Washington University School of Medicine
Digital Commons@Becker

Open Access Publications

2016

Chitinase-3-like 1 protein (CHI3L1) locus influences cerebrospinal fluid levels of YKL-40

Yuetiva Deming
Washington University School of Medicine in St. Louis

Kathleen Black
Washington University School of Medicine in St. Louis

David Carrell
Washington University School of Medicine in St. Louis

Yefei Cai
Washington University School of Medicine in St. Louis

Jorge L. Del-Aguila
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Authors

Yuetiva Deming, Kathleen Black, David Carrell, Yefei Cai, Jorge L. Del-Aguila, Maria Victoria Fernandez, John Budde, ShengMei Ma, Benjamin Saef, Bill Howells, Sarah Bertelsen, Kuan-lin Huang, Courtney L. Sutphen, Rawan Tarawneh, Anne M. Fagan, David M. Holtzman, John C. Morris, Alison M. Goate, Joseph D. Dougherty, and Carlos Cruchaga

BMC Neurology

Chitinase-3-like 1 protein (*CHI3L1*) locus influences cerebrospinal fluid levels of YKL-40

Yuetiva Deming[1] , Kathleen Black[1], David Carrell[1], Yefei Cai[1], Jorge L. Del-Aguila[1], Maria Victoria Fernandez[1], John Budde[1], ShengMei Ma[1], Benjamin Saef[1], Bill Howells[1], Sarah Bertelsen[2], Kuan-lin Huang[3], Courtney L. Sutphen[4], Rawan Tarawneh[4,5,6], Anne M. Fagan[4,5,6], David M. Holtzman[4,5,6,7], John C. Morris[4,5,6,7], Alison M. Goate[2], Joseph D. Dougherty[1,3] and Carlos Cruchaga[1,6*]

Abstract

Background: Alzheimer's disease (AD) pathology appears several years before clinical symptoms, so identifying ways to detect individuals in the preclinical stage is imperative. The cerebrospinal fluid (CSF) Tau/Aβ_{42} ratio is currently the best known predictor of AD status and cognitive decline, and the ratio of CSF levels of chitinase-3-like 1 protein (*CHI3L1*, YKL-40) and amyloid beta (Aβ_{42}) were reported as predictive, but individual variability and group overlap inhibits their utility for individual diagnosis making it necessary to find ways to improve sensitivity of these biomarkers.

Methods: We used linear regression to identify genetic loci associated with CSF YKL-40 levels in 379 individuals (80 cognitively impaired and 299 cognitively normal) from the Charles F and Joanne Knight Alzheimer's Disease Research Center. We tested correlations between YKL-40 and CSF Tau/Aβ_{42} ratio, Aβ_{42}, tau, and phosphorylated tau (ptau$_{181}$). We used studentized residuals from a linear regression model of the log-transformed, standardized protein levels and the additive reference allele counts from the most significant locus to adjust YKL-40 values and tested the differences in correlations with CSF Tau/Aβ_{42} ratio, Aβ_{42}, tau, and ptau$_{181}$.

Results: We found that genetic variants on the *CH13L1* locus were significantly associated with CSF YKL-40 levels, but not AD risk, age at onset, or disease progression. The most significant variant is a reported expression quantitative trait locus for *CHI3L1*, the gene which encodes YKL-40, and explained 12.74 % of the variance in CSF YKL-40 in our study. YKL-40 was positively correlated with ptau$_{181}$ (r = 0.521) and the strength of the correlation significantly increased with the addition of genetic information (r = 0.573, *p* = 0.006).

Conclusions: CSF YKL-40 levels are likely a biomarker for AD, but we found no evidence that they are an AD endophenotype. YKL-40 levels are highly regulated by genetic variation, and by including genetic information the strength of the correlation between YKL-40 and ptau$_{181}$ levels is significantly improved. Our results suggest that studies of potential biomarkers may benefit from including genetic information.

Keywords: CHI3L1, YKL-40, Cerebrospinal fluid, Alzheimer disease

* Correspondence: ccruchaga@wustl.edu
[1]Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave. B8134, St. Louis, MO 63110, USA
[6]Hope Center for Neurological Disorders, Washington University School of Medicine, 660 S. Euclid Ave. B8111, St. Louis, MO 63110, USA
Full list of author information is available at the end of the article

Deming *et al. BMC Neurology* (2016) 16:217

Page 2 of 8

Background

Since it has been well demonstrated that Alzheimer's disease (AD) pathology is present long before any clinical symptoms, emphasis has been on improving identification of those in this preclinical stage [1–3]. Studies have found that the ratio of cerebrospinal fluid (CSF) tau to $A\beta_{42}$ is a better predictor of AD status and cognitive decline than either of these proteins individually and improves discrimination of AD from other dementias such as vascular dementia and frontotemporal lobar degeneration [4–6]. CSF levels of tau correlate with the amount of neurodegeneration, phosphorylated tau (ptau) levels correlate with tangle pathology, and $A\beta_{42}$ levels inversely correlate with the amount of plaques, which makes these ideal biomarkers for AD pathology [3, 7]. Studies have focused on targeting $A\beta$, tau, or both proteins as potential treatments for AD, making it necessary to find alternative biomarkers that are not being directly targeted [8–11]. CSF levels of YKL-40 are a promising biomarker for AD; they are significantly higher in individuals with AD dementia than in cognitively normal individuals [12]. CSF YKL-40 is also associated with cortical thinning in cognitively normal individuals with low levels of CSF $A\beta_{42}$ who are at risk for AD, and highly correlates with CSF ptau [13]. YKL-40 is a secreted glycoprotein, encoded by chitinase-3-like 1 protein (*CHI3L1*), expressed in astrocytes and associated with neuroinflammatory response. Although the exact function of YKL-40 is unknown, it has been associated with many immune and inflammatory diseases as well as several cancers [14, 15]. Studies have shown that CSF levels of YKL-40 can be used to distinguish AD from non-AD dementia, Parkinson's disease, dementia with Lewy bodies, and to distinguish between mild cognitive impairment (MCI) that progresses to AD vs MCI not due to AD indicating that CSF YKL-40 is not simply a marker for inflammation due to neurodegeneration but may be specific enough for AD [16, 17]. However, due to individual variability and group overlap in these CSF protein levels, further studies are necessary to find ways to improve sensitivity and specificity for AD. We hypothesized that adding genetic information may provide a means to improve specificity and sensitivity of CSF YKL-40 as a biomarker for AD.

One key feature of endophenotypes is that they have a clear genetic connection with the trait of interest, and for a biomarker to be considered an endophenotype it has to be measureable, heritable, and segregate with status. The genetic connection between endophenotypes and complex traits has been demonstrated to provide power in genetic studies to identify novel variants and inform about possible underlying biology [18, 19]. CSF levels of tau, $ptau_{181}$, $A\beta_{42}$ are well-established AD endophenotypes and have allowed us to identify novel variants associated not only with tau and $ptau_{181}$ levels

but also with AD risk, tangle pathology, and cognitive decline [18]. Recent research suggests that neuroinflammation is not merely a by-product of neurodegeneration in AD but may play a key role in pathology [20]. Since CSF levels of YKL-40 can distinguish between AD and non-AD MCI, perhaps the unknown function of YKL-40 in neuroinflammation is involved in the progression of pathology. If polymorphisms regulating YKL-40 also contribute to some aspect of AD such as risk, age at onset, or cognitive decline then understanding their genetic regulation can help provide insights into biological mechanisms underlying AD, and can help to determine whether YKL-40 is really involved in the pathogenesis (endophenotype) or is just a biomarker for the disease.

First we performed single variant genetic analyses of CSF levels of YKL-40 to identify single nucleotide polymorphisms (SNPs) associated with YKL-40 levels. Next we analyzed whether genome-wide significant variants from our analyses were also associated with AD risk, age at onset, or progression. Finally, we tested correlations between YKL-40 and the CSF Tau/$A\beta_{42}$ ratio, $A\beta_{42}$, tau, and $ptau_{181}$, before and after including genetic information for YKL-40 levels, to determine if adding genetic information can improve the correlation with other known biomarkers, especially $ptau_{181}$.

Methods

Ethics statement

The Washington University Institutional Review Board approved this study. Written informed consent was obtained from participants or their family members.

Study participants

All 379 individuals with measured CSF levels of YKL-40 were from the Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight-ADRC; Table 1). Dementia severity was determined using the Clinical Dementia Rating (CDR) where 0 indicates cognitive normality, 0.5 is defined as very mild dementia, 1 is mild dementia, 2 is moderate dementia, and 3 is severe dementia [21]. There were 80 cases, defined as individuals with CDR > 0 at lumbar puncture, and 299 cognitively normal controls (CDR = 0 at lumbar puncture; Table 1). Neuropsychological and clinical assessments were collected for all participants and CSF was collected in the morning after an overnight fast, processed, and stored at -80 °C, as described previously [22]. Individuals were evaluated by Clinical Core personnel at Washington University.

Genotyping and quality control

The samples were genotyped with the Illumina 610 or Omniexpress chip. Stringent quality control criteria were applied to each genotyping array separately before

Deming *et al. BMC Neurology* (2016) 16:217

Page 3 of 8

Table 1 Characteristics of CSF YKL-40 data

	Knight-ADRC
Samples	379
Age in years (mean ± SD)	70.86 ± 8.80
(range)	49–91
Gender (M/F)	148/231
% CDR >0	21.11
% APOE E4	36.68
CSF YKL-40 (ng/mL)	307.25 ± 108.45

combining data. A ≤98 % call rate was applied for single nucleotide polymorphisms (SNPs) and individuals. SNPs not in Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$) or with MAF <0.02 were excluded. X-chromosome SNPs were analyzed to verify gender identification. Pairwise genome-wide estimates of proportion identity-by-descent were used to find duplicate and related individuals which were eliminated from the analysis. Principal components were calculated using EIGENSTRAT [23] to confirm ethnicity of each sample. Imputation was performed as described previously [18]. Briefly, BEAGLE v3.3.1 software [24] and the 1,000 genome data were used to impute up to 6 million SNPs. There were 5,986,883 imputed and genotyped SNPs after removing SNPs with a call rate <95 % or a BEAGLE $r^2 \leq 0.3$.

Analyte measurements and quality control
The Knight-ADRC Biomarker Core measured CSF levels of $A\beta_{42}$, tau, and $ptau_{181}$ using single-analyte enzyme-linked immunosorbent assays (ELISA) as described previously [22]. CSF YKL-40 levels were measured using the MicroVue ELISA (Quidel) as described previously [12].

Statistical analyses
Log-transformed, standardized values for CSF YKL-40 levels were tested for normality using the Shapiro-Wilk test. We used R v3.2.1 [25] to perform linear regression to determine if CSF levels of YKL-40 were influenced by age, gender, or sample batch (Additional file 1: Table S1). Age, gender, and sample batch were used as covariates to test association of YKL-40 with cognitive status. Correlations of CSF levels of YKL-40 with CSF $A\beta_{42}$, tau, $ptau_{181}$, and $Tau/A\beta_{42}$ ratio were calculated using Pearson's correlation.

Genetic association with CSF levels of YKL-40 were tested using an additive model in PLINK v1.9 (http://www.cog-genomics.org/plink2) [26]. Covariates used were sample batch, age, gender, and two principal component factors for population structure. Statistical significance for single variant association was defined as $p < 5 \times 10^{-8}$ based on the commonly used threshold considered appropriate for the likely number of independent tests with Bonferroni

correction. The threshold of $p < 1 \times 10^{-5}$ was defined for suggestive association. The genomic inflation factor for association with YKL-40 levels was 1, suggesting no evidence of inflation due to population stratification. SNP annotation was performed using ANNOVAR version 2015-06-17 [27] and the NCBI Database of Single Nucleotide Polymorphisms (dbSNP) Build ID: 142 (http://www.ncbi.nlm.nih.gov/SNP) [28]. RegulomeDB v1.1 (http://regulome.stanford.edu/index) [29] was used to determine if SNPs of interest were potential regulatory elements. The Genotype-Tissue Expression (GTEx) Analysis Release V6, dbGaP Accession phs000424.v6.p1 (http://www.gtexportal.org) [30] was used to determine if SNPs of interest were potential expression quantitative trait loci (eQTLs).

Disease progression was modeled as the change in CDR Sum of Boxes (CDR-SB) per year. A total of 1,646 individuals from longitudinal studies of AD patients with ≥3 clinical assessments over 1.5 years after being diagnosed with AD were included in the analysis. A mixed-model repeated measure framework was used to account for correlation between repeated measures in the same individual. Age, sex, baseline CDR, follow-up time, level of education, site, and PCs were included as covariates. The appropriate optimal variance-covariance structure that minimizes the Akaike Information Criterion for testing the null model AR1 was selected [31].

To estimate the proportion of variance in CSF levels of YKL-40 explained by genetic variants we used the coefficient of determination (R^2) of a linear model of the log-transformed, standardized CSF protein levels and the additive reference allele counts of the top genome-wide associated SNP (rs10399931) with sample batch, age, and gender as covariates and subtracted the R^2 of a linear model with the log-transformed, standardized CSF protein levels and the covariates in the null model. To adjust CSF levels of YKL-40 for genetic effect, we used studentized residuals from a linear regression model of the log-transformed, standardized protein levels and the additive reference allele counts of rs10399931. Age, gender, and sample batch were used as covariates to test association of the adjusted levels of YKL-40 with AD status. Correlations of the adjusted levels of YKL-40 with CSF $A\beta_{42}$, tau, $ptau_{181}$, and $tau/A\beta_{42}$ ratio were determined using Pearson's correlation (r). The R package cocor version 1.1-1 [32] was used to compare correlations between CSF $A\beta_{42}$, tau, $ptau_{181}$, and $tau/A\beta_{42}$ ratio and adjusted or unadjusted CSF protein levels. We reported the results using the Meng Z-test model [33], but the results for all models reported by cocor were comparable.

See Additional file 2: Supplementary Methods and Results for gene ontology over-representation and tissue-specific expression analyses.

Deming *et al. BMC Neurology* (2016) 16:217

Page 4 of 8

Results

CSF levels of YKL-40 in AD cases vs controls

After applying stringent quality control, our dataset contained CSF levels of YKL-40 for 379 individuals from the Charles F. and Joanne Knight Alzheimer's Disease Research Center (Knight-ADRC; Table 1). We used logistic regression including age, gender, and sample batch as covariates to test whether levels of YKL-40 were associated with AD status defined by CDR. Cases (defined by CDR at lumbar puncture >0) had significantly higher CSF levels of YKL-40 than controls (cases: 366.37 ± 136.48 ng/mL; controls: 290.18 ± 92.44 ng/mL; $p = 0.015$, $\beta = 0.698$; Additional file 3: Figure S1). We used linear regression to test if *APOE* genotype influenced YKL-40 and found there was no effect on YKL-40 levels ($p = 0.704$). Several studies indicate that CSF Tau/Aβ_{42} ratio is a better predictor of disease status than clinical assessment [5, 6]. We tested the correlation between CSF Tau/Aβ_{42} ratio and YKL-40. Levels of YKL-40 were positively correlated with CSF Tau/Aβ_{42} ratio ($p = 2.61 \times 10^{-8}$, r = 0.318), most likely due to the positive correlation with tau ($p = 7.06 \times 10^{-22}$, r = 0.522) since tau and Tau/Aβ_{42} ratio are highly correlated ($p = 3.99 \times 10^{-52}$, r = 0.736) and although there is a high negative correlation between Tau/Aβ_{42} ratio and Aβ_{42} levels ($p = 3.07 \times 10^{-64}$, r = -0.788), YKL-40 was not correlated with Aβ_{42} ($p = 0.838$, r = 0.012; Table 2 and Additional file 4: Figure S2).

Single variant analysis of CSF YKL-40 levels

To determine whether genetic variants are associated with levels of YKL-40, we used linear regression to test the additive genetic model of each single nucleotide polymorphism (SNP) for association with CSF protein levels using age, gender, sample batch, and two principal component factors for population stratification as covariates. We found 14 genome-wide significant SNPs associated with YKL-40, all in the *CHI3L1* locus. The most significant SNP was rs10399931, a variant <200 bp upstream of the transcription start site for *CHI3L1* (genotyped, $p = 1.76 \times 10^{-14}$, $\beta = -0.575$, minor allele frequency (MAF) = 0.244; Fig. 1 and Additional file 5: Table S2). We did not find any additional genome-wide significant loci when we ran the analysis conditioned

on rs10399931. Data from RegulomeDB [29] indicates that rs10399931 is likely to affect binding and gene expression; two RNA-Seq studies and the Genotype-Tissue Expression (GTEx) database have reported rs10399931 is an eQTL for *CHI3L1* (whole blood: $p = 1.7 \times 10^{-11}$, $\beta = -0.35$; transformed fibroblasts: $p = 3.1 \times 10^{-11}$, $\beta = -0.37$; thyroid: $p = 5.7 \times 10^{-10}$, $\beta = -0.42$; lung: $p = 6.7 \times 10^{-9}$, $\beta = -0.37$; tibial nerve: $p = 8.1 \times 10^{-9}$, $\beta = -0.33$; subcutaneous adipose: $p = 5.5 \times 10^{-7}$, $\beta = -0.28$; tibial artery: $p = 2.4 \times 10^{-6}$, $\beta = -0.26$) [30, 34, 35]. *CHI3L1* variants have been associated with YKL-40 levels in serum previously [36], but to our knowledge this is the first time it has been reported to be associated with CSF levels of YKL-40. We calculated what proportion of variance in CSF YKL-40 levels was explained by rs10399931. Age, gender, and sample batch explained 14.89 % of the variance in YKL-40 levels, and addition of the rs10399931 effect explained another 12.74 % of the variance. As a comparison, a previous GWAS study of CSF levels of tau and ptau$_{181}$ estimated that the genetic effect of the genome-wide significant association located in apolipoprotein E (*APOE*) explained only 0.25–0.29 % of the variability of tau and ptau$_{181}$ levels [18]. Additionally, we found a signal in chromosome 13 (rs78081700, $p = 6.26 \times 10^{-8}$, $\beta = 0.636$, Fig. 1 and Additional file 5: Table S2) that almost reached genome-wide significance and another eight loci with suggestive p values (Fig. 1 and Additional file 5: Table S2), indicating that additional loci/genes could be associated with CSF YKL-40, although a larger sample size would be needed to confirm this hypothesis.
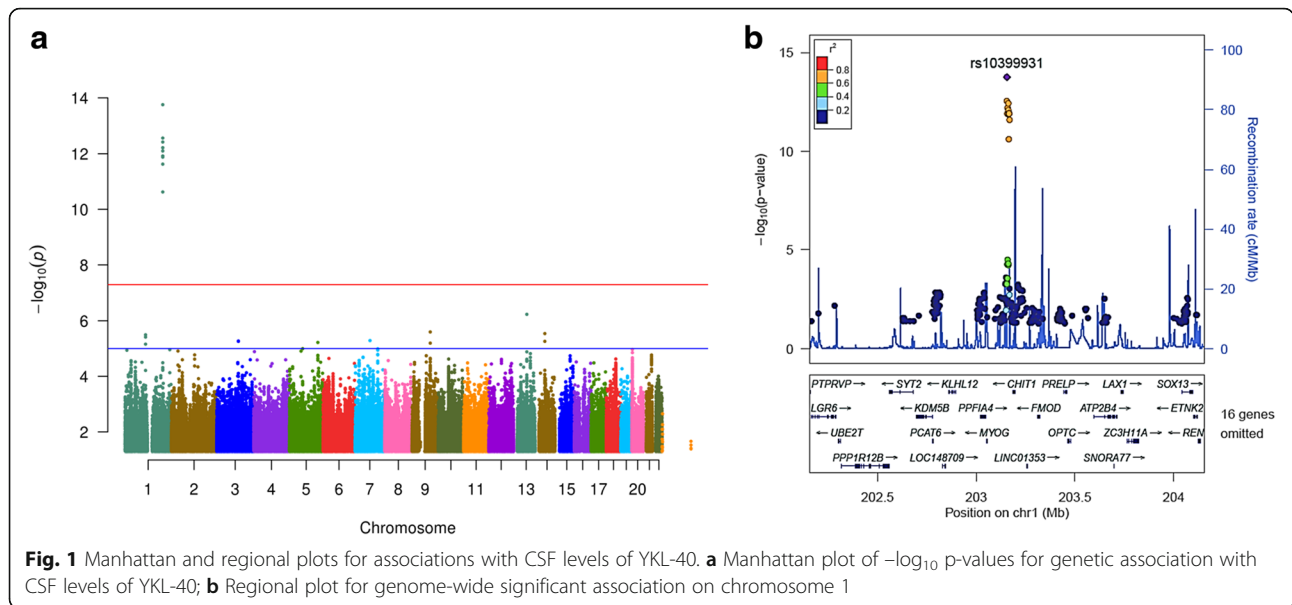
Association of GWAS hits and pathways with AD risk, age at onset, and disease progression

To determine if our genome-wide significant SNPs were also associated with risk for AD we searched the results from a GWAS previously published by the International Genomics of Alzheimer's Project (I-GAP) consisting of a total 25,580 AD cases and 48,466 controls [37]. None of our genome-wide significant SNPs had significant p-values in the I-GAP results (rs10399931: $p = 0.763$, $\beta = 0.006$; Table 3). Based on analyses of data from a previously published GWAS investigating genetic

Table 2 Correlations of CSF levels of YKL-40 with CSF Tau/Aβ ratio, AB42, ptau$_{181}$, and tau levels before and after adjusting for top SNP effect

	CSF Tau/Aβ ratio		CSF Aβ_{42} levels		CSF ptau$_{181}$ levels		CSF tau levels	
	p	r	p	r	p	r	p	r
CSF YKL-40	2.61×10^{-8}	0.318	0.838	0.012	8.98×10^{-22}	0.521	7.06×10^{-22}	0.522
CSF YKL-40-rs10399931 effect	5.28×10^{-10}	0.357	0.972	0.002	2.82×10^{-26}	0.573	1.66×10^{-26}	0.575
	p	95 % CI	p	95 % CI	p	95 % CI	p	95 % CI
Correlation difference (Meng Z-test)	0.071	-0.083, 0.003	0.661	-0.031, 0.049	0.006	-0.116, -0.020	0.006	-0.116, -0.020

95 % CI: 95 % confidence interval, null hypothesis was retained if interval included 0

Deming *et al. BMC Neurology* (2016) 16:217

Page 5 of 8

Fig. 1 Manhattan and regional plots for associations with CSF levels of YKL-40. **a** Manhattan plot of $-\log_{10}$ p-values for genetic association with CSF levels of YKL-40; **b** Regional plot for genome-wide significant association on chromosome 1

variants associated with age at onset for AD [38] our genome-wide significant SNPs did not appear to be associated with age at onset (rs10399931: $p = 0.197$, $\beta = 0.026$; Table 3). We also found that our genome-wide significant SNPs were not significantly associated with advancement in CDR (rs10399931: $p = 0.142$, $\beta = -0.072$; Table 3). Recently the I-GAP published results from a study of biological pathways and gene expression networks associated with AD [39]. None of the gene ontology terms that were enriched in our gene ontology analyses of GWAS results for YKL-40 (Additional file 6: Table S3) were found in the I-GAP report [39]. Together these data suggest that YKL-40 may be a promising biomarker for AD, but probably not an endophenotype.

Improving CSF biomarkers and CSF YKL-40 utility by including genetic information

Because the genome-wide significant SNPs are not associated with disease status but explain a large proportion of the CSF levels of YKL-40, we hypothesized that by accounting for genetic information, it would be possible to improve the efficacy of these measures as biomarkers. Levels of YKL-40 have been reported to be correlated with CSF levels of tau and ptau$_{181}$ but not Aβ_{42}, and this was replicated in our analyses (Table 2 and Additional file 4: Figure S2) [12, 15].

We decided to focus on CSF ptau$_{181}$ and YKL-40 since these were highly correlated and the results were similar

for tau. First we used linear regression to determine whether rs10399931 genotype influenced ptau$_{181}$ levels ($p = 0.782$, $R^2 = -0.005$). Then we included the additive model for rs10399931 in YKL-40 levels and re-analyzed the correlation with ptau$_{181}$. The correlation coefficient for the adjusted values of CSF YKL-40 levels and ptau$_{181}$ was higher than the unadjusted values (adjusted: $p = 2.82 \times 10^{-26}$, $r = 0.573$ vs. unadjusted: $p = 8.98 \times 10^{-22}$, $r = 0.521$; Table 2). We used the R package cocor [32] and determined that this change in correlation was statistically significant ($p = 0.006$, 95 % CI: -0.116, -0.020; Table 2). Similar results were found for tau (Table 2). As expected, we did not find any significant difference in the correlation between Aβ_{42} and the corrected or uncorrected YKL-40 values (Table 2). For CSF Tau/Aβ_{42} ratio, which is a powerful predictor for AD and progression, we found a marginally significant improvement when the genetic information was included (adjusted: r = 0.357 vs. unadjusted: $r = 0.318$; $p = 0.071$, 95 % CI: -0.083, 0.003, Table 2).

Discussion

AD pathology is present long before any clinical symptoms [5, 21, 40–45], and multiple clinical trials are being performed in pre-symptomatic individuals. However, it is necessary to have reliable biomarkers to identify these individuals and to monitor the efficacy of new treatments. CSF ptau$_{181}$ and Aβ_{42} have emerged as the most promising biochemical biomarkers for AD risk and progression

Table 3 Association of genome-wide significant locus from CSF YKL-40 GWAS with AD risk, age at onset, and disease progression

					CSF YKL-40		AD risk		Age at onset		Progression	
Chr	Position	SNP	A1	MAF	p	β	p	β	p	β	p	β
1	203156080	rs10399931	T	0.244	1.76×10^{-14}	-0.575	0.763	0.006	0.197	0.026	0.142	-0.072

Deming *et al. BMC Neurology* (2016) 16:217

Page 6 of 8

[46]. Although these CSF biomarker levels are highly associated with AD, there is large inter-individual variability, and a relatively large overlap in the absolute CSF levels between cases and controls. Additionally, if treatments directly target tau proteins or $A\beta_{42}$, the CSF levels may no longer be an informative biomarker because the treatment could affect those levels separately from disease state. Therefore, the identification of surrogates for $A\beta_{42}$ and tau and novel approaches to improve accuracy of CSF biomarkers that can control for the individual inter-variability can have a large impact on clinical trials as well as in the general practice.

CSF YKL-40 has emerged as a novel potential biomarker for AD, as it is higher in individuals with AD than in cognitively normal individuals and is highly correlated with CSF $ptau_{181}$ levels [2, 12, 13, 15]. However, it is not clear whether CSF YKL-40 is simply a biomarker or also an endophenotype for AD. Biomarkers are measurable biological characteristics that can be used as indicators of complex traits, but don't necessarily provide information about the biology of the trait. They may simply be influenced by the same biological processes rather than being part of those processes. Endophenotypes are biomarkers that are heritable traits with a genetic connection with disease, can be measured in all individuals regardless of disease status, and therefore can be highly informative about the biological causes of the disease.

In this study, we used genomic approaches to determine whether CSF YKL-40 is an endophenotype for AD, and whether CSF YKL-40 becomes more informative by adding genetic information. Here we reported for the first time a genetic analysis for CSF YKL-40 levels. We found a genome-wide significant signal on *CHI3L1*, which encodes YKL-40 (rs10399931: $p = 1.76 \times 10^{-14}$, MAF = 0.24), and multiple suggestive signals on other chromosomes. Interestingly, rs10399931 alone explains almost as much of the variance in YKL-40 levels ($R^2 = 0.127$) as both age and gender combined ($R^2 = 0.149$). Neither this variant or any of the suggestive variants were associated with AD risk, age at onset, or disease progression, indicating that CSF YKL-40 is probably not an endophenotype for AD.

Because of the large proportion of CSF YKL-40 variability explained by rs10399931, we analyzed whether the correlation with CSF $ptau_{181}$ improves when including the genotypic information for this variant. We hypothesize that in order to use genetic information to improve biomarker efficacy, the genetic association must be strong and replicable. We also predict that the improvement of the efficacy of the biomarker could be correlated with the proportion of biomarker level variability explained by the genetic variant. Additionally, if multiple loci are significantly associated with biomarker levels, a polygenic risk score should provide better performance than single locus analyses. As hypothesized, we found a significant increase in the correlation of CSF YKL-40 with tau and $ptau_{181}$ when genetic information for rs10399931 was included in the model. The correlation with the Tau/$A\beta_{42}$ ratio also improved, although it did not reach statistical significance. Based on these results, we hypothesize that the addition of genotypic information could significantly increase the sensitivity and specificity of biomarkers, but strong genetic associations are needed. When multiple loci are associated with biomarker levels, genetic risk scores would probably provide better results.

Additionally, we think this approach for improving biomarker efficacy by adjusting for genetic effect would only be effective for biomarkers, as may be the case for YKL-40, and not for informative endophenotypes such as CSF levels of $A\beta_{42}$ and tau. Genetic variants associated with endophenotype values are very likely to be associated with disease risk as well, as has been found previously in the case of AD endophenotypes [18, 47]. Therefore, correcting for genetic variants associated with endophenotype levels would generate a co-linearity problem and not improve, but actually worsen, the biomarker performance in that case.

Conclusions

Our genetic analyses indicate that although CSF YKL-40 levels are a promising biomarker for AD, there is no evidence they are an AD endophenotype. Genetic variation highly regulates CSF YKL-40 levels, and by including genetic information, the correlations with CSF tau and $ptau_{181}$ levels increase. Although additional studies are needed to confirm our hypothesis, our results suggest that studies of potential biomarkers for complex traits may benefit from correction of genetic effects on biomarker levels. This could be particularly important when using biomarkers as surrogate endpoints in clinical trials.

Additional files

Additional file 1: Table S1. Covariate associations with CSF YKL-40. Results (*p*-value and adjusted R^2) from regression of potentially confounding covariates: age at lumbar puncture, gender, and sample batch. Only age appeared significantly associated with CSF levels of YKL-40 ($p = 1.19 \times 10^{-18}$, $R^2 = 0.184$). (DOCX 16 kb)

Additional file 2: Supplementary Methods and Results. Methods and results for the gene ontology over-representation analyses of top SNPs from the single variant analysis of CSF YKL-40 levels. 70 genes mapped to the top SNPs ($p < 1 \times 10^{-4}$) were significantly enriched in human brain and pituitary tissue at all levels of specificity. Figure illustrates the different human tissues with expression data available with Benjamini-Hochberg corrected $p < 0.10$ in the enrichment analysis and the Table shows the results for all tissues with uncorrected $p < 0.05$ in at least one specificity index. The specificity index represents how specific a set of genes are to a particular tissue. (DOCX 163 kb)

Additional file 3: Figure S1. Beeswarm plot of the normalized CSF YKL-40 levels in cases (defined as CDR > 0 at time of lumbar puncture)

Deming *et al. BMC Neurology* (2016) 16:217

Page 7 of 8

compared to controls (CDR = 0 at time of lumbar puncture). $p = 0.015$. (DOCX 119 kb)

Additional file 4: Figure S2. Scatterplots of correlations between normalized values of CSF YKL-40 and Tau/Aβ_{42} ratio, levels of Aβ_{42}, ptau$_{181}$, and tau. Pearson's correlation (r). CSF YKL-40 positively correlated with Tau/Aβ_{42} ratio (a), ptau$_{181}$ (c), and tau (d), but was not correlated with Aβ_{42} (b). (DOCX 89 kb)

Additional file 5: Table S2. YKL-40 single variant analysis top loci ($p < 1 \times 10^{-5}$). Chr = chromosome, bp position = base pair position, SNP = rs ID for single nucleotide polymorphism, MAF = minor allele frequency (also effect allele), Gene = nearest gene. Chromosome and base pair position based on Build 37 of reference genome. (DOCX 14 kb)

Additional file 6: Table S3. Top gene ontology categories ($p < 0.05$) in both CPDB and PANTHER over-representation analyses of YKL-40 GWAS results ($p < 1 \times 10^{-4}$). GO ID = Identification number from the Gene Ontology Consortium, CPDB = Consensus Path Database, PANTHER = Protein Analysis Through Evolutionary Relationships. Thirteen gene ontology terms had $p < 0.05$ in both analyses. (DOCX 13 kb)

Abbreviations
AD: Alzheimer's disease; APOE: Apolipoprotein E; Aβ_{42}: Amyloid beta (1-42); CDR: Clinical dementia rating; CDR-SB: CDR Sum of Boxes; CHI3L1: Chitinase-3-like 1; CSF: Cerebrospinal fluid; ELISA: Enzyme-linked immunosorbent assay; eQTL: Expression quantitative trait locus; GWAS: Genome-wide association study; Knight-ADRC: Charles F. and Joanne Knight Alzheimer's Disease Research Center; MAF: Minor allele frequency; ptau$_{181}$: Phosphorylated tau (181); SNP: Single nucleotide polymorphism

Acknowledgements
We appreciate all the patients, control subjects, and their families for their participation in or support of this study.

Availability of data and materials
The phenotypic and genetic data for the Knight-ADRC are available to qualified investigators through http://knightadrc.wustl.edu/Research/ResourceRequest.htm.

Authors' contributions
YD analyzed data and wrote the manuscript. KB and DC performed genotyping. JLD-A, YC, SB, MVF, JB, SM, BS, and BH prepared genetic data: performed imputation, cleaning, and calculated principal components. CLS, and RT measured CSF levels of YKL-40. AMF, DH, JCM, KH, and AG provided data. JDD contributed conceptually to the analysis. CC prepared manuscript and supervised the project. All authors read and approved the final version of this manuscript.

Competing interests
JCM reported having participated in or currently participating in clinical trials of antidementia drugs sponsored by Janssen Immunotherapy, Pfizer, Eli Lilly and Co/Avid Radiopharmaceuticals, SNIFF (Study of Nasal Insulin to Fight Forgetfulness), and A4 Study (Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease) and serving as a consultant for Lilly USA, ISIS Pharmaceuticals, and the Charles Dana Foundation. DMH reported being a cofounder of C2N Diagnostics LLC; serving on the scientific advisory boards of AstraZeneca, Genentech, Neurophage, and C2N Diagnostics; and serving as a consultant for Eli Lilly and Co. Washington University receives grants to the laboratory of DMH from the Tau Consortium, Cure Alzheimer's Fund, the JPB Foundation, Eli Lilly and Co, Janssen, and C2N Diagnostics. AMF reported serving on the scientific advisory boards of IBL International and Roche and serving as a consultant for AbbVie and Novartis. The other co-authors reported no potential conflicts of interest.

Consent for publication
Not applicable – this manuscript does not contain any individual person's data.

Ethics approval and consent to participate
This research was approved by the Washington University Institutional Review Board. Written informed consent was obtained from participants and their family members by the Clinical and Genetics Core of the Knight-ADRC. The approval number for the Knight-ADRC Genetics Core family studies is 201104178.

Author details
[1]Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave. B8134, St. Louis, MO 63110, USA. [2]Ronald M. Loeb Center for Alzheimer's disease, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Genetics, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA. [4]Department of Neurology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA. [5]Knight Alzheimer's Disease Research Center, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA. [6]Hope Center for Neurological Disorders, Washington University School of Medicine, 660 S. Euclid Ave. B8111, St. Louis, MO 63110, USA. [7]Department of Developmental Biology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, USA.

References
1. Tarawneh R, Head D, Allison S, Buckles V, Fagan AM, Ladenson JH, Morris JC, Holtzman DM. Cerebrospinal fluid markers of neurodegeneration and rates of brain atrophy in early Alzheimer disease. JAMA Neurol. 2015;72(6):656–65.
2. Perrin RJ, Craig-Schapiro R, Malone JP, Shah AR, Gilmore P, Davis AE, Roe CM, Peskind ER, Li G, Galasko DR, et al. Identification and validation of novel cerebrospinal fluid biomarkers for staging early Alzheimer's disease. PLoS One. 2011;6(1):e16032.
3. Price JL, Morris JC. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. Ann Neurol. 1999;45(3):358–68.
4. de Jong D, Jansen RW, Kremer BP, Verbeek MM. Cerebrospinal fluid amyloid beta42/phosphorylated tau ratio discriminates between Alzheimer's disease and vascular dementia. J Gerontol A Biol Sci Med Sci. 2006;61(7):755–8.
5. Harari O, Cruchaga C, Kauwe JS, Ainscough BJ, Bales K, Pickering EH, Bertelsen S, Fagan AM, Holtzman DM, Morris JC, et al. Phosphorylated tau-Abeta42 ratio as a continuous trait for biomarker discovery for early-stage Alzheimer's disease in multiplex immunoassay panels of cerebrospinal fluid. Biol Psychiatry. 2014;75(9):723–31.
6. Fagan AM, Roe CM, Xiong C, Mintun MA, Morris JC, Holtzman DM. Cerebrospinal fluid tau/beta-amyloid (42) ratio as a prediction of cognitive decline in nondemented older adults. Arch Neurol. 2007;64(3):343–9.
7. Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol. 2010;6(3):131–44.
8. Lansdall CJ. An effective treatment for Alzheimer's disease must consider both amyloid and tau. Bioscience Horizons. 2014;7: hzu002 doi:10.1093/biohorizons/hzu002. published online June 17, 2014.
9. Gotz J, Ittner A, Ittner LM. Tau-targeted treatment strategies in Alzheimer's disease. Br J Pharmacol. 2012;165(5):1246–59.
10. Hong-Qi Y, Zhi-Kun S, Sheng-Di C. Current advances in the treatment of Alzheimer's disease: focused on considerations targeting Aβ and tau. Transl Neurodegener. 2012;1(1):1–12.

Deming *et al. BMC Neurology* (2016) 16:217

Page 8 of 8

11. Wischik CM, Harrington CR, Storey JMD. Tau-aggregation inhibitor therapy for Alzheimer's disease. Biochem Pharmacol. 2014;88(4):529–39.

12. Craig-Schapiro R, Perrin RJ, Roe CM, Xiong C, Carter D, Cairns NJ, Mintun MA, Peskind ER, Li G, Galasko DR, et al. YKL-40: a novel prognostic fluid biomarker for preclinical Alzheimer's disease. Biol Psychiatry. 2010;68(10):903–12.

13. Alcolea D, Vilaplana E, Pegueroles J, Montal V, Sanchez-Juan P, Gonzalez-Suarez A, Pozueta A, Rodriguez-Rodriguez E, Bartres-Faz D, Vidal-Pineiro D, et al. Relationship between cortical thickness and cerebrospinal fluid YKL-40 in predemential stages of Alzheimer's disease. Neurobiol Aging. 2015;36(6):2018–23.

14. Bonneh-Barkay D, Wang G, Starkey A, Hamilton RL, Wiley CA. In vivo CHI3L1 (YKL-40) expression in astrocytes in acute and chronic neurological diseases. J Neuroinflammation. 2010;7:34.

15. Antonell A, Mansilla A, Rami L, Llado A, Iranzo A, Olives J, Balasa M, Sanchez-Valle R, Molinuevo JL. Cerebrospinal fluid level of YKL-40 protein in preclinical and prodromal Alzheimer's disease. J Alzheimers Dis. 2014;42(3):901–8.

16. Hellwig K, Kvartsberg H, Portelius E, Andreasson U, Oberstein TJ, Lewczuk P, Blennow K, Kornhuber J, Maler JM, Zetterberg H, et al. Neurogranin and YKL-40: independent markers of synaptic degeneration and neuroinflammation in Alzheimer's disease. Alzheimers Res Ther. 2015;7(1):74.

17. Wennstrom M, Surova Y, Hall S, Nilsson C, Minthon L, Hansson O, Nielsen HM. The inflammatory marker YKL-40 is elevated in cerebrospinal fluid from patients with Alzheimer's but Not Parkinson's disease or dementia with lewy bodies. PLoS One. 2015;10(8):e0135458.

18. Cruchaga C, Kauwe JS, Harari O, Jin SC, Cai Y, Karch CM, Benitez BA, Jeng AT, Skorupa T, Carrell D, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for Alzheimer's disease. Neuron. 2013;78(2):256–68.

19. Cruchaga C, Kauwe JS, Nowotny P, Bales K, Pickering EH, Mayo K, Bertelsen S, Hinrichs A, Alzheimer's Disease Neuroimaging I, Fagan AM, et al. Cerebrospinal fluid APOE levels: an endophenotype for genetic studies for Alzheimer's disease. Hum Mol Genet. 2012;21(20):4558–71.

20. Nazem A, Sankowski R, Bacher M, Al-Abed Y. Rodent models of neuroinflammation for Alzheimer's disease. J Neuroinflammation. 2015;12:74.

21. Morris JC, Price JL. Pathologic correlates of nondemented aging, mild cognitive impairment, and early-stage Alzheimer's disease. J Mol Neurosci. 2001;17(2):101–18.

22. Fagan AM, Mintun MA, Mach RH, Lee SY, Dence CS, Shah AR, LaRossa GN, Spinner ML, Klunk WE, Mathis CA, et al. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. Ann Neurol. 2006;59(3):512–9.

23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9.

24. Browning BL, Browning SR. Efficient multilocus association testing for whole genome association studies using localized haplotype clustering. Genet Epidemiol. 2007;31(5):365–75.

25. R Core Team: R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing; 2015.

26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

27. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

28. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308–11.

29. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790–7.

30. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648–60.

31. Cruchaga C, Kauwe JS, Mayo K, Spiegel N, Bertelsen S, Nowotny P, Shah AR, Abraham R, Hollingworth P, Harold D, et al. SNPs associated with cerebrospinal fluid phospho-tau levels influence rate of decline in Alzheimer's disease. PLoS Genet. 2010;6(9):e1001101.

32. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. PLoS One. 2015;10(3):e0121945.

33. Meng XL, Rosenthal R, Rubin DB. Comparing correlated correlation-coefficients. Psychol Bull. 1992;111(1):172–5.

34. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. Plos Genetics. 2008;4(10):e1000214.

35. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010;464(7289):768–72.

36. Ober C, Tan Z, Sun Y, Possick JD, Pan L, Nicolae R, Radford S, Parry RR, Heinzmann A, Deichmann KA, et al. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. N Engl J Med. 2008;358(16):1682–91.

37. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452–8.

38. Naj AC, Jun G, Reitz C, Kunkle BW, Perry W, Park YS, Beecham GW, Rajbhandary RA, Hamilton-Nelson KL, Wang LS, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. JAMA Neurol. 2014;71(11):1394–404.

39. International Genomics of Alzheimer's Disease C. Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimers Dement. 2015;11(6):658–71.

40. Gomez-Isla T, Price JL, McKeel Jr DW, Morris JC, Growdon JH, Hyman BT. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. J Neurosci. 1996;16(14):4491–500.

41. Hulette CM, Welsh-Bohmer KA, Murray MG, Saunders AM, Mash DC, McIntyre LM. Neuropathological and neuropsychological changes in "normal" aging: evidence for preclinical Alzheimer disease in cognitively normal individuals. J Neuropathol Exp Neurol. 1998;57(12):1168–74.

42. Price JL, Ko AI, Wade MJ, Tsou SK, McKeel DW, Morris JC. Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. Arch Neurol. 2001;58(9):1395–402.

43. Markesbery WR, Schmitt FA, Kryscio RJ, Davis DG, Smith CD, Wekstein DR. Neuropathologic substrate of mild cognitive impairment. Arch Neurol. 2006;63(1):38–46.

44. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack Jr CR, Kaye J, Montine TJ, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7(3):280–92.

45. Aschenbrenner AJ, Balota DA, Fagan AM, Duchek JM, Benzinger TL, Morris JC. Alzheimer disease cerebrospinal fluid biomarkers moderate baseline differences and predict longitudinal change in attentional control and episodic memory composites in the adult children study. J Int Neuropsychol Soc. 2015;21(8):573–83.

46. Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, Marcus DS, Cairns NJ, Xie X, Blazey TM, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N Engl J Med. 2012;367(9):795–804.

47. Kauwe JS, Cruchaga C, Bertelsen S, Mayo K, Latu W, Nowotny P, Hinrichs AL, Fagan AM, Holtzman DM, Alzheimer's Disease Neuroimaging I, et al. Validating predicted biological effects of Alzheimer's disease associated SNPs using CSF biomarker levels. J Alzheimers Dis. 2010;21(3):833–42.