Genome editing of upstream open reading frames enables translational control in plants

Huawei Zhang[1,4], Xiaomin Si[1,2,4], Xiang Ji[1,2,4], Rong Fan[2,3], Jinxing Liu[1], Kunling Chen[1], Daowen Wang[1,2] & Caixia Gao[1,2]

Translational regulation by upstream open reading frames (uORFs) is becoming established as a general mechanism for controlling the amount of protein that is synthesized from downstream primary ORFs (pORFs)[1–5]. We found that genome editing of endogenous uORFs in plants enabled the modulation of translation of mRNAs from four pORFs that are involved in either development or antioxidant biosynthesis. A single-guide RNA that targeted the region harboring a uORF initiation codon can produce multiple mutations. Following uORF editing, we observed varying amounts of mRNA translation in four pORFs. Notably, editing the uORF of _LsGGP2,_ which encodes a key enzyme in vitamin C biosynthesis in lettuce, not only increased oxidation stress tolerance, but also increased ascorbate content by ~150%. These data indicate that editing plant uORFs provides a generalizable, efficient method for manipulating translation of mRNA that could be applied to dissect biological mechanisms and improve crops.

The control of gene expression by transcriptional regulation has a fundamental role in the generation of phenotypic diversity in living organisms. Post-transcriptional regulation of translation is also involved in regulating protein levels[6]. Substantial efforts have been made to develop methods to manipulate mRNA translation to alter cellular protein levels for therapeutic purposes or to improve traits. For example, reducing the translation of a subset of mRNAs in leukemia cells by manipulating the level of heterotrimeric eukaryotic initiation complex eIF4F is being tested as a potential cancer treatment[7], and antisense long non-coding RNAs are being used to increase the production of specific cellular proteins[8,9]. Translation enhancers have been identified for many animal genes and viruses, and some have been used to increase protein expression from cloned transgenes[10–12].

These applications suggest that targeting mRNA translation could be generally applied to manipulate the amount of protein produced by a wide range of genes. To date, however, a simple, reliable and generalizable approach to manipulate the translation of either plant or animal mRNA is lacking.

We were intrigued by reports investigating the role(s) of uORFs, which are short protein-coding elements located in the 5′ leader region of downstream pORFs[1,2]. uORFs are common in eukaryotic genes. Large proportions of human (49%) and mouse (44%) transcripts harbor one or more uORFs[1]. In 11 plant species, including _Arabidopsis thaliana_, maize and rice, the proportion of transcripts that harbors uORFs varies from ~30% to more than 40%[2]. uORFs with the canonical initiation codon ATG are most common, but non-ATG uORFs are also widespread[3]. Molecular and functional analyses have revealed that uORFs substantially influence the translation of pORF mRNAs, and have measurable effects on phenotype[1–5].

uORFs are frequently detected in the transcripts of plant metabolic genes and have been reported to regulate metabolic functions. In addition, many plant genes that encode protein kinases and transcription factors also harbor uORFs, and some of these uORFs have been reported to regulate crucial developmental processes[2]. For example, a previous study reported that insertion of a uORF upstream of a pORF could be used to control the translation of a gene that regulates plant immunity in rice, providing disease resistance without any reduction in grain yield[5].

Given the precision of CRISPR/Cas9 technology[13], we hypothesized that genome editing of uORFs might prove to be a generally applicable and efficient approach for increasing the translation of mRNAs transcribed from important eukaryotic genes (**Fig. 1a**). To test our hypothesis, we selected four genes that are important in plant development: _AtBRI1_, _AtVTC2_, _LsGGP1_ and _LsGGP2_. _AtBRI1_ (phytohormone brassinosteroid (BR) receptor in _Arabidopsis_[14]) has one ATG-uORF (uORF$_{AtBRI1}$) in its 5′ leader sequence (**Supplementary Fig. 1a**). _AtVTC2_ (GDP-L-galactose phosphorylase (GGP), functions in production of ascorbic acid in plant cells[15]) has a uORF located in the 5′ leader region (uORF$_{AtVTC2}$) with two putative non-canonical initiation codons (ATCACG) (**Supplementary Fig. 1b**), but only the second one (ACG) is surrounded by a Kozak consensus (with a purine at position −3 and a G at +4), and is important for initiating the translation of uORF$_{AtVTC2}$[15]. _LsGGP1_ and _LsGGP2_ are lettuce homologs of _AtVTC2_, and each has a single uORF with a conserved ACG codon (uORF$_{LsGGP1}$ and uORF$_{LsGGP2}$, respectively) similar to uORF$_{AtVTC2}$ (**Supplementary Fig. 1c,d**).

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Hebei Laboratory of Agricultural Water-Saving, Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang, China. [4]These authors contributed equally to this work. Correspondence should be addressed to C. G. (cxgao@genetics.ac.cn).

First, we tested whether mutating the region harboring a uORF initiation codon increases translation of a downstream pORF mRNA. We generated two defined deletions and one mutation in the uORF initiation codon of each of the four genes tested (*AtBRI1*, *AtVTC2*, *LsGGP1* and *LsGGP2*). The read-out for translation was a dual-luciferase reporter system[16] (**Fig. 1b**). The wild-type (WT) 5′ leader,
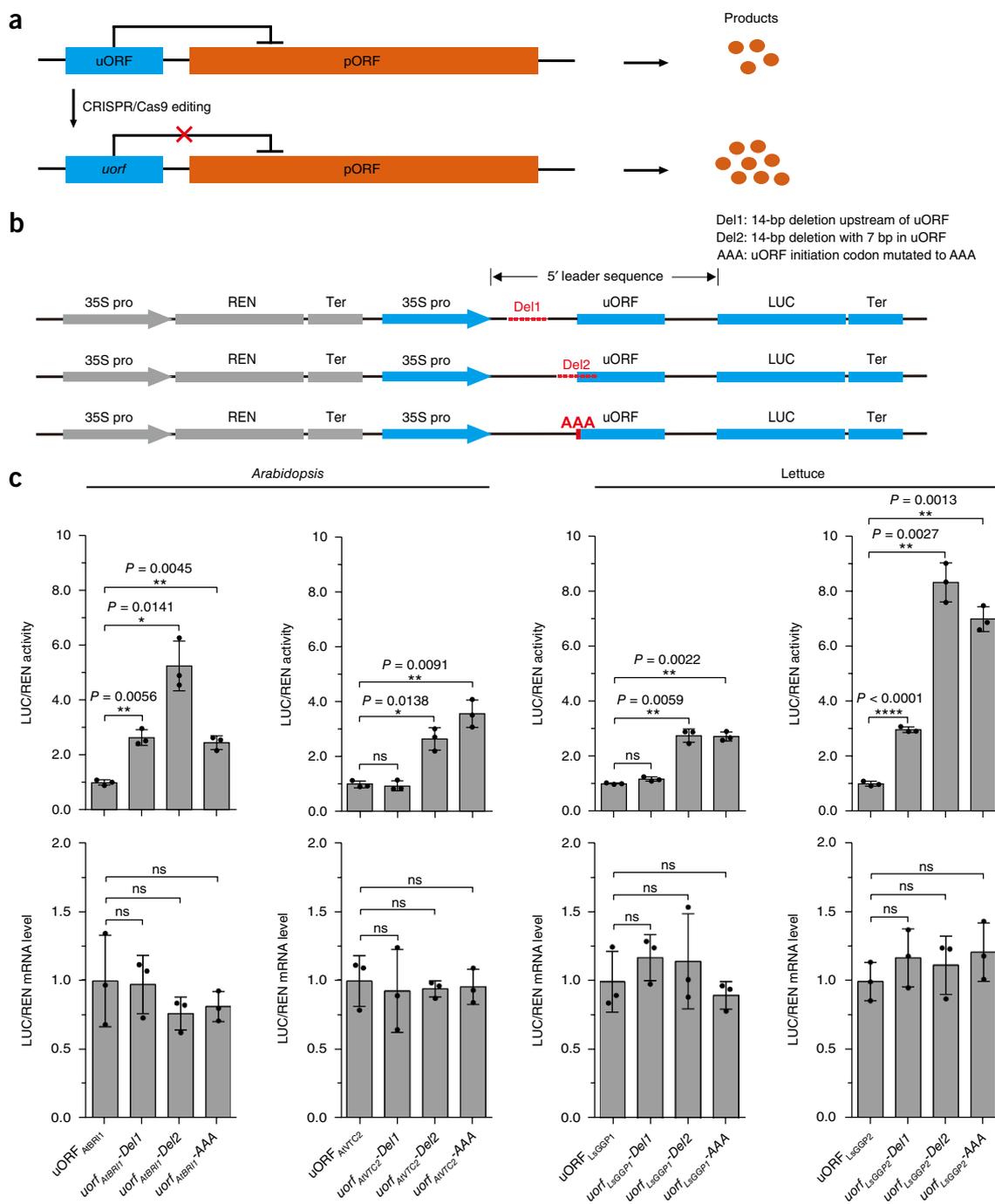

Figure 1 CRISPR/Cas9-mediated genome editing of uORFs. (**a**) Diagram of CRISPR/Cas9-mediated editing of a WT uORF that represses translation of downstream pORF mRNA. The mutant *uorf* reduced translation inhibition, leading to increased proteins production from the pORF. (**b**) Three types of mutant constructs used to test whether altering the region harboring uORF initiation codon is generally effective in disrupting mRNA translation inhibition by uORFs in the dual-luciferase assays. Del1 and Del2 were defined as deletions, whereas AAA involved the mutation of uORF initiation codon to AAA. (**c**) Effects of the three types of mutations on LUC/REN activity (top) and mRNA level (bottom) in dual-luciferase assays. uORF$_{AtBRI1}$, uORF$_{AtVTC2}$, uORF$_{LsGGP1}$ and uORF$_{LsGGP2}$ came from *AtBRI1*, *AtVTC2*, *LsGGP1* and *LsGGP2*, respectively. The WT constructs and those carrying Del1, Del2 or AAA mutations were expressed in *Arabidopsis* or lettuce protoplasts, and the mean LUC/REN activity and mRNA levels conferred by each mutant construct were normalized to those of WT control (*n* = 3 biologically independent experiments). All values represent means ± s.d. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$; ns, no significant difference by two-tailed Student's *t* test.

with intact uORF, and three mutant 5′ leaders (Del1, a 14-bp deletion, 43 bp upstream of the uORF initiation codon; Del2, a 14-bp deletion encompassing the uORF initiation codon; and AAA, uORF initiation codon mutated from ATG or ACG to AAA to prevent uORF translation) were each cloned upstream of the luciferase (LUC) coding region in an expression cassette driven by the 35S promoter. The resulting constructs also harbored a second 35S-promoter-driven cassette expressing *Renilla reniformis* luciferase (REN) as an internal vector control (**Fig. 1b**). Four constructs for each gene were expressed in protoplasts in parallel transfections, and we calculated the mean LUC/REN activities and LUC/REN mRNA levels using data from multiple assays (**Fig. 1c**).

Compared with WT 5′ leaders, mutant leaders with Del2 consistently generated substantially higher LUC/REN activity levels, with fold increases ranging from approximately 2.65 (*uorf*$_{AtVTC2}$-*Del2*) to 8.33 (*uorf*$_{LsGGP2}$-*Del2*) (**Fig. 1c**). The 5′ leaders with the uORF initiation codon mutated to AAA were the second most effective for increasing LUC/REN activity, whereas those with Del1 resulted in little to no increase in LUC/REN activity (**Fig. 1c**). Quantitative RT-PCR assays revealed that the LUC/REN mRNA levels transcribed from the various constructs did not differ significantly (**Fig. 1c**). These data indicate that small deletions that remove the uORF initiation codon might be generally effective at upregulating translation of pORF transcripts without affecting transcription of the pORF. Thus, we chose regions harboring the uORF initiation codons as target sites for single-guide RNAs (sgRNAs) in subsequent genome editing experiments.

We edited uORF$_{AtBRI1}$ and obtained more than 30 mutant lines. Two homozygous transgene-free lines (*uorf*$_{AtBRI1}$-*1* and *uorf*$_{AtBRI1}$-*2*) of *Arabidopsis* were produced, each with a single nucleotide (G or T) insertion in the uORF (uATG) initiation codon (**Fig. 2a**, **Supplementary Fig. 2b** and **Supplementary Tables 1–3**). Transcript levels did not differ between WT and mutant plants (**Fig. 2b**, **Supplementary Fig. 3a,b**), but AtBRI1 protein was upregulated in both mutants, particularly in *uorf*$_{AtBRI1}$-*2* (**Fig. 2c** and **Supplementary Fig. 3a,b**). Overexpression of *AtBRI1* reduces inhibition of *Arabidopsis* hypocotyl growth by brassinazole, which has been used as a chemical agent for inhibiting BR biosynthesis[14]. Consistent with this, in the presence of exogenous brassinazole, the hypocotyls of both mutant seedlings were longer than those of the WT, with the effect being more obvious for *uorf*$_{AtBRI1}$-*2* (**Fig. 2d** and **Supplementary Fig. 4**). These results reveal for the first time, to the best of our knowledge, a role for uORF$_{AtBRI1}$ in regulating the level of AtBRI1 *in vivo* and provide evidence that editing this uORF can modulate pORF mRNA translation. It should be noted that editing the uATG in *uorf*$_{AtBRI1}$-*1* inadvertently created an in-frame GTG codon; this might result in reduced translation of the uORF rather than abolition of translation, as translation can initiate from variant codons, such as GTG, albeit with reduced efficiency[3]. We hypothesize that incomplete inactivation of uORF$_{AtBRI1}$ might be responsible for the reduced effect of the *uorf*$_{AtBRI1}$-*1* mutation in promoting AtBRI1 mRNA translation. Consistent with this notion, the *uorf*$_{AtBRI1}$-*1* mutant leader was substantially less effective than the *uorf*$_{AtBRI1}$-*2* mutant leader, or a leader with the uATG replaced by AAA, at increasing LUC/REN activities (**Supplementary Fig. 5b**). Note that the four leader sequences produced similar LUC/REN mRNA levels (**Supplementary Fig. 5c**).

Next, we tested whether our uORF-editing strategy could be applied to an agronomically relevant plant trait in *Arabidopsis*. Ascorbic acid is an essential nutrient for humans and livestock. Substantial effort has been devoted to increasing ascorbic acid concentrations in plants by bioengineering and breeding[17,18]. In *Arabidopsis*, although uOR-F$_{AtVTC2}$ has been implicated in negative feedback control of *AtVTC2*
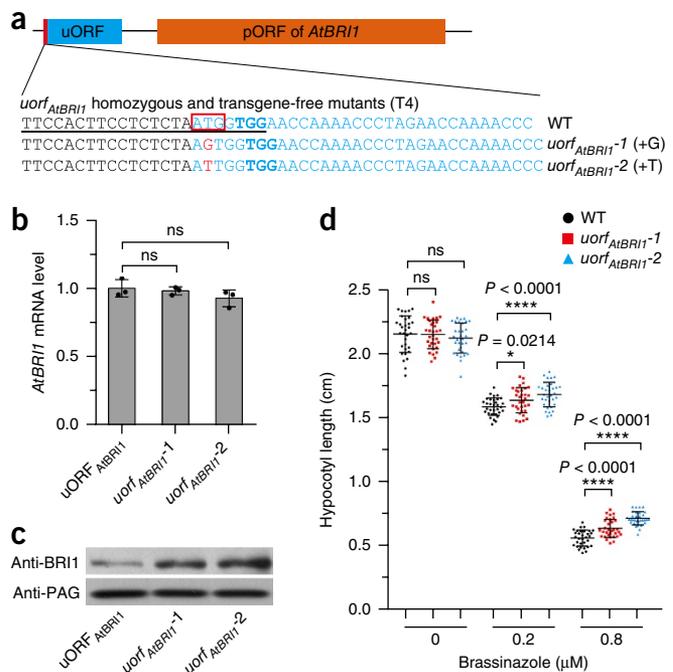

Figure 2 Effect of editing the *AtBRI1* uORF. (**a**) Two homozygous transgene-free mutants (T4 generation) of uORF$_{AtBRI1}$ obtained by CRISPR/Cas9 editing. The uORF sequence (blue) is shown with the sgRNA target site underlined and the protospacer-adjacent motif shown in bold. The G or T nucleotide insertion in the two mutants is labeled in red. The putative initiation codon of uORF$_{AtBRI1}$ is boxed. (**b**) Comparison of mRNA levels among WT control, *uorf*$_{AtBRI1}$-*1* and *uorf*$_{AtBRI1}$-*2*. Actin2 (At3g18780) was used as internal control for the quantitative real-time PCR assay (*n* = 3 biologically independent experiments). (**c**) Comparison of *AtBRI1* protein levels among WT control, *uorf*$_{AtBRI1}$-*1* and *uorf*$_{AtBRI1}$-*2*. PAG (20S proteasome α-subunit G1, At2g27020) was used as loading control in the immunoblot assay. Similar results obtained from two more independent experiments are shown in **Supplementary Figure 3**. (**d**) Hypocotyl lengths of WT control (black dot), *uorf*$_{AtBRI1}$-*1* (red square) and *uorf*$_{AtBRI1}$-*2* (blue triangle) seedlings (*n* = 32 biologically independent samples) grown on 1/2 MS medium containing 0, 0.2 or 0.8 µM brassinazole in the dark for 8 d. All values represent means ± s.d. *$P < 0.05$, ****$P < 0.0001$; ns, no significant difference by two-tailed Student's *t* test.

mRNA translation by ascorbic acid[15], it is not known whether disruption of uORF$_{AtVTC2}$ increases ascorbic acid content. We edited uOR-F$_{AtVTC2}$ and produced two homozygous, transgene-free mutant plants, one with a 37-nucleotide deletion upstream of the initiation codon ACG (*uorf*$_{AtVTC2}$-*1*) and a second with the ACG codon removed by an indel event (*uorf*$_{AtVTC2}$-*2*) (**Fig. 3a**, **Supplementary Fig. 2c** and **Supplementary Tables 1–3**).

We found that the ascorbic acid content of *uorf*$_{AtVTC2}$-*2* leaves was >70% higher than that of WT controls, whereas the ascorbic acid content of *uorf*$_{AtVTC2}$-*1* leaves was unchanged (**Fig. 3d**). Given that the levels of *AtVTC2* transcripts in the WT, *uorf*$_{AtVTC2}$-*1* and *uorf*$_{AtVTC2}$-*2* leaves were similar (**Supplementary Fig. 6**), we conclude that increased translation of *AtVTC2* transcripts is responsible for the large increase of ascorbic acid in *uorf*$_{AtVTC2}$-*2*. In dual-luciferase reporter assays, the 5′ *uorf*$_{AtVTC2}$-*2* leader increased LUC/REN activity by 261%, whereas the *uorf*$_{AtVTC2}$-*1* leader generated a similar level of LUC/REN activity as WT. Notably, the LUC/REN mRNA levels containing these three leader sequences did not differ significantly
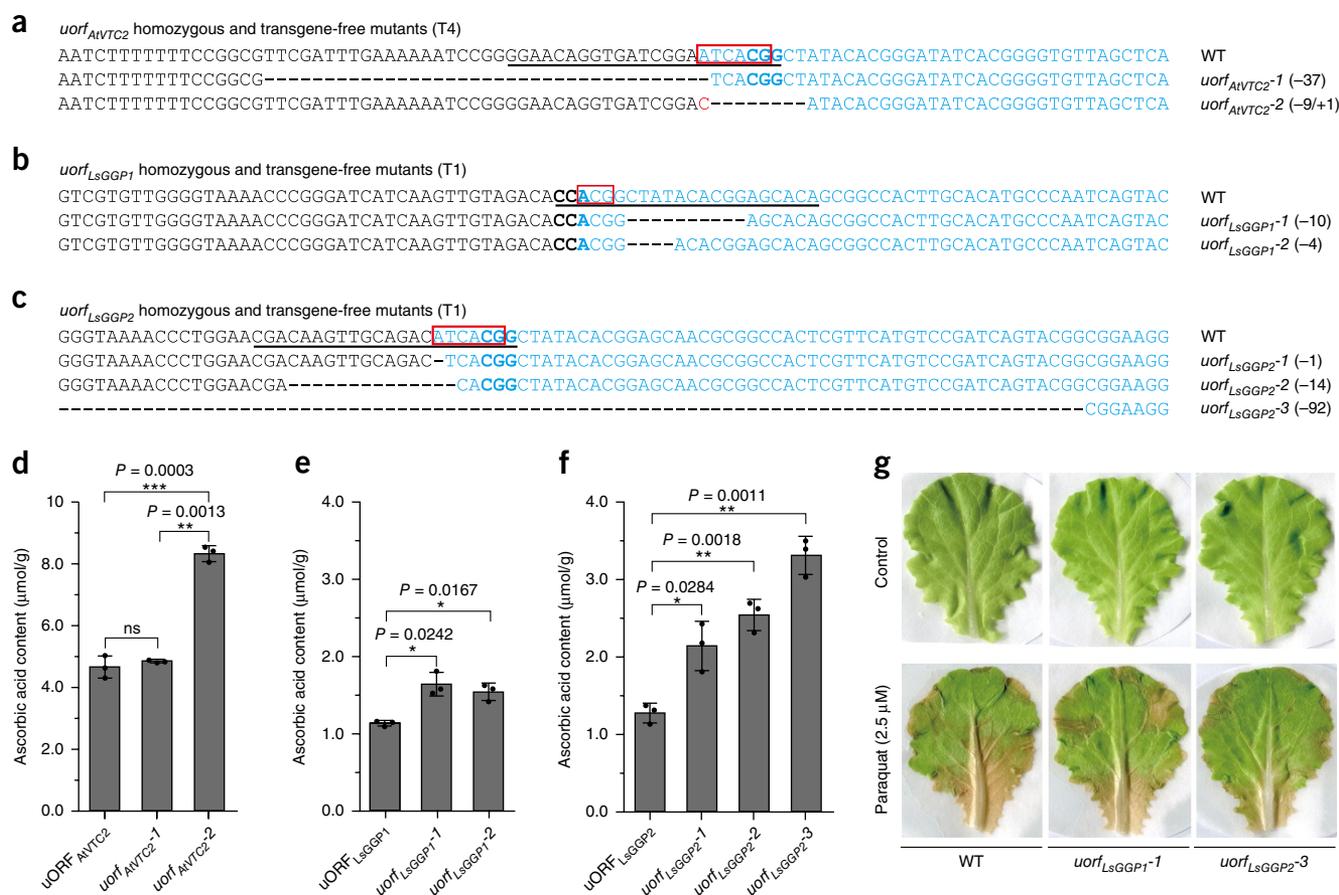
Figure 3 Editing uORFs of *Arabidopsis AtVTC2* and lettuce homologs *LsGGP1* and *LsGGP2* increases foliar ascorbic acid production. (**a–c**) CRISPR/Cas9 induced mutations in the homozygous transgene-free uORF mutants for *AtVTC2* (T4 generation; **a**), *LsGGP1* (T1 generation; **b**) and *LsGGP2* (T1 generation; **c**). The uORF sequences are colored blue and the sgRNA target sites are underlined with the protospacer-adjacent motif shown in bold. Inserted nucleotides are red, deleted nucleotides are shown by hyphens. Putative initiation codons of uORFs are in boxes. (**d–f**) Comparison of foliar ascorbic acid concentration between WT control and the relevant mutants of *AtVTC2* (**d**), *LsGGP1* (**e**) and *LsGGP2* (**f**). Rosette leaves of 14-d-old *Arabidopsis* plants (**d**) or leaves of 18-d-old lettuce plants (**e,f**) were using measurements of three biologically independent experiments. For *Arabidopsis,* about ten seedlings were mixed for each biological repeat. For lettuce, each biological repeat contained the leaves from one plant. (**g**) Oxidation-induced leaf bleaching by the oxidant paraquat in WT lettuce and the uORF mutants *uorf$_{LsGGP1}$-1* and *uorf$_{LsGGP2}$-3*. Leaves were detached from 18-d-old plants and incubated in distilled water containing 0 or 2.5 μM paraquat under continuous light for 24 h before being photographed. The photographs shown are typical of three separate tests, with each test using three leaves from three different plants for each genotype. All values in **d–f** are means ± s.d. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$; ns, no significant difference by two-tailed Student's *t* test.

(**Supplementary Fig. 7**). These data confirm that *AtVTC2* mRNA translation is boosted by the *uorf$_{AtVTC2}$-2* mutation, but not by the *uorf$_{AtVTC2}$-1* mutation. The effect of the *uorf$_{AtVTC2}$-2* mutation on *AtVTC2* mRNA translation was similar to that resulting from deletion of the initiation codon (ACG) (**Fig. 3a**).

We then investigated whether editing the conserved uORF in *AtVTC2* homologs could be used to increase the ascorbic acid content of lettuce, which is a cheap and convenient source of vitamins for human consumption. Lettuce (*Lactuca sativa* L.) is a globally popular vegetable[19]. Using sequence alignment, we identified two *AtVTC2* homologs, *LsGGP1* and *LsGGP2*, the protein products of which were more than 70% identical to AtVTC2 (**Supplementary Fig. 8**). Single uORFs were identified upstream of *LsGGP1* (uORF$_{LsGGP1}$) and *LsGGP2* (uORF$_{LsGGP2}$) (**Supplementary Fig. 1c,d**), and two different sgRNAs (**Fig. 3b,c** and **Supplementary Table 4**) were designed to mutate them.

We analyzed two homozygous transgene-free mutant lines for uORF$_{LsGGP1}$ (*uorf$_{LsGGP1}$-1* and *uorf$_{LsGGP1}$-2*; **Fig. 3b, Supplementary Fig. 9** and **Supplementary Table 5**). In all of the plant lines that we tested,

the ACG codon was retained, but deletions introduced by genome editing disrupted the reading frame and we expected that they would reduce translation from the uORF. Compared with WT lettuce, the foliar ascorbic acid contents of *uorf$_{LsGGP1}$-1* and *uorf$_{LsGGP1}$-2* were increased by 43.1% and 34.5%, respectively (**Fig. 3e**), whereas *LsGGP1* mRNA levels were unaffected (**Supplementary Fig. 10a**). Thus, the mutations in *uorf$_{LsGGP1}$-1* and *uorf$_{LsGGP1}$-2* reduced translation of uORF$_{LsGGP1}$ and enhanced translation of *LsGGP1* mRNA.

For uORF$_{LsGGP2}$, three homozygous transgene-free mutants were analyzed in more detail (**Fig. 3c, Supplementary Fig. 9** and **Supplementary Table 5**). In *uorf$_{LsGGP2}$-1* and *uorf$_{LsGGP2}$-2*, small deletions (up to 14 bp) were present proximal to the initiation codon (ACG), but a large deletion (92 bp) in *uorf$_{LsGGP2}$-3* led to the loss of both the initiation codon and a portion of the uORF coding region (**Fig. 3c**). Foliar ascorbic acid content was significantly increased in all three mutants and, as expected, the increase was greater for *uorf$_{LsGGP2}$-3* (156.6%) than for *uorf$_{LsGGP2}$-1* and *uorf$_{LsGGP2}$-2* (66.6% and 97.7%, respectively; **Fig. 3f**). Transcript levels of *LsGGP2* did

not differ markedly between the WT control and $uorf_{LsGGP2}$-1 and $uorf_{LsGGP2}$-2; however, *LsGGP2* was significantly upregulated in $uorf_{LsGGP2}$-3 (**Supplementary Fig. 10b**) and we speculate that the large deletion in the uORF of the $uorf_{LsGGP2}$-3 mutant promoted the accumulation of the mRNA, as well as increasing translation of the pORF. We note that the ascorbic acid contents of the *LsGGP2* uORF mutants were generally higher than those of the *LsGGP1* uORF mutants (**Fig. 3e,f**), which suggests that *LsGGP2* may encode the major GGP isozyme of ascorbic acid biosynthesis in lettuce.

Ascorbic acid is known to increase antioxidation activity[20]. We treated two uORF mutants of *LsGGP1* and *LsGGP2* with paraquat, an herbicide that kills plant cells by inducing superoxide anions[20], and found that they were more tolerant to this herbicide than the WT (**Fig. 3g**). Quantitative comparison revealed that the percentages of chlorotic area in $uorf_{LsGGP2}$-3 (35.7%) and $uorf_{LsGGP1}$-1 (44.1%) leaves were both significantly lower than that in WT controls (56.1%) when treated with paraquat for 24 h (**Supplementary Fig. 11**). The higher tolerance of $uorf_{LsGGP2}$-3 leaves to paraquat is consistent with the greater accumulation of ascorbic acid in this line (**Fig. 3e,f**). One concern is the occurrence of off-target mutations. Using the standard T7E1 assay[21], we did not find mutations in the potential off-target sites in the mutant lines of *AtBRI1*, *AtVTC2*, *LsGGP1* and *LsGGP2* that we generated (**Supplementary Figs. 12** and **13** and **Supplementary Table 6**).

On the basis of our data, we propose that genome editing of uORFs can, in general, increase mRNA translation, thereby increasing the amounts of protein synthesized. We found that a single sgRNA targeting the region harboring a uORF initiation codon can produce multiple mutants with varying levels of translation of the mRNA of the relevant pORF, which makes this strategy simple. The resulting mutants can provide insights into key biological processes. Another important advantage of our method is that transgene-free lines of plants with improved traits, as exemplified by the $uorf_{LsGGP1}$ and $uorf_{LsGGP2}$ mutants, are readily obtained, which might expedite crop improvement.

The availability of a suite of genome-editing tools, including base editors and the availability of Cas9 variants with expanded PAM recognition[22–24], means that manipulation of uORFs by genome editing to fine-tune mRNA translation and in turn protein concentrations is likely to become easier over time. Owing to the presence of uORFs in many eukaryotic genes in both plants and animals, it is conceivable that our method could be applied more widely in the future.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

H.Z., X.S., X.J. and C.G. designed the experiments. X.S., H.Z., X.J., R.F. and J.L. performed the experiments. X.J. and K.C. analyzed the results. C.G. supervised the project and C.G., D.W. and H.Z. wrote the manuscript.

1. Calvo, S.E., Pagliarini, D.J. & Mootha, V.K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. USA* **106**, 7507–7512 (2009).
2. von Arnim, A.G., Jia, Q. & Vaughn, J.N. Regulation of plant translation by upstream open reading frames. *Plant Sci.* **214**, 1–12 (2014).
3. McGillivray, P. *et al.* A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Res.* **46**, 3326–3338 (2018).
4. Liang, X.H. *et al.* Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat. Biotechnol.* **34**, 875–880 (2016).
5. Xu, G. *et al.* uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* **545**, 491–494 (2017).
6. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
7. Pelletier, J., Graff, J., Ruggero, D. & Sonenberg, N. Targeting the eIF4F translation initiation complex: a critical nexus for cancer development. *Cancer Res.* **75**, 250–263 (2015).
8. Carrieri, C. *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**, 454–457 (2012).
9. Zucchelli, S., Patrucco, L., Persichetti, F., Gustincich, S. & Cotella, D. Engineering translation in mammalian cell factories to increase protein yield: the unexpected use of long non-coding SINEUP RNAs. *Comput. Struct. Biotechnol. J.* **14**, 404–410 (2016).
10. Simon, A.E. & Miller, W.A. 3′ cap-independent translation enhancers of plant viruses. *Annu. Rev. Microbiol.* **67**, 21–42 (2013).
11. Vivinus, S. *et al.* An element within the 5′ untranslated region of human Hsp70 mRNA which acts as a general enhancer of mRNA translation. *Eur. J. Biochem.* **268**, 1908–1917 (2001).
12. Pfeiffer, B.D., Truman, J.W. & Rubin, G.M. Using translational enhancers to increase transgene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **109**, 6626–6631 (2012).
13. Yin, K., Gao, C. & Qiu, J.L. Progress and prospects in plant genome editing. *Nat. Plants* **3**, 17107 (2017).
14. Wang, Z.Y., Seto, H., Fujioka, S., Yoshida, S. & Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **410**, 380–383 (2001).
15. Laing, W.A. *et al.* An upstream open reading frame is essential for feedback regulation of ascorbate biosynthesis in *Arabidopsis*. *Plant Cell* **27**, 772–786 (2015).
16. Hellens, R.P. *et al.* Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* **1**, 13 (2005).
17. Agius, F. *et al.* Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nat. Biotechnol.* **21**, 177–181 (2003).
18. Bulley, S. *et al.* Enhancing ascorbate in fruits and tubers through over-expression of the L-galactose pathway gene GDP-L-galactose phosphorylase. *Plant Biotechnol. J.* **10**, 390–397 (2012).
19. Reyes-Chin-Wo, S. *et al.* Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* **8**, 14953 (2017).
20. Qian, W. *et al.* Molecular and functional analysis of phosphomannomutase (PMM) from higher plants and genetic evidence for the involvement of PMM in ascorbic acid biosynthesis in *Arabidopsis* and Nicotiana benthamiana. *Plant J.* **49**, 399–413 (2007).
21. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395–2410 (2014).
22. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
23. Gaudelli, N.M. *et al.* Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
24. Hu, J.H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

ONLINE METHODS

Plasmid construction. To generate mutants of uORFs in *AtBRI1* and *AtVTC2*, *pYAO*:hSpCas9-*AtBRI1uORF*-sgRNA and *pYAO*:hSpCas9-*AtVTC2uORF*-sgRNA were prepared as previously reported[25]. Briefly, AtU6-26-*AtBRI1uORF*-sgRNA and AtU6-26-*AtVTC2uORF*-sgRNA were constructed with the primers listed in **Supplementary Table 4**, and BsaI-digested AtU6-26-SK. They were digested using SpeI and NheI, followed by cloning into SpeI digested *pYAO*:hspCas9.

To produce uORF mutants for *LsGGP1* and *LsGGP2*, pKSE401-*LsGGP1uORF*-sgRNA and pKSE401-*LsGGP2uORF*-sgRNA were constructed as previously described[26]. BsaI-digested pKSE401 was used to insert sgRNAs, which were prepared with the primers listed in **Supplementary Table 4**.

For developing the constructs used in the dual-luciferase assay, the 35S promoter-fused WT and mutated forms of 5′-leader sequences of each gene (**Supplementary Fig. 14**) were synthesized commercially (Generaybio). They were then cloned into the pGreenII0800-LUC vector digested with HindIII and NcoI[16]. Cloning was verified by DNA sequencing. All sequence information and constructs are available upon request.

Generation of uORF mutants for *AtBRI1* and *AtVTC2*. The uORF mutants of *AtBRI1* and *AtVTC2* were generated by genetic transformation of *Arabidopsis* (ecotype Col-0) using flower dipping with the vectors *pYAO*:hSpCas9-*AtBRI1uORF*-sgRNA and *pYAO*:hSpCas9-*AtVTC2uORF*-sgRNA, respectively. The resulting mutants were verified by sequencing the PCR products amplified with targeting site-specific primers (**Supplementary Table 7**). Transgene-free mutants were selected by PCR using four pairs of primers (**Supplementary Table 7**), and verified by lack of antibiotic resistance.

Protoplast culture and transfection. *Arabidopsis* and lettuce protoplasts were isolated from 14-d-old and 10-d-old seedlings grown on 1/2 MS medium, respectively[27,28]. They were transfected by pGreenII0800-*p35S*:*AtBRI1*-5′-leader-LUC, pGreenII0800-*p35S*:*AtVTC2*-5′-leader-LUC, pGreenII0800-*p35S*:*LsGGP1*-5′-leader-LUC, pGreenII0800-*p35S*:*LsGGP2*-5′-leader-LUC or the constructs carrying the mutated forms of the relevant 5′ leader sequences following a previously reported protocol[29]. In each transfection, 20 µg plasmid DNA and approximately 5×10^5 protoplasts were used. Two days after transfection, the protoplasts were harvested by centrifugation at 100 g for 5 min. LUC/REN activity was measured with the Dual-Luciferase Reporter Assay System (Promega). The LUC/REN levels conferred by the constructs with mutated 5′ leader sequences were calculated relative to those produced by pGreenII0800-*p35S*:*AtBRI1*-5′-leader-LUC, pGreenII0800-*p35S*:*AtVTC2*-5′-leader-LUC, pGreenII0800-*p35S*:*LsGGP1*-5′-leader-LUC or pGreenII0800-*p35S*:*LsGGP2*-5′ leader-LUC.

RNA preparation and quantitative real-time PCR. Total RNA was extracted from the desired protoplast and plant samples with the eZNA plant RNA kit (Omega bio-tek). Reverse transcription was performed using M-MLV Reverse Transcriptase (Promega). Subsequently, quantitative real-time PCR was performed using SsoFast EvaGreen Supermix kit (Bio-Rad) following the supplier's instruction. The primers used are listed in **Supplementary Table 7**.

Brassinazole treatment. Brassinazole treatment was carried out using a previously reported protocol[30]. *Arabidopsis* seeds were surface-sterilized with 70% ethanol for 1 min, followed by 10% bleach for 15 min. Seeds were washed three times with sterile water. Sterile seeds were sown on half-strength MS plates, which were wrapped with aluminum foil and placed in the dark at 4 °C for 3 d. Afterwards, plates were unwrapped and placed horizontally in the light at 22 °C for 3 h. These plates were then wrapped with aluminum foil and placed back in dark at 22 °C for 20 h. Geminated seeds were transferred to fresh half-strength MS plates with 0, 0.2 or 0.8 µM brassinazole. These plates were wrapped with aluminum foil and placed horizontally in the dark at 22 °C. After 8 d, the hypocotyl lengths of seedlings were measured and analyzed using ImageJ (https://imagej.nih.gov/ij/).

Agrobacterium-mediated transformation of lettuce and preparation of uORF mutants for *LsGGP1* and *LsGGP2*. Iceberg lettuce (*Lactuca sativa L.* var. *capitata*) seeds were surface-sterilized with 70% ethanol for 1 min, followed by submersion in 1.0% sodium hypochlorite solution for 15 min, and then sown on the MS medium solidified with 0.8% Bacto agar (BD) and supplemented with 3% sucrose. The plates were incubated under a photoperiod of 16 h light (150 µmol m^{-2} s^{-1}) and 8 h dark at 25 °C for 7 d. The cotyledon explants were excised aseptically from germinated seedlings and placed upside down on the MS co-cultivation medium (supplemented with 30 g/l sucrose, 0.8% plant agar, 0.1 mg/l α-naphthalaneacetic acid, and 0.5 mg/l 6-benzylaminopurine) for 2 d. Then the explants were incubated for 10 min with the *Agrobacterium* (EHA105) suspension carrying the desired construct (pKSE401-*LsGGP1uORF*-sgRNA, pKSE401-*LsGGP2uORF*-sgRNA or the empty vector pKSE401). Following co-cultivation, excess *Agrobacterium* cells in the explants were removed with sterile filter paper. The treated explants were placed upside down on MS co-cultivation medium again and incubated at 25 °C in dark for 48 h. Afterwards, explants were transferred to MS selection medium (supplemented with 30 g/l sucrose, 0.8% plant agar, 0.1 mg/l α-naphthalaneacetic acid, 0.5 mg/l 6-benzylaminopurine, 40 mg/l kanamycin monosulfate, and 250 mg/l carbenicilin), and incubated under a 16 h light (150 µmol m^{-2} s^{-1}) and 8 h dark cycle at 25 °C. After 15 d, calli (4-8 mm in diameter) were subcultured on fresh MS selection medium. 10 d later, calli with regenerated shoots were transferred to the MS selection medium containing reduced amounts of α-naphthalaneacetic acid (0.026 mg/l) and 6-benzylaminopurine (0.046 mg/l). When the shoots reached 3 cm, they were transferred to MS rooting medium (1/2 MS supplemented with 15 g/l sucrose, 0.1 mg/l 3-indole acetic acid, and 250 mg/l carbenicilin) for root induction. The plantlets with well-developed shoot and root were each examined for uORF mutations as described above.

Protein extraction and protein gel blot analysis. Protein was extracted from 14-d-old *Arabidopsis* seedlings with an extraction buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% NP40, 4 M urea, and 1 mM PMSF. Protein gel blot analysis was performed with an anti-AtBRI1 antibody[31] (1:1,500 dilution) or an anti-PAG1 antibody[32] (1:10,000 dilution). The secondary antibody was a goat anti-rabbit antibody conjugated to horseradish peroxidase, and reaction signals were visualized using an enhanced chemiluminescence solution (Millipore).

Measurement of ascorbic acid. Ascorbic acid concentration was measured using high-performance liquid chromatography (HPLC) following a previously published protocol[33]. In brief, leaf tissues were ground to powder in liquid nitrogen. The powder was solubilized using an extraction buffer, which contained 74.45 mg EDTA, 286.65 mg TCEP and 5 ml of 98% orthophosphoric acid in a final volume of 100 ml Milli-Q water. The suspension was vortexed for 30 s, followed by incubation at 25 °C for 2 min. Samples were then placed on ice for 10 min. Subsequently, they were centrifuged at 12,000 *g* at 4 °C for 30 min, with the supernatant retained and filtered using 4 mm hydrophilic PTFE syringe filter. Filtered samples were assayed using Pursuit XRs C18 A2000250X046 column (Agilent), and detected by ultraviolet (244 nm).

Oxidation stress tolerance assessment for lettuce mutant leaves. This assessment, involving three repeated tests, was carried out using WT, $uorf_{LsGGP1}$-1 and $uorf_{LsGGP2}$-3 plants. In each test, three leaves from three independent 18-d-old plants were collected from each of the three genotypes, and placed into the petri dishes containing only water or 2.5 µM paraquat in water (Sigma). The treated leaves were incubated under continuous light for 24 h at 25 °C to induce leaf bleaching by paraquat[20]. After 24 h, the leaves were placed on wet filter paper and imaged by a digital scanner, with the total and chlorotic areas in each leaf measured using the Image-J software (https://imagej.nih.gov/ij/). Highly similar results were obtained from the three separate tests.

Analysis of potential off-target edits. The off-target candidates were predicted using the online tool Cas-OFFinder[34]. For the uORF mutants of *AtBRI1*, *AtVTC2*, *LsGGP1* or *LsGGP2*, the transgene-free lines $uorf_{AtBRI1}$-2, $uorf_{AtVTC2}$-2, $uorf_{LsGGP1}$-1 or $uorf_{LsGGP2}$-3 were used as representatives for the analysis. Three separate tests were performed using three different plants. Genomic DNA was extracted from the relevant plants, and amplified by the primers listed in **Supplementary Table 7**. The T7E1 assay was performed as previously reported[21].

Statistical analysis. All numerical values were presented as means ± s.d. Statistical differences in between WT control and the relevant mutants were tested using two-tailed Student's *t* test.

Life Sciences Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All data supporting the findings of this study are available in the article or in Supplementary Information files, or are available from the corresponding author upon request. Sequence data is present in The *Arabidopsis* Information Resource (TAIR) or Phytozome databases under the following accession numbers: *AtBRI1* (At4g39400), *AtVTC2* (At4g26850), *LsGGP1* (Lsat_1_v5_gn_5_3140) and *LsGGP2* (Lsat_1_v5_gn_7_113861).

25. Yan, L. *et al.* High-efficiency genome editing in *Arabidopsis* using YAO promoter-driven CRISPR/Cas9 system. *Mol. Plant* **8**, 1820–1823 (2015).
26. Xing, H.L. *et al.* A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* **14**, 327 (2014).
27. Zhai, Z., Jung, H.I. & Vatamaniuk, O.K. Isolation of protoplasts from tissues of 14-d-old seedlings of *Arabidopsis* thaliana. *J. Vis. Exp.* **30**, e1149 (2009).
28. Woo, J.W. *et al.* DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.* **33**, 1162–1164 (2015).
29. Yoo, S.D., Cho, Y.H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
30. Espinosa-Ruiz, A., Martínez, C. & Prat, S. Protocol to treat seedlings with brassinazole and measure hypocotyl length in *Arabidopsis* thaliana. *Bio Protoc.* **5**, e1568 (2015).
31. Cui, F. *et al.* *Arabidopsis* ubiquitin conjugase UBC32 is an ERAD component that functions in brassinosteroid-mediated salt stress tolerance. *Plant Cell* **24**, 233–244 (2012).
32. Zhang, H. *et al.* The RING finger ubiquitin E3 ligase SDIR1 targets SDIR1-INTERACTING PROTEIN1 for degradation to modulate the salt stress response and ABA signaling in *Arabidopsis*. *Plant Cell* **27**, 214–227 (2015).
33. Kovács, L. *et al.* Quantitative determination of ascorbate from the green alga *Chlamydomonas reinhardtii* by HPLC. *Bio Protoc.* **6**, e2067 (2016).
34. Bae, S., Park, J. & Kim, J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

naturereseach

Corresponding author(s): Caixia Gao

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	7500 software v2.0.6 was used to collect the qPCR data.
Data analysis	Graphpad prism 6 was used to analyze the data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available in the article and its Supplementary Information files or are available from the corresponding author on request. Sequence data in this article can be found in The Arabidopsis Information Resource (TAIR) or Phytozome databases under the following accession numbers: AtBRI1 (At4g39400), AtVTC2 (At4g26850), LsGGP1 (Lsat_1_v5_gn_5_3140) and LsGGP2 (Lsat_1_v5_gn_7_113861).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All the experiments were performed with three biological repeats. And the sample size of hypocotyl growth measurement for each experiment was 32. For Arabidopsis and lettuce protoplast assay, about 500,000 protoplasts were used for each transfection. The number of mutants were confirmed by T7E1 assay and Sanger sequencing.
Data exclusions	No data exclusion.
Replication	All attempts at replication were successful. A minimum of 3 biological reps were included.
Randomization	The protoplasts of Arabidopsis and lettuce were were isolated and randomly separated to each transformation.
Blinding	Samples were not blinded as the data did not require blinding.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	The following first and second antibodies were used in this immunoblot assay: rabbit anti-BRI1 (1:1500, from Qi Xie's lab, Institute of Genetics and Developmental Biology, Chinese Academy of Science), rabbit anti-PAG (1:15000, from Qi Xie's lab, Institute of Genetics and Developmental Biology, Chinese Academy of Science) and horseradish peroxidase conjugated goat-anti-rabbit IgG (H+L) (1:2500, SA00001-2, Proteintech). These antibodies were validated in previous reports (Cui et al., 2012, Zhang et al., 2015).