Queensland University of Technology
Brisbane Australia

This may be the author's version of a work that was submitted/accepted for publication in the following source:

He, Ji, Tang, Lu, Benyamin, Beben, Shah, Sonia, Hemani, Gibran, Liu, Rong, Ye, Shan, Liu, Xiaolu, Ma, Yan, Zhang, Huagang, Cremin, Katie, Leo, Paul, Wray, Naomi, Visscher, Peter, Xu, Huji, Brown, Matthew, Bartlett, Perry, Mangelsdorf, Marie, & Fan, Dongsheng
(2015)
C9orf72 hexanucleotide repeat expansions in Chinese sporadic amyotrophic lateral sclerosis.
Neurobiology of Aging, *36*(9), 2660.e1-2660.e8.

This file was downloaded from: https://eprints.qut.edu.au/110832/

© Consult author(s) regarding copyright matters

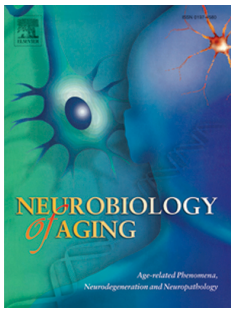
This work is covered by copyright. Unless the document is being made available under a Creative Commons Licence, you must assume that re-use is limited to personal use and that permission from the copyright owner must be obtained for all other uses. If the document is available under a Creative Commons License (or other specified license) then refer to the Licence for details of permitted re-use. It is a condition of access that users recognise and abide by the legal requirements associated with these rights. If you believe that this work infringes copyright please provide details by email to qut.copyright@qut.edu.au

License: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Notice: *Please note that this document may not be the Version of Record (i.e. published version) of the work. Author manuscript versions (as Submitted for peer review or as Accepted for publication after peer review) can be identified by an absence of publisher branding and/or typeset appearance. If there is any doubt, please refer to the published source.*

https://doi.org/10.1016/j.neurobiolaging.2015.06.002

Accepted Manuscript



*C9orf72*Hexanucleotide Repeat Expansions In Chinese Sporadic Amyotrophic Lateral Sclerosis

Ji He, Lu Tang, Beben Benyaminx, Sonia Shah, Gib Hemani, Rong Liu, Shan Ye, Xiaolu Liu, Yan Ma, Huagang Zhang, Katie Cremin, Paul Leo, Naomi R. Wray, Peter M. Visscher, Huji Xu, Matthew A. Brown, Perry F. Bartlett, Marie Mangelsdorf, Dongsheng Fan

PII: S0197-4580(15)00313-9

DOI: 10.1016/j.neurobiolaging.2015.06.002

Reference: NBA 9289

To appear in: *Neurobiology of Aging*

Received Date: 16 December 2014

Revised Date: 5 May 2015

Accepted Date: 1 June 2015

Please cite this article as: He, J., Tang, L., Benyaminx, B., Shah, S., Hemani, G., Liu, R., Ye, S., Liu, X., Ma, Y., Zhang, H., Cremin, K., Leo, P., Wray, N.R., Visscher, P.M., Xu, H., Brown, M.A., Bartlett, P.F., Mangelsdorf, M., Fan, D., *C9orf72*Hexanucleotide Repeat Expansions In Chinese Sporadic Amyotrophic Lateral Sclerosis, *Neurobiology of Aging* (2015), doi: 10.1016/j.neurobiolaging.2015.06.002.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

TITLE:

*C9orf72*Hexanucleotide Repeat Expansions In Chinese Sporadic Amyotrophic Lateral Sclerosis

Ji He,[*,§,†,2] Lu Tang,[*,2] Beben Benyamin,[§] Sonia Shah,[§] GibHemani,[§,†] Rong Liu,[*] Shan Ye,[*] Xiaolu Liu,[*] Yan Ma,[*] Huagang Zhang,[*] Katie Cremin,[†] Paul Leo,[†] Naomi R. Wray,[§] Peter M.Visscher,[§,†] HujiXu, [‡], Matthew A. Brown,[†] Perry F. Bartlett,[§] Marie Mangelsdorf[§] and Dongsheng Fan[*,1]

[2]**Contributed equally in the article**

[*]Department of Neurology, Peking University Third Hospital, Beijing, China

[§]Queensland Brain Institute, University of Queensland, St Lucia, Queensland, Australia, 4072

[†]University of Queensland Diamantina Institute, Translational Research Institute, Woolloongabba, Queensland, Australia, 4102

[‡]Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai, China

[1]**Corresponding author:** Dongsheng Fan

Address: 49 North Garden Road, Haidian District, Beijing 100191, China

Phone: + 86 10 82266699

Fax: + 86 10 62017700

Email: dsfan2010@aliyun.com

ACCEPTED MANUSCRIPT

ABSTRACT

A hexanucleotide repeat expansion (HRE) in the *C9orf72* gene has been identified as the most common mutation in amyotrophic lateral sclerosis (ALS) among Caucasian populations. We sought to comprehensively evaluate genetic and epigenetic variants of *C9orf72* and the contribution of the HRE in Chinese ALS cases. We performed fragment-length and repeat-primed PCR to determine GGGGCC copy number and expansion within the *C9orf72*gene in 1,092 sporadic ALS (sALS) and 1,062 controls from China. We performed haplotype analysis of 23 SNPs within and surrounding *C9orf72*. The *C9orf72* HRE was found in three sALS patients (0.3%) but not in control subjects (p = 0.25). For two of the cases with the HRE, genotypes of 8 SNPs flanking the HRE were inconsistent with the haplotype reported to be strongly associated with ALS in Caucasian populations. For these two individuals we found hypermethylation of the CpGisland upstream of the repeat, an observation not detected in other sALS patients ($p < 10^{-8}$) or controls. The detailed analysis of the*C9orf72* locus in a large cohort of Chinese samples provides robust evidence that may not be consistent with a single Caucasian founder event. Both the Caucasian and Chinese haplotypes associated with HRE were highly associated with repeat lengths greater than 8 repeats implying that both haplotypes may confer instability of repeat length.

Keywords: amyotrophic lateral sclerosis; C9orf72 gene; hexanucleotide repeat expansion; CpG methylation; Chinese population

ACCEPTED MANUSCRIPT

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder characterized by loss of upper and lower motor neurons. The disorder occurs as sporadic ALS (sALS) in most cases, while 5-10% of cases are familial (fALS).

Since *SOD1* mutations were linked to fALS in 1993 (Rosen, et al., 1993), a number of causative genes have been identified. A large hexanucleotide (GGGGCC) repeat expansion (HRE) in the first intron of the *C9orf72* gene has been identified as the most common mutation detected in ALS and frontotemporal dementia (FTD) patients in Caucasian populations (DeJesus-Hernandez, et al., 2011,Millecamps, et al., 2012,Renton, et al., 2011). Mutations in these genes may also be found in sporadic ALS, but other than*C9orf72*HRE each accounts for less than 1% of cases (Renton, et al., 2014).

The *C9orf72*HRE mutation has been comprehensively screened and assessed in Caucasian populations. The HRE frequency amongst sALS cases ranges from 3.2%-21% in different populations (Beck, et al., 2013,Galimberti, et al., 2014,Majounie, et al., 2012,Ratti, et al., 2012,Renton, et al., 2011). However in Asian sALS, it was found to be rare in Japanese cohorts (Ishiura, et al., 2012,Konno, et al., 2012,Majounie, et al., 2012,Ogaki, et al., 2012) and absent in cases from mainland China (Jiao, et al., 2014,Liu, et al., 2013,Majounie, et al., 2012,Zou, et al., 2013). Additionally, genotypes consistent with a common founder risk haplotype have been reported in most *C9orf72* HRE-carrying ALS cases in diverse populations, and it has

ACCEPTED MANUSCRIPT

been proposed that the HRE arose in a single common founder ~1,500 years ago (Majounie, et al., 2012).

The function of the protein encoded by *C9orf72* remains unclear, however there are several hypotheses about functional consequences of the HRE including toxicity of the transcribed repeat, toxicity of protein dipeptides translated from the transcribed repeat, or loss of function (DeJesus-Hernandez, et al., 2011,Donnelly, et al., 2013,Gijselinck, et al., 2012,Mori, et al., 2013). Abnormal methylation of the CpGisland upstream of the repeat was found in 73% of HRE carriers and was associated with downregulation of *C9orf72* transcripts (Xi, et al., 2013).

To date, large cohorts of Caucasian ALS cases have been assessed for the *C9orf72* HRE, however comparatively few studies of small cohorts of Asian cases have been conducted. We aimed to more accurately determine the frequency of the *C9orf72* HRE in a large cohort of Chinese sALS subjects, and investigate its relationship with SNP haplotypes and nearby CpG methylation.

MATERIALS AND METHODS

Participants

Patients attending the ALS specialty clinic at the Department of Neurology of the Peking University Third Hospital, Beijing, China, from 2003-2013 were recruited. Patients in the case cohort were diagnosed with ALS according to El Escorial revised criteria (Brooks, et al., 2000) by a neurologist specializing in ALS. In the

ACCEPTED MANUSCRIPT

present analysis, only sALS cases were included based on self-report and clinical interview. The cohort consisted of 65% males, while mean age of onset of sALS was 49.7 ± 12.2 years. The proportion of patients with bulbar onset was 17%. The control samples were from individuals who attended the same hospital and Shanghai Changzheng Hospital, and who had no medical or family history of neurological disorders. All patients and control subjects were from mainland China and of Chinese origin. Control cohorts were age- and sex-matched neurologically healthy individuals. Table 1 describes the demographic characteristics of the cases and controls. All patients and controls provided written informed consent for the clinical and genetic studies during their visit to the neurologist. The Peking University Third Hospital and Changzheng Hospital ethics committees approved the collection of DNA samples from case and control subjects.

Repeat-Primed PCR

Genomic DNA from 1,092 patients with sALS and 1,062 controls subjects was extracted from blood using standard protocols. Two-step polymerase chain reaction (PCR) was performed to detect *C9orf72* HRE as previously described (DeJesus-Hernandez, et al., 2011). Briefly, fluorescent fragment-length analysis was performed with genotyping primers. The samples with a homozygous peak pattern were analyzed by fluorescent repeat-primed PCR to identify HREs. The HRE was defined as repeat number greater than 30 (Renton, et al., 2011) indicated by the typical “saw-tooth” pattern seen by repeat-primed PCR.

Haplotype SNP Analysis, Principal-component Analysis and Validation

ACCEPTED MANUSCRIPT

Genome-wide genotyping was performed using IlluminaHumanOmniZhongHua DNA analysis arrays (900,015 SNPs/individual) and manufacturer's protocols. Bead intensity data was processed and normalized for each sample, and genotypes extracted in Genome Studio (Illumina). In order to predict haplotypes with greater accuracy a larger cohort of 3,115 control samples was used for haplotype analysis. 939 cases and 2,850 controls survived after QC with 784,352 SNPs/individual. SNP genotypes from HRE carriers were validated by direct sequencing. The SNPs tested included 20 from the consensus risk founder haplotype in several ALS populations (K. Mok, et al., 2012) and three additional SNPs flanking the repeat (previously included by (Smith, et al., 2013)). Haplotype analysis of the 23 SNPs within and surrounding *C9orf72* was performed by direct investigation and data imputation (using IMPUTE2)(Howie, et al., 2011), with the 1,000 Genomes Project reference panel, Phase I version 3 to determine whether the Chinese patients carried the founder haplotype associated with a risk for ALS. Haplotype reconstruction and frequency estimations were performed using PHASE (version 2.1.1, Chicago, IL) (Stephens, et al., 2001). Principal component analysis (PCA) was conducted by using the software package EIGENSTRAT(Price, et al., 2006). Detailed methods are provided in Tables S1and S3in the supplementary materials.

DNA methylation analysis

Genome-wide methylation levels in DNA extracted from blood of 461 sALS cases and 198 controls (a subset of those with genome-wide genotyping) were measured using the HumanMethylation450BeadChip (Illumina) following manufacturer's

ACCEPTED MANUSCRIPT

protocols. Bead intensity data was background-corrected and normalized using internal controls, and methylation beta-values extracted using the minfi package in R CRAN (http://cran.r-project.org/). Low quality samples and probes were removed. To assess the level of DNA methylation in the *C9orf72*region, beta-values for all eight probes annotated to the *C9orf72* region (based on Illumina probe annotation) were extracted, which detected two predicted CpG islands (From UCSC database) closely flanking the GGGGCC repeat.

RESULTS

C9orf72 genotypes in Chinese ALS cases and controls

The HRE was identified in three Chinese patients with sALS (Ch_MND_11134; Ch_MND_8446; Ch_MND_9059) of the 1,092 (0.3%) cases genotyped, and was not found in controls (p=0.25, Fisher's exact test). The average repeat number in the 1,089 ALS patients without the HRE was 3.79 ± 2.59 (range 2-25), while the repeat number in the 1,062 control subjects was 3.84± 2.60 (range 2-23). The repeat number distribution of the alleles is shown in (Figure 1). PCA showed minimal evidence for population stratification. The three patients carrying the HRE were not ethnic outliers, and are not closely related to one another (genomic relationship matrix elements<0.01)(Yang, et al., 2010).

All three HRE carriers showed classic clinical manifestation of ALS with progressive respiratory dysfunction. Ch_MND_8446 had spinal-onset ALS characterized by right arm weakness beginning at age 58. He presented with

ACCEPTED MANUSCRIPT

dysarthria and difficulty swallowing, and the patient died of ALS three years later. Ch_MND_9059 initially suffered from weakness in both hands at age 58 and died from respiratory failure within 15 months after the onset of symptoms. Ch_MND_11134 presented with weakness in the left leg at age 54. He presented with progressive respiratory dysfunction and received a tracheotomy three years later, but the patient died within six months after the tracheotomy.

Haplotype analysis of the *C9orf72* locus

Genotypes were generated from 23 SNP loci located within 121 kilobases of *C9orf72*. In single marker analysis, no SNPs or haplotypes showed any significant association with ALS risk. The total set of consensus European risk alleles and risk allele A of rs2814707 was not found in the three HRE carriers (Table 2). Five alleles emphasized in defining the 20-SNP European founder risk haplotype (A of rs2814707, A of rs3849942, C of rs774359, G of rs2282241, and G of rs1982915) were not present in two of the three cases (Ch_MND_9059 and Ch_MND_11134). For the two SNPs flanking the repeat (rs2282441 and rs11789520), these two cases were homozygous for T and C alleles, while the European founder carries G and T alleles respectively. Thus for these two individuals their genotypes were inconsistent with the haplotype attributed to the European founder chromosome (Table 2, grey shading on genotypes).

Analysis of the genotype data together with data from 1000 Genomes CHB subjects generated inferred haplotypes for the three HRE carriers (Table S2) that were robust

ACCEPTED MANUSCRIPT

to different PHASE parameters (deviations from the stepwise mutation model or inclusion of the repeat genotypes,Table S3). The inferred haplotype solution allocates a shared 23-SNP haplotype harboring the HRE in Ch_MND_11134 and Ch_MND_8446; this haplotypes (excluding the HRE, TTGGAAACCCGTGTCACCAAAGT) has a frequency of 0.049 in cases and, 0.056 in controls). With only three individuals identified with the *C9orf72*HRE we believe the statistically inferred haplotype should be interpreted with caution. To add insight we explored alternative haplotype solutions, for example all three HRE have genotypes that could be consistent with a shared 22-SNP haplotype. Such an alternative haplotype could be shared by Ch_MND_11134 and Ch_MND_9059, while the HRE is harboured on the European published haplotype for Ch_MND_8446. We calculated the frequencies of the haplotype possibilities in 1000 Genomes and cases and controls (Table S2), and examined reported genotypes for other published studies for Chinese or Japanese cohorts (Table S4) but were unable to draw conclusions.

The frequency of the Finnish 20-SNP haplotype was 1.8% in cases and 1.9% in controls (p=0.89). The *C9orf72* repeat copy number in cases was significantly higher for the 29 Finnish 20-SNP haplotypes compared to all other haplotypes (Table S6); Fisher's Exact test comparing the frequency of alleles less than 8 repeats vs alleles of8 or more repeats (3 vs1459 with fewer than 8, and 26 vs108 with 8 or morerepeats, p =0).The frequency of the haplotype shared by the two HRE carriers Ch_MND_11134 and Ch_MND_9059 ("Chinese haplotype") has similar frequency

ACCEPTED MANUSCRIPT

in cases and controls (4.9% vs 5.6%, p=0.26). However, as for the Finnish haplotype, alleles with 8 or more repeats were more common inthe80“Chinese haplotypes” found in cases (alleles with fewer than 8 repeats (62vs 1398) vs alleles with 8 or more repeats (14vs118) $p = 4.6x10^{-3}$ or $p=4.8x10^{-4}$ after excluding Finnish haplotypes). These results imply that the Chinese haplotype like the Finnish haplotype is associated with repeat length instability.

CpG Methylation in the vicinity of *C9orf72*

We next performed DNA methylation analysis of the three HRE carriers and 656 non-carriers (458 sALS cases, 198 healthy controls). Three probes located within the CpG island immediately upstream of the repeat (cg14363787, cg23074747 and cg05990720; Figure 2,Table S5), showed increased CpG methylation for two HRE carriers (Ch_MND_11134, Ch_MND_9059); the two individuals whose SNP genotypes are inconsistent with the reported European risk haplotype) and no significant change in the third patient (Ch_MND_8446) (2/3 carriers, 0/656 non-carriers).Ch_MND_11134 showed the greatest increase in methylation amongst all 659 samples tested. All non-carriers showed low methylation levels at the four CpG sites closest to the repeat and in these individuals the correlation between methylation levels and repeat length was not significantly different from zero. A conservative estimate of the probability of the observed results occurring by chance is to consider the results of only cg14363787, cg23074747 for which Ch_MND_11134 and Ch_MND_9059 were ranked first and second out of all cases.

ACCEPTED MANUSCRIPT

Such low probability (((3x8)/(461*460))2 = 1.2 x10^{-8}) is consistent with a causal functional role for these expansion carriers.

DISCUSSION

This study confirms that *C9orf72* repeat expansions are less prevalent in patients of Chinese compared to Caucasian ethnicity. The expansion was identified in 3/1,092 cases compared to 0/1,062 controls. Given the low prevalence, this difference doesnot achieve statistical significance (Fisher's exact test, p =0.25). The HRE has previously been identified in Caucasian control subjects with an observed frequency of 0-0.4% (Beck, et al., 2013,Galimberti, et al., 2014,Majounie, et al., 2012,Renton, et al., 2011), almost as small as the frequency found in our sALS cohort. We are unable to provide statistical or functional proof that the HREs detected in the Chinese cohort are the cause of ALS in the three identified cases, rather than the equivalent to HREs found in non-ALS cohorts of European ancestry. However, the HRE was not found in 1,062 Chinese controls in our study or Chinese healthy controls reported by others (Jiao, et al., 2014,Zou, et al., 2013). Increased methylation of the CpGisland upstream of the repeat supports that, at least for the two patients who do not have a diplotype consistent with the European haplotype, the HREs are full expansions rather than large alleles indistinguishable from full expansions by the limitations of repeat-primed PCR. It should be noted that full expansions could be definitively confirmed using Southern blot analysis (Beck et al, 2013), however, asufficient quantity of DNA was unavailable for each of the three cases. Identification of further *C9orf72* HRE carriers in China, and confirmation by

ACCEPTED MANUSCRIPT

Southern blotting in the future will be necessary to strengthen our conclusions that the HRE may occurindependent from the Finnish founder. We note that the HRE testing and the methylation analysis of the samples occurred independently at two different locations (Beijing and Brisbane respectively). Similarly, the SNP genotyping was conducted in Beijing and the genome-wide SNP genotyping was conducted in Brisbane. This strategy of convenience also provides confidence that our results are protected from technical error. The chance of the methylation outliers occurring in the same samples as those with HRE by chance is unlikely ($P<10^{-8}$) and we conclude that methylation is the result of GGGGCC expansion.

The frequency of the *C9orf72* HRE in several Caucasian ALS cohorts range from 10.95%- 47% in fALS and 3.20%- 21.1% in sALS(Garcia-Redondo, et al., 2013,Gijselinck, et al., 2012,Majounie, et al., 2012,Millecamps, et al., 2012,K.Y. Mok, et al., 2012,Ratti, et al., 2012,Sabatelli, et al., 2012). The reported frequencies in Asian samples have been lower (Ishiura, et al., 2012,Konno, et al., 2012,Majounie, et al., 2012,Ogaki, et al., 2012) but with a higher frequency reported in cases from Taiwan of Han Chinese ancestry (18.2% in fALS and 2.0% in sALS) (Tsai, et al., 2012). In the present study, we report that the frequency of the *C9orf72* HRE in Chinese sALS patients is 0.3%. To date, this is the largest cohort screened in non-Caucasian ALS cases. Previous reports about *C9orf72* HRE in different populations support the hypothesis that the pathogenic HRE arose from a single founder haplotype both in European and Asian populations (Ishiura, et al., 2012,Jiao, et al.,

ACCEPTED MANUSCRIPT

2014,Konno, et al., 2012,Majounie, et al., 2012,K.Y. Mok, et al., 2012,Ogaki, et al., 2012,Tsai, et al., 2012) and for cases in Taiwan it was postulated that the HRE was introduced in the 17th century when Taiwan was under Dutch and Spanish colonial rule. The conjecture that Asian HREs arose on the same European founder seems inconsistent with the mutation occurring in a founder 1,500 years ago (Majounie, et al., 2012), since our study identified individuals with SNP alleles closely flanking the HRE that are not part of the reported haplotype. The A allele of rs2814707 has been used as a significant genetic association marker of chromosome 9p21 in the Flanders-Belgian cohort of patients with FTLD, FTLD-ALS, and ALS (odds ratio 2.6, 95% CI 1.5-4.7; p=0.001) (Gijselinck, et al., 2012). However, this allele was not found in our cases carrying the *C9orf72* HRE, corresponding with the absence of association between rs2814707 on 9p21 and sALS in Japanese or Chinese populations (Iida, et al., 2011). The A allele of rs3849942 has also been associated with an increased risk of ALS (Beck, et al., 2013), and this variant was not present in Ch_MND_9059 or Ch_MND_11134. Four additional SNP alleles present on the European founder haplotype were also not found in these two patients.

Only one report has suggested that the HRE has occurred on multiple haplotypes in Europeans (Beck, et al., 2013) but others have concluded that the risk haplotype holds worldwide including in the Chinese and Japanese ALS populations (Pliner, et al., 2014). The information provided by Chinese samples we report here conclusively shows that in Chinese cases, ALS-associated *C9orf72* repeats are found on different haplotypes to the European risk haplotypes. The mutation on the

ACCEPTED MANUSCRIPT

European founder remains the most common known cause of ALS in Caucasian populations, but the possibility of rare, independent mutations should not be ruled out.

Our exploration of the relationship between presence of the HRE and aberrant CpG methylation should be interpreted with caution given only 3 HRE carriers were identified. None-the-less, hypermethylation of CpGdinucleotides is an epigenetic modification that could lead to gene silencing or reduction of the amount of transcript produced. The previously reported methylation study on a larger cohort of Caucasian HRE carriers showed heterogeneity in the level of methylation (Xi, et al., 2013). Consistent with this observation, only two of the three Chinese HRE carriers showed hyper-methylation at CpG sites assessed. In the Caucasian study, methylation of the CpGisland was not found in samples containing normal or intermediate sized alleles (up to 43 repeats). The presence of the HRE along with methylation of the upstream CpGisland suggests that the HREs, for at least two of the Chinese ALS patients are full repeat expansions rather than large alleles indistinguishable from full expansions by the limitations of repeat-primed PCR. Methylation status at these sites is of importance given the hypotheses of RNA toxicity of the transcribed repeat and/or toxicity of dipeptides translated from the transcribed repeat. Both could be influenced by the amount of HRE-containing transcript produced, which is influenced by methylation of the HRE and nearby CpG sites (Xi, et al., 2013)).

ACCEPTED MANUSCRIPT

As well as reduced frequency of repeat expansions, the size distribution of normal alleles also differs between the Han Chinese cohort and that reported in Caucasians. Caucasian chromosomes contain a range of alleles with up to43 copies (Beck, et al., 2013,Xi, et al., 2013) whereas in the Chinese samples alleles with more than 12 repeats were rarely observed (Figure 1). For other diseases for which repeat expansion mutations are causal a lower frequency of larger, normal alleles in a population was related to a reduced incidence of repeat expansion associated disease(e.g. myotonic dystrophy (Zerylnick, et al., 1995)), implying that higher repeat lengths are associated with genomic instability. Germline mutation from a normal range or a pre-mutation range allele to a full expansion is yet to be observed for the *C9orf72* repeat. However, a sporadic ALS case has recently been described with 90 copies of the CCGGG repeat in blood, but over 3000 copies in brain and spinal cord (Fratta, et al., 2014). This somatic instability suggests that alleles with 90 copies of the repeat are possibly within the pre-mutation range for *C9orf72* e.g. although perhaps not pathogenic as 90 copies, upon transmission (both germline and somatic) can give rise to a larger pathogenic expansion. Thus the rules for dynamic mutation appear to apply to the *C9orf72* repeat where upon transmission, a dynamic mutation allele has a higher chance of mutation that its predecessor (Sutherland and Richards, 1995). Three HREs have been identified in our Han Chinese ALS patients, and at least two do not carry the reported European haplotype, which could imply that alleles of the *C9orf72*hexanucleotide repeat susceptible to dynamic mutation, independent of the European founder chromosome, do occur but are rare. The

ACCEPTED MANUSCRIPT

European haplotype associated with pathological HRE does exist in the Chinese population, albeit at lower frequency (<2% in Chinese controls, compared to ~9% in European controls) and in our case sample the haplotype is highly associated with repeat length greater than 8 as in Europeans (90% of Finnish haplotypes vs7% of other haplotype, p=0, Table S6). Interestingly, the haplotype associated with two of the three Chinese HRE carrierswhich has frequency ~5% in both cases and controls is also associated with repeat alleles of 8 or more copies (20% vs 8% in other haplotypes, $p=6.3 \times 10^{-4}$; Table S7). These results imply both the Finnish and Chinese haplotypes associated with HRE may confer instability of repeat length.

Bulbar onset and cognitive impairment are more common in ALS patients with the *C9orf72* HRE (Snowden, et al., 2013,Stewart, et al., 2012). The three Chinese patients harboring the *C9orf72* expansion presented with spinal onset but no cognitive impairment. Although the age of onset has been reported to be younger in sALS patients with the HREs compared to those without expansions, our patients developed the symptoms at a relatively older age. Given that we report only three HRE carriers, it is not possible for us to make conclusions regarding the role of the HRE or CpG methylation in disease onset or progression in Chinese ALS.

In conclusion, the present study demonstrates the low frequency of the *C9orf72* HRE in Chinese sALS patients from mainland China. HRE in Chinese sALS was associated with methylation of the upstream CpGisland in a pattern that appears similar to that in Caucasian ALS. The distinct alleles in SNPs flanking the HRE at ~1kb in Chinese sALS patients compared to the European 20-SNP consensus risk

ACCEPTED MANUSCRIPT

haplotype does not support the hypothesis that the *C9orf72* expansion arose in a single founder about 100 generations ago (Pliner, et al., 2014).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ACKNOWLEDGMENTS

This work was supported by the National Natural Sciences Foundation of China [81030019] [DF]; The Australian Research Council Linkage Grant [LPl10200926]; National Health and Medical Research Council grants [APP1048853, APP1046880 to P.M.V.], Senior Research Fellowship [613602to N.R.W.], Senior Principal Research Fellowships to M.A.B., P.M.V., The Ross Maclean Senior Research Fellowship and the Peter Goodenough Bequest [MM] and the Mick Rodger Benalla Grant from the Motor Neuron Disease Research Institute of Australia [B.B]

CONFLICT OF INTEREST STATEMENT

None declared

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P., Orrell, R.W., Malaspina, A., Rowe, J., Brown, J., Hodges, J., Sidle, K., Polke, J.M., Houlden, H., Schott, J.M., Fox, N.C., Rossor, M.N., Tabrizi, S.J., Isaacs, A.M., Hardy, J., Warren, J.D., Collinge, J., Mead, S. 2013. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. American journal of human genetics 92(3), 345-53. doi:10.1016/j.ajhg.2013.01.011.

Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L. 2000. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases 1(5), 293-9.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.Y., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72(2), 245-56. doi:10.1016/j.neuron.2011.09.011.

Donnelly, C.J., Zhang, P.W., Pham, J.T., Heusler, A.R., Mistry, N.A., Vidensky, S., Daley, E.L., Poth, E.M., Hoover, B., Fines, D.M., Maragakis, N., Tienari, P.J., Petrucelli, L., Traynor, B.J., Wang, J., Rigo, F., Bennett, C.F., Blackshaw, S., Sattler, R., Rothstein, J.D. 2013. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron 80(2), 415-28. doi:10.1016/j.neuron.2013.10.015.

Fratta, P., Polke, J.M., Newcombe, J., Mizielinska, S., Lashley, T., Poulter, M., Beck, J., Preza, E., Devoy, A., Sidle, K., Howard, R., Malaspina, A., Orrell, R.W., Clarke, J., Lu, C.H., Mok, K., Collins, T., Shoaii, M., Nanji, T., Wray, S., Adamson, G., Pittman, A., Renton, A.E., Traynor, B.J., Sweeney, M.G., Revesz, T., Houlden, H., Mead, S., Isaacs, A.M., Fisher, E.M. 2014. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiology of aging. doi:10.1016/j.neurobiolaging.2014.07.037.

Galimberti, D., Arosio, B., Fenoglio, C., Serpente, M., Cioffi, S.M., Bonsi, R., Rossi, P., Abbate, C., Mari, D., Scarpini, E. 2014. Incomplete Penetrance of the C9ORF72 Hexanucleotide Repeat Expansions: Frequency in a Cohort of Geriatric Non-Demented Subjects. Journal of Alzheimer's disease : JAD 39(1), 19-22. doi:10.3233/JAD-131172.

Garcia-Redondo, A., Dols-Icardo, O., Rojas-Garcia, R., Esteban-Perez, J., Cordero-Vazquez, P., Munoz-Blanco, J.L., Catalina, I., Gonzalez-Munoz, M., Varona, L., Sarasola, E., Povedano, M., Sevilla, T., Guerrero, A., Pardo, J., de Munain,

ACCEPTED MANUSCRIPT

A.L., Marquez-Infante, C., de Rivera, F.J., Pastor, P., Jerico, I., de Arcaya, A.A., Mora, J.S., Clarimon, J., Gonzalo-Martinez, J.F., Juarez-Rufian, A., Atencia, G., Jimenez-Bautista, R., Moran, Y., Mascias, J., Hernandez-Barral, M., Kapetanovic, S., Garcia-Barcina, M., Alcala, C., Vela, A., Ramirez-Ramos, C., Galan, L., Perez-Tur, J., Quintans, B., Sobrido, M.J., Fernandez-Torron, R., Poza, J.J., Gorostidi, A., Paradas, C., Villoslada, P., Larrode, P., Capablo, J.L., Pascual-Calvet, J., Goni, M., Morgado, Y., Guitart, M., Moreno-Laguna, S., Rueda, A., Martin-Estefania, C., Cemillan, C., Blesa, R., Lleo, A. 2013. Analysis of the C9orf72 Gene in Patients with Amyotrophic Lateral Sclerosis in Spain and Different Populations Worldwide. Human mutation 34(1), 79-82. doi:10.1002/humu.22211.

Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., Engelborghs, S., Sieben, A., De Jonghe, P., Vandenberghe, R., Santens, P., De Bleecker, J., Maes, G., Baumer, V., Dillen, L., Joris, G., Cuijt, I., Corsmit, E., Elinck, E., Van Dongen, J., Vermeulen, S., Van den Broeck, M., Vaerenberg, C., Mattheijssens, M., Peeters, K., Robberecht, W., Cras, P., Martin, J.J., De Deyn, P.P., Cruts, M., Van Broeckhoven, C. 2012. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet neurology 11(1), 54-65. doi:10.1016/S1474-4422(11)70261-7.

Howie, B., Marchini, J., Stephens, M. 2011. Genotype imputation with thousands of genomes. G3 (Bethesda) 1(6), 457-70. doi:10.1534/g3.111.001198.

Iida, A., Takahashi, A., Deng, M., Zhang, Y., Wang, J., Atsuta, N., Tanaka, F., Kamei, T., Sano, M., Oshima, S., Tokuda, T., Morita, M., Akimoto, C., Nakajima, M., Kubo, M., Kamatani, N., Nakano, I., Sobue, G., Nakamura, Y., Fan, D., Ikegawa, S. 2011. Replication analysis of SNPs on 9p21.2 and 19p13.3 with amyotrophic lateral sclerosis in East Asians. Neurobiology of aging 32(4), 757 e13-4. doi:10.1016/j.neurobiolaging.2010.12.011.

Ishiura, H., Takahashi, Y., Mitsui, J., Yoshida, S., Kihira, T., Kokubo, Y., Kuzuhara, S., Ranum, L.P., Tamaoki, T., Ichikawa, Y., Date, H., Goto, J., Tsuji, S. 2012. C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. Archives of neurology 69(9), 1154-8. doi:10.1001/archneurol.2012.1219.

Jiao, B., Tang, B., Liu, X., Yan, X., Zhou, L., Yang, Y., Wang, J., Xia, K., Shen, L. 2014. Identification of C9orf72 repeat expansions in patients with amyotrophic lateral sclerosis and frontotemporal dementia in mainland China. Neurobiology of aging 35(4), 936 e19-22. doi:10.1016/j.neurobiolaging.2013.10.001.

Konno, T., Shiga, A., Tsujino, A., Sugai, A., Kato, T., Kanai, K., Yokoseki, A., Eguchi, H., Kuwabara, S., Nishizawa, M., Takahashi, H., Onodera, O. 2012. Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. Journal of neurology,

ACCEPTED MANUSCRIPT

neurosurgery, and psychiatry 84(4), 398-401. doi:10.1136/jnnp-2012-302272.

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R., Nalls, M.A., Heckerman, D., Tienari, P.J., Traynor, B.J. 2010. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet neurology 9(10), 978-85. doi:10.1016/S1474-4422(10)70184-8.

Liu, R., Tang, L., Cai, B., Liu, X., Ye, S., Ma, Y., Zhang, H., Fan, D. 2013. C9orf72 repeat expansions are not detected in Chinese patients with familial ALS. Amyotrophic lateral sclerosis and frontotemporal degeneration 14(7-8), 630-1. doi:10.3109/21678421.2013.817588.

Majounie, E., Renton, A.E., Mok, K., Dopper, E.G., Waite, A., Rollinson, S., Chio, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., van Swieten, J.C., Abramzon, Y., Johnson, J.O., Sendtner, M., Pamphlett, R., Orrell, R.W., Mead, S., Sidle, K.C., Houlden, H., Rohrer, J.D., Morrison, K.E., Pall, H., Talbot, K., Ansorge, O., Hernandez, D.G., Arepalli, S., Sabatelli, M., Mora, G., Corbo, M., Giannini, F., Calvo, A., Englund, E., Borghero, G., Floris, G.L., Remes, A.M., Laaksovirta, H., McCluskey, L., Trojanowski, J.Q., Van Deerlin, V.M., Schellenberg, G.D., Nalls, M.A., Drory, V.E., Lu, C.S., Yeh, T.H., Ishiura, H., Takahashi, Y., Tsuji, S., Le Ber, I., Brice, A., Drepper, C., Williams, N., Kirby, J., Shaw, P., Hardy, J., Tienari, P.J., Heutink, P., Morris, H.R., Pickering-Brown, S., Traynor, B.J. 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet neurology 11(4), 323-30. doi:10.1016/S1474-4422(12)70043-1.

Millecamps, S., Boillee, S., Le Ber, I., Seilhean, D., Teyssou, E., Giraudeau, M., Moigneu, C., Vandenberghe, N., Danel-Brunaud, V., Corcia, P., Pradat, P.F., Le Forestier, N., Lacomblez, L., Bruneteau, G., Camu, W., Brice, A., Cazeneuve, C., Leguern, E., Meininger, V., Salachas, F. 2012. Phenotype difference between ALS patients with expanded repeats in C9ORF72 and patients with mutations in other ALS-related genes. Journal of medical genetics 49(4), 258-63. doi:10.1136/jmedgenet-2011-100699.

Mok, K., Traynor, B.J., Schymick, J., Tienari, P.J., Laaksovirta, H., Peuralinna, T., Myllykangas, L., Chio, A., Shatunov, A., Boeve, B.F., Boxer, A.L., DeJesus-Hernandez, M., Mackenzie, I.R., Waite, A., Williams, N., Morris, H.R., Simon-Sanchez, J., van Swieten, J.C., Heutink, P., Restagno, G., Mora, G., Morrison, K.E., Shaw, P.J., Rollinson, P.S., Al-Chalabi, A., Rademakers, R., Pickering-Brown, S., Orrell, R.W., Nalls, M.A., Hardy, J. 2012. Chromosome 9 ALS and FTD locus is probably derived from a single founder. Neurobiology of aging 33(1), 209 e3-8. doi:10.1016/j.neurobiolaging.2011.08.005.

Mok, K.Y., Koutsis, G., Schottlaender, L.V., Polke, J., Panas, M., Houlden, H. 2012. High frequency of the expanded C9ORF72 hexanucleotide repeat in

ACCEPTED MANUSCRIPT

familial and sporadic Greek ALS patients. Neurobiology of aging 33(8), 1851 e1-5. doi:10.1016/j.neurobiolaging.2012.02.021.

Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C., Haass, C., Edbauer, D. 2013. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. Science 339(6125), 1335-8. doi:10.1126/science.1232927.

Ogaki, K., Li, Y., Atsuta, N., Tomiyama, H., Funayama, M., Watanabe, H., Nakamura, R., Yoshino, H., Yato, S., Tamura, A., Naito, Y., Taniguchi, A., Fujita, K., Izumi, Y., Kaji, R., Hattori, N., Sobue, G. 2012. Analysis of C9orf72 repeat expansion in 563 Japanese patients with amyotrophic lateral sclerosis. Neurobiology of aging 33(10), 2527 e11-6. doi:10.1016/j.neurobiolaging.2012.05.011.

Pliner, H.A., Mann, D.M., Traynor, B.J. 2014. Searching for Grendel: origin and global spread of the C9ORF72 repeat expansion. Acta neuropathologica 127(3), 391-6. doi:10.1007/s00401-014-1250-x.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38(8), 904-9. doi:10.1038/ng1847.

Ratti, A., Corrado, L., Castellotti, B., Del Bo, R., Fogh, I., Cereda, C., Tiloca, C., D'Ascenzo, C., Bagarotti, A., Pensato, V., Ranieri, M., Gagliardi, S., Calini, D., Mazzini, L., Taroni, F., Corti, S., Ceroni, M., Oggioni, G.D., Lin, K., Powell, J.F., Soraru, G., Ticozzi, N., Comi, G.P., D'Alfonso, S., Gellera, C., Silani, V. 2012. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. Neurobiology of aging 33(10), 2528 e7-14. doi:10.1016/j.neurobiolaging.2012.06.008.

Renton, A.E., Chio, A., Traynor, B.J. 2014. State of play in amyotrophic lateral sclerosis genetics. Nature neuroscience 17(1), 17-23. doi:10.1038/nn.3584.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A.M., Kaganovich, A., Scholz, S.W., Duckworth, J., Ding, J., Harmer, D.W., Hernandez, D.G., Johnson, J.O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R.J., Orrell, R.W., Neal, J., Murray, A., Pearson, J., Jansen, I.E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J.B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M.A., Peuralinna, T., Jansson, L., Isoviita, V.M., Kaivorinne, A.L., Holtta-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chio, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J.D., Sendtner, M., Drepper, C., Eichler, E.E., Alkan, C., Abdullaev, Z., Pack, S.D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N.M., Heutink, P., Pickering-Brown, S., Morris, H.R., Tienari, P.J., Traynor, B.J. 2011. A hexanucleotide repeat

ACCEPTED MANUSCRIPT

expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72(2), 257-68. doi:10.1016/j.neuron.2011.09.010.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X., et al. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362(6415), 59-62. doi:10.1038/362059a0.

Sabatelli, M., Conforti, F.L., Zollino, M., Mora, G., Monsurro, M.R., Volanti, P., Marinou, K., Salvi, F., Corbo, M., Giannini, F., Battistini, S., Penco, S., Lunetta, C., Quattrone, A., Gambardella, A., Logroscino, G., Simone, I., Bartolomei, I., Pisano, F., Tedeschi, G., Conte, A., Spataro, R., La Bella, V., Caponnetto, C., Mancardi, G., Mandich, P., Sola, P., Mandrioli, J., Renton, A.E., Majounie, E., Abramzon, Y., Marrosu, F., Marrosu, M.G., Murru, M.R., Sotgiu, M.A., Pugliatti, M., Rodolico, C., Moglia, C., Calvo, A., Ossola, I., Brunetti, M., Traynor, B.J., Borghero, G., Restagno, G., Chio, A. 2012. C9ORF72 hexanucleotide repeat expansions in the Italian sporadic ALS population. Neurobiology of aging 33(8), 1848 e15-20. doi:10.1016/j.neurobiolaging.2012.02.011.

Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M., Jones, A., Gray, I., Wright, J., Hortobagyi, T., Al-Sarraj, S., Rogelj, B., Powell, J., Lupton, M., Lovestone, S., Sapp, P.C., Weber, M., Nestor, P.J., Schelhaas, H.J., Asbroek, A.A., Silani, V., Gellera, C., Taroni, F., Ticozzi, N., Van den Berg, L., Veldink, J., Van Damme, P., Robberecht, W., Shaw, P.J., Kirby, J., Pall, H., Morrison, K.E., Morris, A., de Belleroche, J., Vianney de Jong, J.M., Baas, F., Andersen, P.M., Landers, J., Brown, R.H., Jr., Weale, M.E., Al-Chalabi, A., Shaw, C.E. 2013. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. European journal of human genetics : EJHG 21(1), 102-8. doi:10.1038/ejhg.2012.98.

Snowden, J.S., Harris, J., Richardson, A., Rollinson, S., Thompson, J.C., Neary, D., Mann, D.M., Pickering-Brown, S. 2013. Frontotemporal dementia with amyotrophic lateral sclerosis: a clinical comparison of patients with and without repeat expansions in C9orf72. Amyotrophic lateral sclerosis and frontotemporal degeneration 14(3), 172-6. doi:10.3109/21678421.2013.765485.

Stephens, M., Smith, N.J., Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. American journal of human genetics 68(4), 978-89. doi:10.1086/319501.

Stewart, H., Rutherford, N.J., Briemberg, H., Krieger, C., Cashman, N., Fabros, M., Baker, M., Fok, A., DeJesus-Hernandez, M., Eisen, A., Rademakers, R., Mackenzie, I.R. 2012. Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. Acta neuropathologica 123(3), 409-17. doi:10.1007/s00401-011-0937-5.

ACCEPTED MANUSCRIPT

Sutherland, G.R., Richards, R.I. 1995. Simple tandem DNA repeats and human genetic disease. Proceedings of the National Academy of Sciences of the United States of America 92(9), 3636-41.

Tsai, C.P., Soong, B.W., Tu, P.H., Lin, K.P., Fuh, J.L., Tsai, P.C., Lu, Y.C., Lee, I.H., Lee, Y.C. 2012. A hexanucleotide repeat expansion in C9ORF72 causes familial and sporadic ALS in Taiwan. Neurobiology of aging 33(9), 2232 e11- e18. doi:10.1016/j.neurobiolaging.2012.05.002.

Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J., Robertson, J., Rogaeva, E. 2013. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. American journal of human genetics 92(6), 981-9. doi:10.1016/j.ajhg.2013.04.017.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42(7), 565-9. doi:10.1038/ng.608.

Zerylnick, C., Torroni, A., Sherman, S.L., Warren, S.T. 1995. Normal variation at the myotonic dystrophy locus in global human populations. American journal of human genetics 56(1), 123-30.

Zou, Z.Y., Li, X.G., Liu, M.S., Cui, L.Y. 2013. Screening for C9orf72 repeat expansions in Chinese amyotrophic lateral sclerosis patients. Neurobiology of aging 34(6), 1710 e5-6. doi:10.1016/j.neurobiolaging.2012.11.018.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

LEGENDS TO FIGURES

Figure 1.

The frequency of alleles with different repeat copy numbers in sporadic ALS patients (n=1,092) and control subjects (n=1,062).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure 2.

Boxplots of the methylation signal in cases and controls for the*C9orf72*probes andDNA methylation levels in *C9orf72* gene region.

The boxes show individuals with beta values between the 1st and 3rd quartiles, the black horizontal line shows the median methylation value in cases and controls, and outliers (methylation values more than 1·5 times inter-quartile range) are shown as black circles. The methylation level for each probe in the three individuals with repeat expansion is shown in red (together with sample ID). The distance of the CpG probe to the transcription start site (TSS) is given above each figure. Below the box plots is a physical map of the HRE interval and positions of CpG islands. The zoomed in interval shows position of CpGs with increased methylation associated with the HRE. One asterisk indicates probes with increased methylation in Ch_MND_11134 only, two asterisks indicates probes with increased methylation in both Ch_MND_11134 and Ch_MND_9059. The positions of CpGs tested by Xi et al, are shown below the scale bar using the numbering from their report (Xi, et al., 2013). The hashed lines connecting the probes assessed in this report and CpGs below the scale, indicated CpGs common to both studies. The numbering of the chromosomal location above the scale bar are from the UCSC Genome Browser, build GRCh37/hg19. In our study, at the four probes closest to the TSS (from cg05990720 to cg23074747), two of all three cases(Ch_MND_11134, Ch_MND_9059)have increased levels of methylation. All656 non-carriers showed very low methylation levels at the 4 CpG sites closest to upstream of the

ACCEPTED MANUSCRIPT

repeat.Despite the very small sample size, our results showed clear and similar differences in methylation levels at CpG sites near the HRE in carriers.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

TABLES

Table 1. Demographics of the case and control cohorts analyzed for the frequency of the HRE in *C9orf72*.

	Sporadic ALS patients[a]	**Controls for mutation screening[b]**	**Controls for haplotype analysis[d]**
Total	1,092	1,062	2,850
Male	697 (65.20%)	561 (54.05%)	1,213(42.57%)
Female	372 (34.80%)	477 (45.95%)	1,637(57.43%)
Age at onset			
Mean ±SD	49.70 ± 12.22	50.25 ± 14.90[c]	-
Site of onset			
Bulbar	131 (17.08%)	-	-
Spinal	636 (82.92%)	-	-

[a] Data not available for the gender of 23 patients, age of onset for 311 patients, site of onset for 325 patients.

[b]Data not available for the gender for 24 control subjects, age at enrollment for 24 control subjects.

[c] Age at enrollment for control cohort.

[d]Healthy controls survived after quality control.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2. The genotyping data of the single nucleotide polymorphisms (SNPs) surrounding *C9orf72*.

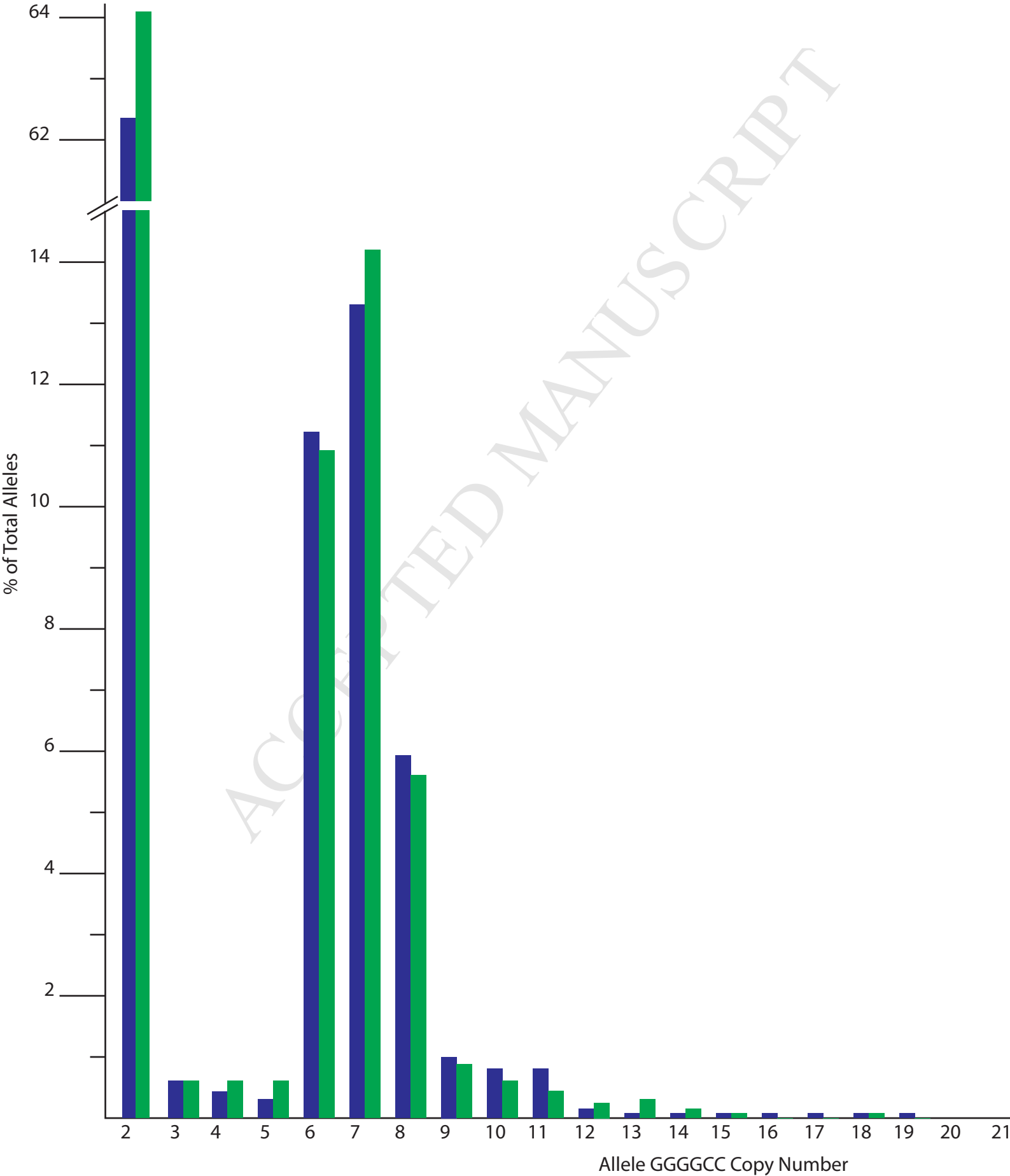
	Marker[a]	Position (bp)	Reference/ Alternative Allele	Risk Haplotype[a]	Frequency of Risk Allele[b]		Genotypes of the 3 HRE individuals[c]: IDs		
					ASN	EUR	11134	8446	9059
1	rs1822723	27478052	C/T	C	0.63	0.73	T/T	C/T	C/C
2	rs4879515	27482235	C/T	T	0.67	0.46	T/T	T/T	T/T
3	rs868856	27489251	T/C	T	0.32	0.28	T/C	T/C	T/T
4	rs7046653	27490967	A/G	A	0.32	0.28	A/G	A/G	A/A
5	rs1977661	27502986	C/A	C	0.69	0.89	C/A	C/A	C/C
6	rs903603	27529316	C/T	C	0.38	0.49	C/T	C/T	C/T
7	rs10812610	27533984	C/A	C	0.39	0.48	C/A	C/A	C/A
8	rs2814707	27536397	G/A	**A**	0.06	0.25	G/G	G/G	G/G
9	rs3849942	27543281	A/G	**A**	0.08	0.25	G/G	A/G	G/G
10	rs700791	27545960	C/A	A	0.08	0.25	C/C	A/C	C/C
11	rs10122902	27556780	G/A	G	0.69	0.84	G/A	G/G	G/A
12	rs10757665	27557919	T/C	T	0.89	0.75	T/T	T/T	T/T
13	rs1565948	27559733	G/A	G	0.57	0.53	G/A	G/G	G/A
14	rs774359	27561049	T/C	**C**	0.08	0.27	T/T	C/T	T/T
15	rs2453554	27561800	C/T	T	0.08	0.26	C/C	T/C	C/C
16	rs2282241	27572255	G/T	**G**	0.32	0.57	T/T	G/T	T/T
	***C9orf72*HR**	27573483	30+	**30+**			30+/2	30+/2	30+/2
17	rs11789520	27574515	C/T	**T**	0.08	0.26	C/C	C/T	C/C
18	rs1948522	27575785	C/T	C	0.88	0.82	C/C	C/C	C/C
19	rs1982915	27579560	A/G	**G**	0.22	0.47	A/A	G/A	A/A
20	rs2453556	27586162	A/G	G	0.46	0.37	G/A	G/A	G/A
21	rs702231	27588731	C/A	A	0.65	0.78	A/C	A/C	A/A
22	rs696826	27589657	A/G	G	0.99	0.86	G/G	G/G	G/G
23	rs2477518	27599746	T/C	T	0.85	0.75	T/T	T/T	C/T

a: Of the 23 SNPs genotyped 20 contributed to the risk haplotype of (Laaksovirta, et al., 2010). The SNPs shaded in grey were included in some subsequent studies e.g. (Smith, et al., 2013).

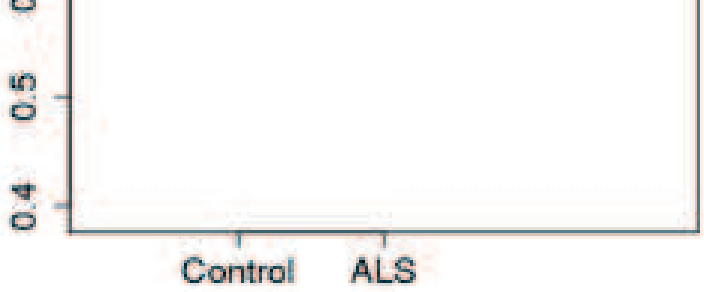
b. The frequency of the haplotype risk alleles derived from the 1,000 Genomes database (www.1000genomes.org/); EUR = frequency estimated from 379 individuals from 1000G populations labeled CEU, FIN, GER, IBS and TSI;ASN = frequency estimated from 286 individuals from 1000G populations labeled CHB.

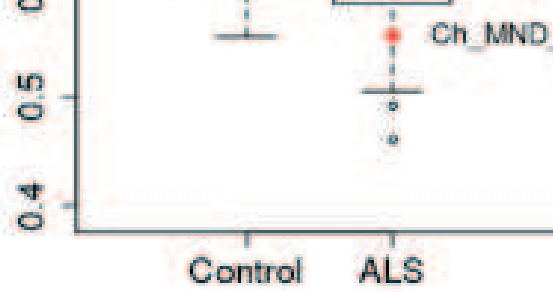
c. Genotypes of the Chinese individuals that are incompatible with the European risk haplotype are shaded grey.

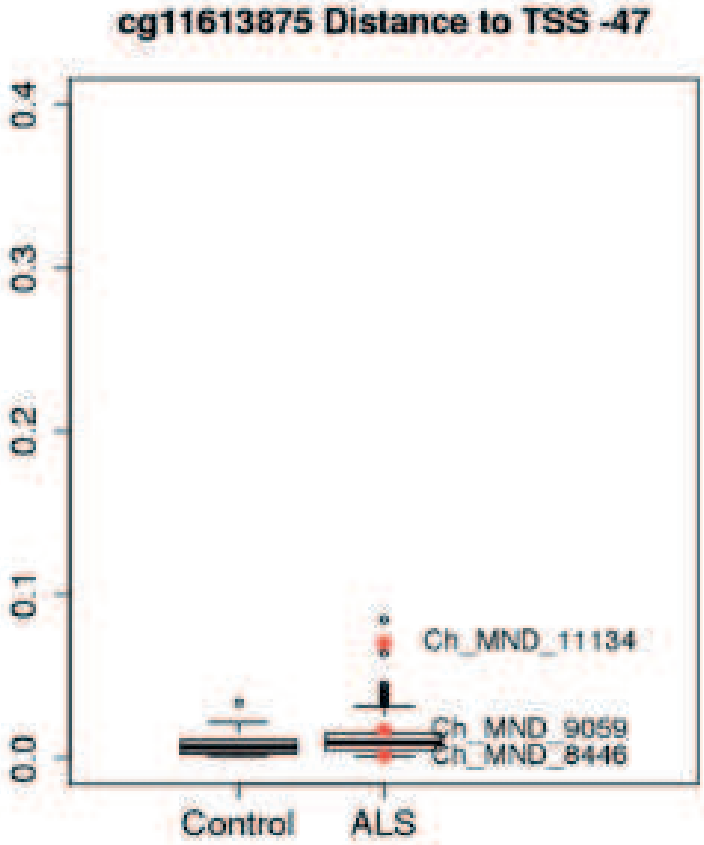
ACCEPTED MANUSCRIPT

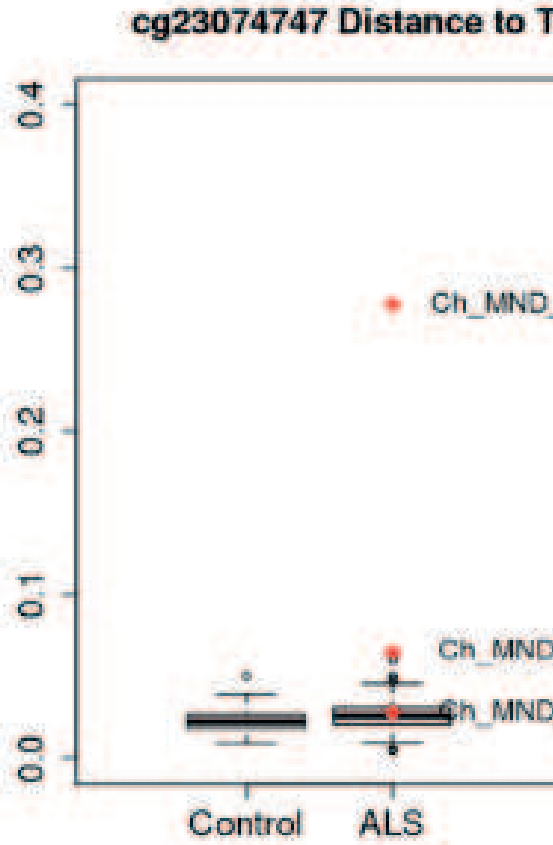


ACCEPTED MANUSCRIPT

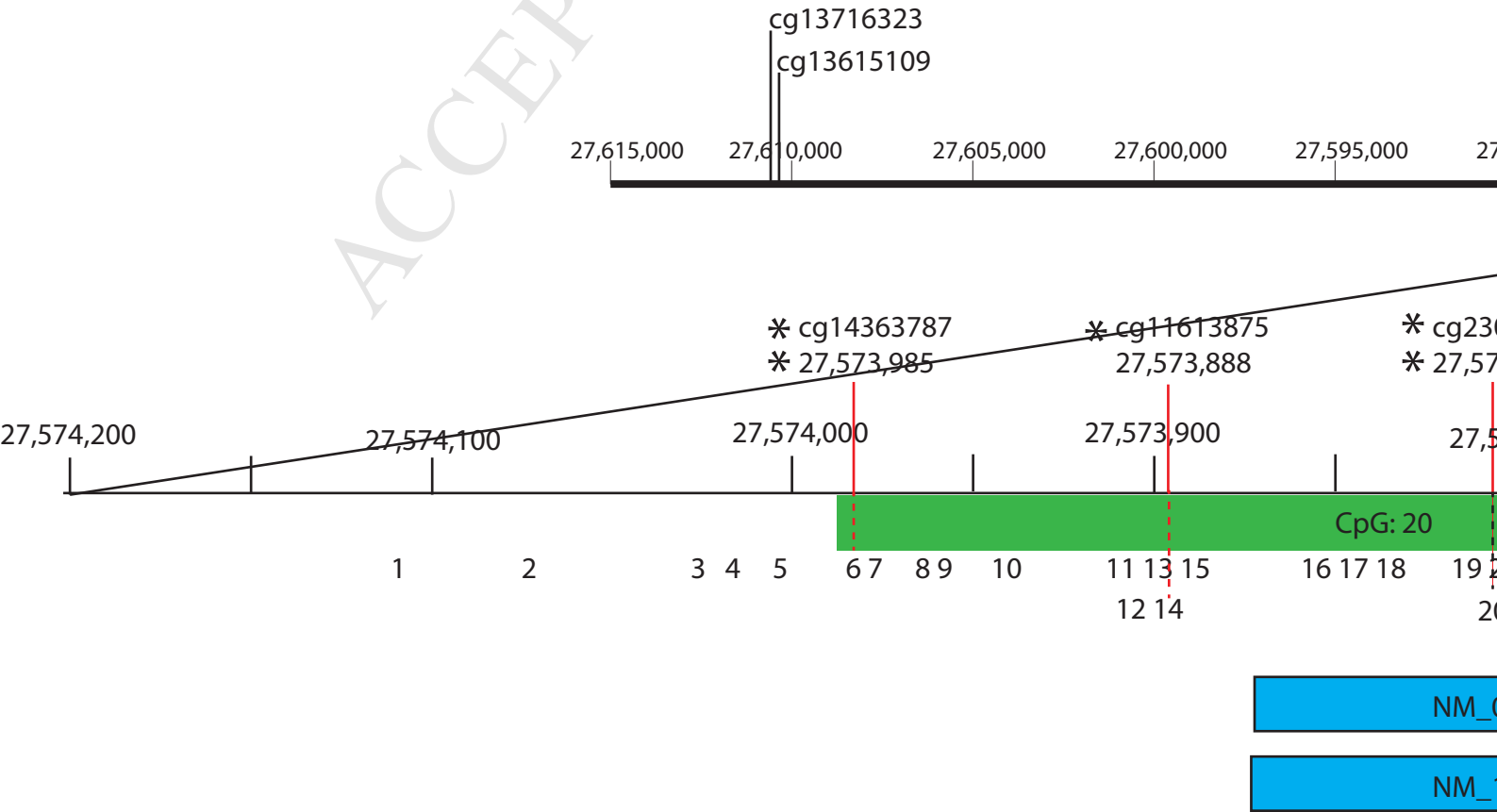

0.4
0.5
Control
ALS


0.4
0.5
Ch_MND
Control
ALS


cg11613875 Distance to TSS -47
0.0
0.1
0.2
0.3
0.4
Ch_MND_11134
Ch_MND_9059
Ch_MND_8446
Control
ALS


cg23074747 Distance to T
0.0
0.1
0.2
0.3
0.4
Ch_MND
Ch_MND
Ch_MND
Control
ALS

ACCEPT


cg13716323
cg13615109
27,615,000
27,610,000
27,605,000
27,600,000
27,595,000
cg14363787
27,573,985
cg11613875
27,573,888
27,574,200
27,574,100
27,574,000
27,573,900
CpG: 20
1
2
3 4 5
6 7
8 9
10
11 13 15
12 14
16 17 18
19

ACCEPTED MANUSCRIPT

Highlights

- We have examined the *C9orf72* gene in the largest cohort of sporadic ALS cases from a non-Caucasian population.
- We conducted a series of comprehensive genetic tests including *C9orf72* hexanucleotide repeat expansion screening, haplotype analysis and methylation analysis of CpG sites surrounding *C9orf72*.
- When compared with Caucasian populations there is a low frequency of the repeat expansion mutation. The Chinese cases we describe are the first expansions that conclusively do not occur on the same founder haplotype as European cases.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table S1 Primers and protocols for repeat PCR

A. Primer sequences for the repeat-primed PCR(DeJesus-Hernandez, et al., 2011)

Step	Primer name	Sequence
Fragment length PCR	Chr9:27563580F	FAM-CAAGGAGGGAAACAACCGCAGCC
	Chr9:27563465R	GCAGGCACCGCAACCGCAG
Repeat-primed PCR	MRX-F	FAM-TGTAAAACGACGGCCAGTCAAGGAGGGAAACAACCGCAGCC
	MRX-M13R	CAGGAAACAGCTATGACC
	MRX-R1	CAGGAAACAGCTATGACCGGGCCCGCCCCGACCACGCCCCGGCCCCGGCCCCGG

B. PCR cycling conditions

Step	Temperature	Time	Repeat
Fragment length PCR	95°	5 min	
	95°	1 min	30 cycles
	64°	1 min	
	72°	1 min	
	72°	60 min	
Repeat-primed PCR	95°	5 min	
	95°	1 min	30 cycles
	62°	1 min	
	72°	1 min	
	72°	60 min	

C. Primers used to amplify the set of 23 SNPs within and surrounding the *C9orf72 gene*.

rs number	Position	Forward primer	Reverse primer
rs1822723*	27478052	TGCAAGGTAAGATGCTGGAA	TTGCTTCTTCCCTGGTTCTG
rs4879515*	27482235	TGAACGAAAAGCACGGTAAA	AGTAAGCCCCCAAAGGAGAC
rs868856*	27489251	GGTCTGACAGAAGCCTTTGC	TGGCTCCCACGGTAGTTTTA
rs7046653*	27490967	CATCGATGCTGTAGGCACAA	TCTCATCGTTCCGAAGAAAT
rs1977661*	27502986	TCACAGGAAAGTAGGCAGGAA	TCCTCCAGTCTAGTCACCCCTA
rs903603*	27529316	CTGCTCGCTCCCGATCTC	CCCAGGCGCACTTTTGTA
rs10812610*	27533984	GGTTTCTATCCAGTGTCCCAGA	GCATGTTTGGGTCTCCATCT
rs2814707*	27536397	GCCTCCTGTAATTGCTCACC	TCCCCTTTTCACCTCTCTCA
rs3849942*	27543281	TCAGAGTGCTGTGGTACGATT	GCGCTCCACAGATGGATACT
rs700791	27545960	ACAGCAGGGATTCACACTCG	ACACTTTCATCTGCAAAGCTAGG
rs10122902*	27556780	TGAAGTGGCTCTCCAGAAGG	TTTGCACACACTCATGAATAGC
rs10757665*	27557919	TTGAAAATGCACAAAAGCCTAA	GGATTTAAGCAATTCCTATGTATGTGT
rs1565948*	27559733	GGTTTGCTGAAATTGAAAGATACA	TCATCTTGATTTCATTGTCTTCG
rs774359*	27561049	GGCACTCAACAAATACTGGCTA	CTTTTTCCATCCAGCAGTGG
rs2453554	27561800	AGCAAGGGGCCATGATTTCT	GTGGGAGGATCCCAGCTCTT
rs2282241*	27572255	TTCTCAGACTTTGGGAAACTTTTA	TGCCAACATAGCCAAACAAA
C9orf72 repeat expansion	27573483		

ACCEPTED MANUSCRIPT

rs11789520	27574515	GCGTCATCTTTTACGTGGGC	GAACAGTAGGAAAAGGGTCTGT
rs1948522*	27575785	CACCCCATAATCCACTCACC	CCTTTCGGAATGGCAGTATC
rs1982915*	27579560	GGCCAGTTTCATGCAATTCT	TGTTTCAGGGTTGATATCACAAA
rs2453556*	27586162	ATCATCCACTCCCCTCCTCT	AACCAAGCAGCCATGAAAAG
rs702231*	27588731	AACCATCTGTCTTCCTTCCATGTC	AGGCAGTGTTTACTCGCCAC
rs696826*	27589657	TTGCTGGAGTGTTTGCAGAG	ATGAACCCAGCTGTCTCACC
rs2477518*	27599746	CGTTTTGGGAAATCGCTTAG	CAGGTGAAACCCATTTGTCA

* from the 20 SNP consensus risk founder haplotype.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table S2: Diplotypes and inferred haplotypes from PHASE analysis in Chinese ALS.

	Marker[a]	Position (bp)	European 23-SNP Risk Haplotype	Genotypes of the HRE carriers[b]			PHASE inferred haplotypes						Finnish 20-SNP risk haplotype[c]	23-SNP haplotype shared by 2 HRE carriers[d]	Haplotype shared by 3 HRE carriers	Haplotype shared by European and 3 HRE carriers
				11134	8446	9059	11134		8446		9059					
1	rs1822723	27478052	C	T/T	C/T	C/C	T	T	T	C	C	C	C	T		
2	rs4879515	27482235	T	T/T	T/T	T/T	T	T	T	T	T	T	T	T	T	T
3	rs868856	27489251	T	T/C	T/C	T/T	C	T	C	T	T	T	T	C		
4	rs7046653	27490967	A	A/G	A/G	A/A	G	A	G	A	A	A	A	G		
5	rs1977661	27502986	C	C/A	C/A	C/C	A	C	A	C	C	C	C	A		
6	rs903603	27529316	C	C/T	C/T	C/T	T	C	T	C	C	T	C	T		
7	rs10812610	27533984	C	C/A	C/A	C/A	A	C	A	C	C	A	C	A		
8	rs2814707	27536397	A	G/G	G/G	G/G	G	G	G	G	G	G	A	G	G	
9	rs3849942	27543281	A	G/G	A/G	G/G	G	G	G	A	G	G	A	G	G	
10	rs700791	27545960	A	C/C	A/C	C/C	C	C	C	A	C	C		C	C	
11	rs10122902	27556780	G	G/A	G/G	G/A	G	A	G	G	A	G	G	G		
12	rs10757665	27557919	T	T/T	T/T	T/T	T	T	T	T	T	T	T	T	T	T
13	rs1565948	27559733	G	G/A	G/G	G/A	G	A	G	G	A	G	G	G		
14	rs774359	27561049	C	T/T	C/T	T/T	T	T	T	C	T	T	C	T	T	
15	rs2453554	27561800	T	C/C	T/C	C/C	C	C	C	T	C	C		C	C	
16	rs2282241	27572255	G	T/T	G/T	T/T	T	T	T	G	T	T	G	T	T	
	C9orf72 HR	27573483	30+	30+/2	30+/2	30+/2	30+	2	30+	2	30+	2				
17	rs11789520	27574515	T	C/C	C/T	C/C	C	C	C	T	C	C		C	C	
18	rs1948522	27575785	C	C/C	C/C	C/C	C	C	C	C	C	C	C	C	C	C
19	rs1982915	27579560	G	A/A	G/A	A/A	A	A	A	G	A	A	G	A	A	
20	rs2453556	27586162	G	G/A	G/A	G/A	A	G	A	G	A	G	G	A	A	
21	rs702231	27588731	A	A/C	A/C	A/A	A	C	A	C	A	A	A	A	A	A
22	rs696826	27589657	G	G/G	G/G	G/G	G	G	G	G	G	G	G	G	G	G
23	rs2477518	27599746	T	T/T	T/T	C/T	T	T	T	T	T	C	T	T	T	T
							Cases (n=939)						33	89	374	470
							Controls (n=2,850)						105	310	1,174	1,469
							Haplotype Frequency in cases						0·018	0·049	0·23	0·30
							Haplotype Frequency in controls						0·019	0·056	0·23	0·30
							p value (χ^2-test)						0·89	0·26	0·50	0·47
							1000G Phase 1 v3 (EUR: n=379)						68	31	49	182
							1000G Phase 1 v3 (ASN; n=286)						17	21	131	192
							Haplotype Frequency in EUR						0.09	0.04	0.07	0.24
							Haplotype Frequency in ASN						0.03	0.04	0.23	0.34
							p value difference (χ^2-test)						3.80E-10	0.58	0	5.71E-08

a: The 20 from the risk haplotype (Laaksovirta, et al., 2010) and 3 additional SNPs (shaded) use by others e.g. (Smith, et al., 2013).
b: Genotypes of the Chinese HRE carriers that are incompatible with the European risk haplotype are shaded in grey.
c: The 20- SNP risk haplotype associated with the C9orf72 gene and shared by most of the patients with ALS and FTD carrying C9orf72 expanded repeats in populations from Finland (Mok et al. 2012).
d. PHASE was used to identify a 23-SNP haplotype shared between CH_MND_11134 and CH_MND_8447.

ACCEPTED MANUSCRIPT

Table S3 The results of the haplotype analysis of the 23 SNPs within and surrounding the *C9orf72* gene in the Chinese Han samples

A. The Output with short burn-in (default; iteration: 100; thinning interval: 1; burn-in: 100)

Ch_MND_11134	TTGGAAACCCGTGTCA 30 CCAAAGT	TTAACGCCCCATATCA 2 CCAGCGT
Ch_MND_8446	TTGGAAACCCGTGTCA 30 CCAAAGT	CTAACGCCTAGTGCTC 2 TCGGCGT
Ch_MND_9059	CTAACGCCCCATATCA 30 CCAAAGT	CTAACAACCCGTGTCA 2 CCAGAGC

B. The Output with long burn-in (iteration: 100; thinning interval: 1; burn-in: 500)

Ch_MND_11134	TTGGAAACCCGTGTCA 30 CCAAAGT	TTAACGCCCCATATCA 2 CCAGCGT
Ch_MND_8446	TTGGAAACCCGTGTCA 30 CCAAAGT	CTAACGCCTAGTGCTC 2 TCGGCGT
Ch_MND_9059	CTAACGCCCCATATCA 30 CCAAAGT	CTAACAACCCGTGTCA 2 CCAGAGC

C. The output without microsatellite loci (iteration: 100; thinning interval: 1; burn-in: 100)

Ch_MND_11134	TTGGAAACCCGTGTCA CCAAAGT	TTAACGCCCCATATCA CCAGCGT
Ch_MND_8446	TTGGAAACCCGTGTCA CCAA(C)GT	CTAACGCCTAGTGCTC TCGG(A)GT
Ch_MND_9059	CTAACGCCCCATATCA CCA(G)AG(C)	CTAACAACCCGTGTCA CCA(A)AG(T)

Different parameters in the analyses were used to check for the consistency. The haplotype with the posterior probability >0.90 was accepted. Brackets: The alleles with probability <0.90.

D. The output with long burn in with varying delta(Stephens, et al., 2001).

Using PHASE analysis for microsatellite loci where the stepwise model is a good assumption, using that model (Delta = 0, default) would provide more power to reconstruct haplotypes than other models (Delta > 0). However, using a small value for (Delta = 0.05 – 0.1) provides greater robustness to occasional deviations from the stepwise model(Stephens, et al., 2001). Using small values of delta (Delta = 0.05-0.25) yielded the same haplotypes. This result showed that the inferred haplotypes are quite robust to the occasional deviations from the stepwise mutation model.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table S4 Reported genotypes of ALS cases carrying the HRE from Asia

			Japan(Konno, et al., 2012,Ogaki, et al., 2012)				Taiwan(Tsai, et al., 2012)				China(Jiao, et al., 2014)	
Marker	European Haplotype	ChMND Haplotype	fALS 1	fALS 2	Family A	Family B	fALS/FTD A Phased	fALS/FTD B Unphased	fALS/FTD C Phased	fALS D Unphased	fALS/FTD	FTD
rs1822723*	C		C/T	C/T	C	C	-	-	-	-	C	C
rs4879515*	T	T	T/C	T	T	T	-	-	-	-	C/T	T
rs868856*	T		T/C	T/C	T	T	-	-	-	-	C/T	T
rs7046653*	A		A	A/G	A	A	A	G/A	A	A	G/A	A
rs1977661 *	C		C	C/A	C	C	-	-	-	-	C	C
rs903603*	C		C/T	C/T	C	C/T	C	T/C	C	C	C	C
rs10812610*	C		C/A	C/A	C	C/A	-	-	-	-	C/A	C
rs2814707*	A	G	A/G	G	A	A/G	G	G/A	A	G/A	G/A	G
rs3849942*	A	G	A/G	A/G	A	A/G	A	G/A	A	G/A	G/A	G/A
rs700791	A	C	-	-	-	-	-	-	-	-	-	-
rs10122902*	G		G	G	G	G	-	-	-	-	G	G/A
rs10757665*	T	T	T/C	T	T	T	-	-	-	-	T	T/T
rs1565948*	G		G/A	G	G	G	-	-	-	-	G	G/A
rs774359*	C	T	C/T	C/T	C	C/T	C	T/C	C	T/C	C/T	C/T
rs2453554	T	C	-	-	-	-	-	-	-	-	-	-
rs2282241*	G	T	G	G/T	G	G/T	-	-	-	-	G/T	G/T
C9orf72 repeat expansion												
rs11789520	T	C	-	-	-	-	-	-	-	-	-	-
rs1948522*	C	C	C	C	C	C	-	-	-	-	C	C
rs1982915*	G	A	G/A	G/A	G	G/A	G	A/G	G	A/G	G/A	G/A
rs2453556*	G	A	G/A	G/A	G	G/A	G	A/G	G	A/G	G/A	G/A
rs702231*	A	A	A	A/C	A	A	C	C/A	A	A	A/C	A
rs696826*	G	G	G	G	G	G	-	-	-	-	G/C	G
rs2477518*	T	T	T	T	T	T	-	-	-	-	C/T	T

The table includes ALS cases carrying a *C9orf72* HRE and reported genotypes for SNPs in the region. ChMND haplotype is the haplotype from Table S3 that is shared by all 3 HRE carriers. SNPs that were homozygous in reported HRE carriers are recorded as a single allele. If phase was determined only the alleles from the haplotype associated with the HRE are shown. Both alleles are shown for heterozygous SNPs where phase was not determined. The blue and grey shading indicate SNPs that distinguish the European haplotype from that found in our ALS cases from China (1 and 3), green shading indicates an allele not present on either haplotype, and "–" indicates SNPs for which no genotypes were published.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table S5 Beta-values of quantifying methylation levels by microarray analysis in the Chinese cohorts.

ID	Position	TSS Position	Distance to TSS	Mean Beta value (Non-carriers)	Ch_MND_11134 Beta value	Ch_MND_8446 Beta value	Ch_MND_9059 Beta value
cg13958452	27571483	27573841	1997	0.93	0.92	0.89	0.98
cg05990720	27573649	27573841	192	0.04	0.10	0.04	0.05
cg23074747	27573816	27573841	25	0.03	0.28	0.03	0.06
cg11613875	27573888	27573841	-47	0.01	0.07	0.00	0.02
cg14363787	27573985	27573841	-144	0.05	0.27	0.06	0.10
cg14634447	27576471	27573841	-2630	0.85	0.83	0.82	0.85
cg13615109	27610825	27573841	-36984	0.62	0.56	0.63	0.66
cg13716323	27610876	27573841	-37035	0.95	0.95	0.95	0.94

TSS: transcription start site

Table S6. The distribution of HRE repeat sizes of Finnish 20-SNP haplotype vs other haplotypes in cases. Fisher's Exact test comparing the frequency of the repeat allele < 8 vs alleles of length 8 or greater (3 vs 1459 with fewer than 8 repeat alleles, and 26 vs 108 with 8 or more repeat allele, p = 0). The threshold of 8 was selected based on the repeat lengths associated with the Finnish haplotype in Europeans (Beck et al, 2013).

Repeat Size	**Finnish Haplotype**	**%**	**Other Haplotypes**	**%**	**All Haplotypes**
2	0	0.0	1016	64.8	1016
3	0	0.0	12	0.8	12
4	0	0.0	9	0.6	9
5	0	0.0	11	0.7	11
6	2	6.9	182	11.6	184
7	1	3.4	229	14.6	230
8	21	72.4	70	4.5	91
9	2	6.9	11	0.7	13
10	1	3.4	7	0.4	8
11	1	3.4	5	0.3	6
12	1	3.4	3	0.2	4
13	0	0.0	6	0.4	6
14	0	0.0	2	0.1	2
18	0	0.0	1	0.1	1
30	0	0.0	3	0.2	3
	29	100	1567	100	1596

We could not provide the similar comparisons in controls because the control samples with repeat size information were not genotyped for the 23 flanking SNPs.

ACCEPTED MANUSCRIPT

Table S7. The distribution of repeat size in cases who carry 23 SNPs haplotype that are shared between 2 Chinese cases with HRE. Fisher's Exact test comparing the frequency of the repeat allele < 8 vs alleles of length 8 or greater alleles (excluding the two HRE carriers, 62 vs 1398 with fewer than 8 repeat alleles, and 14 vs 118 with 8 or more repeat allele, $p = 4.6 \times 10^{-3}$ or $p=4.8\times10^{-4}$ after excluding the Finnish haplotype carriers). The threshold of 8 was selected based on the repeat lengths associated with the Finnish haplotype in Europeans (Beck et al, 2013).

Repeat Size	Chinese Haplotype	%	Other Haplotypes	%	All Cases
2	41	51.3	975	64.3	1016
3	2	2.5	10	0.7	12
4	0	0.0	9	0.6	9
5	1	1.3	10	0.7	11
6	9	11.3	175	11.5	184
7	11	13.8	219	14.4	230
8	8	10.0	83	5.5	91
9	1	1.3	12	0.8	13
10	0	0.0	8	0.5	8
11	1	1.3	5	0.3	6
12	2	2.5	2	0.1	4
13	2	2.5	4	0.3	6
14	0	0.0	2	0.1	2
18	0	0.0	1	0.1	1
30	2	2.5	1	0.1	3
	80	100	1516	100	1596

We could not provide the similar comparisons in controls because the control samples with repeat size information were not genotyped for the 23 flanking SNPs.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References:

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.Y., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72(2), 245-56. doi:10.1016/j.neuron.2011.09.011.

Jiao, B., Tang, B., Liu, X., Yan, X., Zhou, L., Yang, Y., Wang, J., Xia, K., Shen, L. 2014. Identification of C9orf72 repeat expansions in patients with amyotrophic lateral sclerosis and frontotemporal dementia in mainland China. Neurobiology of aging 35(4), 936 e19-22. doi:10.1016/j.neurobiolaging.2013.10.001.

Konno, T., Shiga, A., Tsujino, A., Sugai, A., Kato, T., Kanai, K., Yokoseki, A., Eguchi, H., Kuwabara, S., Nishizawa, M., Takahashi, H., Onodera, O. 2012. Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. Journal of neurology, neurosurgery, and psychiatry 84(4), 398-401. doi:10.1136/jnnp-2012-302272.

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R., Nalls, M.A., Heckerman, D., Tienari, P.J., Traynor, B.J. 2010. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet neurology 9(10), 978-85. doi:10.1016/S1474-4422(10)70184-8.

Ogaki, K., Li, Y., Atsuta, N., Tomiyama, H., Funayama, M., Watanabe, H., Nakamura, R., Yoshino, H., Yato, S., Tamura, A., Naito, Y., Taniguchi, A., Fujita, K., Izumi, Y., Kaji, R., Hattori, N., Sobue, G. 2012. Analysis of C9orf72 repeat expansion in 563 Japanese patients with amyotrophic lateral sclerosis. Neurobiology of aging 33(10), 2527 e11-6. doi:10.1016/j.neurobiolaging.2012.05.011.

Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M., Jones, A., Gray, I., Wright, J., Hortobagyi, T., Al-Sarraj, S., Rogelj, B., Powell, J., Lupton, M., Lovestone, S., Sapp, P.C., Weber, M., Nestor, P.J., Schelhaas, H.J., Asbroek, A.A., Silani, V., Gellera, C., Taroni, F., Ticozzi, N., Van den Berg, L., Veldink, J., Van Damme, P., Robberecht, W., Shaw, P.J., Kirby, J., Pall, H., Morrison, K.E., Morris, A., de Belleroche, J., Vianney de Jong, J.M., Baas, F., Andersen, P.M., Landers, J., Brown, R.H., Jr., Weale, M.E., Al-Chalabi, A., Shaw, C.E. 2013. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. European journal of human genetics : EJHG 21(1), 102-8. doi:10.1038/ejhg.2012.98.

Stephens, M., Smith, N.J., Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. American journal of human genetics 68(4), 978-89. doi:10.1086/319501.

Tsai, C.P., Soong, B.W., Tu, P.H., Lin, K.P., Fuh, J.L., Tsai, P.C., Lu, Y.C., Lee, I.H., Lee, Y.C. 2012. A hexanucleotide repeat expansion in C9ORF72 causes familial and sporadic ALS in Taiwan. Neurobiology of aging 33(9), 2232 e11- e18. doi:10.1016/j.neurobiolaging.2012.05.002.