



RESEARCH PAPER

 OPEN ACCESS 

## Guanine Nucleotide-Binding Protein-Like 1 (GNL1) binds RNA G-quadruplex structures in genes associated with Parkinson's disease

Marc-Antoine Turcotte, Jean-Michel Garant , H el ene Cossette-Roberge, and Jean-Pierre Perreault

Department of Biochemistry, Pavillon de Recherche Appliqu ee Sur le Cancer, Universit e de Sherbrooke, Sherbrooke, Qu ebec, Canada

### ABSTRACT

RNAs are highly regulated at the post-transcriptional level in neurodegenerative diseases and just a few mutations can significantly affect the fate of neuronal cells. To date, the impact of G-quadruplex (G4) regulation in neurodegenerative diseases like Parkinson's disease (PD) has not been analysed. In this study, *in silico* potential G4s located in deregulated genes related to the nervous system were initially identified and were found to be significantly enriched. Several G4 sequences found in the 5' untranslated regions (5'UTR) of mRNAs associated with Parkinson's disease were demonstrated to in fact fold *in vitro* by biochemical assays. Subcloning of the full-length 5'UTRs of these candidates upstream of a luciferase reporter system led to the demonstration that the G4s of both Parkin RBR E3 Ubiquitin Protein Ligase (PRKN) and Vacuolar Protein Sorting-Associated Protein 35 (VPS35) significantly repressed the translation of both genes in SH-SY5Y cells. Subsequently, a strategy of using label-free RNA affinity purification assays with either of these two G4 sequences as bait isolated the Guanine Nucleotide-Binding Protein-Like 1 (GNL1). The latter was shown to have a higher affinity for the G4 sequences than for their mutated version. This study sheds light on new RNA G-quadruplexes located in genes dysregulated in Parkinson disease and a new G4-binding protein, GNL1.

### ARTICLE HISTORY

Received 2 September 2020  
Revised 23 October 2020  
Accepted 5 November 2020

### KEYWORDS

G-quadruplex; translation;  
RNA binding proteins;  
transcriptome;  
neurosciences

### Introduction

G-quadruplexes (G4) are non-canonical secondary structures found in both DNA and RNA molecules. In cells, as well as *in vitro*, four guanine residues can be linked together by Hoogsteen bonds in order to form a structure called a G-quartet that can be stabilized by the presence of a monovalent cation, usually potassium [1]. Canonical G4 structures respect the motif G<sub>3</sub>-N<sub>1-7</sub>-G<sub>3</sub>-N<sub>1-7</sub>-G<sub>3</sub>-N<sub>1-7</sub>-G<sub>3</sub> (where N indicates A, U, C or G). However, some studies have shown that non-canonical G4s may have longer loops, only two quartets, or the presence of bulges in G-runs [2–4]. While DNA G4s can be present in many conformations, RNA G4s, which are more stable than their DNA counterparts, are restricted to folding into parallel conformations [5].


Studies on G4s have demonstrated that they exhibit many regulatory functions in cells. For example, DNA G4s have been demonstrated to be implicated in DNA replication, transcription and in telomere elongation [6–8]. In mRNA, G4s have been located in the 5' untranslated region (UTR), the coding sequence (CDS) and the 3'UTR [1]. G4s located in the 5'UTR have been shown to impair ribosome scanning, leading to repression of the cap-dependent translation [9,10]. However, some G4s have been also demonstrated to enhance cap-independent translation by the stabilization of internal ribosome entry sites [11]. On the other hand, CDS G4s have been shown to impair translational elongation and to induce frameshifting [12,13]. Finally, G4s located in the 3'UTR have been shown to be associated with both translational repression and the regulation of polyadenylation [14,15].

Furthermore, G4s were found to have an impact on other types of RNAs, like pre-miRNA, where they have been shown to regulate RNA processing [16]. Lastly, in long non-coding RNA, they often sequester G4 RNA binding proteins [17].

It has been proposed that RNA G4s are globally unfolded in eukaryotes [18]. Since they are known to be relatively stable structures, intracellular mechanisms must regulate their formation. Yet only a few proteins have been shown to be able to fold, unfold or stabilize G4s [19]. Examples of these RNA binding proteins are the G-Rich RNA Sequence Binding Factor 1 (GRSF1), a protein known to unfold RNA G4s and to facilitate degradosome-mediated decay, and nucleolin, a protein known to stabilize RNA G4s [20,21]. Since each G4 function is thought to be regulated by specific binding proteins, targeting them appears to be an attractive avenue for controlling the associated post-translational events [22].

Over the years, several G4 motifs have been associated with important mechanisms in neurological diseases [23,24]. For example, the RNA foci C9orf72, which forms multiple repeats of G4s, has been linked to both frontotemporal dementia and amyotrophic lateral sclerosis [25]. A mutated miRNA associated with Alzheimer's disease has been demonstrated to fold into a G4 [26]. In fragile X syndrome, fragile X mental retardation protein, a protein known to both bind to mRNAs containing a G4 structure and to regulate their transport, was shown to be less abundant. This reduction also directly affects the translation of several other mRNAs [27]. In the

**CONTACT** Jean-Pierre Perreault  [jean-pierre.perreault@usherbrooke.ca](mailto:jean-pierre.perreault@usherbrooke.ca)  Department of Biochemistry, Pavillon de Recherche Appliqu ee Sur le Cancer, Universit e de Sherbrooke, Sherbrooke, Qu ebec J1E 4K8, Canada

 Supplemental data for this article can be accessed [here](#).

  2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

present study, the original identification of the Guanine Nucleotide-Binding protein-like 1 (GLN1) as an RNA binding protein (RBP) that binds to the G4 structures located in the 5'UTRs of both VPS35 and PRKN, two genes that are deregulated in Parkinson's disease, is reported.

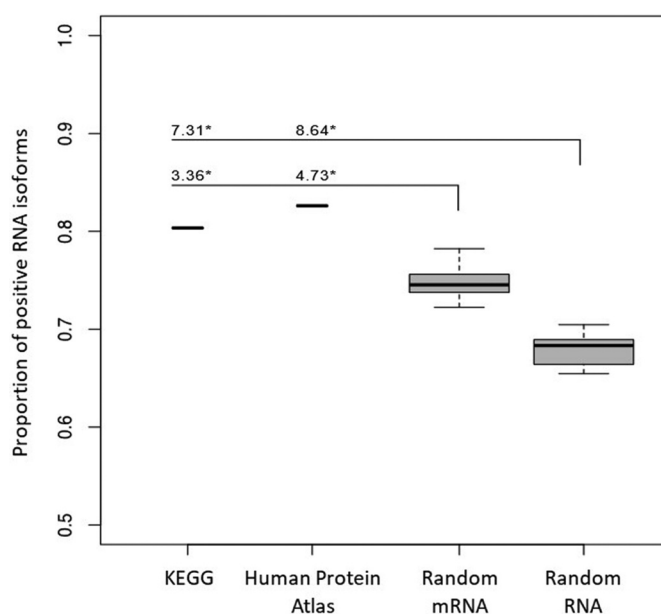
## Results

### *G-quadruplexes are globally enriched in diseases associated with the nervous system*

Initially, potential G4s (pG4) related to nervous system diseases were investigated using the disease database in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [28]. In this database, each gene annotated to be deregulated for a specific disease is presented. The sequences were scanned using the G4Screener software, a predictor of potential G4 that uses three scores [29]. Two of these scores (specifically the G4H and cG/cC scores) estimate the G4 propensity, including measuring the competing nucleotide context, while the third score is a result of a neuronal network training using experimentally proved G4 sequences (G4NN) [30–32]. The use of a threshold of 0.5 or higher for the G4NN score (the default threshold proposed for the G4Screener) showed that 85% of the 758 genes analysed included at least one pG4 sequence. However, the different isoforms resulting from alternative splicing were not taken into account with the KEGG database. In order to remove this limitation, a new search for all isoforms of these genes was performed in the Refseq database using the accession numbers associated with each gene from the disease database. According to this search, 1 985 out of the 2 460 mRNA isoforms include at least one pG4 sequence (i.e. 81% of all of the isoforms include at least one pG4). The level of mRNAs that include at least one pG4 appears to be significantly higher than expected. In order to verify whether or not this was a true enrichment of pG4 in the mRNAs of nervous system diseases, the mRNA isoforms were compared to random sets of both mRNA and all other types of RNA (Fig. 1). Alternative splicing isoforms from genes deregulated in nervous system diseases that include at least one pG4 are statistically enriched when compared to both random sequence RNA sets. The use of a second database, specifically the Human Protein Atlas, also led to the conclusion that there is a higher proportion of pG4 within the enriched alternatively spliced isoform variants that are related to the nervous system (Fig. 1).

Since pG4 are abundant in the genes associated with nervous system diseases, it was decided to restrict the analysis to the neurodegenerative diseases (ND), a sub-group in the KEGG disease database. Out of the 380 genes in the ND sub-group, 255 of them included at least one pG4 sequence (67%). In terms of the alternative splicing isoforms from the RefSeq database, 846 variants out of the 1038 in the ND subgroup included at least one pG4 sequence (82%).

The location of a G4 in an mRNA is a predictor of its potential contribution in terms of post-transcriptional regulation. In order to investigate the location of the pG4 within the mRNA, their positions were established for each of the isoform variants associated with the neurodegenerative diseases:



**Figure 1.** Enrichment of pG4 in genes associated with the nervous system.

pG4s with a G4NN threshold of at least 0.5 for one window were considered as being positive pG4s. pG4s from both the nervous system section in the KEGG disease (KEGG) and the proteins enriched in the nervous system from the Human Proteins Atlas were compared with a set of random mRNAs and with random RNA. All distributions of random passed the 'One-sample Kolmogorov-Smirnov test' for normal distribution with a p-value below 0.01. Within these samples, a value must have a z-score above 2.262 in order to be considered as not being part of the distribution with a confidence level of 95% ( $\alpha = 0.05$ ), or above 3.250 in order to be considered as not being part of the distribution with a confidence level of 99% ( $\alpha = 0.01$ ).

21% were located in the 5'UTR, 46% in the CDS region and 33% in the 3'UTR. At first view, the abundance of pG4 located within the CDS region seems surprising, because G4s are globally known to be enriched in both the 5' and 3'UTRs [33].

Interestingly when one looks more closely at each mRNA associated with neurodegenerative diseases, an important enrichment of pG4 was observed for those associated with the Parkinson disease (PD) sub-group. Fifteen out of the sixteen associated genes (94%) included at least one pG4 sequence (46 out of the 49 isoform variants). More specifically, 29% of the G4s were located in the 5'UTRs, 51% in the coding regions and 20% in the 3'UTRs.

When the densities of G4s were analysed instead of the percentages, 4.3 positive windows (pw) per kilobase (kb) in the 5'UTR, 1.5 pw/kb in the CDS and 1.2 pw/kb in the 3'UTR were observed. Specifically, for the PD associated mRNA, we determined densities of 9.4 pw/kb for the 5'UTR, 2.6 pw/kb for the CDS, and 1.2 pw/kb for the 3'UTR regions. This is higher than the ND density results, highlighting the potential importance of G4s in PD. In brief, a G4 enrichment in the 5'UTRs was detected.

A G4Screener search using the default windows of 60 nucleotides (nts) yielded an output that consisted of a score indicating the probability of a pG4 sequence being located within the scanned region. Consequently, when moving from one position to another the output looks like a belt curve for which the maximal value should be indicative of the sequences most likely to fold into a G4 structure. The most

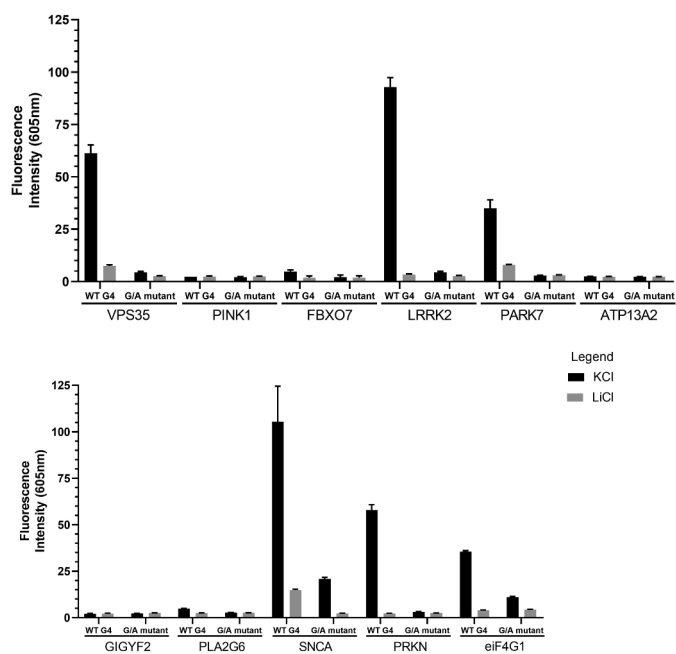
probable sequence of the pG4 located in the 5'UTR (i.e. that located at the top of the belt shape), were selected for further characterization (see Table 1 for the detailed nucleotide sequences). Surprisingly, visual analysis of these eleven sequences revealed that almost all sequences did not fulfill the definition of a canonical pG4 ( $G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$ ). Instead, only the pG4 sequence of the PLA2G6 mRNA appeared to satisfy the criteria of three G-quartets and three loops of 7 nucleotides or less (see Table 1). Most of the other sequences included pG4 sequences with only two G-quartets (eIF4G1 and SNCA). Several include central loops longer than 7 nucleotides (VPS35 and GIGYF2), and one was proposed to be formed by three G-quartets that included a bulge (PRKN). All of these features are proposed to be unfavourable to the stability of a G4 structure. Moreover, it is not clear which runs of guanosines would be forming the G4 structures. The demonstration of the folding of these pG4s into stable G4 structures was therefore mandatory before any further work could be undertaken.

### In vitro validation of the pG4

The binding of fluorescence ligands *in vitro* is a technique that is frequently used for the validation of G4 structures [34]. Initially, the binding of the fluorescent G4-specific ligand N-Methyl Mesoporphrine (NMM) was tested in order to verify whether or not the eleven pG4 sequence candidates from the 5'UTRs of the genes related to PD can fold into a G4 structure. Briefly, in addition to the candidate sequences that can potentially form pG4s, the sequences that are located up to 25 nucleotides both upstream and downstream of the pG4 were also considered in the *in vitro* transcription synthesis (See Tables S1 and S2 for the detailed sequences). The goal of this was to conserve the natural context of the tested pG4s as the neighbouring nucleotides are known to influence the folding of a G4 [30]. For example, it has been shown that a surrounding sequence rich in cytosine residues may form CG Watson-Crick base pairs that prevent the formation of the G4 structures. Alternatively, a surrounding sequence including a run of guanosines may also contribute to the formation

**Table 1.** 5'UTR pG4s located in genes related to Parkinson's disease and manual analysis of the G4 structures. The 5'UTR pG4s associated with Parkinson's disease are listed in the first column. In the second column, the boxes indicate the G4 possessing the highest score, and the most abundant alternative mRNAs are shown. A minimum of one G4 structure had to have been manually found in each sequence. In order to improve the clarity, a maximum of 2 potential structures are shown per sequence. These structures are not exhaustive, many others could occur. The positions of the nucleotides are shown respective to their location in the 5'UTR. The G4NN scores are shown in the third column.

Genes	Nucleotide sequence 5' to 3'	G4NN Scores
VPS35	GCGGGGCUUGGAGGGGCGCAGCGUCACAUGACCGCGGGAGGCUCACGCGCGGGGCGGGUG 2 G-Quartets, long central loop	0,843
PINK1	CCGGCGGGGACGCGCGGUGGUGGCGGCAGCGCGCGGCGUGCGGGGGCACCGGGCCGCGGGCGC 2 G-Quartets	0,821
FBXO7	CCGGCGUGCGCGGAGCGGAGGGUGCAGGCGACGGGAAGCGCGGGUGGUGGCGUGGGGUC 2 G-Quartets	0,817
LRRK2	GGAGGGCGCGGGUUGGAAGCAGGUGCCACCAUGGCUAGUGGCAGCUGUCAGGGGUGCGA 2 G-Quartets	0,663
PARK7	UGAGUCUGCGCAGUGUGGGGUGAGGGAGGCGGAGCGCGCGUGCGUGCGUGGGGUGCGG 2 G-Quartets	0,626
ATP13A2	AGAGGGCGGGGCGGGCTTGCGGCGCGCAAGGAGGACTGCGGCAGTGTGGAGCCGCGC 2 G-Quartets	0,624
GIGYF2	GUGACGUGCGUGGCGACGUGUGGCAUCUUGUGUUGUUGAGGCGUGAGGACUGACUGGGG 2 G-Quartets, central long loop	0,618
PLA2G6	GGAAGUAGAAGUGCUGAGUAAGCCGAGGUGAGUGACCUCGGGGUGGGGGGCGCCUGGGG 3 G-Quartets	0,599
SNCA	GAAAGGAGGACUAGGAGGAGGAGGACGCGCAGCAGACAGAAAGGGCCCAAGAGGGGG 2 G-Quartets	0,558
PRKN	CGGGCGCGGGGCGGGAGGCCUGGAGGAUUUAACCGAGAGAGCCGUGGUGGGAGGC 3 G-Quartets, 1 bulge	0,549
eIF4G1	GGGUGGGGGGUGGGGACGCCAGGCCGAAGCAGCUGACCCGUUCGUGAUCCGGGAGCC 2 G-Quartets	0,542



**Figure 2.** Characterization of RNA G-quadruplexes found in genes deregulated in PD by fluorescence assay.

The fluorescent assays were performed with either WT or G/A mutant G4 RNAs (300 pmol) of the 11 5'UTR pG4 candidates. The RNAs were incubated for 30 min with 0.5 mM NMM and 100 mM of either KCl or LiCl at room temperature. The samples were excited at 399 nm and the emission was detected 605 nm. The error bars represent the SEM. All experiments were performed in duplicate.

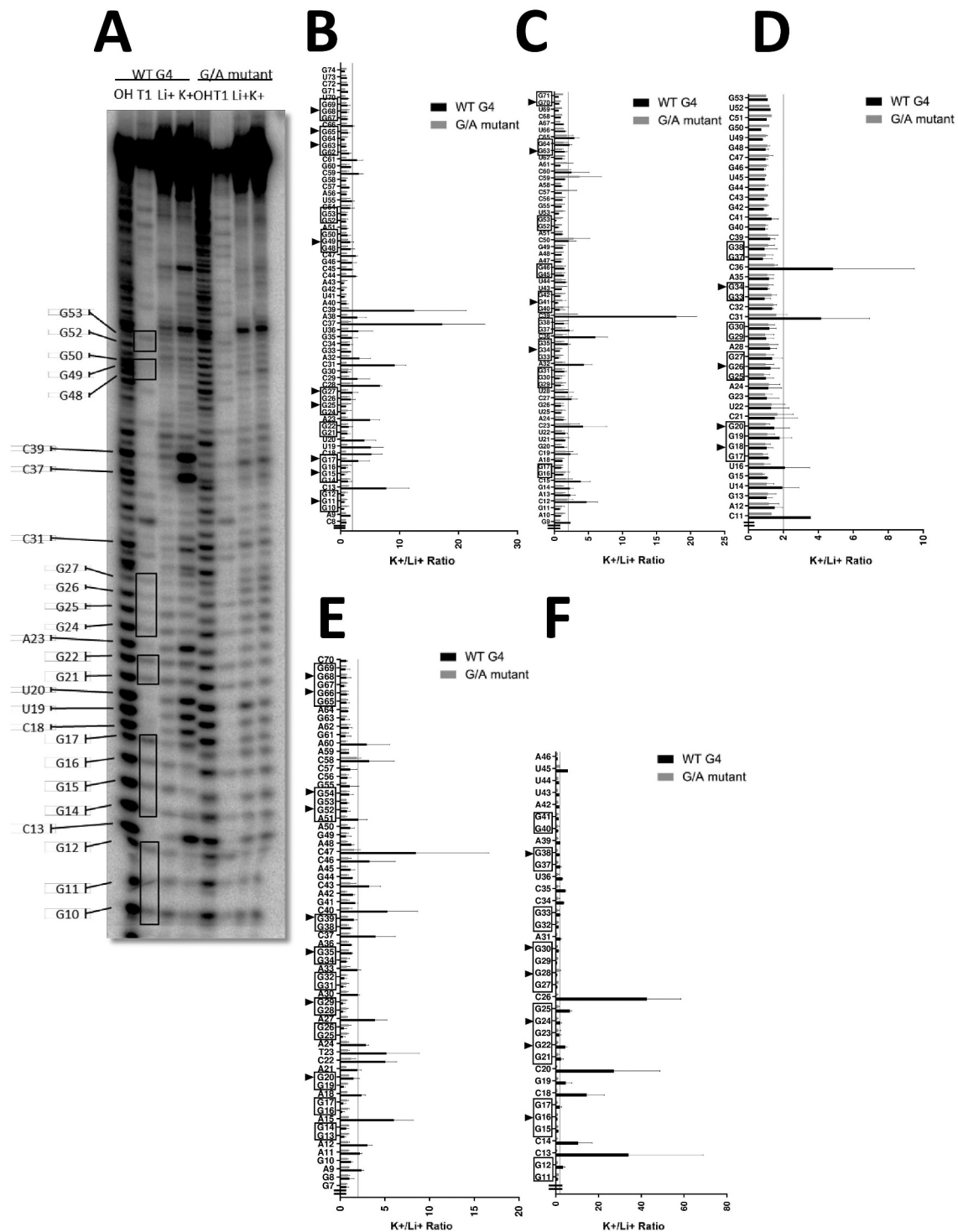
of an unexpected G4 structure. For each pG4, a G/A mutant version was also constructed. The mutants consist in the substitution of several guanosine residues by adenosines in order to prevent the formation of the G4 motifs. The RNA species were dissolved in a buffer that contained either LiCl or KCl, that is to say conditions that do and do not support G4 folding, respectively. The emission signal of NMM was detected with a peak at 605 nm [35]. G4s were considered as being positive when the signal was higher than 25 fluorescence units, which is an arbitrarily identified value for the wildtype (WT) sequence in the presence of  $K^+$ . The results for each of the pG4 candidates are presented in the histogram in Fig. 2. Six out of the eleven WT sequences exhibited a signal over 25 fluorescence units (VPS35, LRRK2, PARK7, SNCA, PRKN, and eIF4G1A). In all cases, the WT sequences in the presence of  $Li^+$  had a signal lower than 10 units, with the exception of the SNCA candidate that exhibited an increase of the fluorescent signal of around 15 units. The fluorescent signals for the G/A mutants were at background levels no matter what the salt conditions were. The only exception was the SNCA candidate where a signal of 22 units was detected in the presence of  $K^+$ , while it remained at the background level in the presence of  $Li^+$ . Most likely this is the result of complex mixture of the many distinct structures formed by this RNA species. In brief, NMM binding suggested that more than half of the pG4 sequences do indeed fold into G4 structures.

In order to increase the confidence of the pG4 folding into a G4 structure, as well as to identify which guanosine residues are involved in the G-quartets, *in-line* probing analysis were performed for each of the six retained candidates. This technique, which is often used in the characterization of RNA G4

structures, is based on the spontaneous hydrolysis of flexible RNA nucleotides (e.g. those located in single-stranded positions) in the presence of  $Mg^{2+}$  [36]. Nucleotides involved in interactions (usually those located in double-stranded positions) are less susceptible to hydrolysis. In the case of a G4 structure, the guanines involved in the G-quartets are usually not flexible, while the nucleotides that form the loops are and are therefore highly susceptible to hydrolysis. As compared to classical biophysical approaches like the circular dichroism spectrum, this approach requires only trace amounts of RNA, a fact which is considerably more biologically relevant. A typical gel for the VPS35 candidate is shown in Fig. 3A. The hydrolysis of the nucleotides involved in the loops, specifically  $C_{13}$  of loop 1,  $C_{18}$ ,  $U_{19}$  and  $U_{20}$  of loop 2 (the central loop) and  $A_{23}$  of loop 3, was observed for the WT sequence in the presence of  $K^+$ , but not in the presence of  $Li^+$ , nor for the G/A mutant. The cleavage intensities under the different conditions were evaluated by densitometry, and the  $K^+/Li^+$  ratios for each nucleotide are reported in the associated histograms (Figs. 3B and S1). A nucleotide was considered to be significantly accessible when the ratio  $K^+/Li^+$  was greater than 2. The  $K^+/Li^+$  ratio was also compared to that of the G/A mutant sequence in order to show that the increased accessibility was due to the presence of the G4 structure. For VPS35, the G4 involves the nucleotides located between positions 10 to 27 that are located near the 5' end of the synthetic transcript. In the case of LRRK2, the G4 involves the residues  $G_{29}$  to  $G_{42}$  and most likely forms a structure that contains only two quartets (Fig. 3C). The pG4 of PARK7 was not folded according to the *in-line* probing data (Fig. 3D), while the band profile of eIF4G1 was so complex that it could not be analysed and most likely results from the presence of a mixture of different structures in solution (data not shown). In the latter cases, it is important to mention that these two were the candidates with the lowest fluorescent NMM signals (see Fig. 2). For SNCA, a mixture of two quartets can be observed from positions  $G_{13}$  to  $G_{39}$  (Fig. 3E) resulting from the presence of many consecutive guanines. Finally, for PRKN, a 3 quartet G4 was formed by the residues  $G_{10}$  to  $G_{30}$  (Fig. 3F). Thus, the *in-line* probing results confirmed the folding of four out of the six candidates tested: VPS35, LRRK2, SNCA and PRKN.

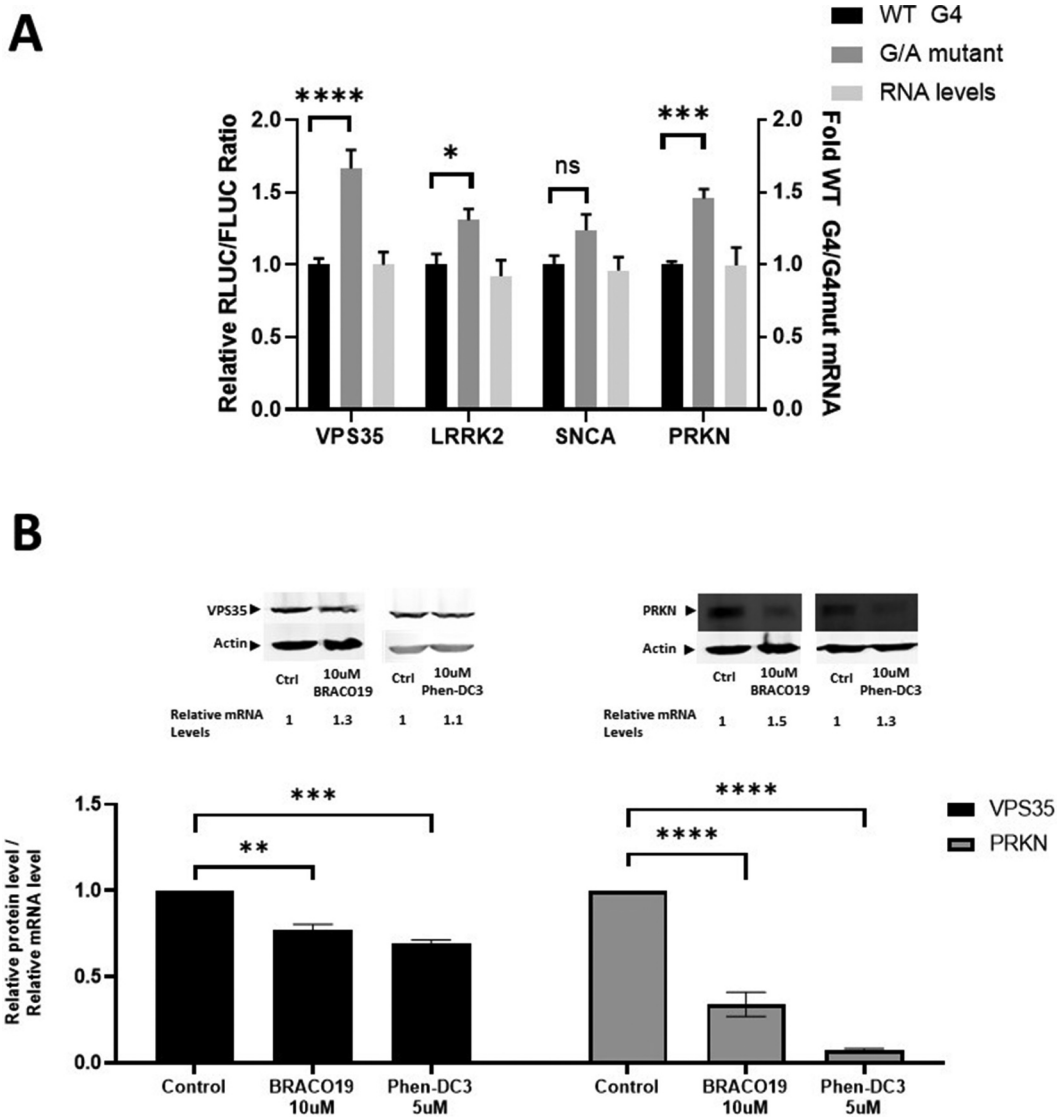
### **In cellulo evaluation of the G4 folding**

The four G4 validated by the biochemical approaches were further investigated *in cellulo*. Briefly, the full-length 5'UTR, as well as the G/A mutant versions, of the four genes were cloned upstream of the *Renilla* luciferase (RLuc) open reading frame (ORF) in the PsiCHECK-2 plasmid, which also includes a *Firefly* luciferase (FLuc) ORF for normalization purposes (see Table S1 and Fig. S2 for the detailed sequences). The resulting plasmids were transfected into SH-SY5Y cells, a well-documented model cell line for Parkinson's disease studies. This cell line originates from humans, is catecholaminergic and is easy to work with [37]. Twenty-four hours post-transfection, the RLuc and FLuc luciferase levels were evaluated by measuring their respective activities in the different cellular lysates. Interestingly, a significant decrease (~1.5 fold) in the



**Figure 3.** Characterization of the RNA G-quadruplexes found in genes that are deregulated in PD by *in-line* probing.

The *in-line* probing experiments were analysed by both polyacrylamide gel electrophoresis and densitometry. (A) Denaturing polyacrylamide gel of the *in-line* probing reactions of VPS35. Lanes 1 to 4 are for the WT, while lanes 5 to 8 are for the G/A mutant. Lanes 1 and 5: NaOH Ladder; Lanes 2 and 6: RNase T1 Ladder; Lanes 3 and 7: the Li<sup>+</sup> condition; and, Lanes 4 and 8: the K<sup>+</sup> condition. Histograms for the (B) VPS35; (C) LRRK2; (D) PARK7; (E) SNCA; and, (F) PRKN RNA G4s were traced from the *in-line* probing data. The results were derived from the densitometric analysis of each band using the SAFA software [61]. The sequence is indicated on the y-axis. The X-axis represents the K<sup>+</sup>/Li<sup>+</sup> ratios of each band's intensity for both the WT and the G/A mutant G4s. The K<sup>+</sup>/Li<sup>+</sup> ratios are shown in black for the WT G4s and in grey for the G/A mutant G4s. The nucleotides are identified by their position with respect to the 5' end of the synthetic transcript. The nucleotides marked with arrowheads are those substituted for adenine in the G/A mutant. The square guanines are those which could be expected to form a G4. "||" indicates the minimal limit of gel quantification. The dotted line represents the twofold threshold that denotes a significant gain in flexibility. Each bar represents the average of two independent experiments, and the error bars represent the standard deviations.



**Figure 4.** *In cellulo* characterization of RNA G-quadruplexes in SH-SY5Y cells.

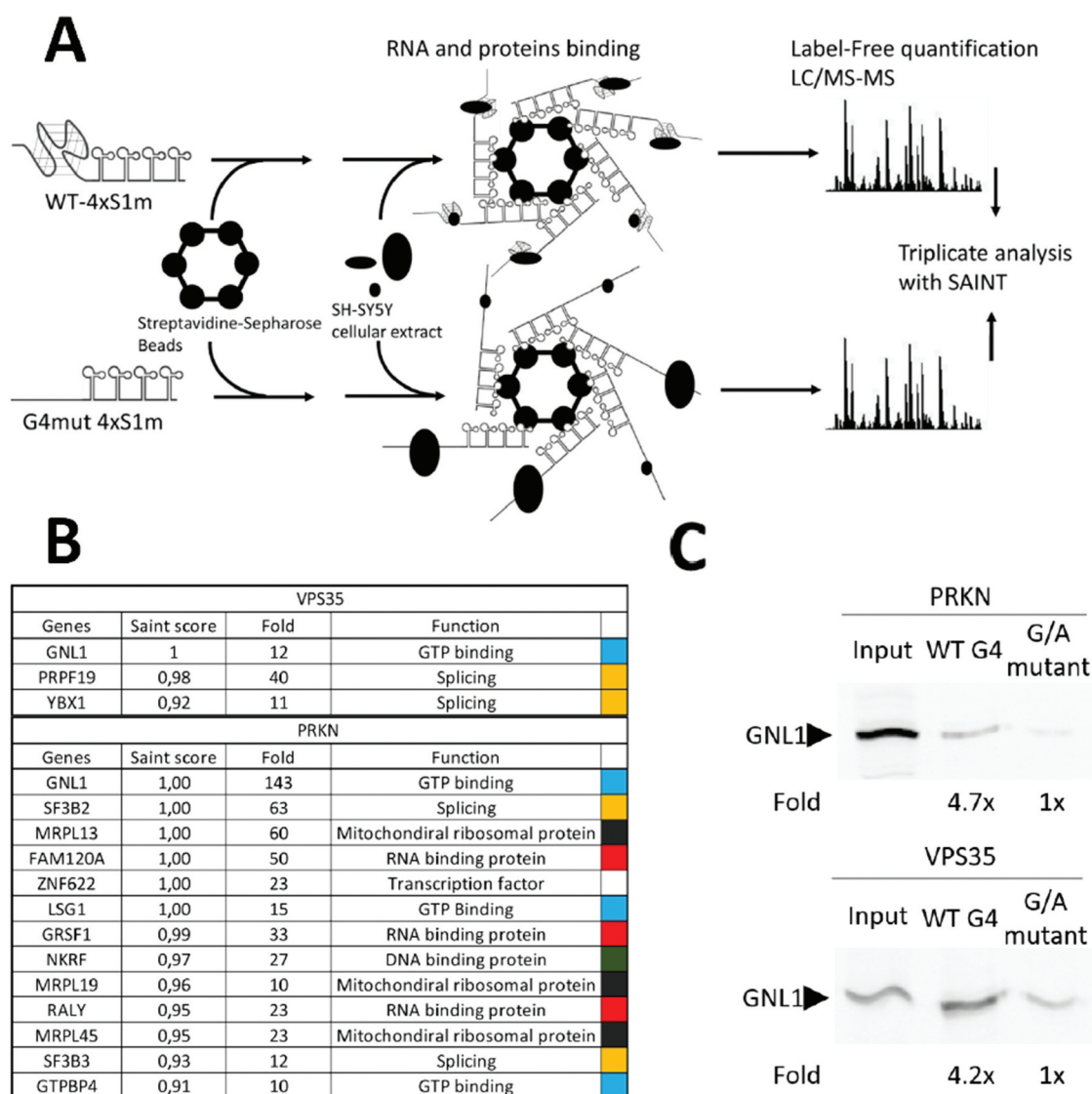
The impact of the G4s on translation was analysed using both luciferase assays and specific G4 ligands. (A) The relative RLuc/FLuc ratios for the luciferase activities of the 5'UTRs of VPS35, LRRK2, SNCA and PRKN are shown on the left y-axis. The relative RNA levels for both the WT G4 and G/A mutant, as compared to the RNA level of the control FLuc, are shown on the right y-axis. The results shown are the means of at least three independent experiments, and the error bars represent the SEM. The P-values were calculated by two-way ANOVA. (B) Upper section: Endogenous protein expression levels of VPS35, PRKN and  $\beta$ -actin were detected by western blots in the presence of BRACO19 or PHEN-DC. The mRNAs were detected by ddPCR and the relative levels are indicated. Lower section: The histogram presents the relative level of proteins over the relative levels of mRNA. The results shown are the means of two independent experiments, and the error bars represent the SEM. The P-values were calculated by two-way ANOVA. (\*)  $P < 0.05$ , (\*\*)  $P < 0.01$ , (\*\*\*)  $P < 0.001$ , (\*\*\*\*)  $P < 0.0001$ .

activity of the RLuc was observed for the WT G4s of both the PRKN and the VPS35 5'UTRs as compared to those of their respective G/A mutant versions. RT-ddPCR monitoring showed that the RNA quantities remained unchanged, confirming that the observed differences occurred at the translational and not the transcriptional level. No significant impact was observed for the G4 of the 5'UTR of SNCA, and a slight decrease was observed for the LRRK2 5'UTR G4.

The same experiment was repeated using HEK293T cells, which are often used for neuronal cell studies because of their similarity to neuronal cells in terms of development [38]. Similar results were obtained: the presence of the WT 5'UTR G4s of both PRKN and VPS35 significantly decreased the RLuc level (2.5-fold). Unlike the experiments performed in

the SH-SY5Y cells, those using the HEK293T cells revealed a slight decrease in luciferase expression in the presence of the SNCA 5'UTR G4, while no effect was observed in the case of the LRRK2 5'UTR G4 (Fig. S3).

Since the G4s of the 5'UTRs of both PRKN and VPS35 were the only ones which successfully passed both the *in vitro* and the *in vivo* tests, all subsequent experiments were limited to these two candidates. At this point, it became important to confirm that the folding of the G4 in the 5'UTR has a similar effect on the endogenous PRKN and VPS35. In order to do so, the expression levels of both proteins were evaluated and normalized to the mRNA level in both the presence and the absence of two distinct G4 ligands, BRACO19 and Phen-DC3. Both of these ligands are known to stabilize G4 structures in



**Figure 5.** RNA pull-down of proteins by the PRKN and VPS35 G-quadruplexes.

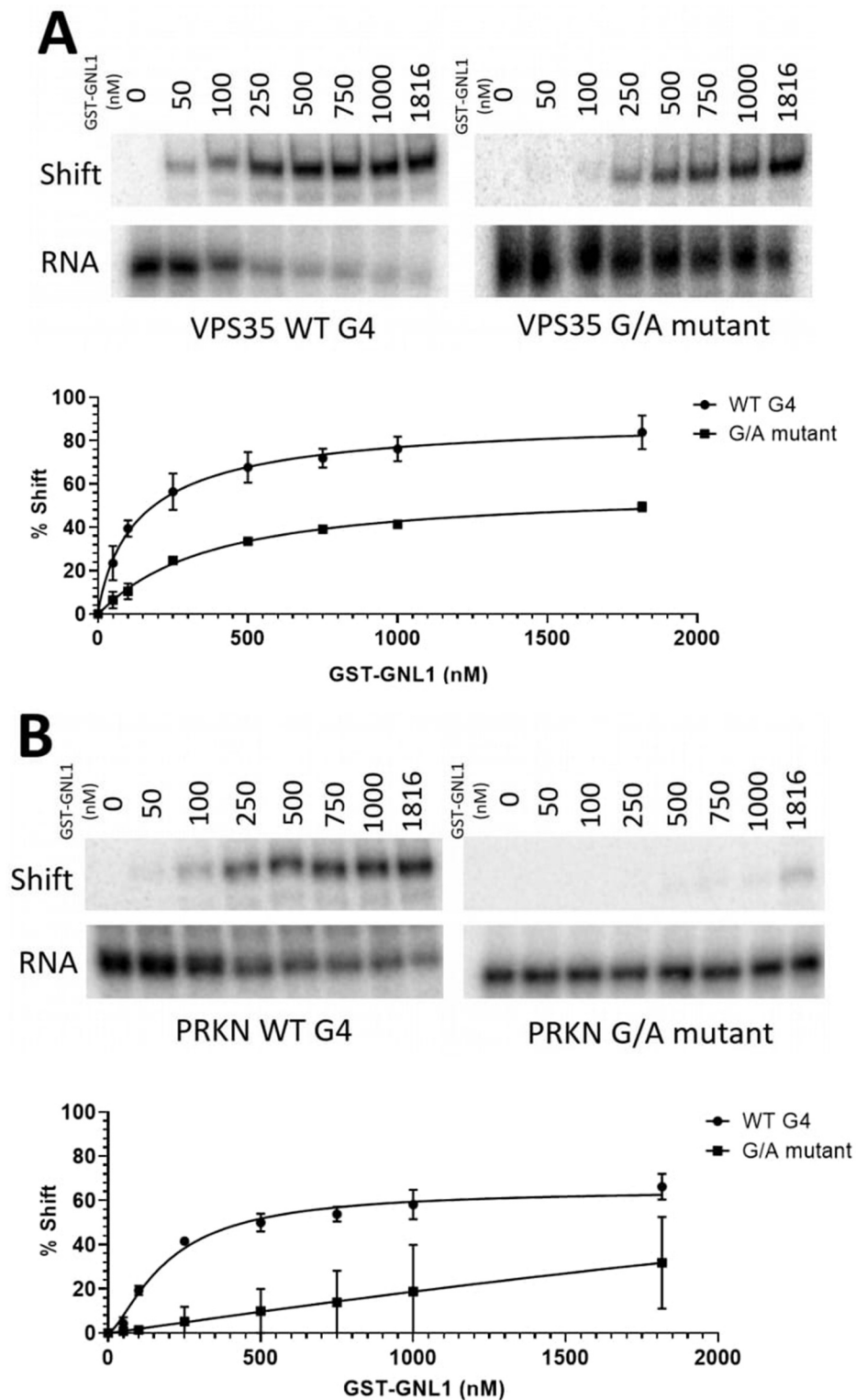
Proteins able to bind to the VPS35 and PRKN RNA G4s were identified by a pull-down experiment that used the 4xS1m aptamer and the total proteins extract from SH-SY5Y cells. (A) Illustration of the pull-down experiments. (B) The proteins that were statistically enriched by the VPS35 and PRKN RNA transcripts. Only the genes with a SAINT score >0.9 and a  $\geq 10$ -fold enrichment were kept. The G/A mutants were considered as being the control group. The fold enrichment was determined by the SAINT software. The colours facilitate the visual association of the different functions: GTP binding (Blue), Splicing (Orange), Mitochondrial ribosomal protein (Black), RNA binding protein (Red), DNA binding protein (Green) and Transcription factor (White). (C) A western blot using anti-GNL1 antibody was performed on 30  $\mu$ g of the protein input (Lane 1) and 50% of the proteins eluted from the beads (Lanes 2 and 3). The folds were calculated using Image Studio Lite (Licor).

cells [39–42]. Western blot analyses with the mRNA relative expressions revealed that the presence of Phen-DC3 and BRACO19 significantly reduced the expression levels of both proteins (Fig. 4B). These results confirmed not only that the presence of the G4 can repress the translation of both the endogenously expressed VPS35 and PRKN, but also the equivalent results obtained in the from the luciferase assays.

#### Isolation of proteins binding the G4 of either VPS35 or PRKN

It has been proposed that G4 motifs are only transiently folded in eucaryotes [41], highlighting the fact that RNA

binding proteins (RBP) are important regulators in both their folding and their unfolding. Consequently, it is of interest to identify proteins able to bind the G4s of the VPS35 and/or PRKN 5'UTRs. Label-free RNA affinity purification assays were performed in order to identify the repertoire of potential binding partners (see Fig. 5A). Initially, both the WT and G/A mutant G4 sequences of both VPS35 and PRKN were fused to four repeats of the S1m aptamer (4xS1m; see Fig. S4). This RNA aptamer has a high affinity for streptavidin [43] and its presence does not impair the G4 folding as the NMM fluorescence is not affected (Fig. S5). Moreover, *in-line* probing experiments also confirmed the proper folding of both G4s when they are fused to the S1m aptamer (Fig. S6). In order to identify any potential protein partners, total proteins from



**Figure 6.** Electrophoretic mobility shift assays of the GNL1 protein.

Electrophoretic mobility shift assays of the A) VPS35 and B) PRKN RNA G4s. The EMSA were performed using 0, 50, 100, 250, 500, 750, 1000 and 1816 nM of protein. The bands were analysed by densitometry (on the top), and the data plotted graphically (on the bottom). The GraphPad software (Prism) 'specific binding with hill slope nonlinear regression' function was used to determine the  $K_d$  values. Each assay was performed in duplicate.



SH-SY5Y cells were extracted and then pulled down using either the chimeric WT G4 or the G/A mutant RNA bound to streptavidin-coated sepharose beads. After the removal of any non-specific interactions, the proteins bound to the RNAs were recovered, trypsin digested and analysed by MS/MS. As an RNA binding control to ensure an equal quantity of RNA, we confirmed that a similar quantity of WT and G/A mutant G4 RNA was retained on the beads. These label-free RNA affinity purification assays were performed in triplicate, and were analysed using the MaxQuant software [44], followed by the probabilistic scoring of affinity purification software using the SAINT software [45] by comparing proteins observed in the WT G4 fraction to those found in the G/A mutant fraction. The SAINT output consists of a protein fold enrichment and a calculated enrichment score from 0 to 1 where the most significant proteins are at 1. The proteins with a SAINT score higher than 0.9 and with a fold enrichment of at least 10-fold were selected for both PRKN and VPS35 [45]. The SAINT data identified 3 proteins that were able to bind VPS35: specifically GNL1, a GTPase; and, two splicing factors, PRPF19 and YBX1. The later RBP is a protein already known to bind G4 [46], which increased the confidence in the results presented here (Fig. 5B). In the case of PRKN, 13 proteins were identified: three GTP binding proteins (GNL1, LSG1 and GTPGF4); two splicing factors (SF3B2 and SF3B3); three mitochondrial ribosomal proteins (MRPL13, MRPL19 and MRPL45; all of which are highly unlikely and most likely simply result from the fact that cellular extracts without prior compartment isolation were used); three other RBP (FAM120, RALY and GRSF1; the later has been previously reported to bind a G4 RNA structure [21]); one transcription factor (ZNF622); and, one DNA binding protein (NKR1) (Fig. 5B). Interestingly, GNL1 was the only protein common to both datasets of G4 binding proteins. This protein also has one of the highest enrichments and SAINT scores in both cases. GNL1 is not identified as a RBP in the literature. However, we were able to observe a low level of this protein in some RBP high-throughput experiments, confirming that GNL1 could be an uncharacterized RBP [47,48]. To confirm that GNL1 was indeed enriched, western blots were performed after the 4xS1m protein pull-down. As expected, a clear enrichment of GNL1 for VPS35 and PRKN was observed when the WT G4 was compared to the G/A mutants (i.e. ~ 4.5 folds; Fig. 5C). It was therefore decided to further investigate the interaction of GNL1 with both VPS35 and PRKN.

### **Characterization of the interaction between GNL1 and either the VPS35 or PRKN proteins**

In order to confirm that GNL1 indeed binds to both the VPS35 and the PRKN G4s, electrophoretic mobility shift assays (EMSA) were performed. Initially, recombinant GST-GNL1 proteins were produced in HEK293T cells and were purified on glutathione-Sepharose-Beads (Fig. S7). The production of recombinant proteins in human cells increases the probability that GNL1 has the post-translational modifications found in SH-SY5Y cells. The EMSA assays were performed using the same RNA molecules as were used for both

the NMM assays and the *in-line* probings. As expected, GST-GNL1 had a significantly higher affinity for the WT G4s of both the VPS35 and PRKN 5'-UTRs as compared to those for their respective G/A mutant versions (Fig. 6). The dissociation constant of the WT VPS35 was 2.5-fold greater than that for the G/A mutant (152 nM and 360 nM, respectively), while that for the WT PRKN was 35 folds smaller than that of its G/A mutant (194 nM versus 7339 nM, respectively). The results were also confirmed using a commercial GST-GNL1 obtained from Abnova (data not shown). In addition, it was confirmed that neither the GST protein alone, nor the impurities retrieved with the GST beads during the purification process, induced any shift with the various RNA constructs (data not shown). In brief, these data confirm that GNL1 binds directly both the VPS35 and PRKN G4s.

### **Discussion**

The neuronal network is a complex system with many different levels of post-transcriptional and post-translational mechanisms of regulation. Several post-transcriptional or post-translational deregulations at either the RNA or the protein levels have been shown to result in neurodegenerative diseases [49]. Hence, the observation of an enrichment of pG4 sequences in mRNA isoforms associated with neuronal diseases, as compared to what is seen in both a random RNA and mRNA samples, was not surprising (Fig. 1). Indeed G4 motifs have so far been implicated in a wide variety of post-transcriptional regulation types, including splicing, pre-miRNA processing and both cap-dependant and -independent translations, to name only a few examples [1,50]. In fact, the general belief of the research community is that G4 structures are intimately associated with fine-tuned mechanisms of both transcriptional and post-transcriptional regulation. Therefore, it was expected that many pG4 would be found in the present study since the nervous system is the most differentiated system in the human body. More specifically in PD, pG4 were detected in fifteen genes, highlighting the potential importance of gene regulation through the presence of RNA G4 structures in PD. Prior to this bioinformatic search, only a single G4 motif located in the SNCA mRNA had been reported to be associated with PD [23]. This candidate was also retrieved in the present study. The RNA G4 subgroup associated with PD includes pG4s located in both the 3'UTR and in the coding region, but the majority were located in the 5'UTR of the mRNA (11 out of 16 genes). It was therefore decided to focus on these 11 candidates as they provided the advantage that the mechanism is relatively easy to study using a luciferase reporter assay. However, it is important to keep in mind the possibility that other candidates may contribute to the regulation of other important post-transcriptional mechanisms.

The predicted G4 sequences located in the 5'UTRs of PD associated genes appeared to be relatively unstable. Only the pG4 from the PLA2G6 candidate fulfilled the criteria for a canonical G4 motif, all others did not (Table 1). Specifically, all of the others harboured several variations, variations such as the presence of a 2-quartet instead of the classical 3-quartet core, the presence of a mismatch between

the G-quartets, or the presence of a larger than 7-nts loop, to name but a few examples. This level of non-canonical pG4 is quite surprising, and may be considered as being a potential characteristic of the transcriptome of nervous cells. Further biochemical, as well as cellular and molecular biology studies, have the potential of providing physical support for this hypothesis. That said, considering that G4s have been suggested to be ‘events’ and not be ‘things’ (i.e. they are transiently folded in human cells), the abundance of non-canonical RNA G4s suggests that RBP may play a significant role in either the folding or the unfolding of these structures as most of the time non-canonical RNA G4s are less stable than canonical ones. Such RBP proteins may manage the RNA G4s by being a dynamic switch for post-transcriptional regulation.

Because of the above, the finding of a large proportion of non-canonical G4 structures makes mandatory the identification of the pG4 candidates that actually fold into a G4 structure. NMM fluorescent binding revealed that six of the candidates did indeed fold into G4 structures, although they did so in different proportions (according to the fluorescent signals). The *in-line* probing experiments confirmed that four of these six did indeed fold into G4 structures (Figs. 2 and 3). The *in vitro* validation permitted the screening of a large number of potential candidates in order to further focus the efforts on the more time consuming *in cellulo* experiments. Unfortunately, only two out the four candidates were in fact confirmed to have a significant impact *in cellulo*. This could be an indication that RBP in cells may, in fact, have an impact on the post-transcriptional regulation. While the *in vitro* assays were performed using pG4 sequences surrounded by up to 25 nts, in order to not be restricted to the positions that may specifically contribute to the G4 structures, the *in cellulo* luciferase assays were performed using the full-length 5'UTR sequences in order to be as relevant as possible to the natural genomic context. The luciferase assays revealed that the G4 motifs of both the VPS35 and PRKN 5'UTRs significantly repressed the translation (Fig. 4A). In this study, the pG4 of the SNCA 5'UTR that had previously been reported to show some limited repression of translation [23], exhibited only a light repression that cannot be deemed as being significant. This discrepancy may simply depend on the construction used, the experimental conditions, or, alternatively, on the fact that only a small proportion of the pG4 is actually folded (maybe because it is a motif based solely on 2G-quartets; Table 1). More importantly, the folding of the G4 from both the endogenous VPS35 and PRKN mRNAs received physical support from ligand binding experiments in cells (Fig. 4B). All of these experiments provide much information on the stability of the G4s related to PD. The vast majority of them seem to show that these G4 are globally unstable. Most are formed by non-canonical G4s, thus reducing the stability of the G4 structures. Also, some structures seem to be formed by an equilibrium of a mixture of different G4s (e.g. SNCA and eIF4G1).

PRKN and VPS35 were the two most promising candidates identified by the screening methodology used here. These proteins exhibit distinct functions and are differently deregulated in PD [51]. On one hand, Parkin RBR E3 Ubiquitin

Protein Ligase (PRKN) is an E3 ubiquitin ligase that is activated by PINK1 to promote the ubiquitination of the outer mitochondrial membrane proteins, leading to mitophagy. In PD, there is an inactivation and/or reduction of the ubiquitination activity of PRKN that alters the mitochondria recycling pathways [52,53]. On the other hand, Vacuolar Protein Sorting-Associated Protein 35 (VPS35) is a component of the the WASP and SCAR homologue (WASH) retromer that is implicated in endosomal trafficking. In PD, VPS35 deficiency or mutation impairs both mitochondrial fusion in dopaminergic neurons and endosome transportation [54]. In brief, both are important proteins in PD and controlling their expression may provide a new therapeutic avenue, or at least minimally provide a functional genomic approach for learning more about their contribution to the disease state. To do so, targeting G4 structures may *a priori* be an interesting approach [22]. However, one of the important hurdles of such an approach is that G4s are structurally similar and, consequently, targeting sequence specifically a G4 motif will likely be difficult to achieve using small molecules. Moreover, the presence of non-canonical RNA G4 does not simplify the situation.

Clearly, the results obtained both *in vitro* and *in cellulo*, as well as the fact that it appears to be complicated to specifically target a G4 structure in the nervous system, campaign for the identification of the RBPs that interact with both candidates. In order to do so, a pull-down approach independently using the RNA G4s of PRKN and VPS35 was performed (Fig. 5). Even though crude cellular extracts were used (i.e. there was no cellular compartment pre-isolation), most of the identified proteins were RBP, and only a few DNA binding proteins were retrieved (e.g. NKRF; Fig. 5B) even though RNA and DNA G4 structures are globally similar [55]. The use of a crude cellular extract is also most likely the reason that the mitochondrial proteins MRPL13, MRPL19 and MRPL45 were retrieved (Fig. 5B). Clearly, this is a limit with this approach. Importantly, the finding that YBX1 bound to the G4 of VPS35 and that GRSF1 bound to the G4 of PRKN, two RBP previously reported to interact with other G4 structures [21,46], provided confidence in the results presented here. There was also another interesting protein, FAM120A, which had previously been identified as a protein binding to G-Rich regions located in 3'UTRs [56]. In the latter report, the authors showed that mutating some of the guanine residues of the G-Rich regions reduced the binding of FAM120A, a characteristic of the G4 binding proteins. The only protein that was retrieved which bound with both the VPS35 and PRKN G4 structures was GNL1, a protein known to be a GTPase [57]. To our knowledge GNL1 has only been reported once, recently, in the literature to interact with RNA G-quadruplex [58]. However, this was a high-throughput experiment and the interaction had not been characterized. The sequence of GNL1 includes a Di-RG domain, a domain type previously described as being a G4 RNA binding domain [59–61]. The direct binding of GNL1 to both the VPS35 and PRKN RNA G4s was confirmed by EMSA. A higher affinity of GNL1 was observed for the WT G4 versions as compared to the respective G/A mutant ones. Interestingly, in both cases, the  $K_d$  for the WT sequence was

in the nanomolar range, supporting a strong affinity of GNL1 towards the RNAs. Another interesting fact was that the G/A mutation seems to affect the binding of GNL1 to PRKN more (35-fold versus 2.5-fold for VPS35). PRKN harbours a canonical G4, while VPS35 has a non-canonical one. Therefore, the difference between the WT and the G/A mutant versions of PRKN is more significant in terms of stability, and, most likely, a higher fraction of the G4 is truly folded. Once again, this hypothesis pointed out the importance of further investigating the consequences associated with the canonical and non-canonical G4 motifs that appear to be a particular feature of the transcriptome of the nervous cells.

This study reinforces the possible importance of G4 structures as potential *cis*-acting features in the post-transcriptional regulation of the nervous system, more specifically that in neurological diseases. A new RBP has also been identified, GNL1, that is able to bind to the 5'UTR RNA G4s of two genes associated with Parkinson's disease (VPS35 and PRKN).

## Material and methods

### Bioinformatic analyses

In order to retrieve genes associated with nervous system diseases, BRITe hierarchy files (br08402) were downloaded from the disease data-oriented entry points of the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [28]. The gene symbols from each line associated with the nervous system were conserved and then transformed into a list file. Next, the symbols were corrected using the Multi-symbol checker of the HUGO Gene Nomenclature Committee (HGNC) [62]. Then, the RNA.fa.gz files were downloaded from Refseq (NCBI, March 2018), and each gene from the HGNC list was associated with a Fasta transcript sequence. The Fasta files were then submitted to the G4 RNA Screener software using the default parameters, and all windows with a G4NN score >0.5 were considered as being positive [32]. Finally, pG4 windows were annotated by their transcript position (5'UTR, CDS and 3'UTR) using the RefSeq CDS annotation "ncbiRefSeqCds.txt.gz" (May 2018). The positive window densities were calculated with the mean length of the 5'UTR, CDS, and 3'UTR in the human transcriptome [63].

For the purpose of comparing the KEGG results to other databases, tab-separated values files were retrieved from the Human Protein Atlas for the enriched neuronal genes. They were then associated with a Fasta transcript file and submitted to the G4 RNA screener. Genes with a G4NN score >0.5 were considered as being positive.

### Design of the oligonucleotides

Oligonucleotides corresponding to the potential G4 sequences were designed to include both upstream and downstream wild type flanking sequences of up to 20 to 25 nucleotides (nts) in length. G/A mutants (several Gs substituted by As) were also synthesized for each WT G4 sequence (see in Tables S1 and S2 for sequence details). When required for *in vitro* transcription, a T7 promoter sequence (5'-TAATACGACTCACTATAG<sub>1-3</sub>

-3', where G<sub>1-3</sub> indicates a stretch of 1 to 3 consecutive guanines) was added to the sense strand. All oligonucleotides were purchased from Biobasic.

### *In vitro* transcription

In order to reconstitute the full-length WT G4 and G/A mutant sequences (which are 70–90 nts in size), DNA templates were synthesized via a PCR filling strategy using both forward and reverse primers. Then, an *in vitro* transcription was performed, as described previously [64]. Subsequently, the proteins were removed by phenol/chloroform extraction and the RNAs were purified through electrophoresis, using denaturing (8 M urea) 5% polyacrylamide gels (19:1). For the transcriptions of both pMA-RQ VPS35-4xS1m and pMA-RQ PRKN-4xS1m, the plasmids were digested with *Eco*R1 and *Hind*III prior to the transcription reaction.

### N-methyl mesoporphyrin (NMM) fluorescent assays

With the aim of probing G4 formation, fluorescence assays using NMM were performed as described previously [64]. Briefly, either the WT or a G/A mutant G4 RNA (200 pmol) was added to folding buffer (20 mM Li-cacodylate (pH 7.5) and 100 mM of either LiCl or KCl) in a final volume of 50  $\mu$ L. The reactions were then heated to 70°C and allowed to slowly cool for 1 h at room temperature. Next, 50  $\mu$ L of 2X *in-line* probing buffer (40 mM Li-cacodylate (pH 8.5), 40 mM MgCl<sub>2</sub> and 200 mM of either LiCl or KCl) were added, yielding a final volume of 100  $\mu$ L. The NMM ligand (1  $\mu$ L at a concentration of 0.5 mM; N-Methyl-Mesoporphyrin IX, NMM580, Frontier Scientific Inc., Logan, Utah) was then added and the reaction incubated for 30 min at room temperature while being protected from light. The fluorescence intensity was monitored using an Hitachi F-2500 fluorescence spectrophotometer with an excitation wavelength of 399 nm and the emission wavelength being measured between 500 nm and 650 nm in a 10 mm quartz cuvette. The fluorescence at 605 nm was used for quantification. All NMM assays were performed at least in duplicate.

### Dephosphorylation and radioactive labelling

The RNAs were dephosphorylated with the Antarctic Phosphatase kit (according to the manufacturer, New England Biolabs) and the resulting RNAs (10 pmol) were phosphorylated with [ $\gamma$ -<sup>32</sup>P]ATP (2  $\mu$ L; 6 000 Ci (222 TBq)/mmol in 50 mM Tricine (pH 7.6), PerkinElmer) using the T4 Polynucleotide Kinase (PNK) kit (according to the manufacturer, Promega). Then, the radioactive RNAs were purified through electrophoresis using denaturing (8 M urea) 5% polyacrylamide gels (19:1).

### *In-line* probing

In order to confirm G4 folding, *in-line* probing was performed as described previously [36,64]. The results were visualized using a Typhoon Trio imaging system (GE Healthcare), and the band quantification was performed using the SAFA semi-

automated software [65]. The hydrolysis ratio for each nucleotide position is represented by the band intensity in the presence of KCl divided by that in the presence of LiCl. All experiments were performed at least in duplicate.

### Cell cultures

SH-SY5Y cells (ATCC, CRL-2266) were cultivated in Dulbecco's modified Eagle medium/F12 with Glutamax (Gibco) supplemented with 10% foetal bovine serum (FBS; Wisent). Human Kidney Embryo 293T (HEK293T) cells (ATCC, CRL-3216) were cultivated in DMEM (Multicell) supplemented with 10% FBS. When the cells reached a confluence of 80–90% they were subcultured at dilutions of 1:2 and 1:20, respectively. All cultures were incubated at 37°C with a 5% CO<sub>2</sub> atmosphere.

### Molecular cloning

The PsiCHECK-2 plasmids used in the luciferase assays were prepared as follows. Puc57 plasmids containing either the full-length WT or G/A mutant 5'UTR G4 sequences of VPS35, LRRK2, SNCA and PRKN were purchased from Biobasics (see Table S1 for the full sequences). The 5'UTR sequences were codigested with *SpeI* and *SallI*, and were then ligated upstream of the Renilla luciferase in a modified version of the PsiCHECK-2 plasmid (Fig. S2) that had previously been digested with the same restriction enzymes. The insertion was then confirmed by PCR, and the plasmids were sequenced at the Plateforme de Séquençage of Université Laval.

The PsiCHECK-2 plasmids prepared previously were used as templates in PCR reactions in order to generate G4 regions containing *EcoRI* and *BamHI* restriction sites located in 5' and 3', respectively. The resulting DNA templates were codigested with *EcoRI* and *BamHI* and ligated into pMA-RQ plasmid containing the 4xS1m aptamer [43]. The insertion was confirmed by PCR, and all constructs were sequenced at the Plateforme de Séquençage of Université Laval.

### Luciferase assay

SH-SY5Y and HEK293 cells were subcultured at densities of 80 000 and 22 500 cells per cm<sup>2</sup> 24 h, respectively, before transfection. They were then transfected in triplicate with Lipofectamine 2000 as described by the manufacturer (Thermo Fisher Scientific) using 500 ng of pUC19 and 50 ng of PsiCHECK-2 in each well of 24 well plates. A day later, the medium was removed and the cells were washed once with PBS. The cells were then dislodged and suspended by scraping in the presence of 500 µL of PBS. Four hundred microlitres of the resulting suspension were then used to extract proteins, while the remaining 100 µL were used to extract total RNA. Both fractions were centrifuged at 4°C for 10 min at 2400 x g and the supernatants removed. The cell pellets for RNA isolation were frozen at –80°C until treated, while 100 µL of protein lysis buffer (Dual-luciferase reporter assay system, Promega) was added to the cell pellets destined for protein extraction. In order to lyse the cells, the solutions

were then incubated for 15 min at room temperature. Only a fraction (5 µL) of the resulting solution was then used for the luciferase assays as suggested by the manufacturer's protocol. Luminescence was then monitored on a Glomax 20/20 luminometer with an integration time of 5 sec. The ratios of the *Renilla/Firefly* luciferases were then calculated and presented as the means and standard deviation of 3 experiments.

RNAs were extracted with Trizol according to the manufacturer's protocol (Thermo Fisher Scientific). When required, the RNAs were quantified by digital RT-PCR at the Laboratoire de Génomique Fonctionnel de Université de Sherbrooke.

### G4 ligands

SH-SY5Y cells at a confluence of 70% were incubated for 24 h with either 10 µM of BRACO19 or 5 µM of Phen-DC3. Ultrapure water was used as negative control for BRACO19 and DMSO for Phen-DC3. After the 24 h were complete, the cells were washed once with 1X PBS and were then harvested by centrifugation for 5 min at 600 x g. The resulting supernatants were discarded and the cells were incubated with 20 µL of KCl-lysis buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 mM NaCl, 1 mM DTT, 0,1% Triton X-100, 10% glycerol and 1 mM MgCl<sub>2</sub>) for 30 min at 4°C. The cellular lysates were then centrifuged at 17 000 x g for 20 min at 4°C, and the resulting supernatants were kept. The protein concentrations were determined by the method of Bradford (BIO-RAD). Western blots were then performed as described previously [11] in order to monitor the expression of VPS35 and PRKN. Anti-PRKN (ab77924; Abcam), anti-VPS35 (ab226180; Abcam) and anti-β-Actin (A5441; Sigma-Aldrich) were used as primary antibodies, and both anti-Mouse (A-21057; Thermo Fisher Scientific) and anti-Rabbit (926-32213; Licor) were used as secondary antibodies. The densities of the resulting bands were determined using the Image Studio Lite software (Licor). Both the PRKN and VPS35 protein levels were adjusted with the β-Actin protein levels. RNAs were extracted with Trizol according to the manufacturer's protocol (Thermo Fisher Scientific). The RNAs were quantified by digital RT-PCR at the Laboratoire de Génomique Fonctionnel de Université de Sherbrooke and normalized to β-Actin mRNA. Proteins relative levels were then normalized to mRNA relative levels.

### Pull-down assays and mass spectrometry

Proteins able to bind PRKN and VPS35 G4 were isolated using a previously described strategy [43,58]. Briefly, cellular protein extracts were prepared from 4 confluent 15 cm dishes of untransfected SH-SY5Y cells. The cell culture dishes were placed on ice and washed once with ice-cold PBS (10 mL). The cells were then lysed in ice-cold KCl-lysis buffer (5 mL; 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 50 mM NaCl, 1 mM DTT, 0,1% Triton X-100, 10% glycerol, 2 mM MgCl<sub>2</sub>, Ribonucleoside Vanadyl complex RNase inhibitors (New England Biolabs) and Mini Complete Protease inhibitors (Roche)) for 30 min at 4°C. The lysate was then clarified by centrifugation at 17 000 x g for 20 min at 4°C. The supernatant was then recovered, divided into 2 fractions and

concentrated using Amicon Ultra-4 MWCO 3 000 centrifuge tubes (EMD Millipore) at 1 600 x g until the final volumes reached 0.5–1 mL. The protein concentrations of the cellular extracts were then determined using the method of Bradford (BIO-RAD). The extracts were then pre-cleared by the addition of 50  $\mu$ L of a 50% slurry of Streptavidin Sepharose High Performance beads (GE Healthcare) and incubated for 16 h at 4°C. At this point, the beads were discarded. Next, either the *in vitro* transcribed 4xS1m WT or 4xS1m G/A mutant G4 RNA (30  $\mu$ g) was dissolved in 140  $\mu$ L of KCl-lysis buffer without MgCl<sub>2</sub> or RNase inhibitors and then heated to 70°C prior to cooling over a period of 1 h until it reached room temperature. Subsequently, the solution was supplemented to a final concentrations of 1 mM MgCl<sub>2</sub> and 120 U of RNase inhibitors were added (Protein purification platform from Université de Sherbrooke). The RNA solutions were then mixed with 100  $\mu$ L of a pre-washed (with KCl-lysis buffer) Streptavidin Sepharose bead slurry and incubated for 2 h at 4°C. After this incubation, the beads were centrifuged at 20 x g for 1 min and the supernatant was discarded. The beads were then washed once with ice-cold KCl-lysis buffer. Next, the pre-cleared lysate (e.g. the one incubated 16 h at 4°C) was supplemented with 1.5  $\mu$ L of RNase inhibitor (Protein purification platform from Université de Sherbrooke), added to the prepared RNA-coupled Streptavidin Sepharose beads and incubated for 3 h at 4°C. Lastly, the beads were washed five times with 1 mL of KCl-wash buffer (50 mM Tris-HCl (pH 7.5), 100 mM KCl, 200 mM NaCl, 1 mM DTT, 0.1% Triton X-100, 10% glycerol, 5 mM MgCl<sub>2</sub>, Ribonucleoside Vanadyl complex RNase inhibitors (NEB) and Mini Complete Protease inhibitors (Roche)).

Samples were then prepared for mass spectrometry, following on-beads digestion and analysed by mass spectrometry, as previously described [66,67]. The Xcalibur software was used to acquire the data (Thermo Fisher Scientific). Quantification was performed using the MaxQuant version 1.5.2.8 software, and the Uniprot human database (10/04/2018, 74,811 entries) as described previously [67,68]. Proteins with a coverage >5% and a score of >15 were conserved for analysis. PRKN was analysed with the triplicate pull-downs, while VPS35 was analysed with 2 of the triplicate pull-downs because of the lower amount of protein present. The MS/MS counts were then analysed using SAINT (a probabilistic scoring of affinity purification) in order to compare the WT G4 and G/A mutant G4 conditions [45]. Proteins with a score higher than 0.9 and with a 10-fold or more enrichment were kept.

In order to confirm the proteins enrichment of GNL1 by western blot, 5% of the protein input was removed from the cellular lysate, dosed with DC reagent (BIO-RAD) and mixed with Laemmli buffer in order to obtain a final concentration of 50 mM Tris-HCl (pH 6.8), 1% SDS, 10% glycerol, 0.002% bromophenol blue and 2.5%  $\beta$ -mercaptoethanol. After the five washes of the proteins bound to the RNA-bead complexes, 50  $\mu$ L of Laemmli buffer (50 mM Tris-HCl (pH 6.8), 1% SDS, 10% glycerol, 0.002% bromophenol blue and 2.5%  $\beta$ -mercaptoethanol) were added to the beads to dissolve the proteins. An SDS-PAGE was then performed on 30  $\mu$ g of the protein input and 50% of each pull-down. Western blots

were performed using anti-GNL1 as primary antibody (ab245634; Abcam) and anti-rabbit as secondary antibody (926-32213; Licor).

### Protein purification

In order to produce recombinant protein, the GNL1 sequence was fused to an N-terminal glutathione-S-transferase (GST) gene in the plasmid pcDNA3.1 (-) and the resulting plasmid was then transfected into ten 15-cm dishes of 60% confluent SH-SY5Y cells. Two days post transfection, the cells were washed once with PBS and then were scraped in PBS. The cells were then harvested by centrifugation at 1 600 x g for 10 min, and the resulting supernatant discarded. The cell pellet was resuspended in 20 mL of lysis buffer (1X PBS, 1 mM EDTA, 250 mM NaCl, 10% glycerol, 1 mM DTT and 1 tablet of Mini Complete Protease inhibitors (Roche)). Triton X-100 (210  $\mu$ L) was then added and the solution was sonicated at 20% power for 20 cycles of 5 sec ON – 5 sec OFF using a Branson Digital Sonifier 450. The resulting solution was then incubated for 30 min at 4°C prior to being centrifuged at 17 000 x g for 10 min at 4°C. The supernatant was then carefully transferred into a new tube. In parallel, of glutathione Sepharose 4B beads (250  $\mu$ L, GE healthcare) were washed three times with lysis buffer. The beads were then added to the cellular lysate and the sample incubated for 3 h at 4°C. The sample was then centrifuged at 225 x g for 2 min at 4°C and the supernatant discarded. The beads were then washed twice with lysis buffer containing 1% Triton X-100, twice with cold PBS and once with wash buffer (50 mM Tris-HCl pH 8.0, 250 mM of NaCl and 0.1% Triton X-100). The GST-GNL1 was then eluted by incubating for 15 min with 250  $\mu$ L of elution buffer (10 mM reduced glutathione, 50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10% Glycerol and 0.1% Triton X-100). The elution step was repeated four times. The different fractions were then combined and dialysed for 16 h using a 3 mL Slide-A-Lyzer Dialysis Cassette (10 K MWCO; Thermo Fisher) with dialysis buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 2 mM DTT and 10% glycerol). Finally, the protein concentration was estimated using the DC protein assay (BIO-RAD), and the purity was evaluated by SDS-PAGE.

### Electrophoretic mobility shift assay (EMSA)

In order to confirm the binding of GST-GNL1 to the G4 structures of both the PRKN and the VPS35 5'UTRs, EMSA were performed with the same 5'-labelled transcripts as were used in the *in-line* probing experiments. For each assay, 500 CPM of labelled RNA were mixed with 1 nM of unlabelled transcript in folding buffer (20 mM Tris-HCl pH 8.0 and 100 mM KCl), and the samples were then incubated for 5 min at 70°C prior to slow cooling at room temperature over a period of 1 h. Next, for every 7.5  $\mu$ L of RNA solution of, 11.25  $\mu$ L of binding buffer (200 mM Tris-HCl pH 8.0, 900 mM KCl, 10 mM MgCl<sub>2</sub> and 50% glycerol), 0.94  $\mu$ L of RNase inhibitor (40 U/ $\mu$ L), 7.5  $\mu$ L of BSA (3 mg/ $\mu$ L), 7.5  $\mu$ L of yeast total RNA (2  $\mu$ g/ $\mu$ L) and 6.5  $\mu$ L of water were added.

The recombinant protein was then added at concentrations ranging from 0 to 1816 nM. The samples were then incubated for 20 min at room temperature and then electrophoresed through a 5% native polyacrylamide gel in TBE buffer at 20 V/cm for 1 h at 4°C. The shift percentages were evaluated using the formula «Shift/(Shift + No shift RNA)». The dissociation constants ( $K_d$ ) were calculated with the GraphPad software (Prism) using the ‘specific binding with hill slope nonlinear regression function’.

### Statistical analysis

Statistical analyses were performed using the GraphPad software (Prism). Depending on the experiments, either the One-sample Kolmogorov-Smirnov test or the Two-way ANOVA test was used (see the figure legends for a description). A p-value of less than 0.05 was considered statically significant. Proteins statistically enriched in the pull-down experiments were identified with SAINT [45]. \* represented a p-value < 0.05, \*\* a p-value < 0.01, \*\*\* a P-value < 0.001 and \*\*\*\* a p-value < 0.0001.

### Authors contributions

M.A.T and J.P.P designed all of the experiments. MAT, J.M.G and H.R.C performed the experiments. MAT, J.M.G, H.R.C and J.P.P analysed data. M.A.T and J.P.P wrote the manuscript.

### Acknowledgments

We thank D. Lévesque for his support with the MS instrument and for the label-free analysis. J.P.P. holds the Research Chair of the Université de Sherbrooke in RNA Structure and Genomics and is a member of the Centre de Recherche du CHUS.

### Disclosure statement

The authors declare no conflicts of interest for this work.

### Funding

This project was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) [155219-17] to JPP. MAT received student fellowships from the Canadian Institutes of Health Research (CIHR), the Fonds de Recherche Québec Santé (FRQS) and from the Faculty of medicine and health sciences of Université de Sherbrooke. J.M.G received a student fellowship from NSERC. H.C.R received a student fellowship from Fonds de Recherche Québec Nature et Technologie (FRQNT). The funders had no role in study design, data collection and analysis, decision to publish, nor in the preparation of the manuscript.

### ORCID

Jean-Michel Garant  <http://orcid.org/0000-0002-0559-5598>

### References

- [1] Rouleau S, Jodoin R, Garant J-M, et al. RNA G-quadruplexes as key motifs of the transcriptome. In: Seitz H, Stahl F, Walter J-G, editors. *Catalytically active nucleic acids*. Vol. 170. Cham: Springer International Publishing; 2017. p. 1–20.
- [2] Bolduc F, Garant J-M, Allard F, et al. Irregular G-quadruplexes found in the untranslated regions of human mRNAs influence translation. *J Biol Chem*. 2016;291:21751–21760.
- [3] Jodoin R, Bauer L, Garant J-M, et al. The folding of 5'-UTR human G-quadruplexes possessing a long central loop. *RNA N Y N*. 2014;20:1129–1141.
- [4] Xiao C-D, Shibata T, Yamamoto Y, et al. An intramolecular antiparallel G-quadruplex formed by human telomere RNA. *Chem Commun*. 2018;54:3944–3946.
- [5] Małgowska M. Overview of RNA G-quadruplex structures. *Acta Biochim Pol*. 2016;63:609–621.
- [6] Kim N. The interplay between G-quadruplex and transcription. *Curr Med Chem*. 2019;26:2898–2917.
- [7] Lerner LK, Sale JE. Replication of G quadruplex DNA. *Genes (Basel)*. 2019;10:95.
- [8] Neidle S, Parkinson GN. The structure of telomeric DNA. *Curr Opin Struct Biol*. 2003;13:275–283.
- [9] Beaudoin J-D, Perreault J-P. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res*. 2010;38:7022–7036.
- [10] Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*. 2012;40:4727–4741.
- [11] Jodoin R, Carrier JC, Rivard N, et al. G-quadruplex located in the 5'UTR of the BAG-1 mRNA affects both its cap-dependent and cap-independent translation through global secondary structure maintenance. *Nucleic Acids Res*. 2019;47:10247–10266.
- [12] Endoh T, Kawasaki Y, Sugimoto N. Suppression of gene expression by g-quadruplexes in open reading frames depends on G-quadruplex stability. *Angew Chem Int Ed*. 2013;52:5522–5526.
- [13] Yu C-H, Teulade-Fichou M-P, Olsthoorn RCL. Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res*. 2014;42:1887–1892.
- [14] Arora A, Suess B. An RNA G-quadruplex in the 3' UTR of the proto-oncogene PIM1 represses translation. *RNA Biol*. 2011;8:802–805.
- [15] Beaudoin J-D, Perreault J-P. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res*. 2013;41:5898–5911.
- [16] Rouleau SG, Garant J-M, Bolduc F, et al. G-quadruplexes influence pri-microRNA processing. *RNA Biol*. 2017;15:198–206.
- [17] Matsumura K, Kawasaki Y, Miyamoto M, et al. The novel G-quadruplex-containing long non-coding RNA GSEC antagonizes DHX36 and modulates colon cancer cell migration. *Oncogene*. 2017;36:1191–1199.
- [18] Guo JU, Bartel DP. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*. 2016;353:5371.
- [19] Mendoza O, Bourdoncle A, Boulé J-B, et al. G-quadruplexes and helicases. *Nucleic Acids Res*. 2016;44:1989–2006.
- [20] Masuzawa T, Oyoshi T. Roles of the RGG domain and RNA recognition motif of nucleolin in G-quadruplex stabilization. *ACS Omega*. 2020;5:5202–5208.
- [21] Pietras Z, Wojcik MA, Borowski LS, et al. Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nat Commun*. 2018;9:2558.
- [22] Sun Z-Y, Wang X-N, Cheng S-Q, et al. Developing novel G-quadruplex ligands: from interaction with nucleic acids to interfering with nucleic acid–protein interaction. *Molecules*. 2019;24:396.
- [23] Koukouraki P, Doxakis E. Constitutive translation of human  $\alpha$ -synuclein is mediated by the 5'-untranslated region. *Open Biol*. 2016;6:160022.
- [24] Ishiguro A, Kimura N, Watanabe Y, et al. TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells*. 2016;21:466–481.
- [25] Schludi MH, Edbauer D. Targeting RNA G-quadruplexes as new treatment strategy for C9orf72 ALS/FTD. *EMBO Mol Med*. 2018;10:4–6.
- [26] Imperatore JA, Then ML, McDougal KB, et al. Characterization of a G-quadruplex structure in pre-miRNA-1229 and in its

- Alzheimer's disease-associated variant rs2291418: implications for miRNA-1229 maturation. *Int J Mol Sci.* **2020**;21:767.
- [27] Melko M, Bardoni B. The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P. *Biochimie.* **2010**;92:919–926.
- [28] Kanehisa M, Sato Y, Kawashima M, et al. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2015**;44:D457–D462.
- [29] Garant J-M, Perreault J-P, Scott MS. G4RNA screener web server: user focused interface for RNA G-quadruplex prediction. *Biochimie.* **2018**;151:115–118.
- [30] Beaudoin J-D, Jodoin R, Perreault J-P. New scoring system to identify RNA G-quadruplex folding. *Nucleic Acids Res.* **2014**;42:1209–1223.
- [31] Bedrat A, Lacroix L, Mergny J-L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016**;44:1746–1759.
- [32] Garant J-M, Perreault J-P, Scott MS. Motif independent identification of potential RNA G-quadruplexes by G4RNA screener. *Bioinformatics.* **2017**;33:3532–3537.
- [33] Vannutelli A, Belhamiti S, Garant J-M, et al. Where are G-quadruplexes located in the human transcriptome? *NAR Genomics Bioinf.* **2020**;2.
- [34] Umar MI, Ji D, Chan C-Y, et al. G-quadruplex-based fluorescent turn-on ligands and aptamers: from development to applications. *Molecules.* **2019**;24:2416.
- [35] Nicoludis JM, Barrett SP, Mergny J-L, et al. Interaction of human telomeric DNA with N-methyl mesoporphyrin IX. *Nucleic Acids Res.* **2012**;40:5432–5447.
- [36] Beaudoin J-D, Jodoin R, Perreault J-P. In-line probing of RNA G-quadruplexes. *Methods.* **2013**;64:79–87.
- [37] Xicoy H, Wieringa B, Martens GJM. The SH-SY5Y cell line in Parkinson's disease research: a systematic review. *Mol Neurodegener.* **2017**;12:10.
- [38] Stephanenko AA, Dmitrenko VV. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene.* **2015**;569:182–190.
- [39] Burger AM, Dai F, Schultes CM, et al. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **2005**;65:1489–1496.
- [40] Chung WJ, Heddi B, Hamon F, et al. Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew Chem Int Ed.* **2014**;53:999–1002.
- [41] Yang SY, Lejault P, Chevrier S, et al. Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat Commun.* **2018**;9:1–11.
- [42] Zyner KG, Mulhearn DS, Adhikari S, et al. Genetic interactions of G-quadruplexes in humans. *eLife.* **2019**;8:e46793.
- [43] Leppek K, Stoecklin G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.* **2014**;42:e13.
- [44] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* **2008**;26:1367–1372.
- [45] Choi H, Larsen B, Lin Z-Y, et al. SAINT: probabilistic scoring of affinity purification - mass spectrometry data. *Nat Methods.* **2011**;8:70–73.
- [46] Lyons SM, Achorn C, Kedersha NL, et al. YB-1 regulates tRNA-induced Stress Granule formation but not translational repression. *Nucleic Acids Res.* **2016**;44:6949–6960.
- [47] Queiroz RML, Smith T, Villanueva E, et al. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nat Biotechnol.* **2019**;37(2):169–178.
- [48] Trendel J, Schwarzl T, Horos R, et al. The human RNA-binding proteome and its dynamics during translational arrest. *Cell.* **2019**;176(1–2):391–403.e19.
- [49] Liu EY, Cali CP, Lee EB. RNA metabolism in neurodegenerative disease. *Dis Model Mech.* **2017**;10:509–518.
- [50] Varshney D, Spiegel J, Zyner K, et al. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol.* **2020**;21:1–16.
- [51] Nguyen M, Wong YC, Ysselstein D, et al. Synaptic, mitochondrial, and lysosomal dysfunction in Parkinson's disease. *Trends Neurosci.* **2019**;42:140–149.
- [52] Sliter DA, Martinez J, Hao L, et al. Parkin and PINK1 mitigate STING-induced inflammation. *Nature.* **2018**;561:258.
- [53] Gatica D, Lahiri V, Klionsky DJ. Cargo recognition and degradation by selective autophagy. *Nat Cell Biol.* **2018**;20:233.
- [54] Tang F-L, Liu W, Hu J-X, et al. VPS35 deficiency or mutation causes dopaminergic neuronal loss by impairing mitochondrial fusion and function. *Cell Rep.* **2015**;12:1631–1643.
- [55] Asamitsu S, Takeuchi M, Ikenoshita S, et al. Perspectives for applying G-quadruplex structures in neurobiology and neuropharmacology. *Int J Mol Sci.* **2019**;20:2884.
- [56] Kelly TJ, Suzuki HI, Zamudio JR, et al. Sequestration of microRNA-mediated target repression by the Ago2-associated RNA-binding protein FAM120A. *RNA.* **2019**;25:1291–1297.
- [57] Boddapati N, Anbarasu K, Suryaraja R, et al. Subcellular distribution of the human putative nucleolar GTPase GNL1 is regulated by a novel arginine/lysine-rich domain and a GTP binding domain in a cell cycle-dependent manner. *J Mol Biol.* **2012**;416:346–366.
- [58] Bolduc F, Turcotte M-A, Perreault J-P. The small nuclear ribonucleoprotein polypeptide A (SNRPA) binds to the G-quadruplex of the BAG-1 5'UTR. *Biochimie.* **2020**;176:122–127.
- [59] Thandapani P, O'Connor TR, Bailey TL, et al. Defining the RGG/RG motif. *Mol Cell.* **2013**;50:613–623.
- [60] Ozdilek BA, Thompson VF, Ahmed NS, et al. Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* **2017**;45:7984–7996.
- [61] Huang Z-L, Dai J, Luo W-H, et al. Identification of G-quadruplex-binding protein from the exploration of RGG motif/G-quadruplex interactions. *J Am Chem Soc.* **2018**;140:17945–17955.
- [62] Braschi B, Denny P, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **2019**;47:D786–D792.
- [63] Piovesan A, Caracausi M, Antonaros F, et al. GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database.* **2016**;2016:baw153.
- [64] Jodoin R, Perreault J-P. G-quadruplexes formation in the 5'UTRs of mRNAs associated with colorectal cancer pathways. *PLoS One.* **2018**;13:e0208363–e0208363.
- [65] Laederach A, Das R, Vicens Q, et al. Semi-automated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nat Protoc.* **2008**;3:1395–1401.
- [66] Boisvert F-M, Lam YW, Lamont D, et al. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol Cell Proteomics MCP.* **2010**;9:457–470.
- [67] Telekawa C, Boisvert F-M, Bachand F. Proteomic profiling and functional characterization of post-translational modifications of the fission yeast RNA exosome. *Nucleic Acids Res.* **2018**;46:11169–11183.
- [68] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc.* **2016**;11:2301–2319.