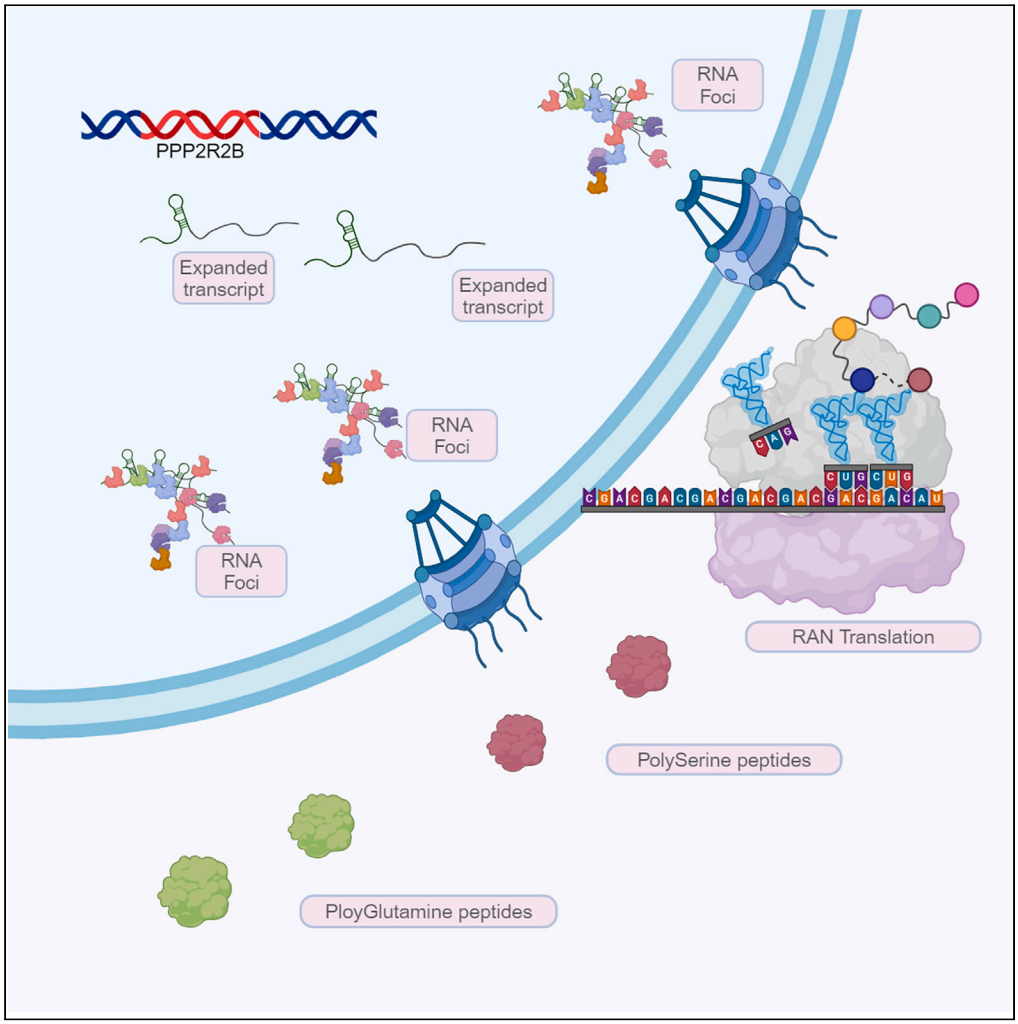
iScience

CellPress
OPEN ACCESS

Article

Molecular clues unveiling spinocerebellar ataxia type-12 pathogenesis



Manish Kumar, Shweta Sahni, Vivekanand A, ..., Himanshi Kapoor, Achal K. Srivastava, Mohammed Faruq

faruq.mohd@igib.res.in

Highlights

RNA foci are found in the patient-derived neural cell lines of SCA12

Expanded CAG in PPP2R2B transcript binds key nuclear proteins

Polyglutamine and polyserine RAN proteins are expressed in SCA12

IPA and mRNA sequencing analysis revealed affected pathways in SCA12

Kumar et al., iScience 27, 109768
May 17, 2024
https://doi.org/10.1016/j.isci.2024.109768

iScience

CellPress
OPEN ACCESS

Article

Molecular clues unveiling spinocerebellar ataxia type-12 pathogenesis

Manish Kumar,[1,2,3] Shweta Sahni,[1,4] Vivekanand A,[1,2,3] Deepak Kumar,[5,6] Neetu Kushwah,[1] Divya Goel,[7] Himanshi Kapoor,[1] Achal K. Srivastava,[4] and Mohammed Faruq[1,2,3,8,*]

SUMMARY

Spinocerebellar Ataxia type-12 (SCA12) is a neurodegenerative disease caused by tandem CAG repeat expansion in the 5′-UTR/non-coding region of *PPP2R2B*. Molecular pathology of SCA12 has not been studied in the context of CAG repeats, and no appropriate models exist. We found in human SCA12-iPSC-derived neuronal lineage that expanded CAG in *PPP2R2B* transcript forms nuclear RNA foci and were found to sequester variety of proteins. Further, the ectopic expression of transcript containing varying length of CAG repeats exhibits non-canonical repeat-associated non-AUG (RAN) translation in multiple frames in HEK293T cells, which was further validated in patient-derived neural stem cells using specific antibodies. mRNA sequencing of the SCA12 and control neurons have shown a network of crucial transcription factors affecting neural fate, in addition to alteration of various signaling pathways involved in neurodevelopment. Altogether, this study identifies the molecular signatures of SCA12 disorder using patient-derived neuronal cell lines.

INTRODUCTION

Spinocerebellar Ataxia-12 (SCA12) is a progressive late-onset autosomal dominantly inherited neurodegenerative disorder characterized by progressive hand tremors, mild to moderate gait ataxia, dysarthria, bradykinesia, hyperreflexia, and other neuropsychiatric symptoms.[1] SCA12 is caused by a CAG repeat expansion within the 5′-UTR region of the *PPP2R2B* gene on chr5q, which in healthy control is 4–32 repeats, whereas the pathogenic threshold for the disease onset is 43.[2] Clinically, SCA12 is a unique ataxia subtype with characteristic features of action tremor of hands in addition to the cerebellar and neuronal features. SCA12 disease exhibits two broad phenotypes in patients viz. the tremor dominant or the gait dominant. Typically, the tremor and ataxia syndrome ensues in the 4th and 5th decade of the patient's life.[3] Although most cases of SCA12 have been reported from the Indian subcontinent, a variety of rare presentations have been reported from different parts of the world.[4]

The only study on the *postmortem* brain of SCA12 patients reported Purkinje cell degeneration with lesions in the cerebellar cortical regions.[1] The same study also reported ubiquitin-positive inclusions in pars compacta and Purkinje cells, negative for polyglutamine, tau, and transactive response DNA-binding protein 43 (TDP-43).[1] Also, it was reported that the CAG repeats act as a promoter element and can increase the expression of the *PPP2R2B* gene.[1] Another study on the plasma of SCA12 suggested the altered protein abundance of transportation and lipid metabolism proteins.[5] The exact molecular pathology of CAG expansion in *PPP2R2B* is unknown; however, a gain-of-function mechanism has been hypothesized. Lin *et al.* (2010), reported that CAG repeats in a reporter construct exhibit an increase in the expression akin to the CGG repeat expansion effect on the FMR1 locus of fragile-X-associated tremor/ataxia syndrome.[6] At the protein level, the effect of CAG repeats has not been elucidated. At the clinical level, it has similarities with fragile-X-associated tremor/ataxia syndrome as well as spinocerebellar ataxias;[7] however, at the molecular level no similar mechanism has been so far elucidated.

We sought to decipher the molecular pathology of CAG repeat expansion within *PPP2R2B* in SCA12 using patient-Induced Pluripotent Stem Cell (iPSC)-derived models of neuronal lineage. The pertinent questions were (1) to explore the molecular correlates of tremor-ataxia phenotype, if it has a matching molecular mechanism as reported for CGG repeats in the 5′UTR of the *FMR1* gene where an RNA gain-of-function mechanism is well established, (2) to explore if the presence of CAG repeat motif exhibits polyglutamine (polyQ) pathology with protein level toxicity and thus, phenotypic equivalence with other SCA types i.e., SCA1-3 and Huntington disease (HD), etc. In the present

[1]Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB), Mall Road, Delhi 110007, India
[2]CSIR-HRDC Campus, Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
[3]Genomics and Molecular Medicine Division, CSIR - Institute of Genomics and Integrative Biology, New Delhi, India
[4]Department of Neurology, All India Institute of Medical Sciences, New Delhi 110029, India
[5]Division of Genomics and Molecular Medicine, CSIR - Institute of Genomics and Integrative Biology (IGIB), New Delhi 110007, India
[6]Department of Zoology, University of Allahabad, Prayagraj, Uttar Pradesh 211002, India
[7]Department of Pharmacology, School of Pharmaceutical Education & Research (SPER), Jamia Hamdard, New Delhi 110062, India
[8]Lead contact
*Correspondence: faruq.mohd@igib.res.in
https://doi.org/10.1016/j.isci.2024.109768

Table 1. Sample details and associated metadata

Sample ID	Cell line ID	CAG Repeat	Design	Condition	Cell line	Sex	Batch
10223NSC-1	10223 NSC	09/16	Control_NSC	Control	NSC	MALE	Batch 2
10223NSC-2	10223 NSC	09/16	Control_NSC	Control	NSC	MALE	Batch 2
21-14NSC-2	21/14 NSC (IGIBi004-A)	17/65	Diseased_NSC	Diseased	NSC	MALE	Batch 2
21-14NSC-1	21/14 NSC (IGIBi004-A)	17/65	Diseased_NSC	Diseased	NSC	MALE	Batch 2
2555NEU	2555 Neuron (IGIBi002-A)	14/59	Diseased_Neuron	Diseased	Neuron	MALE	Batch 2
2555NSC-1	2555 NSC (IGIBi002-A)	14/59	Diseased_NSC	Diseased	NSC	MALE	Batch 2
2555NSC-2	2555 NSC (IGIBi002-A)	14/59	Diseased_NSC	Diseased	NSC	MALE	Batch 2
D161NSC-1	D161 NSC (ADBSi001-A)	10/13	Control_NSC	Control	NSC	FEMALE	Batch 2
D161NSC-2	D161 NSC (ADBSi001-A)	10/13	Control_NSC	Control	NSC	FEMALE	Batch 2
SRR5754278	10223 iPSC	09/16	Control_iPSC	Control	iPSC	MALE	Batch 1
SRR5754279	10223 iPSC	09/16	Control_iPSC	Control	iPSC	MALE	Batch 1
SRR5754280	10223 NSC	09/16	Control_NSC	Control	NSC	MALE	Batch 1
SRR5754281	10223 NSC	09/16	Control_NSC	Control	NSC	MALE	Batch 1
SRR5754282	10223 Neuron	09/16	Control_Neuron	Control	Neuron	MALE	Batch 1
SRR5754283	10223 Neuron	09/16	Control_Neuron	Control	Neuron	MALE	Batch 1
SRR5754576	2555 iPSC (IGIBi002-A)	14/59	Diseased_iPSC	Diseased	iPSC	MALE	Batch 1
SRR5754577	2555 iPSC (IGIBi002-A)	14/59	Diseased_iPSC	Diseased	iPSC	MALE	Batch 1
SRR5754626	2555 NSC (IGIBi002-A)	14/59	Diseased_NSC	Diseased	NSC	MALE	Batch 1
SRR5754627	2555 Neuron (IGIBi002-A)	14/59	Diseased_Neuron	Diseased	Neuron	MALE	Batch 1
SRR5754628	2555 Neuron (IGIBi002-A)	14/59	Diseased_Neuron	Diseased	Neuron	MALE	Batch 1
SRR5754716	11/08 iPSC (IGIBi003-A)	10/67	Diseased_iPSC	Diseased	iPSC	FEMALE	Batch 1
SRR5754717	11/08 iPSC (IGIBi003-A)	10/67	Diseased_iPSC	Diseased	iPSC	FEMALE	Batch 1
SRR5754718	21/14 iPSC (IGIBi004-A)	17/65	Diseased_iPSC	Diseased	iPSC	MALE	Batch 1

study, we utilized *in vitro* iPSC-derived Neural Stem Cell (NSC) lineage models of SCA12 cell lines to reveal the occurrence of RNA-mediated gain-of-function through RNA-foci and non-canonical non-ATG-mediated translation of *PPP2R2B* with peptide frame coding for CAG repeats. We were able to detect the occurrence of RNA foci formation in the nucleus of the SCA12-patient-derived NSC model. We further explored the pathways that might be affected in the SCA12 disease biology through RNA pull-down assay and transcriptomics analysis. In addition, we depicted the process of Repeat Associated Non-AUG-mediated (RAN) translation, which has been reported in other repeat expansion disorders including fragile-X-associated tremor/ataxia syndrome,[8,9] in the *PPP2R2B* gene driven by the expanded CAG repeat motif, both in the transient overexpression model of *PPP2R2B* and in the patient-derived cellular model of SCA12. In summary, our study provides preliminary evidence linking the molecular mechanisms of disease pathology to the tremor-ataxia phenotype as in fragile X-associated tremor/ataxia syndrome and other PolyQ diseases and also provides clues toward the global molecular disruptions caused by the pathogenic CAG repeat.

RESULTS

Pathogenic SCA12 repeats induce RNA foci formation in the nucleus

Repeat-containing transcripts are known to form nuclear aggregates in the form of RNA foci in many repeat-mediated CAG/CUG diseases.[10] To check for the occurrence of similar repeat-containing transcript aggregates in our SCA12 NSC model, we designed probes against the CAG-repeat region. We utilized the RNA FISH technique to determine the location of CAG-containing transcript within the NSCs generated from three patients and two controls with varying repeat lengths (Table 1). SCA12 positive cell lines showed the presence of RNA foci or aggregate-like formations inside the nucleus of cells (Figures 1A–1C). These aggregates were lacking in the secondary neuronal cell lines (SK-N-SH) (data not shown) and control cell lines (Figures 1D and 1E).

RNA foci are considered as a major hallmark of RNA-mediated toxicity in many other neurodegenerative diseases; their presence in the SCA12 disease model links it to the disease pathogenesis. Further, we checked the colocalization of MBNL1(muscleblind-like 1) protein, the protein most commonly known to be associated with RNA foci in other repeat-mediated diseases[10] with the *PPP2R2B* RNA foci. RNA FISH combined with Immuno Fluorescence (IF) for MBNL1 protein using anti-MBNL1 antibody did not find colocalization of MBNL1 with SCA12 RNA foci (Mander's overlap equation 0.5) (data not shown).

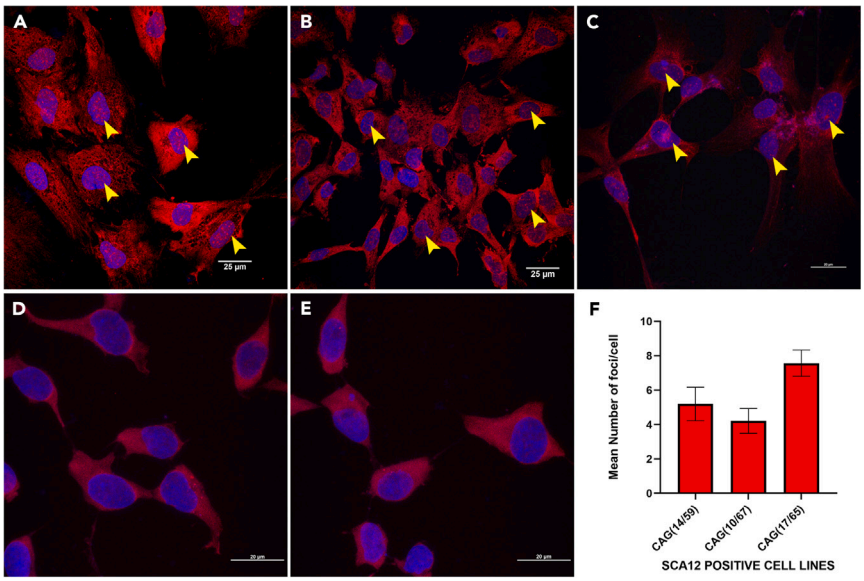


Figure 1. RNA FISH analysis of SCA12 positive neural stem cells shows RNA aggregates inside the nucleus of the cells

Where (A–C) are SCA12 positive neural stem cell lines with CAG(14/59), CAG (10/67), and CAG (17/65), respectively.

(D) and (E) are two healthy control-derived NSCs with CAG (09/16) and CAG (10/13), respectively.

(F) Bar chart illustrating the average count of nuclear foci per cell across patient-derived cell lines. CTG probe labeled with CY3 is shown in red, and DAPI is shown in blue. Yellow arrow indicates the localization of RNA FOCI inside the nucleus. All of the images exhibited here have been converted as Maximum Instensity Projection (MIP) images for enhanced visualization using Nikon's NIS elements software. Scale bar = A-C (25 µM [micrometer]), D-E (20 µM), the error bars in the graph depicts standard error.

Expanded SCA12 CAG repeats induce RNA foci to bind key nuclear proteins

To determine if the SCA12 nuclear RNA foci in NSCs interact with or bind to any specific proteins within the cellular milieu, we performed RNA pull-down assay. CAG repeats with their flanking regions upstream and downstream of the *PPP2R2B* gene from the patient (CAG-54) and control DNA (CAG-10) was cloned into pcDNA3.1 vector and amplified using primers against T7 promoter and PPP2R2B region. Next, this DNA template was used to make RNA through the Hi-scribe transcription kit using biotinylated CTP (Thermo Fisher) instead of dCTP. Both the RNAs (CAG-54 and CAG-10) were incubated with nuclear proteins extracted from SCA12 positive NSCs and control NSCs, and a pull-down assay was performed using streptavidin beads to check for interaction through mass spectrometry. The high-score peptides obtained from the mass-spectrometry-based detection of interacting proteins revealed that 378 proteins from control NSC nuclear protein extract were bound in high concentration with the CAG-10 construct (CT_10 = 378), whereas 623 proteins from the SCA12 positive NSC protein extract were retained by the CAG-10 construct (SCA12_10 = 623). Of the total interacting proteins observed, 303 proteins were common in both the groups that were found to bind to the CAG-10 repeat (CT_10 ∩ SCA12_10 = 303). Similarly, we looked for the identity and number of proteins retained by the pathogenic CAG repeat construct (CAG-54). Three hundred sixty-two proteins were found to be associated with CAG-54 repeats from the nuclear extract of control NSC in high concentration (CT_54 = 362); 179 proteins from SCA12 positive NSC nuclear protein extract were found to bind with CAG-54 repeats (SCA12_54 = 179), of which 115 proteins were common in both the groups (CT_54 ∩ SCA12_54 = 115). Further analysis revealed that 13 proteins exclusively bind to CAG-54 repeats (KRT27, HSPA1B, H1-3, DYNLL2, C1QBP, S100A16, CTSA, KIAA0100, PSMA2, FN1, SNRPB, GSN, and PSMA1). We endeavored to validate selected proteins through western blot analysis and successfully confirmed the binding of GSN1 protein to the expanded CAG RNA, featuring the flanking sequence of *PPP2R2B*. Notably, this binding was absent in transcripts containing only 10 repeats of CAG with the same flanking *PPP2R2B* sequence (Figure S1).

RAN translation in *PPP2R2B* ORF

To check if the expanded *PPP2R2B* leads to RAN translation in transfected cells, multiple constructs with different CAG/AGC repeats sizes of *PPP2R2B* mini genes (60, 50, 43, and 10) were generated with three epitope tags in each reading frame (Figure 2A) and characterized using the Sanger sequencing method (data not shown). As reported previously, CAG repeat length decreases when amplified inside a plasmid using bacterial culture.[11] We also observed much heterogeneity in the sequences of picked colonies after transformation, hence we proceeded with CAG repeat lengths of 54, 51, 39, and 10 repeats. According to the NCBI and UniProt databases, the CAG repeat (Glutamine) falls under the reading frame of AGC (Serine), hence we have referred to the CAG repeat sequence as the AGC repeat sequence in Figure 2B. Transient overexpression of RAN constructs in the HEK293T cell line revealed the occurrence of RAN translation within the CAG/AGC extended reading frame, which emits positive signals unique to polyGln (poly Glutamine) and polySer (poly Serine) proteins. Surprisingly, polyGln and polySer proteins were expressed in constructs lacking the ATG start codon in the enlarged CAG/AGC repeats. Also, polyglutamine was found to be

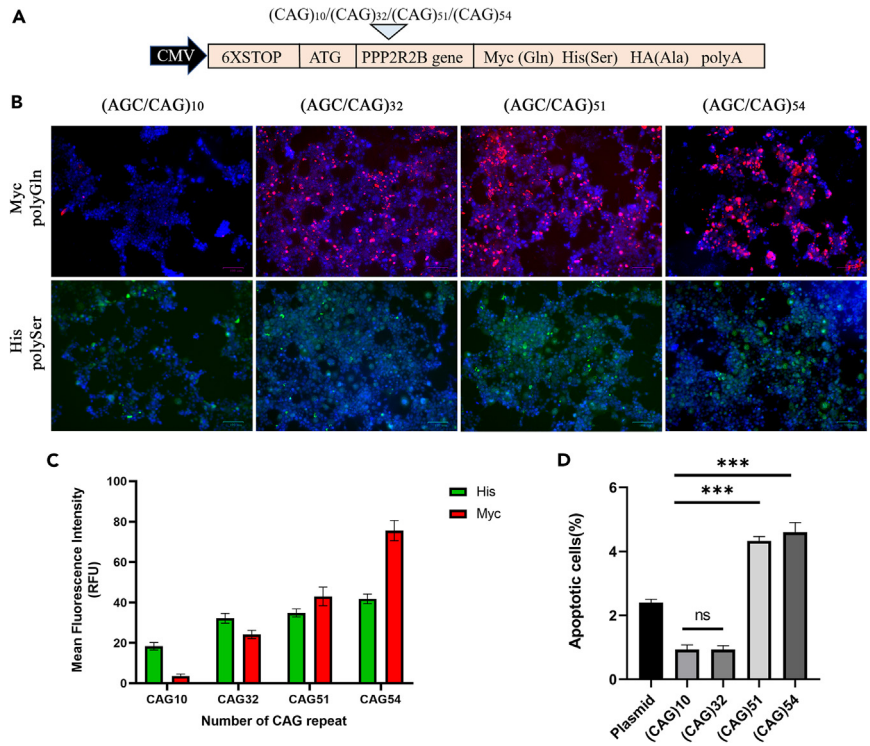


Figure 2. Repeat-dependent RAN protein expression in HEK293T cell line

(A) RAN construct having 6x stop codon in 5 prime region, variable CAG repeat length followed by three different epitopes respective for each reading frame.

(B) Immunofluorescence microscopy results of HEK293T cells after transfection with indicated constructs, where DAPI is shown in blue, Myc/polyGln (polyglutamine) signal is shown in red, and His/polySer (polyserine) is shown in green with magnification bar of 100 μM (micrometer).

(C) Bar graph depicting the average fluorescence intensity of Myc and His tags following transfection with RAN constructs featuring varying repeat lengths of CAG in the HEK293T cell line.

(D) The bar graph illustrates the percentage of apoptotic cells (Annexin V-positive cells) measured through flow cytometry. Each bar represents the average of three independent experimental values. A one-way ANOVA test was conducted, with *** indicating a *p* value <0.001 and "ns" denoting non-significance.

expressed when the only start codon ATG was mutated to GGG, this suggests that RAN translation of the polyglutamine frame is favored in the *PPP2R2B* gene inside the HEK293T cell line (Figure S2). CAG/AGC with unexpanded repeats (10), on the other hand, did not create any RAN proteins. We did not see any signal with the HA tag antibodies because the *PPP2R2B* gene sequence is predicted to induce a stop codon in the Alanine frame.

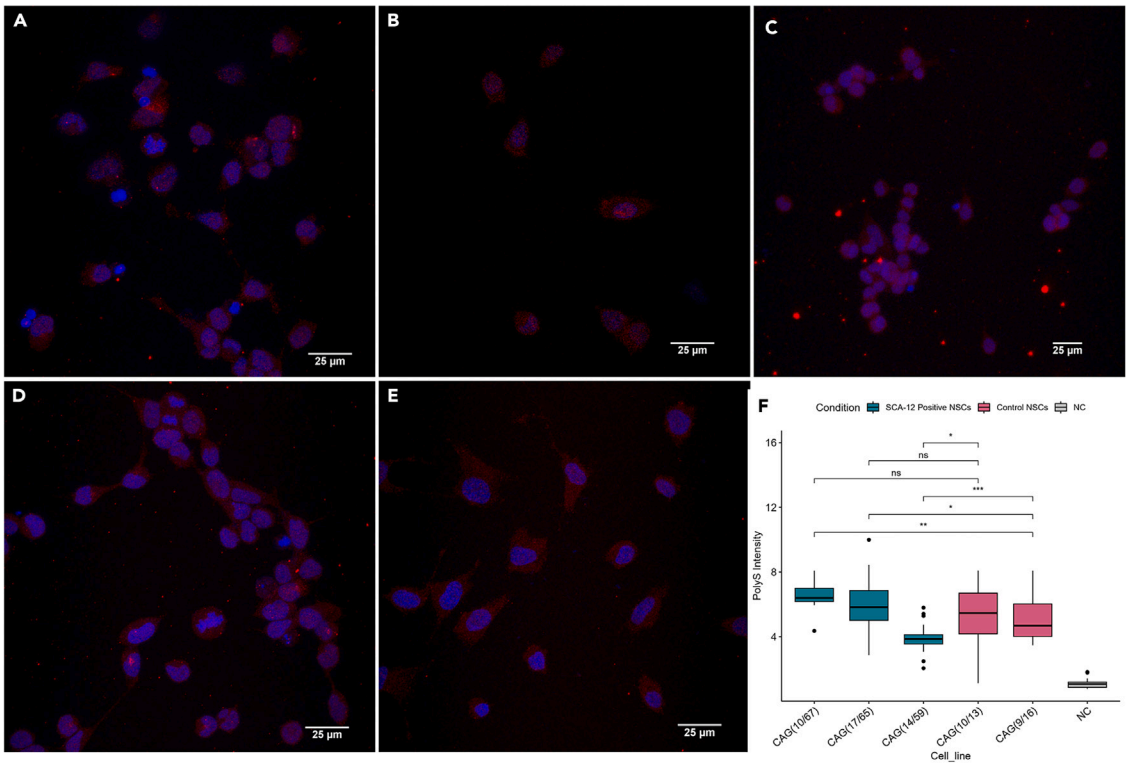
Expression of expanded CAG repeats induces cell death

Annexin-V-mediated apoptosis assay was used to determine if RAN proteins induce apoptosis in cells showing RAN translation. We transfected the HEK293T cell line with multiple RAN constructs of different CAG lengths (54, 51, 32, 10), and after 36 h, cells were treated with staurosporine for 6 h. Cells were then trypsinized, and 5 μL of Annexin V along with Propidium Iodide (PI) with 20 μL of binding buffer was added; then the percentage of apoptotic cells was detected using BD Accuri C6 plus flow cytometer (Figure 2C), and the percentage of apoptotic cells was found to be significantly higher in expanded CAG construct (54, 51) in comparison to the normal repeat length of CAG (10). This indicates that a mutated *PPP2R2B* mini gene with a CAG length of >51 upon expression through RAN translation potentially induces cellular death in HEK293T cell lines.

RAN proteins accumulate in neural stem cells with expanded CAG repeats

To confirm the occurrence of RAN translation and its associated proteins in SCA12 disease, we designed polyclonal antibodies against the anticipated C-terminal portions of the *PPP2R2B* gene that were specific to polySer (AGC) and polyGln (CAG) proteins (Figure S3) that were commercially obtained to evaluate the presence of RAN proteins in our disease model i.e., SCA12 positive NSCs. These antibodies were used to perform immunocytochemistry (ICC) and confocal microscopy on SCA12 NSCs and control NSCs (Figure 3). Confocal microscopy analysis showed significant difference in both PolySer and PolyGln expression (Figures 3A–3E and 3G–3K). PolySer intensity profile showed little higher expression of PolySer in two SCA12 NSCs (CAG 10/67 and 17/65) with less significance, whereas significantly lower expression in one SCA12 NSC (14/59), which suggests repeat number variation may also play role in RAN Proteins expressions (Figure 3F). However, PolyGln intensity

Expression of Poly Serine in SCA12 patient and control derived NSCs



Expression of Poly Glutamine in SCA12 patient and control derived NSCs

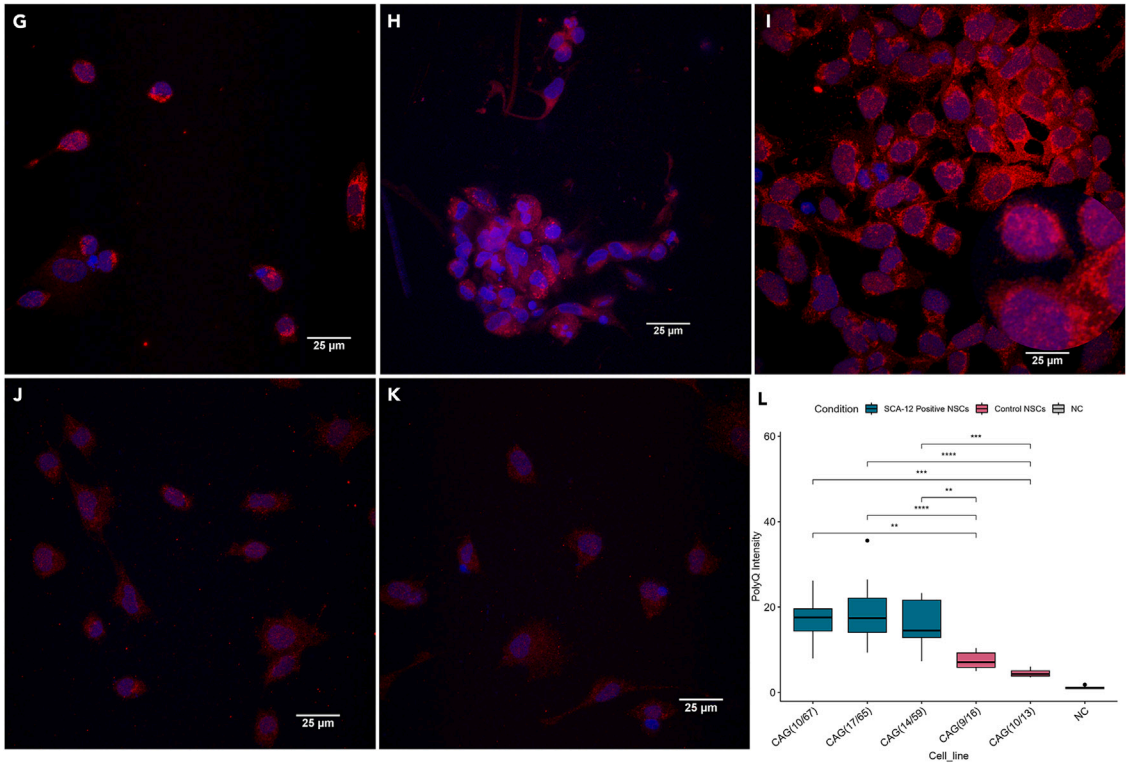


Figure 3. Expression of RAN proteins in SCA12 vs. control neuronal cell lines

Where (A–E) are cell lines showing positive staining for red (PolySerine) and blue (DAPI). (A–C) are SCA12 positive neural stem cell lines with CAG (14/59), CAG (10/67), and CAG (17/65), respectively. (D) and (E) are two healthy control-derived NSCs with CAG (09/16) and CAG (10/13), respectively.

(F) Boxplot comparing PolyS (Poly Serine) intensity between case and control cell lines, analyzed using a Student's t test. (* indicates a significant p value <0.05)

(G–E) are cell lines showing positive staining for red (polyglutamine) and blue (DAPI). (G–I) are SCA12 positive neural stem cell lines with CAG(14/59), CAG (10/67), and CAG (17/65), respectively.

(J) and (K) are two healthy control-derived NSCs with CAG (09/16) and CAG (10/13), respectively. All of the images exhibited here have been converted as MIP images for enhanced visualization using Nikon's NIS elements software by Nikon with magnification bar of 25 μM (micrometer).

(L) Boxplot comparing PolyQ (polyglutamine) intensity between case and control cell lines, analyzed using a Student's t test. (* indicates a significant p-value <0.05).

profile showed a strikingly significant difference in the intensity of signals for poly-Glu RAN-specific antibodies in SCA12 NSCs, which was comparatively trivial in control NSCs (Figure 3L). Our data provide evidence for RAN-mediated translation of the *PPP2R2B* gene in a repeat length-dependent non-canonical manner, resulting in the synthesis of enlarged homopolymer proteins in multiple frames, which might be responsible for cellular toxicity as observed in other repeat-containing diseases such as fragile-X-associated tremor/ataxia syndrome, Huntington disease, etc.

Widespread transcriptional alterations are present in SCA12-derived pan neurons and neuronal progenitors

To study the changes in global transcriptional signatures induced by the pathogenic CAG repeats, mRNA sequencing was performed. The sample set consisted of four, four, and two biological replicates for iPSC, NSC, and neuronal lineage, respectively, along with their technical replicates, details appended in Table 1. RNA isolated from all the samples was subjected to PolyA-enrichment-based mRNA sequencing. An average of 38 million sequencing reads were generated for all samples taken together, and approximately 83% of the reads were mapped to the human genome (Table 1). After processing the sequencing data to reduce noise, remove outliers, and normalize, batch effect correction was performed using limma to remove the variation introduced due to sample collection and sequencing in two batches. Principal-Component Analysis (PCA) showed a separation of patient and control samples along the PC2 (y axis) based on the expression of the 500 most variable genes (Figure 4). We next evaluated the RNA-seq data in pairwise comparisons for each cell lineage using DESeq2.[12] A total number of 20,525; 20,283; and 22,150 genes were detected in the iPSC, NSC, and neurons, respectively. Between developmental stages, the number of differentially expressed genes between patients and controls was the highest for iPSCs and the least for neurons (Figure 5). Patient neurons and iPSCs showed more genes to be significantly upregulated, whereas patient NSCs showed greater downregulation of genes in comparison to controls (Figure 5). Unsupervised hierarchical clustering of the samples demonstrated a clear separation between patient- and control-derived cell types (Figure 4B). To further query the nature of global transcriptome alterations induced by the pathogenic SCA12 CAG repeats in the *PPP2R2B* gene, we performed pathway analysis using Qiagen's Ingenuity Pathway Analysis (IPA) software. A comprehensive list of the top 30 significant pathways for neurons, NSC, and iPSC datasets are listed in Table S2. In the neuronal dataset, IPA identified signaling pathways as the most significant canonical pathways. Neurovascular coupling signaling emerged as the top most significant pathway predicted to be upregulated in SCA12 neurons, which has previously been linked to Alzheimer disease.[13,14] The other top significant pathways in SCA12 neurons include the GABA receptor signaling pathway, WNT/β-catenin signaling, and CREB signaling in neurons. IPA results also highlight the involvement of pathways of immune response i.e., neuroinflammation signaling pathway and the antigen presentation pathway to be altered in the SCA12 neurons.

Different cell lineages exhibit differential PPP2R2B transcript isoform profiles under the SCA12 disease context

The *PPP2R2B* gene is known to exhibit alternative splicing, and various repeat-containing and non-repeat-containing variants have been reported. *PPP2R2B* isoforms differ in their 5′ sequence and therefore code for proteins with variable N-terminal regions, thus affecting their subcellular localization and substrate specificity of the PP2A enzyme.[15,16] We utilized the RNA-seq data to study the variation in transcript isoform profile of *PPP2R2B* gene across different cell lineages induced by the presence of pathogenic CAG repeats. Mean-of-inferential replicates were used to quantify the expression difference in fold change between SCA12 and control cell lines (Figure 5). The transcripts that were statistically different between patients and controls are NM_181674.2 (Variant 2), NM_181677.2 (Variant 5), and NR_073527.1 (Variant 12). NM_181674.2 (Variant 2) and NM_181677.2 (Variant 5) are both coding transcripts, and their expression was reduced in the SCA12-derived pan neurons. On the other hand, NR_073527.1 (Variant 12) is a non-coding transcript of the *PPP2R2B* gene but its expression was found to be increased in SCA12 neurons. Overall, in the spinocerebellar ataxia type-12 neurons, most isoforms of the *PPP2R2B* gene showed a downregulation, except NM_001271900.2 (Variant 8) and NR_073527.1 (Variant 12) in comparison to controls. In contrast to the SCA12 neurons, transcript profiles in SCA12 NSCs showed a striking difference in terms of more than 2.5-fold upregulation of two isoforms viz. NM_181674.2 (Variant 2) and NM_181676.2 (Variant 4). Interestingly, both these are coding transcript isoforms and show a statistically significant upregulation in the patient NSCs when compared with controls. We were able to validate NM_181676.2 (Variant 4) transcript with the help of RT-qPCR (Figure S4). The other transcript isoforms identified in the NSCs did not show any significant change between patients and controls. We also looked at the isoform profile at the iPSC stage and found two isoforms NM_001271948.1 (Variant 10) and NR_073527.1 (Variant 12) to be strongly upregulated (L2Fc > 4) in the SCA-12 iPSCs. NR_073527.1 (Variant 12) is a non-coding isoform of *PPP2R2B* but NM_001271948.1 (Variant 10) is a coding mRNA of 2578 bp. Despite limitations on the number of samples, the transcript isoform level data from our study gives interesting insights

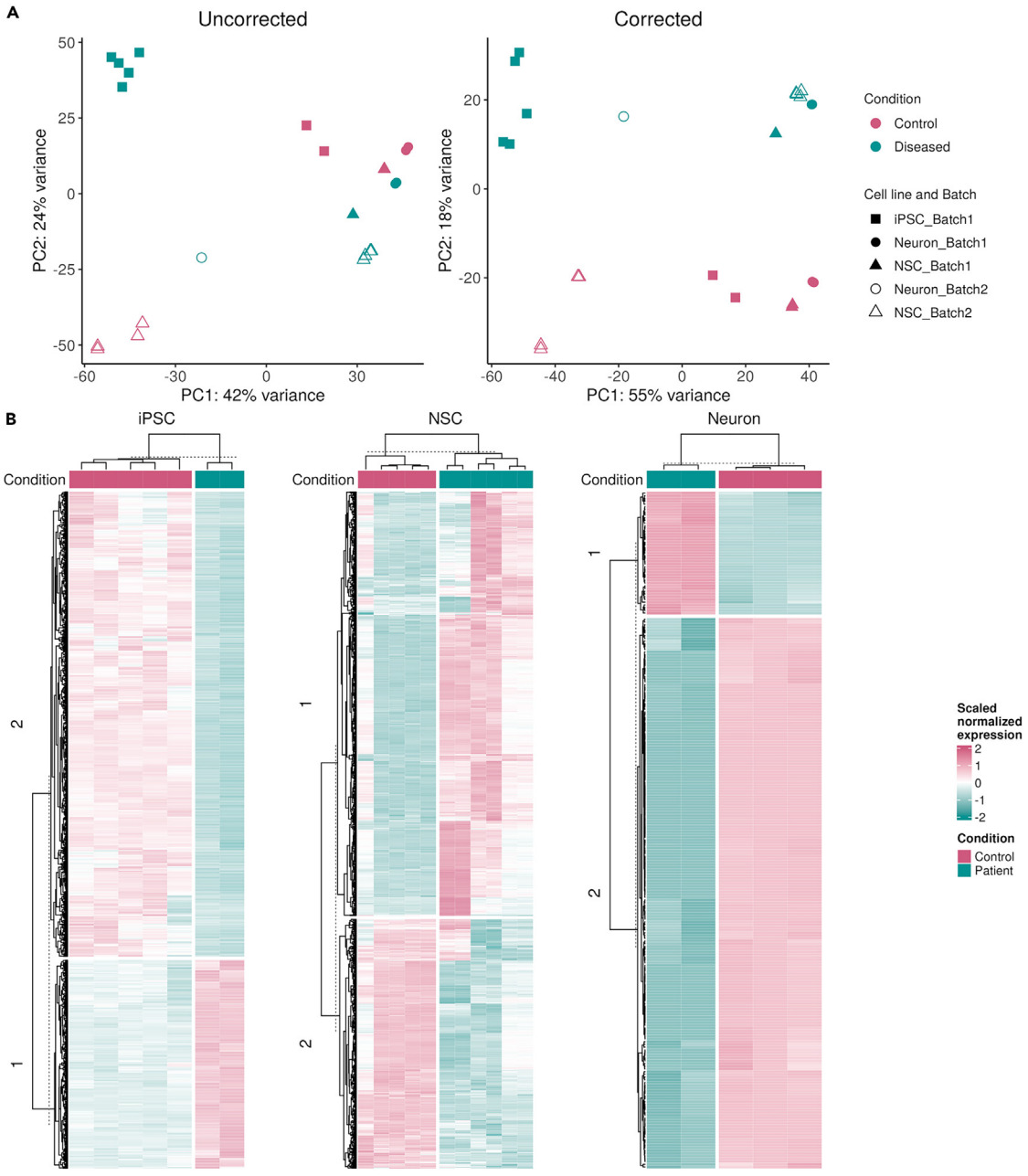


Figure 4. Principal-component analysis and heatmap of top differentially expressed genes

(A) PCA of before and after correction of the batch effect.

(B) Heatmap of genes filtered based on |log2FoldChange| > 2 and padj <0.01. The values in the heatmap are scaled variant stabilization transformation data.

into consistent downregulation of *PPP2R2B* gene isoforms in the mature SCA12 neurons, whereas many transcript isoforms show abundant expression at the progenitor stages.

Transcriptional dysregulation in SCA12 involves a network of transcription factor genes responsible for nervous system development and function

The list of DEGs was used to infer biologically and disease-relevant networks among represented genes based on their known interactions using the IPA. One of the top networks identified by the IPA is defined by the terms nervous system development and function, organ morphology, and tissue morphology and has a top score of 42 (Figure 6). This network has a total of 35 molecules with 26 of them picked up from our DEG gene list. Interestingly, *PPP2R2B* emerges as one of the hub genes in the network where most other genes are involved

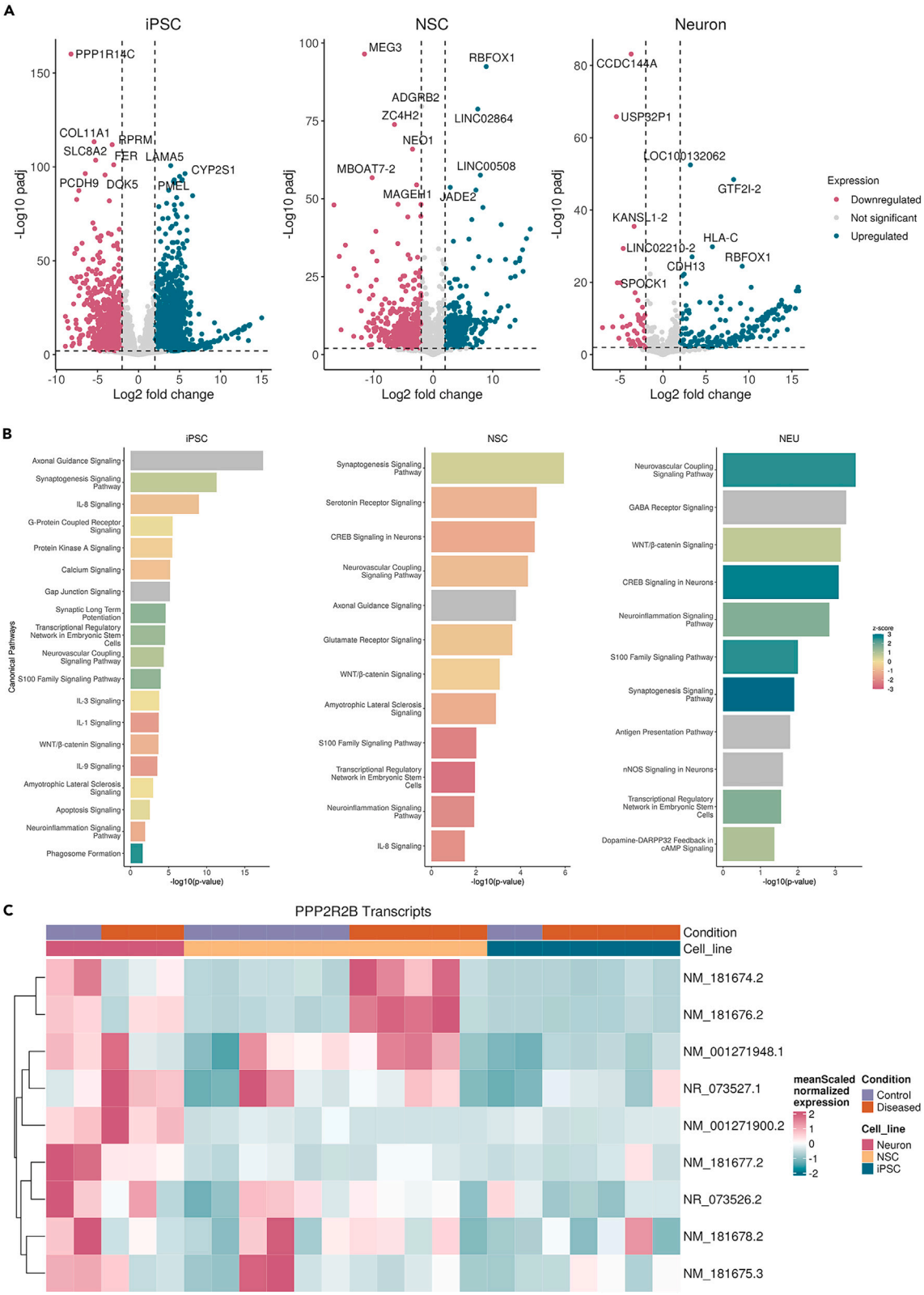

A
iPSC
NSC
Neuron
-Log10 padj
Log2 fold change
PPP1R14C
COL11A1
SLC8A2
RPRM
FER
LAMA5
CYP2S1
PCDH9
DOK5
PMEL
MEG3
RBFOX1
ADGRB2
ZC4H2
LINC02864
NEO1
MBOAT7-2
LINC00508
MAGEH1
JADE2
CCDC144A
USP32P1
LOC100132062
GTF2I-2
KANSL1-2
HLA-C
LINC02210-2
RBFOX1
CDH13
SPOCK1
Expression
Downregulated
Not significant
Upregulated
B
iPSC
NSC
NEU
Canonical Pathways
Axonal Guidance Signaling
Synaptogenesis Signaling Pathway
IL-8 Signaling
G-Protein Coupled Receptor Signaling
Protein Kinase A Signaling
Calcium Signaling
Gap Junction Signaling
Synaptic Long Term Potentiation
Transcriptional Regulatory Network in Embryonic Stem Cells
Neurovascular Coupling Signaling Pathway
S100 Family Signaling Pathway
IL-3 Signaling
IL-1 Signaling
WNT/β-catenin Signaling
IL-9 Signaling
Amyotrophic Lateral Sclerosis Signaling
Apoptosis Signaling
Neuroinflammation Signaling Pathway
Phagosome Formation
Serotonin Receptor Signaling
CREB Signaling in Neurons
Glutamate Receptor Signaling
GABA Receptor Signaling
Antigen Presentation Pathway
nNOS Signaling in Neurons
Dopamine-DARPP32 Feedback in cAMP Signaling
-log10(p-value)
z-score
C
PPP2R2B Transcripts
Condition
Cell_line
NM_181674.2
NM_181676.2
NM_001271948.1
NR_073527.1
NM_001271900.2
NM_181677.2
NR_073526.2
NM_181678.2
NM_181675.3
meanScaled normalized expression
Condition
Control
Diseased
Cell_line
Neuron
NSC
iPSC

Figure 5. Differential expression and enrichment analysis of iPSCs, NSCs, and neurons in SCA12 disorder

(A) Volcano plots of the differentially expressed genes within the cutoff of | log2Fold change| > 2 and padj< 0.01 in iPSC, NSC, and neuron. Labeled the top 10 genes based on padj values.

(B) Barplot of enrichment of canonical pathways identified from IPA. x axis represents -log10(pvalue) and colors scaled with *Z* score, representing enrichment (*Z* score >0) or depletion (*Z* score <0) of the pathway. Gray color represents not enough genes present to calculate *Z* score or the direction of enrichment.

(C) Heatmap of the PPP2R2B transcript expression in iPSC, NSC, and neuron. Values in the heatmap are scaled Mean counts of the inferential replicates.

in functions related to the development, differentiation, and morphology of the nervous system and neurons. The network contains eight genes for transcription factors viz. *BMAL1*, *GBX2*, *HIF3A*, *MYT1*, *NEUROD4*, *NEUROG2*, *PAX6*, and *SOX2*. NEUROG2 is a transcription factor that induces the differentiation of progenitor cells into terminal neuronal cells.[17] This lineage commitment is driven by sequential transcriptional activation of *PAX6* followed by *NEUROG2*. NEUROG2 forms a heterodimer with NEUROD4, which regulates neurogenesis in the cortex by accelerating specific transcriptional patterns.[18] Both NEUROG2 and NEUROD4 are predominantly expressed in progenitor cells. *MYT1* is a zinc family pan-neuron-specific transcription factor that safeguards neuronal identity by repressing non-neuronal cell fates.[19] Although SOX2 is essentially recognized for its role in maintaining stemness or pluripotency of somatic cells,[20] recent studies have revealed its fundamental role in differentiated neurons and glia, such as the development of proper neuronal connectivity,[21] axon growth, and motor neuron maturation.[22] Despite their primary role in neural progenitor cells, the overexpression of most of these transcription factors in mature SCA12 neurons raises the question of their role in differentiated neuronal cells.

DISCUSSION

In this study, we were able to generate disease models of the SCA12 disorder using human iPSC-derived cell lines in the form of repeat-containing NSCs and neurons. Our extensive analysis using these disease models has shed new insights into the pathological mechanism of SCA12 at the molecular level. The occurrence of RNA foci and RAN translation and its associated proteins are two phenomena known to be integral in the pathology of many repeat-containing diseases. Their presence in SCA12 models provides an inkling of the underlying mechanisms that result in the similitude of clinical symptoms in SCA12, FXTAS, and HD.

Deciphering the impact of RNA-pull-down-identified proteins on neurodegenerative pathways

From pull-down assay we identified 13 proteins, which were found specifically binding to expanded CAG (54) transcript, which was KRT27, HSPA1B, H1-3, DYNLL2, C1QBP, S100A16, CTSA, KIAA0100, PSMA2, FN1, SNRPB, GSN, and PSMA1. Further investigation of these proteins revealed that their accumulation had previously been associated with several neurodegenerative disorders such as Parkinson disease, Alzheimer disease, and Amyotrophic Lateral Sclerosis (ALS). Hsp70 proteins, particularly HSPA1B, are important in protein folding and preventing misfolded proteins from aggregating. These proteins aid in the refolding of damaged proteins and direct their degradation by the proteasome. It has also been proposed that Hsp70 associates with protein aggregates as an endogenous effort to alleviate the toxic effects of the aggregates in Alzheimer and Parkinson brains.[23] Similarly, C1QBP expression was found to be considerably lower in Alzheimer in microarray analysis, despite being elevated in normal aging. C1QBP has been demonstrated to interact with Aβ and stimulate fibrillogenesis, the process by which Aβ aggregates into fibrils. This interaction indicates that C1QBP may be involved in the production and buildup of amyloid plaques in Alzheimer disease.[24] Furthermore, Proline-Arginine Dipeptide Repeat (PR-DPR), the most toxic protein formed by RAN translation in ALS patients with C9ORF72 gene expansion mutation, is known to interact with C1QBP molecule to activate intracellular NOD-, LRR-, and Pyrin Domain-Containing 3 (NLRP3) inflammasome activation of microglia cells.[25] Dynein Light Chain LC3 type 2 DYNLL2 is a motor protein involved in vesicle and organelle axonal-retrograde transport. Interruption of axonal transport is a common hallmark of neurodegenerative diseases. Second, autophagy and the clearance of aggregation-prone proteins, both of which rely on dynein proteins, are reported to be defective in Alzheimer, Parkinson, and motor neuron disorders.[26] The FN1 Fibronectin 1 (FN1) molecule is a glycoprotein found in the extracellular matrix that is involved in cell adhesion, tissue organization, and migration. It participates in clotting, which may influence amyloid-beta fibrillization.[27] Furthermore, FN1 expression was shown to be higher in the plasma of Alzheimer patients with dementia compared with Alzheimer patients without dementia symptoms.[27]

Small Nuclear Ribonucleoprotein Polypeptides B (SNRPB), on the other hand, is involved in the assembly and functioning of the spliceosome complex. Unavailability or dysfunction of this protein may result in abnormal splicing processes, leading to hyperexcitability and cognitive impairment in Alzheimer disease.[28] Gelsolin GSN is an actin-binding protein that plays a role in cytoskeleton dynamics, cell motility, and cellular responses to extracellular signals. GSN is known to inhibit amyloid aggregation and cellular apoptosis in Alzheimer disease. Also, GSN is observed to have reduced expression in advanced disease states of Alzheimer, leading to exacerbation of disease.[29] Similarly, S100A6 (calcyclin or calmodulin-like protein 3) is a member of the S100 family of calcium-binding proteins whose expression level has been observed to be altered in Alzheimer disease, Parkinson disease, ALS, and HD. Recently, its role has been discovered in the differentiation of astrocytes.[30] Furthermore, it has recently been discovered that RNA foci emerge from the nucleus to cytoplasm over time and that RNA and RAN protein interaction and aggregation are associated with RNA-binding protein dispositioning and cellular toxicity.[31] Also, our protein pull-down data showed the entrapment of proteins that are involved in the clearance mechanism of aggregated proteins (HSPA1B, PSMA1, PSMA2). This shows that RNA and RAN proteins interact to cause and aggravate disease over time. Overall, these are critical proteins with important functions to perform in the cell, and their absence or altered expression can cause serious disruptions in cellular processes, as shown in other neurodegenerative disorders.

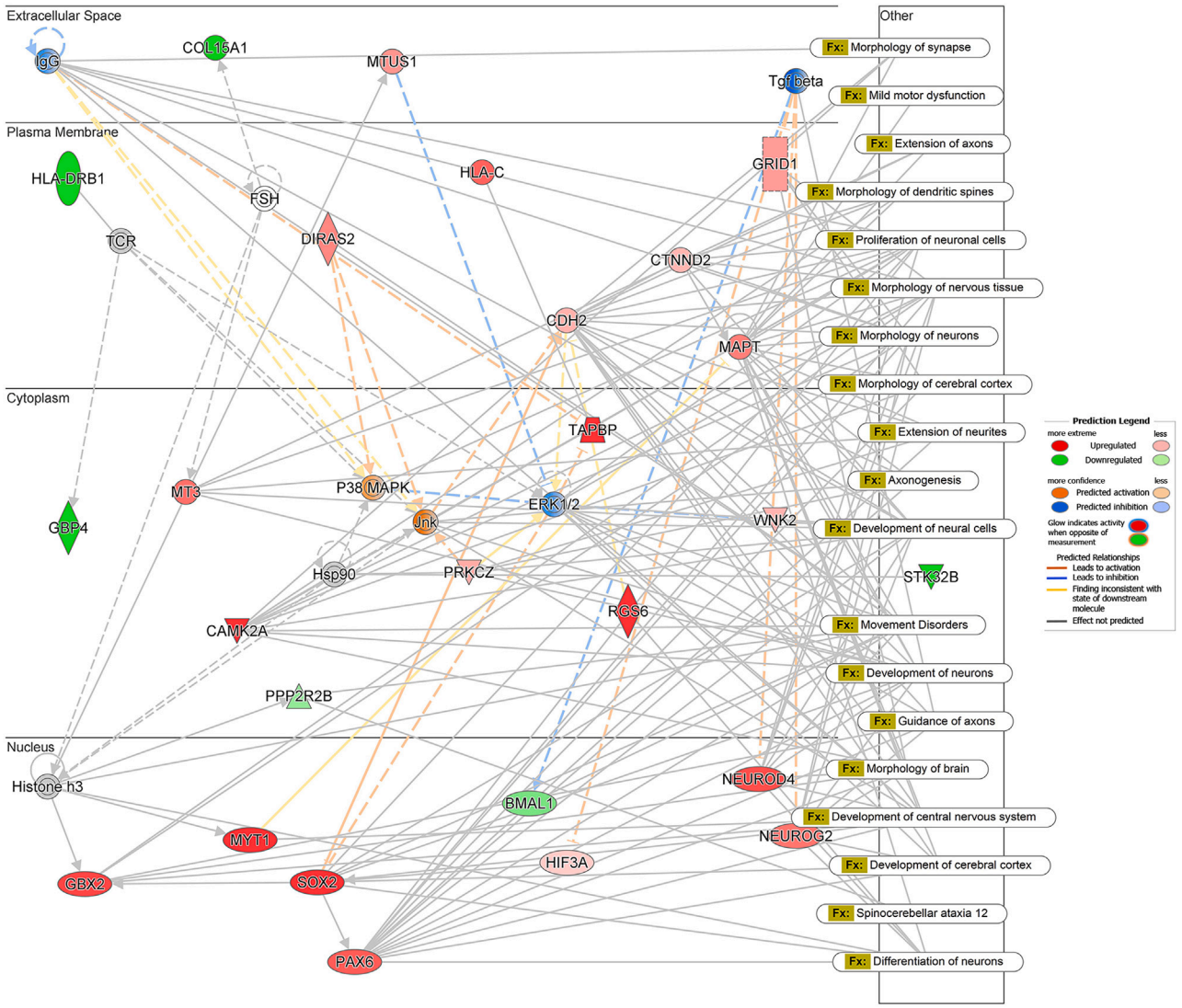


Figure 6. Functionally significant network of relevant genes in spinocerebellar ataxia-12 neurons

IPA was used to identify top functionally and disease relevant networks with link to SCA-12 pathology using gene expression profile of SCA-12 neurons.

Insights from IPA and mRNA seq analysis

IPA of the NSC dataset revealed several pathways involved in the transition of NSC to the neuronal lineage. CREB signaling in neurons, neurovascular coupling signaling pathway, and WNT/β-catenin signaling pathways that were identified in the neuronal dataset also appeared in the NSC top pathways by significance. Other top significant pathways identified in the NSC dataset include synaptogenesis signaling, endocannabinoid neuronal synapse, synaptic long-term depression, serotonin receptor signaling, axonal guidance signaling, and gap junction signaling. This is indicative of the role of pathogenic repeats in affecting the development of neuronal junctions and synapses. NSC pathways also exhibit signatures indicative of other neurodegenerative diseases such as amyotrophic lateral sclerosis signaling and multiple sclerosis signaling pathways. NSCs showed a predicted inhibition of pathways involved in inflammatory responses such as HMGB1 signaling, interleukin-17 (IL-17) signaling, pathogen-induced cytokine storm signaling pathway, and wound healing signaling pathway. Necessitated by the presence of pathogenic repeats and their associated pathology, i.e., RNA foci and RAN-translated proteins, inhibition of the inflammatory immune responses seems to be crucial to allow the development of NSCs into neurons.

Interestingly, *BMAL1* is considered an irreplaceable clock gene that regulates circadian rhythm and rhythmic behaviors in mammals.[32,33] However, recent studies in preclinical models indicate the role of disruption of the *BMAL1* gene to be associated with cerebellar damage and ataxia.[34] The role of the *BMAL1* gene is being currently investigated in neurodegenerative pathologies such as Parkinson disease and Alzheimer disease.[35] Furthermore, defects in the *BMAL1* gene are associated with the regulation of inflammatory and immune responses and the regulation of neuroinflammation. Disruption of the *BMAL1*-regulated dopamine signaling and its consequent impact on

neuroinflammation in the brain has been proposed as a causative factor for the onset of Parkinson disease.[36] A possible role of the *BMAL1* gene in neuronal function could be the synchronization of the rhythmic control of transcription of other genes controlling neuronal activity and function.[37] The downregulation of the *BMAL1* gene in SCA12 neurons as observed in our data provides evidence for the above hypothesis and warrants further investigation. Future studies on some of these candidate genes might be able to provide better insights into the pathogenesis of neurodegeneration in general and SCA12-associated pathology in particular.

Conclusion

We were able to generate disease models of the SCA12 disorder using human iPSC-derived cell lines in the form of repeat-containing NSCs and neurons. Our extensive analysis using these disease models has shed new insights into the pathological mechanism of SCA12 at the molecular level. The occurrence of RNA foci and RAN translation and its associated proteins are two phenomena known to be integral to the pathology of many repeat-containing diseases. Their presence in SCA12 models provides an inkling of the underlying mechanisms that result in the similitude of clinical symptoms in SCA12, fragile-X-associated tremor/ataxia syndrome and Huntington disease. Recently, Das et al. (2023), showed that RAN translation occurs after extended retention of nuclear foci, and after localization of RNA foci in cytoplasm it interacts with RAN products to form aggregation-like molecules.[31] Our protein pull-down data showed the entrapment of proteins that are involved in the clearance mechanism of aggregated proteins (HSPA1B, PSMA1, PSMA2). Going forth, evaluation of the transcriptome in SCA12 neurons and NSCs highlights the nature of global transcriptional alterations brought about by the dysregulation of important transcription factors. Also, pathogenic phenomena in SCA12, including the occurrence of RNA foci aggregates, RAN translation products, and its associated gene expression changes, possibly induce inflammatory immune cascades in these cells, providing a link to the neuroinflammation associated with neurodegeneration, as evident from the transcriptomic data. In concordance with this, C1QBP, which activates inflammasomes in the microglia, has also been found to bind specifically to the expanded CAG transcript. Further, our data provide new evidence for the linkage of *BMAL1* gene dysregulation with cerebellar dysfunction and ataxia. Overlapping results from protein pull-down assay and RNA seq analysis in our data suggest new molecular insights into the pathogenesis of SCA12 disorder. These results are, however, of preliminary nature and need to be explored further in future studies in sync with other CAG/CTG repeat expansion diseases.

Study limitations

Size limitation: the findings of the study may have limited applicability to a wider population due to the use of a relatively small sample size. RT-PCR validation: notably, validation of *PPP2R2B* gene transcripts via RT-PCR was omitted. This decision stemmed from the high conservation (approximately 90%) across all transcripts. Despite efforts to design transcript-specific primers, the RT-PCR results displayed considerable non-specific amplification, posing challenges to the accuracy of the validation process.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Cell culture
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Ingenuity pathway analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109768.

ACKNOWLEDGMENTS

The funding support for the study was obtained from Council of Scientific and Industrial Research (CSIR)-funded projects: GenCODE (BSC0123) and the Indian Council of Medical Research (ICMR)-funded project, 5/4–5/5/Ad-hoc/Neuro/220/NCD-I (GAP240).

We are sincerely thankful to all the patients and their families for their participation and cooperation in this study. We are also thankful to Dr. Himanshi Kapoor for helping with Confocal microscopy and NCBS for providing us the one control iPSC D161 (ADBSi001-A) for our experiments. We acknowledge the efforts of Mr. Bharathram Uppili for helping with figure compilations and Shreya Bari for helping with data analysis.

AUTHOR CONTRIBUTIONS

Study conceptualization and design: M.F.; all experimental design and execution: M.K.; iPSC cell line generation: M.K. and D.K.; RNAseq analysis: S.S., V.A., and M.F.; RNA fish analysis: M.K. and N.K.; cellular assays: D.G.; imaging: H.K.; manuscript writing: M.K., S.S., A.K.S., and M.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 27, 2023
Revised: January 31, 2024
Accepted: April 15, 2024
Published: April 18, 2024

REFERENCES

1. O'Hearn, E.E., Hwang, H.S., Holmes, S.E., Rudnicki, D.D., Chung, D.W., Seixas, A.I., Cohen, R.L., Ross, C.A., Trojanowski, J.Q., Pletnikova, O., et al. (2015). Neuropathology and Cellular Pathogenesis of Spinocerebellar Ataxia Type 12. Mov. Disord. *30*, 1813–1824. https://doi.org/10.1002/mds.26348.
2. Srivastava, A.K., Takkar, A., Garg, A., and Faruq, M. (2017). Clinical behaviour of spinocerebellar ataxia type 12 and intermediate length abnormal CAG repeats in PPP2R2B. Brain *140*, 27–36. https://doi.org/10.1093/brain/aww269.
3. Srivastava, A., Kumar, D., Faruq, M., and Gundluru, V. (2019). Spinocerebellar ataxia type 12: An update. Ann. Mov. Disord. *2*, 48–57. https://doi.org/10.4103/AOMD.AOMD_5_19.
4. Sharma, P., Sonakar, A.K., Tyagi, N., Suroliya, V., Kumar, M., Kutum, R., Asokchandran, V., Ambawat, S., Shamim, U., Anand, A., et al. (2022). Genetics of Ataxias in Indian Population: A Collative Insight from a Common Genetic Screening Tool. Adv. Genet. *3*, 2100078. https://doi.org/10.1002/ggn2.202100078.
5. Swarup, V., Srivastava, A.K., and Rajeswari, M.R. (2012). Identification and quantification of differentially expressed proteins in plasma of spinocerebellar ataxia type 12. Neurosci. Res. *73*, 161–167. https://doi.org/10.1016/j.neures.2012.03.002.
6. Lin, C.H., Chen, C.M., Hou, Y.T., Wu, Y.R., Hsieh-Li, H.M., Su, M.T., and Lee-Chen, G.J. (2010). The CAG repeat in SCA12 functions as a cis element to up-regulate PPP2R2B expression. Hum. Genet. *128*, 205–212. https://doi.org/10.1007/s00439-010-0843-2.
7. Faruq, M., Srivastava, A.K., Suroliya, V., Kumar, D., Garg, A., Shukla, G., and Behari, M. (2014). Identification of FXTAS presenting with SCA 12 like phenotype in India. Parkinsonism Relat. Disord. *20*, 1089–1093. https://doi.org/10.1016/j.parkreldis.2014.07.001.
8. 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., et al. (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. USA *108*, 260–265. https://doi.org/10.1073/pnas.1013343108.
9. Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V., et al. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440–455. https://doi.org/10.1016/j.neuron.2013.03.026.
10. Zhang, N., and Ashizawa, T. (2017). RNA toxicity and foci formation in microsatellite expansion diseases. Curr. Opin. Genet. Dev. *44*, 17–29. https://doi.org/10.1016/j.gde.2017.01.005.
11. Kim, S.H., Pytlos, M.J., Rosche, W.A., and Sinden, R.R. (2006). (CAG)*(CTG) repeats associated with neurodegenerative diseases are stable in the Escherichia coli chromosome. J. Biol. Chem. *281*, 27950–27955. https://doi.org/10.1074/jbc.M601129200.
12. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
13. Stackhouse, T.L., and Mishra, A. (2021). Neurovascular Coupling in Development and Disease: Focus on Astrocytes. Front. Cell Dev. Biol. *9*, 702832. https://doi.org/10.3389/fcell.2021.702832.
14. Zhu, W.M., Neuhaus, A., Beard, D.J., Sutherland, B.A., and DeLuca, G.C. (2022). Neurovascular coupling mechanisms in health and neurovascular uncoupling in Alzheimer's disease. Brain *145*, 2276–2292. https://doi.org/10.1093/brain/awac174.
15. Strack, S., Zaucha, J.A., Ebner, F.F., Colbran, R.J., and Wadzinski, B.E. (1998). Brain protein phosphatase 2A: developmental regulation and distinct cellular and subcellular localization by B subunits. J. Comp. Neurol. *392*, 515–527. https://doi.org/10.1002/(SICI)1096-9861(19980323)392:4<515::AID-CNE8>3.0.CO;2-3.
16. Dagda, R.K., Zaucha, J.A., Wadzinski, B.E., and Strack, S. (2003). A developmentally regulated, neuron-specific splice variant of the variable subunit Bbeta targets protein phosphatase 2A to mitochondria and modulates apoptosis. J. Biol. Chem. *278*, 24976–24985. https://doi.org/10.1074/jbc.M302832200.
17. Ge, W., He, F., Kim, K.J., Blanchi, B., Coskun, V., Nguyen, L., Wu, X., Zhao, J., Heng, J.I.T., Martinowich, K., et al. (2006). Coupling of cell migration with neurogenesis by proneural bHLH factors. Proc. Natl. Acad. Sci. USA *103*, 1319–1324. https://doi.org/10.1073/pnas.0510419103.
18. Mattar, P., Langevin, L.M., Markham, K., Klenin, N., Shivji, S., Zinyk, D., and Schuurmans, C. (2008). Basic helix-loop-helix transcription factors cooperate to specify a cortical projection neuron identity. Mol. Cell Biol. *28*, 1456–1469. https://doi.org/10.1128/MCB.01510-07.
19. Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake, S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. Nature *544*, 245–249. https://doi.org/10.1038/nature21722.
20. Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. *17*, 183–193. https://doi.org/10.1038/nrm.2016.8.
21. Mercurio, S., Serra, L., and Nicolis, S.K. (2019). More than just Stem Cells: Functional Roles of the Transcription Factor Sox2 in Differentiated Glia and Neurons. Int. J. Mol. Sci. *20*, 4540. https://doi.org/10.3390/ijms20184540.
22. Gong, J., Hu, S., Huang, Z., Hu, Y., Wang, X., Zhao, J., Qian, P., Wang, C., Sheng, J., Lu, X., et al. (2020). The Requirement of Sox2 for the Spinal Cord Motor Neuron Development of Zebrafish. Front. Mol. Neurosci. *13*, 34. https://doi.org/10.3389/fnmol.2020.00034.
23. Leak, R.K. (2014). Heat shock proteins in neurodegenerative disorders and aging. J. Cell Commun. Signal. *8*, 293–310. https://doi.org/10.1007/s12079-014-0243-9.
24. Cribbs, D.H., Berchtold, N.C., Perreau, V., Coleman, P.D., Rogers, J., Tenner, A.J., and Cotman, C.W. (2012). Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study. J. Neuroinflammation *9*, 179. https://doi.org/10.1186/1742-2094-9-179.
25. Fu, R.H., Tsai, C.W., Chiu, S.C., Liu, S.P., Chiang, Y.T., Kuo, Y.H., Shyu, W.C., and Lin, S.Z. (2022). C9-ALS-Associated Proline-Arginine Dipeptide Repeat Protein Induces Activation of NLRP3 Inflammasome of HMC3 Microglia Cells by Binding of Complement Component 1 Q Subcomponent-Binding Protein (C1QBP), and Syringin Prevents This Effect. Cells *11*, 3128. https://doi.org/10.3390/cells11193128.
26. Eschbach, J., and Dupuis, L. (2011). Cytoplasmic dynein in neurodegeneration. Pharmacol. Ther. *130*, 348–363. https://doi.org/10.1016/j.pharmthera.2011.03.004.
27. Kim, Y., Kim, J., Son, M., Lee, J., Yeo, I., Choi, K.Y., Kim, H., Kim, B.C., Lee, K.H., and Kim, Y. (2022). Plasma protein biomarker model for

screening Alzheimer disease using multiple reaction monitoring-mass spectrometry. Sci. Rep. *12*, 1282. https://doi.org/10.1038/s41598-022-05384-8.

28. Chen, P.C., Han, X., Shaw, T.I., Fu, Y., Sun, H., Niu, M., Wang, Z., Jiao, Y., Teubner, B.J.W., Eddins, D., et al. (2022). Alzheimer's disease-associated U1 snRNP splicing dysfunction causes neuronal hyperexcitability and cognitive impairment. Nat. Aging *2*, 923–940. https://doi.org/10.1038/s43587-022-00290-0.
29. Jiang, Y., Wan, M., Xiao, X., Lin, Z., Liu, X., Zhou, Y., Liao, X., Lin, J., Zhou, H., Zhou, L., et al. (2023). GSN gene frameshift mutations in Alzheimer's disease. J. Neurol. Neurosurg. Psychiatry *94*, 436–447. https://doi.org/10.1136/jnnp-2022-330465.
30. Cheng, Y.C., Huang, C.J., Ku, W.C., Guo, S.L., Tien, L.T., Lee, Y.J., and Chien, C.C. (2022). Downregulated Calcium-Binding Protein S100A16 and HSP27 in Placenta-Derived Multipotent Cells Induce Functional Astrocyte Differentiation. Stem Cell Rev. Rep. *18*, 839–852. https://doi.org/10.1007/s12015-021-10319-3.
31. Das, M.R., Chang, Y., Anderson, R., Saunders, R.A., Zhang, N., Tomberlin, C.P., Vale, R.D., and Jain, A. (2023). Repeat-associated non-AUG translation induces cytoplasmic aggregation of CAG repeat-containing RNAs. Proc. Natl. Acad. Sci. USA *120*, e2215071120. https://doi.org/10.1073/pnas.2215071120.
32. Bunger, M.K., Wilsbacher, L.D., Moran, S.M., Clendenin, C., Radcliffe, L.A., Hogenesch, J.B., Simon, M.C., Takahashi, J.S., and Bradfield, C.A. (2000). Mop3 is an essential component of the master circadian pacemaker in mammals. Cell *103*, 1009–1017. https://doi.org/10.1016/s0092-8674(00)00205-1.
33. Welz, P.S., Zinna, V.M., Symeonidi, A., Koronowski, K.B., Kinouchi, K., Smith, J.G., Guillén, I.M., Castellanos, A., Furrow, S., Aragón, F., et al. (2019). BMAL1-Driven Tissue Clocks Respond Independently to Light to Maintain Homeostasis. Erratum in. Cell *177*, 1436–1447.e12. https://doi.org/10.1016/j.cell.2019.05.009.
34. Liu, D., Nanclares, C., Simbriger, K., Fang, K., Lorsung, E., Le, N., Amorim, I.S., Chalkiadaki, K., Pathak, S.S., Li, J., et al. (2023). Autistic-like behavior and cerebellar dysfunction in Bmal1 mutant mice ameliorated by mTORC1 inhibition. Mol. Psychiatry *28*, 3727–3738. https://doi.org/10.1038/s41380-022-01499-6.
35. Sharma, A., Sethi, G., Tambuwala, M.M., Aljabali, A.A.A., Chellappan, D.K., Dua, K., and Goyal, R. (2021). Circadian Rhythm Disruption and Alzheimer's Disease: The Dynamics of a Vicious Cycle. Curr. Neuropharmacol. *19*, 248–264. https://doi.org/10.2174/1570159X18666200429013041.
36. Liu, W.W., Wei, S.Z., Huang, G.D., Liu, L.B., Gu, C., Shen, Y., Wang, X.H., Xia, S.T., Xie, A.M., Hu, L.F., et al. (2020). BMAL1 regulation of microglia-mediated neuroinflammation in MPTP-induced Parkinson's disease mouse model. FASEB J. *34*, 6570–6581. https://doi.org/10.1096/fj.201901565RR.
37. Zheng, Y., Pan, L., Wang, F., Yan, J., Wang, T., Xia, Y., Yao, L., Deng, K., Zheng, Y., Xia, X., et al. (2023). Neural function of Bmal1: an overview. Cell Biosci. *13*, 1. https://doi.org/10.1186/s13578-022-00947-8.
38. Iyer, S., Bhatia, P., Rao, M., and Mukherjee, O. (2018). Developing two reference control samples for the Indian population. Stem Cell Res. *30*, 38–42. https://doi.org/10.1016/j.scr.2018.05.001.
39. Zhang, S.C., Wernig, M., Duncan, I.D., Brüstle, O., and Thomson, J.A. (2001). *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. Nat. Biotechnol. *19*, 1129–1133. https://doi.org/10.1038/nbt1201-1129.
40. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
41. Morgan, M., Obenchain, V., Hester, J., and Pagès, H. (2023). SummarizedExperiment: SummarizedExperiment Container. https://doi.org/10.18129/B9.bioc.SummarizedExperiment.
42. Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521. https://doi.org/10.12688/f1000research.7563.2.
43. Zhu, A., Ibrahim, J.G., and Love, M.I. (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.
44. Kumar, D., Hussain, A., Srivastava, A.K., Mukerji, M., Mukherjee, O., and Faruq, M. (2018). Generation of three spinocerebellar ataxia type-12 patients derived induced pluripotent stem cell lines (IGIBi002-A, IGIBi003-A and IGIBi004-A). Stem Cell Res. *31*, 216–221. https://doi.org/10.1016/j.scr.2018.08.008.
45. Omi, N., Tokuda, Y., Ikeda, Y., Ueno, M., Mori, K., Sotozono, C., Kinoshita, S., Nakano, M., and Tashiro, K. (2017). Efficient and reliable establishment of lymphoblastoid cell lines by Epstein-Barr virus transformation from a limited amount of peripheral blood. Sci. Rep. *7*, 43833. https://doi.org/10.1038/srep43833.
46. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10. https://doi.org/10.14806/ej.17.1.200.
47. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.
48. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
49. Zhu, A., Srivastava, A., Ibrahim, J.G., Patro, R., and Love, M.I. (2019). Nonparametric expression analysis using inferential replicate counts. Nucleic Acids Res. *47*, e105. https://doi.org/10.1093/nar/gkz622.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Pluripotent Stem Cell 4-Marker Immunocytochemistry Kit	ThermoFisher Scientific	A24881
Anti-Myc tag antibody - ChIP Grade	Abcam	Cat# ab9132; RRID: AB_307033
Anti-6X His tag® antibody	Abcam	Cat# ab9108; RRID: AB_307016
Anti-POLYQ-CT	This Paper	N/A
Anti-POLYS-CT	This Paper	N/A
CY3- (CTG)6-CT LNA Probes	This Paper	N/A
MBNL1 Polyclonal Antibody	ThermoFisher Scientific	Cat# PA5-29541; RRID: AB_2547017
Anti-Nestin	abcam	Cat# ab22035; RRID: AB_446723
Rabbit Polyclonal GATA-4 Antibody	Novus biologicals	NBP2-16613
Anti- Brachyury	abcam	Cat# ab209665; RRID: AB_2750925
Bacterial and virus strains		
Epstein Barr virus (EBV)	ATCC	VR-1492-5ml
Biological samples		
Patient derived cell lines	Kumar et al.[38]	IGIBi002-A, IGIBi003-A and IGIBi004-A
Healthy control derived cell line (D161)	Iyer et al.[39]	ADBSi001-A
Healthy control derived cell line (10223)	This paper	N/A
Mouse Embryonic Fibroblast (MEF)	ThermoFisher Scientific	A24903
Chemicals, peptides, and recombinant proteins		
SFM-NSC- Kit	ThermoFisher Scientific	A1050901
Human FGF-basic	ThermoFisher Scientific	PHG6015
Sodium butyrate	Merck	B5887-1G
Gibco™ 2-Mercaptoethanol	ThermoFisher Scientific	21985023
Ascorbic Acid	Merck	A4544
B-27 SUPPLEMENT	ThermoFisher Scientific	17504044
N2 Supplement	ThermoFisher Scientific	17502048
Rock Inhibitor	Merck	Y0503-1MG
Critical commercial assays		
STELLARIS RNA FISH PROBE KIT	Biosearch technologies	N/A
FITC Annexin V/Dead Cell Apoptosis Kit	ThermoFisher Scientific	V13242
Lipofectamine™ 3000 Transfection Reagent	ThermoFisher Scientific	L3000001
HiScribe T7 RNA synthesis kit	NEB	E2040S
Deposited data		
RNA sequencing analysis	This Paper	PRJNA391759, NCBI
Experimental models: Cell lines		
HEK293T cell line	Dr. Binu kumar Lab,CSIR-IGIB	N/A
Oligonucleotides		
Primers for cloning, see Table S1	This Paper	N/A
Recombinant DNA		
RAN plasmid	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
tximport (v.1.22.0)	Patro et al.[40]	N/A
DESeq2 (v.1.34)	Ewels et al.[41]	N/A
apeglm (v1.8.0)	Morgan et al.[42]	N/A
Limma (v3.42.2	Soneson et al.[43]	N/A
Fiji ImageJ software	https://imagej.net/software/fiji/	N/A

RESOURCE AVAILABILITY

Lead contact

Requests for resources, reagents, protocols and any information should be directed to the lead contact Dr. Mohammed Faruq (Faruq.mohd@igib.res.in, faruq.mohd@igib.in).

Materials availability

The resources generated in the study (Anti-POLYQ-CT, Anti-POLYS-CT, Anti-POLYS-CT, CY3- (CTG)6-CT LNA Probes, Healthy control derived cell line, Primers for cloning and RAN plasmid) are available upon request directed to directed the lead contact, and as per institutional material transfer policy.

Data and code availability

- The data used for RNA sequencing analysis and figures are available at NCBI with Bioproject PRJNA391759 and GitHub: https://github.com/viv3kanand/SCA12-RNA-Seq-Analysis.
- All the software used are mentioned in key resources table, further analysis codes are available from the lead author upon request.
- Comprehensive details on experiment design, execution protocols, and any additional information are readily accessible from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Three patients of SCA12: IGIBi002-A (age 49, Male), IGIBi003-A (age 50, Female) and IGIBi004-A (age 49, Male), and 2 Controls- ADBSi001-A[38] (age 88, Female) and 10223 (Male) Lymphoblastoid Cell line (LCL)s were selected for cell line derivation. The study was approved by the Institute Ethics Committee & Institutional Committee for Stem Cell Research, AIIMS, New Delhi (Ref no. IEC/NP-26/2013 & IC-SCRT/19/14R) and the Human Ethics committee of CSIR-IGIB for GENCODE (BSC0123). Informed written consent was obtained from each subject. The samples were collected in strict compliance with the ethical guidelines of Indian Council of Medical Research (ICMR).

SK-NSH cell lines: RNA foci; HEK293: Demonstration of RAN; NSCs (iPSC derived): 2 NSCs from control subjects (10223, D161), 3 SCA12 patients [21/14 (IGIBi004-A), 2555 (IGIBi002-A), 11/08 (IGIBi003-A)]; iPSC: reprogrammed LCLs;[44] Differentiated neurons (10223-control and 2555-case [IGIBi002-A]). Refer Table 1.

METHOD DETAILS

Cell culture

Peripheral blood mononuclear cells (PBMC) isolation and Lymphoblastoid Cell Line (LCL) generation

PBMCs were isolated from one healthy (control) individual using the Histopaque-1077 (Sigma- Aldrich) gradient method according to the manufacturer's protocol. Isolated PBMCs were then transformed to generate lymphoblastoid cell lines (LCLs) using Epstein Barr virus (EBV) following previously published protocol.[45] These LCLs were maintained in RPMI 1640 medium (Gibco) supplemented with 20% FBS (Gibco), 2 mM glutamine, and 1X Penicillin & Streptomycin (Pen-Strep) (Gibco).

Generation of induced pluripotent stem cells (iPSCs) from LCL

iPSCs generation was done utilizing Okita's episomal plasmids, with some modifications. In brief, LCL cell lines were passaged 2 days before nucleofection and maintained in an iPSC medium containing knockout DMEM F-12 (Gibco), 20% knockout serum replacement (Gibco), 55 mM beta-mercaptoethanol (Sigma), 10 mM nonessential amino acids (Gibco), 1X glutamax (Gibco), 1X Pen-Strep, 10 ng/ml recombinant human bFGF (Gibco). On day 0, cells were counted manually, by using a hemocytometer, and nucleofection was done with 2μg of plasmid cocktail on 1 million LCLs. After which, cells were maintained in an iPSC medium containing Vitamin C (50μg/ml) + NaB (0.5Mm) on day 2. The medium was changed every alternate day. Cells were transferred to Mouse Embryonic Fibroblast (MEF) feeder cell line on the 8th day in the same culture medium till the 12th day. After which, sodium butyrate was withdrawn from the medium. On day 22, Human ES-like colonies were

observed. Following that, iPSC colonies were passaged manually using insulin syringes through a dissection microscope, to a new MEF feeder layer dish with iPSC medium containing 10uM rock inhibitor. The iPSCs were characterized using 4 pluripotency markers. (Figure S5). Further, Embryoid bodies (EBs) generated with these iPSCs showed all three germ layer markers (ectoderm, mesoderm, and endoderm) (Figure S6).

Neural stem cells (NSCs) generation and pan neuronal differentiation

Neural stem cells (NSCs) were differentiated from three Spinocerebellar ataxia type-12 iPSCs (IGIBi002-A, IGIBi003-A, and IGIBi004-A) (published previously),[44] and two Control iPSCs, ADBSi001-A,[38] and one in house iPSC generated from a healthy individual (Cell line ID. 10223) (Figures S5 and S6) by using previously published protocol[39] with slight modifications. Briefly, the iPSCs were cut manually by an insulin syringe making 100-150 clumps, then these clumps were grown in an adherent free culture dish in the presence of an iPSC medium without basic FGF. The medium was changed every alternate day, and after 5 days floating iPSCs or embryoid bodies (EBs) were observed. After the 5th day, the medium was changed to neural induction medium (DMEM/F12 (Gibco) 1X glutamax (Gibco), 1X MEM non- essential amino acids (Gibco), N2 supplement (1X) (Gibco), B27 supplement without vitamin A (2X) (Gibco), basic FGF (20ng/ml) (Gibco), EGF (10ng/ml) (Gibco), heparin (2ug/ml) (sigma), pen-strep (1X) (Gibco). After 4-5 days of culture in a neural induction medium, EBs generated rosette-like structures at their center. These structures were marked and cut manually and transferred onto laminin-coated dishes. On the 10th day, the medium was changed to neural expansion medium having knock-out DMEM/F12, 1X Stem-pro neural supplement, 1X pen- strep, 1X glutamax, FGF-2 (10ng/μl), EGF (10ng/μl). On the 12th day, the dish was covered with a homogenous population of neural stem cells (NSCs). Further, NSCs were passaged with 30-40% confluency in another laminin-coated dish (Day 0) and left overnight for cells to adhere to the surface. For pan neuron generation, on day 1, the medium was changed to neural differentiation medium containing neurobasal medium (Gibco), B27 supplement with vitamin A (2X) (Gibco), N2 supplement (1X) (Gibco), glutamax (1X) (Gibco), MEM non-essential amino acids (1X) (Gibco), pen-strep (1X) (Gibco). The medium was changed every 3rd day, and mature pan neurons were observed on day 15th. Further, the neurons were characterized using mature neuronal markers TUJ1 (Beta III tubulin) and MAP2 (Figure S7).

Secondary cell line maintenance

HEK293T cell line and SK-N-SH Cell line were maintained in 12% DMEM medium containing, DMEM (Gibco), 12% FBS (Fetal bovine Serum) (Gibco), 1X glutamax (Gibco), 1X penstrep (Gibco).

Plasmid transfection and transient overexpression

Using the manufacturer's protocol, RAN constructs plasmids with a variable number of CAG repeats that were transfected in the HEK293T cell line using Lipofectamine TM 3000 transfection reagent (Invitrogen). In Brief, HEK293T cells were seeded in 6 well culture plates. After the cells reached 50% confluency, for each well, a mixture of Lipofectamine, P3000 reagent, and Opti-MEM (Gibco) was mixed (1.5ul + 1.5 ul + 100ul) with 1.5ug of plasmid construct. The medium was changed with fresh 10% DMEM after 6 hours of transfection.

RAN construct generation

To investigate if RAN translation can occur from the *PPP2R2B* gene ORF, different CAG lengths of *PPP2R2B* were amplified using SCA12 positive patient's DNA and cloned in pcDNA 3.1 vector including three epitope tags in each reading frame at the C-terminus. Also, the plasmid was further mutated at the ATG start codon using site-directed mutagenesis. To investigate expression without the AUG start codon in expanded CAG-expressing cells.

Designing of novel RAN-specific antibodies & Immunofluorescence (IF) microscopy

To test if RAN proteins are getting expressed in Spinocerebellar ataxia type-12 patient-derived neural cells, we designed two polyclonal antibodies from C-terminal regions of the *PPP2R2B* gene to generate peptide for frame-specific polyclonal antibodies, obtained from Biomatik. Cells were fixed with 4% paraformaldehyde (PFA) for 20 minutes followed by permeabilization with 0.1% Triton- X 100 for 10 minutes. The cells were kept in a blocking solution (1% BSA) for one hour at RT, incubated overnight at 4°C with diluted primary antibodies. The cells were washed thrice with 1X PBS (5 minutes per wash) and incubated for 1 hour with diluted secondary antibodies at RT. Counterstaining for nuclei was done with Slowfade gold reagent DAPI (Invitrogen) for 20 minutes at RT. Imaging was done using confocal microscopy (NIKON A1R) and Zoe Cell Imager microscope.

RNA fluorescence in situ hybridization (FISH)

To detect RNA/Nuclear foci inside the nucleus of Patient-derived cell lines, we used CY3- (CTG)6-CT LNA modified at nucleotide positions 2, 5, 8, 13, 16, and 19 (Takara Bio), and Stellaris RNA FISH kit (Biosearch Technologies) was used to perform RNA FISH (Fluorescence *in situ* hybridization) experiments. In brief, Neural Stem Cells (NSC) were grown on confocal dishes (Thermo) and then washed with 1X DPBS, fixed with fixation buffer (Stellaris RNA FISH kit) for 10 min at RT, and permeabilized with 70% ethanol for 1 hour at 4°C. Subsequently, washing was done using Wash Buffer A, and 100μl of Hybridization Buffer containing CY3-(CTG)6-CT probe (150pm/μl) was added and the confocal dish was transferred in a humidified chamber and incubated in the dark at 37°C for 16 hours. Thereafter, washing was done by Wash Buffer A, and DAPI nuclear stain (5ng/μl) was added, and incubated in the dark at 37°C for 30 min. Later, Wash Buffer B was added to the dish, and imaging was done on a confocal microscope (NIKON A1R).

Protein pull down assay

To identify different RNA binding proteins with our *PPP2R2B* specific CAG sequence, we used one expanded RAN CAG construct (CAG = 54) and one normal length RAN CAG construct (CAG= 10), as mentioned earlier. Using that template, first linear DNA was amplified using T7 promoter-specific primer and *PPP2R2B* specific Reverse primer (Table S1). RNA was synthesized by *in vitro* transcription, using HiScribe T7 RNA synthesis kit (NEB) using manufacturer's protocol but using Biotinylated-CTP instead of dCTP, to make the resulting amplified product to bind with streptavidin. Further streptavidin was added and nuclear protein (isolated from Spinocerebellar ataxia type-12 derived NSCs and Control NSCs) was allowed to interact with RNA, and incubated at Room Temperature for 30 min, then bound proteins were isolated using the magnetic stand, and sent for Mass Spectrometry.

RNA seq analysis quality control and read alignment

The initial quality check was performed using FastQC (v0.11.9), which also served to identify potential batch effects due to different sequencing sources The samples had an average of 35 million reads and a 56% GC content. FastQC's different modules were used to analyze all files and to verify read quality, read length, GC content, N content, adapter content, and potential sources of contamination. Substandard samples, or those exhibiting adapter content, were filtered out using Cutadapt (v.3.4).[46] The trimmed reads were aligned to the reference transcripts (Gencode Release 19 - GRCh37.p13) using the pseudo aligner Salmon (v.1.7.0)[47] with options (numGibbsSamples=30) for inferential replicates. Salmon provided quantification of each transcript in *.quant files. Post-alignment QC was performed using MultiQC (v1.11)[40] to verify the quality of alignment based on Salmon logs, FastQC reports, and Cutadapt logs.

Differential expression analysis

The *.quant files were gene length scale transformed and imported into R as a SummarizedExperiment object (v.1.24.0)[41] using tximport (v.1.22.0).[42] The imported read counts were pre-filtered with a read-count cut-off of at least 5 reads mapped on a gene for a minimum of 3 samples. We used the DESeq2 (v.1.34)[12] R Bioconductor package for the quantification of differentially expressed genes, considering sex and batch as confounding factors. Finally, the apeglm (v1.8.0)[43] method was used to estimate a shrunken fold change for better ranking and accurate effect size of fold changes.

DESeq2 corrects batch effects based on the model design mentioned as confounding factors before performing differential expression analysis. However, the variance-stabilizing transformation function in DESeq2 does not use the model design for variance removal. Therefore, it is anticipated that variance associated with the batch or other confounding variables will persist. For data representation purposes, the transformed data was batch corrected using the removeBatchEffect function from the Limma (v3.42.2)[48] R Bioconductor package. The quality of batch correction was assessed using Principal Component Analysis (PCA).

Differential transcript expression

Salmon outputs were imported without summarizing to gene level and filtered the data for any transcripts having 10 reads for at least 3 samples. The batch effect was corrected using Limma (v3.42.2)[48] on each inferential replicate and used fishpond (2.6.0)[49] for transcript-level differential expression.

Image acquisition and analysis

Characterization images of iPSC, RNA- FISH, RAN translation in NSCs, and neuronal characterization images were taken on A1R confocal microscope by Nikon, and analysis was done using raw TIFF files on Fiji ImageJ software, Images of Embryoid bodies characterization, having 3 germ layer markers and HEK293T cell line images for RAN translation were taken on ZOE cell imager microscope.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical tests used are mentioned in the figure legends, and the results were acquired from at least three independent cell culture procedures. Student t-test was used to calculate the significance between Spinocerebellar ataxia type-12 positive and Control cell lines in PolyGlutamine and PolySerine analysis.

Ingenuity pathway analysis

Ingenuity Pathway Analysis (IPA) for all the samples was done using Qiagen's Ingenuity Pathway Analysis (IPA) software (Table S2).