Article

A novel genetic marker for the *C9orf72* repeat expansion in the Finnish population

Hannah **Rostalski**[a], Ville **Korhonen**[b], Teemu **Kuulasmaa**[c], Eino **Solje**[d,e], Johanna **Krüger**[f,g], FinnGen, Karri **Kaivola**[h], Per Kristian **Eide**[i], Jean-Charles **Lambert**[j], Valtteri **Julkunen**[d,e], Pentti J. **Tienari**[h], Anne M. **Remes**[f,g], Ville **Leinonen**[b], Mikko **Hiltunen**[c]*, Annakaisa **Haapasalo**[a]*

[a]A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, P. O. Box 1627 (Neulaniementie 2), 70211 Kuopio, Finland.

[b]Neurocenter, Neurosurgery, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland.

[c]Institute of Biomedicine, Yliopistonranta 1E, University of Eastern Finland, 70211 Kuopio, Finland.

[d]Institute of Clinical Medicine - Neurology, University of Eastern Finland.

[e]Neuro Center, Neurology, Kuopio University Hospital, Kuopio, Finland.

[f]Research Unit of Clinical Neuroscience, Neurology, University of Oulu, Oulu, Finland.

[g]Medical Research Center (MRC), Oulu University Hospital, Oulu, Finland.

[h]Department of Neurology, Helsinki University Hospital and Translational Immunology Program, Biomedicum, University of Helsinki, Helsinki, Finland.

[i]Oslo University Hospital-Rikshospitalet; and Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway.

[j]Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167-RID-AGE facteurs de risque et déterminants moléculaires des maladies liés au vieillissement, Lille, France.

*These authors contributed equally to this work

Corresponding authors:

*Annakaisa Haapasalo, PhD, Adjunct Professor, Research Director

Email**:** annakaisa.haapasalo@uef.fi; Tel: +358403552768

and

*Mikko Hiltunen, PhD, Professor

Email: mikko.hiltunen@uef.fi; Tel: +358403552014

Running title: A novel *C9orf72* repeat expansion-associated SNP

Disclosure of support for the work:

This work was supported by the Academy of Finland, under grant numbers 315459 (AH), 315460 (AMR), and 307866 (MH); the Strategic Neuroscience Funding of the University of Eastern Finland; Finnish Brain Foundation (ES), Sigrid Juselius Foundation (ES, MH), Instrumentarium Science Foundation (ES), Orion Research Foundation (ES). This publication is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 740264.

Abstract

Background: *C9orf72* repeat expansion ($C9_{exp}$) is the most common genetic cause underlying frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS). However, detection of the $C9_{exp}$ requires elaborative methods. **Objective:** Identification of $C9_{exp}$ carriers from genotyped cohorts could be facilitated by using single nucleotide polymorphisms (SNPs) as markers for the $C9_{exp}$. **Methods:** We elucidated the potential of the previously described Finnish risk haplotype, defined by the SNP rs3849942, to identify potential $C9_{exp}$ carriers among 218,792 Finns using the FinnGen database. The haplotype approach was first tested in an idiopathic normal pressure hydrocephalus (iNPH) patient cohort containing $C9_{exp}$ carriers by comparing intermediate (15-30) and full-length (> 60 repeats) $C9_{exp}$ carriers (n = 41) to $C9_{exp}$ negative patients (< 15 repeats, n=801). **Results:** In this analysis, rs3849942 was associated with carriership of $C9_{exp}$ (OR 8.44, $p < 2 \times 10^{-15}$), while the strongest association was found with rs139185008 (OR 39.4, $p < 5\times10^{-18}$). Unbiased analysis of rs139185008 in FinnGen showed the strongest association with FTLD (OR 4.38, 3×10^{-15}) and motor neuron disease ALS (OR 5.19, 3×10^{-21}). rs139185008 was the top SNP in all diseases (iNPH, FTLD, ALS). **Conclusion:** Our findings suggest that rs139185008 is a useful marker to identify potential $C9_{exp}$ carriers in the genotyped cohorts and biobanks originating from Finland.

Keywords: Amyotrophic lateral sclerosis; *C9orf72*; DNA Repeat Expansion; Frontotemporal lobar degeneration; Motor neuron disease; Polymorphism, Single Nucleotide

Introduction

Frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS) are neurodegenerative disorders sharing genetic and neuropathological similarities [1]. *C9orf72* hexanucleotide repeat expansion ($C9_{exp}$), the most common genetic cause of FTLD and ALS [2], [3], is exceptionally prevalent in Finnish FTLD and ALS patients [4], [5]. Previous studies have suggested that more than 30 units of the $C9_{exp}$ are pathogenic [3], [6]. However, even shorter, so called intermediate $C9_{exp}$ (10 units of repeats) may associate with disease [7]. Because it is not possible to sequence the expanded region using whole genome-sequencing, the presence and estimated length of the $C9_{exp}$ can only be determined using repeat-primed PCR, Southern blotting or long-read sequencing [8], [9]. The discovery of specific single nucleotide polymorphisms (SNPs) and groups of SNPs (haplotypes) associating with the $C9_{exp}$ would enable identification of potential $C9_{exp}$ carriers from large genotyped cohorts from which the $C9_{exp}$ cannot be detected using current methods. Previously, a Finnish risk haplotype of 42 single nucleotide polymorphisms (SNP) was reported to associate with ALS in Finland [10]. Moreover, another risk haplotype of 20 SNPs has been shown to associate with FTLD, ALS, and the $C9_{exp}$ in other European and U.S. cohorts [11]. Here, we aimed to identify a SNP that could be used as a genetic marker to identify $C9_{exp}$ carriers in Finnish cohorts. Our findings showed that the variant rs139185008 distinguishes $C9_{exp}$ carriers from non-carriers and associates with the clinical diagnoses of FTLD and ALS in the large population-based FinnGen database. This suggests that rs139185008 might be a powerful genetic marker for the identification of $C9_{exp}$ carriers in other Finnish cohorts as well.

Subjects and Methods

Cohorts, genotyping and imputations of EADB samples, and clinical endpoints

This study includes GWAS data from the European Alzheimer's Disease DNA BioBank (EADB) and the FinnGen database. EADB data were processed as previously described [12]. Finnish idiopathic normal pressure hydrocephalus (iNPH) patients included in the EADB GWAS were diagnosed according to published guidelines and procedures [13], [14]. $C9_{exp}$ genotyping was performed using repeat-primed PCR and amplicon length analysis [3]. The iNPH cohort contains 41 $C9_{exp}$ carriers [7 full-length (> 60 repeats) and 34 intermediate $C9_{exp}$ carriers (15-30 repeats)] and 801 controls (< 15 repeats). Forty-eight percent of $C9_{exp}$ carriers and controls were male.

Detailed information of the FinnGen data is described at https://www.finngen.fi/fi. Genome and clinical data from 218,792 individuals were obtained from FinnGen study data release 5. Clinical diagnoses were derived from the International Statistical Classification of Diseases and Related Health Problems, version 10 (ICD-10) codes in Finnish national hospital registries and cause-of-death registry as part of FinnGen project. UK Biobank data were used for validation of the identified SNPs and haplotypes.

Generation of risk haplotypes associating with $C9_{exp}$

Trans-Omics for Precision Medicine (TOPMed) imputed genotype data from EADB consortium was used [12]. In short, samples were genotyped by using Illumina Infinium Global Screening Array. The samples and variants passing the QC were used as the input of the imputation process. The imputation was performed by the Michigan Imputation Server [12] and TOPMed Freeze5 reference panel [12]. Genotypes were phased with Eagle v2.4105 and imputed with Minimac4 v4-1.0.2. Only SNPs having Hardy-Weinberg equilibrium $p > 10^{-5}$ and imputation quality greater than 0.6 were considered. The imputation quality score for rs139185008 was 0.75. The previously published 20-SNP Finnish risk haplotype [11] was used

to test for association with the $C9_{exp}$ (iNPH cohort) and clinical diagnoses ("motor neuron disease" for ALS, "circumscribed brain atrophy" for FTLD; FinnGen). Additional upstream and downstream SNPs were added to create larger haplotypes that were able to distinguish $C9_{exp}$ carriers from non-carriers. SNP selection was conducted based on inspection of individual $C9_{exp}$ carrier haplotypes of the phased and imputed most probable genotype data. Minor and major alleles included in the haplotypes are presented in Supplementary Table 1.

Analysis of SNP and haplotype association with $C9_{exp}$ and clinical endpoints

Both LD-statistics (D') and case *vs.* control logistic regression analysis with covariates were conducted on pre-processed imputed genotypes using PLINK software [version 1.9; [15]]. For iNPH cohort, only principal component (PC) 1-2, and for FinnGen PC1-5 were used as covariates. UK Biobank data were extracted through http://big.stats.ox.ac.uk/variant/9:27491942-T-C.

Data Presentation

Manhattan and regional association plots were drawn using LocusZoom software (v0.12.0). For LD calculation, European reference population was used. Images were modified using LibreOffice Draw (version: 6.0.2.1). Bar graphs and geographical plot of minor allele frequencies (MAFs) were generated, and Pearson's Chi-square test on minor *vs.* major allele counts among Finnish regions was performed using RStudio software (version: 1.1.463) and ggplot2 [16] and geofi packages [17].

Data availability

Data are available on reasonable request from the corresponding authors. Due to privacy policies, the data are not publicly available.

Ethics statement

All experimental procedures complied with the standards of the Declaration of Helsinki. The Ethics Committee of Hospital District of Northern Savo approved the iNPH study and all patients provided an informed consent. Patients and controls in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act (https://www.finngen.fi/fi). All DNA samples and data were pseudonymized (iNPH cohort and FinnGen cohort).

Results

$C9_{exp}$ associates with SNPs near MOB3B and C9orf72 genes

The most common form of hydrocephalus, iNPH, can be characterized by progressive gait impairment, cognitive decline, and loss of bladder control [18]. Recently, the $C9_{exp}$ was shown to be an important genetic etiology for iNPH [19]. Thus, based on genotype data obtained from a global screening array, SNP association analysis was performed in a well-characterized iNPH patient cohort, comprising intermediate (15-30) and full-length (> 60 repeats) $C9_{exp}$ carriers (n = 41) who were compared to non-carriers (< 15 repeats, n = 801). Previous studies have suggested a pathological threshold of >30 units [3], [6] or >45 units [20] for the $C9_{exp}$. However, intermediate repeats of <30 units may also associate with disease [4], [6], [7], [20]. In these studies, the minimum lengths of the $C9_{exp}$ among intermediate repeat expansion carriers have been identified as 7 [4] and 17 [7], [20] repeats on the longer allele. Here, we chose a threshold of 15 repeats to define individuals positive for the $C9_{exp}$. Except for two SNPs, all significantly $C9_{exp}$-associated SNPs ($p < 5\times10^{-8}$) were located on chromosome 9 (Figure 1A). Several of these were close or within the MOB kinase activator 3B (*MOB3B*) or *C9orf72* genes, spanning an approx. 94 kb region, and showed a strong linkage disequilibrium (LD) ($r^2 \geq 0.8$) with the reference SNP rs3849942 (Figure 1B), a previously reported surrogate marker for the chromosome 9p risk haplotype [2], [11]. Interestingly, rs139185008 (odds ratio, OR = 39.4, 95% CI [17.2-90.5], $p = 4.6\times10^{-18}$), localizing within a recombination-poor region 81 541 bp upstream of the $C9_{exp}$, showed the strongest single SNP association with $C9_{exp}$ carriership. Also, rs139185008 (MAF 0.016) was in complete LD (D' = 1.00) with the reference SNP rs3849942 (MAF 0.17), which showed a weaker association with $C9_{exp}$ carriership (OR 8.44, 95% CI [4.99-140.29], $p = 2.0\times10^{-15}$). Importantly, rs139185008 was highly abundant in $C9_{exp}$ carriers (minor allele frequency, MAF for full-length and intermediate carriers = 0.21 and 0.19, respectively), but rare in non-carriers (<15 repeats, MAF = 0.008). Several $C9_{exp}$-associated

haplotypes were significantly overrepresented in $C9_{exp}$ carriers as compared to non-carriers in the iNPH cohort (Table 1). rs139185008 was part of the haplotypes 2, 5, 8 and 10 showing the most prominent risk effects (OR > 42.0). Moreover, as compared to the previously reported 20-SNP Finnish risk haplotype, including e.g., rs868856, rs7046653, rs2814707, rs3849942, and rs774359 [11] (Figure 1B), the inclusion of rs139185008 to haplotypes ("haplo") 2, 5, 8 and 10 markedly improved the specificity to identify $C9_{exp}$ carriers from non-carriers in the iNPH cohort (Table 1, Supplementary Table 1). E.g., the OR for haplotype 2 (OR = 11.33, 95% CI [6.38-20.14], $p = 1.28\times10^{-16}$) substantially increased after the inclusion of rs139185008 (OR = 42.74, 95% CI [18.35-99.53], $p = 3.16\times10^{-18}$).

rs139185008 strongly associates with FTLD and ALS in FinnGen

Next, we unbiasedly examined in the FinnGen database which clinical diagnoses associate with the SNPs and haplotypes identified in the iNPH cohort. The FinnGen database contains comprehensive genome-wide genotype data and life-long medical history from >200,000 Finns. However, FinnGen does not include genetic data on complex genomic alterations, such as $C9_{exp}$. rs139185008 and haplotypes 2, 5, 8, and 10 containing the minor allele of rs139185008 strongly associated with ALS and FTLD (Table 1). In comparison, the previously reported 20-SNP Finnish risk haplotype [11] showed a weaker association with ALS and FTLD (haplo 2, Table 1). In FinnGen, rs139185008 was the top SNP that associated with ALS and FTLD, confirming the result obtained in the iNPH GWAS (Figure 1B). Importantly, rs139185008 also significantly associated with ALS in UK Biobank ($p = 9.0\times10^{-8}$), but it was not among the top SNPs associated with ALS (beta value = -0.4; p values < 4.0×10^{-19}).

rs139185008 is regionally enriched to South-Eastern Finland

Finally, we used FinnGen data to calculate the MAF of the rs139185008 according to the region of birth in Finland (Figure 2). Geographically, the rs139185008 minor allele showed the highest

prevalence in Southern Savonia (MAF = 0.025) and the lowest in Ostrobothnia (MAF = 0.008) (Figure 2). Pearson's Chi-square test of the frequency of rs139185008 minor allele revealed significant differences in the geographic distribution of rs139185008 in Finland ($p < 2.2\times10^{-16}$, $X^2 = 282.43$, df = 18).

Discussion

We report that rs139185008 strongly associates with $C9_{exp}$ in a cohort of iNPH cases, suggesting surrogate marker potential for identifying $C9_{exp}$ carriers in large population-based cohorts and biobank databases. rs139185008 indicated stronger association with FTLD and ALS clinical diagnoses in FinnGen (OR 4.4 and 5.2, respectively) as compared to the previously reported $C9_{exp}$ proxy marker rs3849942 (OR 1.2 and 1.6 respectively). The top SNPs differed in FinnGen and UK Biobank, which indicates that there are differences in the $C9_{exp}$ haplotype structures among European populations. In Finland, the frequency of rs139185008 minor allele was highest in South-Eastern Finland, and lowest in the west-coastal Ostrobothnia, which represent genetically different geographical regions. The regional distribution of rs139185008 in Finland is consistent with the most enriched areas of haplotypes of Finnish Heritage Diseases [21], a phenomenon traceable back to the population migration history within Finland and the resulting genetic isolation due to bottleneck events and founder effects. However, rs139185008 is also highly prevalent in panmictic Helsinki and surrounding areas [22] and showed a link to ALS in the UK Biobank, consisting of a more heterogeneous population. In this context, however, it should be emphasized that the beta-value provided by UK biobank for rs139185008 was negative, indicating an odds ratio below one for this SNP. Importantly, similar results (OR < 1) were also observed with some other SNPs significantly associated with ALS in the UK biobank in the *MOB3B*/*C9orf72* region. Thus, further investigations on the prevalence of rs139185008 and its association with $C9_{exp}$-linked diseases in other populations and cohorts are warranted in the future to evaluate its translational potential beyond Finland and the UK biobank.

Collectively, the present data suggest that specific haplotypes containing rs139185008 are useful proxy markers to identify potential $C9_{exp}$ carriers. Since gene-based therapies are emerging in $C9_{exp}$-linked diseases, the present findings may be utilized in the identification of potential carriers already at an early phase from biobanks and population cohorts for confirmatory $C9_{exp}$ genotyping and subsequent clinical trials in the future.

Acknowledgements

HR is a PhD student in the GenomMed and Molecular Medicine (DPMM) Doctoral Programs of the University of Eastern Finland, Kuopio, Finland. This study is part of the research activities of the Finnish FTD Research Network (FinFTD).

Conflict of Interest

Pentti J. Tienari holds a patent on C9orf72 in diagnostics and treatment of ALS/FTLD. The other authors have no conflicts of interest to disclose.

Funding

This work was supported by the Academy of Finland, under grant numbers 315459 (AH), 315460 (AMR), and 307866 (MH); the Strategic Neuroscience Funding of the University of Eastern Finland; Finnish Brain Foundation (ES), Sigrid Juselius Foundation (ES, MH, PJT), Finnish Cultural Foundation (KK) Instrumentarium Science Foundation (ES), Orion Research Foundation (ES). This work was supported by a grant (European Alzheimer DNA BioBank, EADB) from the EU Joint Programme – Neurodegenerative Disease Research (JPND). Inserm UMR1167 is also funded by Inserm, Institut Pasteur de Lille, the Lille Métropole Communauté Urbaine, the French government's LABEX DISTALZ program (development of innovative strategies for a transdisciplinary approach to Alzheimer's disease). This publication is part of a

233 project that has received funding from the European Union's Horizon 2020 research and
234 innovation programme under the Marie Skłodowska-Curie grant agreement No 740264.

235

236

237

References

[1] T. Van Langenhove, J. van der Zee and C. Van Broeckhoven, "The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum," *Ann. Med.,* vol. 44, *(8),* pp. 817-828, Dec, 2012.

[2] M. DeJesus-Hernandez *et al,* "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS," *Neuron,* vol. 72, *(2),* pp. 245-256, Oct 20, 2011.

[3] A. E. Renton *et al,* "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD," *Neuron,* vol. 72, *(2),* pp. 257-268, Oct 20, 2011.

[4] J. van der Zee *et al,* "A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats," *Hum. Mutat.,* vol. 34, *(2),* pp. 363-373, Feb, 2013.

[5] E. Majounie *et al,* "Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study," *Lancet Neurol.,* vol. 11, *(4),* pp. 323-330, Apr, 2012.

[6] S. Van Mossevelde *et al,* "Relationship between C9orf72 repeat size and clinical phenotype," *Curr. Opin. Genet. Dev.,* vol. 44, pp. 117-124, Jun, 2017.

[7] C. P. Cali *et al,* "C9orf72 intermediate repeats are associated with corticobasal degeneration, increased C9orf72 expression and disruption of autophagy," *Acta Neuropathol.,* vol. 138, *(5),* pp. 795-811, Nov, 2019.

[8] J. N. Cochran *et al,* "Genome sequencing for early-onset or atypical dementia: high diagnostic yield and frequent observation of multiple contributory alleles," *Cold Spring Harb Mol. Case Stud.,* vol. 5, *(6),* pp. 10.1101/mcs.a003491. Print 2019 Dec, Dec 13, 2019.

[9] M. T. W. Ebbert *et al,* "Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: implications for clinical use and genetic discovery efforts in human disease," *Mol. Neurodegener,* vol. 13, *(1),* pp. 46-018-0274-4, Aug 21, 2018.

[10] H. Laaksovirta *et al,* "Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study," *Lancet Neurol.,* vol. 9, *(10),* pp. 978-985, Oct, 2010.

[11] K. Mok *et al,* "Chromosome 9 ALS and FTD locus is probably derived from a single founder," *Neurobiol. Aging,* vol. 33, *(1),* pp. 209.e3-209.e8, Jan, 2012.

[12] C. Bellenguez *et al,* "New insights on the genetic etiology of Alzheimer's and related dementia," *medRxiv,* pp. 2020.10.01.20200659, 01/01, 2020.

[13] A. Junkkari *et al,* "The Kuopio idiopathic normal pressure hydrocephalus protocol: initial outcome of 175 patients," *Fluids Barriers CNS,* vol. 16, *(1),* pp. 21-019-0142-9, Jul 25, 2019.

[14] N. Relkin *et al,* "Diagnosing idiopathic normal-pressure hydrocephalus," *Neurosurgery,* vol. 57, *(3 Suppl),* pp. S4-16; discussion ii-v, Sep, 2005.

[15] C. C. Chang *et al*, "Second-generation PLINK: rising to the challenge of larger and richer datasets," *Gigascience*, vol. 4, pp. 7-015-0047-8. eCollection 2015, Feb 25, 2015.

[16] H. Wickham *et al*, Ed., *Ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag, 2016.

[17] M. Kainu *et al*, "geofi: Access Finnish Geospatial Data," 2015-2021.

[18] J. M Das and M. C. Biagioni, "Normal pressure hydrocephalus," in *StatPearls*Anonymous Treasure Island (FL): StatPearls Publishing LLC, 2021.

[19] V. E. Korhonen *et al*, "Prevalence of C9ORF72 Expansion in a Large Series of Patients with Idiopathic Normal-Pressure Hydrocephalus," *Dement. Geriatr. Cogn. Disord.*, vol. 47, *(1-2)*, pp. 91-103, 2019.

[20] K. Kaivola *et al*, "Carriership of two copies of C9orf72 hexanucleotide repeat intermediate-length alleles is a risk factor for ALS in the Finnish population," *Acta Neuropathologica Communications*, vol. 8, *(1)*, pp. 187, 11/09, 2020.

[21] A. R. Martin *et al*, "Haplotype Sharing Provides Insights into Fine-Scale Population History and Disease in Finland," *Am. J. Hum. Genet.*, vol. 102, *(5)*, pp. 760-775, May 3, 2018.

[22] A. de la Chapelle and F. A. Wright, "Linkage disequilibrium mapping in isolated populations: the example of Finland revisited," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, *(21)*, pp. 12416-12423, Oct 13, 1998.

Table 1. Haplotypes and individual SNPs significantly associating with *C9orf72* repeat expansion ($C9_{exp}$) in the iNPH cohort and ALS and FTLD in the FinnGen cohort

Haplotype/ SNP	iNPH cohort (*C9orf72* repeat expansion)				Motor neuron disease ALS (FinnGen)				Frontotemporal lobar degeneration (FinnGen)			
	OR (95% CI)	p	MAF in $C9_{exp}$ carriers	MAF in control	OR (95% CI)	p	MAF in disease	MAF in control	OR (95% CI)	p	MAF in disease	MAF in control
haplo 1	10.04 (5.79-17.39)	2.01×10^{-16}	0.5119	0.1038	1.68 (1.34-2.12)	8.40×10^{-06}	0.1933	0.1239	1.39 (1.09-1.77)	7.74×10^{-3}	0.1653	0.1248
haplo 2	11.33 (6.38-20.14)	1.28×10^{-16}	0.5000	0.0943	1.81 (1.43-2.28)	7.05×10^{-07}	0.1849	0.1110	1.49 (1.16-1.90)	1.8×10^{-3}	0.1570	0.1115
haplo 2*	42.74 (18.35-99.53)	3.16×10^{-18}	0.1905	0.0071	5.26 (3.72-7.42)	4.05×10^{-21}	0.0777	0.0156	4.41 (3.06-6.36)	1.97×10^{-15}	0.0641	0.0159
haplo 3	7.08 (4.04-12.41)	7.59×10^{-12}	0.3452	0.0802	1.94 (1.51-2.49)	2.72×10^{-07}	0.1534	0.0847	1.59 (1.22-2.08)	0.7×10^{-3}	0.1281	0.0857
haplo 4	7.08 (4.04-12.41)	7.59×10^{-12}	0.3452	0.0802	1.94 (1.51-2.5)	2.57×10^{-07}	0.1534	0.0846	1.59 (1.22-2.08)	6.77×10^{-4}	0.1281	0.0856
haplo 5	8.23 (4.64-14.63)	6.45×10^{-13}	0.3690	0.0820	2.12 (1.66-2.71)	1.65×10^{-09}	0.1660	0.0860	1.62 (1.24-2.11)	4.23×10^{-4}	0.1302	0.0862
haplo 5*	46.76 (19.71-110.9)	2.67×10^{-18}	0.1905	0.0065	5.12 (3.61-7.26)	5.40×10^{-20}	0.0756	0.0155	4.57 (3.19-6.56)	1.43×10^{-16}	0.0661	0.0158
haplo 6	8.23 (4.64-14.63)	6.45×10^{-13}	0.3690	0.0820	2.12 (1.66-2.71)	1.64×10^{-09}	0.1660	0.0860	1.62 (1.24-2.11)	4.23×10^{-4}	0.1302	0.0862
haplo 6*	46.76 (19.71-110.9)	2.67×10^{-18}	0.1905	0.0065	5.12 (3.61-7.26)	5.40×10^{-20}	0.0756	0.0155	4.57 (3.19-6.56)	1.43×10^{-16}	0.0661	0.0158
haplo 7	7.12 (4.07-12.47)	6.73×10^{-12}	0.3452	0.0796	1.94 (1.51-2.5)	2.40×10^{-07}	0.1534	0.0844	1.57 (1.2-2.05)	1.15×10^{-3}	0.1260	0.0854
haplo 8	9.48 (5.12-17.56)	8.87×10^{-13}	0.3333	0.0696	2.15 (1.66-2.79)	8.82×10^{-09}	0.1429	0.0714	1.74 (1.32-2.31)	9.32×10^{-5}	0.1178	0.0720
haplo 8*	46.76 (19.71-110.9)	2.67×10^{-18}	0.1905	0.0065	5.12 (3.61-7.27)	5.27×10^{-20}	0.0756	0.0155	4.42 (3.07-6.38)	1.71×10^{-15}	0.0641	0.0158
haplo 9	9.87 (5.32-18.32)	3.94×10^{-13}	0.3333	0.0666	2.23 (1.72-2.89)	1.85×10^{-09}	0.1429	0.0690	1.81 (1.37-2.39)	3.33×10^{-5}	0.1178	0.0697
haplo 10	7.68 (4.13-14.3)	1.24×10^{-10}	0.2500	0.0495	2.39 (1.83-3.13)	2.24×10^{-10}	0.1303	0.0589	1.86 (1.39-2.5)	3.17×10^{-5}	0.1054	0.0597
haplo 10*	46.51 (19.09-113.3)	2.88×10^{-17}	0.1786	0.0059	5.68 (3.98-8.1)	1.12×10^{-21}	0.0735	0.0137	4.84 (3.33-7.02)	1.13×10^{-16}	0.0620	0.0140
rs139185008	39.41 (17.17-90.49)	4.58×10^{-18}	0.1905	0.007665	5.19 (3.69-7.3)	2.57×10^{-21}	0.0798	0.0163	4.38 (3.05-6.28)	1.08×10^{-15}	0.0661	0.0166
rs3849942	8.44 (4.99-14.29)	2.04×10^{-15}	0.5595	0.1486	1.58 (1.28-1.95)	1.94×10^{-05}	0.2437	0.1698	1.22 (0.97-1.53)	8.53×10^{-2}	0.1983	0.1704

Notes: iNPH cohort $N_{carriers/control}$ = 41/801; Motor neuron disease ALS $N_{cases/control}$ = 238/111,855; Frontotemporal lobar degeneration $N_{cases/control}$ = 242/214,474; * denotes that variant chr9_27491944_T_C (rs139185008) was added to haplotype analysis; chromosomal positions of SNPs constituting haplotypes are listed in Supplementary Table 1

Abbreviations: CI, confidence interval; FTLD, frontotemporal lobar degeneration; haplo, haplotype; iNPH, idiopathic normal pressure hydrocephalus; MAF, minor allele frequency; N, number of subjects; OR, odds ratio; SNP, single-nucleotide polymorphism

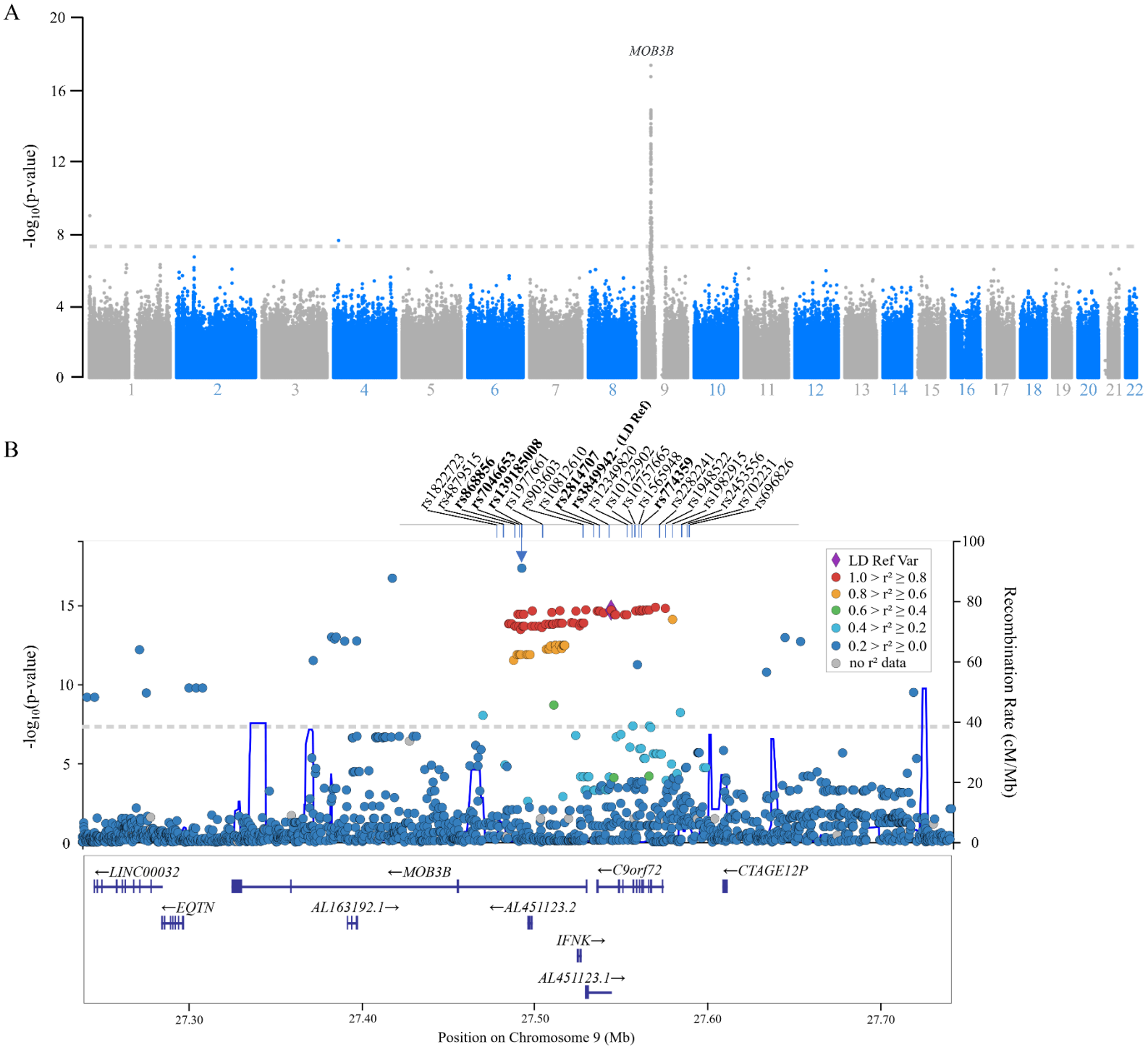
A
MOB3B
-log10(p-value)
20
16
12
8
4
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
B
rs1822723
rs4879515
rs868856
rs70466653
rs139185008
rs1977661
rs903603
rs10812610
rs2814707
rs38449942 (LD Ref)
rs12349820
rs10122902
rs10757665
rs15659948
rs774359
rs2382241
rs1948522
rs1982915
rs2453556
rs702231
rs696826
LD Ref Var
1.0 > r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0.0
no r² data
-log10(p-value)
Recombination Rate (cM/Mb)
←LINC00032
←MOB3B
←C9orf72
←CTAGE12P
←EQTN
AL163192.1→
←AL451123.2
IFNK→
AL451123.1→
27.30
27.40
27.50
27.60
27.70
Position on Chromosome 9 (Mb)

Figure 1: SNPs associating with the *C9orf72* repeat expansion in iNPH cohort locate near the *MOB3B* and *C9orf72* genes. Manhattan plot of genome-wide association (GWA) of SNPs associated with the *C9orf72* expansion in a Finnish iNPH cohort. Chromosome numbers are indicated below the x-axis (**A**). Regional association plot of chromosome 9 locus, which contained significant association from the GWA study. SNPs of the previously described Finnish risk haplotype [11] and rs139185008 (arrow) are indicated above the plot. Significantly associated SNPs are indicated in bold. Linkage disequilibrium is indicated as color-coded r^2 values. Recombination rates are depicted by continuous line. The reference variant rs3849942 is shown as a diamond. (**B**). Gray lines indicate significance level ($p < 5\times10^{-8}$). Abbreviations: iNPH, idiopathic normal pressure hydrocephalus; SNP, single nucleotide polymorphism

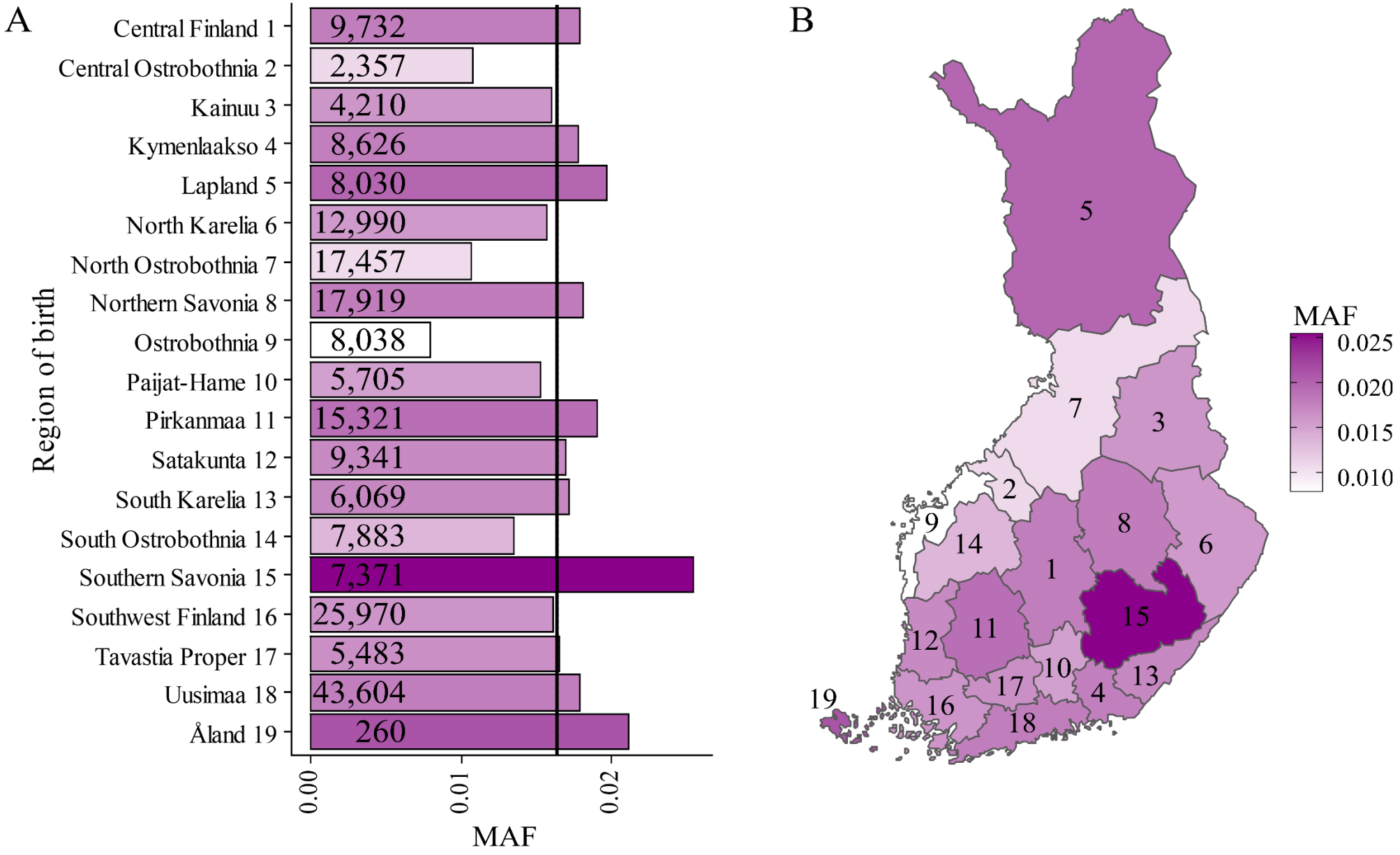
A
Central Finland 1
Central Ostrobothnia 2
Kainuu 3
Kymenlaakso 4
Lapland 5
North Karelia 6
North Ostrobothnia 7
Northern Savonia 8
Ostrobothnia 9
Paijat-Hame 10
Pirkanmaa 11
Satakunta 12
South Karelia 13
South Ostrobothnia 14
Southern Savonia 15
Southwest Finland 16
Tavastia Proper 17
Uusimaa 18
Åland 19
9,732
2,357
4,210
8,626
8,030
12,990
17,457
17,919
8,038
5,705
15,321
9,341
6,069
7,883
7,371
25,970
5,483
43,604
260
Region of birth
0.00
0.01
0.02
MAF
B
MAF
0.025
0.020
0.015
0.010
5
7
3
2
9
14
1
8
6
15
12
11
10
13
17
4
19
16
18

Figure 2: Geographical distribution of rs139185008 minor allele frequencies in different regions in Finland. Minor allele frequencies (MAF) of rs139185008 in Finland. Pearson's Chi-square test revealed statistically significant deviation ($p < 2.2\times10^{-16}$, $X^2 = 282.43$, df = 18) of the geographical distribution of the minor allele counts of rs139185008. Genotyped population sizes are given for each region. Mean MAF for all regions is indicated as black vertical line (**A**). MAF of rs139185008 within Finnish regions showed geographical clustering of high (dark magenta) and low (white) frequencies (**B**).

Supplementary file

Supplementary Table 1. List of SNPs included in haplotypes

Supplementary Information. List of FinnGen authors and their affiliation

Haplotypes listed in Table 1; SNPs which are included in haplotypes are marked with alternative allele (gray); "*" in haplotype name denotes that SNP rs139185008 was added to haplotype analysis

position	SNP	reference	alternative	**haplo 1**	**haplo 2**	**haplo 2***	**haplo 3**	**haplo 4**	**haplo 5**	**haplo 5***	**haplo 6**	**haplo 6***	**haplo 7**	**haplo 8**	**haplo 8***	**haplo 9**	**haplo 10**	**haplo 10***
27451484	rs4879507	C	T														C	C
27453329	rs10812599	C	T														C	C
27455825	rs10967945	C	T														C	C
27456930	rs7021930	G	A														G	G
27461738	rs76444167	A	G														A	A
27463312	rs62542379	G	A														G	G
27467672	rs10757663	G	A						A	A	A	A				A	A	A
27468264	rs61349511	A	G						A	A	A	A				A	A	A
27474216	rs10967952	T	C						T	T	T	T				T	T	T
27477876	rs1444533	C	T						T	T	T	T				T	T	T
27478054	rs1822723	C	T		C	C								C	C	C	C	C
27478711	rs4879514	T	C						C	C	C	C		C	C	C	C	C
27482237	rs4879515	C	T	T	T	T	T	T					T	T	T	T	T	T
27484498	rs10812602	A	G					A	A	A	A	A	A	A	A	A	A	A
27488094	rs17779457	T	G						G	G	G	G	G	G	G	G	G	G
27489253	rs868856	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
27490969	rs7046653	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
27491944	rs139185008	T	C			C				C		C			C			C
27495475		G	A								G	G		G	G	G	G	G
27502988	rs1977661	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
27508689	rs2166128	T	C						C	C	C	C	C	C	C	C	C	C
27510494	rs2477522	C	T					T	T	T	T	T	T	T	T	T	T	T
27513838	rs74439636	C	T						C	C	C	C	C	C	C	C	C	C
27521398	rs75741240	C	T						C	C	C	C	C	C	C	C	C	C
27527739	rs117257581	T	C						T	T	T	T	T	T	T	T	T	T
27529318	rs903603	G	A	G	G	G	G	G					G	G	G	G	G	G
27533986	rs10812610	C	A	C	C	C	C	C					C	C	C	C	C	C
27536399	rs2814707	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
27543283	rs3849942	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
27543384	rs3849943	C	T						C	C	C	C	C	C	C	C	C	C
27546892	rs13691	G	A						G	G	G	G	G	G	G	G	G	G
27549487	rs80067552	G	T						G	G	G	G	G	G	G	G	G	G
27553878	rs12349820	T	C	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
27556782	rs10122902	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
27556833	rs62538126	C	A						C	C	C	C	C	C	C	C	C	C
27557532	rs76925759	T	C						T	T	T	T	T	T	T	T	T	T
27557921	rs10757665	T	C	T	T	T	T	T					T	T	T	T	T	T
27559735	rs1565948	G	A	G	G	G	G	G					G	G	G	G	G	G
27561051	rs774359	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
27572257	rs2282241	C	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
27575787	rs1948522	C	T	C	C	C	C	C					C	C	C	C	C	C
27579562	rs1982915	A	G	G	G	G	G	G					G	G	G	G	G	G
27580676	rs12350076	A	C					C	C	C	C	C	C	C	C	C	C	C
27582805	rs17769370	C	T						C	C	C	C	C	C	C	C	C	C
27583130	rs7864502	G	C				C	C					C	C	C	C	C	C
27583296	rs702230	T	A				A	A					A	A	A	A	A	A
27583756	rs28522676	G	T				T	T					T	T	T	T	T	T
27583819	rs4879585	A	C						A	A	A	A	A	A	A	A	A	A
27584061	rs34555425	G	C				G											
27585699	rs36062268	G	A					G			G	G	G	G	G	G	G	G
27586164	rs2453556	A	G	G	G	G	G	G					G	G	G	G	G	G
27586447	rs7848063	G	A					A	A	A	A	A	A	A	A	A	A	A
27587790	rs3910852	T	C						T	T	T	T	T	T	T	T	T	T
27588733	rs702231	C	A	A	A	A	A	A					A	A	A	A	A	A
27589659	rs696826	A	G	G	G	G	G	G					G	G	G	G	G	G
27589698	rs112001877	T	C						T	T	T	T	T			T	T	T
27591128	rs9792688	T	A					T					T					
27601619	rs75747155	G	A														G	G

FINNGEN

Aarno Palotie[1], Mark Daly[1], Howard Jacob[2], Athena Matakidou[3], Heiko Runz[4], Sally John[4], Robert Plenge[5], Mark McCarthy[6], Julie Hunkapiller[6], Meg Ehm[7], Dawn Waterworth[7], Caroline Fox[8], Anders Malarstig[9], Kathy Klinger[10], Kathy Call10, Tim Behrens[11], Patrick Loerch[12], Tomi Mäkelä[13], Jaakko Kaprio[1], Petri Virolainen[14], Kari Pulkki[14], Terhi Kilpi[15], Markus Perola[15], Jukka Partanen[16], Anne Pitkäranta[17], Riitta Kaarteenaho[18], Seppo Vainio[18], Miia Turpeinen[18], Raisa Serpi[18], Tarja Laitinen[19], Johanna Mäkelä[19], Veli-Matti Kosma[20], Urho Kujala[21], Outi Tuovila[22], Minna Hendolin[22], Raimo Pakkanen[22], Jeff Waring[2], Bridget Riley-Gillis[2], Jimmy Liu[4], Shameek Biswas[5], Julie Hunkapiller[6], Dorothee Diogo[8], Catherine Marshall[9], Xinli Hu[9], Matthias Gossel[10], Robert Graham[11], Tim Behrens[11], Beryl Cummings[12], Samuli Ripatti[1], Johanna Schleutker[14], Mikko Arvas[16], Olli Carpén[17], Reetta Hinttala[18], Johannes Kettunen[18], Arto Mannermaa[20], Jari Laukkanen[21], Hilkka Soininen[23], Valtteri Julkunen[23], Anne Remes[23], Reetta Kälviäinen[23], Jukka Peltola[24], Pentti Tienari[25], Juha Rinne[26], Adam Ziemann[2], Jeffrey Waring[2], Sahar Esmaeeli[2], Nizar Smaoui[2], Anne Lehtonen[2], Susan Eaton[4], Sanni Lahdenperä[4], Janet van Adelsberg[5], Shameek Biswas[5], John Michon[6], Geoff Kerchner[6], Natalie Bowers[6], Edmond Teng[6], John Eicher[8], Vinay Mehta[8], Padhraig Gormley[8], Kari Linden[9], Christopher Whelan[9], Fanli Xu[7], David Pulford[7], Martti Färkkilä[25], Sampsa Pikkarainen[25], Airi Jussila[27], Timo Blomster[28], Mikko Kiviniemi[29], Markku Voutilainen[26], Bob Georgantas[2], Graham Heap[2], Fedik Rahimov[2], Keith Usiskin[5], Tim Lu[6], Danny Oh[6], Kirsi Kalpala[9], Melissa Miller[9], Linda McCarthy[7], Kari Eklund[25], Antti Palomäki[26], Pia Isomäki[27], Laura Pirilä[26], Oili Kaipiainen-Seppänen[29], Johanna Huhtakangas[28], Bob Georgantas[2], Fedik Rahimov[2], Apinya Lertratanakul[2], Marla Hochfeld[5], Kirsi Kalpala[9], Nan Bing[9], Jorge Esparza Gordillo[7], Nina Mars[1], Margit Pelkonen[29], Paula Kauppi[25], Hannu Kankaanranta[24], Terttu Harju[28], David Close[3], Steven Greenberg[5], Hubert Chen[6], Jo Betts[7], Soumitra Ghosh[7], Veikko Salomaa[30], Teemu Niiranen[30], Markus Juonala[26], Kaj Metsärinne[26], Mika Kähönen[27], Juhani Junttila[28], Markku Laakso[23], Jussi Pihlajamäki[23], Juha Sinisalo[25], Marja-Riitta Taskinen[25], Tiinamaija Tuomi[25], Ben Challis[3], Andrew Peterson[6], Audrey Chu[8], Jaakko Parkkinen[9], Melissa Miller[9], Anthony Muslin[10], Dawn Waterworth[7], Heikki Joensuu[25], Tuomo Meretoja[25], Lauri Aaltonen[25], Johanna Mattson[25], Annika Auranen[24], Peeter Karihtala[28], Saila Kauppila[28], Päivi Auvinen[23], Klaus Elenius[26], Relja Popovic[2], Jennifer Schutzman[6], Andrey Loboda[8], Aparna Chhibber[8], Heli Lehtonen[9], Stefan McDonough[9], Marika Crohns[10], Diptee Kulkarni[7], Kai Kaarniranta[23], Joni A Turunen[25], Terhi Ollila[25], Sanna Seitsonen[25], Hannu Uusitalo[24], Vesa Aaltonen[26], Hannele Uusitalo-Järvinen[24], Marja Luodonpää[28], Nina Hautala[28], Stephanie Loomis[4], Erich Strauss[6], Hao Chen[6], Anna Podgornaia[8], Joshua Hoffman[7], Kaisa Tasanen[28], Laura Huilaja[28], Katariina Hannula-Jouppi[25], Teea Salmi[27], Sirkku Peltonen[25], Leena Koulu[25], Ilkka Harvima[23], Kirsi Kalpala[9], Ying Wu[9], David Choy[6], Fedik Rahimov[2], Dawn Waterworth[7], Pirkko Pussinen[25], Aino Salminen[25], Tuula Salo[25], David Rice[25], Pekka Nieminen[25], Ulla Palotie[25], Maria Siponen[23], Liisa Suominen[23], Päivi Mäntylä[23], Ulvi Gursoy[26], Vuokko Anttonen[28], Kirsi Sipilä[28], Justin Wade Davis[2], Bridget Riley-Gillis[2], Danjuma Quarless[2], Fedik Rahimov[2], Sahar Esmaeeli[2], Slavé Petrovski[3], Eleonor Wigmore[3], Chia-Yen Chen[4], Paola Bronson4, Ellen Tsai[4], Yunfeng Huang[4], Joseph Maranville[5], Elmutaz Shaikho Elhaj Mohammed[5], Samir Wadhawan[31], Erika Kvikstad[31], Minal Caliskan[31], Diana Chang[6], Tushar Bhangale[6], Natalie Bowers[6], Sarah Pendergrass[6], Emily Holzinger[8], Xing Chen[9], Åsa Hedman[9], Karen S King[7], Clarence Wang[10], Ethan Xu[10], Franck Auge[10], Clement Chatelain[10], Deepak Rajpal[10], Dongyu Liu[10], Katherine Call[10], Tai-he Xia[10], Matt Brauer[11], Mitja Kurki[1], Samuli Ripatti[1], Juha Karjalainen[1], Aki Havulinna[1], Anu Jalanko[1], Priit Palta[1], Pietro della Briotta Parolo[1], Wei Zhou[32], Susanna Lemmelä[1], Manuel Rivas[33], Jarmo Harju[1], Arto Lehisto[1], Andrea Ganna[1], Vincent Llorens[1], Hannele Laivuori[1], Sina Rüeger[1], Mari E Niemi[1], Taru Tukiainen[1], Mary Pat Reeve[1], Henrike Heyne[1], Nina Mars[1], Kimmo Palin[34], Javier Garcia-Tabuenca[35], Harri Siirtola[35], Tuomo Kiiskinen[1], Tuomo Kiiskinen[1], Jiwoo Lee[1], Kristin Tsuo[1], Amanda Elliott[1], Kati Kristiansson[15], Kati Hyvärinen[36], Jarmo Ritari[36], Miika Koskinen[17], Katri Pylkäs[18], Marita Kalaoja[18], Minna Karjalainen[18], Tuomo Mantere[18], Eeva Kangasniemi[19], Sami Heikkinen[20], Sami Heikkinen[21], Eija Laakkonen[21], Csilla Sipeky[37], Samuel Heron[37], Antti Karlsson[14], Dhanaprakash Jambulingam[37], Venkat Subramaniam Rathinakannan[37], Anu

Jalanko[1], Risto Kajanne[1], Mervi Aavikko[1], Manuel González Jiménez[1], Mitja Kurki[1], Juha Karjalainen[1], Pietro della Briotta Parola[1], Sina Rüeger[1], Arto Lehistö[1], Masahiro Kanai[32], Hannele Laivuori[1], Aki Havulinna[1], Susanna Lemmelä[1], Tuomo Kiiskinen[1], Mari Kaunisto[1], Jarmo Harju[1], Elina Kilpeläinen[1], Timo P. Sipilä[1], Georg Brein[1], Ghazal Awaisa[1], Anastasia Shcherban[1], Kati Donner[1], Timo P. Sipilä[1], Anu Loukola[17], Päivi Laiho[15], Tuuli Sistonen[15], Essi Kaiharju[15], Markku Laukkanen[15], Elina Järvensivu[15], Sini Lähteenmäki[15], Lotta Männikkö[15], Regis Wong[15], Hannele Mattsson[15], Kati Kristiansson[15], Susanna Lemmelä[1], Tero Hiekkalinna[15], Teemu Paajanen[15], Priit Palta[1], Kalle Pärn[1], Harri Siirtola[35], Javier Gracia-Tabuenca[35]

[1]Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland. [2]Abbvie, Chicago, IL, United States. [3]Astra Zeneca, Cambridge, United Kingdom. [4]Biogen, Cambridge, MA, United States. [5]Celgene, Summit, NJ, United States. [6]Genentech, San Francisco, CA, United States. [7]GlaxoSmithKline, Brentford, United Kingdom. [8]Merck, Kenilworth, NJ, United States. [9]Pfizer, New York, NY, United States. [10]Sanofi, Paris, France. [11]Maze Therapeutics, San Francisco, CA, United States. [12]Janssen Biotech, Beerse, Belgium. [13]HiLIFE, University of Helsinki, Finland, Finland. [14]Auria Biobank / University of Turku / Hospital District of Southwest Finland, Turku, Finland. [15]THL Biobank / The National Institute of Health and Welfare Helsinki, Finland. [16]Finnish Red Cross Blood Service / Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. [17]Helsinki Biobank / Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [18]Northern Finland Biobank Borealis / University of Oulu / Northern Ostrobothnia Hospital District, Oulu, Finland. [19]Finnish Clinical Biobank Tampere **/** University of Tampere / Pirkanmaa Hospital District, Tampere, Finland. [20]Biobank of Eastern Finland / University of Eastern Finland / Northern Savo Hospital District, Kuopio, Finland. [21]Central Finland Biobank / University of Jyväskylä / Central Finland Health Care District, Jyväskylä, Finland. [22]Business Finland, Helsinki, Finland. [23]Northern Savo Hospital District, Kuopio, Finland. [24]Pirkanmaa Hospital District, Tampere, Finland. [25]Hospital District of Helsinki and Uusimaa, Helsinki, Finland. [26]Hospital District of Southwest Finland, Turku, Finland. [27]Pirkanmaa Hospital District, Tampere, Finland. [28]Northern Ostrobothnia Hospital District, Oulu, Finland. [29]Northern Savo Hospital District, Kuopio, Finland. [30]The National Institute of Health and Welfare Helsinki, Finland. [31]Bristol-Meyers-Squibb. [32]Broad Institute, Cambridge, MA, United States. [33]University of Stanford, Stanford, CA, United States. [34]University of Helsinki, Helsinki, Finland. [35]University of Tampere, Tampere, Finland. [36]Finnish Red Cross Blood Service, Helsinki, Finland. [37]University of Turku, Turku, Finland.