

# OsSPL13 controls grain size in cultivated rice

Lizhen Si<sup>1</sup>, Jiaying Chen<sup>1</sup>, Xuehui Huang<sup>1</sup>, Hao Gong<sup>1</sup>, Jianghong Luo<sup>1</sup>, Qingqing Hou<sup>1</sup>, Taoying Zhou<sup>1</sup>, Tingting Lu<sup>1</sup>, Jingjie Zhu<sup>1</sup>, Yingying Shangguan<sup>1</sup>, Erwang Chen<sup>1</sup>, Chengxiang Gong<sup>1</sup>, Qiang Zhao<sup>1</sup>, Yufeng Jing<sup>1</sup>, Yan Zhao<sup>1</sup>, Yan Li<sup>1</sup>, Lingling Cui<sup>1</sup>, Danlin Fan<sup>1</sup>, Yiqi Lu<sup>1</sup>, Qijun Weng<sup>1</sup>, Yongchun Wang<sup>1</sup>, Qilin Zhan<sup>1</sup>, Kunyan Liu<sup>1</sup>, Xinghua Wei<sup>2</sup>, Kyungsook An<sup>3</sup>, Gynheung An<sup>3</sup> & Bin Han<sup>1</sup>

Although genetic diversity has a cardinal role in domestication, abundant natural allelic variations across the rice genome that cause agronomically important differences between diverse varieties have not been fully explored. Here we implement an approach integrating genome-wide association testing with functional analysis on grain size in a diverse rice population. We report that a major quantitative trait locus, *GLW7*, encoding the plant-specific transcription factor OsSPL13, positively regulates cell size in the grain hull, resulting in enhanced rice grain length and yield. We determine that a tandem-repeat sequence in the 5' UTR of *OsSPL13* alters its expression by affecting transcription and translation and that high expression of *OsSPL13* is associated with large grains in tropical *japonica* rice. Further analysis indicates that the large-grain allele of *GLW7* in tropical *japonica* rice was introgressed from *indica* varieties under artificial selection. Our study demonstrates that new genes can be effectively identified on the basis of genome-wide association data.

To meet the food demands of a rapidly growing world population, it is critical to increase crop productivity through efficient breeding<sup>1</sup>. Diverse rice varieties carrying a huge amount of genetic variation will be valuable resources for improving rice yield<sup>2</sup>. Asian cultivated rice (*Oryza sativa* L.) is classified into *indica* and *japonica* subspecies. *Japonica* rice was first domesticated from a small population of a wild rice species (*Oryza rufipogon*) in southern China, and *indica* rice was subsequently developed from cross-hybridizations between the *japonica* cultivars and other *O. rufipogon* ecotypes, as cultivated *japonica* rice from China was spread into Southeast and South Asia<sup>3</sup>. Many of the domestication-related genes responsible for marked morphological differences between wild and cultivated rice have been identified<sup>4–7</sup>. Meanwhile, domesticated rice varieties have been differentiated to adapt to local environments and to acquire unique traits under recent genetic improvements. However, the natural allelic variations responsible for rice varietal differences have not been fully explored.

*Indica* varieties are widely grown in tropical and subtropical regions. Most *japonica* varieties are grown in temperate regions, although some are grown in tropical and subtropical regions. Thus, *japonica* varieties can be classified into temperate and tropical ecotypes on the basis of their ecological, morphological and physiological characteristics. Although tropical and temperate *japonica* ecotypes have a closer genetic relationship with each other than with *indica* varieties, their phenotypes are distinct from one another. Typical *indica* rice has long grains and temperate *japonica* rice has short, round grains. But typical tropical *japonica* varieties (also called *javanica* rice) have larger grains

than the temperate *japonica* varieties<sup>8,9</sup>. The genetic basis of these morphological differences remains unknown.

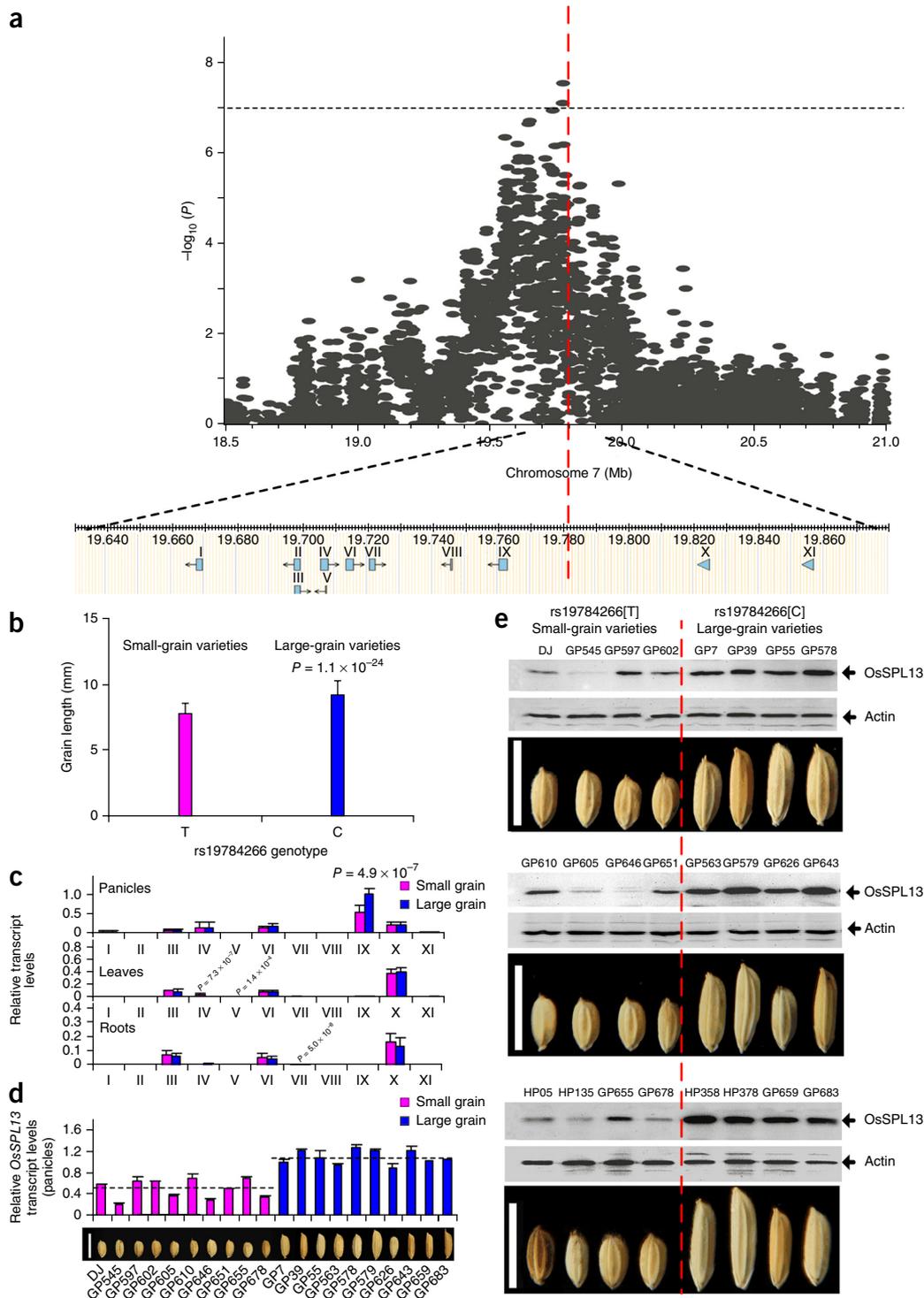
Grain weight is controlled by a combination of grain length, grain width and grain thickness. Thus far, several genes for grain length and grain width, such as *GW2*, *qSW5*, *GS5*, *GW8*, *GS3* and *GL7/GW7*, have been found to regulate grain shape through activation of the cell cycle machinery in promoting cell division<sup>10–16</sup>. Growth of plant organs results from the combination of cell division and cell expansion<sup>17</sup>. It is therefore worthwhile to explore new genetic loci that promote grain size through regulation of cell size. Genome-wide association studies (GWAS) of complex traits in rice have been successful in identifying hundreds of robust association signals<sup>18–20</sup>. However, as a self-fertilizing plant, rice has modest rates of linkage disequilibrium (LD) decay, limiting mapping resolution to regions of approximately 100 kb in length that usually contain several genes<sup>18</sup>. To identify genes underlying grain size traits in rice, we integrated the GWAS approach with analyses of expression patterns, genetic variations and transfer DNA (T-DNA)-derived mutants for functional characterization of a grain shape-related quantitative trait locus (QTL). We identified a major QTL, *GLW7*, that was a key element in producing larger grains and panicles and eventually improved grain yield in cultivated rice.

## RESULTS

### GWAS on grain length and weight in *japonica* rice

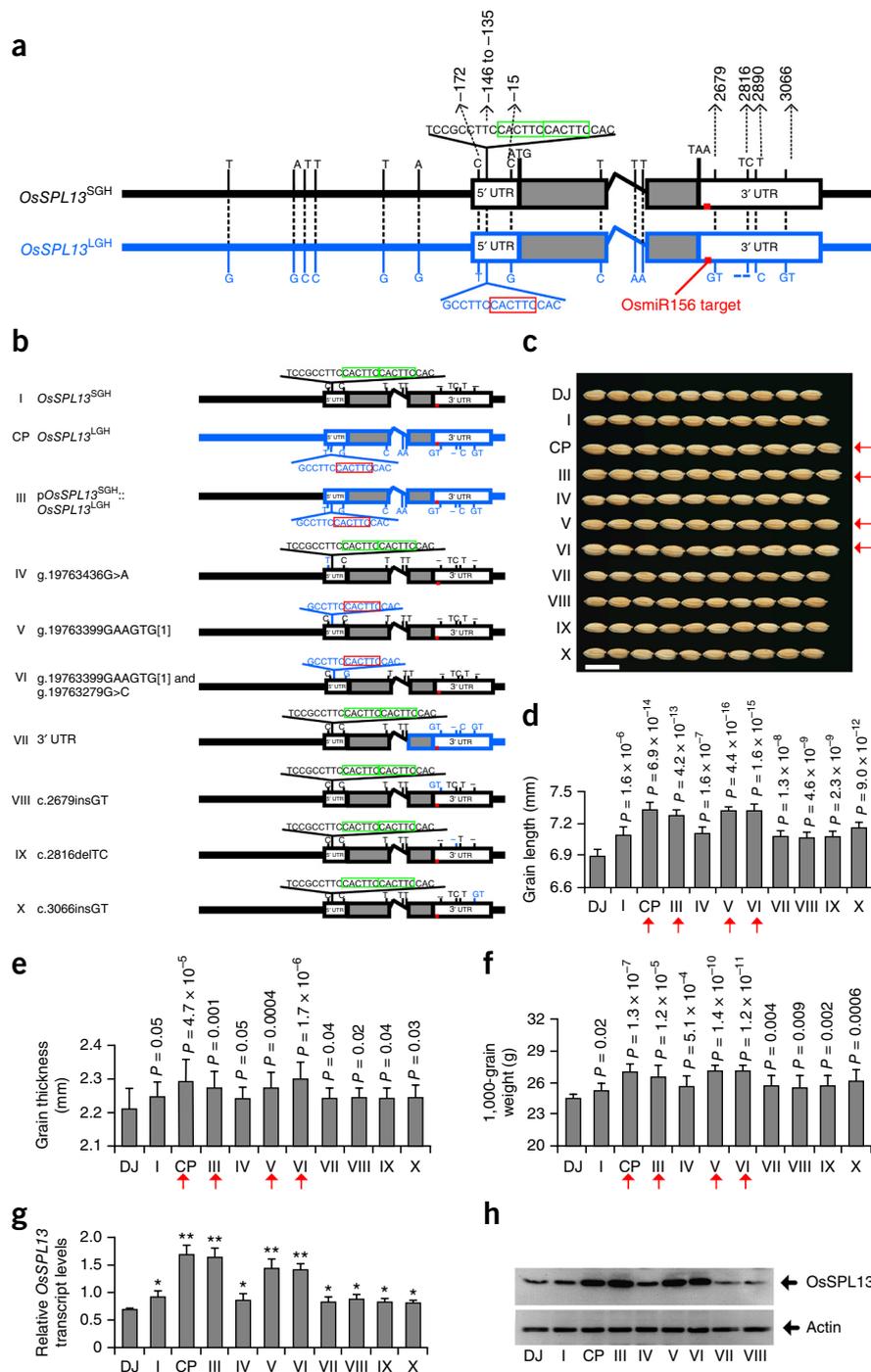
We carried out GWAS of grain size in a diverse collection of worldwide rice germplasm. We rephenotyped 381 *japonica* varieties to map the traits of grain size, including 40 tropical and 341 temperate varieties

<sup>1</sup>National Center for Gene Research, Chinese Academy of Sciences Center for Excellence of Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. <sup>2</sup>State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, China. <sup>3</sup>Crop Biotech Institute, Kyung Hee University, Yongin, Republic of Korea. Correspondence should be addressed to B.H. (bhan@ncgr.ac.cn).



**Figure 1** GWAS analysis and molecular characterization of the large-grain trait. (a) The genome-wide association signals for large grain (grain length and weight) are shown in the region at 18.5–21.0 Mb on chromosome 7 (x axis). The position of the peak SNP is indicated by the red vertical line. Negative  $\log_{10}$ -transformed  $P$  values from the compressed mixed linear model are plotted on the y axis. The horizontal dashed line indicates the genome-wide significance threshold. The locations of the predicted ORFs in the *japonica* Nipponbare genome are indicated: I to XI represent 11 predicted genes (I, *Os07g0502900*; II, *Os07g0503200*; III, *Os07g0503300*; IV, *Os07g0503500*; V, *Os07g0503600*; VI, *Os07g0503700*; VII, *Os07g0503900*; VIII, *Os07g0504601*; IX, *Os07g0505200* (*OsSPL13*); X, *Os07g0506000*; XI, *Os07g0506366*). (b) Analysis of grain length for *japonica* varieties stratified by genotype at rs19784266. The T allele represents the small-grain genotype ( $n = 288$ ), and the C allele represents the large-grain genotype ( $n = 46$ ). Data represent means  $\pm$  s.d. (c) The expression levels of the 11 candidate genes in panicles, leaves and roots are shown. The expression levels for ten small-grain varieties (pink) and ten large-grain varieties (blue) were calibrated to rice *UBIQUITIN* gene expression. Data represent means  $\pm$  s.d. ( $n = 10$ ). (d) Expression analysis of candidate gene IX (*OsSPL13*) in panicles from different varieties and grain appearance for the selected varieties. Data represent means  $\pm$  s.d. ( $n = 3$ ). The dashed horizontal lines represent the average expression levels of *OsSPL13* in ten small-grain varieties and ten large-grain varieties. (e) Analysis of *OsSPL13* protein levels in 12 small-grain and 12 large-grain varieties (separated by the red vertical line). Scale bar, 1 cm.

**Figure 2** Comparative analyses of the *OsSPL13* locus between the small-grain and large-grain haplotypes. **(a)** Structural variations of the *OsSPL13* haplotypes in small-grain and long-grain varieties. White boxes represent the 5' and 3' UTRs, gray boxes represent coding sequences and the line between the gray boxes represents the intron. The red bar indicates the OsmiR156 target site. A minus sign indicates nucleotide deletion, and nucleotide polymorphisms are indicated. **(b)** The constructs used for transformation. Construct I, *OsSPL13*<sup>SGH</sup>, contains the 8-kb genomic sequence of *OsSPL13* from the small-grain Dongjing variety. Construct CP, *OsSPL13*<sup>LGH</sup>, contains the 8-kb genomic sequence of *OsSPL13* from the large-grain GP7 variety. Construct III, p*OsSPL13*<sup>SGH</sup>::*OsSPL13*<sup>LGH</sup>, contains the 2.4-kb promoter region from small-grain variety Dongjing and the 3.6-kb transcript region from large-grain variety GP7. Constructs IV and V have different mutations, as indicated. Genomic coordinates are from IRGSP4. **(c)** Grain sizes of wild-type Dongjing (DJ) and the Dongjing plants transformed with constructs I–X. Scale bar, 1 cm. **(d–f)** Grain length ( $n = 15$ ) **(d)**, grain thickness ( $n = 20$ ) **(e)** and 1,000-grain weight ( $n = 10$ ) **(f)** of plants transgenic for constructs I–X. The red arrows in c–f indicate transgenic plants with dramatic improvement in grain size. **(g,h)** Comparison of *OsSPL13* transcript ( $n = 3$ ) **(g)** and *OsSPL13* protein **(h)** levels in panicles from the different transgenic plants. Panicles were collected when they reached 50% of full panicle length. Actin was used as the loading control. All data in the graphs represent means  $\pm$  s.d.; \* $P < 0.05$ , \*\* $P < 0.01$ .



genotyped previously<sup>19</sup>. A GWAS performed in this *japonica* population for grain length using a mixed linear model approach<sup>21</sup> identified four associated loci meeting a threshold for suggestive evidence of association ( $P < 1 \times 10^{-6}$  in the linear mixed model; false discovery rate (FDR)  $< 0.05$ ), of which only one signal passed a stricter significance cutoff ( $P < 1 \times 10^{-7}$  in the linear mixed model, after Bonferroni correction; FDR  $< 0.01$ ) (Supplementary Fig. 1a and Supplementary Table 1). In contrast to a GWAS in an *indica* population, which identified GS3 as a major locus that controlled grain length and grain weight<sup>19</sup>, we identified a single major locus on chromosome 7 ( $P = 3.98 \times 10^{-7}$ ) responsible for both grain length and grain weight in this *japonica* population. We therefore designated this locus as *GRAIN LENGTH AND WEIGHT ON CHROMOSOME 7* (*GLW7*). Further analysis showed that the *GLW7* locus explains about 30% of the variance in grain length and 25% of the variance in grain weight in this *japonica* population (Fig. 1a and Supplementary Fig. 1b). Two major haplotypes based on the lead SNP (rs19784266 on chromosome 7) of the association signal—large-grain haplotype (LGH) and small-grain haplotype (SGH)—were determined to be associated with large-grain and small-grain phenotypes in *japonica*, respectively. Grain length when stratified by genotype at rs19784266 (T/C) was significantly

higher in LGH varieties than in SGH varieties (Fig. 1b). Interestingly, we found that most tropical *japonica* varieties had the LGH haplotype and most temperate *japonica* varieties had the SGH haplotype in this region.

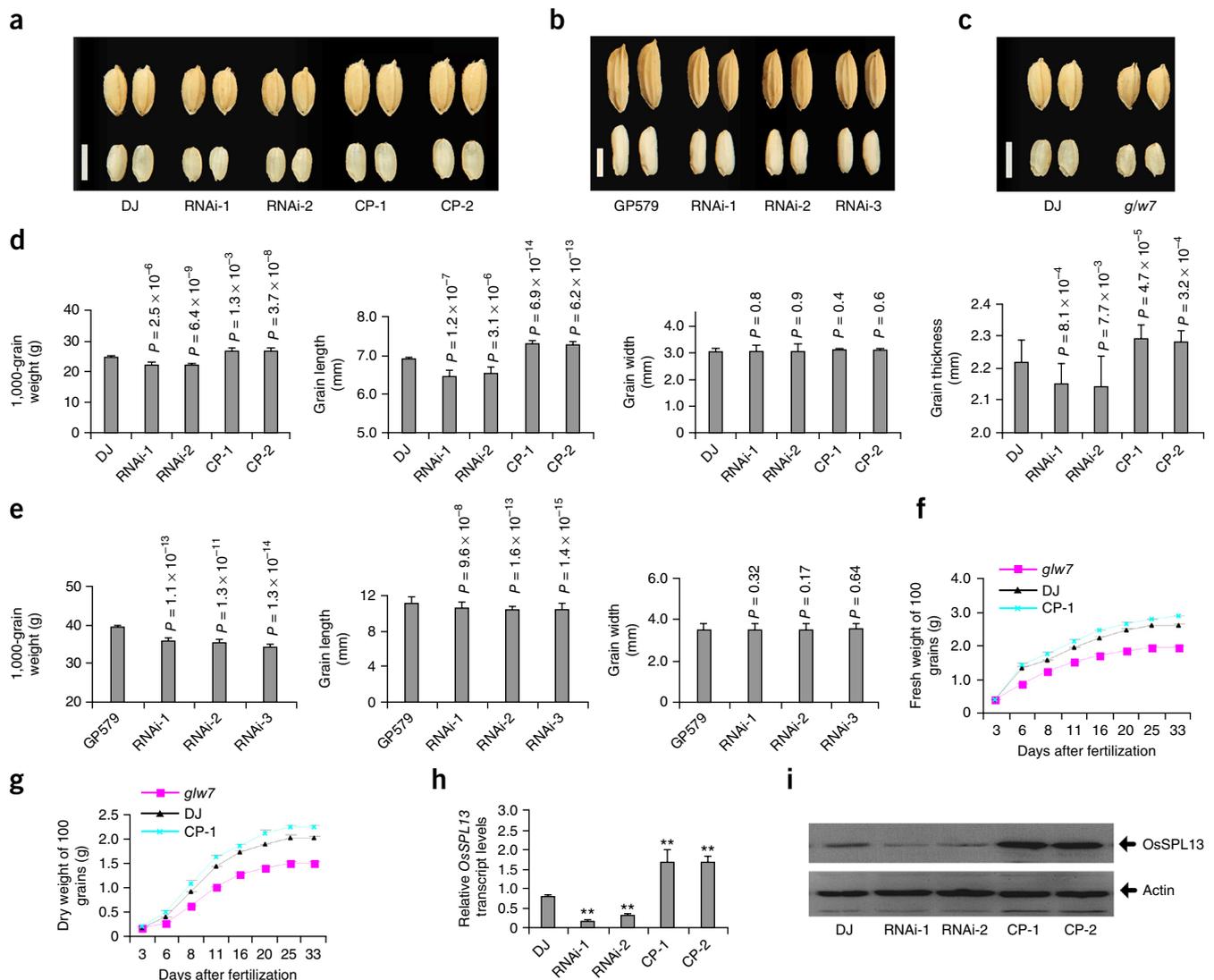
Given the estimated LD decay rate of about 100 to 200 kb in *japonica* landraces<sup>18</sup>, we carefully analyzed the 11 predicted genes within the 260-kb interval centered on the index SNP from the GWAS (Fig. 1a and Supplementary Table 2). We randomly selected ten lines each of large-grain and small-grain *japonica* varieties to measure the expression levels of these 11 genes in panicles, leaves and roots by quantitative RT-PCR (qRT-PCR) (Fig. 1c). Of the 11 genes, 10 showed no difference in expression levels in panicles

from the small-grain and large-grain varieties. Only gene IX (*Os07g0505200*) showed about twofold higher expression in large-grain varieties as compared to small-grain varieties ( $P = 4.9 \times 10^{-7}$ ) (Fig. 1c,d, Supplementary Fig. 2a and Supplementary Data Set). Most of the 11 genes exhibited similar levels of expression in roots and leaves for the large-grain and small-grain varieties (Fig. 1c and Supplementary Fig. 2b,c).

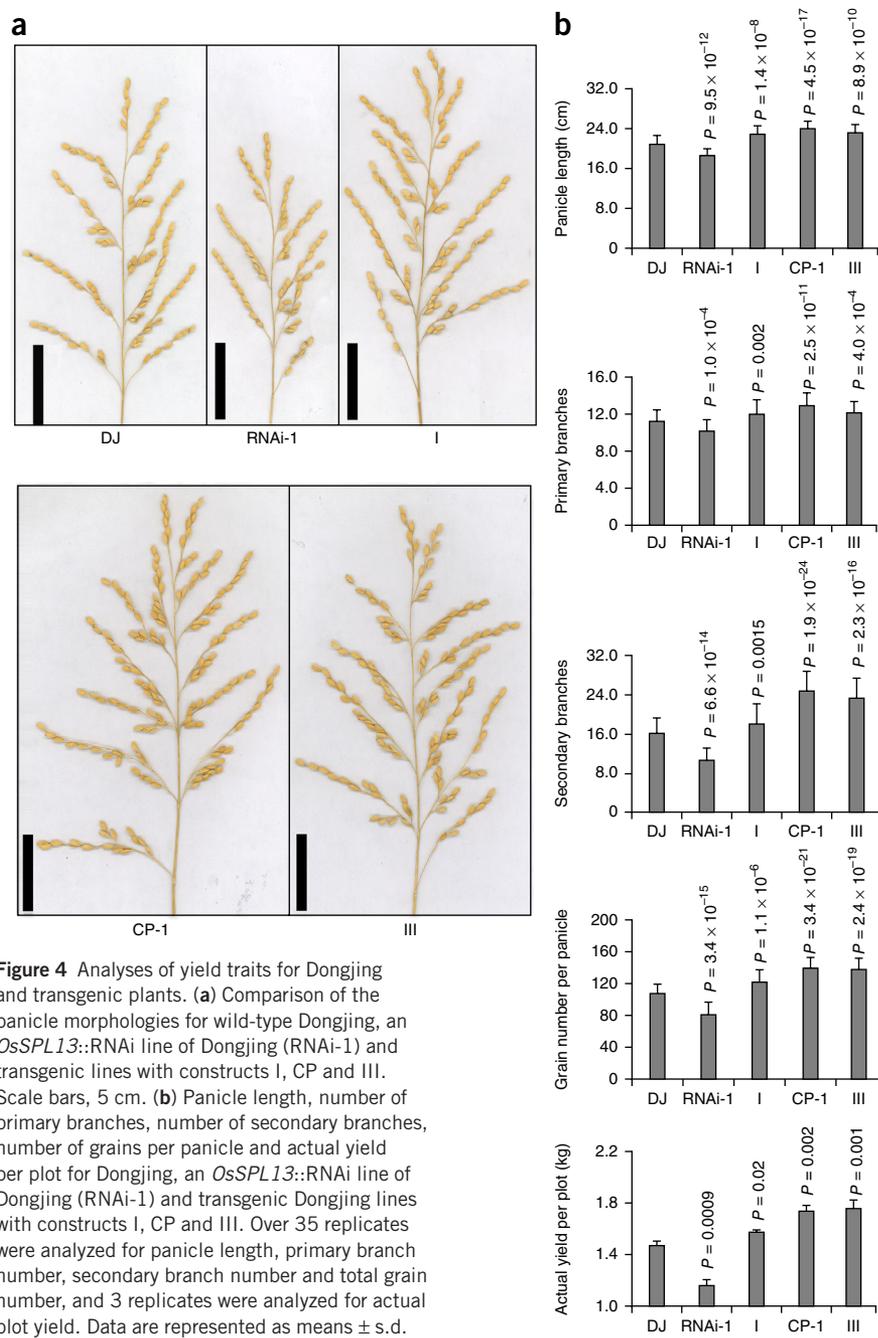
Previous reports indicated that grain shape is mainly determined by the shape of the lemma and palea in rice<sup>11,12</sup>. We found that *Os07g0505200* expression levels in panicles from large-grain varieties were higher than those in small-grain varieties. *Os07g0505200* was also preferentially expressed in developing panicles in comparison to the other ten genes in the examined interval. Therefore, the

*Os07g0505200* gene (also predicted as *LOC\_Os07g32170* and referred to hereafter as *OsSPL13*) was the most likely candidate gene for *GLW7*. *OsSPL13* encodes the plant-specific transcription factor OsSPL13 (ref. 22) (Supplementary Fig. 3). Phylogenetic analysis indicated that OsSPL13 belongs to the SQUAMOSA PROMOTER BINDING PROTEIN (SBP) family of transcription factors (Supplementary Fig. 4). SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes have been shown to have numerous important roles during plant growth and development<sup>15,23–27</sup>.

To investigate whether OsSPL13 protein levels are associated with grain size in the natural *japonica* population, we prepared total protein extracts from the panicles of 24 large-grain and small-grain varieties. We developed a polyclonal antibody against OsSPL13



**Figure 3** Analyses of grain shape for wild-type Dongjing, transgenic plants and a T-DNA mutant. (a) Grain sizes of wild-type Dongjing, *OsSPL13*::RNAi plants (RNAi-1 and RNAi-2) and complementation plants (CP-1 and CP-2) with *OsSPL13*<sup>LGH</sup>. (b) Grain sizes of the large-grain variety GP579 and *OsSPL13*::RNAi lines of GP579 (RNAi-1, RNAi-2 and RNAi-3). (c) Grain sizes of Dongjing and the T-DNA insertion mutant of Dongjing (*glw7*). Scale bars in a–c, 5 mm. (d) Phenotypic data of 1,000-grain weight ( $n = 12$ ), grain length ( $n = 12$ ), grain width ( $n = 12$ ) and grain thickness ( $n = 20$ ) for Dongjing, two *OsSPL13*::RNAi lines (RNAi-1 and RNAi-2) and two complementation transgenic lines (CP-1 and CP-2) with *OsSPL13*<sup>LGH</sup>. (e) Phenotypic data of 1,000-grain weight ( $n = 8$ ), grain length ( $n = 100$ ), grain width ( $n = 100$ ) and grain thickness ( $n = 20$ ) for GP579 and three *OsSPL13*::RNAi lines (RNAi-1, RNAi-2 and RNAi-3) of the GP579 variety. (f,g) Assays of the fresh (f) and dry (g) weight of 100 grains for grain milk filling analysis of Dongjing, the Dongjing *glw7* mutant and Dongjing transformed with *OsSPL13*<sup>LGH</sup> (CP-1) ( $n = 6$ ). (h,i) Comparison of *OsSPL13* transcript ( $n = 3$ ) (h) and *OsSPL13* protein (i) levels in panicles from the different transgenic plants. Panicles were collected when they reached 50% of full panicle length. Actin was used as the loading control. All data in the graphs represent means  $\pm$  s.d.; \*\* $P < 0.01$ .



**Figure 4** Analyses of yield traits for Dongjing and transgenic plants. **(a)** Comparison of the panicle morphologies for wild-type Dongjing, an *OsSPL13*::RNAi line of Dongjing (RNAi-1) and transgenic lines with constructs I, CP and III. Scale bars, 5 cm. **(b)** Panicle length, number of primary branches, number of secondary branches, number of grains per panicle and actual yield per plot for Dongjing, an *OsSPL13*::RNAi line of Dongjing (RNAi-1) and transgenic Dongjing lines with constructs I, CP and III. Over 35 replicates were analyzed for panicle length, primary branch number, secondary branch number and total grain number, and 3 replicates were analyzed for actual plot yield. Data are represented as means  $\pm$  s.d.

by immunizing rabbits with His-tagged OsSPL13 protein purified from *Escherichia coli* (**Supplementary Fig. 5a–c**). Applying this antibody, we found much higher levels of OsSPL13 protein in the large-grain varieties than in the small-grain varieties, although OsSPL13 levels varied in different genetic backgrounds (**Fig. 1e** and **Supplementary Fig. 6a–c**), indicating that OsSPL13 levels are strongly associated with grain size.

#### A tandem repeat in the 5' UTR alters *OsSPL13* expression

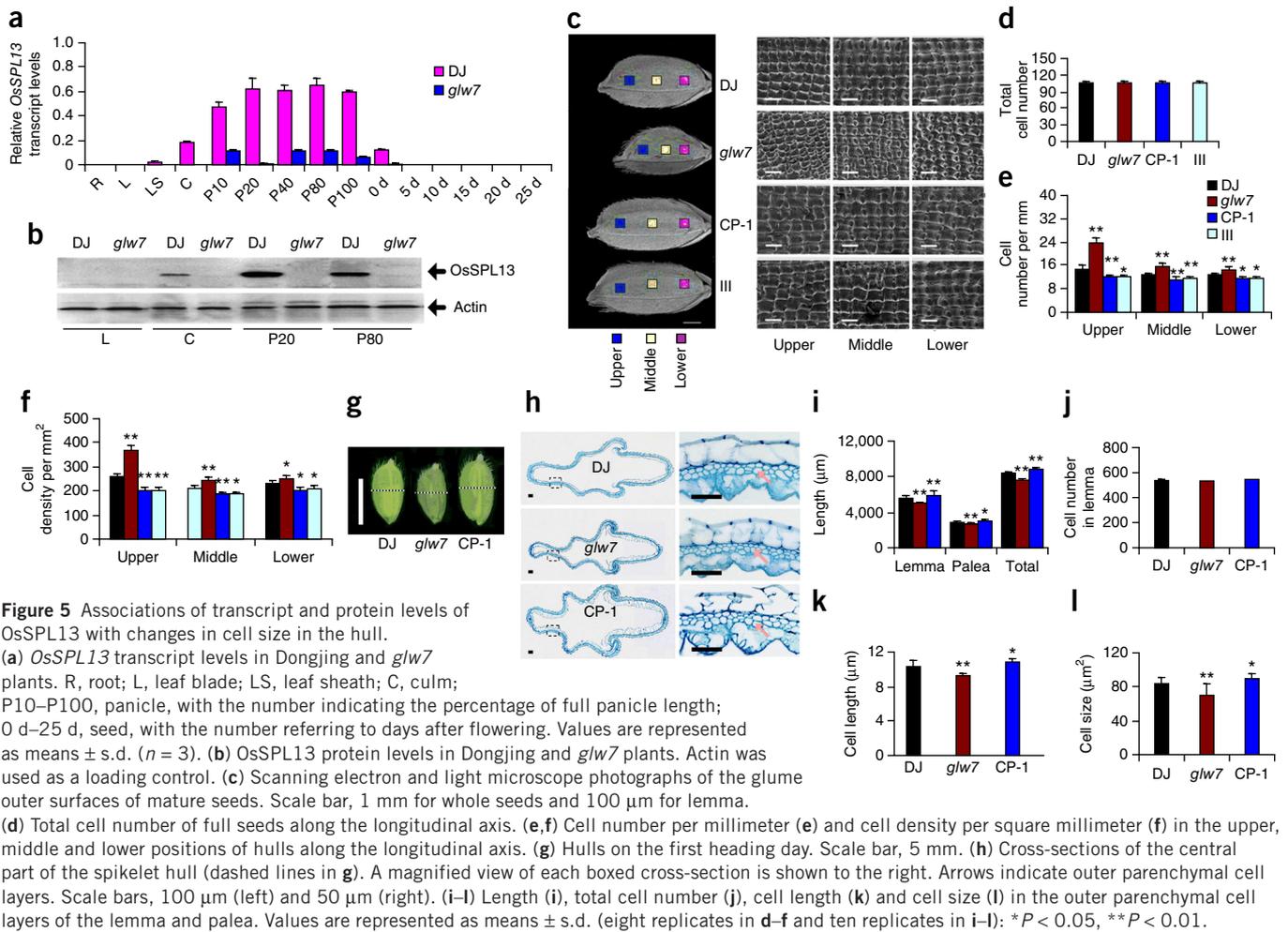
To investigate functional allelic variations in the *OsSPL13* locus, we sequenced the *OsSPL13* gene in 26 small-grain and 21 large-grain *japonica* varieties. Sequencing analysis further confirmed that the genotypes in the *OsSPL13* locus in *japonica* varieties can

be classified into two different haplotypes, *OsSPL13*<sup>LGH</sup> for the large-grain phenotype and *OsSPL13*<sup>SGH</sup> for the small-grain phenotype (**Fig. 2a**). Comparative analysis of the *OsSPL13*<sup>SGH</sup> and *OsSPL13*<sup>LGH</sup> sequences showed that only 16 of 29 polymorphisms were tightly associated with the lead SNP (rs19784266 on chromosome 7) (**Supplementary Table 3**): 6 SNPs in the promoter region (**Fig. 2a**), 3 polymorphisms in the 5' UTR (g.19763436G>A (C>T variant at position –172), g.19763399GAAGTG[1] (one copy of the repeat sequence beginning at –135), g.19763279G>C (C>G variant at position –15)), 1 synonymous polymorphism in exon 1, 2 SNPs in the intron, and 1 SNP and 3 indels in the 3' UTR. No associated amino acid changes were detected in the coding region of *OsSPL13* between the *OsSPL13*<sup>LGH</sup> and *OsSPL13*<sup>SGH</sup> haplotypes.

To determine whether the functional differences in regulation of grain size between the two *OsSPL13* haplotypes are attributable to the polymorphisms in the promoter or coding regions, we generated three transgene constructs—construct I (*OsSPL13*<sup>SGH</sup>), construct CP (*OsSPL13*<sup>LGH</sup>) and construct III (*pOsSPL13*<sup>SGH</sup>::*OsSPL13*<sup>LGH</sup>) (**Fig. 2b**)—which were used to generate transgenic plants (Online Methods). In comparison with wild-type Dongjing plants, the construct CP and III transgenic lines had significant increases in both grain length and grain thickness, resulting in approximately 10% and 9% increases in 1,000-grain weight, respectively (**Fig. 2c–f**). Although transgenic plants with construct I had increased grain size, their grains were smaller than those of transgenic plants with construct CP. No obvious alterations in grain width were observed in any of these transgenic plants (**Supplementary Fig. 6d**). These transgene studies show that the polymorphisms in the promoter region of *OsSPL13* are not responsible for the differences in grain size among *japonica* varieties. We further mutated different sites in the *OsSPL13*<sup>SGH</sup> construct to generate additional constructs

(IV–X; **Fig. 2b**) and carried out transgenic analysis in Dongjing plants. Both construct V and VI transgenic plants had significantly improved grain size, which was similar to that of transgenic plants with construct CP. Transgenic plants with construct IV and four 3' UTR mutations (constructs VII–X) had grain phenotypes similar to those of transgenic plants with construct I (**Fig. 2c–f** and **Supplementary Fig. 7**). Analyses of the various transgenic plants showed higher levels of OsSPL13 transcript and protein in the transgenic lines with constructs CP, III, V and VI than in wild-type Dongjing and the transgenic lines with the other constructs (**Fig. 2g,h**).

Taking these results together, we identify a tandem repeat of the CCATTC sequence from –146 to –135 bp in the 5' UTR of *OsSPL13* as the major cause of the *GLW7* effect on variation in grain size among



**Figure 5** Associations of transcript and protein levels of *OsSPL13* with changes in cell size in the hull. (a) *OsSPL13* transcript levels in Dongjing and *glw7* plants. R, root; L, leaf blade; LS, leaf sheath; C, culm; P10–P100, panicle, with the number indicating the percentage of full panicle length; 0 d–25 d, seed, with the number referring to days after flowering. Values are represented as means  $\pm$  s.d. ( $n = 3$ ). (b) *OsSPL13* protein levels in Dongjing and *glw7* plants. Actin was used as a loading control. (c) Scanning electron and light microscope photographs of the glume outer surfaces of mature seeds. Scale bar, 1 mm for whole seeds and 100  $\mu$ m for lemma. (d) Total cell number of full seeds along the longitudinal axis. (e, f) Cell number per millimeter (e) and cell density per square millimeter (f) in the upper, middle and lower positions of hulls along the longitudinal axis. (g) Hulls on the first heading day. Scale bar, 5 mm. (h) Cross-sections of the central part of the spikelet hull (dashed lines in g). A magnified view of each boxed cross-section is shown to the right. Arrows indicate outer parenchymal cell layers. Scale bars, 100  $\mu$ m (left) and 50  $\mu$ m (right). (i–l) Length (i), total cell number (j), cell length (k) and cell size (l) in the outer parenchymal cell layers of the lemma and palea. Values are represented as means  $\pm$  s.d. (eight replicates in d–f and ten replicates in i–l). \* $P < 0.05$ , \*\* $P < 0.01$ .

*japonica* rice (Supplementary Fig. 8), with two CCATTC copies in the *OsSPL13* 5' UTR causing reduced expression levels of *OsSPL13* and resulting in small grains.

### *OsSPL13* positively increases cell size and grain yield

We then carried out RNA interference (RNAi) to suppress the expression of *OsSPL13* in the small-grain *japonica* variety Dongjing and the large-grain *japonica* variety GP579. The corresponding transgenic plants showed significantly decreased 1,000-grain weight, grain length and grain thickness when compared with the wild-type plants. No changes were detected in grain width (Fig. 3a,b,d,e). Control plants with pTCK303 did not show any differences in grain shape for the transgenic plants (Supplementary Table 4).

To elucidate how *OsSPL13* regulates grain size, we isolated a *glw7* mutant with a T-DNA insertion in the intron of *OsSPL13* (1B-14102) from a mutant library of the *japonica* Dongjing variety<sup>28</sup> (Supplementary Fig. 9a). *glw7* plants had significantly lower grain weight (28%), grain length (12.3%) and grain thickness (9%) as compared to wild-type Dongjing plants (Fig. 3c and Supplementary Fig. 9b–g). Transgenic analysis of *glw7* plants using the 8-kb genomic fragment of wild-type *OsSPL13* from Dongjing (construct I) showed that this construct could completely restore the grain phenotype of *glw7* to that of the wild-type Dongjing strain (Supplementary Fig. 10a,b). The grain size of transgenic plants with construct CP (CP-1 and CP-2) was significantly higher than that of wild-type Dongjing plants (Fig. 3d). Consistent with these results, higher levels

of *OsSPL13* transcript and protein were detected in CP-1 and CP-2 plants and lower levels were observed in RNAi plants as compared to wild-type Dongjing plants (Fig. 3h,i).

We measured the rate of grain milk filling for the mutants and transgenic lines. Significant differences in fresh and dry endosperm weight between the *glw7* plants and the transgenic lines were detected on day 8 after fertilization. These differences reached a maximum on day 25 after fertilization (Fig. 3f,g). These results show that *OsSPL13* has an important role in the accumulation of dry matter in rice grains.

*OsSPL13* also has an important role in panicle development. Panicle length and the number of primary branches for plants transformed with constructs I, CP and III were significantly higher than for Dongjing plants (Fig. 4a,b). The number of secondary branches and grains per panicle of transgenic plants with construct CP were 55.9% and 28.8% higher, respectively, as compared to Dongjing plants (Fig. 4a,b). Consistently, panicle length, branch number and the number of grains per panicle were greatly reduced in the *glw7* and Dongjing RNAi-1 lines (Fig. 4b and Supplementary Fig. 11a–e). No differences were observed in plant height and total tiller number per plant for transgenic plants with constructs I, CP and III (Supplementary Fig. 12a,b). Taking all these effects into account, transgenic plants with constructs I, CP and III had an advantage of 7%, 23% and 19% in plot grain yield, respectively, with respect to wild-type Dongjing (Fig. 4b), whereas the grain yield of *glw7* plants decreased by 80% per plant (Supplementary Fig. 11f).

We investigated the grain quality of wild-type Dongjing and *glw7* plants and the transgenic lines with constructs I, CP and III. With respect to *glw7* plants, protein content was higher than in the wild-type Dongjing plants, whereas amylose content and gel consistency were lower. The percentage of grains with chalkiness was lower in all these transgenic plants than in wild-type plants (Supplementary Table 5). Scanning electron microscopy showed little difference in the starch granules of mature endosperm from *glw7* plants relative to Dongjing plants, and no obvious differences were detected in transgenic plants with construct CP (Supplementary Fig. 13). OsSPL13 transcript and protein levels were considerably higher in developing Dongjing panicles than in culms; no OsSPL13 protein was detected in *glw7* plants (Fig. 5a,b).

OsSPL13 regulates grain length and grain thickness but does not regulate grain width. We investigated cell number and cell size in the outer epidermis of mature grains and found no difference in total cell number for the lemma along the longitudinal axis among *glw7* plants, wild-type Dongjing plants and transgenic lines with constructs CP and III (Fig. 5c,d), suggesting that the increase in grain length associated with *GLW7* might result mainly from cell expansion and not from an increase in cell number.

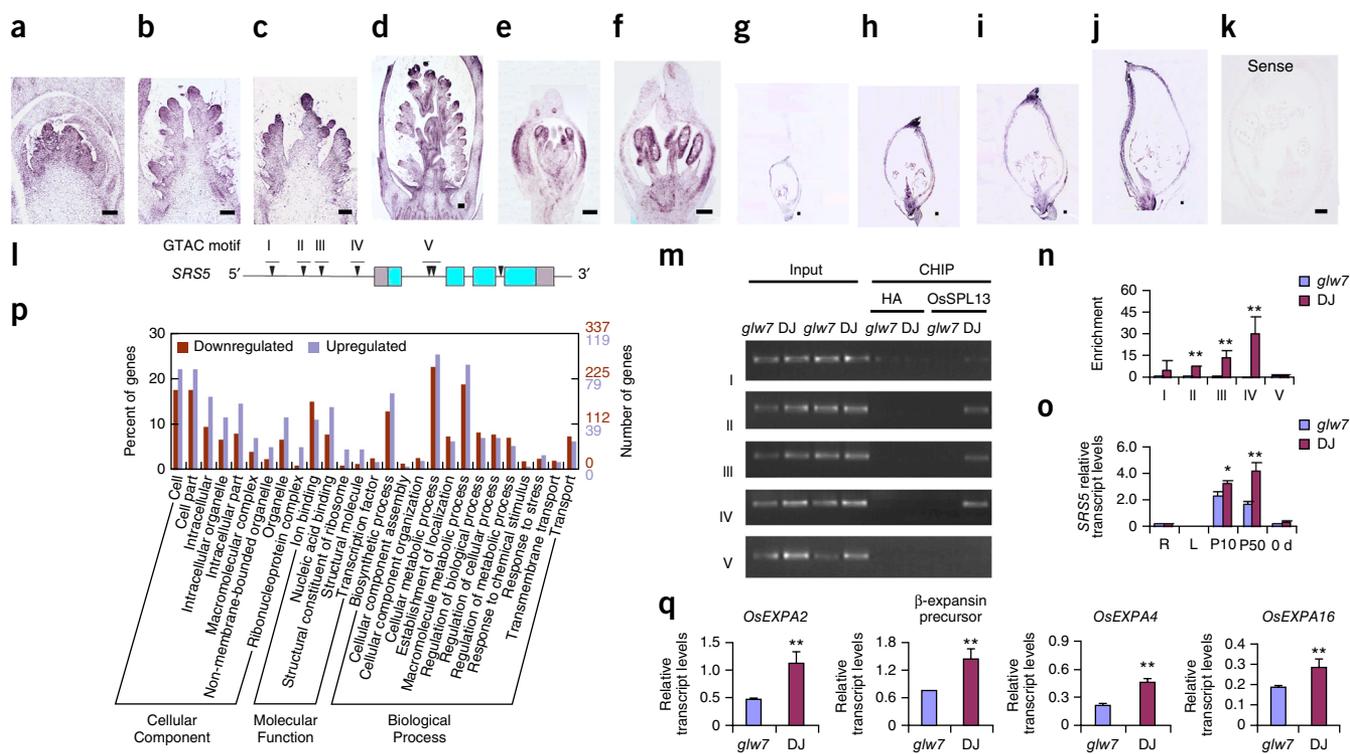
We divided the grain hull into upper, middle and lower parts (Fig. 5c) to measure cell number and cell density for the lemma

in these three locations. Cell number per millimeter along the longitudinal axis was significantly higher in the three parts of the lemma for *glw7* plants than for wild-type Dongjing plants (Fig. 5e). The lemmas of CP-1 plants contained 21%, 9% and 10% fewer cells per millimeter in the upper, middle and lower parts, respectively, than the corresponding sites in wild-type Dongjing (Fig. 5e). The cell density per square millimeter was markedly higher in *glw7* plants than in wild-type Dongjing plants. Consistently, cell density was significantly lower in transgenic plants with constructs CP and III (Fig. 5f). All these results indicate that *OsSPL13* regulates grain shape through activation of cell size regulation machinery.

Cross-sections of central parts of the spikelet showed that the outer parenchyma cell layer in transgenic CP-1 plants was longer and contained substantially larger cells than the corresponding layer in Dongjing and *glw7* plants, with a minor difference in cell number (Fig. 5g–l). Cross-sections of the uppermost internodes also showed *OsSPL13* mainly involved in cell size regulation during organ development (Supplementary Figs. 14a–i and 15a,b).

### OsSPL13 is expressed specifically during floret development

According to rice spikelet developmental stages<sup>29</sup>, *OsSPL13* was detected when the primary- and secondary branch primordia initiated (Fig. 6a–c) and was strongly expressed in floral organ



**Figure 6** *In situ* hybridization of *OsSPL13* in hulls during floret development and identification of the *SRS5* gene as a direct target of *OsSPL13*. (a–k) *In situ* hybridization for *OsSPL13* in primary branch formation (a), secondary branch formation (b), later secondary branch formation (c), the spikelet formation stage (d), floret 1 mm in length (e), floret 1.6 mm in length (f), floret 2.2 mm in length (g), floret 3.1 mm in length (h), floret 4 mm in length (i) and floret 5 mm in length (k). (l) Negative-control hybridization with *OsSPL13* sense probe. Scale bars for a–k, 100  $\mu$ m. (I) Diagram of the *SRS5* genomic region. Triangles indicate GTAC motifs in the promoter and introns. The gray boxes represent the 5' and 3' UTRs of the *SRS5* gene, and the blue boxes represent the coding regions. (m) ChIP analysis of *SRS5* genomic fragments in the panicles of wild-type Dongjing and *glw7* plants. (n) ChIP-qPCR analysis of *SRS5* genomic fragments in the panicles of Dongjing and *glw7* plants. Relative enrichment of fragments was calculated by comparing samples immunoprecipitated with antibodies to *OsSPL13* and HA. (o) *SRS5* expression in panicles from wild-type Dongjing and *glw7* plants. Relative expression levels were calibrated to rice *UBIQUITIN* gene expression. P10 and P50, young panicles with panicle lengths 10% and 50%, respectively, of full panicle length in wild-type Dongjing and *glw7* plants. (p) GO term analysis of the genes downregulated and upregulated in *glw7* panicles relative to wild-type Dongjing panicles from RNA sequencing data. (q) Transcript analysis of *OsEXPA2*,  $\beta$ -expansin precursor, *OsEXPA4* and *OsEXPA16* in wild-type Dongjing and *glw7* plants. Values are represented as means  $\pm$  s.d. (three replicates in n, o and q): \* $P < 0.05$ , \*\* $P < 0.01$ .

**Figure 7** The level of genetic differentiation ( $F_{ST}$ ) across chromosome 7 between different subspecies. **(a)** Scan of genetic differentiation between temperate *japonica* and tropical *japonica*. **(b)** Scan of genetic differentiation between *indica* and tropical *japonica*. In total, 1,040 rice varieties were used for analysis, including 520 *indica*, 409 temperate *japonica* and 75 tropical *japonica*.

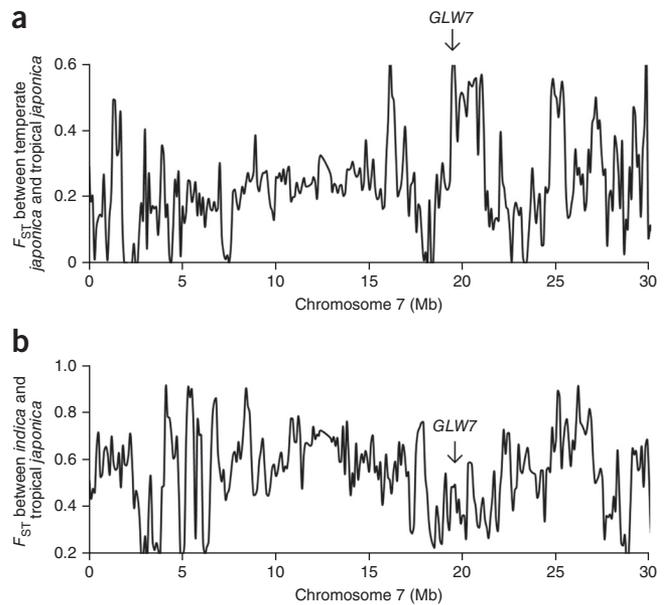
primordia (**Fig. 6d**). *OsSPL13* was highly expressed in the middle of the lemma and palea when florets were about 1 mm in length (**Fig. 6e**), and expression gradually decreased during the growth of the florets and stamen (**Fig. 6f**). When florets reached 2.2 mm in length, only weak *OsSPL13* expression was detected in the hulls of the florets (**Fig. 6g**). Interestingly, *OsSPL13* expression was strong in the apices of the lemma and palea when the hulls were 4–5 mm in length and gradually decreased from apices to the middle of the hull (**Fig. 6h–j**). No signal was detected in the negative control (**Fig. 6k**). These results are consistent with the finding of markedly different cell sizes in the upper part of the lemma in comparison of CP-1 and *glw7* plants.

Bioinformatics analysis indicated that the 3' UTR of *OsSPL13* has an OsmiR156 complementary site<sup>20</sup> (**Supplementary Fig. 3**). Transgenic plants (pDJ::GUS-m*OsSPL13*, with expression driven by the promoter of *OsSPL13* from Dongjing) containing *OsSPL13* with a mutated OsmiR156-targeted site showed high levels of  $\beta$ -glucuronidase (GUS) activity in leaves and in lemma and palea. In contrast, transgenic plants lacking a mutation in the OsmiR156-targeted site (pDJ::GUS-*OsSPL13*) had no GUS activity in leaves, but GUS activity remained high in lemma and palea (**Supplementary Fig. 16a,b,d**). Consistent with the above results, *OsSPL13* transcript levels in transgenic lines overexpressing OsmiR156 were about fourfold lower than those in control plants, whereas overexpression of an OsmiR156 target mimic, MIM156 (ref. 30), resulted in a substantial increase in *OsSPL13* expression in panicles (**Supplementary Fig. 16c**). These results suggest that *OsSPL13* is one of the targets of OsmiR156.

### *OsSPL13* directly targets *SRS5*

The nuclear localization of an *OsSPL13*-GFP fusion protein is consistent with the notion that *OsSPL13* encodes a putative transcription factor (**Supplementary Fig. 16e**)<sup>31</sup>. *OsSPL13* contains a putative bipartite nuclear localization signal at the C-terminal end of the highly conserved SBP domain (**Supplementary Fig. 3**), which is necessary and sufficient for DNA binding and nuclear localization of SPLs according to several previous reports<sup>32–34</sup>.

Chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) analysis showed that *SMALL AND ROUND SEED 5* (*SRS5*) might be a candidate for a direct target of *OsSPL13*. *SRS5* encodes  $\alpha$ -tubulin, and an amino acid substitution (p.Arg308Leu) in *SRS5* reduced seed length by altering the regulation of cell elongation in rice seeds<sup>35</sup>. In *Arabidopsis thaliana* and rice, SBPs were reported to directly target a core binding motif (GTAC). Four GTAC motifs are located 1,873, 1,208, 933 and 143 bp upstream of the *SRS5* transcription start site (TSS), corresponding to motifs I–IV, respectively (**Fig. 6l**). A fifth region (V), in the first exon and containing two GTAC motifs, was included as a negative control. In ChIP analysis using two different antibodies to *OsSPL13*, both antibodies detected no or only weak signal for motifs I and V. Fragments encompassing motifs II–IV were clearly amplified from wild-type Dongjing samples but not from *glw7* samples (**Fig. 6m**). More notably, binding to motifs III (–933 bp) and IV (–143 bp) was significantly enriched by about 10-fold and 37-fold, respectively, in wild-type Dongjing panicles as compared to *glw7* panicles. No enrichment of fragments corresponding to any of the



motifs was detected in negative-control experiments where ChIP was performed with antibody to HA, either in wild-type Dongjing or *glw7* samples (**Fig. 6n**). Further investigation showed that *SRS5* transcript levels were significantly higher in panicles from Dongjing plants and transgenic plants with the CP construct than in *glw7* mutants (**Fig. 6o** and **Supplementary Fig. 17a,b**). These results suggest that *OsSPL13* directly targets *SRS5* and functions as a positive regulator of *SRS5* expression in the regulation of hull cell size in rice.

ChIP-qPCR analysis showed that *OsSPL13* was able to bind the promoter of *DEP1*, a gene known to regulate panicle length and grain number<sup>36</sup>, and upregulate its expression in very young panicles (**Supplementary Fig. 18a–c**). To further address the function of *OsSPL13*, we performed RNA sequencing analysis with panicles from wild-type Dongjing and *glw7* plants. A total of 1,915 differentially expressed genes were detected (FDR  $P < 0.01$ ), of which 25.4% (486 genes) were upregulated and 74.6% (1,429 genes) were downregulated in *glw7* plants as compared to wild-type Dongjing plants (**Supplementary Table 6**). In Gene Ontology (GO) analysis, these genes were significantly ( $P < 0.01$ ) enriched for Cellular Component, Molecular Function and Biological Process GO terms (**Fig. 6p**). We also compared the transcript levels of four expansin-related genes—*OsEXPA2*,  $\beta$ -expansin precursor, *OsEXPA4* and *OsEXPA16*—by qPCR. All four genes had higher expression levels in wild-type Dongjing panicles than in panicles from *glw7* plants (**Fig. 6q**). Collectively, these results support the hypothesis that *OsSPL13* functions as a regulator of cell size.

### Tropical *japonica* *OsSPL13* allele was introgressed from *indica*

Phylogenetic analysis showed that all of the SGH varieties, which are temperate *japonica*, grouped together. The LGH varieties, including most tropical *japonica* and some temperate *japonica*, were clustered with *indica* varieties. Among the LGH varieties, several tropical *japonica* were clustered in a small group (**Supplementary Fig. 19** and **Supplementary Table 7**). The *OsSPL13*<sup>SGH</sup> allele was only detected in the *japonica* group and not in the *indica* group (**Supplementary Fig. 20** and **Supplementary Table 7**). Moreover, we examined *OsSPL13* evolution by estimating the level of population difference ( $F_{ST}$ ) on chromosome 7 between different subspecies. Pairwise measurements of  $F_{ST}$  on chromosome 7 showed that  $F_{ST}$  levels were truly higher between temperate *japonica* and tropical *japonica* in the *OsSPL13* gene; however,

markedly weaker differences were found between tropical *japonica* and *indica* at the *GLW7* locus as compared to the mean  $F_{ST}$  over chromosome 7 as a whole (Fig. 7a,b). Considering the fact that both *indica* and tropical *japonica* varieties had intermixed planting in tropical regions, we propose that the large-grain haplotype of *OsSPL13* in tropical *japonica* varieties was introgressed from *indica* varieties under artificial selection for increased grain size and yield.

### *GLW7* and *GS3* regulate grain shape independently

Previous reports indicated that *qGS3* is a major determinant of grain length in rice<sup>16</sup>. Sequence analysis showed that wild-type Dongjing contains full-length *GS3* and has a short-seed phenotype. When decreasing *GS3* expression using RNAi, we found increased grain length in transgenic lines (Supplementary Fig. 21a,b). We generated  $F_2$  progeny for two hybrid combinations (*GS3*-RNAi-2 crossed with CP-1 plants and *GS3*-RNAi-2 crossed with *glw7* plants). Of the plants analyzed, CP-1 × *GS3*-RNAi-2 plants produced the heaviest and longest grains (Supplementary Fig. 22a–e). Transcript analysis showed that *OsSPL13* RNA levels were not altered in *GS3*-RNAi plants; consistent with this finding, no changes were detected in *GS3* transcript levels in the *glw7* and CP-1 plants (Supplementary Fig. 23a,b). Therefore, *GLW7* and *GS3* likely regulate grain shape independently during seed formation.

### DISCUSSION

Introgression between different rice ecotypes is a key process in evolution and domestication, as it may contribute to increased grain yields and adaptation to new environments<sup>37</sup>. We report here that the large-grain allele of *OsSPL13* in tropical *japonica* rice was introgressed from *indica* varieties. Our results demonstrate that GWAS analysis is a feasible approach to quickly identify new genes associated with complex traits in diverse plants.

*OsSPL13* encodes conserved domains characteristic of the SPL family. Proteins encoded by *SPL* genes have been shown to regulate multiple important and divergent biological processes, and there are 17 *SPL* genes in *Arabidopsis* and 19 *SPL* genes in rice<sup>22</sup>. *IPA1* (*OsSPL14*) has a critical role in plant architecture, and upregulation of *OsSPL14* increases grain number per panicle<sup>23,24</sup>. Our results demonstrate that *OsSPL13* could promote the development of secondary branches to enhance grain number per panicle. The rice *GW8* locus directly binds to *GL7/GW7* and represses its expression<sup>14</sup>. Both *GW8* and *GW7* regulate grain width and grain length through regulation of cell cycle machinery. In contrast, *GLW7* increased grain weight in this study by regulating both grain length and grain thickness, without affecting grain width. Our results show that *GLW7* mainly promotes grain size by increasing cell size. Overall, these findings suggest that *GW8* and *GLW7* may bind to different target genes and have different roles in regulating grain size.

Our ChIP analysis showed that *SRS5* is directly regulated by *OsSPL13*. *SRS5* encodes  $\alpha$ -tubulin<sup>35</sup>. In plants,  $\alpha$ -tubulin and  $\beta$ -tubulin form a heterodimer that is the basic structural unit for microtubules. Alteration of tubulin gene expression affects cell elongation both in plants and yeast<sup>38–40</sup>. An increasing body of evidence has shown that expansins are plant cell wall proteins that can induce cell wall extension and modify cell growth<sup>41–43</sup>. Considering these findings as a whole, both the  $\alpha$ -tubulin and expansin genes are necessary for cell growth and elongation; these results suggest that *OsSPL13* regulates cell size mainly by fine-tuning microtubule and cell wall pathways (Supplementary Fig. 24). Thus, we are able to shape rice grains by pyramiding *OsSPL13* and *GS3* loci in rice breeding.

After *japonica* varieties were first domesticated from wild rice in southern China, they moved northward to become the temperate ecotype; at the same time, they also moved southward to Southeast Asia to become the tropical ecotype<sup>8</sup>. Because *indica* varieties grow throughout the tropics and subtropics, it is possible that the *japonica* varieties were introduced to alien genes from *indica* during the tropical *japonica* breeding process. Although preference for rice grain shape varies with consumer groups, the majority of rice consumers in Southeast Asia prefer long grains<sup>44</sup>. This may be another reason that the large-grain haplotype of *OsSPL13* was selected in tropical *japonica* varieties. *OsSPL13* can substantially improve grain size, as well as grain number and panicle size, by increasing the number of secondary branches and panicle length. These results indicate that *OsSPL13* might have been a critical factor in the divergence of tropical *japonica* and temperate *japonica*, thus providing the opportunity to improve grain size and grain yield for a majority of temperate *japonica* varieties by introducing the large-grain *OsSPL13*<sup>LGH</sup> allele when breeding new elite varieties.

### METHODS

Methods and any associated references are available in the online version of the paper.

**Accession codes.** Full-length *OsSPL13* cDNA from Dongjing, [KT899581](#); full-length *OsSPL13* cDNA from GP7, [KT899582](#). Microarray data are available under BioProject accession [PRJEB11401](#). *OsSPL13* genomic sequences are available under accessions [LT159866](#)–[LT159966](#).

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

### ACKNOWLEDGMENTS

We thank J. Wang and W. Cai for help in the construction of the MIM156 vector, K. Chong (Institute of Botany, Chinese Academy of Sciences) for providing the pTCK303 vector, P. Hu and G. Jiao for the measurements of grain quality, and J. Li, Xiaoyan Gao and Xiaoshu Gao for microscopy experiments. This work was supported by the Chinese Academy of Sciences (XDA08020101), the National Natural Science Foundation of China (91535202 and 31421093) and the Ministry of Science and Technology of China (2013CBA01404).

### AUTHOR CONTRIBUTIONS

B.H. and L.S. designed and conducted the research. L.S., J.C., J.L., Q.H., T.Z., J.Z., Y.S., E.C., C.G., Y.J., Y.W., D.F., Y. Lu, Q.W., Q. Zhan, L.C. and Y.W. performed research. L.S., X.H., H.G., T.L., Y.Z., Y. Li, K.L. and Q. Zhao analyzed data. X.W. collected rice cultivars. K.A. and G.A. performed the generation of mutant rice. L.S. and B.H. wrote the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Godfray, H.C. *et al.* Food security: the challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010).
- Huang, X. & Han, B. Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **65**, 531–551 (2014).
- Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
- Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
- Tan, L. *et al.* Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **40**, 1360–1364 (2008).
- Jin, J. *et al.* Genetic control of rice plant architecture under domestication. *Nat. Genet.* **40**, 1365–1369 (2008).
- Zhu, B.F. *et al.* Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* **155**, 1301–1311 (2011).

8. Khush, G.S. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**, 25–34 (1997).
9. Ali, M.L. *et al.* A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Sci.* **51**, 2021–2035 (2011).
10. Song, X.J., Huang, W., Shi, M., Zhu, M.Z. & Lin, H.X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–630 (2007).
11. Shomura, A. *et al.* Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40**, 1023–1028 (2008).
12. Li, Y. *et al.* Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* **43**, 1266–1269 (2011).
13. Wang, Y. *et al.* Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat. Genet.* **47**, 944–948 (2015).
14. Wang, S. *et al.* The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* **47**, 949–954 (2015).
15. Wang, S. *et al.* Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* **44**, 950–954 (2012).
16. Mao, H. *et al.* Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. USA* **107**, 19579–19584 (2010).
17. Ben-Haj-Salah, H. & Tardieu, F. Temperature affects expansion rate of maize leaves without change in spatial distribution of cell length (analysis of the coordination between cell division and cell expansion). *Plant Physiol.* **109**, 861–870 (1995).
18. Huang, X. *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–967 (2010).
19. Huang, X. *et al.* Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39 (2012).
20. Huang, X. *et al.* Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* **6**, 6258 (2015).
21. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).
22. Xie, K., Wu, C. & Xiong, L. Genomic organization, differential expression, and interaction of *SQUAMOSA* promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* **142**, 280–293 (2006).
23. Jiao, Y. *et al.* Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nat. Genet.* **42**, 541–544 (2010).
24. Miura, K. *et al.* *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* **42**, 545–549 (2010).
25. Wang, J.W., Czech, B. & Weigel, D. *miR156*-regulated *SPL* transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* **138**, 738–749 (2009).
26. Yu, S. *et al.* Gibberellin regulates the *Arabidopsis* floral transition through *miR156*-targeted *SQUAMOSA* PROMOTER BINDING-LIKE transcription factors. *Plant Cell* **24**, 3320–3332 (2012).
27. Xing, S., Salinas, M., Höhmann, S., Berndtgen, R. & Huijser, P. *miR156*-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in *Arabidopsis*. *Plant Cell* **22**, 3935–3950 (2010).
28. Jeon, J.S. *et al.* T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.* **22**, 561–570 (2000).
29. Itoh, J. *et al.* Rice plant development: from zygote to spikelet. *Plant Cell Physiol.* **46**, 23–47 (2005).
30. Franco-Zorrilla, J.M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**, 1033–1037 (2007).
31. Lu, G., Gao, C., Zheng, X. & Han, B. Identification of *OsbZIP72* as a positive regulator of ABA response and drought tolerance in rice. *Planta* **229**, 605–615 (2009).
32. Birkenbihl, R.P., Jach, G., Saedler, H. & Huijser, P. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J. Mol. Biol.* **352**, 585–596 (2005).
33. Klein, J., Saedler, H. & Huijser, P. A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. *Mol. Gen. Genet.* **250**, 7–16 (1996).
34. Yamasaki, K. *et al.* A novel zinc-binding motif revealed by solution structures of DNA-binding domains of *Arabidopsis* SBP-family transcription factors. *J. Mol. Biol.* **337**, 49–63 (2004).
35. Segami, S. *et al.* *Small and round seed 5* gene encodes  $\alpha$ -tubulin regulating seed cell elongation in rice. *Rice (N.Y.)* **5**, 4 (2012).
36. Huang, X. *et al.* Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat. Genet.* **41**, 494–497 (2009).
37. McNally, K.L. *et al.* Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* **106**, 12273–12278 (2009).
38. Catterou, M. *et al.* Brassinosteroids, microtubules and cell elongation in *Arabidopsis thaliana*. II. Effects of brassinosteroids on microtubules and cell elongation in the *bul1* mutant. *Planta* **212**, 673–683 (2001).
39. Ji, S. *et al.* A  $\beta$ -tubulin-like cDNA expressed specifically in elongating cotton fibers induces longitudinal growth of fission yeast. *Biochem. Biophys. Res. Commun.* **296**, 1245–1250 (2002).
40. Lee, Y. & Kende, H. Expression of  $\beta$ -expansins is correlated with internodal elongation in deepwater rice. *Plant Physiol.* **127**, 645–654 (2001).
41. Cho, H.T. & Kende, H. Expression of expansin genes is correlated with growth in deepwater rice. *Plant Cell* **9**, 1661–1671 (1997).
42. Park, C.H. *et al.* Brassinosteroids control *AtEXPA5* gene expression in *Arabidopsis thaliana*. *Phytochemistry* **71**, 380–387 (2010).
43. Zenoni, S. *et al.* Overexpression of *PhEXPA1* increases cell size, modifies cell wall polymer composition and affects the timing of axillary meristem development in *Petunia hybrida*. *New Phytol.* **191**, 662–677 (2011).
44. Calingacion, M. *et al.* Diversity of global rice markets and the science required for consumer-targeted rice breeding. *PLoS One* **9**, e85106 (2014).

## ONLINE METHODS

**Materials.** Approximately 36 seeds for each accession from a collection of 342 accessions were germinated and planted in an experimental field (Hangzhou, China; 30.32° N, 120.12° E). GWAS analysis of seed length was performed according to Huang *et al.*<sup>19</sup>. The *OsSPL13* T-DNA insertion line (1B-14102) in the *japonica* Dongjing background was obtained from Kyung Hee University, Korea<sup>28,45</sup>, and confirmed using PCR with the primers listed in **Supplementary Table 8**. The 74 varieties of cultivated rice (*O. sativa*) and the 30 accessions of wild rice used in the phylogenetic analysis are listed in **Supplementary Table 7**.

**GWAS analysis.** In total, we used 493,777 SNPs with a minor allele frequency of >0.05 in the *japonica* population for association analysis. Association analysis using a mixed model was performed with the EMMAX software package to account for sample structure. The matrix of pairwise genetic distances derived from simple matching coefficients was used as the variance-covariance matrix for random effect, and the first ten principal components were included as fixed effects. Principal-component analysis of the population was performed using EIGENSOFT software. With the number of SNPs analyzed ( $n = 493,777$ ), the threshold for significance was estimated to be approximately  $P = 1 \times 10^{-7}$  (that is, 0.05/493,777) by the Bonferroni correction method.

**Primers.** The sequences for all primers used in this study are listed in **Supplementary Table 8**.

**Gene cloning, construct generation and transformation.** The 8-kb genomic region including the 2.4-kb promoter and coding sequence for *OsSPL13* was amplified from Dongjing and cloned into plant binary vector pCAMBIA 2300 to generate construct I. Construct CP contains the 8-kb genomic fragment from large-grain variety GP7. The 2.4-kb promoter region from Dongjing was fused with the transcript region (including the 5' UTR, coding region and 3' UTR) from GP7 to generate construct III. Control plants were generated by introducing the empty vector into wild-type Dongjing by *Agrobacterium tumefaciens*-mediated transformation as described previously<sup>46</sup>. For the pDJ::GUS-*OsSPL13*<sup>DJ</sup> and pDJ::GUS-*mOsSPL13*<sup>DJ</sup> constructs, the *GUS* gene was fused to *OsSPL13* with a normal or mutated microRNA target site, respectively, with expression driven by the Dongjing *OjSPL13* promoter. The RNAi construct contained an inverted repeat harboring the 491-bp *OsSPL13* cDNA fragment in vector pTCK303 (ref. 46). All plasmid constructs were introduced into *A. tumefaciens* strain EHA105 and subsequently transferred into the small-grain *japonica* variety Dongjing<sup>45</sup>.

The rice gene *OsmiR156* was PCR amplified and subcloned into the binary vector pTCK303 (ref. 46). The MIM156 construct was generated as described by Franco-Zorrilla *et al.*<sup>30</sup>; the expression of both constructs was controlled by the same *UBIQUITIN* promoter.

**Trait measurements.** Grain length, width and thickness and 1,000-grain weight were measured when plants were completely matured. The entire spikelet hull, including the lemma and palea, was included in these measurements. To determine fresh weight during different stages, 100 grains were weighed, and measurements were repeated six times with the same 100 grains. After the seeds were dried at 80 °C for 3 d, dry weight was measured. The entire spikelet hull was included in these measurements.

For plot yield analysis, a total of 96 plants for each line were grown in a paddy field with three replicates, and 60 plants were collected for yield measurement. Plant height, tiller number, panicle length and grain number were measured and analyzed.

**Subcellular localization and histological observation.** The coding sequence of *OsSPL13* from Dongjing was fused with the *GFP* gene to generate the *GFP-OsSPL13* construct. The fusion protein was transcribed from the cauliflower mosaic virus (CaMV) 35S promoter. Next, the *GFP-OsSPL13* construct was bombarded into onion epidermis with the PDS-1000/He biolistic system (Bio-Rad). The subcellular localization and colocalization of these proteins was evaluated using a confocal laser scanning microscope (FluoView FV1000, Olympus).

The outer surface of lemma from mature seeds was observed with a scanning electron microscope at an acceleration voltage of 15 kV. Cell number and cell density were calculated along the longitudinal axis.

**RNA *in situ* hybridization and measurement of GUS activity.** Panicles and florets from wild-type Dongjing plants were fixed in 4% paraformaldehyde and embedded in Paraplast. A 792-bp region of *OsSPL13* cDNA was used as the probe, which was labeled by digoxigenin (Roche). RNA *in situ* hybridization was performed on 8- $\mu$ m sections; sense probe was used as a negative control<sup>47</sup>.

To measure GUS activity, rice samples were incubated in a solution of 1 mM 5-bromo-4-chloro-3-indolyl- $\beta$ -D-glucuronic acid (X-gluc), 10% methanol, 0.5% Triton X-100 and 50 mM NaPO<sub>4</sub> at 37 °C in the dark for 12 h. After GUS staining, chlorophyll was removed using 75% ethanol before taking pictures.

**Antibody generation and protein blot analysis.** *OsSPL13* cDNA was subcloned into pET28a and pMAL-C2. Recombinant His-*OsSPL13* protein was expressed in BL21 *E. coli* and purified using Ni-NTA agarose beads under native conditions. A total of 5 mg of purified His-*OsSPL13* per rabbit was used to raise polyclonal antibody. Maltose-binding protein (MBP)-*OsSPL13* protein was purified by amylose resin column (New England BioLabs, 8021S), and antibody to *OsSPL13* was then purified by SulfoLink affinity column (Pierce, 20401) coupled to purified MBP-*OsSPL13* protein. The specificity of the antibody to *OsSPL13* was determined by protein blot analysis.

For protein blot analysis, plant materials were ground into powder in liquid nitrogen, and protein extraction buffer was added. Samples were boiled and centrifuged, and the supernatants were resolved by SDS-PAGE. Protein blot analysis was performed according to a general procedure. Antibody to *OsSPL13* (1:4,000 dilution) and commercial antibody to actin (Earth Ox, E021080; 1:2,500 dilution) were used to detect *OsSPL13* and actin protein levels, respectively.

**RNA sequencing and RT-PCR.** Panicles from Dongjing and *japonica* varieties were collected when they reached about 30–60% of their full length. RNA was extracted from the roots, leaves and panicles of 20 varieties that were selected on the basis of the GWAS data. Spikelet tissues included the palea and lemma. Total RNA was isolated using TRIzol reagent (Invitrogen), and RNase-free DNase I treatment was performed to remove any genomic DNA contamination. For transcriptome analysis, panicles from wild-type Dongjing and *glw7* plants when they reached 30–60% of their full length were used for library construction and sequencing according to the instructions from Illumina. Pearson's  $\chi^2$  was applied to assess the differences between samples. The *P* value was computed for each gene<sup>48,49</sup>.

RT-PCR was performed on an ABI 7500 instrument using SYBR Green PCR Mastermix according to the manufacturer's instructions (Takara, RR420). The cycling conditions included incubation for 30 s at 95 °C followed by 40 cycles of amplification (95 °C for 5 s and 60 °C for 31 s). The *UBIQUITIN* gene was used as an endogenous control to normalize detected gene expression. The primers used in qRT-PCR are listed in **Supplementary Table 8**.

**ChIP-qPCR analysis.** Panicles from wild-type Dongjing and *glw7* plants were used as materials for ChIP assays, which were performed according to a previously described protocol<sup>48</sup>. Briefly, panicles were cross-linked with 1% formaldehyde under vacuum for 10 min and ground in liquid nitrogen, and nuclei were isolated. DNA was sonicated into 200- to 500-bp fragments. Chromatin complexes were incubated with polyclonal antibody to *OsSPL13*, and immune complexes were pulled down using Protein A beads. The precipitated DNA was purified and dissolved in water for further analysis. Commercial polyclonal antibody to HA (Santa Cruz Biotechnology, sc-305) was used as a negative control.

45. Jeong, D.H. *et al.* Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice. *Plant J.* **45**, 123–132 (2006).
46. Wang, Z. *et al.* A practical vector for efficient knockdown of gene expression in rice (*Oryza sativa* L.). *Plant Mol. Biol. Rep.* **22**, 409–417 (2004).
47. Luo, D., Carpenter, R., Vincent, C., Cosey, L. & Coen, E. Origin of floral asymmetry in *Antirrhinum*. *Nature* **383**, 794–799 (1996).
48. Saleh, A., Alvarez-Venegas, R. & Avramova, Z. An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in *Arabidopsis* plants. *Nat. Protoc.* **3**, 1018–1025 (2008).
49. Lu, T. *et al.* Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* **20**, 1238–1249 (2010).