Prediction of environmental response in field-grown rice using expression-dynamics-QTL

Makoto Kashima[1], Ryota L. Sakamoto[2], Hiroki Saito[3,4], Satoshi Ohkubo[3,5], Ayumi Tezuka[1], Ayumi Deguchi[1], Yoichi Hashida[1], Yuko Kurita[1] & Atsushi J. Nagano[6]*

[1] Research Institute for Food and Agriculture, Ryukoku University, Shiga Japan

[2] Seibi Senior High School, Gifu, Japan

[3] Graduate school of Agriculture, Kyoto University, Kyoto, Japan

[4] Tropical Agriculture Research Front, Japan International Research Center for Agricultural Sciences, Okinawa, Japan

[5] Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan

[6] Faculty of Agriculture, Ryukoku University, Shiga, Japan

*Corresponding author: Atsushi J. Nagano (email: anagano@agr.ryukoku.ac.jp)

Organisms respond to the fluctuations in field environments differently throughout their lives depending on their genetic variations. However, how such genetic variations affect environmental responses of organisms under field conditions, that is, the genotype × environment interaction (G×E), remains unclear. In the present study, we developed a method to identify the loci influencing gene expression dynamics (GED) in field-grown rice. We cultivated chromosome segment substitution lines (CSSLs) from Koshihikari and Takanari rice varieties[1] in the field, and conducted approximately 900 time-series RNA sequencing (RNA-Seq) (bihourly-sampling for 16 d during 4 months). This novel analytical approach identified more than one thousand expression dynamics quantitative trait loci (edQTLs), which cause differences in GED between Koshihikari and Takanari. Based on the edQTLs, we successfully predicted the GED under environments different from training environments and in rice cultivars with more complex genotypes than the CSSLs. Results demonstrated that integrating genotype and environment information via the transcriptome can help predict the environmental response of GED under field conditions. We anticipate that edQTLs identification will help understand the G×E of organisms under natural conditions, which is difficult to assess with laboratory equipments[2]. As genetic variation and

environmental conditions converge on transcriptome and transcriptome-based trait prediction functions effectively[3,4], the prediction of GED based on genetic variations and environmental information will contribute to crop breeding by increasing the accuracy of trait prediction under diverse conditions.

To understand the responses of different plant lines under various conditions, it is essential to elucidate G×E. Although the expression quantitative trait locus (eQTL) approach is useful to assess the linkage between genetic variation and transcriptome, it requires transcriptome data for the complete set of a QTL mapping population under given conditions[5–9]. Due to high cost and labor, this approach can cover G×E under limited conditions. Contrarily, mathematical models have succeeded in predicting GED under fluctuating field conditions based on meteorological information[10–12], although only predicting transcriptome dynamics in genotypes used as training data[10–12]. Thus, no study has clarified the relationship between GED and G×E.

Rice (*Oryza sativa* L.) is a major food crop[13–15]. The CSSLs and backcross inbred lines (BILs) between Takanari (a high-yield indica variety)[1,16] and Koshihikari (a leading variety in Japan) are available[1,17], and useful for examining G×E. To understand the broad range of G×E in field-grown rice (Fig. 1a), we conducted mathematical analysis of time-series RNA-Seq of Koshihikari, Takanari, and their reciprocal CSSLs

53 under field conditions to identify edQTLs and predict transcriptome dynamics under
54 field conditions. Our approach consisted of four steps (See Methods and Fig. 1b). We
55 first cultivated 78 reciprocal CSSLs between Koshihikari and Takanari[1], in which a part
56 of a chromosome was partially substituted with the other parental chromosome, and the
57 two parental lines (Fig. 1b). Each substituted fragment was categorized into Koshihikari
58 and Takanari haplotypes using 141 DNA markers (Fig. 1b). To consider the effects of
59 developmental stage, we transplanted seedlings on four different days (Fig. 1c), and 16
60 sets of bihourly sampling for 24 h were conducted from May to September 2015
61 (cropping season in Japan). Two genotypes were collected from each transplant set per
62 sampling time point along corresponding local meteorological data (air temperature and
63 global solar radiation) (Fig. 1c). The youngest fully expanded leaf from individual
64 plants was used for RNA-Seq, and transcriptome data was obtained for 926 samples.
65 After filtering samples and genes (Extended Data Fig. 1a, b) and confirming genotypes
66 (Extended Data Fig. 1c), 23,924 expressed genes (EGs) from 854 samples were used for
67 edQTL detection. Correlation analysis of the transcriptome data showed an obvious
68 transcriptome-wide diurnal variation in EGs, albeit genotype variation (Fig. 1d).
69 Second, because lines were not sampled simultaneously, it was difficult to directly
70 compare expression between genotypes at each time point, and we therefore constructed

71 predictive models for GED in Koshihikari and Takanari. As most of the genome in the
72 CSSLs is not substituted, GED of most genes is expected to be the same as in a
73 background genotype. Thus, the RNA-Seq data of Koshihikari- and
74 Takanari-background (BG) samples were used to construct BG models in which GED of
75 Koshihikari and Takanari could be predicted based on meteorological and circadian
76 clock data (Extended Data Fig. 2a). In addition, "scaled age" was applied to adjust
77 differences in heading date among rice genotypes (Extended Data Fig. 2). A comparison
78 of the prediction by the two BG models showed expression dynamics polymorphism
79 (EDP) in 3696 genes (15.4% of the EGs; Extended Data Fig. 1e). For example, the
80 predicted GED of *Os09g0343200* differed between the two BG models. Notably, in
81 some samples, *Os09g0343200* expression appeared to be similar to the prediction of the
82 other BG model, rather than their own BG model (Fig. 2a), suggesting that chromosome
83 substitution affects *Os09g0343200* expression in such samples.

84 Third, to detect edQTLs, we then evaluated the improvement in the residual error of
85 each gene by assuming the existence of edQTLs around each marker (Extended Data
86 Fig. 3a). In *Os09g0343200*, the sum of residual errors (SRE) was only significantly
87 improved assuming edQTLs on chromosome 9 (Fig. 2b and Extended Data Fig. 3b, c),
88 indicating that *Os09g0343200* edQTLs are cis-edQTLs. As expected, edQTL-based

89 prediction of *Os09g0343200* was better than BG-based prediction (Fig. 2c). Overall,
90 edQTLs were identified for 1675 genes, including 222 genes affected by *trans*-edQTLs
91 (Fig. 2d) and 43 genes affected by multiple edQTLs. For 33 genes, the SRE of GED
92 based on all edQTLs was smaller than that based on the most significant edQTL only
93 (Extended Data Fig. 4), suggesting accuracy is higher when considering all edQTLs.

94 To assess if our approach could detect the edQTLs affecting GED at a specific time
95 of day and developmental stage, we calculated the variation in median expression level
96 of each gene with time and scaled age (Extended Data Fig. 5a). The edQTLs affected
97 393 fluctuating genes depending on the time of day (standard deviation of median
98 expression at each hour in either BG model was > 1) and 275 genes depending on
99 scaled age (standard deviation of median expression in every 10 scaled age in either BG
100 model was > 1) (Extended Data Fig. 5). For example, the GED of *Os01g0537250*
101 depended on time in both BG models but on scaled age only in the Koshihikari-BG
102 model, which was caused by the *cis*-edQTL (Fig. 2e, f and Extended Data Fig. 5, 6).
103 The expression level of *Os01g0537250* in the Koshihikari-BG model differed from that
104 in the Takanari-BG model during the early developmental stages, but not during the late
105 stage (Fig. 2e). Thus, the edQTL model can only explain the difference in GED under
106 limited conditions.

107 Fourth, we then evaluated the predictive performance of the edQTL model in
108 different environments from where training data were obtained for model construction.
109 Koshihikari, Takanari, and the CSSLs of two different transplant sets were cultivated in
110 a field different from that used in 2015 and sampled in August 2016; 139 RNA-Seq
111 datasets were obtained. Environmental factors differed between the 2015 and 2016
112 fields, especially on the third sampling date (Fig. 3a). The prediction of 1537 genes
113 among the 1675 genes influenced by edQTL was improved by the edQTL model
114 compared with BG models (Fig. 3b, c and Extended Data Fig. 7). Furthermore, the
115 prediction of EGs for the validation dataset showed comparable accuracy to that for the
116 training data set (Extended Data Fig. 8). Thus, our edQTL-based approach can predict
117 GED in different years and locations.

118 Finally, GED was predicted for more complex genotypes than CSSLs. For this
119 purpose, two BILs between Koshihikari and Takanari (HP-a and HP-b)[18] were
120 cultivated and sampled from June to September 2015, and five and six RNA-Seq
121 datasets were obtained for HP-a and HP-b, respectively. Although HP-a and HP-b are
122 Takanari-BG lines, 16.3% and 19.9% of their markers were Koshihikari alleles (Fig. 4a),
123 respectively. As the genotype of edQTL for *Os12g0406000* presented the Koshihikari
124 allele in the genome of HP-a and the Takanari allele in the genome of HP-b (Fig. 4a),

125 Koshihikari- and Takanari-BG models were used to predict *Os12g0406000* expression
126 in HP-a and HP-b edQTL models, respectively. The GED of *Os12g0406000* fluctuated
127 with time and it was higher in Koshihikari than in Takanari (Extended Data Fig. 9a).
128 The edQTL-based prediction of *Os12g0406000* expression was better than the
129 BG-based prediction (Fig. 4b). Regarding *OsKS3* (*Os04g0611700*)[19] in HP-b, the
130 edQTL-based prediction was worse than the BG-based prediction (Extended Data Fig.
131 9b, c). Probably, true edQTLs do not exist around the substituted DNA markers in HP-b
132 (Extended Data Fig. 9b). Because, in CSSLs, several flanking markers are linked to the
133 substituted regions of genome, with respect to some genes, the resolution of edQTL
134 detection in CSSLs was not sufficient for BILs. To evaluate the predictive performance
135 of the edQTL model, we calculated the sum of prediction errors of all
136 edQTL-influenced genes using edQTL, Koshihikari-BG, and Takanari-BG models.
137 Overall, edQTL-based prediction was best (Fig. 4c, Extended Data Fig. 10).
138 Permutation analysis of markers in HP-a and HP-b genomes showed the significant
139 advantage of edQTL-based prediction ($p < 0.001$) (Fig. 4c, Extended Data Fig. 10).
140 Thus, we succeeded in predicting environmental response by integrating genotype and
141 environment information.

142 There is a complementary relationship between edQTL and eQTL approaches
143 regarding sensitivity and environmental conditions (Fig. 1a). Although eQTLs in all
144 samples showed the same influence on gene expression at a sampling time point, some
145 edQTLs showed different influences depending on their environments. Generally,
146 *trans*-QTLs are involved in regulating sets of genes under a specific environment and/or
147 at a specific developmental stage. Therefore, the eQTL approach might be more
148 sensitive for *trans*-QTL than the edQTL approach. This is reflected by the observed
149 differences in the ratios of *trans*-QTLs to all QTLs (13.3% for edQTL approach; 62–
150 71% for eQTL approach[6,8,9]).

151 Phenotypic plasticity via G×E has key roles in the adaptation of plants to field
152 environment[20,21]. Nevertheless, due to fluctuating environmental factors, the genetic
153 architecture underlying G×E and phenotypic plasticity under field conditions remains
154 largely unexplored. By identifying thousands of edQTLs, our novel method provides a
155 comprehensive picture of the genetic architecture underlying G×E in Koshihikari and
156 Takanari in field. In addition, our method demonstrated that rice GED can be predicted
157 based on genotypic and meteorological data. In crop breeding, genetic variations in
158 response to environment result in unexpected performance of a bred crop under
159 environments differing from experimental fields. Considering that transcriptome-based

160 trait prediction functioned effectively in recent studies[3,4], our approach will contribute
161 for crop breeding by actualizing transcriptome prediction under various conditions.
162
163 METHODS
164 **Overview of our approach to identify edQTLs and predict GED.** Our approach
165 consisted of four steps (Fig. 1b). First, rice lines were cultivated in a paddy field and
166 sparsely sampled at several time points (environments) for RNA-Seq. Second,
167 predictive models were constructed for GED of the parental lines based on
168 environmental information (BG-based prediction). Third, the dependency of EDP on
169 genetic variation was statistically evaluated based on the comparison between RNA-Seq
170 data and prediction results (edQTL detection). Fourth, by integrating BG-based
171 prediction and edQTL information, GED was predicted based on environmental
172 information and genome-wide genotype data (edQTL-based prediction).
173 **Biological samples**. The japonica variety Koshihikari, the indica variety Takanari, and
174 78 CSSLs (40 Koshihikari BG lines, except for SL1213 plus 38 Takanari BG lines,
175 except for SL1306) were used[1]. Each variety and line was sown in nursery trays. After
176 approximately one month from sowing, seedlings were transplanted to a paddy field at
177 Takatsuki, Osaka, Japan (34°51′19″N, 135°37′51″E) in 2015 and at Kizugawa, Kyoto,

178 Japan (34°4′05″N, 135°50′33″E) in 2016. To integrate the effects of developmental
179 stages (age) into the prediction of GED with the R package 'FIT'[10], four transplant sets
180 of the 80 genotypes were prepared (Fig. 1c). Seed sowing date, transplanting date, and
181 sampling time for each sample are listed in
182 https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-prediction/Os2
183 015/Os2015_input/SampleAttribute_2015_Takatsuki.txt and
184 https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-prediction/Os2
185 016/Os2016_input/SampleAttribute_2016_Kizu.txt. Briefly, seed sowing and
186 transplanting were conducted according to the following schedules (Fig. 1c). In 2015,
187 Transplant set 1: Seed sowing date "2015-04-03", Transplantation date "2015-05-01";
188 Transplant set 2: Seed sowing date "2015-04-17", Transplantation date "2015-05-08";
189 Transplant set 3: Seed sowing date "2015-05-01", Transplantation date "2015-05-22";
190 Transplant set 4: Seed sowing date "2015-05-15", Transplantation date "2015-06-05". In
191 2016, Transplant set 1: Seed sowing date "2016-04-21", Transplantation date
192 "2016-05-16"; Transplant set 2: Seed sowing date "2016-05-19", Transplantation date
193 "2016-06-08".

194 **Sampling and RNA extraction**. Sixteen sets (2015) and three sets (2016) of bihourly
195 sampling for 24 h were conducted on the following dates: 5–6 May, 16–17 June, 22–23

196 June, 29–30 June, 6–7 July, 13–14 July, 20–21 July, 27–28 July, 3–4 August, 19–20
197 August, 24–25 August, 31 August to 1 September, 7–8 September, 14–15 September,
198 21–22 September, and 28–29 September in 2015; and 4–5, 11–12, and 25–26 August in
199 2016. Two plants from the 80 genotypes of each transplant set were randomly selected
200 at each sampling point (Fig. 1c). The youngest fully expanded leaf from each plant was
201 collected, immediately frozen in liquid nitrogen, and stored at -80 °C until RNA
202 isolation for RNA-Seq. Individual plants were only sampled once. Thus, RNA-Seq data
203 derived from independent plants. Frozen samples were homogenized with TissueLyser
204 II (Qiagen, Hilden, Germany), and total RNA was then extracted using the Maxwell 16
205 LEV Plant RNA Kit (Promega, Tokyo, Japan) and Maxwell 16 Automated Purification
206 System (Promega, Tokyo, Japan). The concentration of RNA was measured using the
207 Quant-iT RNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA).
208 **RNA-Seq library preparation and sequencing**. One microgram of total RNA was
209 used to prepare each RNA-Seq library. An automated liquid handling system Freedom
210 EVO 150 (TECAN, Zurich, Switzerland) and a thermal cycler ODTC 384 (INHECO,
211 Martinsried, Germany) were utilized to prepare the RNA-Seq library of 384 samples
212 simultaneously. The protocol of RNA-Seq library preparation is described below.
213 Aliquots of 252 and 192 kinds of 100-μM 60-mer antisense DNA (IDT, San Jose, CA,

214 USA) covering rRNA and the top eight most frequently-detected transcripts
215 (*Osp1g00110.1*, *Osp1g00180.1*, *Osp1g00420.1*, *Osp1g00170.1*, *Osp1g00340.1*,
216 *Osp1g00330.1*, *Osp1g00600.1*, and *Os11t0707000-02*) in RNA-Seq of *O. sativa* were
217 mixed in equal amounts (SDRNA oligo pool)[22,23]. One microgram of total RNA, 1.5 μL
218 of the SDRNA oligo pool, 3 μL of 5× hybridization buffer (0.5 M Tris-HCl (pH 7.4)
219 and 1 M NaCl), and RNase-free water were mixed to a total volume of 10 μL. The
220 mixture was incubated to anneal the antisense DNA with the target transcripts using the
221 thermal cycler ODTC 384 with the following program: 95 °C for 2 min, slow cooled to
222 45 °C (0.1 °C/s), followed by 45 °C for 5 min. Subsequently, 1 μL of Tli Thermostable
223 RNase H (TaKaRa, Kusatsu, Japan), 2 μL of 10× RNase H digestion buffer [500 mM
224 Tris-HCl (pH 7.4), 1 M NaCl, and 200 mM MgCl], and 7 μL of RNase-free water were
225 added and mixed well. This mixture was incubated at 45 °C for 30 min to selectively
226 digest the RNA of the DNA/RNA hybrid double strand. Then, 24 μL of AMpure XP
227 beads (Beckman Coulter, Brea, CA, USA) was added and the mixture was purified in a
228 384-well magnetic plate following the manufacturer's instructions. The RNA was then
229 eluted with 14 μL of RNase-free water. Ten microliters of RNA-containing water, 1 μL
230 of DNaseI (5 unit/μL) (Promega, Tokyo, Japan), 1 μL of 10× DNaseI buffer (Promega,
231 Tokyo, Japan), 1 μL of 100 mM DTT (Invitrogen, Carlsbad, CA, USA), and 7 μL of

232 RNase-free water were mixed well. This mixture was incubated at 37 °C for 30 min to
233 degrade the SDRNA oligo pool. Subsequently, 24 μL of AMpure XP beads (Beckman
234 Coulter, Brea, CA, USA) was added and purified according to the manufacturer's
235 instructions. The RNA was eluted with 14 μL of RNase-free water. Five microliters of
236 the purified mRNA obtained was mixed with 4 μL of 5× SS buffer (Invitrogen,
237 Carlsbad, CA, USA), and 1 μL of 100 mM DTT (Invitrogen, Carlsbad, CA, USA).
238 Fragmentation of mRNA was carried out at 94 °C for 4.5 min and immediately cooled
239 on ice. Subsequently, 0.6 μL of 100 μM random primer $(N)_6$ (Promega, Tokyo, Japan)
240 and 0.9 μL of RNase-free water were added to the mixture, which was incubated at
241 50 °C for 5 min and immediately chilled on ice to relax the secondary structure of the
242 RNA. The fragmented RNA with random hexamer and the reverse transcription master
243 mix [1 μL of 100 mM DTT, 0.4 μL of dNTPs (25 mM each; Promega, Tokyo, Japan),
244 0.1 μL of SuperScript (Invitrogen, Carlsbad, CA, USA), 0.2 μL of actinomycin D (1000
245 ng/μL) (Nacalai Tesque, Kyoto, Japan), and 5.9 μL of RNase-free water] were mixed.
246 For the reverse transcription step, the mixture was incubated at 25 °C for 10 min,
247 followed by 50 min at 50 °C. SuperScript was inactivated by heating the mixture at
248 75 °C for 15 min. Subsequently, 24 μL of AMPure XP (Beckman Coulter, Brea, CA,
249 USA) and 12 μL of 99.5% ethanol were added, and the mixture was purified according

250 to the manufacturer's protocol. The reverse transcription product was eluted with 14 μL
251 of RNase-free water. The purified DNA/RNA hybrid solution without beads and the
252 second strand synthesis master mix [2 μL of 10× Blue Buffer (Enzymatics, Beverly,
253 MA, USA), 1 μL of dUTP/NTP mix (Fermentas, Burlington, Canada), 0.5 μL of 100
254 mM DTT, 0.5 μL of RNase H (Enzymatics, Beverly, MA, USA), 1 μL of DNA
255 polymerase I (Enzymatics, Beverly, MA, USA), and 5 μL of RNase-free water] were
256 mixed. This mixture was incubated at 16 °C for 4 h, and then purified with 24 μL of
257 AMPure XP (Beckman Coulter, Brea, CA, USA) according the manufacturer's protocol.
258 The purified dsDNA was eluted with 14 μL of RNase-free water. Five microliters of the
259 dsDNA solution was used in the following step. End-repair, A-tailing, and adapter
260 ligation were carried out using the KAPA Hyper prep kit (KAPA BIOSYSTEMS,
261 Wilmington, MA, USA) with 1/10× volume of the solutions according the
262 manufacturer's protocol. One microliter of 0.1 μM Y-shape adapter[22] was used in the
263 adapter ligation step for 15 min. Then, size selection of the ligation product was
264 performed with 5.5 μL of AMPure XP (Beckman Coulter, Brea, CA, USA). The
265 purified dsDNA was eluted with 10 μL of RNase-free water and a second round of size
266 selection was performed with 10 μL of AMPure XP (Beckman Coulter, Brea, CA,
267 USA). The size-selected ligation product was eluted with 15 μL of 10 mM Tris-HCl

268 (pH 8.0). One microliter of uracil DNA glycosylase (UDG; Enzymatics, Beverly, MA,
269 USA) was added to the size-selected ligation product. The mixture was incubated at
270 37 °C for 30 min to exclude the second-strand DNA. For library amplification, 2 μL of
271 the UDG-digested DNA, 1 μL of 2.5 μM index primer[22], 1 μL of 10 μM universal
272 primer[22], 0.5 μL of RNase-free water, and 5 μL of KAPA HiFi HotStart ReadyMix (2×)
273 (KAPA BIOSYSTEMS, Wilmington, MA, USA) were mixed. The DNA fragments
274 with the adapters and an index sequence were amplified using a thermal cycler with the
275 following program: denaturation at 94 °C for 2 min, 16 cycles at 98 °C for 10 s, 65 °C
276 for 30 s, and 72 °C for 30 s for replication, and 72 °C for 5 min for the final extension.
277 Two rounds of size selection were then performed to remove adapter dimers with equal
278 volume of AMPure XP (Beckman Coulter, Brea, CA, USA) to the library solution. The
279 purified library was eluted with 10 μL of RNase-free water. The concentration of each
280 library was measured by qPCR with the KAPA SYBR FAST qPCR Master Mix (2×)
281 Kit (KAPA BIOSYSTEMS, Wilmington, MA, USA) and LightCycler 480 II (Roche
282 Diagnostics, Basel, Switzerland) to mix equal amounts of libraries. One microliter of
283 the mixed library was used for electrophoresis using a Bioanalyzer 2100 with Agilent
284 High Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA) to check for
285 quality. Sequencing of 50-bp single-ends using HiSeq 2500 (Illumina, San Diego, CA,

286 USA) was carried out by Macrogen Co (Seoul, South Korea).

287 **Calculation and normalization of RNA-Seq count data**. All reads obtained were
288 trimmed with Trimmomatic (version 0.3.3)[24] using the following parameters:
289 TOPHRED33 ILLUMINACLIP: TruSeq3-SE.fa:2:30:10 LEADING:19 TRAILING:19
290 SLIDINGWINDOW:30:20 AVGQUAL:20 MINLEN:40. This procedure allowed
291 obtaining reads with length > 39 nucleotides and average quality score > 19. The
292 trimmed reads were then mapped to the reference sequences of
293 IRGSP-1.0_transcript[13,14], rRNAs, and transcripts coded in the mitochondria and
294 chloroplast genomes (NC_001320.1, NC_011033.1) with RSEM (version 1.3.0)[25], using
295 Bowtie (version 1.1.2)[26] with default parameters. Expected counts of each gene in the
296 RSEM outputs were used with R (version 3.4.2)[27] in the following analysis. According
297 to a previous study[28], we removed the effects of misidentified reads in multiplex
298 sequencing experiments with HiSeq 2500 (Illumina, San Diego, CA, USA). Of the 926
299 samples from 2015 and 143 samples from 2016, we used 887 and 139 RNA-Seq
300 transcriptome datasets from 2015 and 2016, respectively, with more than 10^5 total read
301 counts for all genes, except for the targets of selective depletion (Extended Data Fig. 1a).
302 We then filtered out rarely detected genes (number of samples with read count > 0; ≤
303 20% of all samples in the 2015 dataset) from the following analyses to focus on the

304 23,924 EGs only (Extended Data Fig. 1b).

305 **Identification of Koshihikari and Takanari specific single nucleotide**
306 **polymorphisms.** The paired-end reads of Koshihikari (SRR1630927) and Takanari
307 (DRR018360) genomes were trimmed with Trimmomatic (version 0.3.3)[24] using the
308 parameters stated in the previous section, except for TruSeq3-PE.fa for adapter
309 trimming. The reads were then mapped to the reference genome (IRGSP-1.0, plastid:
310 NC_001320.1, mitochondrion: NC_011033.1)[13] using Bowtie2 with option -N 1, which
311 allows a mismatch in seed alignment (version 2.2.9)[29]. By using SAMtools (version
312 1.3.1)[30], the reads mapped at multiple loci and the duplicated reads, except for each
313 representative, were discarded. Picard (version 2.1.0)
314 (https://broadinstitute.github.io/picard/) was used to add read group to each SAM file.
315 All SAM files of each sample were merged and a VCF file was generated using
316 SAMtools (version 1.3.1)[30] and BCFtools (version 1.1)[30]. Single nucleotide
317 polymorphisms (SNPs) were filtered using VCFtools (version 0.1.14)[31] with option
318 --minDP 5, which indicates that all alleles supported by more than four reads are
319 reported. The VCF file generated was analysed using R (version 3.4.2)[27] and the R
320 package "vcfR" (version 1.5.0)[32]. We used 15,117 homozygous SNPs distinguishing
321 Koshihikari and Takanari for genotype confirmation of each sample based on RNA-Seq

322 data.

323 **Genotype confirmation**. To confirm the genotype of each sample, we typed each DNA
324 marker based on the SNPs obtained by RNA-Seq. First, to type Koshihikari- or
325 Takanari-specific SNPs in each RNA-Seq sample, the trimmed RNA-Seq reads of each
326 sample were mapped to the reference genome as described in the previous section using
327 Bowtie2. The SNPs were filtered using VCFtools (version 0.1.14)[31] with option
328 --minDP 3, which indicates that all alleles supported by more than two reads are
329 reported. To type each DNA marker in samples, Koshihikari- and Takanari-type SNPs
330 were counted from each DNA marker until one of the counts reached "7" or all the
331 SNPs in 1.5 Mbp from each DNA marker had been counted. In addition, imputation for
332 typing alleles was performed for SL1201, SL1229, and SL1321 due to the low density
333 of SNPs around the DNA markers distinguishing genotypes (See
334 https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-prediction/anal
335 ysis.R for details). Each DNA marker was categorized into Koshihikari-type,
336 Takanari-type, or undetermined based on the majority of SNP counts. Each sample was
337 labeled with the genotype with the most similar DNA marker composition. Thus, 865
338 samples were confirmed as Koshihikari, Takanari, or CSSLs and 11 samples as BILs in
339 2015; and 139 samples as Koshihikari, Takanari, or CSSLs in 2016.

340 **Correlation plot of transcriptomes**. Pearson's correlations of transcriptome data for
341 854 samples in 2015 in the time-series order were calculated with the "cor" function of
342 R, and the heatmap representing such correlations was generated with the "image.plot"
343 function in the R package "fields" (version 9.0)[33].

344 **Meteorological data**. Data on average global solar radiation per 10 min at the Hirakata
345 Weather Station in 2015 for Takatsuki and Kyotanabe Weather Station in 2016 for
346 Kizugawa were obtained from the Japan Meteorological Agency. The average air
347 temperature per minute was then interpolated with the "approxfun" function in R. Data
348 on average global solar radiation per minute at the Osaka Weather Station in 2015 for
349 Takatsuki and Nara Weather Station in 2016 for Kizugawa were obtained from the
350 Japan Meteorological Agency. The time points with global solar radiation < 0.3 kJ m^{-2}
351 min^{-1} were defined as 'dark condition'.

352 **Conversion of age to scaled age.** To consider the difference in heading date among
353 genotypes, we defined scaled age (*SA*) based on sampling dates (*d*), seeding dates (*SD*),
354 and genotypes (*G*). Under the assumption that rice grows lineally before heading at
355 various speeds among genotypes, flowers at scaled age 85, and ages at the same speed
356 after flowering, *SA* (*d, SD, G*) was calculated using the formula:

357 $$SA(d, SD, G) = \begin{cases} 85 \times \dfrac{d - SD}{FD(G, SD)} & ((d - SD) \leq FD(G, SD)) \\ 85 + (d - SD) - FD(G, SD) & \big((d - SD) > FD(G, SD)\big) \end{cases}$$

358 where *FD(G,SD)* is the flowering date of a genotype *G* seeded at seeding date *SD*.

359 **BG-based prediction of GEDs in Koshihikari and Takanari**. A predictive model for

360 GED in Koshihikari and Takanari was constructed with the R package "FIT" (version

361 0.0.4)[10], based on the meteorological data of Takatsuki in 2015, log2reads per million

362 (log2rpm) in 2015, and scaled age of each sample. The GED data of Koshihikari- and

363 Takanari-BG genotypes were used for the Koshihikari and Takanari models,

364 respectively. Precision weight was calculated according to the voom method[10,34] and

365 using data from 887 samples in 2015, and then used for FIT. Details on other parameters

366 are listed in

367 https://github.com/naganolab/Rice_edQTL-analysis_and_edQTL-based-prediction/anal

368 ysis.R. Further, log2rpm of each gene in each sample was predicted with the constructed

369 predictive models corresponding to all possible combinations of the meteorological data

370 and scaled ages of each genotype for the four transplant sets in Takatsuki in 2015 and

371 two transplant sets in Kizugawa in 2016.

372 **Detection of edQTL**. We focused on 3696 genes with EDP (Euclidean distance > 40

373 between Koshihikari and Takanari based on expression levels at the observed time

374 points in 2015). According to the genotype of backgrounds or each marker, the SRE

375 between the observed log2rpm and predicted log2rpm in the Koshihikari- or
376 Takanari-BG models were calculated (Extended Data Fig. 3a, b). To evaluate the
377 statistical significance of improvement in SRE, 1000 permutations of samples with
378 Koshihikari- and Takanari-BG were conducted, and SRE was calculated for each test.
379 Lognormal distribution was then fitted to the obtained distributions of SREs using the
380 fitdist function in the R package "fitdistrplus"[35], with the fitting method of quantile
381 matching of 70% and 1% (Extended Data Fig. 3c). For almost all genes with EDP, the
382 top 0.1% values in the fitted lognormal distribution were higher than those in the
383 distributions generated by permutation (Extended Data Fig. 3d), indicating that the
384 p-value obtained from lognormal distribution is conservative. The p-values for
385 improvement in the SRE were calculated with the fitted lognormal distribution for each
386 gene. The adjustment for multiple comparisons with Benjamini-Hochberg method[36] was
387 performed with the p.adjust function in R. Finally, we defined the markers with the
388 smallest adjusted p-value (< 0.05) in each peak for each gene as the edQTL affecting
389 each GED.

390 **edQTL-based prediction of GED**. For edQTL-affected genes, the Koshihikari- or
391 Takanari-BG model was selected based on the genotype of the edQTL(s). For other
392 genes, the Koshihikari- or Takanari-BG model was selected based on the genotype of

393 backgrounds.

394 **Statistical test for the accuracy of GED prediction in BILs**. The SPEs of all genes
395 affected by each edQTL were calculated based on BG-based prediction (the
396 Koshihikari- and Takanari-BG models) and edQTL-based prediction (edQTL model).
397 To evaluate the statistical significance of improvement in the SPEs, 10,000
398 permutations were performed for the markers in each BIL genome, and the value
399 corresponding to 0.1% percentile in each permutation was calculated.

400 **Code availability**. The scripts used in this study are available from PRJDB7234.

401 **Data availability.** All datasets generated and/or used in this study are available from
402 PRJDB7234.

403

404 1. Takai, T. et al. Genetic mechanisms underlying yield potential in the rice
405 high-yielding cultivar Takanari, based on reciprocal chromosome segment
406 substitution lines. *BMC Plant Biol.* **295**, 14 (2014).

407 2. Poorter, H. et al. Pampered inside, pestered outside? Differences and similarities
408 between plants growing in controlled conditions and in the field. *New Phytol.* **212**,
409 838–855 (2016).

410 3. Kremling, K. A. G. G. et al. Dysregulation of expression correlates with
411 rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).

412 4. Horinouchi, T. et al. Prediction of cross-resistance and collateral sensitivity by
413 gene expression profiles and genomic mutations. *Sci. Rep.* **14009**, 7 (2017).

414 5. Horiuchi, Y. et al. Global expression differences and tissue specific expression
415 differences in rice evolution result in two contrasting types of differentially
416 expressed genes. *BMC Genomics* **1099**, 16 (2015).

417 6. Wang, J. et al. An expression quantitative trait loci-guided co-expression analysis
418 for constructing regulatory network using a rice recombinant inbred line
419 population. *J. Exp. Bot.* **65**, 1069–1079 (2014).

420 7. Jansen, R. C. & Nap, J. P. Genetical genomics: The added value from segregation.
421 *Trends Genet.* **17**, 388–391 (2001).

422 8. Wang, J. et al. A global analysis of QTLs for expression variations in rice shoots
423 at the early seedling stage. *Plant J.* **63**, 1063–1074 (2010).

424 9. Kuroha, T. et al. eQTLs regulating transcript variations associated with rapid
425 internode elongation in deepwater rice. *Front. Plant Sci.* **8**, 1753 (2017).

426 10. Iwayama, K., Aisaka, Y., Kutsuna, N. & Nagano, A. J. FIT: Statistical modeling
427 tool for transcriptome dynamics under fluctuating field conditions.
428 *Bioinformatics* **33**, 1672–1680 (2017).

429 11. Nagano, A. J. et al. Deciphering and prediction of transcriptome dynamics under
430 fluctuating field conditions. *Cell* **151**, 1358–1369 (2012).

431 12. Matsuzaki, J., Kawahara, Y. & Izawa, T. Punctual transcriptional regulation by
432 the rice circadian clock under fluctuating field conditions. *Plant Cell* **27**, 633–648
433 (2015).

434 13. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference
435 genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).

436 14. Sakai, H. et al. Rice annotation project database (RAP-DB): an integrative and
437 interactive database for rice genomics. *Plant Cell Physiol.* **54**, 1–11 (2018).

438 15. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated
439 rice. *Nature* **490**, 497–501 (2012).

440 16. Ookawa, T. et al. Precise estimation of genomic regions controlling lodging
441 resistance using a set of reciprocal chromosome segment substitution lines in rice.
442 *Sci. Rep.* **30572**, 6 (2016).

443 17. Adachi, S. et al. Introgression of two chromosome regions for leaf photosynthesis
444 from an indica rice into the genetic background of a japonica rice. *J. Exp. Bot.* **65**,
445 2049–2056 (2014).

446 18. Adachi, S. et al. The mesophyll anatomy enhancing CO_2 diffusion is a key trait
447 for improving rice photosynthesis. *J. Exp. Bot.* **64**, 1061–1072 (2013).

448 19. Sakamoto, T. An overview of gibberellin metabolism enzyme genes and their
449 related mutants in rice. *Plant Physiol.* **134**, 1642–1653 (2004).

450 20. Sultan, S. E. Phenotypic plasticity and plant adaptation. *Acta Bot. Neerl.* **44**, 363–
451 383 (1995).

452 21. Fournier-Level, A. et al. Predicting the evolutionary dynamics of seasonal
453 adaptation to novel climates in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S.*
454 *A.* **113**, E2812-21 (2016).
455

456 **Supplementary Information** is available in the online version of the paper.

457

458 **Acknowledgements.** We thank the field work support team for help at field sites, Fumie
459 Kobayashi for assistance with laboratory experiments, and Yasuhiro Sato for valuable
460 discussions. We thank Toshio Yamamoto (Institute of Crop Science, NARO) for
461 providing the CSSLs and BILs between Koshihikari and Takanari. This work was
462 supported by the JST CREST JPMJCR15O2 and JSPS KAKENHI Grants JP16H06171
463 and JP16H01473 to A.J.N.

464

465 **Author contributions.** M.K. analysed the data. R.L.S., H.S., and S.O. performed field
466 work. M.K., A.T., A.D., Y.H., and Y.K. conducted RNA-Seq experiments. A.J.N.
467 designed the study. M.K. and A.J.N. wrote the manuscript with input from all
468 co-authors.

469

470 **Declaration of interests.** The authors declare no competing interests.

471

472 22. Nagano, A. J., Honjo, M. N., Mihara, M., Sato, M. & Kudoh, H. Detection of
473 plant viruses in natural environments by using RNA-Seq. *Plant Virol. Protoc.*

474 *Methods Mol. Biol.* **1236**, 89–98 (2015).

475 23. Morlan, J. D., Qu, K. & Sinicropi, D. V. Selective depletion of rRNA enables
476 whole transcriptome profiling of archival fixed tissue. *PLoS One* **7**, e42882
477 (2012).

478 24. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
479 Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

480 25. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq
481 data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

482 26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and
483 memory-efficient alignment of short DNA sequences to the human genome.
484 *Genome Biol.* **10**, R25 (2009).

485 27. R Core Team. R: A language and environment for statistical computing. R
486 Foundation for Statistical Computing, Vienna, Austria. (2017).

487 28. Kamitani, M., Nagano, A. J., Honjo, M. N. & Kudoh, H. RNA-Seq reveals
488 virus-virus and virus-plant interactions in nature. *FEMS Microbiol. Ecol.* **92**, 1–
489 11 (2016).

490 29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie2. *Nat.*
491 *Methods* **9**, 357–359 (2012).

492 30. Li, H. et al. The sequence alignment / map (SAM) format and SAMtools.
493 *Bioinformatics* **25**, 2078–2079 (2009).

494 31. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**,
495 2156–2158 (2011).

496 32. Knaus, B. J. & Grünwald, N. J. VCFR: a package to manipulate and visualize
497 variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).

498 33. Nychka, D., Furrer, R., Paige, J. & Sain, S. fields: Tools for spatial data. (2017).
499 doi:10.5065/D6W957CT

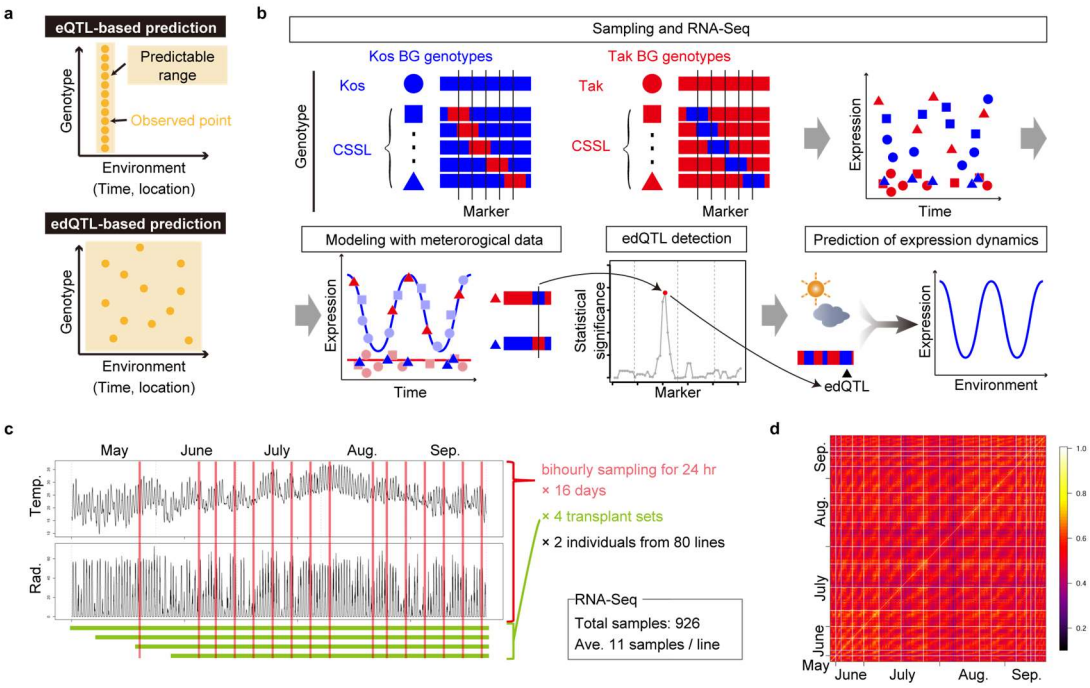
500 34. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. VOOM: precision weights unlock
501 linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29
502 (2014).

503 35. Delignette-Muller, M. L. & Dutang, C. fitdistrplus : An R package for fitting
504 distributions. *J. Stat. Softw.* **64**, 1–34 (2015).

505 36. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate : a practical

and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

508 **Figures**



509

510 **Figure 1 | Concept and workflow of expression dynamics quantitative trait loci**

511 **(edQTLs) detection. a**, Conceptual differences between expression quantitative trait

512 locus (eQTL)- and edQTL-based prediction of genotype × environment interaction

513 (G×E). **b**, Workflow of edQTL detection using chromosome segment substitution lines

514 (CSSLs) and prediction of expression dynamics. **c**, Summary of the sampling design for

515 edQTL detection. The left panels show plots of meteorological data [air temperature

516 (Temp., °C) and global solar radiation (Rad., kJ m^{-2} min^{-1})] and sampling time points in

517 Takatsuki from May to September in 2015. Vertical red lines represent the sampling

518 time points. **d**, Pearson's correlations among the 854 samples used for constructing the

519 predictive model based on expression data of 23,294 expressed genes. White lines

520 indicate the border of each bihourly sampling set.

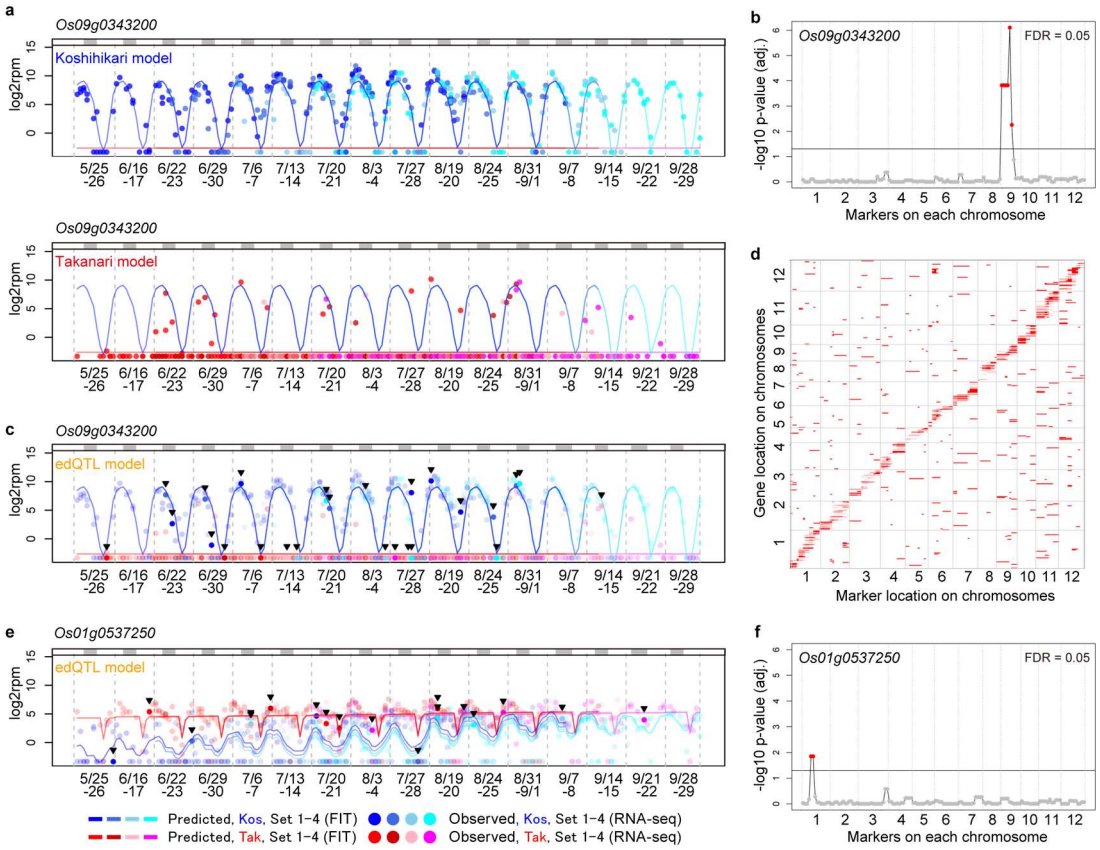


Figure 2 | Expression dynamics quantitative trait loci (edQTLs) detection revealed *cis*- and *trans*-edQTLs are involved in environmental responses. a, c, e, Prediction of *Os09g0343200* (**a,c**) and *Os01g0537250* (**e**) expression levels in Takatsuki in 2015 based on environmental information (time, air temperature, and global solar radiation). and indicate predicted expression levels by 'FIT' in Koshihikari and Takanari in Transplant sets 1, 2, 3, and 4, respectively. and indicate the expression level obtained by RNA sequencing for samples in Transplant sets 1, 2, 3, and 4 of individuals with Koshihikari background (BG) and Takanari BG in (**a**) or Koshihikari- and Takanari-type edQTLs in (**c,e**), respectively. In (**c,e**), strong colors

531 emphasized by the arrow heads indicate samples harboring edQTLs different from BG,

532 and light colors indicate the samples harboring edQTLs identical to BG. The upper

533 indicates dark conditions (global solar radiation < 0.3 kJ m^{-2} min^{-1}). **b,f**, edQTLs

534 regulating *Os09g0343200* (**b**) and *Os01g0537250* (**f**). **d**, X-axis and Y-axis represent

535 statistical significance of each marker (FDR = 0.05) and the genes influenced by the

536 markers in order of position on each chromosome, respectively.

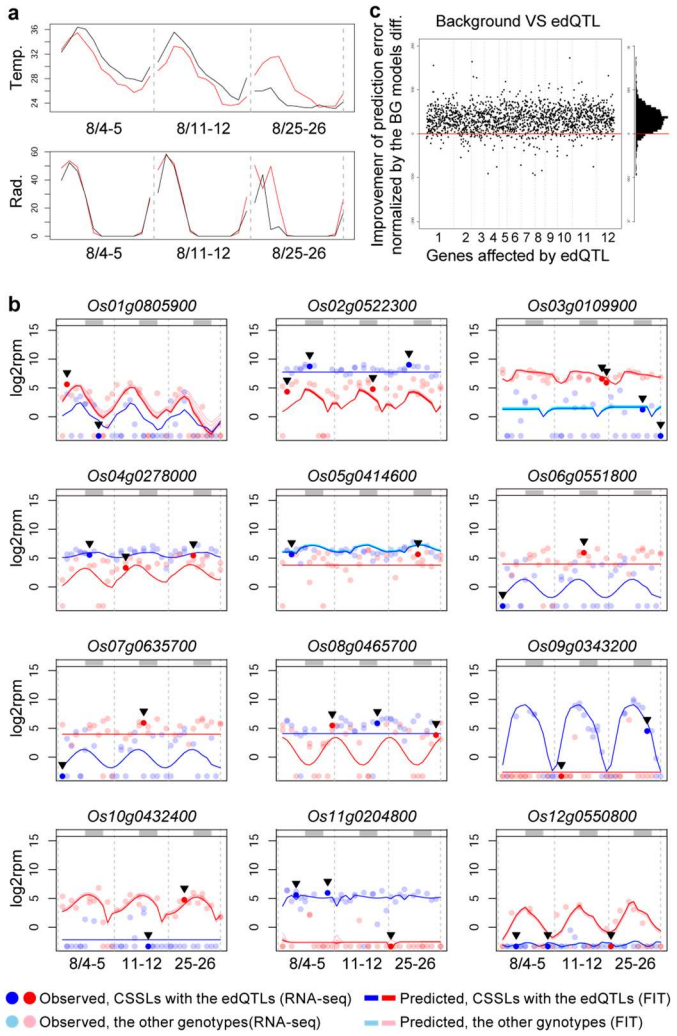


Figure 3 | Expression dynamics quantitative trait locus (edQTL)-based prediction of gene expression dynamics in different environments. a, Plots of meteorological data [air temperature (Temp., °C) and global solar radiation (Rad., kJ m^{-2} min^{-1})] on sampling dates in August. ─ indicates data in Takatsuki in 2015 and ─ indicates data in Kizugawa in 2016. **b,** Examples of prediction of expression dynamics in Kizugawa in 2016 based on environmental information and edQTL in Transplant set 1. Points in strong colors emphasized by arrowheads indicate samples harboring edQTL different

545 from background (BG) genotypes and light colors indicate samples harboring edQTL

546 identical to BG genotypes. The upper ▬ indicates dark conditions (global solar radiation

547 < 0.3 kJ m^{-2} min^{-1}). **c,** Effects of edQTL over prediction of gene expression dynamics in

548 Kizugawa in 2016. Genes influenced by edQTL are in order of position on the

549 chromosomes along the horizontal axis.

550

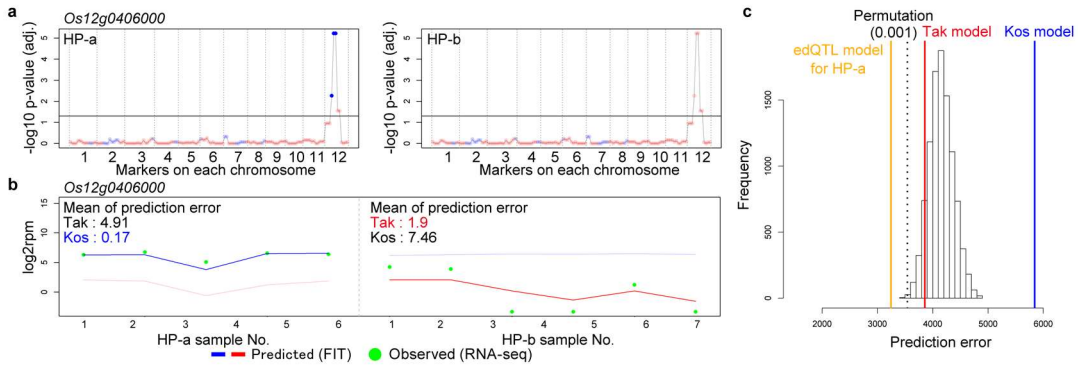


Figure 4 | Expression dynamics quantitative trait locus (edQTL)-based prediction of gene expression dynamics with more complex genotypes than the chromosome segment substitution lines (CSSLs). a, edQTL for *Os12g0406000* and genotypes of HP-a and HP-b. ● indicate significant Koshihikari-type markers. **b,** Prediction of *Os12g0406000* expression in HP-a and HP-b. Strongly colored lines are applied models for HP-a and HP-b. **c,** Prediction accuracy of edQTL model for HP-a. The blue, red, and orange vertical lines indicate the sums of prediction errors based on Koshihikari, Takanari, and edQTL models. The histogram shows the distribution of sums of prediction errors based on the edQTL model in 10,000 permutations of markers in HP-a genome. The dashed vertical line indicates the 0.1% percentile of the distribution.