the node on the overall flow of information in the network[16]. The local connectivity metrics used in this work, for a single node i, are defined as follows;

$$E_{loc,i} = \frac{\sum_{j,h \in N, j \neq i} a_{ij} a_{ih} \left[d_{jh}(N_i) \right]^{-1}}{k_i(k_i - 1)}, \tag{5}$$

where, $d_{jh}(N_i)$, is the length of the shortest path between node j and h - as defined in Equation, and contains only neighbours of h 1.

$$C_i = \frac{2t_i^w}{k_i(k_i - 1)}. \tag{6}$$

$$b_i = \frac{1}{(n-1)(n-2)} \sum_{h,j \in N, h \neq j, h \neq i, i \neq j} \frac{\rho_{hj}(i)}{\rho_{hj}}, \tag{7}$$

where $\rho_{hj}(i)$ is the weights of shortest path between h and j that passes through i.

Statistical Analysis

We used different statistical methods as described below; however, for the multiple testing we relied on the Bonferroni correction[57,58]. Where applicable, the thresholds were obtained by dividing 0.05 by the number of independent tests. We stated the corrected threshold and number of independent tests at the caption of each table in the Results section.

Quantifying the Change in CDR and Connectivity Metrics

To determine the longitudinal change in CDR, local and global connectivity metrics, we calculated the absolute difference between the first visit (the baseline visit) and the first visit after 12 months (the follow-up visit). Unless stated otherwise, this is the primary way of quantifying this longitudinal change used in the analysis.

Estimation of Gene Expression from Multiple Probe Sets

Different probe set expression values were present for each gene in the data. To estimate a representative gene expression out of the probe set expression, we conducted a non-parametric Mann-Whitney U test to evaluate whether the expression in AD was different from those of controls. For each gene, we selected the probe set expression that has the lowest Mann-Whitney U p-value. In this way, we selected the most differentially expressed probe sets in our data and considered those for the remaining analysis.

Spearman's Rank Correlation Coefficient

To test the statistical significance of pair-wise undirected relationships, we used the Spearman's rank correlation coefficient (ρ). The Spearman coefficient is a non-parametric method which ranks pairs of measurements and assesses their monotonic relationship. We report here the coefficient ρ along with the corresponding p-value to evaluate the significance of the relationship. A ρ of ± 1 indicates a very strong relationship, while $\rho = 0$ means there is no relationship.

Quantile Regression

To model the directed relationship between two variables, we used the quantile regression model[59]. This model is used as an alternative to the linear regression model when the linear regression assumptions are not met. This fact allows the response and predictor variables to have a non-symmetric distribution. The quantile regression model estimates the conditional median of the dependent variable given the independent variables. Besides, it can be used to estimate any conditional quantile; and is therefore robust to outliers. Accordingly, in this work we used the quantile regression to model the directed relationship between two variables. As an alternative to the linear regression, we chose the 50th percentile (the median) as our quantile and estimated the conditional median (rather than the conditional mean in case of the ordinary linear regression) of the dependent variable across given values of the independent variable.

Ridge Regression

For estimating the relationship between more than two variables, we used ridge regression[60]. The basic idea behind this model is that it solves the least square function penalizing it using the l_2 norm regularization. More specifically, the ridge regression minimizes the following objective function:

$$\begin{aligned} &||y - X\beta||_2^2 + \alpha||\beta||_2^2, \\ &\text{i.e.,} \\ \beta^{Ridge} =& \underset{\beta \in \mathbb{R}}{argmin} ||y - X\beta||_2^2 + \alpha||\beta||_2^2, \end{aligned} \tag{8}$$

where y is the dependent (or response) variable, X is the independent variable (feature, or predictor), β is the ordinary least square coefficient (or, the slope), α is the regularization parameter, β^{Ridge} is the ridge regression coefficient, *argmin* is the argument of minimum and it is responsible for making the function attain the minimum and is $L_2(v) = ||v||_2$ represents the L2 norm function[61]. Moreover, we normalized the predictors to get a more robust estimation of our parameters.

Software

We used python 3.7.1 for this work; our code has been made available under the MIT License https://choosealicense.com/licenses/mit/, and is accessible at https://github.com/elssam/RGLCG.

References

1. Barnes, D. E. & Yaffe, K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurol.* **10**, 819–828 (2011).
2. Gatz, M. *et al.* Heritability for Alzheimer's disease: the study of dementia in swedish twins. *The Journals Gerontol. Ser. A: Biol. Sci. Med. Sci.* **52**, M117–M125 (1997).
3. Gaiteri, C., Mostafavi, S., Honey, C. J., De Jager, P. L. & Bennett, D. A. Genetic variants in Alzheimer disease—molecular and brain network approaches. *Nat. Rev. Neurol.* **12**, 413 (2016).
4. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. genetics* **45**, 1452 (2013).
5. Escott-Price, V. *et al.* Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. *PloS one* **9**, e94661 (2014).
6. Elsheikh, S. S., Chimusa, E. R., Mulder, N. J. & Crimi, A. Genome-wide association study of brain connectivity changes for alzheimer's disease. *Sci. Reports* **10**, 1–16 (2020).
7. Corder, E., Saunders, A. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
8. Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704 (1991).
9. Levy-Lahad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973–977 (1995).

10. Rogaev, E. *et al.* Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. *Nature* **376**, 775 (1995).

11. Shaw, P. *et al.* Cortical morphology in children and adolescents with different apolipoprotein e gene polymorphisms: an observational study. *The Lancet Neurol.* **6**, 494–500 (2007).

12. Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS biology* **6**, e159 (2008).

13. Thompson, P. M. *et al.* The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging behavior* **8**, 153–182 (2014).

14. Alexander, A. L., Lee, J. E., Lazar, M. & Field, A. S. Diffusion tensor imaging of the brain. *Neurotherapeutics* **4**, 316–329 (2007).

15. Deco, G., Tononi, G. *et al.* Rethinking segregation and integration: contributions of whole-brain modelling. *Nat. Rev. Neurosci.* **16**, 430–439 (2015).

16. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).

17. Brown, J. *et al.* Brain network local interconnectivity loss in aging apoe-4 allele carriers. *Proc. Natl. Acad. Sci.* **108**, 20760–20765 (2011).

18. Chen, Y., Chen, K. *et al.* Disrupted functional and structural networks in cognitively normal elderly subjects with the APOE ε4 allele. *Neuropsychopharmacology* **40**, 1181 (2015).

19. Jahanshad, N. *et al.* Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc. Natl. Acad. Sci.* **110**, 4768–4773 (2013).

20. Chiesa, P. A. *et al.* Revolution of resting-state functional neuroimaging genetics in Alzheimer's disease. *Trends neurosciences* **40**, 469–480 (2017).

21. van den Heuvel, M. P., Mandl, R. C., Stam, C. J., Kahn, R. S. & Pol, H. E. H. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *J. Neurosci.* **30**, 15915–15926 (2010).

22. Rudie, J. D. *et al.* Altered functional and structural brain network organization in autism. *NeuroImage: clinical* **2**, 79–94 (2013).

23. Griffa, A., Baumann, P. S., Thiran, J.-P. & Hagmann, P. Structural connectomics in brain diseases. *Neuroimage* **80**, 515–526 (2013).

24. Cohen, J. R. & D'Esposito, M. The segregation and integration of distinct brain networks and their relationship to cognition. *J. Neurosci.* **36**, 12083–12094 (2016).

25. Rubinov, M. Constraints and spandrels of interareal connectomes. *Nat. communications* **7**, 1–11 (2016).

26. Morris, J. C. *et al.* Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's disease cooperative study experience. *Neurology* **48**, 1508–1510 (1997).

27. Szklarczyk, D. *et al.* The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research* gkw937 (2016).

28. Prescott, J. W. *et al.* The Alzheimer structural connectome: changes in cortical network topology with increased amyloid plaque burden. *Radiology* **273**, 175–184 (2014).

29. Alloza, C. *et al.* Polygenic risk score for schizophrenia and structural brain connectivity in older age: a longitudinal connectome and tractography study. *Neuroimage* **183**, 884–896 (2018).

30. Robson, K. *et al.* Synergy between the c2 allele of transferrin and the c282y allele of the haemochromatosis gene (hfe) as risk factors for developing Alzheimer's disease. *J. medical genetics* **41**, 261–265 (2004).

31. Pilling, L. C. *et al.* Common conditions associated with hereditary haemochromatosis genetic variants: cohort study in uk biobank. *bmj* **364**, k5222 (2019).

32. Pujol, J. *et al.* Anatomical variability of the anterior cingulate gyrus and basic dimensions of human personality. *Neuroimage* **15**, 847–855 (2002).

33. Wilson, R. S. *et al.* Harm avoidance and risk of Alzheimer's disease. *Psychosom. Medicine* **73**, 690 (2011).

34. Galton, C. J. *et al.* Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology* **57**, 216–225 (2001).

35. Finckh, U. *et al.* Association of late-onset Alzheimer disease with a genotype of plau, the gene encoding urokinase-type plasminogen activator on chromosome 10q22. 2. *Neurogenetics* **4**, 213–217 (2003).

36. Namba, Y., Ouchi, Y., Takeda, A., Ueki, A. & Ikeda, K. Bleomycin hydrolase immunoreactivity in senile plaque in the brains of patients with Alzheimer's disease. *Brain research* **830**, 200–202 (1999).

37. Papassotiropoulos, A. *et al.* Confirmation of the association between bleomycin hydrolase genotype and Alzheimer's disease. *Mol. psychiatry* **5**, 213 (2000).

38. Farrer, L. A. *et al.* Association between bleomycin hydrolase and Alzheimer's disease in caucasians. *Annals neurology* **44**, 808–811 (1998).

39. Kajiya, A., Kaji, H., Isobe, T. & Takeda, A. Processing of amyloid β-peptides by neutral cysteine protease bleomycin hydrolase. *Protein peptide letters* **13**, 119–123 (2006).

40. Grothe, M. J. *et al.* Molecular properties underlying regional vulnerability to alzheimer's disease pathology. *Brain* **141**, 2755–2771 (2018).

41. Cho, H. *et al.* Tau pet in alzheimer disease and mild cognitive impairment. *Neurology* **87**, 375–383 (2016).

42. Suszyńska-Zajczyk, J., Łuczak, M., Marczak, Ł. & Jakubowski, H. Hyperhomocysteinemia and bleomycin hydrolase modulate the expression of mouse brain proteins involved in neurodegeneration. *J. Alzheimer's Dis.* **40**, 713–726 (2014).

43. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391 (2012).

44. Richiardi, J. *et al.* Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).

45. Palmqvist, S. *et al.* Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nat. communications* **8**, 1–13 (2017).

46. Canter, R. G. *et al.* 3d mapping reveals network-specific amyloid progression and subcortical susceptibility in mice. *Commun. biology* **2**, 1–12 (2019).

47. Tzourio-Mazoyer, N., Landeau, B. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).

48. Saykin, A. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimer's & Dementia* **11**, 792–814 (2015).

49. Irizarry, R. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

50. Vawter, M. P. *et al.* Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology* **29**, 373 (2004).

51. Schadt, E. E., Woo, S. & Hao, K. Bayesian method to predict individual snp genotypes from gene expression data. *Nat. genetics* **44**, 603 (2012).

52. Smedley, D. *et al.* The biomart community portal: an innovative alternative to large, centralized data repositories. *Nucleic acids research* **43**, W589–W598 (2015).

53. Garyfallidis, E. *et al.* Dipy, a library for the analysis of diffusion mri data. *Front. neuroinformatics* **8**, 8 (2014).

54. Aganj, I. *et al.* Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magn. resonance medicine* **64**, 554–566 (2010).

55. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion mri streamlines tractography through effective use of anatomical information. *Neuroimage* **62**, 1924–1938 (2012).

56. Gordon, E. M. *et al.* Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. cortex* bhu239 (2014).

57. White, T., van der Ende, J. & Nichols, T. E. Beyond bonferroni revisited: concerns over inflated false positive research findings in the fields of conservation genetics, biology, and medicine. *Conserv. Genet.* 1–11 (2019).

58. Narum, S. R. Beyond bonferroni: less conservative analyses for conservation genetics. *Conserv. genetics* **7**, 783–787 (2006).

59. Koenker, R. & Hallock, K. F. Quantile regression. *J. economic perspectives* **15**, 143–156 (2001).

60. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).

61. Strang, G., Strang, G., Strang, G. & Strang, G. *Introduction to linear algebra*, vol. 3 (Wellesley-Cambridge Press Wellesley, MA, 1993).

Acknowledgements

We would like to acknowledge our funders, the Organisation for Women in Science for the Developing World (OWSD), the Swedish International Development Cooperation Agency (SIDA) and the University of Cape Town for their continuous support. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern

California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author Contributions

S.S.M.E. conducted the integrated imaging-genetics analysis and wrote the paper. A.C. preprocessed all neuroimaging data and set the connectome pipeline. N.J.M. and A.C. gave constructive feedback on this work continuously and followed up the analysis and writing progress closely. E.R.C. gave feedback on the overall paper. The final version of the paper was proofread by all authors.

Additional Information

Supplementary information: Supplementary materials available for this paper.

Competing interests: The authors declare no competing interests.

Data Availability: The data used in this work are available at the ADNI repository (http://adni.loni.usc.edu/).