



HAL
open science

Genes containing hexanucleotide repeats resembling C9ORF72 and expressed in the central nervous system are frequent in the human genome

Patrick Vourc'h, François Wurmser, Céline Brulard, Kevin Mouzat, Sandra Kassem, Audrey Dangoumau, Frédéric Laumonier, H  l  ne Blasco, Philippe Corcia, Christian R. Andres

► To cite this version:

Patrick Vourc'h, Fran  ois Wurmser, C  line Brulard, Kevin Mouzat, Sandra Kassem, et al.. Genes containing hexanucleotide repeats resembling C9ORF72 and expressed in the central nervous system are frequent in the human genome. *Neurobiology of Aging*, 2021, 97, pp.148.e1 - 148.e7. 10.1016/j.neurobiolaging.2020.07.027 . hal-03493473

HAL Id: hal-03493473

<https://hal.science/hal-03493473>

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin  e au d  p  t et    la diffusion de documents scientifiques de niveau recherche, publi  s ou non,   manant des   tablissements d'enseignement et de recherche fran  ais ou   trangers, des laboratoires publics ou priv  s.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Genes containing hexanucleotide repeats resembling *C9ORF72* and expressed in the central nervous system are frequent in the Human genome

Vourc'h Patrick^{1,2}, Wurmser François¹, Brulard Céline^{1,2}, Mouzat Kevin³, Kassem Sandra², Dangoumau Audrey¹, Frédéric Laumonnier¹, Blasco Hélène^{1,2}, Corcia Philippe^{1,4}, Andres Christian R.^{1,2}

¹ UMR 1253, Université de Tours, Inserm, Tours, France

² CHRU de Tours, Laboratoire de Biochimie et de Biologie moléculaire, Tours, France

³ CHRU de Nîmes, Laboratoire de Biochimie, Nîmes, France

⁴ CHRU de Tours, Service de Neurologie, Tours, France

*Corresponding author P Vourc'h, UMR 1253 iBRAIN Université de Tours, Inserm, Team Genomics and Neuronal physiopathology, 10 bd Tonnellé, 37032 Tours, France. Tel: 33 (0)2 34 37 89 10; fax: 33 (0)2 47 36 61 85.

E-mail address: vourch@med.univ-tours.fr (P. Vourc'h)

Keywords: Repeats, Neurodegenerative, Neurodevelopmental, Ubiquitin, ALS

ABSTRACT

More than 40 Human diseases, mainly diseases affecting the central nervous system, are caused by the expansion of unstable nucleotide repeats. Repeats of sequences like (CAG)_n present in different genes, can be responsible for various diseases of the central nervous system. An expanded hexanucleotide repeat (GGGGCC)_n in the *C9ORF72* gene has been characterized as the most frequent genetic cause of Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Lobar Dementia (FTLD). In this study, we performed a genome-wide analysis in the Human genome and identified 74 genes containing this precise hexanucleotide repeat, with a preference for a location in exon 1 or intron 1, similar to the *C9ORF72* gene. 36 of these 74 genes may be of interest as candidates in neurodevelopmental or neurodegenerative diseases, based on their function.

1. INTRODUCTION

Several human diseases, and particularly neurological and neuromuscular diseases, are caused by the expansion of nucleotide repeats named microsatellites in different genes (Nelson et al., 2013). These DNA repeats seem to be predisposed to such expansion, certainly because of unusual structural features. They are located in various regions of the genes, including promoters, 5'-or 3'-untranslated regions (UTR), coding sequences, and introns (Mirkin, 2007). For example, the exonic triplet repeats (CAG)_n described in several genes encode expanded polyglutamine-repeats in proteins. This expansion is implicated in Huntington's disease (HD), spinal and bulbar muscular atrophy (SBMA), and spinocerebellar ataxia (He et al., 2010). The exonic triplet repeats (GCN)_n also reported in several genes encode expanded polyalanine-repeats in proteins. Intronic triplet repeats such as (GAA)_n are observed in Friedreich's ataxia (FRDA) for example. 5'UTR triplet repeats such as (CGG)_n and 3'UTR triplet repeats such as (CTG)_n are implicated in Fragile-X mental retardation (FRAXA) and myotonic dystrophy type 1 respectively. Other expansions are reported in myotonic dystrophy type 2 (CCTG)_n, spinocerebellar ataxia type 10 (ATTCT)_n and type 36 (GGCCTG)_n.

An hexanucleotide repeat (HR) expansion (GGGGCC)_n located in the first intron of the *C9ORF72* gene was identified as the most frequent mutation in Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Lobar Degeneration (FTLD) (DeJesus-Hernandez et al., 2011; Gijssels et al., 2012; Renton et al., 2011). Non-expanded repeat size ranges from 2 to 22 in the normal population. It generally largely exceeds 30 repeats in pathological populations. The (GGGGCC)_n repeat is located in the first intron corresponding to two

transcripts (1 and 3) of *C9ORF72*, and in the 5'UTR corresponding to a third transcript (2) (DeJesus-Hernandez et al., 2011). Mutations in *C9ORF72* induce neurodegeneration via loss-of-function and/or gain-of-function mechanisms (Martin et al., 2020). Indeed, the expansion was associated with a reduction in *C9ORF72* gene expression (van der Zee et al., 2013), suggesting a haploinsufficiency mechanism. *C9ORF72* may have a role in autophagy and vesicular trafficking (Levine et al., 2013; Zhang et al., 2012). Recent studies indicated that it is as a guanine nucleotide exchange factor (GEF) that regulates specific Rab GTPases (Iyer et al., 2018). Two other disease mechanisms, both by toxic gain of function, have been proposed (Gendron et al., 2014, Haessler et al., 2014). They are based on the observation that the expansion is also associated with a production of both sense and antisense transcribed expanded repeats ((5'GGGGCC3')_n and (5'GGCCCC3')_n), which form RNA-G-quadruplex structures aberrantly interacting with various RNA-binding proteins (RBPs) and normal RNA in cells, and forming RNA foci (DeJesus-Hernandez et al., 2011, Gendron et al., 2013). These repeat-containing RNAs can also be the site of a translation leading to the production of toxic dipeptide-repeat containing proteins (Ash et al., 2013; Mori et al., 2013). This mechanism was first described in trinucleotide repeat expansions in spinocerebellar ataxia 8 and myotonic dystrophy type 1 (Zu et al., 2011). The use of antibodies generated against HR-translated peptides detected neuronal inclusions throughout the central nervous system of patients with ALS/FTLD (Ash et al., 2013). These dipeptide-repeat containing proteins could be associated to various cellular defects through mechanisms such as heterochromatin anomalies (Zhang et al., 2019). Interestingly, recent works indicate that the mechanisms of loss of function and gain of function could act in a combined manner. Indeed, HR expansion in *C9ORF72* is associated with reduced autophagy, which results in less efficient elimination of DPR proteins. This same HR expansion is also responsible for a production of DPR proteins, which

are toxic to cells (Boivin et al., 2020).

In this paper, we hypothesized that the HR 5'GGGGCC3'_n or its antisense sequence 5'GGCCCC3'_n, encoding RNA-G-quadruplex structures and/or expanded poly dipeptide-repeat proteins, could be present not only in *C9ORF72*, but also in other genes in the Human genome. Similarly to what was observed in ALS/FTDP with *C9ORF72*, these repeats could be subject to expansion and cause neurological diseases. We analysed the presence and distribution of these HRs in the Human genome. Based on data on gene expression and function, we established a list of genes that would be of interest for repeat-primed polymerase chain reaction (PCR) studies in neurological and neuromuscular diseases.

2. METHODS

The genome-wide analysis was done using the human genome assembly hg19. Only perfect repeats of three or more repeats were searched using an *ad hoc* pipeline. The precise location of the HR (GGGGCC or GGCCCC) in genomic regions such as promoters, introns and exons was identified using the genome browsers Ensembl and UCSC (Flicek et al., 2014; Kent et al., 2002). We assessed if this HR pattern was statistically overrepresented in the human genome using a Monte Carlo test with 1,000 iterations.

The expression and function of genes with HR were analysed using the gene annotation portal bioGPS (bioGPS.org) and Medline database (pubmed.org). Variant Call Format (VCF) files extracted for each gene from 1000 Genome database (<http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>) were used to search

polymorphisms describes at repeat positions. Pathway enrichment analysis was performed with GORILLA software (Gene Ontology enRIchment anaLysis and visuaLizAtion tool : <http://cbl-gorilla.cs.technion.ac.il/>).

3. RESULTS

3.1. Loci with repeats in the human genome.

A genome-wide analysis was performed to identify and locate all GGGGCC and GGCCCC HRs in the Human genome. The number of repeats of the hexanucleotide GGGGCC in *C9ORF72* in the human genome of reference in Ensembl is 3. We therefore set the lower limit at 3 repetitions in our analysis. We found the repetitions GGGGCC or GGCCCC in 101 loci in the Human genome. These repeats were not over-represented in the genome compared with randomly chosen six-base repeats (p-value: 0.096, Monte Carlo test with 1,000 iterations). Repeats were observed on all chromosomes except chromosome 14 and Y (Figure 1A). The average number of repeats per chromosome was 4.6. 22 intergenic regions contained the repeats. They were considered intergenic sequences because of their location at more 1000 bp of a coding sequence. We considered in our analysis only perfect HR, (GGGGCC GGGGCC GGGGCC) or (GGCCCC GGCCCC GGCCCC). These perfect repeats were often surrounded by very similar sequences, as for example GGGGCG in *C9ORF72*.

3.2. Genes with repeats in the Human genome.

We next focused on repeats within genes in the Human genome. We observed the repeats (GGGGCC)_n or (GGCCCC)_n, with $n \geq 3$ in 74 genes distributed over all

chromosomes, except chromosome 14, 21 and Y (Supplementary Table 1). We identified the (GGGGCC)_n repeat in intron 1 of *C9ORF72* gene, validating our bioinformatic analysis. For 15 genes, repeats were located in proximity (<1000pb) of 5' regions of coding sequences, therefore in promoters (Figure 1B). For 64 genes, repeats were in RNA coding sequences in translated or untranslated regions, such as UTRs (Supplementary Table 1). They were located in intronic sequences for 31 genes (42%). We observed the HR mostly in intron 1 (23 genes) compared to other introns, as in *C9ORF72* involved in ALS/DFT, but this was mainly due to the large size of this first intron in many genes. Indeed the repetitions were not more frequent in the intron 1 than in other regions of the genes if we take into account the length of exons and introns (p=0.48).

23 genes (31%) contained the HRs in exons, and again mostly in the first exon (20 genes) (Supplementary Table 1). For 7 of these genes the repeats were in the translated coding sequence, and in a hydrophobic region of the protein. This region was formed of perfect repeats and surrounded by sequences highly similar to the repeats. For example, the coding sequence of the *TAF4* gene contains 5 perfect HR surrounded by three highly similar sequences, encoding a hydrophobic domain of 8 repeats of the dipeptide Glycine-Proline in the protein. These 7 genes were *STK39* (Ala-Pro)₄, *FGF2* (Gly-Arg)₃, *RRP36* (Ala-Gly)₄, *CDKN1C* (Ala-Pro)₈, *PAPD5* (Ala-Pro)₃, *RBM12* (Gly-Pro)₆ and *TAF4* (Gly-Pro)₈.

3.3. Genes with repeats expressed in the nervous and neuromuscular systems

From the list of 74 genes we extracted all genes expressed in the central nervous system (CNS) or in muscles (bioGPS.org), and with known functions (Table 1) (Wu et al., 2009). The 36 selected genes were located on 18 chromosomes, with 5 genes in 19q13. VCF

files extracted from the 1000 genomes database and corresponding to these 36 genes showed two genes with a different number of HR compared to the reference genome. *RRP36* lost 2 out of 4 repeats in the reference genome in 27% of cases, and *ADGRA1* lost 1 out of 4 repeats in the reference genome in 29% of cases.

Given this set of 36 genes, a GO analysis found over-representations in four pathways with non-corrected p value between $1.05e^{-5}$ and $4.31e^{-5}$: regulation of ion transmembrane transport, calcium ion import and membrane depolarization during action potential, and protein linear polyubiquitination (Table 2). Many genes in this set associated with ion transport and membrane depolarization are involved in neuronal or muscle excitability. For example *SNC4A* encodes a skeletal muscle sodium channel implicated in muscle disorders characterized by episodic abnormalities of membrane excitability (McClatchey et al., 1992). Several other genes encode proteins of the Ubiquitin pathway, such as *SHARPIN* and *RBCK1* also named *HOIL1*, involved in linear polyubiquitination. The protein SHARPIN (SHANK-associated RH domain-interacting protein), in association RBCK1 and HOIP, forms a complex called LUBAC (linear ubiquitin chain assembly complex) which regulates the NF κ B pathway and plays a major role in immune and inflammatory responses (Tokunaga et al., 2012). A truncating mutation was identified in *RBCK1* in a family with a novel Mendelian disease with neuromuscular and cardiac involvement (Wang et al., 2013). Other genes involved in the Ubiquitin pathway are present in the list, such as *ARID1B* encoding a protein of an E3 ubiquitin ligase complex member of the SWI/SNF-A chromatin-remodelling complex and *HUWE1* producing an E3 ubiquitin ligase regulating axon branching. The haploinsufficiency of *ARID1B* impairs interneuron development in mouse and causes Intellectual Disability (Li et al., 2010; Hoyer et al., 2012; Jung et al., 2017). Duplication or

point mutations in *HUWE1* contribute to Intellectual Disability (Froyen et al., 2008; Vandewalle et al., 2013).

Others genes of interest in the CNS are present in this list of 36 genes. The protein product of *TFDPI* negatively regulates the neuroprotective function of Cdk5 kinase in mature neurons (Zhang and Herrup, 2011). *CDKN1C* negatively controls the expansion of cortical intermediate progenitor cells. It is directly regulated by *ARX*, a gene whose mutations are found in disorders with epilepsy and intellectual disability (Colasante et al., 2013). Together these data support the idea that other genes than *C9ORF72* could be candidate genes to be tested in diseases affecting the CNS.

4. DISCUSSION

Instability characterized by the expansion of simple DNA repeats are involved in more than 40 human hereditary diseases to date, and the list is still growing (Mirkin, 2007; Paulson, 2018). A particular repeat expansion such as $(CAG)_n$ resulting in polyglutamine tract in proteins was observed in many genes, and was implicated in several diseases. One particular observations is the expansion of a HR in *C9ORF72* and its implication in ALS and FTD (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Gijssels et al., 2012). Several disease mechanisms associated with this particular repeat expansion have emerged. They include RNA toxicity and abnormal translation of expanded repeat sequences (poly-dipeptides). Here, we present a list 74 genes in the human genome that contain a HR (GGGGCC or reverse sequence), with many of them expressed in the central nervous system or in muscles and representing interesting candidates in neurological or neuromuscular diseases.

The mechanism of instability of this type of HR expansions is still to be characterized. Studies on *C9ORF72* support the idea that a *de novo* pathological expansion does not exist (Beck et al., 2013; Ratti et al., 2012; Smith et al., 2013). Further studies will indicate whether or not the HRs in *C9ORF72* are instable and if intermediate repeat length may be pathological (Byrne et al., 2014; Ng et al., 2017, Cali et al., 2019). Interestingly, small deletions appear significantly enriched in the region directly flanking the *C9ORF72* repeat in carriers of expanded HR (van der Zee et al., 2013). The consequence is an extension of the perfect repeats (GGGGCC) surrounded with imperfect ones, and consequently a higher vulnerability to expansion by strand slippage for example. In our list of 74 genes, we observed that many genes contained GC-rich low complexity sequences and imperfect repeats flanking the perfects ones.

Disorders of the nervous system characterized by expansion of a nucleotide repeat account for a number of neurodevelopmental and neurodegenerative diseases. It is interesting to note that our screening identified several genes already involved in neurological diseases. Abnormalities in these genes consist mainly in point mutations and copy number variants. For example abnormalities in *HUWE1* (E3 ubiquitin ligase) and *ARID1B* (epigenetic modifier of chromatin structure) are associated with intellectual disability (Backx et al., 2011; Froyen et al., 2008; Santen et al., 2012). In patients with epilepsy, mutations are observed in *CACNA1H* and *HCN2* encoding respectively the Cav3.2 T-type Ca²⁺ channel and the hyperpolarization-activated cyclic nucleotide-gated channel 2 (Chen et al., 2003; Eckle et al., 2013; Nakamura et al., 2013; Tang et al., 2008). Mutations in the adult skeletal muscle sodium channel gene *SCN4A* are responsible for periodic paralysis and myotonic disorders

(Ptáček et al., 1991; Yoshinaga et al., 2012). The presence of HRs in these genes, often in their 5' region such as in *C9ORF72*, suggests that it could be of interest to look for their possible expansion in patients with these clinical signs of diseases but without classical mutations in these genes.

We observed a HR in 29 other genes non-genetically involved in diseases so far. Several of them are interesting candidates for diseases of the nervous system based on their function. Recent studies have shown that the protein of the ubiquitin pathway NUB1, known to suppress Lewy body-like inclusions, is a suppressor of mutant huntingtin toxicity and reduces tau aggregation (Lu et al., 2013; Richet et al., 2012; Tanji et al., 2006). MARCKSL1 is highly expressed in spinal cord and encodes the myristolated alanine-rich C kinase substrate implicated in cytoskeletal plasticity during nerve degeneration (Björkblom et al., 2012; McNamara and Lenox, 2000). *RRP36* and *TTP* are two other interesting genes because of their direct implication in RNA homeostasis, another major theme in neurodegenerative diseases (Renoux and Todd, 2012). *TTP* encodes an AU-rich element binding protein that regulates the stability of many mRNA and is present in stress granules in tauopathies (Vanderweyde et al., 2012). *RRP36* encodes the Ribosomal RNA processing 36 homolog involved in early cleavage of pre-rRNA (Gérus et al., 2010). *ADGRA1* encodes a post-synaptic protein (Gérus et al., 2010; Pandia et al., 2017). These genes contain a HRs in intron for *NUB1*, *MARCKSL1*, and *ADGRA1* or in exonic coding sequence for *RRP36* (Supplementary Figure 1). We also noted two interesting genes encoding synaptic proteins to study in neurodevelopmental diseases. *SHARPIN* encodes a postsynaptic density protein involved in ubiquitination and directly interacting with the shank family of proteins (Lim et al., 2001). *BAI-1* encodes a synaptic receptor that activates the Rho signaling pathway

implicated in intellectual disability (Stephenson et al., 2013; van Bokhoven, 2011).

Our results indicate that *C9ORF72*, implicated in ALS/FTD, is not the only gene in the Human genome containing an HR GGGGCC. The discovery of its expansion in ALS required to analyze intronic regions and to use repeat-primed PCR protocols. The present study provides indirect support for the idea that HR expansions could be present in other genes in the genome. If one of the 36 genes listed in Table 1 is to be studied in a central nervous system disease as a candidate gene, the study of its HR should not be forgotten.

ACKNOWLEDGMENTS

We thank the Department Genomics, PST ASB, of the University of Tours.

REFERENCES

- Ash, P.E.A., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.-L., Dejesus-Hernandez, M., van Blitterswijk, M.M., Jansen-West, K., Paul, J.W., 3rd, Rademakers, R., Boylan, K.B., Dickson, D.W., Petrucelli, L., 2013. Unconventional translation of *C9ORF72* GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77, 639–646.
- Backx, L., Seuntjens, E., Devriendt, K., Vermeesch, J., Van Esch, H., 2011. A balanced translocation t(6;14)(q25.3;q13.2) leading to reciprocal fusion transcripts in a patient with intellectual disability and agenesis of corpus callosum. *Cytogenet. Genome Res.* 132, 135–143.
- 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 HR expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am. J. Hum. Genet.* 92, 345–353.
- Björkblom, B., Padzik, A., Mohammad, H., Westerlund, N., Komulainen, E., Hollos, P., Parviainen, L., Papageorgiou, A.C., Iljin, K., Kallioniemi, O., Kallajoki, M., Courtney, M.J., Mågård, M., James, P., Coffey, E.T., 2012. c-Jun N-terminal kinase phosphorylation of MARCKSL1 determines actin stability and migration in neurons and in cancer cells. *Mol. Cell. Biol.* 32, 3513–3526.
- Boivin, M., Pfister, V., Gaucherot, A., Ruffenach, F., Negroni, L., Sellier, C., Charlet-Berguerand, N., 2020. Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. *EMBO J.* 39(4):e100574.
- Byrne, S., Heverin, M., Elamin, M., Walsh, C., Hardiman, O., 2014. Intermediate repeat expansion length in C9orf72 may be pathological in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener.*;15-148-50.
- Cali, CP, Patino, M, Tai, YK, Ho, WY, McLean, CA, Morris, CM, Seeley, WW, Miller, BL, Gaig, C, Vonsattel, JPG, White, CL 3rd, Roeber, S, Kretzschmar, H, Troncoso, JC, Troakes, C, Gearing, M, Ghetti, B, Van Deerlin, VM, Lee, VM, Trojanowski, JQ, Mok, KY, Ling, H, Dickson, DW, Schellenberg, GD, Ling, SC, Lee, EB. C9orf72 intermediate repeats are associated with corticobasal degeneration, increased C9orf72 expression and disruption of autophagy. *Acta Neuropathol.* 2019 ;138(5):795-811.
- Chen, Y., Lu, J., Pan, H., Zhang, Y., Wu, H., Xu, K., Liu, X., Jiang, Y., Bao, X., Yao, Z., Ding,

- K., Lo, W.H.Y., Qiang, B., Chan, P., Shen, Y., Wu, X., 2003. Association between genetic variation of CACNA1H and childhood absence epilepsy. *Ann. Neurol.* 54, 239–243.
- Colasante, G., Simonet, J.C., Calogero, R., Crispi, S., Sessa, A., Cho, G., Golden, J.A., Broccoli, V., 2013. ARX Regulates Cortical Intermediate Progenitor Cell Expansion and Upper Layer Neuron Formation Through Repression of Cdkn1c. *Cereb. Cortex N. Y. N* 1991.
- 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.R., 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, 245–256.
- Eckle, V.-S., Shcheglovitov, A., Vitko, I., Dey, D., Yap, C.C., Winckler, B., Perez-Reyes, E., 2013. Mechanisms by which a CACNA1H mutation found in epilepsy patients increase seizure susceptibility. *J. Physiol.*
- Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S., Johnson, N., Juettemann, T., Kähäri, A.K., Keenan, S., Kulesha, E., Martin, F.J., Maurel, T., McLaren, W.M., Murphy, D.N., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ruffier, M., Sheppard, D., Taylor, K., Thormann, A., Trevanion, S.J., Vullo, A., Wilder, S.P., Wilson, M., Zadissa, A., Aken, B.L., Birney, E., Cunningham, F., Harrow, J., Herrero, J., Hubbard, T.J.P., Kinsella, R.,

- Muffato, M., Parker, A., Spudich, G., Yates, A., Zerbino, D.R., Searle, S.M.J., 2014. Ensembl 2014. *Nucleic Acids Res.* 42, D749–755.
- Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H., Chelly, J., Sanlaville, D., van Bokhoven, H., Ropers, H.-H., Laumonnier, F., Ranieri, E., Schwartz, C.E., Abidi, F., Tarpey, P.S., Futreal, P.A., Whibley, A., Raymond, F.L., Stratton, M.R., Fryns, J.-P., Scott, R., Peippo, M., Sipponen, M., Partington, M., Mowat, D., Field, M., Hackett, A., Marynen, P., Turner, G., Géczy, J., 2008. Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am. J. Hum. Genet.* 82, 432–443.
- Gendron T.F., Bieniek K.F., Zhang Y.J., Jansen-West K., Ash P.E., Caulfield T., Daugherty L., Dunmore J.H., Castanedes-Casey M., Chew J., Cosio D.M., van Blitterswijk M., Lee W.C., Rademakers R., Boylan K.B., Dickson D.W., Petrucelli L., 2013. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* 126, 829-44.
- Gendron, T.F., Belzil, V.V., Zhang, Y.-J., Petrucelli, L., 2014. Mechanisms of toxicity in C9FTLD/ALS. *Acta Neuropathol. (Berl.)*.
- Gérus, M., Bonnart, C., Caizergues-Ferrer, M., Henry, Y., Henras, A.K., 2010. Evolutionarily conserved function of RRP36 in early cleavages of the pre-rRNA and production of the 40S ribosomal subunit. *Mol. Cell. Biol.* 30, 1130–1144.
- Gijssels, I., Van Langenhove, T., van der Zee, J., Slegers, 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 Bleeker, J., Maes, G.,

- Bäumer, V., Dillen, L., Joris, G., Cuijt, I., Corsmit, E., Elinck, E., Van Dongen, J., Vermeulen, S., Van den Broeck, M., Vaerenberg, C., Mattheijssense, 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 Neurol.* 11, 54–65.
- Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A.J., Shaw, P.G., Kim, M.-S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., Rothstein, J.D., Wang, J., 2014. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195–200.
- He, X.-H., Lin, F., Qin, Z.-H., 2010. Current understanding on the pathogenesis of polyglutamine diseases. *Neurosci. Bull.* 26, 247–256.
- Hoyer, J., Ekici, A.B., Endeke, S., Popp, B., Zweier, C., Wiesener, A., Wohlleber, E., Dufke, A., Rossier, E., Petsch, C., Zweier, M., Göhring, I., Zink, A.M., Rappold, G., Schröck, E., Wieczorek, D., Riess, O., Engels, H., Rauch, A., Reis, A., 2012. Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* 90, 565–572.
- Iyer, S, Subramanian, V, Acharya, KR., 2018. C9orf72, a protein associated with amyotrophic lateral sclerosis (ALS) is a guanine nucleotide exchange factor. *PeerJ.* 6:e5815.
- Jung, E.M., Moffat, J.J., Liu, J., Dravid, S.M., Gurumurthy, C.B., Kim, W.Y., 2017. Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat Neurosci.* 20:1694-1707.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. *Genome Res.* 12, 996–1006.

- Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., Hayes, M.J., 2013. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. *Bioinforma. Oxf. Engl.* 29, 499–503.
- Li, X.S., Trojer, P., Matsumura, T., Treisman, J.E., Tanese, N., 2010. Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Mol. Cell. Biol.* 30, 1673–1688.
- Lim, S., Sala, C., Yoon, J., Park, S., Kuroda, S., Sheng, M., Kim, E., 2001. Sharpin, a novel postsynaptic density protein that directly interacts with the shank family of proteins. *Mol. Cell. Neurosci.* 17, 385–397.
- Lu, B., Al-Ramahi, I., Valencia, A., Wang, Q., Berenshteyn, F., Yang, H., Gallego-Flores, T., Ichcho, S., Lacoste, A., Hild, M., Difulgia, M., Botas, J., Palacino, J., 2013. Identification of NUB1 as a suppressor of mutant Huntington toxicity via enhanced protein clearance. *Nat. Neurosci.* 16, 562–570.
- McClatchey, A.I., Trofatter, J., McKenna-Yasek, D., Raskind, W., Bird, T., Pericak-Vance, M., Gilchrist, J., Arahata, K., Radosavljevic, D., Worthen, H.G., 1992. Dinucleotide repeat polymorphisms at the SCN4A locus suggest allelic heterogeneity of hyperkalemic periodic paralysis and paramyotonia congenita. *Am. J. Hum. Genet.* 50, 896–901.
- McNamara, R.K., Lenox, R.H., 2000. Differential regulation of primary protein kinase C substrate (MARCKS, MLP, GAP-43, RC3) mRNAs in the hippocampus during kainic acid-induced seizures and synaptic reorganization. *J. Neurosci. Res.* 62, 416–426.
- Mirkin, S.M., 2007. Expandable DNA repeats and human disease. *Nature* 447, 932–940.
- Mori, K., Arzberger, T., Grässer, F.A., Gijssels, I., May, S., Rentzsch, K., Weng, S.-M., Schludi, M.H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretschmar, H.A., Haass, C., Edbauer, D., 2013. Bidirectional transcripts of the

- expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol. (Berl.)* 126, 881–893.
- Nakamura, Y., Shi, X., Numata, T., Mori, Y., Inoue, R., Lossin, C., Baram, T.Z., Hirose, S., 2013. Novel HCN2 Mutation Contributes to Febrile Seizures by Shifting the Channel's Kinetics in a Temperature-Dependent Manner. *PLoS One* 8, e80376.
- Nelson, D.L., Orr, H.T., Warren, S.T., 2013. The unstable repeats--three evolving faces of neurological disease. *Neuron* 77, 825–843.
- Ng, A.S.L., Tan, E.K., 2017. Intermediate C9orf72 alleles in neurological disorders: does size really matter? *J Med Genet.* 54, 591-597.
- Paulson H., 2018. Repeat expansion diseases. *Handb Clin Neurol.* 147:105-123.
- Ptáček, L.J., George, A.L., Jr, Griggs, R.C., Tawil, R., Kallen, R.G., Barchi, R.L., Robertson, M., Leppert, M.F., 191. Identification of a mutation in the gene causing hyperkalemic periodic paralysis. *Cell* 67, 1021–1027.
- 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., Sorarù, G., Ticozzi, N., Comi, G.P., D'Alfonso, S., Gellera, C., Silani, V., SLAGEN Consortium, 2012. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. *Neurobiol. Aging* 33, 2528.e7–14.
- Renoux, A.J., Todd, P.K., 2012. Neurodegeneration the RNA way. *Prog. Neurobiol.* 97, 173–189.
- Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, 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., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wu, J., Chiò, 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 expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268.

Richet, E., Pooler, A.M., Rodriguez, T., Novoselov, S.S., Schmidtke, G., Groettrup, M., Hanger, D.P., Cheetham, M.E., van der Spuy, J., 2012. NUB1 modulation of GSK3 β reduces tau aggregation. *Hum. Mol. Genet.* 21, 5254–5267.

Santen, G.W.E., Aten, E., Sun, Y., Almomani, R., Gilissen, C., Nielsen, M., Kant, S.G., Snoeck, I.N., Peeters, E.A.J., Hilhorst-Hofstee, Y., Wessels, M.W., den Hollander, N.S., Ruivenkamp, C.A.L., van Ommen, G.-J.B., Breuning, M.H., den Dunnen, J.T., van Haeringen, A., Kriek, M., 2012. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. *Nat. Genet.* 44, 379–380.

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., Hortobágyi, 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.T., 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 Bellerocche, J., Vianney de Jong, J.M.B., 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. *Eur. J. Hum. Genet. EJHG* 21, 102–108.
- St Martin J.L., Wang L., Kaprielian Z., 2019. Toxicity in ALS: TDP-43 modifiers and C9orf72. *Neurosci Lett.* 11,134621.
- Stephenson, J.R., Paavola, K.J., Schaefer, S.A., Kaur, B., Van Meir, E.G., Hall, R.A., 2013. Brain-specific angiogenesis inhibitor-1 signaling, regulation, and enrichment in the postsynaptic density. *J. Biol. Chem.* 288, 22248–22256.
- Tang, B., Sander, T., Craven, K.B., Hempelmann, A., Escayg, A., 2008. Mutation analysis of the hyperpolarization-activated cyclic nucleotide-gated channels HCN1 and HCN2 in idiopathic generalized epilepsy. *Neurobiol. Dis.* 29, 59–70.
- Tanji, K., Tanaka, T., Mori, F., Kito, K., Takahashi, H., Wakabayashi, K., Kamitani, T., 2006. NUB1 suppresses the formation of Lewy body-like inclusions by proteasomal degradation of synphilin-1. *Am. J. Pathol.* 169, 553–565.
- Tokunaga F., Iwai K., 2012. LUBAC, a novel ubiquitin ligase for linear ubiquitination, is crucial for inflammation and immune responses. *Microbes Infect.* 14:563-72.
- Van Bokhoven, H., 2011. Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* 45, 81–104.
- Van der Zee, J., Gijselinck, I., Dillen, L., Van Langenhove, T., Theuns, J., Engelborghs, S., Philtjens, S., Vandenbulcke, M., Sleegers, K., Sieben, A., Bäumer, V., Maes, G., Corsmit, E., Borroni, B., Padovani, A., Archetti, S., Pernecky, R., Diehl-Schmid, J., de Mendonça, A., Miltenberger-Miltenyi, G., Pereira, S., Pimentel, J., Nacmias, B.,

Bagnoli, S., Sorbi, S., Graff, C., Chiang, H.-H., Westerlund, M., Sanchez-Valle, R., Llado, A., Gelpi, E., Santana, I., Almeida, M.R., Santiago, B., Frisoni, G., Zanetti, O., Bonvicini, C., Synofzik, M., Maetzler, W., Vom Hagen, J.M., Schöls, L., Heneka, M.T., Jessen, F., Matej, R., Parobkova, E., Kovacs, G.G., Ströbel, T., Sarafov, S., Tournev, I., Jordanova, A., Danek, A., Arzberger, T., Fabrizi, G.M., Testi, S., Salmon, E., Santens, P., Martin, J.-J., Cras, P., Vandenberghe, R., De Deyn, P.P., Cruts, M., Van Broeckhoven, C., van der Zee, J., Gijssels, I., Dillen, L., Van Langenhove, T., Theuns, J., Philtjens, S., Sleegers, K., Bäumer, V., Maes, G., Corsmit, E., Cruts, M., Van Broeckhoven, C., van der Zee, J., Gijssels, I., Dillen, L., Van Langenhove, T., Philtjens, S., Theuns, J., Sleegers, K., Bäumer, V., Maes, G., Cruts, M., Van Broeckhoven, C., Engelborghs, S., De Deyn, P.P., Cras, P., Engelborghs, S., De Deyn, P.P., Vandenbulcke, M., Vandenbulcke, M., Borroni, B., Padovani, A., Archetti, S., Pernecky, R., Diehl-Schmid, J., Synofzik, M., Maetzler, W., Müller Vom Hagen, J., Schöls, L., Synofzik, M., Maetzler, W., Müller Vom Hagen, J., Schöls, L., Heneka, M.T., Jessen, F., Ramirez, A., Kurzwelly, D., Sachtleben, C., Mairer, W., de Mendonça, A., Miltenberger-Miltenyi, G., Pereira, S., Firmo, C., Pimentel, J., Sanchez-Valle, R., Llado, A., Antonell, A., Molinuevo, J., Gelpi, E., Graff, C., Chiang, H.-H., Westerlund, M., Graff, C., Kinhult Ståhlbom, A., Thonberg, H., Nennesmo, I., Börjesson-Hanson, A., Nacmias, B., Bagnoli, S., Sorbi, S., Bessi, V., Piaceri, I., Santana, I., Santiago, B., Santana, I., Helena Ribeiro, M., Rosário Almeida, M., Oliveira, C., Massano, J., Garret, C., Pires, P., Frisoni, G., Zanetti, O., Bonvicini, C., Sarafov, S., Tournev, I., Jordanova, A., Tournev, I., Kovacs, G.G., Ströbel, T., Heneka, M.T., Jessen, F., Ramirez, A., Kurzwelly, D., Sachtleben, C., Mairer, W., Jessen, F., Matej, R., Parobkova, E., Danel, A., Arzberger, T., Maria Fabrizi, G., Testi, S., Ferrari,

- S., Cavallaro, T., Salmon, E., Santens, P., Cras, P., European Early-Onset Dementia Consortium, 2013. A pan-European study of the C9orf72 repeat associated with FTLTLD: geographic prevalence, genomic instability, and intermediate repeats. *Hum. Mutat.* 34, 363–373.
- Vanderweyde, T., Yu, H., Varnum, M., Liu-Yesucevitz, L., Citro, A., Ikezu, T., Duff, K., Wolozin, B., 2012. Contrasting pathology of the stress granule proteins TIA-1 and G3BP in tauopathies. *J. Neurosci. Off. J. Soc. Neurosci.* 32, 8270–8283.
- Vandewalle, J., Langen, M., Zschaetzsch, M., Nijhof, B., Kramer, J.M., Brems, H., Bauters, M., Lauwers, E., Srahna, M., Marynen, P., Verstreken, P., Schenck, A., Hassan, B.A., Froyen, G., 2013. Ubiquitin ligase HUWE1 regulates axon branching through the Wnt/ β -catenin pathway in a *Drosophila* model for intellectual disability. *PloS One* 8, e81791.
- Wang K., Kim C., Bradfield J., Guo Y., Toskala E., Otieno F.G., Hou C., Thomas K., Cardinale C., Lyon G.J., Golhar R., Hakonarson H. 2013. Whole-genome DNA/RNA sequencing identifies truncating mutations in RBCK1 in a novel Mendelian disease with neuromuscular and cardiac involvement. *Genome Med.* 5:67.
- Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd, Su, A.I., 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10, R130.
- Yoshinaga, H., Sakoda, S., Good, J.-M., Takahashi, M.P., Kubota, T., Arikawa-Hirasawa, E., Nakata, T., Ohno, K., Kitamura, T., Kobayashi, K., Ohtsuka, Y., 2012. A novel mutation in SCN4A causes severe myotonia and school-age-onset paralytic episodes. *J. Neurol. Sci.* 315, 15–19.
- Zhang, D., Iyer, L.M., He, F., Aravind, L., 2012. Discovery of Novel DENN Proteins:

Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. *Front. Genet.* 3, 283.

Zhang, J., Herrup, K., 2011. Nucleocytoplasmic Cdk5 is involved in neuronal cell cycle and death in post-mitotic neurons. *Cell Cycle Georget. Tex* 10, 1208–1214.

Zhang Y.J., Guo L., Gonzales P.K., Gendron T.F., Wu Y., Jansen-West K., O'Raw A.D., Pickles S.R., Prudencio M., Carlomagno Y., Gachechiladze M.A., Ludwig C., Tian R., Chew J., DeTure M., Lin W.L., Tong J., Daugherty L.M., Yue M., Song Y., Andersen J.W., Castanedes-Casey M., Kurti A., Datta A., Antognetti G., McCampbell A., Rademakers R., Oskarsson B., Dickson D.W., Kampmann M., Ward M.E., Fryer J.D., Link C.D., Shorter J., Petrucelli L., 2019. Heterochromatin anomalies and double-stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science.* 15, 363(6428).

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., Nan, Z., Forster, C., Low, W.C., Schoser, B., Somia, N.V., Clark, H.B., Schmechel, S., Bitterman, P.B., Gourdon, G., Swanson, M.S., Moseley, M., Ranum, L.P.W., 2011. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U. S. A.* 108, 260–265.

FIGURES

Figure 1: (A) Number of GGGGCC or GGCCCC repeats on each chromosome of the Human genome. Arrow: chromosome 9 containing the *C9ORF72* gene; (B) Localisation of these repeats in genes.

Table 1: Genes containing hexanucleotide repeats expressed in the central nervous system or neuromuscular system.

Table 2: Pathways over-represented after GO analysis and genes with hexanucleotide repeats in these pathways.

Figure 1

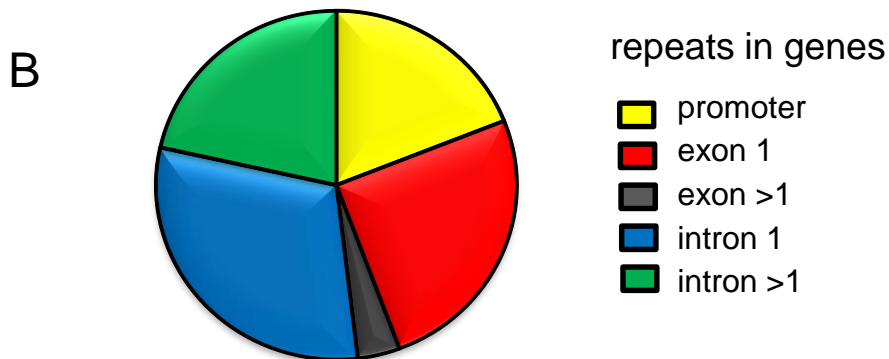
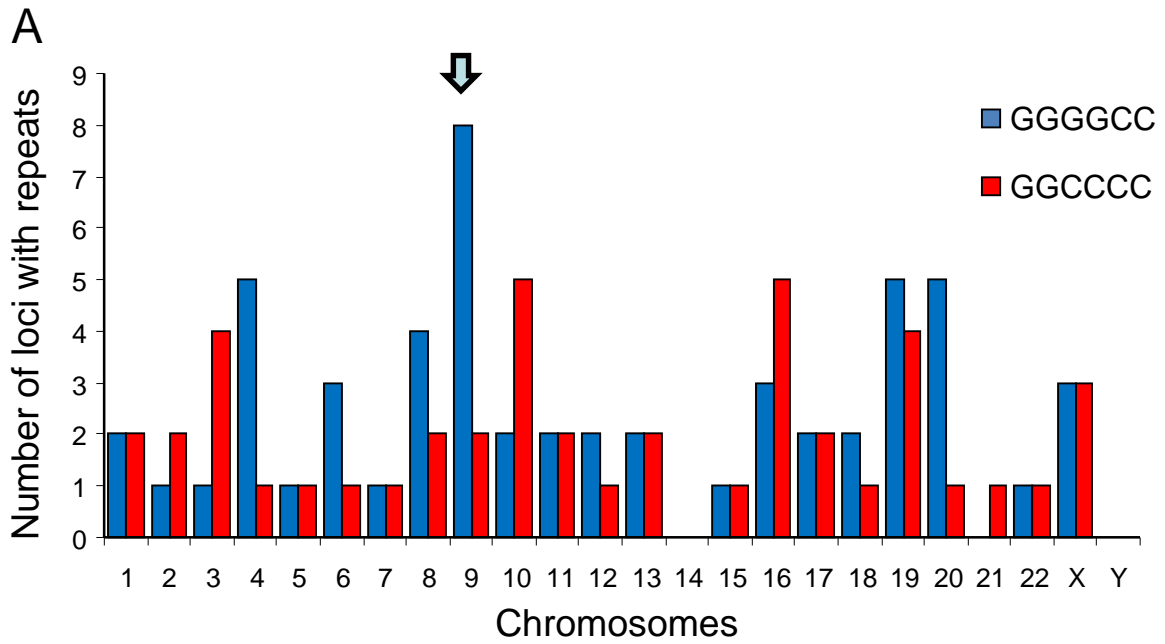


Table 1

Locus (Human)	Gene symbol (synonyms)	Gene name	Repeat units (n)	Localisation	Position (b38)
1p13	<i>KCNA2</i>	Potassium voltage-gated channel, shaker-related subfamily, member 2	3	intron 1	110545223
1p35.1	<i>MARCKSL1</i>	MARCKS-like 1	3	intron 1	32799929
2q24.3	<i>STK39</i>	Serine threonine kinase 39	4	exon 1 CS	168811358
3q13	<i>CASR</i>	Calcium-sensing receptor	3	intron 1	122184297
3q27	<i>CHRD</i>	Chordin	4	UTR5'	184380181
4q26	<i>FGF2</i>	Fibroblast growth factor 2 (basic)	3	exon 1 CS	122827056
6p21	<i>ITPR3</i>	Inositol 1,4,5-trisphosphate receptor, type 3	3	intron 1	33633758
6p21.1	<i>RRP36</i>	Ribosomal RNA processing 36 homolog	4	exon 1 CS	43021685
6q25.3	<i>ARID1B</i>	AT rich interactive domain 1B	3	promoter	156777206
7q36.1	<i>NUB1</i>	Negative regulator of ubiquitin-like protein 1	3	intron 1	151341993
8q22.2	<i>NCALD</i>	Neurocalcin delta	3	intron 1	101689018
8q24.3	<i>BAI1</i>	Brain-specific angiogenesis inhibitor 1	3	intron 3	142475804
8q24.3	<i>SHARPIN</i>	SHANK-associated RH domain interactor	3	intron 5	144108449
9p21.2	<i>C9orf72</i>	Chromosome 9 open reading frame 72	3	intron 1	27546864
9q31	<i>TMEFF1</i>	Tomoregulin-1	3	intron 1	100473999
9q34.3	<i>CACNA1B</i>	Calcium channel, voltage-dependent, N type, alpha 1B subunit	3	intron 1	137878265
9q34.3	<i>GPSM1 / AGS3</i>	G-protein signaling modulator 1	3	intron 1	136327776
10q24.31	<i>LZTS2 / LAPSER1</i>	Leucine zipper tumor suppressor 2	3	UTR5'	100999854
10q26	<i>GPR123</i>	G protein-coupled receptor 123	4	intron 5	133124834
10q26.3	<i>JAKMIP3 / NECC2</i>	Janus kinase and microtubule interacting protein 3	3	intron 10	132142483
11p15.5	<i>AP2A</i>	Adaptor-related protein complex 2, alpha 2 subunit	3	intron 1	926109
11p15.5	<i>CDKN1C</i>	Cyclin-dependent kinase inhibitor 1C (p57, Kip2)	3	exon 1 CS	2888502
12q13.13	<i>POU6F1 / BRN5</i>	POU class 6 homeobox 1	4	promoter	51191300
13q34	<i>TFDP1</i>	Transcription factor Dp-1	3	intron 1	113585152
16p13.3	<i>CACNA1H</i>	Calcium channel, voltage-dependent, T type, alpha 1H subunit	4	UTR5'	1153342
16q12.1	<i>PAPD5</i>	PAP associated domain containing 5	3	exon 2 CS	50153843
17q23.2	<i>TBX2</i>	T-box 2	3	UTR5'	61400135
17q23.3	<i>SCN4A</i>	Sodium channel voltage gated type IV, alpha subunit	3	intron 14	63962421
19p13.3	<i>TNFAIP8L1</i>	Tumor necrosis factor (TNF)-alpha induced protein 8	3	UTR5'	4639516
19p13.3	<i>HCN2</i>	Hyperpolarization-activated cyclic nucleotide-gated channel	4	intron 6	613554
19q13.12	<i>USF2</i>	Upstream transcription factor 2	3	intron 1	35269199
19q13.2	<i>ZFP36 / TTP</i>	Ring finger protein, Tritetraprolin	3	promoter	39406779
19q13.33	<i>LRRRC4B</i>	Leucine rich repeat containing protein 4B	3	promoter	50518355
20p13	<i>RBCK1 / RNF54</i>	Ring finger protein 54	3	intron 1	407959
20q13.33	<i>TAF4 / TAFII130</i>	TBP associated factor 4 RNA polymerase 2	3	exon 1 CS	61955397
Xp11.22	<i>HUWE1</i>	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	5	intron 17	53594344

CS: coding sequence

Table 2

GO terms		Statistical values		Genes
Number	Name	p value	Enrichment	
97039	Protein linear polyubiquitination	1.05e-5	350.22	<i>SHARPIN</i> <i>RBCK1</i>
34765	Regulation of ion transmembrane transport	1.5e-5	10.98	<i>KCNA2</i> <i>STK39</i> <i>HCN2</i> <i>SCN4A</i> <i>CACNA1H</i> <i>CACNA1B</i>
70509	Calcium ion import	2.23e-5	54.34	<i>CASR</i> <i>CACNA1H</i> <i>CACNA1B</i>
86010	Membrane depolarization during action potential	4.31e-5	43.78	<i>SCN4A</i> <i>CACNA1H</i> <i>CACNA1B</i>