

# SCIENTIFIC REPORTS

OPEN

## The *GBAP1* pseudogene acts as a ceRNA for the glucocerebrosidase gene *GBA* by sponging miR-22-3p

Letizia Straniero<sup>1</sup>, Valeria Rimoldi<sup>2</sup>, Maura Samarani<sup>3</sup>, Stefano Goldwurm<sup>4</sup>, Alessio Di Fonzo<sup>5</sup>, Rejko Krüger<sup>6</sup>, Michela Deleidi<sup>7</sup>, Massimo Aureli<sup>3</sup>, Giulia Soldà<sup>1,2</sup>, Stefano Duga<sup>1,2</sup> & Rosanna Asselta<sup>1,2</sup>

Mutations in the *GBA* gene, encoding lysosomal glucocerebrosidase, represent the major predisposing factor for Parkinson's disease (PD), and modulation of the glucocerebrosidase activity is an emerging PD therapy. However, little is known about mechanisms regulating *GBA* expression. We explored the existence of a regulatory network involving *GBA*, its expressed pseudogene *GBAP1*, and microRNAs. The high level of sequence identity between *GBA* and *GBAP1* makes the pseudogene a promising competing-endogenous RNA (ceRNA), functioning as a microRNA sponge. After selecting microRNAs potentially targeting both transcripts, we demonstrated that miR-22-3p binds to and down-regulates *GBA* and *GBAP1*, and decreases their endogenous mRNA levels up to 70%. Moreover, over-expression of *GBAP1* 3'-untranslated region was able to sequester miR-22-3p, thus increasing *GBA* mRNA and glucocerebrosidase levels. The characterization of *GBAP1* splicing identified multiple out-of-frame isoforms down-regulated by the nonsense-mediated mRNA decay, suggesting that *GBAP1* levels and, accordingly, its ceRNA effect, are significantly modulated by this degradation process. Using skin-derived induced pluripotent stem cells of PD patients with *GBA* mutations and controls, we observed a significant *GBA* up-regulation during dopaminergic differentiation, paralleled by down-regulation of miR-22-3p. Our results describe the first microRNA controlling *GBA* and suggest that the *GBAP1* non-coding RNA functions as a *GBA* ceRNA.

The glucocerebrosidase gene (*GBA*) encodes for the enzyme glucocerebrosidase (GCase), which catalyzes the hydrolysis of the membrane glucosylceramide (GlcCer) to ceramide and glucose. GCase is mainly a lysosomal enzyme and only partly associated with the outer surface of the cell membrane<sup>1</sup>. GCase deficiency leads to the accumulation of the substrate, responsible for the multi-organ clinical manifestations of Gaucher's disease (MIM #606463)<sup>2</sup>, one of the most common lysosomal storage disorders<sup>3</sup>. While biallelic mutations in *GBA* are responsible for Gaucher's disease, heterozygous *GBA* variants have been repeatedly associated with susceptibility to Parkinson's Disease (PD)<sup>4,5</sup>. Importantly, Gaucher's and PDs have been connected due to the clinical observation of parkinsonism and Lewy Bodies (LB) pathology in a fraction of patients with Gaucher's disease<sup>6</sup>. Compared with the general population, patients with the milder form of Gaucher's disease (type 1) have a 20-fold increased lifetime risk of developing parkinsonism<sup>7</sup>, whereas the odds ratio for any *GBA* mutation in PD patients compared to controls was greater than 5 in a multi-center analysis including more than 5000 cases and 4000 controls<sup>8</sup>. Several studies confirmed that *GBA* mutations, in particular the two most common ones (p.N370S and p.L444P), are more frequent in PD patients than in healthy controls, demonstrating that genetic lesions in this gene are a

<sup>1</sup>Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini, Pieve Emanuele, Milan, 20090, Italy. <sup>2</sup>Humanitas Clinical and Research Center, Via Manzoni 56, Rozzano, Milan, 20089, Italy. <sup>3</sup>Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, Milano, Italy. <sup>4</sup>Parkinson Institute, ASST "Gaetano Pini-CTO", Milan, Italy. <sup>5</sup>IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Center, Neuroscience Section, Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. <sup>6</sup>Clinical and Experimental Neuroscience, Luxembourg Center for Systems Biomedicine (LCSB), University of Luxembourg and Centre Hospitalier de Luxembourg (CHL), Esch-sur-Alzette, Luxembourg. <sup>7</sup>German Centre for Neurodegenerative Diseases (DZNE) Tübingen within the Helmholtz Association, Tübingen, Germany; Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany. Correspondence and requests for materials should be addressed to G.S. (email: [giulia.solda@hunimed.eu](mailto:giulia.solda@hunimed.eu))

common risk factor for the disease<sup>9,10</sup>. Recently, we proved the strong relationship between *GBA* mutations and PD progression and survival<sup>11</sup>.

Despite many efforts, the mechanism underlying the relation between *GBA* mutations and the development of PD remains unclear. There are studies supporting a gain-of-function effect of the mutated protein (promoting  $\alpha$ -synuclein aggregation), as well as others supporting a loss-of-function mechanism (leading to substrate accumulation, and hence affecting  $\alpha$ -synuclein processing and clearance)<sup>12</sup>. Widespread deficiency of GCase activity has been demonstrated in the brains of PD patients carrying *GBA* mutations, but it is also significant that PD patients without *GBA* mutations were shown to exhibit deficiency of GCase in the *substantia nigra* (SN) as well as in blood<sup>13,14</sup>. Moreover, neurons and brains of PD patients showed accumulation of GlcCer that directly influences the abnormal lysosomal storage of  $\alpha$ -synuclein oligomers, thus resulting in a further inhibition of the GCase activity. These findings suggested that the bi-directional effect of GlcCer and  $\alpha$ -synuclein accumulation forms a positive feedback loop that may lead to a self-propagating disease<sup>15</sup>. Recent data also linked GCase impairment to the cell-to-cell propagation of  $\alpha$ -synuclein aggregates<sup>16</sup>. Based on the above-mentioned evidence, it is plausible that dysregulated *GBA* levels could represent a common feature in PD, whereas loss-of-function *GBA* mutations could constitute the specific trigger responsible for PD development in the *GBA*-associated disease.

Dysregulation of *GBA* expression may, in theory, be due to altered epigenetic, transcriptional, and/or post-transcriptional regulatory mechanisms. In particular, RNA-based networks, characterized by interactions between a specific mRNA, microRNAs (miRNAs), and competitive endogenous RNAs (ceRNAs), are emerging as post-transcriptional regulators of gene expression<sup>17</sup>. Moreover, accumulating evidence points to deregulation of noncoding RNAs as an important and largely unexplored regulatory layer in human neurodegenerative disorders, such as PD<sup>18,19</sup>.

MiRNAs are ~20-nucleotide-long regulatory RNAs that act as post-transcriptional regulators of gene expression by repressing target mRNAs translation and/or by inducing mRNA degradation. About 2000 miRNAs have been experimentally validated in humans and many more have been predicted bioinformatically, making them a major class of regulators<sup>20</sup>. Each miRNA might inhibit the expression of multiple target mRNAs, whose recognition is based on imperfect complementary binding between miRNAs and their target sites, usually located within the 3' untranslated region (3'UTR)<sup>21</sup>. Recently ceRNAs were described as a novel category of regulatory RNAs: these transcripts compete with mRNAs for miRNAs, acting as molecular "sponges" and thus influencing mRNA levels<sup>17</sup>. Pseudogenes are the best ceRNA candidates, since they have a high-sequence identity with the ancestral gene (and, consequently, they could be targets of the same miRNAs), they can be transcribed, but usually they have lost the ability to generate a functional protein product<sup>22</sup>. Interestingly, a highly-homologous (96% sequence identity) expressed *GBA* pseudogene (*GBAP1*) is located 16 kb downstream of the functional gene<sup>23,24</sup>. *GBAP1* originated from a recent duplication event that occurred no more than 40 million years ago, prior to the divergence of the Great Apes and Old World monkeys, and also involved the metaxin (*MTX*) gene<sup>25,26</sup>.

With the aim to better understand *GBA* expression regulation at the post-transcriptional level, we explored the possible existence of a ceRNA-based network involving *GBA* and *GBAP1*. Here, we demonstrated that *GBAP1* may function as a ceRNA to regulate *GBA* expression by sponging miR-22-3p, thus revealing a novel regulatory circuit that can play a role in the pathogenesis of PD.

## Results

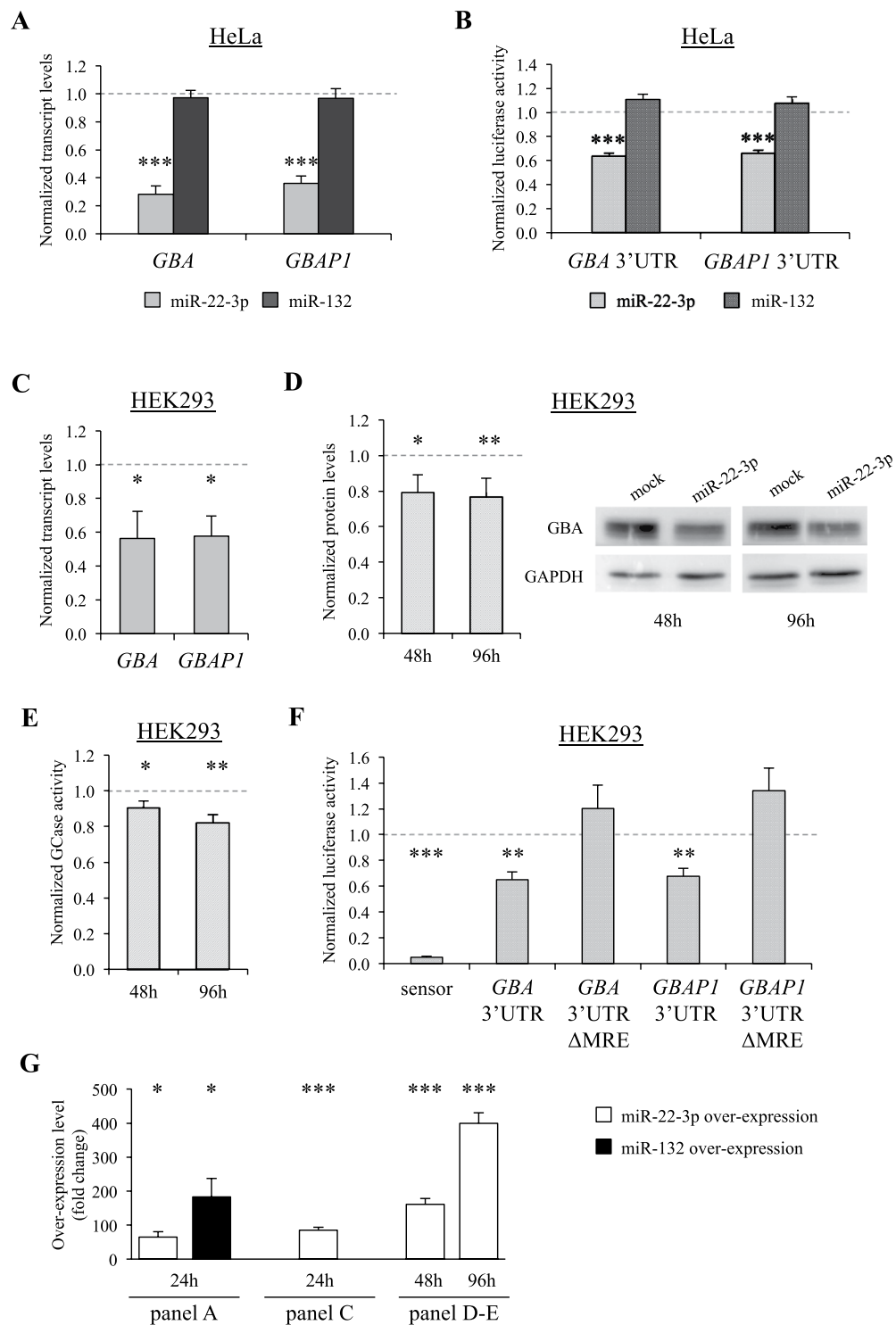
**MiR-22-3p targets *GBA* and *GBAP1*.** Since there is no information on miRNAs modulating *GBA* expression, we searched bioinformatically for miRNAs potentially targeting both *GBA* and its pseudogene. Predictions were performed using eight sources of software; candidate miRNA selection was performed by prioritizing miRNAs: i) predicted by at least five algorithms; ii) containing at least 7-nt perfect seed match with *GBA* and *GBAP1* 3'UTRs; iii) known to be expressed in the brain and previously implicated in neurodegenerative diseases. These filtering steps allowed the selection of three candidate miRNAs: miR-22-3p, miR-132, and miR-212. For functional validation, we prioritized miR-22-3p and miR-132, since they were expressed at a higher level in both the cerebellum and frontal cortex (Supplementary Table 1).

To verify that *GBA/GBAP1* can be targets of miR-22-3p and/or miR-132, we cloned both miRNA precursors in a suitable expression vector, and over-expressed them in HeLa cells for 24 hours. The results of these experiments showed that miR-22-3p over-expression can significantly reduce *GBA* and *GBAP1* endogenous mRNA levels (up to 72%;  $P < 0.0003$ ). Conversely, no *GBA* modulation was detected after miR-132 over-expression (Fig. 1A).

To confirm these results, the 3'UTRs of *GBA* and *GBAP1* were cloned downstream of the luciferase gene in the psiCHECK2 vector. These UTRs differ for only 6 nucleotides, none of them mapping in the predicted binding sites for miR-22-3p and miR-132. We cotransfected in HeLa cells each of these reporter plasmids together with the vector expressing either the miR-22-3p or miR-132 precursor. The results of transfection experiments substantially confirmed previous observations, *i.e.* miR-22-3p was able to target both *GBA* and *GBAP1* UTRs (37% and 34% reduction, respectively;  $P < 0.0001$ ). Conversely, miR-132 did not affect the expression of the reporter gene (Fig. 1B), and was hence not further investigated.

To better unravel the functional impact of miR-22-3p on the expression of *GBA/GBAP1*, we decided to study miR-22-3p/*GBA/GBAP1* expression profiles in 11 cell lines. Real-time reverse-transcription (RT)-PCRs evidenced a ubiquitous expression of *GBA* in the analyzed lines, with highest levels present in HeLa and glioblastoma cells, and lowest levels in HepG2 cells. *GBAP1* was present in all cell lines, though at lower levels than *GBA* (from 186 to 1.8 times less) (Supplementary Figure 1A). MiR-22-3p showed a nearly ubiquitous expression profile, with highest levels in HepG2 and glioblastoma cells, and lowest levels in HEK293 (Supplementary Figure 1B).

Based on these expression profiles, we decided to repeat miR-22-3p over-expression experiments in HEK293 cells. Results were comparable to those observed in the HeLa cell line, with *GBA* and *GBAP1* endogenous mRNA levels significantly decreased, after 24 hours, up to 44% ( $P < 0.05$ ; Fig. 1C). We then confirmed the effects of



**Figure 1.** MiR-22-3p targets *GBA* and *GBAP1*. (A) Endogenous expression levels of *GBA* and *GBAP1* after pre-miR-22-3p or pre-miR-132 over-expression in HeLa cells. Cells were collected 24 hours after transfection and total RNA extracted. Expression levels, measured by real-time RT-PCRs, are shown as normalized rescaled values, setting as 1 the value measured in cell transfected with an empty vector (psiUX, mock). (B) Luciferase reporter assays of *GBA* or the *GBAP1* 3'UTR after pre-miR-22-3p or pre-miR-132 over-expression in HeLa cells. 48 hours after transfection, cells were collected and protein lysates prepared for reporter assays. Renilla luciferase activity was normalized against the firefly luciferase activity, setting as 1 the value measured in cells cotransfected with an empty vector (psiCHECK2, no miRNA overexpression). (C–E) Effect of pre-miR-22-3p over-expression in HEK293 cells. Panel C shows the effect on the endogenous *GBA* and *GBAP1* transcripts, measured by real-time RT-PCR 24 hours after transfection. Panel D shows the reduction of *GBA* protein, as assessed by Western blot analysis, 48 or 96 hours after transfection. A representative blot (right) and the densitometric analysis of three independent experiments (left) are shown. Panel E reports the effect on the

endogenous GBA-specific GCase activity. In all cases, the value measured in cells cotransfected with an empty vector (psiUX, no miRNA over-expression) was set as 1. (F) Luciferase reporter assays of *GBA* or *GBAP1* 3'UTRs, with or without the putative miRNA recognition element ( $\Delta$ MRE), after miR-22-3p over-expression in HEK293 cells. Cells were collected 48 hours after transfection and lysates prepared for reporter assays. Renilla luciferase activity was normalized against the firefly luciferase activity, setting as 1 the value measured in cells cotransfected with an empty vector (psiCHECK2, no miRNA over-expression). The miR-22-3p sensor<sup>27</sup> served as positive control. (G) MiR-22-3p/miR-132 fold increase reached in each over-expression experiment (detailed below histograms). Error bars represent means  $\pm$ SEM of 3 independent biological replicates, each performed at least in triplicate. In all panels, the reference value, set as 1, is indicated by a dotted line. Significance levels of t-tests are shown. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.005$ .

miR-22-3p over-expression also at the protein level. Western blot analysis showed a  $\sim$ 20% reduction in the GBA protein level at both 48 hours and 96 hours after transfection (Fig. 1D). Measurements of endogenous GCase activity demonstrated a 10% and 18% down-regulation after 48 hours and 96 hours of transfection, respectively ( $P = 0.012$  and  $P < 0.003$ ; Fig. 1E).

Finally, the specific binding of miR-22-3p to *GBA* and *GBAP1* 3'UTRs was demonstrated by deleting the miR-22-3p putative miRNA responsive element ( $\Delta$ MRE) in the reporter constructs containing the relevant UTR, and subsequently cotransfecting each mutagenized plasmid together with the miR-22-3p expressing one. In these experiments, a luciferase construct containing miR-22-3p antisense sequences (miR-22-3p sensor) was used as a positive control<sup>27</sup>. Our data showed that miR-22-3p responsiveness strictly depends on the presence of the predicted responsive element in the 3'UTR, since its deletion completely abolishes the miRNA-mediated regulation (Fig. 1F). As expected, the level of luciferase activity in the miR-22-3p sensor control dramatically dropped (95% reduction).

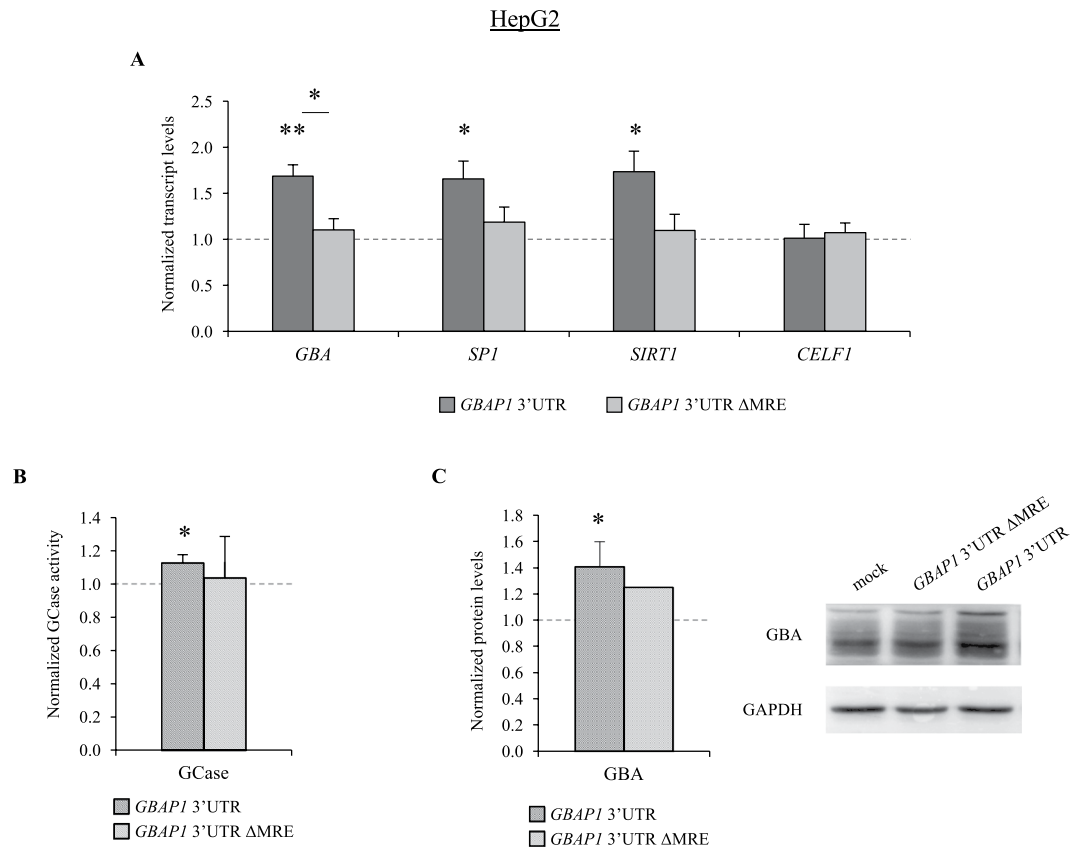
***GBAP1* acts as a ceRNA titrating miR-22-3p and up-regulating *GBA*.** We first verified the co-expression of *GBA/GBAP1/miR-22-3p* in a broad range of samples (20 human tissues as well as 24 different cerebral regions). The three transcripts were all ubiquitously expressed (Supplementary Figure 2A,B). In particular, *GBA* showed minimal expression in the skeletal muscle and the highest level in the medial temporal cortex (17 fold the skeletal muscle). *GBAP1* expression levels weakly correlated with those of *GBA* (Pearson's correlation coefficient of 0.53,  $P < 0.0013$ , Supplementary Figure 2C), in agreement with the possible ceRNA role of *GBAP1*. Interestingly, *GBAP1* highest expression levels were registered in the brain, where the disproportion between the gene and pseudogene levels is one of the lowest (ratio 1:15). Concerning miR-22-3p, although no consistent anti-correlation was found across the analyzed tissues, the highest miR-22-3p levels were detected in the tissues with lowest *GBA/GBAP1* expression (Supplementary Figure 2).

These results prompted us to verify if altered levels of *GBAP1* could indeed modify the expression of *GBA*. First, as a proof of concept, we over-expressed both the 3'UTR of *GBAP1* and the miR-22-3p hairpin in HEK293 cells. Concurrently, the over-expression experiment was conducted using as sponge the 3'UTR of *GBAP1* without the miR-22-3p responsive element. We showed that *GBAP1* 3'UTR over-expression causes a significant increase in the levels of endogenous *GBA* mRNA only in the presence of the miR-22-3p binding site (1.72 fold;  $P = 0.019$ ; Supplementary Figure 3A). We also evaluated the ceRNA effect at the protein level, by measuring the GCase activity upon miR-22-3p and *GBAP1* 3'UTR over-expression. Our data confirmed that the *GBAP1* 3'UTR, containing the miR-22-3p binding site, causes a significant increase of GCase activity (1.11 fold;  $P = 0.013$ ) (Supplementary Figure 3B).

Second, considering the high levels of miR-22-3p measured in HepG2 cells (16-fold the levels measured in HEK293; Supplementary Figure 1), we over-expressed in this cell line the 3'UTR of *GBAP1* alone (with or without the miR-22-3p responsive element) and measured its effect on endogenous *GBA*. We observed a significant increase in the levels of endogenous *GBA* mRNA, once again only in the presence of the miR-22-3p binding site (1.68 fold;  $P = 0.0016$ ; Fig. 2A). The *GBAP1* ceRNA effect through miR-22-3p sponging was confirmed by measuring the expression levels of known miR-22-3p targets, *i.e.* the *SP1* and *SIRT1* genes<sup>28,29</sup>, which both resulted in up-regulation of  $\sim$ 1.7 fold ( $P < 0.015$ ). Conversely, no up-regulation was observed for the *CELF1* transcript (Fig. 2A), which does not contain any miR-22-3p responsive element. These results were corroborated by the measurements of GCase activity and GBA protein levels in HepG2 cells under the same experimental conditions (GCase activity: 1.13 fold increase,  $P = 0.049$ ; GBA protein: 1.40 fold increase,  $P = 0.020$ ) (Fig. 2B). Finally, a similar overexpression experiment was repeated using as ceRNA the miR-22-3p sensor, which, in principle, represents the "perfect" miRNA sponge. As expected, we observed an up-regulation of all miR-22-3p targets (Supplementary Figure 4).

**The *GBAP1* ceRNA effect could be modulated by the nonsense-mediated mRNA decay (NMD) pathway.** To better unravel the reciprocal regulation of the couple *GBA/GBAP1*, we decided to comprehensively study *GBA* and *GBAP1* alternative splicing patterns and the possible regulation of expression of these two genes operated by NMD<sup>30,31</sup>.

To capture the vast majority of all possible splicing events, long-range RT-PCR assays were designed to completely cover both genes (Fig. 3A). The specific amplification of either *GBA* or *GBAP1* in each assay was assured by anchoring one primer to exon 9, in correspondence of the pseudogene-specific 55-bp deletion. RT-PCR assays were performed on RNA extracted from HepG2 cells treated or not with the NMD inhibitor cycloheximide. This analysis allowed the identification of multiple alternatively-spliced isoforms for *GBAP1*; conversely, *GBA* did not show any detectable alternative isoform (Fig. 3A). Notably, the heterogeneity of the splicing pattern of

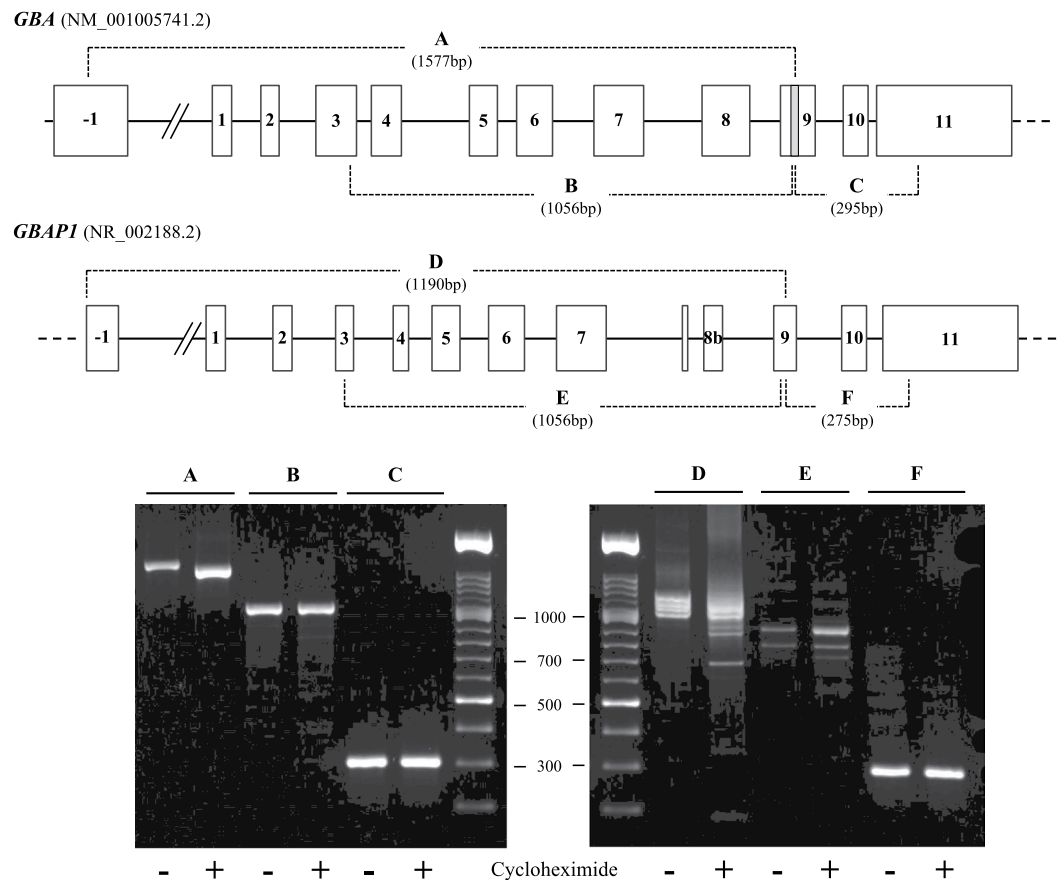
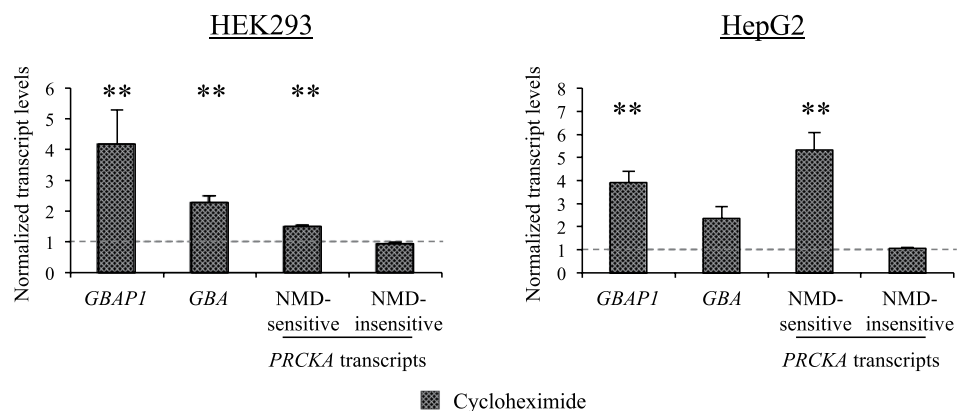


**Figure 2.** *GBAP1* acts as a ceRNA titrating miR-22-3p and up-regulating *GBA*. (A) Effect of *GBAP1* 3'UTR (with or without the miR-22-3p recognition element,  $\Delta$ MRE) over-expression on the endogenous transcript levels of indicated miR-22-3p targets in HepG2 cells. 24 hours after transfections, cells were collected for extracting total RNA for measurements by semi-quantitative real-time RT-PCRs of: i) *GBA*; ii) *SPI* (Sp1 Transcription Factor; known miR-22-3p target, positive control)<sup>28</sup>; iii) *SIRT1* (Sirtuin 1; known miR-22-3p target, positive control)<sup>29</sup>; and iv) *CELF1* (CUGBP, Elav-Like Family Member 1; negative control). The value measured in cells transfected with an empty vector (psiCHECK2, mock) was set as 1. (B) Effect of *GBAP1* 3'UTR (wild type or  $\Delta$ MRE) over-expression on GCase activity, measured 96 hours after transfections. (C) Effect of *GBAP1* 3'UTR (wild type or  $\Delta$ MRE) over-expression on GBA protein level, measured by Western blot 96 hours after transfections. A representative blot (right) and the densitometric analysis (left) are shown. Error bars represent: means +SEM of 3 (A) or 4 (B) independent biological replicates, each performed at least in triplicate; means +SD of 3 independent biological replicates (C, *GBAP1* 3'UTR). In all panels, the reference value, set as 1, is indicated by a dotted line. Significance levels of t-tests are shown. \* $P < 0.05$ ; \*\* $P < 0.01$ .

*GBAP1* increased after cycloheximide treatment, suggesting that multiple pseudogene splicing isoforms may be modulated by NMD. A tentative reconstruction of the main splicing variants of *GBAP1* was performed by a combination of isoform-specific semi-nested RT-PCRs and DNA sequencing, highlighting the presence of multiple transcripts containing a premature termination codon (Supplementary Figure 5).

The global effect of NMD degradation on *GBA* and *GBAP1* levels was also investigated by semi-quantitative real-time RT-PCR. This analysis showed a significant increase in the expression level of *GBAP1* in treated cells (4.18 and 3.92 fold in HEK293 and HepG2 cells,  $P = 0.045$  and  $P = 0.0034$ , respectively), confirming that this pseudogene is down-regulated by NMD (Fig. 3B). Also *GBA* transcripts were up-regulated upon NMD inhibition (2.28 and 2.35 fold in HEK293 and HepG2 cells), a rather unexpected result given the lack of out-of-frame *GBA* isoforms in our preliminary analysis. However, these results suit the hypothesis that *GBAP1* levels may influence *GBA* expression through a ceRNA effect. As control, in-frame and out-of-frame *PRKCA* isoforms, known to be insensitive/sensitive to the NMD blockage<sup>32</sup>, were also analyzed and yielded the expected results (Fig. 3B).

***GBA*, *GBAP1*, and miR-22-3p are expressed in induced pluripotent stem cells (iPSCs)-derived neuronal cells.** To be relevant for the molecular pathogenesis of PD, the *GBA/GBAP1/miR-22-3p* network should work in tissues affected by the disease process, e.g. dopaminergic (DA) neurons. We thus verified the expression of *GBA*, *GBAP1*, and miR-22-3p in iPSCs and iPSC-derived neuronal cells (after 35 days of differentiation). Semi-quantitative real-time RT-PCR assays were performed on total RNA extracted from iPSCs/neurons derived from fibroblasts of six healthy controls and four PD patients (all carrying *GBA* mutations).

**A****B**

**Figure 3.** *GBAPI* codes for multiple alternatively-spliced isoforms and is modulated by NMD. **(A)** Analysis of *GBA* and *GBAPI* splicing patterns. In the upper part of the panel, a schematic representation of *GBA* (reference sequence: NM\_001005741.2) and *GBAPI* (reference sequence: NR\_002188.2) genes is reported. Exons are indicated by boxes, introns by lines. The 55-bp-long sequence characterizing *GBA* exon 9 is specified by a grey rectangle. The scheme is approximately to scale. The overlapping fragments amplified by RT-PCRs to analyze the *GBA* and *GBAPI* splicing patterns are indicated by dashed lines and a letter. In the lower part of the panel, the electrophoretic analysis (agarose gels 2%) of RT-PCR amplicons is shown. RT-PCRs were performed on RNA extracted from HepG2 cells treated (+) or untreated (-) with the NMD inhibitor cycloheximide. On the top of each gel, letters indicate the relevant RT-PCR amplicons. **(B)** Demonstration of the NMD-mediated degradation of *GBAPI* transcripts. The two panel shows expression levels of *GBAPI* and *GBA* isoforms in HEK293 and HepG2 cells, untreated or treated for 8 hours with cycloheximide. Expression levels of endogenous *GBAPI/GBA* isoforms were measured by semi-quantitative real-time RT-PCRs. Results are presented as normalized rescaled values, setting as 1 the value of the untreated samples (dotted line). The expression level of the Connexin 43 or 32 transcripts, known to be insensitive to NMD, were used in the normalization step.

RT-PCRs performed on out-of-frame and in-frame *PRKCA* isoforms, known to be respectively sensitive and insensitive to the NMD blockage<sup>32</sup>, represent the positive and negative control. Error bars represent means  $\pm$  SEM of 3 independent biological replicates, each performed at least in triplicate. Significance levels of t-tests are shown. \* $P < 0.05$ ; \*\* $P < 0.01$ .

All three players of the regulatory circuit were expressed both in iPSCs and iPSC-derived neurons, respectively. The process of differentiation towards neurons is accompanied by a significant up-regulation of *GBA* (8 fold in controls,  $P = 0.024$ ; 3 fold in patients,  $P = 0.029$ ) and by a parallel increase in expression levels of *GBAPI* (Fig. 4A and B). In addition, we detected a trend for down-regulation of the *GBA* transcript in PD patients with respect to controls in DA neurons (0.54 fold,  $P = 0.057$ ). Finally, consistent with the observed up-regulation of *GBA/GBAPI* during neuronal differentiation, we detected lower expression levels of miR-22-3p in DA neurons with respect to their precursors (0.39 fold in controls; 0.18 fold in patients) (Fig. 4C).

## Discussion

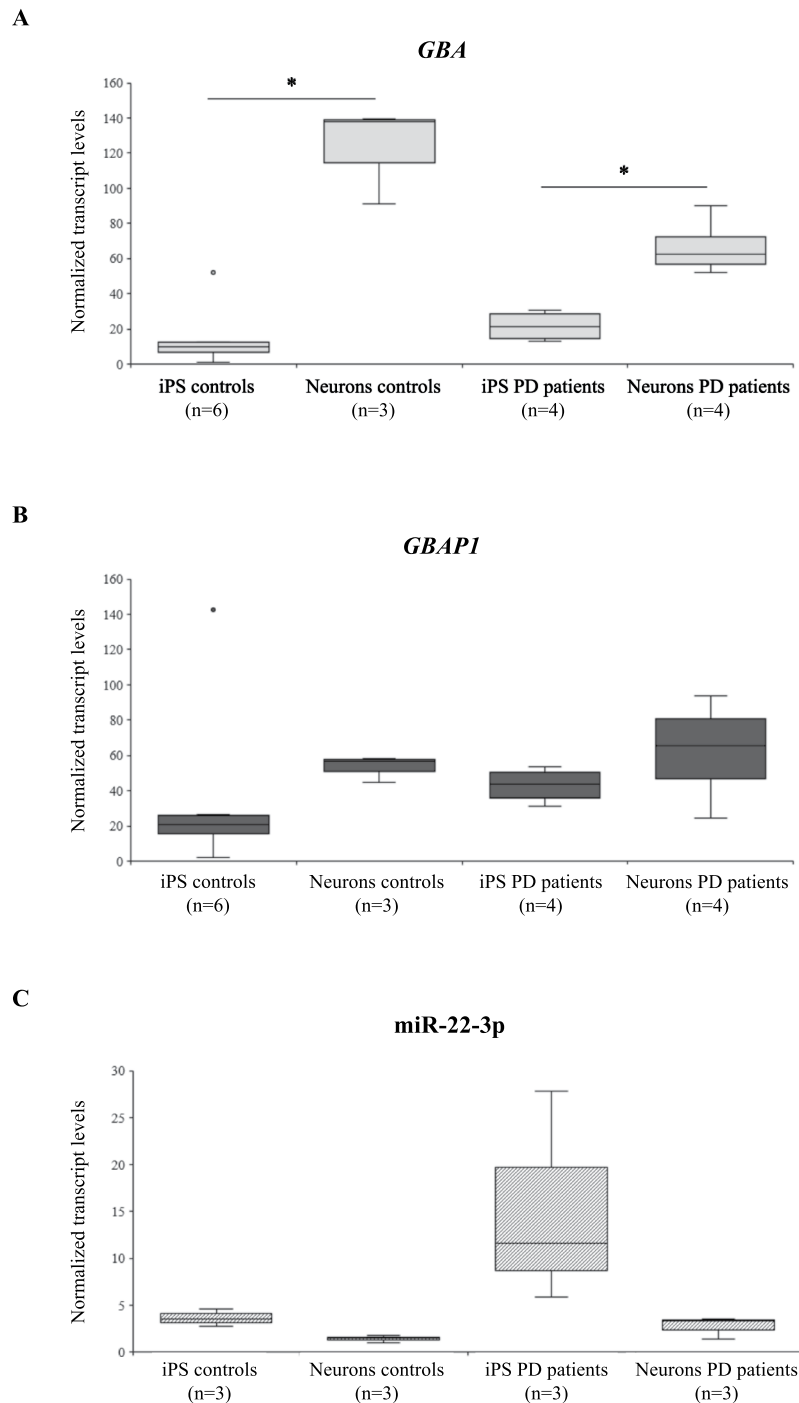
Despite substantial efforts over the past few years to understand the role of long non-coding RNAs (lncRNAs) in health and disease, only some of them have been investigated for their biological function<sup>33</sup>. One promising, although debated, idea assigning to lncRNAs a generalized function is the “ceRNA hypothesis”, based on the fact that specific RNAs can limit miRNA activity through sequestration, thus up-regulating the expression of miRNA target genes<sup>17</sup>. In particular, two classes of lncRNAs are increasingly recognized as main ceRNA contributors, *i.e.* circular RNAs and pseudogene-derived transcripts<sup>34</sup>. Indeed, transcribed pseudogenes, mostly deriving from duplication events, are considered optimal ceRNA candidates, as they share miRNA-binding sites with the ancestral genes<sup>17,34</sup>. To date, a number of pseudogenes have been experimentally demonstrated to act as ceRNAs, including: *PTENP1* and *KRAS1P22*, *OCT4-pg4*<sup>35</sup>, *BRAFPI36*, and *CYP4Z2P37*. In this study, we describe a novel ceRNA-based network involving *GBA*, its pseudogene *GBAPI*, and miR-22-3p (Fig. 5).

The molecular evolution, expression pattern, and mechanisms of transcriptional regulation of *GBA* have been previously investigated<sup>38–40</sup>, mainly because of its direct link with Gaucher’s disease. On the other hand, the few data available on *GBA* post-transcriptional regulation principally stem from a screening aimed to identify miRNAs regulating the GCase activity in p.N370S homozygous Gaucher fibroblasts<sup>41</sup>. This screening involved 875 miRNAs and evidenced at least three candidates (miR-127-5p, miR-16-5p, and miR-195-5p), exhibiting a Z-score of at least  $+/-2$ , with substantial consequences on the GCase activity. However, in all cases, the miRNA effect did not seem to be mediated by a direct binding of the miRNA to *GBA* transcripts; rather, miRNAs acted either on the LIMP-2 receptor, which is involved in the trafficking of GCase from the endoplasmic reticulum to the lysosome, or on the expression levels of known modifiers of the GCase activity<sup>41</sup>. Hence, our work identifies miR-22-3p as the first miRNA directly targeting *GBA*. Interestingly, in the publicly-available dataset of Siebert and coworkers<sup>41</sup> miR-22-3p mimic resulted to down-regulate GCase activity (Z-score =  $-1.5$ ; suggestive  $P = 0.066$ ), according to our results.

Concerning *GBAPI*, sparse information is available to date, and it is primarily focused on the evolution of the *GBAPI* locus as an example of a very recently acquired pseudogene<sup>23,38</sup>. We hence extensively studied *GBAPI* splicing pattern and expression profile, showing that it is subjected to multiple physiologic in-frame and out-of-frame splicing events and that it is broadly expressed, though often at low levels (Supplementary Figure 2). More interestingly, we showed that *GBAPI* is targeted by NMD, which seems to be the main mechanism regulating its expression level: blocking NMD, the ratio between *GBA* and *GBAPI* substantially increased (on average from 1/100 to 1/68 in HepG2 cells, and from 1/9 to 1/4 in HEK293 cells). Of course, the *GBAPI* expression control exerted through NMD raises the question about the pseudogene translation, since RNAs should undergo a pioneer round of translation, associated with an inspection operated by the NMD machinery, before being degraded<sup>29</sup>. Interestingly, it has been recently reported that many lncRNAs, and even 5’UTRs, are translated, and *GBAPI* was identified among non-canonical human translated open reading frames<sup>42</sup>. The remarkable degradation of *GBAPI* operated by NMD not only may cause its low abundance, but also results in an increased degradation of bound miRNAs, possibly enhancing *GBAPI* efficiency as miRNA sponge, as suggested for other pseudogenes<sup>17</sup>.

The relevance of post-transcriptional regulation in determining the low *GBAPI* expression is also suggested by the observation that *GBAPI* proximal and distal promoters show high level of sequence identity with those of *GBA*, and are hence predicted to have similar transcriptional strength. For instance, the presence of two TATA boxes and two CAAT boxes in the proximal promoter of *GBAPI* exactly recapitulates the architecture of *in-cis* regulatory elements characterizing the *GBA* proximal promoter<sup>23</sup>. Moreover, epigenetic marks are not substantially different when comparing the gene and the pseudogene promoters, as inferred from the UCSC Genome Browser ENCODE tracks (<http://genome.ucsc.edu/>; release Feb. 2009, GRCh37/hg19). Indeed, our in-house preliminary data, obtained with reporter constructs, show that the activity of *GBAPI* promoters does reach the transcriptional levels of the corresponding *GBA* promoters (data not shown).

Our overexpression experiments in different cell lines clearly demonstrated that *GBAPI* 3’UTR, at supraphysiological concentrations, can modulate *GBA* mRNA levels through a miR-22-3p-mediated regulatory circuit. However, these results do not necessarily imply that this ceRNA-based regulation may also work in more physiological conditions. To confirm that *GBAPI* can act as a *GBA* ceRNA without overexpression, we exploited the predicted differential sensitivity to NMD of the gene and pseudogene transcripts (see Fig. 3A). Cycloheximide treatment allowed us to increase the relative abundance of the endogenous *GBAPI* mRNA of around 4 times the basal level and was accompanied by a 2-fold increase in *GBA* transcripts, not directly attributable to NMD, and compatible with a ceRNA effect (Fig. 3B). Hence, in specific cells or developmental stages, up-regulation of

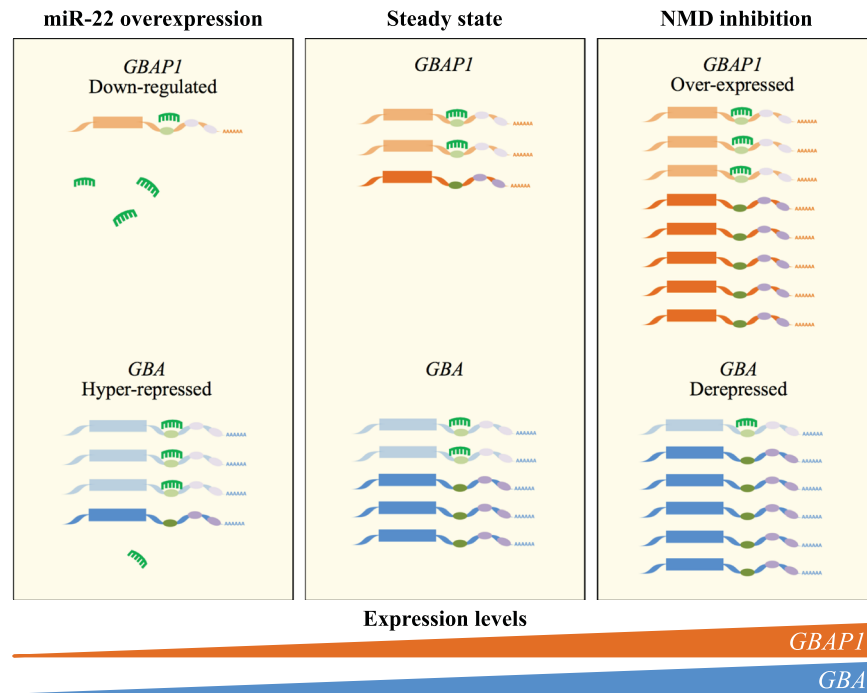


**Figure 4.** *GBA*, *GBAPI*, and miR-22-3p are expressed in iPS cells and iPSC-derived neurons of PD patients and controls. *GBA* (A), *GBAPI* (B), and miR-22-3p (C) expression levels were measured by semi-quantitative real-time RT-PCRs in up to six iPS and iPSC-derived neuronal cells of cases and controls. Boxplots show expression levels according to the disease status; boxes define the interquartile range; the thick line refers to the median. Results are presented as normalized rescaled values. Significance level for differences between groups was calculated by a Wilcoxon-Mann-Whitney test, and showed only if significant. \* $P < 0.05$ .

*GBAPI*, resulting from post-transcriptional or epigenetic regulatory mechanisms, might titrate miRNAs away from the *GBA* protein-coding transcripts, thus providing a physiologic ceRNA effect.

The existence of an RNA-based network controlling *GBA* expression suggests the intriguing possibility that miR-22-3p or *GBAPI* dysregulation could also be associated with PD. In this frame, we investigated *GBA/GBAPI*/miR-22-3p expression pattern in disease-relevant tissues using *in-silico* analyses of microarray datasets publicly available through the Gene Expression Omnibus repository (see Supplementary Materials and Methods), as well as *in-vivo* measurements performed on RNA extracted from iPSCs and iPSC-derived neuronal cells of





**Figure 5.** Schematic representation of the effect of modulating the *GBA/GBAP1/miR-22-3p* RNA-based network on endogenous *GBAP1* and *GBA* levels. Schematic representation of the ceRNA network involving *GBA* (blue transcripts) and *GBAP1* (orange transcripts), harboring the same MRE sites (green and violet ovals). The green MRE sites bind to miR-22-3p (in green), whereas violet ones bind to other not-specified miRNAs. The experimental modulation of the proposed ceRNA network impacts on both coregulated transcripts. In particular, over-expression of miR-22-3p (left part of the figure) determines the down-regulation of both *GBA* and *GBAP1* transcripts. Conversely, over-expression of *GBAP1* (e.g., by inhibiting the NMD pathway, as experimentally verified in the present study; right part of the figure) will increase the cellular concentrations of miR-22-3p MREs, thus resulting in the de-repression of *GBA*. In the scheme, transcripts destined to degradation are colored in lighter shades.

PD cases and controls. In particular, we retrieved three microarray datasets evaluating differential gene expression in the SN of *post-mortem* brains, for a total of 51 cases and 42 controls (Supplementary Table 2). In the meta-analysis, we measured a significant down-regulation of both *GBA* and *GBAP1* transcripts in PD patients ( $P < 0.05$ ; Supplementary Figure 6). Notably, we observed the same significant down-regulation for *GBA* transcripts in iPS-derived DA neurons of PD patients; accordingly, miR-22-3p was slightly, although not significantly, up-regulated in cases vs. controls (on average 1.96 fold,  $P = 0.13$ ; Fig. 4).

A few studies have reported a potential neuroprotective effect of miR-22-3p in rat models of cerebral ischemia-reperfusion injury, as well as in Huntington's and Alzheimer's disease, through a reduction in inflammation and apoptosis<sup>43,44</sup>. However, other studies suggested a pro-senescence role of miR-22 in endothelial progenitor cells, in cancer, and in the aging heart and brain<sup>28,45–47</sup>. While the neuroprotective effects of miR-22 have suggested enhancing its expression as a potential therapeutic strategy for the treatment of neurodegenerative conditions, it may well be that miR-22 overexpression represents a pathophysiologic response to protect the cell from injury and stress also triggering other non-beneficial effects, like increased aging and reduced GCase activity.

In conclusion, we are aware of the fact that the connection between the RNA-based network and PD pathogenesis presented here has not been formally proven. However, one can easily imagine a link between the down-regulation of the sister transcripts *GBA/GBAP1* - or, conversely, the up-regulation of miR-22-3p - and an aberrant  $\alpha$ -synuclein metabolism, as already theorized<sup>12</sup>. A confirmed dysregulation of the *GBA/GBAP1/miR-22-3p* circuit in PD patients would suggest possible novel therapeutic strategies, based either on the direct control of the expression of the miRNA/pseudogene, or on the modulation of the NMD pathway aimed at up-regulating *GBAP1* levels<sup>48</sup>.

## Methods

**Plasmid constructs.** MiR-22-3p and miR-132 precursors were inserted into the psiUX expression vector (kindly provided by Prof. I. Bozzoni, Università di Roma La Sapienza, Rome, Italy). *GBA* and *GBAP1* 3'UTRs were directionally cloned downstream of the renilla luciferase gene in the psiCHECK2 reporter plasmid (Promega, Madison, USA). All constructs were produced by PCR amplifying the relevant genomic region from the DNA of a healthy subject using an appropriate PCR primer couple (Supplementary Table 3), and subsequently by cutting the amplified products with the proper restriction enzyme. Restricted products were ligated into the relevant plasmid.

The constructs carrying the *GBA* and *GBAP1* 3'UTR deleted of the miR-22-3p binding site ( $\Delta$ MRE) were obtained by site-directed mutagenesis, by means of the QuikChange kit (Agilent Technologies, Santa Clara, USA), following the manufacturer protocol.

A pGL3-control luciferase construct containing a single perfectly-complementary miR-22-3p antisense sequence (miR-22-3p sensor), kindly provided by Dr. Da-Zhi Wang (Children's Hospital Boston and Harvard Medical School), was used as a positive control<sup>27</sup>.

All plasmids were purified using the PureYield™ Plasmid Miniprep System kit (Promega) according to the manufacturer's instructions. All recombinant and mutagenized vectors were verified by conventional Sanger sequencing, as described<sup>10</sup>.

**Prediction of *GBA/GBAP1*-targeting miRNAs.** Predictions were performed using publicly-available algorithms: microRNA.org<sup>49</sup>, MicroCosm Targets<sup>50</sup>, PITA<sup>51</sup>, as well as the miRWalk2 suite<sup>52</sup>.

**Cell cultures and transfection experiments.** HEK293, HepG2, and HeLa cells (kind gift of Prof. D. Fornasari and Prof. A. Rollier, University of Milan, Milan, Italy) were cultured according to the standard procedures. All cell lines were routinely tested for mycoplasma contamination.

For **miRNA over-expression** experiments, cells were cotransfected using 3.5  $\mu$ g (HeLa) or 875 ng (HEK293) of the psiUX plasmid expressing either miR-22-3p or miR-132 precursors.

For the **miRNA-target interaction analysis**, HEK293 cells were cotransfected using 300 ng of the psiUX plasmid expressing miR-22-3p together with 720 ng of the psiCHECK2 plasmid containing the relevant 3'UTR.

For the **ceRNA-effect analysis**, HEK293 cells were cotransfected using 300 ng of the psiUX plasmid expressing miR-22-3p together with 300 ng of the psiCHECK2 plasmid containing the *GBAP1* 3'UTR. HepG2 cells were transfected with 300 ng of the *GBAP1* 3'UTR only or with 300 ng of the miR-22-3p sensor (as positive “sponge” control).

In each experiment, an equal number of cells ( $2.5 \times 10^5$  for HeLa,  $3 \times 10^5$  for HEK293,  $4 \times 10^5$  for HepG2) were transfected with the Polyplus jetPRIME (EuroClone, Wetherby, UK) in 6-well plates, as described by the manufacturer. Depending on the measurement to be performed at the end of experiment, cells were collected 24, 48, 72, or 96 hours after transfection (detailed in the relevant figure legend), to obtain either total RNA, or cell lysates (see below).

**RNA samples.** Expression profiles of *GBA*, *GBAP1*, and miR-22-3p were determined using RNA from: a panel of 20 human tissues (First Choice total RNA; Ambion, Austin, USA), a panel of 24 human cerebral regions (Clontech Laboratories, Palo Alto, USA), 11 cell lines, iPSCs, and DA neurons differentiated from iPSCs (see below).

RNA from cell lines, iPSCs, DA neurons, as well as transfected cells was isolated using the Eurozol kit (Euroclone), according to the manufacturer's protocol. RNA concentration/quality was assessed using the Nanodrop ND-1000 (Thermo Fisher Scientific, Waltham, USA).

**Semi-quantitative real-time RT-PCR.** For the evaluation of expression levels of specific genes, random hexamers and the Superscript-III Reverse Transcriptase (Invitrogen, Carlsbad, USA) were used to perform first-strand cDNA synthesis starting from 1  $\mu$ g of RNA extracted from cells, or RNA derived from a panel of human tissues. From a total of 20  $\mu$ L of the RT reaction, 1  $\mu$ L was used as template for amplifications using the FastStart SYBR Green Master Mix (Roche, Basel, Switzerland) on a LightCycler 480 (Roche), following a touch-down thermal protocol. Expression levels were normalized using *HMBBS* (hydroxymethylbilane synthase gene) and *ACTB* ( $\beta$ -actin) as housekeeping genes. To discriminate between the quasi-identical *GBA* and *GBAP1* genes, we took advantage of the 55-bp deletion in exon 9 characterizing *GBAP1* as well as of the few nucleotide differences between *GBA* and *GBAP1* spread along the two genes.

MiR-22-3p and miR-132 levels were measured by real-time RT-PCR by a poly(A) tailing and a universal reverse transcription approach, using the miRNA First Strand Synthesis kit (Agilent Technologies) and starting from 300 ng of total RNA, according to the manufacturer's instructions. RT-PCR reactions were performed using the universal reverse primer (Agilent Technologies) and miRNA-specific forward primers, as described<sup>53</sup>. U6 snRNA was used as housekeeping gene. Real-time reactions were performed as described above.

In all cases, real-time RT-PCR assays were performed at least in triplicate on a LightCycler 480, and expression levels were analyzed by the GeNorm software<sup>54</sup>. Correlation between *GBA/GBAP1*/miR-22-3p expression profiles was calculated using the Pearson's correlation. Pearson's coefficients  $< -0.5$  and  $> 0.5$  are considered as anti-correlation and positive correlation, respectively. P values  $< 0.05$  were considered as statistically significant.

Primer couples used in RT-PCR assays are listed in Supplementary Table 3.

**Luciferase assays.** For miRNA-target interaction assays, the activities of firefly/renilla luciferase were measured in lysates from transfected cells by using the Dual-Luciferase Reporter Assay System (Promega) and the Wallac 1420 VICTOR<sup>3</sup> V reader (PerkinElmer, Waltham, USA). The values of renilla luciferase were normalized against the corresponding values of firefly luciferase.

**Western blot analysis.** Cells were lysed in water containing protease and phosphatase inhibitors (Sigma-Aldrich, Saint Louis, USA) on ice using an ultrasonic homogenizer. Total cell protein content was measured using the DC Protein Assay (Bio-Rad, Hercules, USA). In total, 40–50  $\mu$ g of the protein lysate was loaded on a 10% polyacrylamide gel and transferred on a PVDF membrane (GE Healthcare, Freiburg, Germany). Blots were incubated with primary antibodies overnight at 4 °C on a shaker platform (Anti-GBA ab128879 1:2,500,

Abcam, Cambridge, USA; anti-GAPDH G9545 1:7,000, Sigma) and were then probed with anti-rabbit IgG-HRP secondary antibody (1:2,000, Cell Signaling, Leiden, The Netherlands) for 1 h at room temperature. Visualization was done by using Westar ETA C 2.0 ECL Substrate for Western Blotting (Cyanagen, Bologna, Italy). For quantitative measurements, membranes were acquired using the Uvitec Cambridge technology (Eppendorf, Hamburg, Germany). Image analysis was performed with the Uvitec software.

**GCcase enzymatic activity assays.** Cells to be assayed were washed twice with phosphate buffered saline (PBS), harvested, and then lysed in water containing complete protease inhibitor cocktails (Roche). Total cell protein content was measured using the Micro BCA assay reagent (Pierce, Rockford, USA). Cells lysates were transferred to a 96-well microplate and assays were performed in triplicate. Cell-lysate associated GCcase activity was analyzed using 4-methylumbelliferyl- $\beta$ -D-glucopyranoside (MUB-Glc; Glycosynth, Warrington, UK), solubilized at a final concentration of 6 mM in McIlvaine Buffer (0.1 M Citrate/0.2 M Phosphate, pH 5.2) containing 0.1% Triton X-100. As Triton is a selective inhibitor of  $\beta$ -glucosidase2 (GBA2) activity, these conditions allowed the specific measurement of GBA-related GCcase activity<sup>55,56</sup>. The reaction mixtures were incubated at 37°C under gentle shaking. The fluorescence was recorded after transferring 10  $\mu$ L of the mixture in the microplate and adding 190  $\mu$ L of 0.25 M glycine, pH 10.7. The fluorescence was detected by a Wallac 1420 VICTOR<sup>3</sup> V reader. Data were expressed as pmoles of converted substrate/mg cell proteins  $\times$  hour.

**GBA and GBAP1 splicing pattern and sensitivity to the NMD pathway.** Analysis of GBA/GBAP1 splicing patterns and susceptibility to NMD was undertaken in HepG2 and HEK293 cell lines. Cells were plated at a density of  $4 \times 10^5$  per 6-well dish and, after 72 hours, treated for 8 hours with cycloheximide (100  $\mu$ g/mL; dissolved in dimethyl sulfoxide) or with the vehicle alone. After the treatment, cells were washed with PBS and total RNA extracted.

For the analysis of the splicing pattern, a set of gene-specific or pseudogene-specific RT-PCR assays (Supplementary Table 3) was designed to catch the vast majority of possible alternative splicing events. RT-PCRs were performed as described above. The main amplified products, recovered from the agarose gel using the Wizard SV Gel and PCR Clean-Up System kit (Promega), were directly sequenced to confirm their identity.

Variations in the expression levels of GBA/GBAP1 upon treatment were quantified by real-time RT-PCR assays using as reference an NMD-resistant transcript (*i.e.*, Connexin 43 or Connexin 32 mRNAs, whose coding sequences are all contained in a single exon, for HEK293 and HepG2, respectively). The NMD-sensitive and insensitive PRKCA transcripts were used as controls<sup>32</sup>.

**Fibroblast-derived iPSCs.** iPSC lines derived from skin fibroblasts of six controls and four PD patients carrying heterozygous GBA mutations (p.L444P,  $n = 2$ ; p.N370S,  $n = 2$ ) were previously described<sup>57</sup> and were obtained following the protocol of Takahashi and colleagues<sup>58</sup>. These iPSCs were subjected to neuronal differentiation for 35 days *in vitro*, according to Kriks and collaborators' protocol<sup>59</sup>.

This study has the approval of the local Ethics Committees (Parkinson Institute, ASST "Gaetano Pini-CTO", Milan, Italy; IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; Medical Faculty and the University Hospital Tübingen, Tübingen, Germany) and was performed according to the Declaration of Helsinki. Signed informed consent was obtained from all participants.

**Data availability.** The authors declare that the data supporting the findings of this study are available within the article and its Supplementary material file or from the corresponding authors on request.

## References

- Aureli, M. *et al.* Cell surface associated glycohydrolases in normal and Gaucher disease fibroblasts. *J. Inherit. Metab. Dis.* **35**, 1081–1091 (2012).
- De Fost, M., Aerts, J. M. & Hollak, C. E. Gaucher disease: from fundamental research to effective therapeutic interventions. *Neth. J. Med.* **61**, 3–8 (2003).
- Sidransky, E. Gaucher disease: insights from a rare Mendelian disorder. *Discov. Med.* **14**, 273–281 (2012).
- Pankratz, N. *et al.* Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).
- Nalls, M. A. *et al.* Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
- Lwin, A., Orvisky, E., Goker-Alpan, O., LaMarca, M. E. & Sidransky, E. Glucocerebrosidase mutations in subjects with parkinsonism. *Mol. Genet. Metab.* **81**, 70–73 (2004).
- Westbroek, W., Gustafson, A. M. & Sidransky, E. Exploring the link between glucocerebrosidase mutations and parkinsonism. *Trends Mol. Med.* **17**, 485–493 (2011).
- Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* **361**, 1651–1661 (2009).
- International Parkinson Disease Genomics Consortium. *et al.* Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649 (2011).
- Asselta, R. *et al.* Glucocerebrosidase mutations in primary parkinsonism. *Parkinsonism Relat. Disord.* **20**, 1215–1220 (2014).
- Cilia, R. *et al.* Survival and dementia in GBA-associated Parkinson's disease: The mutation matters. *Ann. Neurol.* **80**, 662–673 (2016).
- Sidransky, E. & Lopez, G. The link between the GBA gene and parkinsonism. *Lancet Neurol.* **11**, 986–998 (2012).
- Gegg, M. E. *et al.* Glucocerebrosidase deficiency in substantia nigra of parkinson disease brains. *Ann. Neurol.* **72**, 455–463 (2012).
- Alcalay, R. N. *et al.* Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. *Brain* **138**, 2648–2658 (2015).
- Mazzulli, J. R. *et al.* Gaucher disease glucocerebrosidase and  $\alpha$ -synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell* **146**, 37–52 (2011).
- Bae, E. J. *et al.* Glucocerebrosidase depletion enhances cell-to-cell transmission of  $\alpha$ -synuclein. *Nat. Commun.* **5**, 4755, <https://doi.org/10.1038/ncomms5755> (2014).

17. Tay, Y., Rinn, J. & Pandolfi, P. P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505**, 344–352 (2014).
18. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
19. Saugstad, J. A. MicroRNAs as effectors of brain function with roles in ischemia and injury, neuroprotection, and neurodegeneration. *J. Cereb. Blood Flow. Metab.* **30**, 1564–1576 (2010).
20. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
21. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
22. Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
23. Horowitz, M. *et al.* The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics* **4**, 87–96 (1989).
24. Imai, K. *et al.* A novel transcript from a pseudogene for human glucocerebrosidase in non-Gaucher disease cells. *Gene*. **136**, 365–368 (1993).
25. Winfield, S. L., Tayebi, N., Martin, B. M., Ginns, E. I. & Sidransky, E. Identification of three additional genes contiguous to the glucocerebrosidase locus on chromosome 1q21: implications for Gaucher disease. *Genome Res.* **10**, 1020–1026 (1997).
26. Martínez-Arias, R. *et al.* Sequence variability of a human pseudogene. *Genome Res.* **11**, 1071–1085 (2001).
27. Huang, Z. P. *et al.* MicroRNA-22 regulates cardiac hypertrophy and remodeling in response to stress. *Circ. Res.* **112**, 1234–1243 (2013).
28. Xu, D. *et al.* miR-22 represses cancer progression by inducing cellular senescence. *J. Cell Biol.* **193**, 409–424 (2011).
29. Zhang, S. *et al.* MicroRNA-22 functions as a tumor suppressor by targeting SIRT1 in renal cell carcinoma. *Oncol. Rep.* **35**, 559–567 (2016).
30. Popp, M. W. & Maquat, L. E. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.* **47**, 139–165 (2013).
31. Mitrovich, Q. M. & Anderson, P. mRNA surveillance of expressed pseudogenes in *C. elegans*. *Curr. Biol.* **15**, 963–967 (2005).
32. Paraboschi, E. M. *et al.* Functional variations modulating PRKCA expression and alternative splicing predispose to multiple sclerosis. *Hum. Mol. Genet.* **23**, 6746–6761 (2014).
33. Quek, X. C. *et al.* lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* **43**, D168–D173 (2015).
34. Thomson, D. W. & Dinger, M. E. Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.* **17**, 272–283 (2016).
35. Wang, L. *et al.* Pseudogene OCT4-pg4 functions as a natural micro RNA sponge to regulate OCT4 expression by competing for miR-145 in hepatocellular carcinoma. *Carcinogenesis* **34**, 1773–1781 (2013).
36. Karreth, F. A. *et al.* The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma *in vivo*. *Cell* **161**, 319–332 (2015).
37. Zheng, L., Li, X., Gu, Y., Lv, X. & Xi, T. The 3'UTR of the pseudogene CYP4Z2P promotes tumor angiogenesis in breast cancer by acting as a ceRNA for CYP4Z1. *Breast Cancer Res. Treat.* **150**, 105–118 (2015).
38. Wafaei, J. R. & Choy, F. Y. Glucocerebrosidase recombinant allele: molecular evolution of the glucocerebrosidase gene and pseudogene in primates. *Blood Cells Mol. Dis.* **35**, 277–285 (2005).
39. Blech-Hermoni, Y. N. *et al.* In silico and functional studies of the regulation of the glucocerebrosidase gene. *Mol. Genet. Metab.* **99**, 275–282 (2010).
40. Svobodová, E. *et al.* Glucocerebrosidase gene has an alternative upstream promoter, which has features and expression characteristic of housekeeping genes. *Blood Cells Mol. Dis.* **46**, 239–245 (2011).
41. Siebert, M. *et al.* Identification of miRNAs that modulate glucocerebrosidase activity in Gaucher disease cells. *RNA Biol.* **11**, 1291–1300 (2014).
42. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *Elife* **4**, e08890, <https://doi.org/10.7554/eLife.08890> (2015).
43. Jovicic, A., Zaldivar Jolissaint, J. F., Moser, R., Silva Santos, MdeF. & Luthi-Carter, R. MicroRNA-22 (miR-22) overexpression is neuroprotective via general anti-apoptotic effects and may also target specific Huntington's disease-related mechanisms. *PLoS One* **8**, e54222, <https://doi.org/10.1371/journal.pone.0054222> (2013).
44. Yu, H. *et al.* Neuroprotective effects of viral overexpression of microRNA-22 in rat and cell models of cerebral ischemia-reperfusion injury. *J. Cell Biochem.* **116**, 233–241 (2015).
45. Li, N., Bates, D. J., An, J., Terry, D. A. & Wang, E. Up-regulation of key microRNAs, and inverse down-regulation of their predicted oxidative phosphorylation target genes, during aging in mouse brain. *Neurobiol. Aging* **32**, 944–955 (2011).
46. Jazbutyte, V. *et al.* MicroRNA-22 increases senescence and activates cardiac fibroblasts in the aging heart. *Age* **35**, 747–762 (2013).
47. Zheng, Y. & Xu, Z. MicroRNA-22 induces endothelial progenitor cell senescence by targeting AKT3. *Cell Physiol. Biochem.* **34**, 1547–1555 (2014).
48. Popp, M. W. & Maquat, L. E. Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell* **165**, 1319–1322 (2016).
49. Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149–D153 (2008).
50. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
51. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39**, 1278–1284 (2007).
52. Dweep, H. & Gretz, N. miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods* **12**, 697 (2015).
53. Soldà, G. *et al.* A novel mutation within the MIR96 gene causes non-syndromic inherited hearing loss in an Italian family by altering pre-miRNA processing. *Hum. Mol. Genet.* **21**, 577–585 (2012).
54. Vandemospele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, Research0034 (2002).
55. van Weely, S., Brandsma, M., Strijland, A., Tager, J. M. & Aerts, J. M. Demonstration of the existence of a second, non-lysosomal glucocerebrosidase that is not deficient in Gaucher disease. *Biochim. Biophys. Acta.* **1181**, 55–62 (1993).
56. Boot, R. G. *et al.* Identification of the non-lysosomal glucosylceramidase as beta-glucosidase 2. *J. Biol. Chem.* **282**, 1305–1312 (2007).
57. Schöndorf, D. C. *et al.* iPSC-derived neurons from GBA1-associated Parkinson's disease patients show autophagic defects and impaired calcium homeostasis. *Nat. Commun.* **5**, 4028, <https://doi.org/10.1038/ncomms5028> (2014).
58. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
59. Kriks, S. *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* **480**, 547–551 (2011).

## Acknowledgements

This study was supported by Fondazione Cariplo grant N°2015–1017, TELETHON Italy (project n. GTB12001), and “Fondazione Grigioni per il Morbo di Parkinson”. We wish to thank Patricia Corcoran, MSED (Tucson, AZ) for her meticulous English proofreading. Simone Digregorio, Francesca Balistreri, Nicole Tonsi, Chiara Baccin, and Giulia Rovaris are acknowledged for their invaluable work, assistance, and enthusiasm. Emanuele Frattini is specifically thanked for contributing to iPSCs and neurons preparation. Finally, we wish to thank Alba Bonetti, Francesca Natuzzi, Rosanna Morini, and all staff of Parkinson Institute for their effort to support the “Parkinson Institute Biobank” (<http://www.parkinsonbiobank.com>).

## Author Contributions

L.S. and V.R. performed the experiments, analyzed the data, and participated in writing the manuscript; M.S. and M.A. performed the GCase activity experiments; R.K., A.D.F., and M.D. produced iPSCs and differentiated neuronal cell lines from patients and controls; S.G. oversaw patient sample collection and selection for the study; G.S., S.D. and R.A. designed and coordinated the study, analyzed the data and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12973-5>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2017