

RESEARCH

Open Access



Genomic, transcriptomic and RNA editing analysis of human MM1 and VV2 sporadic Creutzfeldt-Jakob disease

Martina Tarozzi¹, Simone Baiardi^{1,2}, Claudia Sala¹, Anna Bartoletti-Stella¹, Piero Parchi^{2,3}, Sabina Capellari^{2,3*†} and Gastone Castellani^{1†}

Abstract

Creutzfeldt-Jakob disease (CJD) is characterized by a broad phenotypic spectrum regarding symptoms, progression, and molecular features. Current sporadic CJD (sCJD) classification recognizes six main clinical-pathological phenotypes. This work investigates the molecular basis of the phenotypic heterogeneity of prion diseases through a multi-omics analysis of the two most common sCJD subtypes: MM1 and VV2. We performed DNA target sequencing on 118 genes on a cohort of 48 CJD patients and full exome RNA sequencing on post-mortem frontal cortex tissue on a subset of this cohort. DNA target sequencing identified multiple potential genetic contributors to the disease onset and phenotype, both in terms of coding, damaging-predicted variants, and enriched groups of SNPs in the whole cohort and the two subtypes. The results highlight a different functional impairment, with VV2 associated with higher impairment of the pathways related to dopamine secretion, regulation of calcium release and GABA signaling, showing some similarities with Parkinson's disease both on a genomic and a transcriptomic level. MM1 showed a gene expression profile with several traits shared with different neurodegenerative, without an apparent distinctive characteristic or similarities with a specific disease. In addition, integrating genomic and transcriptomic data led to the discovery of several sites of ADAR-mediated RNA editing events, confirming and expanding previous findings in animal models. On the transcriptomic level, this work represents the first application of RNA sequencing on CJD human brain samples. Here, a good clusterization of the transcriptomic profiles of the two subtypes was achieved, together with the finding of several differently impaired pathways between the two subtypes. The results add to the understanding of the molecular features associated with sporadic CJD and its most common subtypes, revealing strain-specific genetic signatures and functional similarities between VV2 and Parkinson's disease and providing preliminary evidence of RNA editing modifications in human sCJD.

Keywords: Prion diseases, Sporadic Creutzfeldt-Jakob disease, NGS, RNA sequencing, Phenotypic heterogeneity, Prion strains, Multi-omics analysis, RNA editing, Genetic modifiers

Background

Prion diseases are rare, invariably lethal and rapidly progressive neurodegenerative disorders that affect humans and other species of mammals [1]. Sporadic Creutzfeldt-Jakob disease (sCJD) is the most common and best studied human prion disease. It is characterized by a wide phenotypic spectrum regarding first symptoms, disease progression, and histo-molecular features. Current sCJD

[†]Sabina Capellari and Gastone Castellani have equally contributed to this work

*Correspondence: sabina.capellari@unibo.it

² Programma di Neuropatologia delle Malattie, Neurodegenerative, IRCCS Istituto delle Scienze Neurologiche di Bologna, 40139 Bologna, Italy
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

classification recognizes six main clinical and pathological phenotypes that largely correlate at the molecular level with the genotype at *PRNP* codon 129 (methionine, M, or valine, V) and the protein type (1 or 2) [2] accumulated in the brain. Among the six sCJD subtypes, MM1 and VV2 are the most common: MM1 accounts for ~65% of cases, and it is characterized by a rapid disease progression that, on average, has an age of onset of 70 years and a clinical disease duration of 4 months [3, 4]. The VV2 phenotype accounts for 15–20% of sCJD cases and is associated with a slightly longer disease (6 months on average) that appears earlier (64.5 years on average). While rapidly progressive dementia is the main clinical feature of MM1 phenotype, VV2 presents with prominent cerebellar and subcortical impairment [3, 5, 6]. Noteworthy, these subtypes not only show a difference in the clinicopathological phenotype, but transmission studies of the disease in syngeneic hosts demonstrated that sCJD MM1 and VV2 behave biologically as different prion strains, namely the M1 and V2 [7–9].

Recent research aimed mainly at a better understanding of the genetic risk factors and modifiers associated with the onset and phenotypic expression of the sporadic disease: Genome-Wide Association Studies (GWAS) of large sCJD cohorts confirmed the significant association with *PRNP* codon 129, the strongest genetic risk factor [10]. This work also identified two other loci associated with an increased risk of sCJD, in *STX6* (rs3747957) and *GAL3ST1* (rs2267161) genes, indicating intracellular trafficking and sphingolipid metabolism as probable triggering mechanisms and corroborating the likely shared molecular dysregulation with other prion-like disorders [10]. In terms of gene expression, microarray and RNA sequencing have been applied to determine the most affected biological processes and molecular pathways at various disease stages. Most of the available knowledge comes from murine models: according to current literature [11–13], in the early disease stage, prominent changes in gene expression affect immune response through the complement system associated with microglia and astrocyte activation. During the intermediate stages of PrP^{Sc} accumulation, the transcriptional profile seems to alter pathways involving membrane regulation and vesicle traffic, with the activation of sphingolipid, glycosaminoglycans, and cholesterol metabolisms. In the final stage, a transcriptional down-regulation of genes associated with synaptic transmission and axonal growth occurs, followed by activation of cellular processes associated with apoptosis. Only a few studies about gene expression changes in human samples exist [14–17]. Although these studies provide information only about the final stage of the disease, they gain the advantage of focusing on a bona fide sporadic disease, differently

from animal models that can describe only the acquired forms. These works highlight a prominent impairment of gene expression profiles that seem to parallel processes observed in animal models.

Further studies are needed to understand better the molecular mechanisms undergoing the sporadic disease and the biological pathways associated with its different phenotypes. This last aspect is crucial since the strain phenomenon, first described in prion diseases, has now been expanded to other more prevalent proteinopathies, like Parkinson's and Alzheimer's Diseases [18]. In this work, we performed DNA target sequencing on 118 genes in a cohort of 48 MM1 and VV2 sCJD, and RNA sequencing on postmortem brain samples (frontal cortex) in a subset of 16 cases (8 MM1, 8 VV2). On the genomic layer, data science and statistical analysis identified putative genetic modifiers and recurrent genetic patterns in the two classes, while RNA sequencing characterized differentially expressed genes and enriched pathways in the two subtypes/strains. Finally, the integration of these two omics layers also provided proof of RNA editing events in human disease, confirming the previous finding in murine models [19]. These results add to the current understanding of the molecular biology underlying prion diseases and the strain phenomenon.

Results

DNA target sequencing: putative genetic modifiers and recurrent genetic patterns

We analyzed DNAs of 48 Italian patients diagnosed with definite sporadic CJD MM1 and VV2 (Additional file 1: Table S1), with the Neurodegeneration (Illumina) gene panel (Additional file 1: Table S8). After applying filters described in materials and methods, 57,005 different variants were identified in the 118 analyzed genes. We observed 42,169 different variants in the MM1 group and 47,594 in the VV2. The numerical distribution of variants in each gene showed no significant difference between the two sCJD subtypes. On average, each sample carried 14,369 variants ($\sigma=823$). Also, in this case, we found no significant difference in the average number of variants, even though MM1 samples showed a higher homogeneity compared to VV2 samples (MM1: 14,366 $\sigma=528$, VV2: 14,373 $\sigma=1051$).

Variants in VV2 and MM1 affect different genes

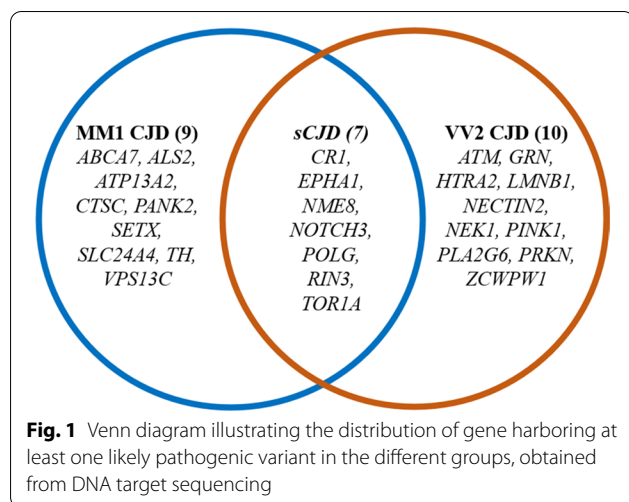
Variants of interest were selected based on their consequence and predicted effect. Thirty-five different missense variants were predicted as “probably/likely pathogenic” or “deleterious” by both Polyphen2 and SIFT predictors. These variants, reported in Table 1, involve 26 genes. Seven variants were found more than once in the dataset, for a total of 62 findings of likely

Table 1 List of 35 probably damaging missense variants identified in this sporadic cohort, result of DNA target sequencing

VID	Gene	HGVSC	HGVSP	SIFT Prediction	PolyPhen2 prediction	Clinvar annotation
19:1046239:G	<i>ABCA7</i>	c.1456C>G	p.(Pro486Ala)	Deleterious	Probably damaging	NA
19:1047336:A	<i>ABCA7</i>	c.2026G>A	p.(Ala676Thr)	Deleterious	Probably damaging	Benign [single provider]
19:1058635:T	<i>ABCA7</i>	c.5168C>T	p.(Ser1723Leu)	Deleterious	Probably damaging	NA
2:202587783:T	<i>ALS2</i>	c.3685T>A	p.(Trp1229Arg)	Deleterious	Possibly damaging	NA
11:108117787:T	<i>ATM</i>	c.998C>T	p.(Ser333Phe)	Deleterious	Probably damaging	Conflicting interpretations: Uncertain significance(1); Benign(2); Likely benign(12)
1:17312787:A	<i>ATP13A2</i>	c.3472C>T	p.(Arg1158Cys)	Deleterious	Possibly damaging	Uncertain significance
1:207680070:T	<i>CR1</i>	c.313C>T	p.(Arg105Cys)	Deleterious	Probably damaging	NA
1:207739203:T	<i>CR1</i>	c.2537C>T	p.(Ser846Phe)	Deleterious	possibly damaging	NA
11:88027209:C	<i>CTSC</i>	c.1357A>G	p.(Ile453Val)	Deleterious	possibly damaging	Benign/Likely benign
7:143088584:T	<i>EPHA1</i>	c.2897G>A	p.(Arg966His)	Deleterious	Probably damaging	NA
7:143092269:A	<i>EPHA1</i>	c.2090C>T	p.(Pro697Leu)	Deleterious	Possibly damaging	NA
17:42427095:A	<i>GRN</i>	c.325G>A	p.(Gly109Arg)	Deleterious	Probably damaging	Uncertain significance
2:74759825:A	<i>HTRA2</i>	c.1195G>A	p.(Gly399Ser)	Deleterious	Probably damaging	Benign/Likely benign
5:126158560:T	<i>LMNB1</i>	c.1474G>T	p.(Ala492Ser)	Deleterious	possibly Damaging	NA
19:45375208:T	<i>NECTIN2</i>	c.577C>T	p.(Arg193Trp)	Deleterious	Probably damaging	NA
4:170398474:C	<i>NEK1</i>	c.2235T>G	p.(Asn745Lys)	Deleterious	Probably damaging	Conflicting interpretations: Likely pathogenic(1); Benign(3); Likely benign(2)
7:37923923:C	<i>NME8</i>	c.1013T>C	p.(Ile338Thr)	Deleterious	Probably damaging	Benign
7:37924854:A	<i>NME8</i>	c.1247G>A	p.(Ser416Asn)	Deleterious	Probably damaging	NA
7:37936557:A	<i>NME8</i>	c.1630G>A	p.(Ala544Thr)	Deleterious	Probably damaging	Benign/Likely benign
19:15273335:T	<i>NOTCH3</i>	c.5854G>A	p.(Val1952Met)	Deleterious	Probably damaging	Benign/Likely benign
19:15289863:A	<i>NOTCH3</i>	c.3691C>T	p.(Arg1231Cys)	Deleterious	Possibly damaging	Conflicting interpretations: Pathogenic(2); Uncertain significance(5)
19:15290917:A	<i>NOTCH3</i>	c.3293C>T	p.(Thr1098Ile)	Deleterious	Possibly damaging	NA
20:3888719:A	<i>PANK2</i>	c.775G>A	p.(Gly259Arg)	Deleterious	Probably Damaging	NA
1:20976976:A	<i>PINK1</i>	c.1538G>A	p.(Gly513Asp)	Deleterious	Probably damaging	NA
22:38539240:A	<i>PLA2G6</i>	c.481C>T	p.(Arg161Cys)	Deleterious	Probably damaging	Uncertain significance
15:89865073:C	<i>POLG</i>	c.2492A>G	p.(Tyr831Cys)	Deleterious	Possibly damaging	Benign/Likely benign
6:161781201:A	<i>PRKN</i>	c.1204C>T	p.(Arg402Cys)	Deleterious	Probably damaging	Conflicting interpretations: Uncertain significance(2); Benign(1)
14:93119136:A	<i>RIN3</i>	c.1742G>A	p.(Arg581Gln)	Deleterious	Probably damaging	NA
14:93142861:C	<i>RIN3</i>	c.2377T>C	p.(Tyr793His)	Deleterious	Possibly damaging	NA
9:135202325:C	<i>SETX</i>	c.4660T>G	p.(Cys1554Gly)	Deleterious	Probably damaging	Benign/Likely benign
14:92905737:C	<i>SLC24A4</i>	c.377T>C	p.(Leu126Pro)	Deleterious	Probably damaging	NA
11:2189817:C	<i>TH</i>	c.484T>G	p.(Phe162Val)	Deleterious	Possibly damaging	Uncertain significance
9:132580901:G	<i>TOR1A</i>	c.646G>C	p.(Asp216His)	Deleterious	Possibly damaging	Benign
15:62269347:C	<i>VPS13C</i>	c.2342T>G	p.(Leu781Trp)	Deleterious	Probably damaging	NA
7:100016781:C	<i>ZCWPW1</i>	c.314A>G	p.(Glu105Gly)	Deleterious	Probably damaging	NA

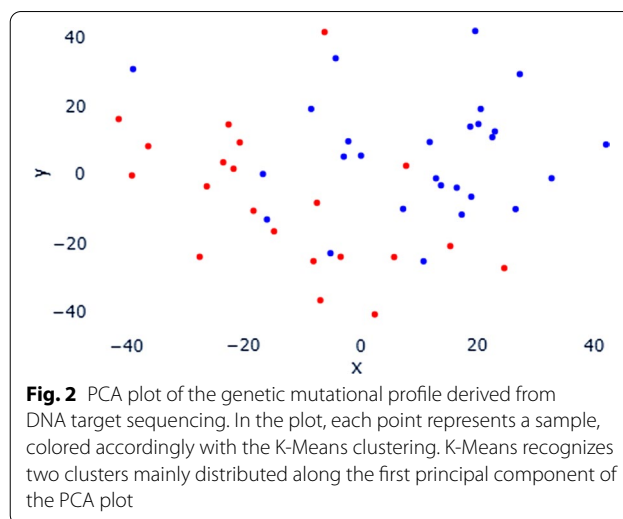
deleterious variants, equally distributed between the two subtypes (31 in MM1 and 31 in VV2). No difference was found in the number of samples carrying at least one putative damaging mutation between the two subtypes (17 MM1 and 16 VV2). Of the 26 genes involved in these possibly damaging variants, seven are affected in both MM1 and VV2 patients (*CR1*, *EPHA1*, *NME8*, *NOTCH3*, *POLG*, *RIN3*, *TOR1A*), while nine genes were affected only in MM1 samples (*ABCA7*,

ALS2, *ATP13A2*, *CTSC*, *PANK2*, *SETX*, *SLC24A4*, *TH*, *VPS13C*) and ten genes only in VV2 samples (*ATM*, *GRN*, *HTRA2*, *LMNB1*, *NECTIN2*, *NEK1*, *PINK1*, *PLA2G6*, *PRKN*, *ZCWPW1*). Therefore, even though the number of likely pathogenic variants shows no difference between subtypes, specific genes seem to be affected exclusively in one subtypes and not in the other (Fig. 1).



Allele frequencies differ between sCJD patients and the European population but not between VV2 and MM1

Allele frequencies were compared between the two considered subtypes and between the whole sCJD cohort and the European population. The comparison between MM1 and VV2 subtypes did not provide any statistically significant result except for codon 129 and three other SNV in its proximity, which was expected given the experimental design. The comparison between sCJD patients versus allele frequencies of the European population reported in the GnomAD database highlighted 237 variants with a statistically significantly different allele frequency ($p\text{-adj} < 0.05$). Of them, 71 are non-intronic (missense, synonymous, 3'/5' UTR, splice region variants) and distributed in 36 different genes. Functional analysis with over representation methods on these 36 genes (Additional file 1: Table S2) showed an enrichment of several biological processes, many of which involved in synaptic regulation, chaperon binding and regulation of cell death, and a strong association with neurodegenerative processes. Three of the overrepresented variants in the sCJD cohort were also predicted as probably pathogenic by SIFT and PolyPhen2 variant predictors: *GRN* p.Gly109Arg ($p\text{-adj} = 0.02$, sample #17, VV2), *NME8* p.Ser416Asn ($p\text{-adj} = 0.03$, sample #40, MM1) and *RIN3* Arg581Gln ($p\text{-adj} = 0.03$, sample #8, VV2). For none of these three variants is available a clinical significance value on ClinVar. Each one of these variants was found in only one sample in the analyzed cohort, thus, a validation in a larger cohort would be necessary to confirm this preliminary finding.



Unsupervised clustering of variants highlights patients' stratification

Principal Component Analysis (PCA) of the whole genetic mutational profile (including 57,005 variants) in this dataset is reported in Fig. 1, where quite a sparse distribution of the 48 sample is visible. The first eighteen principal components explain 50% of the overall variance of the dataset. The main contributor of PC1 is a variant in the genomic region of the gene *PARKN* (chr6-163069504-G-A), while PC2's main contributor is a SNV in the genomic region of the gene *PICALM* (chr11-85710180-G-A). K-means clustering based on PC1 and PC2 identified two numerically similar classes, separating samples along PC1 (Fig. 2). None of the available technical, biological or clinical features, as well as other known sources of variation like sex and geographical origin of the patient, matched the clusterization. Taken together these two findings suggest an overall homogeneous genetic background between the two subtypes, as also depicted by the allele frequencies analysis between subtypes. Despite this, a clusterization based mainly on variants in the *PARKN* genomic region is observed, that anyway does not match any of the most important phenotypic features or confounders.

Intronic variants in PRNP and FERMT2 discriminate MM1 and VV2 patients

Supervised classifiers were used for automatic recognition of genetic patterns among the 57,005 variants identified in this dataset. Decision trees were selected as classifiers due to their previous employment in

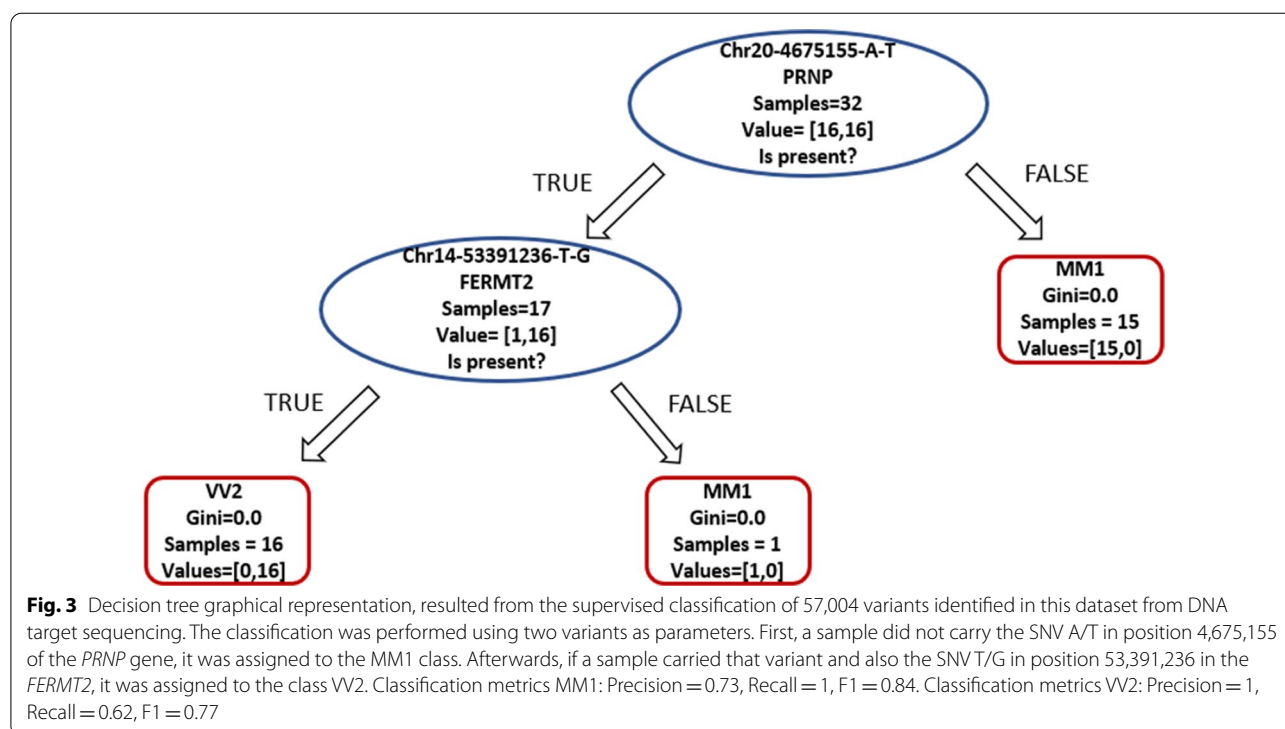
clinical genomics and precision medicine applications to interpret the role of genetic variants in complex diseases [20, 21]. The classification was achieved perfectly, with 100% accuracy on the test set, based on the codon 129 (chr20-4680251-A/G). To test for other recurrent genetic patterns that could characterize the two phenotypic groups, we removed from the input data the variant corresponding to the codon 129. As expected, accuracy decreased both in the training set and in the test set, but interestingly the classifier managed to distinguish the two diseases with good accuracy (training = 1, test 0.81). The classification is based on two intronic variants, in *PRNP* and *FERMT2* genes (Fig. 3). According to common databases and genomic search engines such as VarSome [22], OMIM [23], ClinVar [24] or HGMD [25], the intronic variant in *FERMT2* was never previously reported as a functional intronic variant. The intronic variant in *PRNP* is also referred to with the ID rs6037932, this variant was used in a phylogenetic study about founder effect in another prion disease (FFI) in 2008 [26]. This SNP is in the intronic region between exon 1 and 2, 5 kb away from codon 129. This SNP is not in complete linkage disequilibrium with the allele 129 V, even though in this cohort and in the previously cited work, it is more frequently associated with Valine. On the other hand, the *FERMT2* variant Chr14-53391236-T-G in this cohort is always associated with the allele 129 V.

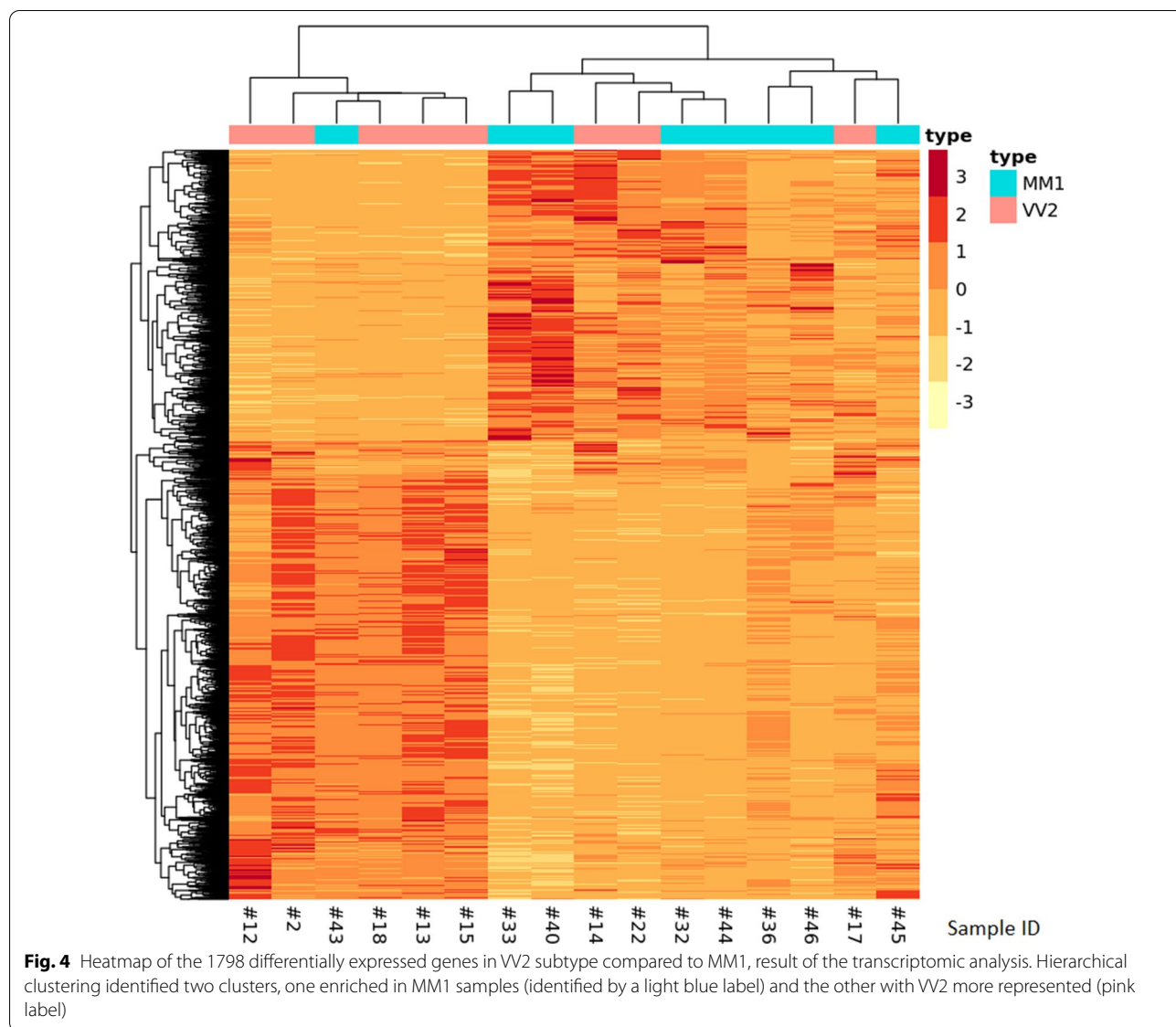
VV2 and MM1 patients show characteristic gene expression profiles

From DGE analysis, 1696 differentially expressed protein coding transcripts were identified in the comparison between VV2 and MM1 (Fig. 4). Among them, 1153 transcripts are significantly over expressed in VV2 compared to MM1 (518 with Log Fold Change (LFC) > 1, Additional file 1: Table S6), while 543 are significantly under over expressed in MM1 compared to VV2 (209 with LFC < -1, Additional file 1: Table S7). Differential expression of the six genes with the most extreme LFC values was validated with digital droplet PCR (ddPCR) (Additional File 1, top 10 Differnetially Expressed Genes (DEG) per subtype and ddPCR), confirming the trend observed in RNAseq. The heatmap of the expression profiles of differentially expressed genes reported in Fig. 4 shows a good clusterization of the two subtypes based on the results of the differential gene expression analysis: two clusters are identified, a smaller one made of five VV2 samples and one MM1 sample, and a larger cluster made of ten samples, seven of which are MM1 and three are VV2.

Functional enrichment analysis shows impairment of pivotal pathways for SCJD pathology

To gain functional insights about the consequences of differentially expressed genes, functional enrichment was performed with three different computational methods: over representation analysis, gene set enrichment





analysis and protein–protein interaction networks. Over representation analysis of MM1 overexpressed genes, identified 21 pathways with a significant enrichment: the most affected pathways were regulatory pathways mediated by guanosine triphosphatases (GTPases), regulation of catabolic processes and maintenance of proper cell morphology and matrix organization (Fig. 5, Additional file 1: Table S5).

In the VV2 group, based on the list of 1153 significantly overexpressed genes, 634 biological processes were significantly enriched, with several pathways referring to synaptic regulation and vesicle trafficking associated with the strongest enrichment (Fig. 5). To highlight the functional modules starting from hundreds of interconnected and overlapping pathways, Cytoscape was used

to summarize non redundant functional modules and visualize interconnections between pathways. As summarized in Fig. 6 and Table 2, pathways involved in synaptic regulation and regulation of endo- and exo-cytosis are not only associated with the lowest p values, but they also represent the most numerous classes of pathways grouped in the functional modules plotted in the network.

Gene set enrichment analysis (GSEA) assumes that also weaker but coordinated changes in sets of functionally related genes can have significant effects, therefore in this type of functional analysis all transcripts are considered. Despite being based on different assumptions compared to the previous approach, these results reinforce the previous findings of increased regulation

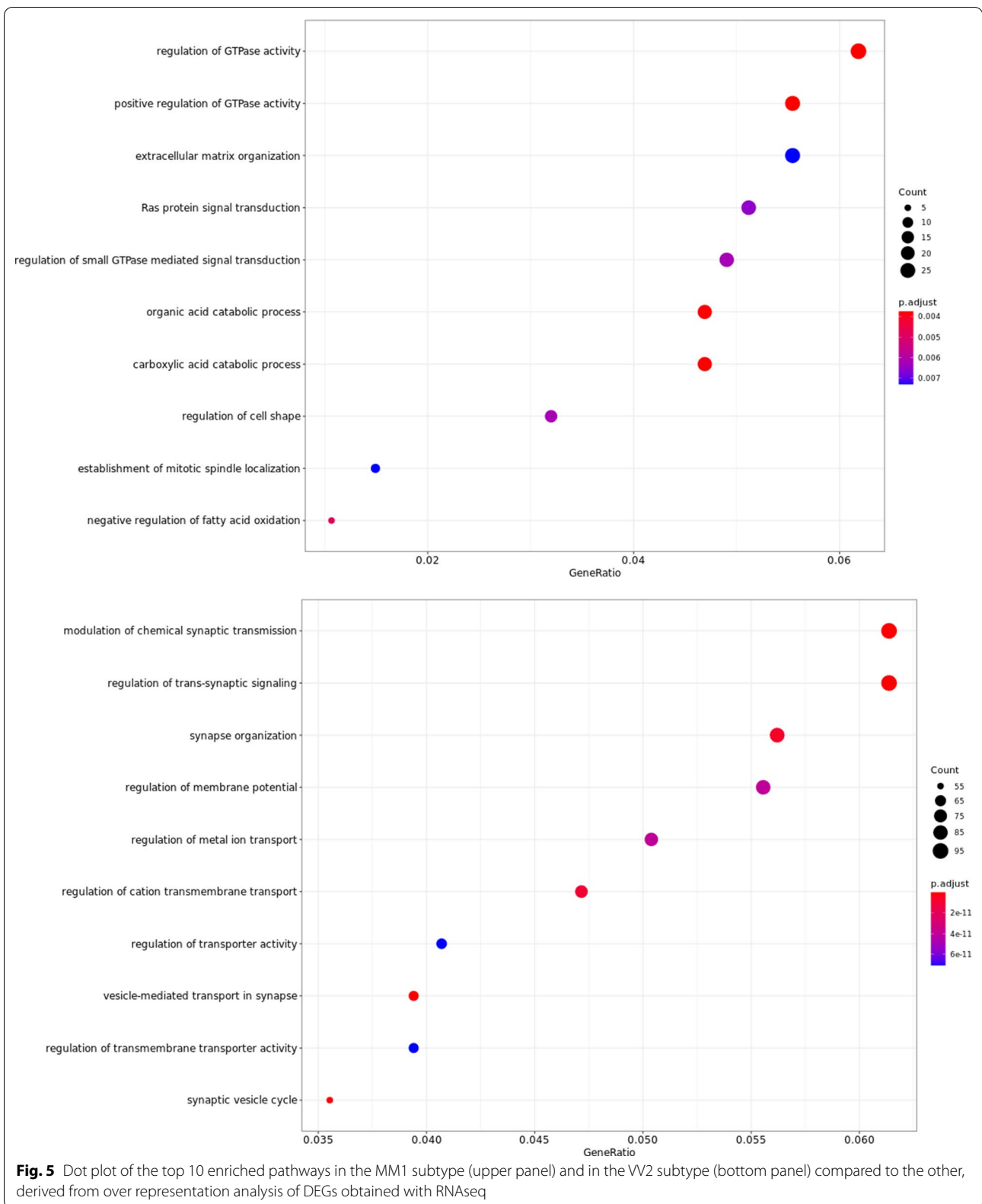
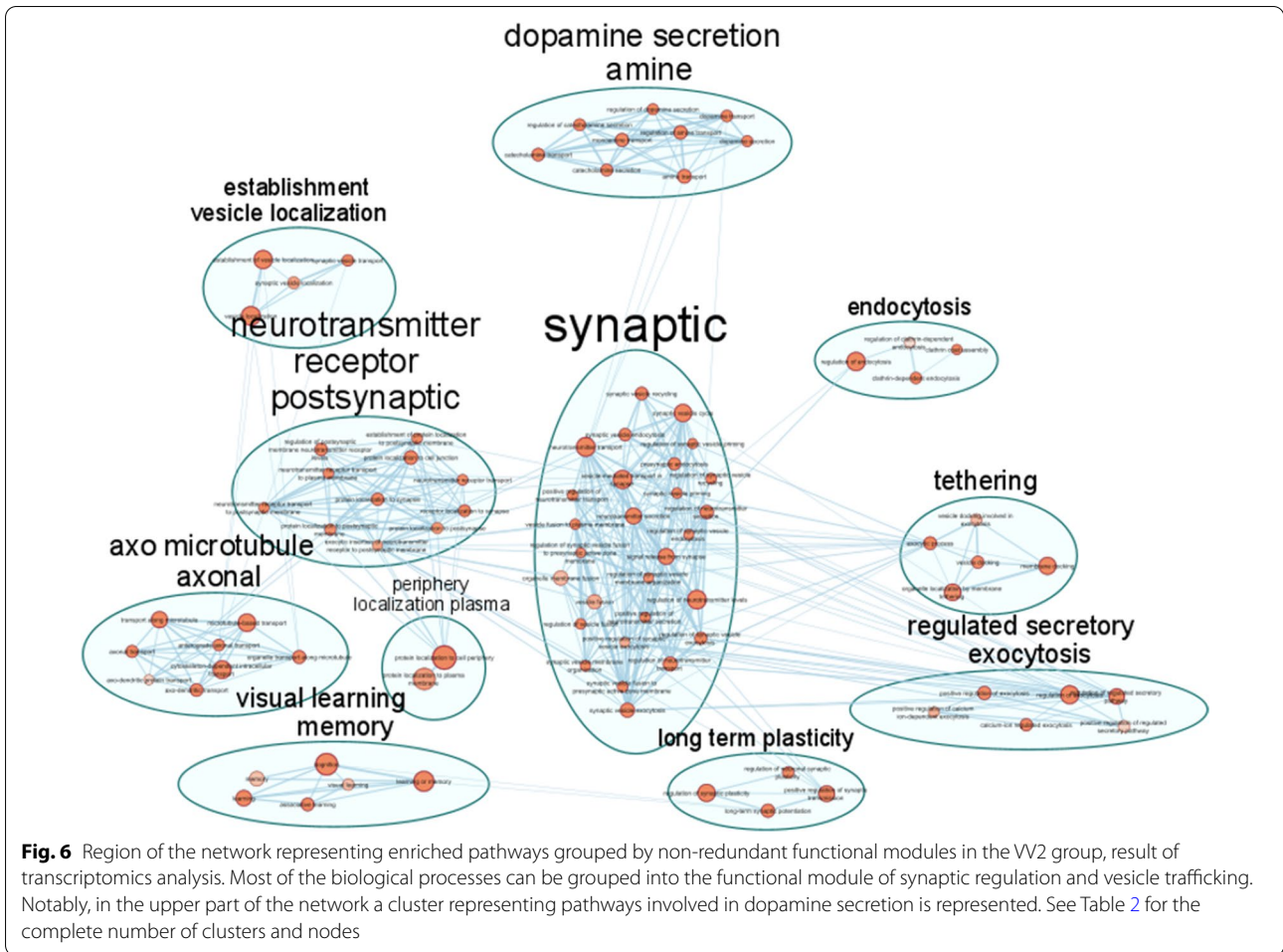


Fig. 5 Dot plot of the top 10 enriched pathways in the MM1 subtype (upper panel) and in the VV2 subtype (bottom panel) compared to the other, derived from over representation analysis of DEGs obtained with RNAseq



of synaptic functionality in the VV2 group compared to the MM1, both in structural terms affecting neuronal projection and axonogenesis, and in functional terms affecting synapse activity with altered regulation of mono and divalent cations transport. Lastly, Protein–Protein Interaction networks (PPIs) provide information on physical interactions of proteins encoded by differentially expressed genes. Here, only direct biophysical interactions were considered (i.e., molecular docking) mapped on STRING interactome and based on these interactions, pathway enrichment was performed on KEGG database. Also with this approach, synapse regulation results to be the most evident pathway differently affected in this comparison between VV2 and MM1 subtypes, and in the MM1 group highlighted an impairment of the regulation of nucleotide binding mechanisms, which is coherent with the previously described result of a positive regulation of GTPase activity.

Integration of DNA and RNA sequencing data highlights RNA editing events

In the genomic regions in which both DNA and RNA sequencing data were available, RNA editing events were analyzed. After applying the quality filters reported in materials and methods, 36 sites in the VV2 and 31 sites in MM1 corresponding to A-to-G changes were identified, consistent with ADAR-mediated RNA editing. The full table of all editing events is reported in Additional File 1: Table S3. We found 15 genes carrying at least one edited site, no significant differences between MM1 and VV2 CJD in terms of number of events and genes involved were observed (Fig. 7). As expected by previous works on RNA editing in aging and neurodegeneration [27], many of the editing events are annotated in regulatory regions such as splicing regions and 3'UTRs, even though also modifications in coding regions were found (Table 3). Interestingly, the genes harboring the highest number of editing events, *TBP* (16), *VPS13C* (12), *HNRNPA2B1*

Table 2 Non-redundant functional modules grouping pathways showing a significant enrichment based on the overexpressed genes in VV2, derived from transcriptomics experiments

Cluster	Nodes	Cluster	Nodes
Vesicle fusion synaptic	28	Regulation dendrite morphogenesis	4
Activity transmembrane transporter	15	Response metal substance	4
Assembly synapse pre-synapse	12	Central nervous neuron	3
Muscle contraction cardiac	11	Chemical postsynaptic excitatory	3
Neurotransmitter receptor postsynaptic	11	Regulation depolarization potential	3
Dopamine secretion amine	9	Amino acid starvation	2
Membrane mitochondrial permeability	9	Peptide hormone insulin	2
Microtubule-mediated axonal transport	8	Periphery localization plasma	2
Anion chloride transmembrane	7	RAC signal transduction	2
Cellular response stimulus	7	Regulation macro autophagy	2
Regulation ph reduction	7	Negative regulation cell	1
Developmental growth extension	6	Negative regulation microtubule	1
Regulated secretory exocytosis	6	Neurofilament cytoskeleton organization	1
Visual learning memory	6	Neuron apoptotic process	1
Adult walking behaviour	5	Neuron recognition	1
Cytosol sarcoplasmic reticulum	5	Post Golgi vesicle	1
Docking organelle tethering	5	Protein folding	1
Positive protein intracellular	5	Protein homo-oligomerization	1
Sodium ion transmembrane	5	Regulation dephosphorylation	1
Clathrin coat endocytosis	4	Regulation proteasomal protein	1
Dendritic spine organization	4	Response temperature stimulus	1
Establishment vesicle localization	4	Spontaneous synaptic transmission	1
Long term plasticity	4	Synaptic transmission gabaergic	1
Potassium ion transmembrane	4	Synaptic transmission glutamatergic	1

The table shows the complete number of clusters and nodes that are graphically represented in Fig. 6

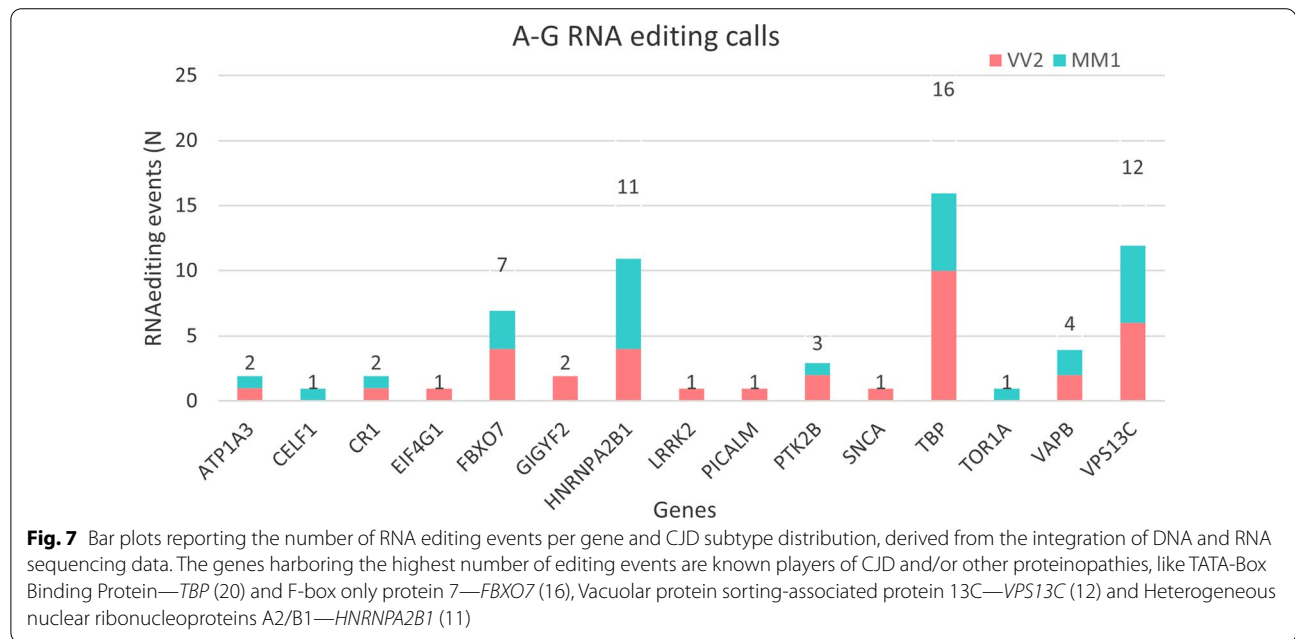


Table 3 List of all the putative RNA editing modifications observed in the cohort of 16 samples, in the 118 genes and regulatory regions covered by both DNA target sequencing and RNA sequencing

Region	Position	Consequence	Class_counts	Gene
Chr19	42,490,154	Splicing, intronic	1 MM1, 1 VV2	<i>ATP1A3</i>
Chr11	47,493,737	Splicing, intronic	1 MM1	<i>CELF1</i>
Chr1	207,741,245	Synonymous	1 VV2	<i>CR1</i>
Chr3	184,039,770	Synonymous	1 VV2	<i>EIF4G1</i>
Chr22	32,871,936	Intronic	3 MM1, 4 VV2	<i>FBXO7</i>
Chr2	233,712,272	Synonymous	2 VV2	<i>GIGYF2</i>
Chr7	26,236,175	Splicing, intronic	4 MM1, 2 VV2	<i>HNRNPA2B1</i>
Chr7	26,236,172	Splicing, intronic	3 MM1, 2 VV2	<i>HNRNPA2B1</i>
Chr12	40,671,989	Missense (Ile723Val)	1 VV2	<i>LRRK2</i>
Chr11	85,742,663	Splicing, intronic	1 VV2	<i>PICALM</i>
Chr8	27,255,263	Synonymous	1 MM1, 2 VV2	<i>PTK2B</i>
Chr4	90,749,343	Splicing, intronic	1 VV2	<i>SNCA</i>
Chr6	170,871,040	Synonymous	2 MM1, 4 VV2	<i>TBP</i>
Chr6	170,871,046	Intronic	2 MM1, 4 VV2	<i>TBP</i>
Chr6	170,871,013	Synonymous	2 MM1, 2 VV2	<i>TBP</i>
Chr9	132,576,302	Synonymous	1 MM1	<i>TOR1A</i>
Chr20	57,022,720	3'UTR	2 VV2 e 2 MM1	<i>VAPB</i>
Chr15	62,306,194	Splicing, intronic	6 MM1, 6 VV2	<i>VPS13C</i>

(11) and *FBXO7* (7), are known players of CJD and/or other proteinopathies. Functional enrichment analysis performed on these 15 genes highlighted a significant ($p_{adj} < 0.05$) involvement of relevant cellular components and biological processes in the disease, such as neuronal cell body and distal axon (GO:0043025, GO:0150034) and synaptic vesicle transport and localization (GO:0048489, GO:0097479) (Additional file 1: Table S4).

Discussion

In this work, we investigated the issue of phenotypic heterogeneity in prion diseases by searching for molecular differences and similarities between the two most common sCJD subtypes, MM1, and VV2, that also represent the principal human prion strains (M1 and V2).

We found several shared features between the sCJD MM1 and VV2. At the genomic level, the two groups showed an overall similar genetic background, as demonstrated by the average number of SNV per sample, the statistical analysis of allele frequencies, and by the principal component analysis. Nevertheless, we also identified several variants that could act as risk factors or phenotypic modifiers. Our results suggest that sCJD could have polygenic contributions able to influence the prevalent strain, opening to the possibility to expand this hypothesis also to carriers of genetic forms of the disease

in future studies. Indeed, the comparison between allele frequencies in the whole sCJD cohort and the healthy European population highlighted several variants with significantly different allele frequencies. Here, functional enrichment provided evidence of a probable downstream impairment of pivotal pathways for CJD pathology -such as chaperon binding, synaptic, and cell death regulation-suggesting a complex genetic background underlying the disease. This is also supported by the finding of probably pathogenic variants in other genes than *PRNP*. In this regard, we found meaningful differences between the two subtypes, with specific genes affected by probably pathogenic variants only in one subtype. In the MM1 subtype, we found variants in genes involved in a heterogeneous group of molecular processes and neurodegenerative diseases, while in the VV2, we found an over-representation of genes involved in PD, such as *HTRA2*, *PINK1*, and *PRKN* [28, 29]. *PINK1* exerts a protective role against mitochondrial dysfunction by activating mitochondrial quality controls mechanisms that mediate mitophagy and lysosomal function through phosphorylation of other mitochondrial proteins such as the E3 ubiquitin-protein ligase *PRKN* [30]. In addition, *PINK1* is also involved in the mitochondrial unfolded protein response (UPR^{mt}) through its interactions with *HTRA2* [31]. Mitochondrial dysfunction has been associated with several neurodegenerative disorders, but very few studies are available in sCJD. Recently, Flønes et al. [32] showed a positive correlation between the level of impairment of the five respiratory complexes in neurons of both MM1 and VV2 CJDs with the severity of other neuropathological changes such as gliosis, vacuolation, and PrP^{Sc} accumulation. The putative pathogenic variants related to PD are found only in the VV2 subtype, suggesting a possible overlapping mechanism regarding mitochondrial quality control dysfunction in Parkinson disease and in VV2 sCJD.

At the transcriptomic level, a comparative analysis was carried out on post-mortem frontal cortex with RNA sequencing (cDNA capture). The similarity between VV2-CJD and PD was also evident at the transcriptomic level, where differential gene expression and functional analysis confirmed and expanded the finding of impairment of some key biological processes associated with PD, such as dopamine secretion, regulation of calcium release, GABA signaling, and mitochondrial permeability in sCJD VV2. This finding is supported by the high prevalence of parkinsonism and other movement disorders in prion diseases [33–37]. Specifically, VV2 and MV2 subtypes exhibit the most severe neuropathological changes, as defined by regional lesion profiles, in the midbrain (substantia nigra) and striatum within the spectrum of sCJD subtypes [2]. In particular, the midbrain and striatum are consistently affected from the earliest stages of

the disease in sCJD VV2 [38]. As a result, within the first 2 months of disease onset, about 15% of patients have parkinsonian signs on neurological examination, rising to 35% considering the entire disease course [38].

We also observed in the VV2 subtype an upregulation of genes involved in synaptic regulation affecting both pre- and post-synaptic terminals. Similarly, VV2 showed overexpression of genes involved in vesicle transport and turnover compared to MM1. Both these biological processes have been described as strongly impaired in all CJD subtypes, especially in the mid and late disease stages [12, 14, 17]. Previous studies comparing CJD cases with controls reported that the intermediate phases of the disease are characterized by a substantial impairment of the transcriptional response directed to pathways involving vesicular trafficking, as well as activation of cholesterol synthesis and efflux, glycosaminoglycan metabolism, and sphingolipid synthesis and degradation [12]. Indeed, membrane composition and membrane microenvironment, in particular lipid rafts where the surface GPI-anchored PrP^C is enriched [39], is involved not only in the normal cellular processing and functions of PrP^C (i.e., localization, internalization, and intracellular trafficking) [40] but also in the conformational conversion of PrP^C into PrP^{Sc}. In fact, evidence suggests that PrP^{Sc} can form pathogenic aggregated while bound to cellular membranes [41] and that PrP^{Sc} formation is most efficient when both PrP^C and PrP^{Sc} are anchored to contiguous membranes [42]. Additionally, it was shown that given their C-terminal attachment site, the GPI anchors twist along one side of the fibril while binding to membranes causing membrane distortions that are likely critical in disease pathogenesis and contribute to explaining the much faster course of CJD compared to other disorders [43].

The observed overexpression in VV2 compared to MM1 was concordant between all functional enrichment methods, suggesting that in the VV2 subtype the downregulation of these processes may occur with less intensity. Previous works have already demonstrated that in CJD genes regulating these pathways are under-expressed compared to controls [14, 44]. In later disease stages, the impairment increases as genes associated with synaptic transmission and axon guidance are downregulated, and ultimately cellular processes related to cell death are activated [11, 12, 17].

In the MM1 group, the most affected pathways involved guanosine triphosphatases (GTPases)-mediated regulatory pathways, regulation of catabolic processes, maintenance of proper cell morphology and matrix organization. The small guanosine triphosphatases (GTPases) of the Ras superfamily are important regulators of key cellular processes such as cell cycle regulation,

proliferation, intracellular trafficking, and apoptosis. Their involvement in neurodegenerative diseases is linked to many processes, particularly impairment of catabolic processes, vesicular trafficking, and regulation of apoptosis [45–47]. Notably, aberrant activity of GTPases and their regulators has been reported and studied extensively in the most common neurodegenerative disorders such as PD, ALS/FTD, and AD for its key role in synaptic maintenance and even as a possible therapeutic target in AD [48, 49]. The other pathway found to be differentially altered in sCJDMM1 is the dysregulation of catabolic process. This is another pathway that has been reported in all proteinopathies [45, 50]. Given their fundamental importance in the functional alteration of neurons and glial cells associated with the accumulation of misfolded proteins, these pathways are shared by several neurodegenerative diseases. Coherently with the genomic findings, MM1-CJD portrayed a gene expression profile where several biological processes shared by different neurodegenerative diseases were involved without an evident distinctive trait.

Interestingly, we found no significant differences between the two subtypes in terms of cell-types involved in immune response and neuroinflammation (see also cell type enrichment in Additional file 1: Table S9). According to the current literature [11, 12, 51], in the early stage of the disease, major changes in gene expression occur involving the immune response through the complement system and leukocyte infiltration, associated with the activation of microglia and astrocytes. Neuroinflammation and activation of the immune response are regulated from the earliest stages of the disease [51], therefore the lack of significant differences could be explained by the fact that in the terminal part of the disease these processes are already fully activated and widespread in both subtypes in a similar manner [52]. Lastly, the integration of genomic and transcriptomic data confirmed the previous findings of RNA editing events in CJD [19]. RNA editing is known to be involved in several neurodegenerative diseases [53, 54], and was described in animal models of CJD [19]. While our results confirm the presence of this type of epigenetic modification, none of the genes identified by Kanata and colleagues in the preclinical stages of the mouse model was found to be modified in the terminal stages of the human disease. The genes harboring the highest number of A-G editing events in our dataset were *TBP* and *FBXO7* (16), *VPS13C* (12), and *HNRNPA2B1*, which are strongly associated with other proteinopathies, such as PD or FTD/ALS. Accordingly, the functional enrichment analysis results on these genes again highlighted the involvement of pivotal pathways impaired in the disease, such as vesicle trafficking and protein quality control. RNA editing is

still scarcely studied in prion diseases, even though it is the object of growing attention in other more common neurodegenerative disorders. All the described genes interested in RNA editing in our cohort are present in the annotation database of A-to-I RNA editing in AD (<https://ccsm.uth.edu/Adeditome>), [54] even though with different loci interested in A-G editing. These findings, together with previous works, highlight the presence of these post-transcriptional modifications in the brains of both human and animal models of CJD; therefore, further studies are needed to understand better the functional effects of these modifications in the disease. Our study is not free of limitations. In interpreting our results, the following should be considered: (1) the relatively small size of the cohorts analyzed; (2) the lack of proteomic data that prevents drawing conclusions about their downstream effects; and (3) given the focus on capturing differences between the two sCJD subtypes, which eliminates shared alterations and consequently impacts on the altered pathways identified, the impossibility of a comprehensive comparison with data sets from other neurodegenerative diseases. Further studies on RNA editing in prion diseases and neurodegeneration coupling genomic, transcriptomic, and proteomic data could clarify these open issues.

Conclusions

In this work, we performed a multi-omics analysis of the two most common subtypes of sCJD in human tissue samples. We identified several putative genetic contributors to the disease onset and phenotype and profiled subtype-specific gene expression alterations, revealing some type-specific genetic signatures and functional similarities only between VV2 CJD and Parkinson's disease. This multi-omics analysis also provided evidence of RNA editing modifications in the disease, confirming results previously obtained in CJD mouse models. These results show a complex misregulation involving alterations at genomic, transcriptomics, and epigenetic (RNA editing) levels in CJD. This work improves the understanding of the molecular pathology of the disease, highlighting novel putative players in the disease onset and its phenotypic heterogeneity, representing a step forward in the state-of-the-art in this field both from a biological and technological perspective.

Methods

Samples and experimental design

In this work, we performed DNA target sequencing on 118 genes in a cohort of 48 MM1 and VV2 sCJD, and, on a subset of 16 samples, RNA sequencing on postmortem brain. Forty-eight patients diagnosed with definite

sporadic CJD, accordingly to the updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease [55], afferent to the IRCCS Institute of Neurological Sciences of Bologna were selected. We selected samples belonging to patients with genotype Met–Met (N=24) or Val–Val (N=24) at codon 129 of the *PRNP* gene, without any co-pathology, and with a clinically “pure” phenotype equally distributed between male and females (Additional file 1: Table S1). Ethical approval was obtained from the ethical board of our institution. For all subjects, written informed consent was provided. All methods were performed in accordance with the relevant guidelines and regulations.

Next generation sequencing

Genomic DNA from peripheral blood or brain tissue was isolated using the Maxwell 16 extractor (Promega, Madison, WI, USA). DNA was quantified using the Quantus Fluorometer (Promega) with QuantiFluor double-stranded DNA system. DNA libraries were prepared with DNA Prep with Enrichment kit (Illumina, CA, USA) performing enrichment with Illumina Neurodegeneration panel (full gene list in Additional file 1: Table S8). We found this product the most appropriate for this work due to the combination between a standardized and field-specific targeted panel with the possibility to also cover also non-coding regulatory regions. Library prep was performed following instruction provided by the vendor. Paired end sequencing (150 × 150) was performed with a NextSeq 500 (Illumina, CA, USA).

On a subset of the previous forty-eight cohort, sixteen brain samples of sCJD patients with subtypes MM1 (N=8) or VV2 (N=8) were selected for RNA sequencing. All selected samples had tissue suitable for this analysis (body kept refrigerated (2–4 °C) before autopsy and with a post-mortem < 36 h to minimize RNA degradation) with mild to moderate lesions in the frontal cortex on histopathological examination. Total RNA was extracted from 50 mg of frozen frontal cortex with RNeasy Lipid Tissue Mini Kit (Qiagen) and quantified with NanoDrop 2000 (Thermo Scientific). Total RNA was subsequently treated with DNase I, RNase-free (Thermo Scientific). Quality assessment of total RNA's quality was performed through capillary electrophoresis with Fragment Analyzer system (Agilent Technologies) with RNA Kit (15nt) (Agilent Technologies). RNA libraries were prepared with Truseq RNA Exome (Illumina) performing enrichment with Illumina Exome Panel—Enrichment Oligos Only, following the protocol provided by the vendor. Paired end sequencing (75 × 75) was performed with a NextSeq 500 (Illumina).

Bioinformatic analysis

Bioinformatic analysis of sequencing data was performed with in-house pipelines using the Snakemake [56] workflow management system for process optimization [57]. In both cases, after demultiplexing, FASTQs quality was checked with FASTQC [58] and trimming was performed with Trimmomatic [59]. For DNA target sequencing data, FASTQ files were mapped to the reference genome (GRCh37/hg19) with the Burrows-Wheeler Aligner [60] using *bwa-mem* algorithm. Sorting, PCR duplicates marking and depth of coverage was computed with GATK [61]. Variant calling of Single Nucleotide Polymorphisms (SNPs) and small indels was performed using Strelka2 [62], setting the analysis for germline variant discovery. Variants were then filtered according to quality parameters and constraints provided by the genomic regions covered by the target sequencing panel. Variant Call Format (VCF) files were annotated with BaseSpace Variant Interpreter (Illumina). Variant discovery was validated on multiple functional loci, such as codon 129 in the *PRNP* gene and/or *APOE* genotype. Missense variants effect prediction was estimated with SIFT [63] and Polyphen2 [64]. RNA sequencing data were mapped to the reference genome (GRCh37) with STAR aligner [65] with default setting and read counts quantification. Aligned bam files are sorted and marked for PCR duplicates with GATK [61]. RNA editing events were assessed with REDIttools [66] (with default setting unless stated) in the 16 samples in which both DNA and RNA sequencing data were acquired. Despite it is in principle possible to perform RNA editing calls on RNAseq data alone, we chose a conservative approach limiting the analysis at the 118 genes covered by both DNA target sequencing and RNA sequencing. Using as reference in this part the work of Wu et al. [54] we applied the following-more stringent- quality filters to keep only well supported RNA editing calls: only calls not present in the DNA data were kept and called with a coverage higher than 30, a Q score higher than 30, and a frequency higher than 0.2 in the RNA data were kept, with the aim of removing calls with too few support on the RNA (possibly technical artifacts). Finally, only A-to-G changes, consistent with A-to-I editing that represent the primary types of ADAR-mediated RNA editing, were kept.

Statistical analysis of genomic variants

The genetic information contained in VCF files was transformed into binary data through an in-house python script (<https://github.com/UniboDIFABiophysics/binarvVCF>), generating a matrix in which each row represents a variant reported in the provided VCF files at least once (which is encoded as chromosome number-position-reference allele- alternate allele) and each column is named

after an ID assigned to each sample. In each cell of the matrix is reported the number of alternative alleles for each locus, thus 0 indicates that the variant is not present in the VCF file of the patient whereas 1 indicates its presence in heterozygosity and 2 the presence of the variant in homozygosity. This matrix was used as input for machine learning methods and statistical analysis, following the workflow recently proposed by Tarozi et al. [21]. Allele frequencies of each variant described at least in one patient of our cohort were calculated and then compared with those reported in the GnomAD database [67] for the non-Finnish European population using Fisher's exact test and Benjamini-Hochberg multiple test correction. The same statistical test was used to compare allele frequencies between samples from MM1 and VV2subtypes.

Machine learning analysis of genomic variants

The overall genomic information of the dataset carried in the ternary matrix was used as input for supervised and unsupervised analysis, using SciKit-Learn [68], Seaborn [69], Plotly express [70], pandas, Numpy [71] and SciPy [72] packages. To visualize such high dimensional data, dimensionality reduction was achieved with Principal component analysis (PCA). SciKit Learn dendrograms and K-Means were used as clustering. As supervised methods, decision trees on binary data labelled accordingly to the different subtypes of each patient were used. The classifier was trained on a random selection of 2/3 of the dataset and adequate branching depth was set to avoid overfitting. The classification rules were tested on a validation set represented by the remaining 1/3 of the dataset. Results are expressed through the parameters Precision, Recall and F1 score. Precision is the ratio of correctly predicted observation to the total predicted positive observations (True Positive/True Positive + False Positive), Recall is the ratio of correctly predicted positive observations to all observations in actual class (True Positive/True Positive + False Negative), F1 Score is the harmonic mean of Precision and Recall ($F1\ Score = 2 \times (Recall \times Precision) / (Recall + Precision)$).

Differential gene expression analysis

Differential gene expression (DGE) was computed on read counts files output of STAR, using DeSeq2 [73]. Based on the gene-level QC performed with default settings, sex, experimental batch, disease duration and post-mortem conservation values were used as parameters of internal sources of variation. Shrinkage parameters of log2FoldChange were estimated with "apeglm" [74] method on the comparison between VV2 and MM1. Significance cut-off of $p < 0.05$ was used and multiple test

correction was performed with Benjamini-Hochberg. Results of DGE analysis were the annotated using the R package Annotables on human genome GRCh37. Validation was performed with digital droplet PCR (Additional File 1, “RNAseq Validation: Digital Droplet PCR”).

Functional enrichment analysis

Functional enrichment analysis was performed with over-representation analysis (ORA) and Functional Class Scoring (FCS) approaches. Functional analysis with both approaches was performed using the Bioconductor packages “ClusterProfiler” [75], “g:profiler” [76], “DOSE” [77] and “Pathview” [78] on Gene Ontology [79] and KEGG [80] databases. Representation of enriched over and under expressed pathways was performed using EnrichmentMap and Annotables on Cytoscape [81]. To further explore the biological interplay of differentially expressed genes, we performed protein–protein interaction analysis (PPI) mapping under and over expressed genes on STRING interactome [82], considering only physical interactions. Results were functionally interpreted with over representation analysis using KEGG database to discover pathways significantly enriched based on the physical interactome.

Abbreviations

SCJD: Sporadic Creutzfeldt-Jakob disease; PD: Parkinson disease; AD: Alzheimer disease; DGE: Differential gene expression; LFC: Log fold change; DEGs: Differentially expressed genes; GO: Gene ontology; GSEA: Gene set enrichment analysis; PCA: Principal component analysis; SNV: Single nucleotide variant; PPI: Protein–Protein interaction.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40478-022-01483-9>.

Additional file 1. Table S1: Clinical information on the dataset used in the genomic layer. For each sample, are reported subtype, disease duration expressed in months, age of onset of the disease (years), markers of co-pathology and biological sex. Males and females were equally distributed in the cohort and in the two subgroups, the average age of onset in the MM1 group was 65.8 years ($\sigma = 7.6$ years) and in the VV2 was 65.7 ($\sigma = 8.5$ years). Disease duration in the MM1 group was 3.1 ($\sigma = 1.8$ months) and in the VV2 6.8 months ($\sigma = 2.4$ months). **Table S2:** most relevant results of the functional analysis performed on the 36 genes harboring at least one variant with a significantly different allele frequency in the sCJD cohort compared to the healthy European population. Legend: GO:BP = Gene Ontology Biological Processes, GO:CC = Gene Ontology Cellular Component, WP = Wiki Pathway database. **Table S3:** Sites in which RNA editing events were observed in this study. “Chromosome” and “Position” define the genomic locus in which the editing event was observed, “Consequence” the predicted functional change. **Table S4:** Top results of functional enrichment analysis of genes involved in RNA editing modifications in the full cohort. **Table S5:** Top biological processes over expressed according to over-representation analysis of differentially expressed genes in the MM1 compared to VV2 samples. **Table S6:** Top fifteen most statistically significant differentially overexpressed genes in the VV2 subtype compared to MM1, according to RNA sequencing data. **Table S7:** Top fifteen most statistically significant differentially overexpressed genes in the MM1 subtype compared to VV2, according to RNA sequencing data.

Table S8: Genes analyzed in the target sequencing analysis with the Neurodegeneration panel (Illumina). **Table S9:** Enrichment scores of cell-type enrichment analysis derived from bulk RNA seq. Here are reported enrichment scores for Astrocytes, Endothelial cells, Macrophages, Macrophages M1, Macrophages M2 and Neurons for each sample and the average value per disease subtype.

Acknowledgements

The authors are grateful to the patients and their families. This work was supported by the Institute of Neurological Sciences (IRCCS) of Bologna and by Alma Mater Studiorum, University of Bologna.

Author contributions

MT sequencing experiments, data analysis, drafting and revising the manuscript; CS contributed to data analysis; ABS, SB, PP and SC data acquisition, GC, SC, PP and MT design of the work and revision of the manuscript. All authors contributed to the article and approved the final manuscript.

Funding

This work is funded by the University of Bologna and the IRCCS Institute of Neurological sciences of Bologna.

Availability of data and materials

The data supporting the conclusions of this article are available in the European Variation Archive - EMBL-EBI (<https://www.ebi.ac.uk/eva/>) and European Nucleotide Archive - EMBL-EBI (<https://www.ebi.ac.uk/ena/>). Genomic data (VCF files) are associated with the ID Project: PRJEB57852. FASTQ files of RNAseq experiment are associated with the ID: Project: PRJEB57720.

Declarations

Ethics approval and consent to participate

For all subjects, written informed consent by patient or next of the kind was provided. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Local Ethics Committee.

Consent for publication

Not applicable.

Competing interests

The authors have no competing interests to declare.

Author details

¹Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40139 Bologna, Italy. ²Programma di Neuropatologia delle Malattie Neurodegenerative, IRCCS Istituto delle Scienze Neurologiche di Bologna, 40139 Bologna, Italy. ³Department of Biomedical and Neuromotor Sciences, University of Bologna, 40139 Bologna, Italy.

Received: 17 October 2022 Accepted: 20 November 2022

Published online: 14 December 2022

References

1. Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. *Science* 216:136–144. <https://doi.org/10.1126/science.6801762>
2. Parchi P, Giese A, Capellari S, Brown P, Schulz-Schaeffer W, Windl O et al (1999) Classification of sporadic Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. *Ann Neurol* 46:224–233. [https://doi.org/10.1002/1531-8249\(199908\)46:2%3c224::AID-ANA12%3e3.0.CO;2-W](https://doi.org/10.1002/1531-8249(199908)46:2%3c224::AID-ANA12%3e3.0.CO;2-W)
3. Baiardi S, Rossi M, Capellari S, Parchi P (2019) Recent advances in the histo-molecular pathology of human prion disease. *Brain Pathol* 29:278–300. <https://doi.org/10.1111/bpa.12695>
4. Rossi M, Baiardi S, Parchi P (2019) Understanding prion strains: evidence from studies of the disease forms affecting humans. *Viruses* 11:1–27. <https://doi.org/10.3390/v11040309>

5. Baiardi S, Magherini A, Capellari S, Redaelli V, Ladogana A, Rossi M, Tagliavini F et al (2017) Towards an early clinical diagnosis of sporadic CJD VV2 (ataxic type). *J Neurol Neurosurg Psychiatry* 88:764–772. <https://doi.org/10.1136/JNPN-2017-315942>
6. Zerr I, Parchi P (2018) Sporadic Creutzfeldt-Jakob disease, 1st edn. Elsevier
7. Bishop MT, Will RG, Manson JC (2010) Defining sporadic Creutzfeldt-Jakob disease strains and their transmission properties. *Proc Natl Acad Sci USA* 107:12005–12010. <https://doi.org/10.1073/PNAS.1004688107>
8. Kobayashi A, Sakuma N, Matsuura Y, Mohri S, Aguzzi A, Kitamoto T (2010) Experimental verification of a traceback phenomenon in prion infection. *J Virol* 84:3230–3238. <https://doi.org/10.1128/JVI.02387-09>
9. Parchi P, Cescatti M, Notari S, Schulz-Schaeffer WJ, Capellari S, Giese A, Zou WQ, Kretzschmar H, Ghetti B, Brown P (2010) Agent strain variation in human prion disease: insights from a molecular and pathological review of the National Institutes of Health series of experimentally transmitted disease. *Brain* 133:3030–3042. <https://doi.org/10.1093/BRAIN/AWQ234>
10. Jones E, Hummerich H, Viré E, Uphill J, Dimitriadis A, Speedy H et al (2020) Identification of novel risk loci and causal insights for sporadic Creutzfeldt-Jakob disease: a genome-wide association study. *Artic Lancet Neurol* 19:840–888. [https://doi.org/10.1016/S1474-4422\(20\)30273-8](https://doi.org/10.1016/S1474-4422(20)30273-8)
11. Hwang D, Lee IY, Yoo H, Gehlenborg N, Cho J-HH, Petritis B et al (2009) A systems approach to prion disease. *Mol Syst Biol* 5:252. <https://doi.org/10.1038/msb.2009.10>
12. Kim TK, Lee I, Cho JH, Canine B, Keller A, Price ND, Hwang D, Carlson G, Hood L (2020) Core transcriptional regulatory circuits in prion diseases. *Mol Brain* 13:1–14. <https://doi.org/10.1186/s13041-020-0551-3>
13. Sorce S, Nuvolone M, Russo G, Chincisan A, Heinzer D, Avar M, Pfam-matter M, Schwarz P, Det al. (2020) Genome-wide transcriptomics identifies an early preclinical signature of prion infection. *bioRxiv* 2020.01.10.901637. <https://doi.org/10.1101/2020.01.10.901637>
14. Bartoletti-Stella A, Corrado P, Mometto N, Baiardi S, Durrenberger PF, Arzberger T et al (2019) Analysis of RNA expression profiles identifies dysregulated vesicle trafficking pathways in Creutzfeldt-Jakob disease. *Mol Neurobiol* 56:5009–5024. <https://doi.org/10.1007/s12035-018-1421-1>
15. Llorens F, Ansoleaga B, Garcia-Esparcia P, Zafar S, Grau-Rivera O, López-González I et al (2013) PrP mRNA and protein expression in brain and PrPc in CSF in Creutzfeldt-Jakob disease MM1 and VV2. *Prion* 7:383–393. <https://doi.org/10.4161/pri.26416>
16. López-González I, Garcia-Esparcia P, Llorens F, Ferrer I (2016) Genetic and transcriptomic profiles of inflammation in neurodegenerative diseases: Alzheimer, Parkinson, Creutzfeldt-Jakob and Tauopathies. *Int J Mol Sci* 17:206. <https://doi.org/10.3390/IJMS17020206>
17. Xiang W, Windl O, Westner IM, Neumann M, Zerr I, Lederer RM, Kretzschmar HA (2005) Cerebral gene expression profiles in sporadic Creutzfeldt-Jakob disease. *Ann Neurol* 58:242–257. <https://doi.org/10.1002/ana.20551>
18. Tian Y, Meng L, Zhang Z (2020) What is strain in neurodegenerative diseases? *Cell Mol Life Sci* 77:665–676. <https://doi.org/10.1007/s00018-019-03298-9>
19. Kanata E, Llorens F, Dafou D, Dimitriadis A, Thüne K, Xanthopoulos K et al (2019) RNA editing alterations define manifestation of prion diseases. *Proc Natl Acad Sci USA* 116:19727–19735. <https://doi.org/10.1073/pnas.1803521116>
20. Machado do Nascimento P, Gomes Medeiros I, Maia Falcão R, Stransky B, Estefano Santana de Souza J (2020) A decision tree to improve identification of pathogenic mutations in clinical practice. *BMC Med Inform Decis Mak*. <https://doi.org/10.1186/s12911-020-1060-0>
21. Tarozzi M, Bartoletti-Stella A, Dall’Olio D, Matteuzzi T, Baiardi S, Parchi P et al (2022) Identification of recurrent genetic patterns from targeted sequencing panels with advanced data science: a case-study on sporadic and genetic neurodegenerative diseases. *BMC Med Genom* 15:1–12. <https://doi.org/10.1186/S12920-022-01173-4>
22. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Aguilera MA, Meyer R, Massouras A, Albarca Aguilera M, Meyer R, Massouras A (2019) VarSome: the human genomic variant search engine. *Bioinformatics* 35:1978–1980. <https://doi.org/10.1093/bioinformatics/bty897>
23. Amberger JS, Bocchini CA, Scott AF, Hamosh A (2019) OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Res* 47:D1038–D1043. <https://doi.org/10.1093/nar/gky1151>
24. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S et al (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46:D1062–D1067. <https://doi.org/10.1093/NAR/GKX1153>
25. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST et al (2003) Human gene mutation database (HGMD): 2003 update. *Hum Mutat* 21:577–581. <https://doi.org/10.1002/HUMU.10212>
26. Rodríguez-Martínez AB, Alfonso-Sánchez MA, Peña JA, Sánchez-Valle R, Zerr I, Capellari S et al (2008) Molecular evidence of founder effects of fatal familial insomnia through SNP haplotypes around the D178N mutation. *Neurogenetics* 9:109–118. <https://doi.org/10.1007/s10048-008-0120-x>
27. Ma Y, Dammer EB, Felsky D, Duong DM, Klein HU, White CC, Zhou M et al (2021) Atlas of RNA editing events affecting protein expression in aged and Alzheimer’s disease human brain tissue. *Nat Commun* 12:1–16. <https://doi.org/10.1038/s41467-021-27204-9>
28. Foroud T, Uniacke SK, Liu L, Pankratz N, Rudolph A, Halter C, Shults C, Marder K et al (2003) Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. *Neurology* 60:796–801. <https://doi.org/10.1212/01.WNL.0000049470.00180.07>
29. Plun-Favreau H, Klupsch K, Moiso N, Gandhi S, Kjaer S, Frith D, Harvey K, Deas E, Harvey RJ, McDonald N, Wood NW, Martins LM, Downward J (2007) The mitochondrial protease HtrA2 is regulated by Parkinson’s disease-associated kinase PINK1. *Nat Cell Biol*. <https://doi.org/10.1038/ncb1644>
30. Kim Y, Park J, Kim S, Song S, Kwon SK, Lee SH, Kitada T, Kim JM, Chung J (2008) PINK1 controls mitochondrial localization of Parkin through direct phosphorylation. *Biochem Biophys Res Commun* 377:975–980. <https://doi.org/10.1016/J.BBRC.2008.10.104>
31. Chen C, Turnbull DM, Reeve AK (2019) Mitochondrial dysfunction in Parkinson’s disease—cause or consequence? *Biology*. <https://doi.org/10.3390/biology8020038>
32. Flønes IH, Ricken G, Klotz S, Lang A, Ströbel T, Dölle C, Kovacs GG, Tzoulis C (2020) Mitochondrial respiratory chain deficiency correlates with the severity of neuropathology in sporadic Creutzfeldt-Jakob disease. *Acta Neuropathol Commun* 8:50. <https://doi.org/10.1186/s40478-020-00915-8>
33. Ragno M, Scarcella MG, Cacchiò G, Capellari S, Di Marzio F, Parchi P, Trojano L (2009) Striatum [123I] FP-CIT SPECT demonstrates dopaminergic deficit in a sporadic case of Creutzfeldt-Jakob disease. *Acta Neurol Scand* 119:131–134. <https://doi.org/10.1111/j.1600-0404.2008.01075.x>
34. Rodríguez-Portel F, Boaratti Ciariariello V, Dwivedi AK, Lovera L, Da Prat G, Lopez-Castellanos R et al (2019) Movement disorders in prionopathies: a systematic review tremor other. *Hyperkinet Mov*. <https://doi.org/10.7916/tohm.v0.712>
35. Sequeira D, Nihat A, Mok T, Coysz T, Rudge P, Collinge J, Mead S (2022) Prevalence and treatments of movement disorders in prion diseases: a longitudinal cohort study. *Mov Disord* 37:1893–1903. <https://doi.org/10.1002/MDS.29152>
36. Tang S, Dou X, Zhang Y (2022) 18F-FP-CIT PET/CT in a case of probable sporadic Creutzfeldt-Jakob disease with parkinsonism as initial symptom. *Prion*. <https://doi.org/10.1080/19336896.2022.2093078>
37. Vital A, Fernagut PO, Cannon MH, Joux J, Bezard E, Martin-Negrier ML, Vital C, Tison F (2009) The nigrostriatal pathway in Creutzfeldt-Jakob disease. *J Neuropathol Exp Neurol* 68:809–815. <https://doi.org/10.1097/NEN.0B013E3181ABDAE8>
38. Baiardi S, Redaelli V, Ripellino P, Rossi M, Franceschini A, Moggio M et al (2019) Prion-related peripheral neuropathy in sporadic Creutzfeldt-Jakob disease. *J Neurol Neurosurg Psychiatry* 90:424–427. <https://doi.org/10.1136/JNPN-2018-319221>
39. Agostini F, Dotti CG, Pérez-Cañamás A, Ledesma MD, Benetti F, Legname G (2013) Prion protein accumulation in lipid rafts of mouse aging brain. *PLoS ONE* 8:e74244. <https://doi.org/10.1371/JOURNAL.PONE.0074244>
40. Campana V, Sarnataro D, Zurzolo C (2005) The highways and byways of prion protein trafficking. *Trends Cell Biol* 15:102–111. <https://doi.org/10.1016/J.TCB.2004.12.002>
41. Rouvinski A, Karniely S, Kounin M, Moussa S, Goldberg MD, Warburg G, Lyakhovetsky R, Papy-Garcia D et al (2014) Live imaging of prions reveals nascent PrPSc in cell surface, raft-associated amyloid strings and webs. *J Cell Biol* 204:423–441. <https://doi.org/10.1083/jcb.201308028>
42. Maxson L, Wong C, Herrmann LM, Caughey B, Baron GS (2003) A solid-phase assay for identification of modulators of prion protein interactions. *Anal Biochem* 323:54–64. <https://doi.org/10.1016/J.AB.2003.07.028>

43. Kraus A, Hoyt F, Schwartz CL, Hansen B, Artikis E, Hughson AG, Raymond GJ, Race B, Baron GS, Caughey B (2021) High-resolution structure and strain comparison of infectious mammalian prions. *Mol Cell* 81:4540–4551.e6. <https://doi.org/10.1016/j.molcel.2021.08.011>
44. Andres Benito P, Dominguez Gonzalez M, Ferrer I (2018) Altered gene transcription linked to astrocytes and oligodendrocytes in frontal cortex in Creutzfeldt-Jakob disease. *Prión* 12:216–225. <https://doi.org/10.1080/19336896.2018.1500076>
45. Gan L, Cookson MR, Petrucelli L, La Spada AR (2018) Converging pathways in neurodegeneration, from genetics to mechanisms. *Nat Neurosci* 21:1300
46. Qu L, Pan C, He SM, Lang B, Gao GD, Wang XL, Wang Y (2019) The ras superfamily of small gtpases in non-neoplastic cerebral diseases. *Front Mol Neurosci*. <https://doi.org/10.3389/FNMOL.2019.00121/FULL>
47. Sastre AA, Montoro ML, Gálvez-Martín P, Lacerda HM, Lucia A, Llavero F, Zugaza JL (2020) Small gtpases of the ras and rho families switch on/off signaling pathways in neurodegenerative diseases. *Int J Mol Sci* 21:1–23. <https://doi.org/10.3390/ijms21176312>
48. Aguilar BJ, Zhu Y, Lu Q (2017) Rho GTPases as therapeutic targets in Alzheimer's disease. *Alzheimer's Res Ther* 9:1–10. <https://doi.org/10.1186/S13195-017-0320-4/TABLES/2>
49. Bolognin S, Lorenzetto E, Diana G, Buffelli M (2014) The potential role of rho GTPases in Alzheimer's disease pathogenesis. *Mol Neurobiol*. <https://doi.org/10.1007/s12035-014-8637-5>
50. Tan SH, Karri V, Tay NWR, Chang KH, Ah HY, Ng PQ, Ho HS, Keh HW, Candasamy M (2019) Emerging pathways to neurodegeneration: dissecting the critical molecular mechanisms in Alzheimer's disease, Parkinson's disease. *Biomed Pharmacother* 111:765–777. <https://doi.org/10.1016/j.biopha.2018.12.101>
51. Sorce S, Nuvolone M, Russo G, Chincisan A, Heinzer D, Avar M, Pfammatter M, Schwarz P et al (2020) Genome-wide transcriptomics identifies an early preclinical signature of prion infection. *PLoS Pathog* 16:e1008653. <https://doi.org/10.1371/JOURNAL.PPAT.1008653>
52. Franceschini A, Strammiello R, Capellari S, Giese A, Parchi P (2018) Regional pattern of microgliosis in sporadic Creutzfeldt-Jakob disease in relation to phenotypic variants and disease progression. *Neuropathol Appl Neurobiol* 44:574–589. <https://doi.org/10.1111/NAN.12461>
53. Costa Cruz PH, Kawahara Y (2021) rna editing in neurological and neurodegenerative disorders. *Methods Mol Biol* 2181:309–330. https://doi.org/10.1007/978-1-0716-0787-9_18
54. Wu S, Yang M, Kim P, Zhou X (2021) A Deditome provides the genomic landscape of A-to-I RNA editing in Alzheimer's disease. *Brief Bioinform* 22:1–12. <https://doi.org/10.1093/bib/bbaa384>
55. Zerr I, Kallenberg K, Summers DM, Romero C, Taratuto A, Heinemann U et al (2009) Updated clinical diagnostic criteria for sporadic Creutzfeldt-Jakob disease. *Brain* 132:2659. <https://doi.org/10.1093/BRAIN/AWP191>
56. Köster J, Rahmann S (2012) Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts480>
57. Dall'Olio D, Curti N, Fonzi E, Sala C, Remondini D, Castellani G, Giampieri E (2021) Impact of concurrency on the performance of a whole exome sequencing pipeline. *BMC Bioinform* 22:1–15. <https://doi.org/10.1186/S12859-020-03780-3>
58. Andrews, Simon, Krueger, Felix , Segonds-Pichon, Anne , Biggins, Laura , Krueger, Christel , Wingett S (2010) FastQC. In: Babraham, UK. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
59. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/BIOINFORMATICS/BTU170>
60. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btp324>
61. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genom Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
62. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X et al (2018) Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. <https://doi.org/10.1038/s41592-018-0051-x>
63. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC (2016) SIFT missense predictions for genomes. *Nat Protoc* 11:1–9. <https://doi.org/10.1038/nprot.2015.123>
64. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. <https://doi.org/10.1002/0471142905.HG0720576>
65. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
66. Picardi E, D'Erchia AM, Montalvo A, Pesole G (2015) Using REDtools to detect RNA editing events in NGS datasets. *Curr Protoc Bioinforma*. <https://doi.org/10.1002/0471250953.bi1212s49>
67. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL et al (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434–443. <https://doi.org/10.1038/s41586-020-2308-7>
68. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825
69. Waskom M (2012) Seaborn: statistical data visualization. In: Seaborn
70. Plotly Technologies Inc. (2015) Collaborative data science, <https://plot.ly>. Plotly Technol. Inc.
71. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D et al (2020) Array programming with NumPy. *Nature* 585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>
72. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E et al (2020) (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17(17):261–272. <https://doi.org/10.1038/s41592-019-0686-2>
73. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:1–21. <https://doi.org/10.1186/s13059-014-0550-8>
74. Zhu A, Ibrahim JG, Love MI (2019) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* 35:2084–2092. <https://doi.org/10.1093/BIOINFORMATICS/BTY895>
75. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J Integr Biol* 16:284–287. <https://doi.org/10.1089/OMI.2011.0118>
76. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A et al (2019) (2019) Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape EnrichmentMap *Nat Protoc* 14(14):482–517. <https://doi.org/10.1038/s41596-018-0103-9>
77. Yu G, Wang L-G, Yan G-R, He Q-Y (2015) DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 31:608–609. <https://doi.org/10.1093/BIOINFORMATICS/BTU684>
78. Luo W, Brouwer C (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 29:1830–1831. <https://doi.org/10.1093/BIOINFORMATICS/BTT285>
79. The Gene Ontology C (2019) The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res*. <https://doi.org/10.17863/CAM.36439>
80. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/NAR/GKV1070>
81. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software Environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>
82. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47:D607–D613. <https://doi.org/10.1093/NAR/GKY1131>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.