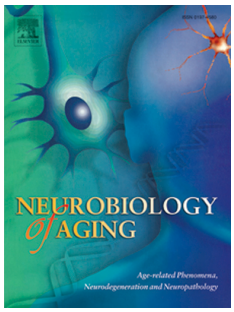
Accepted Manuscript



A *TOMM40* poly-T variant modulates gene expression and is associated with vocabulary ability and decline in non-pathological ageing

A. Payton, P. Sindrewicz, V. Pessoa, H. Platt, M. Horan, W. Ollier, V.J. Bubb, N. Pendleton, J.P. Quinn

PII: S0197-4580(15)00590-4

DOI: 10.1016/j.neurobiolaging.2015.11.017

Reference: NBA 9457

To appear in: *Neurobiology of Aging*

Received Date: 12 June 2015

Revised Date: 11 November 2015

Accepted Date: 25 November 2015

Please cite this article as: Payton, A, Sindrewicz, P., Pessoa, V., Platt, H., Horan, M., Ollier, W., Bubb, V.J., Pendleton, N., Quinn, J.P., A *TOMM40* poly-T variant modulates gene expression and is associated with vocabulary ability and decline in non-pathological ageing, *Neurobiology of Aging* (2016), doi: 10.1016/j.neurobiolaging.2015.11.017.

A *TOMM40* poly-T variant modulates gene expression and is associated with vocabulary ability and decline in non-pathological ageing

A Payton. (Corresponding author). Human Communication and Deafness, The University of Manchester, Manchester, UK, M13 9PT.
tony.payton@manchester.ac.uk. Tel: +44 (0)161 2751654

P. Sindrewicz. Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Crown Street, Liverpool, L69 3BX, UK.
psin86@liv.ac.uk

V. Pessoa. Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Crown Street, Liverpool, L69 3BX, UK.
V.Pessoa@liverpool.ac.uk

H. Platt. Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester, UK, M13 9PT.
hazel.platt@manchester.ac.uk

M. Horan. Centre for Clinical and Cognitive Neuroscience, Salford Royal NHS Hospital, The University of Manchester, Manchester, UK.
michael.horan@manchester.ac.uk

W. Ollier. Centre for Integrated Genomic Medical Research, The University of Manchester, Manchester, UK, M13 9PT.
bill.ollier@manchester.ac.uk

V.J. Bubb. Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Crown Street, Liverpool, L69 3BX, UK.
jillbubb@liverpool.ac.uk

N. Pendleton. Centre for Clinical and Cognitive Neuroscience, Salford Royal NHS Hospital, The University of Manchester, Manchester, UK.
neil.pendleton@manchester.ac.uk

J.P. Quinn. Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Crown Street, Liverpool, L69 3BX, UK.
jquinn@liverpool.ac.uk

Key words: TOMM40, cognition, aging, retrotransposon, APOE, expression

Conflicts of interest: There are no conflicts of interest

Abstract

The TOMM40-APOE locus has been associated with a number of age-related phenotypes in humans including non-pathological cognitive ageing, late-onset Alzheimer's disease and longevity. Here we investigate the influence of the TOMM40 intron 6 poly-T variant (rs10524523) on TOMM40 gene expression and cognitive abilities and decline in a cohort of 1613 community dwelling elderly volunteers who had been followed for changes in cognitive functioning over a period of 14 years (range 12-18 years). We showed that the shorter length poly-T variants were found to act as a repressor of luciferase gene expression in reporter gene constructs. Expression was reduced to approximately half of that observed for the very long variant. We further observed that the shorter poly-T variant was significantly associated with reduced vocabulary ability and a slower rate of vocabulary decline with age compared to the very long poly-T variants. No significant associations were observed for memory, fluid intelligence or processing speed, although the direction of effect, where the short variant was correlated with reduced ability and slower rate of decline was observed for all tests. Our results indicate that the poly-T variant has the ability to interact with transcription machinery and differentially modulate reporter gene expression and influence vocabulary ability and decline with age.

1. Introduction

As human life expectancy is increasing, age-related cognitive impairment is becoming a substantial problem due to its high social (Tannenbaum et al., 2005) and economic (Comas-Herrera et al., 2007) burdens. Consequently the identification of genetic risk factors for cognitive ageing is becoming a high priority research area that aims to prevent or slow down the progression of cognitive decline in later life. Cognitive ageing can be non-pathological (also known as normal ageing) (Davies et al., 2014) or caused by pathological diseases such as Alzheimer's disease (AD) (Bakulski et al., 2012). AD, cognitive ability and cognitive decline are highly heritable with the strongest and most consistent associations being reported within the Translocase of Outer Mitochondrial Membrane 40 Homolog and Apolipoprotein E (TOMM40-APOE) locus which occupies an 18kb region on chromosome 19. Twin and adoption studies suggest that additive genetic effects contribute over half of the adult population variance in intelligence (Deary et al., 2010) with between 0.40 to 0.51 of this variation being accounted for by common SNP markers (Davies et al., 2011). A lifetime (age 11 to old age) study of cognitive change

estimated a narrow-sense heritability of 0.24 with the genetic correlation between intelligence at age 11 and old age to be 0.62 (Davies et al., 2014).

TOMM40 codes for the central component of the translocase of the outer mitochondrial membrane (TOM) which is a multi-subunit complex made up of six TOM proteins (Humphries et al., 2005). TOM is a channel forming protein involved in the transport and sorting of proteins across the mitochondrial membrane. TOMM40 has been reported to be involved in the predisposition to AD and non-pathological cognitive decline possibly through a theory known as the "mitochondrial cascade hypothesis" which proposes a genetic contribution towards mitochondrial durability and function (Swerdlow et al., 2009). SNPs within the TOMM40 gene have been associated with plaque and vascular amyloid deposition (Valant et al., 2012). Several TOMM40 SNPs have been associated with non-pathological cognitive ageing (Davies et al., 2014; Greenbaum et al., 2014) but none of these have yet been shown to be functional and the possibility remains that they may be in linkage disequilibrium with another TOMM40-APOE variant. The TOMM40 gene is 2104 base pairs from APOE and they share the same linkage disequilibrium block (Lyall et al., 2013) making it difficult to identify discreet predisposing functional regions.

Retrotransposons have been less thoroughly studied in disease states than SNPs but have recently been associated with several conditions including Rett syndrome, Ataxia telangiectasia and schizophrenia (Erwin et al., 2014; Coufal et al., 2011). Retrotransposons, also known as mobile DNA elements, encompass up to 45% of the human genome. Their mobilization occurs via 'copy and paste' mechanisms that require the presence of an RNA intermediate which is subjected to reverse transcription and integration into DNA at a locus different than the original sequence (Raiz et al., 2012). They are mobilized in the germline genome during early stages of embryonic development and in somatic cell types of the brain, leading to both *de-novo* germline and somatic mutations (Erwin et al., 2014; Damert et al., 2009). Among them, non-long terminal repeat elements such as long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and SINE-VNTR-Alus (SVAs) constitute approximately 34% of the genome (Bekris et al., 2012). SINEs, such as Alus are very abundant between exons 6 and 7 of TOMM40. Particularly interesting is a block of six adjacent SINEs: AluSx, AluYc3, AluJb, AluJo, another AluJb and FLAM_A. FLAM_A encompasses the poly-T variant rs10524523, (henceforth '523'). This block also constitutes a human block of insertions with only limited homology seen in other primates but not observed in other non-primate species. '523' has been analysed according to its length and is reported as having three variants that are short (S) <20 T residues, long (L) 20-30 and very long (VL) ≥30Ts (Roses et al., 2010). Roses and colleagues have shown that the TOMM40 L variant is in strong LD with APOEε4 and that the S and VL variants are in strong LD with the APOEε3. The longer TOMM40 variants have been linked to a higher risk of

developing late onset Alzheimer's disease (LOAD) and an earlier age of the disease onset compared to the short variant in APOEε3 patients (Roses et al., 2010; Lutz et al., 2010). However, other studies have either not observed this association or reported that the VL variant was associated with lower risk of AD (Jun et al., 2012; Cruchaga et al., 2011). Another study also reported that the VL variant was associated with lower risk of AD and that the L variant was associated with increased risk (Maruszak et al., 2012). This same group also found that the L variant significantly reduced the chance of living to 100 years of age. The '523' S variant has also been associated with improved memory and executive function in several studies (Greenbaum et al., 2014; Caselli et al., 2012; Hayden et al., 2012; Johnson et al., 2011).

The location of these candidate genetic risk factors within non-coding regions of the TOMM40-APOE locus suggest the possible involvement of transcriptional and post-transcriptional regulation. Our group has previously demonstrated that SVAs can serve as transcriptional regulators in a classical reporter gene assay *in vitro* and *in vivo* (Savage et al., 2013; Savage et al., 2014). Here we hypothesise that '523' variants may regulate the expression of TOMM40 and have an influence on cognitive abilities and their decline (over a mean period of 14 years) with age in a cohort of 1613 non-demented community-dwelling older volunteers.

2. Material and Methods

2.1 Cohort

The University of Manchester Longitudinal Studies of Cognition in Normal Healthy Old Age documented longitudinal trajectories in cognitive function in a large sample of older adults in the North of England, UK; the Manchester and Newcastle Longitudinal Studies of Cognitive Ageing Cohorts (Rabbitt et al., 2004). Recruitment took place in Newcastle and Greater Manchester between 1983 and 1992. At the start of the study, 6063 volunteers were available, (1825 men, 4238 women), with a median age of 65 years (range 44 to 93 years). Over the period 1983 to 2003, two alternating batteries of cognitive tasks applied biennially were designed to measure fluid and crystallized aspects of intelligence. The studies have run for over 30 years and have collected a rich archive of demographic (including date and location of birth), lifestyle, health, cognitive and emotional health data. This programme of work continues at Manchester University and it will follow subjects to death. DNA was available on 1613 volunteers. Ethical approval for all projects was obtained from University of Manchester.

2.2 Cognitive measures

Principle components analysis has been used to derive variables for memory, vocabulary ability, fluid intelligence (novel problem solving) and processing speed for the Manchester and Newcastle cohorts. The vocabulary ability (crystallized intelligence) factor was generated using tests that comprised the Mill Hill and Wais vocabulary tests (Rabbitt et al., 2004; Raven, 1965). Cognitive tests used for fluid factors were the two parts of the Alice Heim test 4 and the four subtests of the Culture Fair Test (Heim, 1970; Cattell, 1949). Speed factors were derived from the Alphabet Coding Task and the Random Letters test (Rabbitt et al., 2004; Savage, 1984). Factors for memory were generated from free recall, propositions and spatial memory tests (Rabbitt et al., 2004).

In all cases men and women were examined and scores standardized separately. Longitudinal growth curve models were estimated that took the 0 point on the age scale as age 70, and measured variation about that in units of 10 years. Data were available for up to four occasions (collected at roughly 5 year intervals) of measurement that spanned an average of 14 years (range 12–18 years) for all volunteers. The fixed part of the models that described the overall pattern of change in the construct for the sample as a whole included linear and quadratic age terms. In addition, to account for possible artifactual improvement due to practice effects (a single step function), each test score was allowed to increment between the first and subsequent occasions in which each particular test was taken. Individual differences were accounted for by allowing subject-specific random effects for the intercept, describing the within-sample variation in performance at age 70, and for a linear growth term, describing the individual differences in the trend of cognitive decline over the period of follow-up. The random effects were assumed to be bivariate normally distributed and the models estimated in gllamm by maximum likelihood using adaptive quadrature (www.gllamm.org). Incomplete data observations were included under an assumption that the missing scores were missing at random, allowing attrition to be selective with respect to age, sex, and observed test scores. Subject-specific intercepts and linear trends were estimated using empirical Bayes' methods. The above fully describes the model in the circumstance where a single repeated cognitive test was available. Different tests were allowed different means, scales and error variances but the tests were assumed to reflect a single underlying construct with a common trend.

2.3 '523' Poly T/APOE genotyping and analysis

A polymorphic poly-T variant, rs10524523 ('523'), in the *TOMM40* gene was genotyped by the Roses laboratory using a method described previously (Linnertz et al., 2012). Briefly, *TOMM40* '523' poly-T genotypes were determined based on length variation. The '523'

region of each genomic DNA sample was PCR-amplified using a fluorescently labeled primer. Genotypes were determined on an ABI 3730 DNA Analyzer using GeneMapper, version 4.0 software (Applied Biosystems, Foster City, CA) for allelic size assessment. The '523' allele was assigned according to the length of the PCR product. The convention established by Roses *et al* for determining alleles was used: Short (S), <20 T residues; Long (L) - 20-29; Very Long (VL) ≧30 (Roses et al., 2010). Six genotypes were generated according to the length/combination of the alleles which consisted of: 1.S/S; 2. S/L; 3. S/VL; 4. L/L; 5. L/VL and 6. VL/VL. APOE genotyping was performed using Sequenom technology (Sequenom®, San Diego). This method has been described previously (Ghebranious et al., 2005). The Iplex™ assay was followed according to manufacturer's instructions (http://www.sequenom.com) using 30ng of DNA. Linear regression analysis of genotype and cognitive phenotypes were performed in Stata (StataCorp. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP). Our sample size of 1613 individuals has 89% power to detect a genetic effect size of 0.01 or greater assuming an additive model, stringent significance of 0.005 (two-tailed) and minor allele frequency of 0.13. The minor allele frequencies of the '523' alleles were S, 0.42; L, 0.13; VL, 0.45.

2.4 Bioinformatics

The genetic variation found between exons 6 and 7 of the TOMM40 gene as well as the local environment and conservation profiles were analysed using the UCSC (http://genome-euro.ucsc.edu/index.html) and ECR (http://ecrbrowser.dcode.org/) browsers. The SINEs block (chr19:45,401,766-45,403,217 on Hg19 release) within the intron 6 region of TOMM40 contains 6 Alu domains which all contain T-rich areas (Figure 1). The most 3' of the Alu elements is a FLAM_A which contains the '523' variants. All of the Alu domains contain long poly-T stretches or T-rich regions and are located on the negative strand, with the exception of AluYc3, which is found on the positive strand. The region also encompasses two short T-rich regions, classified as simple repeats, which are composed of 25 and 11 bp and are located on the positive strand.

2.5 Cloning and measuring luciferase expression directed by '523' reporter gene constructs

The TOMM40 fragment containing the SINE block was amplified from human genomic DNA (Promega) using Touchdown (TD)-PCR with Phusion Polymerase (NEB). The TD-PCR products were cloned into pGL3 promoter (pGL3p) and pGL3 control (pGL3c) vectors (Promega) and termed TOMM40/pGL3p and TOMM40/pGL3c respectively. This fragment

contained a 35T variant of rs10524523. Further truncated clones of the SINE were generated by PCR to address more specifically the Flam_A element containing rs10524523 ('523'). Three FLAM_A '523' variants generated; FLAM_A17T, FLAM_A34T and FLAM_A41T. Two of these FLAM_A 17T; FLAM_A 41T possessed identical sequence with the exception of the '523' region which reflected the length of the polymorphic T-stretch encompassing the SNP rs10524523 and comprised either a run of 17 or 41 Ts. Flam_A34T, in contrast, consists of 34 Ts in the '523' region; it is truncated at the 5' end relative to FLAM_A17T & 41T as outlined in Figure 1.

[FIGURE 1 HERE]

Figure 1: Reporter Gene Constructs spanning Hg19 chr19:45,401,688 – 45,403,254 showing polymorphism rs10524523 and the T-rich region.

The human neuroblastoma cell line SK-N-AS (ECACC# 94092302) was maintained in Dulbecco's Modified Eagle medium with glucose (4500 mg/L) (Sigma) supplemented with 10% (v/v) foetal bovine serum, 1% (v/v) MEM non-essential amino acid solution and 100U/ml Penicillin, 100µg/ml Streptomycin. Prior to transfection, cells were plated out in 24-well plates at an approximate density of $6x10^4$ cells per well. Transfection was performed approximately 24 hours post-plating with 1µg of the appropriate reporter gene construct and 10ng of internal control construct TK renilla using TurboFect (Thermo Scientific) following manufacturer's instructions and using serum-free media for transfection. 48 hours post-transfection cells were lysed and the Dual Luciferase Reporter Assay (Promega) was performed. Luminescence was measured using a Glomax 96 Microplate Luminometer (Promega). We performed duplicate experiments each with four repeats for each construct.

Statistical one-way ANOVA with Bonferroni correction for the three FLAM constructs and normal t-test for SINE block in both pGL3p and pGL3c were performed to determine the significance of the fold change of the reporter gene constructs over the pGL3p or pGL3c alone. Significance was scored as follows: *P<0.05, **P<0.01, ***P<0.001. Details regarding the cloning and expression methods can be obtained on request.

3. Results

3.1 Functional activity of the intron 6 SINE block in reporter gene analysis

The generated reporter gene constructs were assessed for their ability to direct reporter gene expression (luciferase) driven by a heterologous SV40 promoter in human neuroblastoma cell line SK-N-AS. As shown in Figure 2 A, the intron SINE containing the 34T variant significantly lowered the reporter gene expression when compared to the minimal promoter (pGL3p) alone ($p<0.001$ and $p<0.01$). In order to confirm the repression observed in the pGL3p vector, the fragment was also analyzed when inserted into pGL3c vector, (which exhibits 7-8 fold higher intrinsic activity than pGL3p). The results for pGL3c were consistent with the results observed for pGL3p and showed that the 17T SINE block significantly reduced the expression of luciferase when compared with the pGL3c vector alone ($p<0.01$ and $p<0.001$) (Figure 2 B).

[FIGURE 2 HERE]

Figure 2: SINE block showed the ability to down-regulate expression in a reporter gene construct. A) The average fold activity of the SINEs block of TOMM40 over the minimal SV40 promoter alone (pGL3p). **B)** The average fold activity of the SINEs block of TOMM40, containing 34T stretch, over the pGL3 control vector which in addition to SV40 promoter also contains an SV40 enhancer. T-tests were performed to measure the significance of the average fold and were scored as follows: *P<0.05, **P<0.01, ***P<0.001

3.2 Functional activity of the FLAM_A '523' variants in reporter gene analysis

To address if a functional difference was supported by the high and low T variants we compared FLAM_A17T and 41T which only vary in the number of Ts in each construct (Figure 2). The FLAM_A17T showed the greatest ability to alter the reporter gene activity and significantly repressed the levels of reporter gene expression, when compared to the minimal promoter alone ($p<0.001$). In contrast, FLAM_A 41T had no effect on transcription. Significant reduction in luciferase activity was also observed for FLAM_A34T (p value, <0.01 and <0.05), however, the repression was not as great as that observed for FLAM_A 17T. Our results would be consistent with a model in which the shorter the '523' variant i.e. the lower the number of thymine residues, within FLAM_A region, the greater its ability to modulate transcription.

[FIGURE 3 HERE]

Figure 3: Distinct polymorphic T runs showed the ability to affect expression in a reporter gene construct. The average fold activity of the three variants over the minimal SV40 promoter alone (pGL3p). Significant difference in fold activity between the fragments: One-way ANOVA with Bonferroni correction was performed and was scored as follows *P<0.05, **P<0.01, ***P<0.001

3.3 Regression analysis of '523' and memory ability and decline with age

'523' genotype frequencies were: S/S (n=261, 16.2%); S/L (n=213, 13.2%); S/VL (n=626, 38.8%); L/L (n=27, 1.7%); L/VL (n=162, 10.0%) and VL/VL (n=324, 20.1%). These frequencies were similar to those reported in another UK cohort (Lyall et al., 2013). Regression analysis suggested that vocabulary ability (a measure of crystallized intelligence) was the most strongly influenced by '523' with the short variant associated with a lower ability. This was most evident (p-value, 0.007) when volunteers who were homozygous for the short allele (n=261) were compared against those homozygous for the very long allele (n=324) (mean residual scores for vocabulary ability at initial testing point S/S, -0.130; VL/VL, 0.090). Significance was also observed for vocabulary ability at initial testing point (p-value, 0.05) when all volunteers underwent regression analysis in the genotype order S/S, S/L, S/VL, L/L, L/VL, VL/VL. It was also found that the rate of vocabulary decline was faster for those who were homozygous for the very long allele compared to those who were homozygous for the short allele, although the significance was marginal (p-value, 0.05). No significant associations were observed for memory, fluid intelligence or processing speed, although the direction of effect, where the S variant was correlated with reduced ability and slower rate of decline, was observed for all tests.

To determine if the association with the TOMM40 short allele and vocabulary ability was being influenced by APOE we analysed APOE (ε2, ε3 and ε4) genotypes in 1539 participants against the vocabulary measure. The allele frequencies were 8.3% (ε2); 78.1% (ε3) and 13.6% (ε4), which is similar to those reported in other studies (Corbo RM and Scacchi R, 1999). Analysis was performed on three genotype groups (volunteers with no ε4, one ε4 and two ε4 alleles) using an additive model. No association between APOE genotype was observed for vocabulary ability (p-value, 0.576) or its decline over time (p-value, 0.464). The '523' homozygous S and VL variants were also investigated in homozygous APOEε3 volunteers but this failed to strengthen significance.

We considered this study an attempt to replicate previous findings and therefore did not adjust for multiple testing.

4. Discussion

We have shown that a primate specific retrotransposon domain containing the genetic variant '523' within the TOMM40 gene can serve as a differential regulator of gene expression based on genotype of the T repeat and has an influence on crystallized intelligence and decline in non-pathological ageing. Specifically, we report that the short poly-T variant was associated with reduced reporter gene expression, reduced vocabulary ability and a slower rate of vocabulary decline in a cohort of 1613 volunteers. The association between the TOMM40-APOE locus and cognitive impairment have been amongst the most extensively reported and consistent findings in the field of complex genetic diseases/traits. The functional relevance of these genes in the aetiology of cognitive impairment adds weight to their involvement. ApoE not only serves in the transport and breakdown of lipoproteins but also influences amyloid-β clearance and competes with amyloid-β for cellular uptake via apoE receptors (Kanekiyo et al., 2014). TOMM40 encodes a translocase of outer mitochondrial membrane 40, a subunit of a larger TOM complex responsible for the translocation of newly synthesized unfolded proteins into the mitochondrial intermembrane space (Humphries et al., 2005). Mitochondria are essential for cell survival and their dysfunction has been linked to pathological and non-pathological cognitive ageing (Bishop et al., 2010).

Until recently the focus of cognitive/dementia genetic research has relied heavily upon SNP data with little emphasis on other variant types. Here we have focused on the function of noncoding DNA associated with SNPs highlighted in our previous association studies for their ability to exert regulatory properties at the locus. Of particular interest at this locus are the properties of the multiple Alu elements which constitute a human specific domain with limited homology seen in other primates. Retrotransposon domains in noncoding regions of the genome can introduce novel splice sites, polyadenylation signals, promoters and regulatory domains that can reorganize gene expression and build new transcription modules (Erwin et al., 2014). The position of retrotransposons within introns and other local environments enhances their ability to alter local architecture and consequently, modulate genome regulation (Quinn et al., 2014). We have previously demonstrated that primate specific SVA retrotransposons in the promoter of the neurodegenerative candidate genes, PARK7 and FUS, similarly to our present study, are both polymorphic and modulate reporter gene expression (Savage et al., 2013; Savage et al., 2014).

Our data (Quinn et al., 2014) indicates that the SINE domain or the FLAM_A domain could mechanistically modulate vocabulary ability and decline by altering TOMM40 gene expression perhaps in response to an additional environmental challenge. However, these

regions could also be hot spots for retrotransposition itself so affecting the function at the locus by multiple mechanisms. With the exception of early embryo development and some malignancies, retrotransposition is suppressed by epigenetic and post-transcriptional modifications in somatic cells, although their mobilization has been reported in the adult brain. Baillie and colleagues have detected somatic L1 retrotransposition of three active retrotransposon families: LINE-1, Alu and SVA, which resulted in alteration of the genetic landscape of the human brain (Baillie et al., 2011). Consistent with CNS mobilisation, next-generation sequencing studies have shown that there is a bias for insertion sites of somatic L1 retrotransposition in neurons of schizophrenia brains and that these events could have an etiological role (Bundo et al., 2014). More recently dysregulated retrotransposon function has been observed in frontal frontotemporal lobar degeneration which the authors' link to activity of the neurodegenerative risk gene TDP-43 (Li et al., 2012). Furthermore, many studies including our own have indicated age as a parameter affecting human cognition. This is consistent with the observation that in the model organism Drosophila, increased activity of transposable elements in the brain contribute to age-dependent loss of neuronal function (Li et al., 2012).

The TOMM40 '523' alleles have a commonly used categorization where they are divided into three groups depending on the length of the T homopolymer: T≤19 (S), Ts=20-29 (L) and T≥30 (VL). The original study by Roses and colleagues reported that longer poly-T length (in particular, T≥27) was associated with earlier age of AD onset and risk compared with individuals with shorter T-homopolymers in people of European descent (Roses et al., 2010). This study was conducted using three independent cohorts with numbers of cases ranging from 34 to 83 and numbers of controls ranging from 31 to 67. In contrast a larger study of 1594 AD cases and 1190 controls (all European descent) found no association between the VL variant in homozygous APOEε3 individuals and age of AD onset but did report that the VL variant was associated with lower AD risk (Cruchaga et al., 2011). This finding was supported by an independent group who investigated the '523' variant using 414 AD cases, 173 centenarians and 305 neurologically healthy controls (Maruszak et al., 2012). This study also reported that the L variant was associated with increased risk of AD and a reduced likelihood of living to 100 years of age. Finally, a meta-analysis study of 11840 AD patients and 10931 controls found no association between '523' variants and AD risk (Jun et al., 2012). The differences in results observed between the above studies may be explained by variation in study design which has been discussed in detail elsewhere (Roses et al., 2013). Accuracy in assigning poly-T length, generating meaningful prognostic haplotypes between TOMM40 and APOE, variation caused by age-dependent risk of AD and

differences in AD diagnosis methodology were some of the reasons highlighted by Roses and colleagues.

The association between the '523' variant and cognition has been a little more consistent than the AD reports, although the cohort sizes have been small. Johnson *et al* found that those homozygous for the VL variant (n=35) scored lower on a test of primary retrieval from a verbal list learning task than S homozygous individuals (n=38) (mean age of cohort, 55 years) (Johnson et al., 2011). Two further studies by Hayden (Hayden et al., 2012) and Greenbaum (Greenbaum et al., 2014) (n=127, mean age 80.6 and n=331, ≥65 years, respectively) observed that those homozygous for the S variant performed better on tests of memory and executive function, although in these studies the results either did not reach significance or were only marginally significant. A larger study of 639 cognitively normal individuals aged 21-97 years found that those homozygous for the VL variant had a flattened test-retest improvement that was only observed in those aged under 60 years (Caselli et al., 2012).

Our results showed a faster rate of cognitive decline in volunteers who carried the VL variant which was marginally significant for vocabulary ability. In contrast to other cognitive studies who measured intelligence at a single time point using neurologically healthy participants we observed that the S variant was associated with reduced scores which reached significance with vocabulary ability but not memory, fluid intelligence or processing speed. The reasons for our observed differences may be related to population stratification as the Greenbaum study investigated TOMM40 in an Israeli Jewish population and the Johnson study did not mention their population ethnicity. The Hayden study was predominantly Caucasian, but their results failed to reach significance after correction for multiple testing. Statistical power is unlikely to be an issue in our analysis as our cohort size of 1613 was larger than the other studies on cognitively healthy individuals combined (89% power to detect a 1% genetic effect size at a stringent p-value of 0.005). The cognitive tests used by the different groups also varied. Hayden used tasks that measured attention, processing speed, memory, executive and visuospatial function whilst the Greenbaum study used a factor analytical approach as we did. However, neither used a test of crystallized intelligence and thus we cannot compare with our vocabulary result.

TOMM40 is located 2104 base pairs from APOE with significant LD between the two genes (Lyall et al., 2013). This strong LD makes it difficult to determine the causative region(s) and therefore we further investigated the influence of both APOE ε3 and ε4 on cognitive ability and decline. Phylogenetic mapping analysis has shown that the APOEε4 allele is linked to the TOMM40 L allele and APOEε3 is linked to either the S or L alleles (Roses et al., 2010). We observed no association between APOEε4 and vocabulary ability (both cross-sectional and longitudinal measures). This observation is likely because of its

link with the TOMM40 L allele and not with the S or LV alleles which were the strongest associated alleles in this study. The '523' S and VL variants were also investigated in homozygous APOEε3 volunteers but this failed to strengthen significance. Our results suggest that APOE ε3 and ε4 do not play a role in vocabulary ability or its decline with age.

5. Conclusion

Retrotransposons have the potential to modulate gene expression via various mechanisms. The presence of a primate specific mobile DNA element within the TOMM40-APOE locus and its potential regulatory nature supports the importance of this key locus in higher cognitive traits and consequently cognitive dysfunction. We demonstrated the ability of '523' polymorphism to affect transcription within a luciferase reporter gene construct and to influence vocabulary in non-pathological elderly volunteers.

Acknowledgements

This work was funded by the Medical Research Council

References

Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature 2011; 479:534-537.

Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, et al. Genome-Wide DNA Methylation Differences Between Late-Onset Alzheimer's Disease and Cognitively Normal Controls in Human Frontal Cortex. J Alzheimers Dis 2012; 29:571-588.

Bekris LM, Lutz F, Yu CE. Functional analysis of APOE locus genetic variation implicates regional enhancers in the regulation of both TOMM40 and APOE. J. Hum Genet 2012; 57:18-25.

Bishop NA, Lu T, Yankner BA. Neural mechanisms of ageing and cognitive decline. Nature 2010; 464:529-35.

Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, et al. Increased L1 retrotransposition in the neuronal genome in schizophrenia. Neuron 2014; 81:306-313.

Caselli RJ, Dueck AC, Huentelman MJ, Lutz MW, Saunders AM, Reiman EM, et al. Longitudinal modeling of cognitive aging and the TOMM40 effect. Alzheimers Dement 2012; 8:490–495.

Cattell R. Culture Free Intelligence Test, Scale 1, Handbook. Champaign [Illinois]: Institute of Personality and Ability; 1949.

Comas-Herrera A, Wittenberg R, Pickard L, Knapp M. Cognitive impairment in older people: future demand for long-term care services and the associated costs. Int J Geriatr Psych 2007; 22:1037-1045.

Corbo RM, Scacchi R. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? Ann Hum Genet 1999; 63:301-10.

Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MC, Muotri AR, Mu Y, et al. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. Proc Natl Acad Sci U S A 2011; 108:20382-20387.

Cruchaga C, Nowotny P, Kauwe JS, Ridge PG, Mayo K, Bertelsen S, et al. Association and expression analyses with single-nucleotide polymorphisms in TOMM40 in Alzheimer disease. Arch Neurol 2011; 68:1013–1019.

Damert A, Raiz J, Horn AV, Lower J, Wang H, Xing J, et al. 5'-Transducing SVA retrotransposon groups spread efficiently throughout the human genome. Genome research 2009; 19:992-2008.

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 2011; 16:996-1005.

Davies G, Harris SE, Reynolds CA, Payton A, Knight HM, Liewald DC, et al. A genome-wide association study implicates the APOE locus in nonpathological cognitive ageing. Mol Psychiatry 2014; 19:76-87.

Deary IJ, Penke L, Johnson, W. The neuroscience of human intelligence differences. Nat Rev Neurosci 2010; 11:201-211.

Erwin JA, Marchetto MC, Gage FH. Mobile DNA elements in the generation of diversity and complexity in the brain. Nat Rev Neurosci 2014; 15:497-506.

Ghebranious N, Ivacic L, Mallum J, Dokken C. Detection of ApoE E2, E3 and E4 alleles using MALDI-TOF mass spectrometry and the homogeneous mass-extend technology. Nucleic Acids Res 2005; 33:e149.

Greenbaum L, Springer RR, Lutz MW, Heymann A, Lubitz I, Cooper I, et al. The TOMM40 poly-T rs10524523 variant is associated with cognitive performance among non-demented elderly with type 2 diabetes. Eur. Neuropsychopharmacol 2014; 24:1492-9.

Hayden KM, McEvoy JM, Linnertz C, Attix D, Kuchibhatla M, Saunders AM, et al. A homopolymer polymorphism in the TOMM40 gene contributes to cognitive performance in aging. Alzheimers Dement 2012; 8:381–388.

Heim AW. The AH 4 group of intelligence tests. Windsor: NFER-Nelson; 1970.

Humphries AD, Streimann IC, Stojanovski D, Johnston AJ, Yano M, Hoogenraad NJ, et al. Dissection of the Mitochondrial Import and Assembly Pathway for Human Tom40. J Biol Chem 2005; 280:11535–11543.

Johnson SC, LaRue A, Hermann BP, Xu G, Koscik RL, Jonaitis EM, et al. The effect of TOMM40 poly-T length on gray matter volume and cognition in middle-aged persons with APOE epsilon3/epsilon3 genotype. Alzheimers Dement 2011; 7:456–465.

Jun G, Vardarajan BN, Buros J, Yu CE, Hawk MV, Dombroski BA, et al. Comprehensive search for Alzheimer disease susceptibility loci in the APOE region. Arch Neurol 2012; 69:1270–1279.

Kanekiyo T, Xu H, Bu G. ApoE and Aβ in Alzheimer's disease: accidental encounters or partners? Neuron 2014; 81:740-54.

Li W, Jin, Y, Prazak L, Hammell M, Dubnau J. Transposable elements in TDP-43-mediated neurodegenerative disorders. PLoS ONE 2012; 7:e44099.

Li W, Prazak L, Chatterjee N, Gruninger S, Krug L, Theodorou D, et al. Activation of transposable elements during aging and neuronal decline in Drosophila. Nature neuroscience 2013; 16:529-531.

Linnertz C, Saunders AM, Lutz MW, Crenshaw DM, Grossman I, Burns DK, et al. Characterization of the Poly-T Variant in the TOMM40 Gene in Diverse Populations. PLoS ONE 2012; 7:e30994.

Lyall DM, Royle NA, Harris SE, Bastin ME, Maniega SM, Murray C, et al. Alzheimer's disease susceptibility genes APOE and TOMM40, and hippocampal volumes in the Lothian birth cohort 1936. PLoS One 2012; 8:e80513.

Lutz MW, Crenshaw DG, Saunders AM, Roses AD. Genetic variation at a single locus and age of onset for Alzheimer's disease. Alzheimers Dement 2010; 6:125-131.

Maruszak A, Pepłońska B, Safranow K, Chodakowska-Żebrowska M, Barcikowska M, Zekanowski C. TOMM40 rs10524523 polymorphism's role in late-onset Alzheimer's disease and in longevity. J Alzheimers Dis 2012; 28:309-22.

Quinn JP, Bubb VJ. SVA retrotransposons as modulators of gene expression. Mob Genet Elements 2014; 4:e32102.

Rabbitt PMA, McInnes L, Diggle P, Holland F, Bent N, Abson V, et al. The University of Manchester Longitudinal Study of Cognition in Normal Healthy Old Age, 1983 through 2003. Aging Neuropsychology and Cognition 2004; 11:245–279.

Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res 2012; 40:1666-1683.

Raven JC. The Mill Hill Vocabulary Scale. London: HK Lewis; 1965.

Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J 2010; 10:375 - 384.

Roses AD, Lutz MW, Amrine-Madsen H, Saunders AM, Crenshaw DG, Sundseth SS, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. Pharmacogenomics J 2010; 10:375-384.

Roses AD, Lutz MW, Crenshaw DG, Grossman I, Saunders AM, Gottschalk WK. TOMM40 and APOE: Requirements for replication studies of association with age of disease onset and enrichment of a clinical trial. Alzheimers Dement 2013; 9:132-6.

Savage AL, Bubb VJ, Breen G, Quinn JP. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. BMC Evol Biol 2013; 13:101.

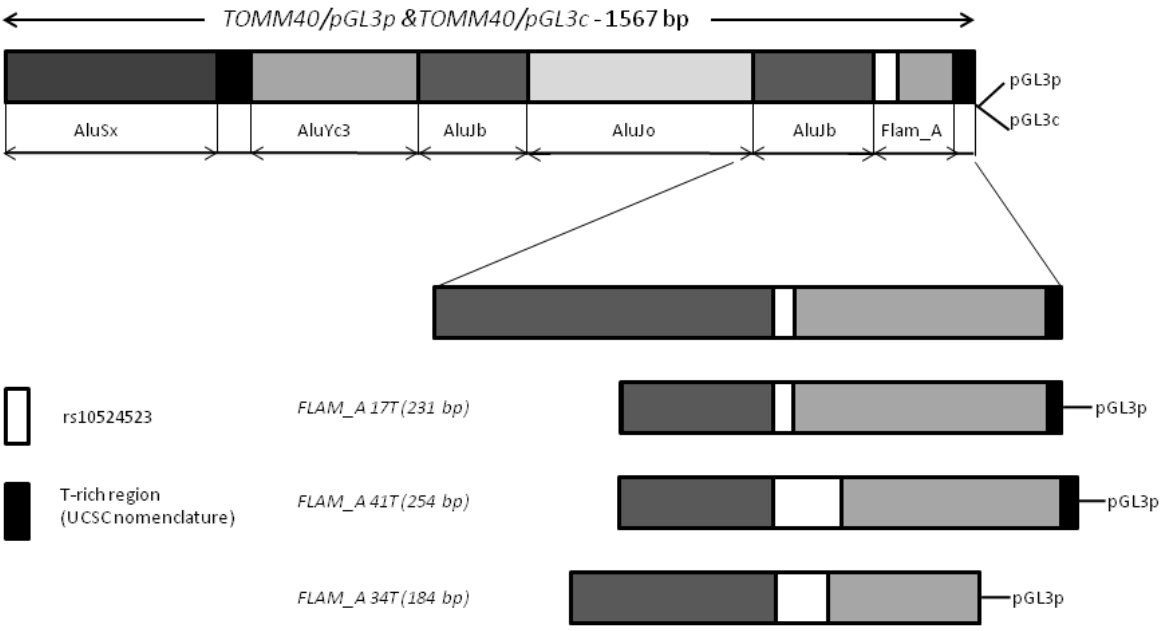
Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ, et al. An Evaluation of a SVA Retrotransposon in the FUS Promoter as a Transcriptional Regulator and Its Association to ALS. PloS one 2014; 9:e90833.

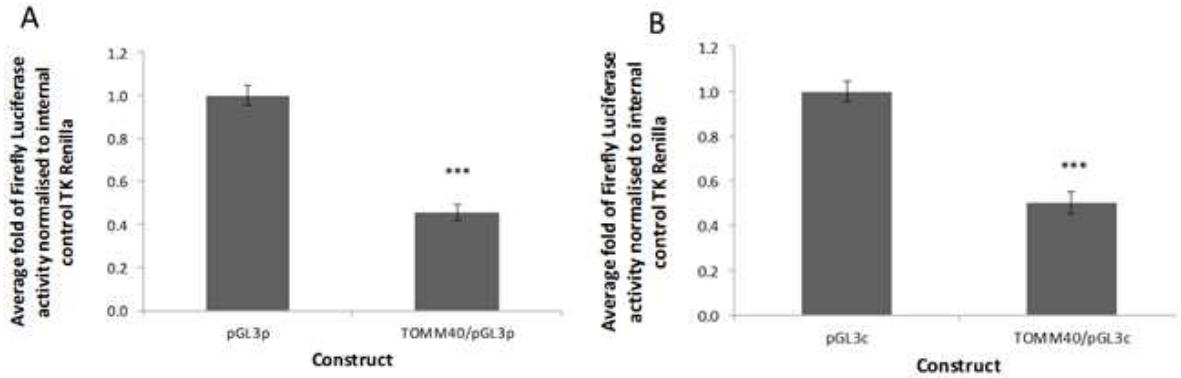
Savage RD. Alphabet Coding Task-15. Murdock University: Perth, W. Australia; 1984.

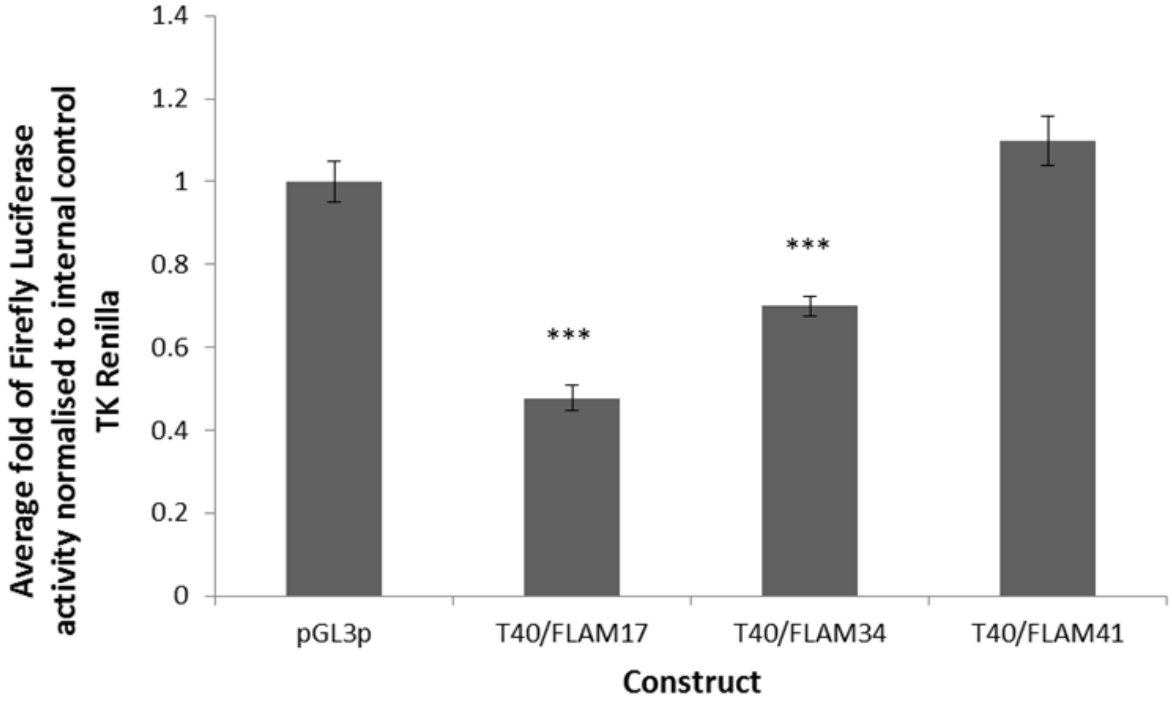
Swerdlow RH, Khan SM. The Alzheimer's disease mitochondrial cascade hypothesis: an update. Exp Neurol 2009; 218:308-15.

Tannenbaum C, Mayo N, Ducharme F. Older women's health priorities and perceptions of care delivery: results of the WOW health survey. Can Med Assoc J 2005; 173:153-159.

Valant V, Keenan BT, Anderson CD, Shulman JM, Devan WJ, Ayres AM, et al. TOMM40 in Cerebral Amyloid Angiopathy Related Intracerebral Hemorrhage: Comparative Genetic Analysis with Alzheimer's Disease. Transl Stroke Res 2012; 3:102-12.







Highlights

- We investigate the influence of the TOMM40 intron 6 poly-T variant (rs10524523) on TOMM40 gene expression and cognitive abilities and decline).
- We showed that the shorter length poly-T variants were found to act as a repressor of luciferase gene expression in reporter gene constructs.
- We further observed that the shorter poly-T variant was significantly associated with reduced vocabulary ability and a slower rate of vocabulary decline with age compared to the very long poly-T variants.