



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/175449/>

Version: Accepted Version

Article:

Woodhouse, Shannon, He, Zhesi, Woolfenden, Hugh et al. (2021) Validation of a novel associative transcriptomics pipeline in *Brassica oleracea*: Identifying candidates for vernalisation response. *BMC Genomics*. 539. ISSN: 1471-2164

<https://doi.org/10.1186/s12864-021-07805-w>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

20 **Abstract**

21 **Background:** Associative transcriptomics has been used extensively in *Brassica napus* to
22 enable the rapid identification of markers correlated with traits of interest. However, within the
23 important vegetable crop species, *Brassica oleracea*, the use of associative transcriptomics
24 has been limited due to a lack of fixed genetic resources and the difficulties in generating
25 material due to self-incompatibility. Within Brassica vegetables, the harvestable product can
26 be vegetative or floral tissues and therefore synchronisation of the floral transition is an
27 important goal for growers and breeders. Vernalisation is known to be a key determinant of
28 the floral transition, yet how different vernalisation treatments influence flowering in *B.*
29 *oleracea* is not well understood.

30 **Results:** Here, we present results from phenotyping a diverse set of 69 *B. oleracea*
31 accessions for heading and flowering traits under different environmental conditions. We
32 developed a new associative transcriptomics pipeline, and inferred and validated a population
33 structure, for the phenotyped accessions. A genome-wide association study identified
34 miR172D as a candidate for the vernalisation response. Gene expression marker association
35 identified variation in expression of *BoFLC.C2* as a further candidate for vernalisation
36 response.

37 **Conclusions:** This study describes a new pipeline for performing associative transcriptomics
38 studies in *B. oleracea*. Using flowering time as an example trait, it provides insights into the
39 genetic basis of vernalisation response in *B. oleracea* through associative transcriptomics and
40 confirms its characterisation as a complex G x E trait. Candidate leads were identified in
41 miR172D and *BoFLC.C2*. These results could facilitate marker-based breeding efforts to
42 produce *B. oleracea* lines with more synchronous heading dates, potentially leading to
43 improved yields.

44

45

46 **Keywords**

47 Associative Transcriptomics, GWAS, Population Structure, *Brassica oleracea*, Flowering,
48 Vernalisation.

49

50 **Introduction**

51 Ensuring synchronous transiting from the vegetative to the reproductive phase is important for
52 maximising the harvestable produce from brassica vegetables. Many cultivated brassica
53 vegetables arose from their native wild form *B. oleracea* var. *oleracea* [1]. Wild cabbage, *B.*
54 *oleracea* L., is a cruciferous perennial growing naturally along the coastlines of Western
55 Europe. From this single species, selective breeding efforts have enabled the production of
56 the numerous subspecies we see today. The specialization of a variety of plant organs has
57 given rise to the large diversity seen within the species. Various parts of brassicas are
58 harvested, including leaves (e.g. leafy-kale and cabbage), stems (e.g. kohlrabi), and
59 inflorescences (broccoli and cauliflower). For all subspecies, the shift from the vegetative to
60 reproductive phase is important and being able to genetically manipulate this transition will aid
61 the development and production of synchronous brassica vegetables.

62 Determining how both environmental and genotypic variation affect flowering time is important
63 for unravelling the mechanisms behind this transition. For many *B. oleracea* varieties, a period
64 of cold exposure, known as vernalisation, is required for the vegetative-to-floral transition to
65 take place. This requirement for vernalisation, or lack thereof, determines whether the plant is
66 a winter annual, perennial or biennial or whether it is rapid-cycling or a summer annual [2]. As
67 a consequence, the response of the plant to vernalisation provides quantifiable variation that

68 has been exploited by breeders to develop varieties with more synchronous heading. Such
69 variation will be key for future breeding in the face of a changing climate.

70 Genome-wide association studies (GWAS) are an effective means of identifying candidate
71 genes for target traits from panels of genetically diverse lines [3]. GWAS has been used
72 successfully in numerous plant species including Arabidopsis, maize, rice and Brassica [4–
73 7]. However, its application is reliant on genomic resources which are not always available for
74 complex polyploid crops. Associative transcriptomics uses the sequences of expressed genes
75 (mRNAseq) aligned to a reference to identify and score molecular markers that correlate with
76 trait data. These molecular markers represent variation in gene sequences and expression
77 levels. Therefore, unlike traditional GWAS analysis, associative transcriptomics also enables
78 identification of associations between traits and gene expression levels [4]. Associative
79 transcriptomics is a robust method for identifying significant associations and is being used
80 increasingly to identify molecular markers linked to trait-controlling loci in crops [8–11].

81 An important factor to account for in association studies is the genetic linkage between loci. If
82 the frequency of association between the different alleles of a locus is higher or lower than
83 what would be expected if the loci were independent and randomly assorted, then the loci are
84 said to be in linkage disequilibrium (LD) [12]. LD will vary across the genome and across
85 chromosomes and it is important to account for this in GWAS analyses. This variation in LD
86 is due to many factors, including selection, mutation rate and genetic drift. Strong selection or
87 admixture within a population will increase LD. Accounting for the correct population structure
88 reduces the risk of detecting spurious associations within GWAS analyses. The population
89 structure can be determined from unlinked markers [13].

90 Here, we develop and validate an associative transcriptomics pipeline for *B. oleracea*. A
91 specific population structure consisting of unlinked markers was generated using SNP data
92 from 69 lines of genetically fixed *B. oleracea* from the Diversity Fixed Foundation Set [14]. The

93 pipeline was successfully used for the identification of candidate leads involved in
94 vernalisation response, identifying a strong candidate in miR172D.

95

96 **Results**

97 **Exposure to different environmental conditions identifies vernalisation** 98 **requirements across the phenotyped accessions**

99 We selected a subset of 69 *B. oleracea* lines, diverse in both eco-geographic origin and crop
100 type, from the *B. oleracea* Diversity Fixed Foundation Set [14]. We used these accessions to
101 evaluate the importance of vernalisation parameters by quantifying flowering time under
102 different conditions (vernalisation start, duration and temperature). Two key developmental
103 stages were monitored: 'days to buds visible' (DTB) and 'days to first flower' (DTF). The
104 variation in flowering time across the different treatments and between the different lines is
105 shown in Fig. 1. The different vernalisation start times demonstrate that exposure to the longer,
106 ten-week pre-vernalisation growth period (10WPG) typically results in earlier flowering,
107 compared to the shorter, six-week pre-growth period (6WPG). The mean DTB for 6WPG was
108 21.0 days (SD = 51.6), compared to 5.8 days (SD = 49.9) for the 10WPG (Wilcoxon Test, W
109 = 17958, $P = 0.004$). Similarly, we found a significant difference in the time taken to reach DTF
110 between the two treatment groups, with a mean of 57.9 days (SD = 55.5) following the 6WPG,
111 in comparison to 35.9 days (SD = 53.1) following the 10WPG (Wilcoxon Test, $W = 17471$, P
112 = $2.96e-05$).

113 Changes in vernalisation duration led to a significant difference in DTB, but not in DTF.
114 Following the six-week vernalisation (6WV), the mean DTB was 9.5 days (SD = 44.5)
115 compared to 5.8 days (SD = 46.8) after exposure to twelve-weeks of vernalisation (12WV)
116 (Wilcoxon Test, $W = 19532$, $P = 0.002$). This difference was coupled with more synchronous

117 heading between lines following the 12WV period. The impact of vernalisation duration on
118 DTB varied across the population, reflecting the numerous factors that can affect DTB
119 depending on crop type, such as stem elongation and developmental arrest.

120 Of the three parameters we investigated, vernalisation temperature resulted in the most
121 pronounced phenotypic differences. The 5°C vernalisation (5 °CV) resulted in the largest DTB
122 (slowest overall bud development), whereas the 10°C vernalisation (10 °CV) treatment
123 resulted in the largest DTF. The distribution between heading dates was distinctly different
124 between the temperatures. Higher vernalisation temperatures resulted in larger the variation
125 in DTB and DTF. The more synchronous heading and flowering for the 5°C treatment
126 suggests that this temperature was able to saturate the vernalisation requirement for a large
127 proportion of the lines. After exposure to the warmer temperatures, the variation in DTB and
128 DTF were greatly increased (Additional File 1), indicating that the cooler vernalisation
129 temperature aided faster transitioning in some lines, but delayed the development of others.
130 This is consistent with differences in *B. oleracea* crop types, for example Brussels Sprouts are
131 known to have a strong vernalisation requirement, whereas Summer Cauliflower have been
132 bred to produce curd rapidly without the need for cold exposure [15,16].

133 The effect of vernalisation temperature on the floral transition is demonstrated clearly between
134 the Broccoli Mar DH and the Brussel Sprout Cavolo Di Bruxelles Precoce (Fig.1A), with polar
135 responses to vernalisation temperature. Mar DH transitioned fastest under the 15 °C
136 vernalisation (15 °CV) treatment, whereas Cavolo Di Bruxelles Precoce transitioned faster
137 under the 5 °CV treatment. Faster transitions at higher vernalisation temperatures as in the
138 case of Mar DH, however, can lead to undesirable phenotypes from a grower's perspective
139 (Fig. 1B).

140

141 **Unlinked markers are required to generate a representative population**
142 **structure**

143 GWAS requires trait, SNP and population data. The correct population structure is important
144 for ensuring that associations are with the trait of interest rather than identified on account of
145 relatedness within the population, in particular for panels of only one species. To generate a
146 representative population structure, it is necessary to ensure the SNPs used are unlinked [13].
147 However, different criteria have been used to select these SNPs [6,17–19]. To evaluate the
148 impact of SNP selection criteria, we generated two population structures and investigated their
149 suitability for representing the panel.

150 Using all markers with a minor allele frequency (MAF) larger than 0.05 [4,20,21], reduced the
151 total number of SNPs from 110,555 to 36,631. Calculation of ΔK showed a maximum value of
152 $K=2$, although a further peak in ΔK was observed at $K=5$ (Additional File 6A), thus identifying
153 substructure within the population. ΔK frequently identifies $K=2$ as the top level of hierarchical
154 structure, even when more subpopulations are present [21,22]. Subsequent phylogenetic
155 analysis (Additional File 7A, 7B) identified clusters representing these sub populations.
156 Therefore, to account for substructure within the population, the value of $K=5$ was used for
157 further analysis [22,23]. A second population structure was generated using stricter
158 parameters, requiring the markers be biallelic, $MAF > 0.05$, one per gene and at least 500 bp
159 apart. A total of 664 SNPs met these requirements, resulting in the identification of four
160 subpopulation clusters (Additional File 4).

161 We assessed the two population structures based on crop type and phenotypic data. Using
162 $K=5$, generated using the less stringent parameters, (Figs. 2A, 2C, 2E) cluster one contained
163 only broccoli and calabrese, both members of the same subspecies var. *italica* [24,25],
164 whereas cluster two mainly comprised cauliflower, subspecies var. *botrytis*. Late flowering
165 accessions were included in both clusters. Interestingly, this population structure grouped the

166 rapid cycling and late flowering kales together with a spread of accessions from other crop
167 types, in cluster four. The remaining two clusters were small by comparison: cluster three
168 comprised of seven accessions, a mixture of broccoli, cauliflower and kale; cluster five
169 consisted of just two lines, one kale and one cauliflower.

170 The four clusters identified using more stringent SNP selection criteria contained all of the
171 rapid cycling kales in cluster one, characterised by their early heading and flowering
172 phenotypes (Figs. 2B, 2D, 2F). This was identified as a clear subgroup within the phylogenetic
173 tree (Additional File 7C). Cluster two was mainly broccoli and calabrese, whilst cluster three
174 consisted largely of the earlier flowering cauliflowers. Cluster four contained the late flowering
175 individuals from all crop types within the population, hence the larger variation in heading and
176 flowering for this cluster.

177 Comparison of the clustering of accessions between the two population structures
178 demonstrated the more stringent SNP criteria gave rise to a population structure in which
179 individuals were grouped with other accessions that would be expected to be genetically
180 similar based on knowledge of crop type and flowering phenotype. Consequently, this
181 population structure was applied in subsequent GWAS analyses.

182 To gauge the extent of linkage disequilibrium we calculated the mean pairwise squared allele-
183 frequency correlation (r^2) for mapped markers. A linkage disequilibrium window of 50
184 (providing > 3 million pairwise values of r^2) resulted in a mean pairwise r^2 of 0.0979, confirming
185 a low overall level of linkage disequilibrium in *B. oleracea*.

186

187 **Associative transcriptomics identifies miR172D as a candidate for controlling**
188 **vernalisation response**

189 SNP associations were compared to the physical positions of orthologues of genes known to
190 be involved in the floral transition in *Arabidopsis*. A total of 43 flowering time related traits
191 (Additional File 2) were analysed using this pipeline, including DTB and DTF for each
192 treatment. A total of 111 significant SNPs were identified, $P < 0.05$, six of which demonstrated
193 clear association peaks and were investigated further (Table 1).

194 We first sought to identify genetic associations with the trait data for the non-vernalised
195 experiment. Whilst no significant association peaks were identified for DTB, a single marker
196 association at Bo8g089990.1:453:T was identified ($P = 2.29E-06$) for DTF under non-
197 vernalising conditions. This marker was within a region demonstrating good synteny to
198 *Arabidopsis*, despite there being a number of unannotated gene models present.
199 Conservation between *Arabidopsis* and *B. oleracea* suggests that this region contains an
200 orthologue of microRNA172D, AT3G55512, which has been linked to the floral transition in *A.*
201 *thaliana* [26,27] (Fig. 3A). Furthermore, the difference in DTB between 10WPG6WV5 °CV
202 and 10WPG12WV15 °CV, identified a significant association on C07 at Bo7g104810.1:204:T
203 (FDR, $P < 0.05$). This association was in the vicinity of a second orthologue of miR172D (Fig.
204 3C).

205 We then analysed the association with traits relating to the timing of vernalisation. No
206 significant associations were identified for traits after 6WPG12WV5 °CV. However, a strong
207 association was identified on C07 at the marker Bo7g026810.1:124:G, for DTF for
208 6WPG12WV10 °CV. Synteny with *Arabidopsis* suggests that an orthologue of *FRI*
209 *INTERACTING PROTEIN 1*, (*FIP1*), AT2G06005.1 (Fig. 3D) is present within this region.
210 Within *Arabidopsis* it has been demonstrated that *FIP1* interacts with *FRIGIDA (FRI)* [28]
211 which is a major source of natural variation in flowering time in *Arabidopsis* and has been
212 shown to be important in determining vernalisation requirement. Additionally, significant
213 associations (FDR, $P < 0.05$), were found for DTB for 6WPG12WV10 °CV. An association was
214 identified at Bo9g179000.1:2589:G, which is in the vicinity of an orthologue of *Early Flowering*

215 6 (*ELF6*), AT5G04240.1 (Fig. 3B), a nuclear targeted protein able to affect flowering time
216 irrespective of *FLC*.

217 The differences in flowering phenotype between the SNP variants for the four strongest
218 associations were analysed (Fig. 4). There were significant differences in the traits associated
219 with miR172D (DTF with no vernalisation and the difference in DTB for plants grown under 5
220 °CV and 15 °CV) for different alleles (Fig. 4A and B). For Bo7g104810.1:204:T (difference in
221 DTB after exposure to 5 °CV and 15 °CV), five individuals, four broccoli and one cauliflower,
222 contained the A variant. The alternate variant, a T allele, and was present in 50 individuals.
223 Conversely, Bo8g089990.1:453:T (DTF with no vernalisation) had 11 individuals with a C
224 allele at this locus, whilst 51 had a T allele. Interestingly, individuals with the C allele were
225 present in every crop type.

226

227 **Associative transcriptomics identifies *BoFLC.C2* as a candidate gene involved in** 228 **vernalisation requirement in *B. oleracea***

229 An advantage of performing associative transcriptomics as opposed to GWAS, is the
230 additional ability to identify associations between gene expression and the trait of interest.
231 GWAS analysis identified an association of the difference between DTB and DTF with a
232 10WPG6WV5 °CV with a candidate marker in the well characterized flowering time gene,
233 *BoFLC.C2* (Table 1). Using gene expression marker (GEM) analysis, *BoFLC.C2* expression
234 was also identified as being significantly associated with both the DTB and DTF under non-
235 vernalising conditions (Fig. 5). *BoFLC.C2* exhibited both low and high expression within the
236 population. As expected, all five rapid cycling accessions demonstrated no *BoFLC.C2*
237 expression. Recently, a Brassica consortium developed targeted sequence capture for a set
238 of relevant genes, including *FLC*. DNA from four of the five rapid cycling accessions had been
239 enriched with that capture library and sequenced. Lacking a reference sequence for *B.*

240 *oleracea* that contains *BoFLC.C2*, we used *B. napus* (cv. Darmor) [29] as a reference to map
241 the captured sequence data from the four rapid cycling accessions to. Comparison of *B.*
242 *oleracea* transcript data [30] to this Darmor genome reference revealed a 99.54 % identity in
243 coding sequence, allowing Darmor to be used as a surrogate reference. Indeed, we found that
244 *BoFLC.C2* was absent from all four rapid cycling accessions, GT050381, GT080767,
245 GT100067 and GT110222, revealed by a lack of read mapping (Additional File 10). *BoFLC.C2*
246 is known to be involved in vernalisation response [30] and rapid cycling varieties do not require
247 a period of vernalisation in order to transition to the floral state. As a control, we investigated
248 mapping for 49 non-rapid cycling accessions where we expect *BoFLC.C2* to be present. For
249 all 49 we found the expected read mapping evidence, confirming that use of the polyploid *B.*
250 *napus* reference is appropriate (Additional File 10). The control of flowering is a complex,
251 multigenic trait, therefore we would not expect a single locus to explain all variation across the
252 entire dataset. Indeed, only a weak positive correlation (DTB $R^2 = 0.024$, DTF $R^2 = 0.036$)
253 between flowering phenotype and *BoFLC.C2* expression was identified. A strong positive
254 correlation (DTB $R^2 = 0.871$, DTF $R^2 = 0.891$) was found for the phenotypic extremes (rapid
255 cycling lines with no expression and the late flowering lines with high levels of *BoFLC.C2*),
256 Fig. 6, confirming a role for *BoFLC.C2*.

257

258 **Discussion**

259 Determining which genes underly phenotypic traits is a key step for crop improvement. A
260 powerful approach for identifying candidates is associative transcriptomics, which has been
261 implemented for several crops. However, for the important vegetable crop *B. oleracea*, no
262 such pipeline has been published to date. Here we present a validated associative
263 transcriptomics pipeline for *B. oleracea* and use it to identify gene candidates for vernalisation.

264 To reduce the risk of false positives, we developed stringent criteria to identify unlinked
265 markers for the determination of the population structure. The population structure was
266 validated using crop type and phenotypic information on heading and flowering, this example
267 was chosen as producing synchronous *B. oleracea* vegetables is a key goal for growers and
268 breeders. Quantifying vernalisation responses for different varieties is an important step
269 towards this goal, providing a foundation for targeted breeding.

270 Phenotyping for both DTB and DTF under different environmental conditions revealed a varied
271 response within the population and identified some general trends. Altering the timing of
272 vernalisation demonstrated that a shorter growth period prior to the exposure to cold extended
273 the time taken to reach DTB and DTF. This could be attributed to the presence of a juvenile
274 phase in many of the lines, which has been widely documented in *B. oleracea* [14,31,32]. A
275 juvenile plant is described as being unable to respond to floral inductive cues. The fact that
276 many lines were able to flower much faster following longer pre-vernalisation growth, suggests
277 they had reached the adult vegetative phase and were receptive to cold as a floral inductive
278 cue. Further experimental work would be needed to test this hypothesis.

279 Increasing vernalisation length and reducing vernalisation temperature resulted, on average,
280 in faster and more synchronous heading and flowering. This was a predicted outcome, as
281 current knowledge suggests that increased vernalisation duration and cooler vernalisation
282 temperatures would saturate the vernalisation requirement of a larger proportion of
283 accessions.

284 Using our validated population structure with associative mapping, we identified candidates
285 orthologous to known Arabidopsis floral regulators, including miR172D. In Arabidopsis, the
286 miR172 family post-transcriptionally suppress a number of *APETALA1*-like genes, including
287 *TARGET OF EAT1, 2* and *3*, which in turn aids the promotion of floral induction [27,33–35].
288 Furthermore, the SNP variant data for both associations implicating miR172D, exhibit

289 significant phenotypic differences. Two orthologues of *Arabidopsis miR172D* have been
290 identified in *B. oleracea* [36] but their functional roles have yet to be determined.

291 GWAS analysis identified a significant association with *BoFLC.C2* and the difference in DTB
292 and DTF following a ten-week pre-growth period, with six weeks of vernalisation at 5 °C.
293 *BoFLC.C2* is a well characterized flowering time gene [30] and the ability of the GWAS pipeline
294 to identify a known candidate gives confidence in the method. Furthermore, GEM analysis
295 identified *BoFLC.C2* expression as being significantly associated with both DTB and DTF
296 under non-vernalising conditions, which can be attributed to the extreme phenotypes within
297 the population (Fig. 6). No *BoFLC.C2* expression was detected in five lines. A loss-of-function
298 mutation at *BoFLC.C2* in cauliflower has been associated with an early flowering phenotype
299 [37], indicating that *BoFLC.C2* has an equivalent role in cauliflower to *FLC* in *Arabidopsis*.
300 Four of the five lines for which *BoFLC.C2* expression could not be detected did not have the
301 *BoFLC.C2* paralogue according to the bait capture sequencing data. These four lines were all
302 kales and demonstrated an early flowering phenotype, suggesting that *BoFLC.C2* has a
303 similar role to *AtFLC* in kales, and potentially across *B. oleracea*. Although DTB and DTF were
304 highly correlated with *BoFLC.C2* expression under non-vernalising conditions for the
305 phenotypic extremes, for the whole population the correlation was low. This is to be expected
306 as *BoFLC.C2* is just one of many genes that we expect to be involved in the floral transition
307 within *B. oleracea* and therefore is unlikely to account for all the observed variation.

308 The expression data used for the GEM analysis was generated from leaf tissue at one
309 timepoint. As a consequence, any genes which are not expressed in the leaf at this time will
310 not be identified in this analysis. Use of transcriptome data from other tissues in addition to
311 the leaf data could identify a greater number of associations.

312

313 **Conclusion**

314 Identifying genes underlying phenotypic traits in *B. oleracea* is an important step for the
315 improvement of brassica vegetables. Here, we generate and validate a novel pipeline for
316 associative transcriptomics analysis in *B. oleracea* and show that this pipeline is effective in
317 identifying genetic regulators of complex traits, such as flowering time, demonstrating this
318 approach can be utilised for other traits of agronomic importance, such as germination, quality
319 traits and disease resistance. GWAS analysis identified miR172D as a candidate for
320 vernalisation response, whilst GWAS and GEM analysis identified a significant marker at
321 *BoFLC.C2*, an important gene in the vernalisation pathway of *B. oleracea*. Our results provide
322 insight into the genetic control of flowering in *B. oleracea*, and candidates which could provide
323 a foundation for future breeding strategies.

324

325 **Methods**

326

327 **Plant Materials and Growth Conditions**

328 A subset of 69 lines fixed as doubled haploids (DH) or at S4 and above were chosen from the
329 *Brassica oleracea* Diversity Fixed Foundation Set [14] (Additional File 1) comprising
330 accessions from seven different *B. oleracea* crop types; cabbage, cauliflower, calabrese,
331 broccoli, kohlrabi, kale and Brussels sprout. Plants were grown in cereals mix (40 % Medium
332 Grade Peat, 40 % Sterilised Soil, 20 % Horticultural Grit, 1.3 kg/m³ PG Mix 14-16-18 + Te
333 Base Fertiliser, 1 kg/m³ Osmocote Mini 16-8-11 2 mg + Te 0.02 % B, Wetting Agent, 3 kg/m³
334 Maglime, 300 g/m³ Exemptor) and given a pre-growth period of either six or ten weeks in a
335 glasshouse under natural light supplemented with LED lighting (16h daylength 21/18 °C
336 day/night). At the end of the pre-growth period, three plants of each line for each treatments
337 were transferred to Conviron controlled environment rooms for six or twelve weeks

338 vernalisation at 5, 10 or 15 °C (16 h daylength LED, 60 % humidity). Following vernalisation,
339 plants were re-potted into 2 L pots and placed into a polytunnel under natural light using a
340 randomised block design. All plants came out of vernalisation and into the polytunnel on the
341 same day due to staggered sowing to control for post-vernalisation environmental conditions.
342 Three replicates of each line were grown without vernalisation as a non-vernalised control
343 group. The plants were scored at buds visible (DTB) and upon opening of first flower (DTF)
344 [38]. A summary of pre-growth and vernalisation conditions and traits analysed is given in
345 Additional File 2.

346

347 **SNP Calling**

348 The growth conditions, sampling of plant material, RNA extraction and transcriptome
349 sequencing was carried out as described by He et al. [39]. The RNA-seq data from each
350 accession were mapped on to CDS models from the *Brassica oleracea* pangenome [40] as
351 reference sequences, using Maq v0.7.1 [41]. SNPs were called by the meta-analysis of
352 alignments as described in Bancroft et al. [42]. SNP positions were excluded if they had a
353 read depth < 10, a base call quality < Q20, missing data > 0.25, and > 3 alleles. This resulted
354 in a SNP file containing 110,555 SNPS, and 65017 unigene sequences with associated RPKM
355 values.

356

357 **Population Structure and GWAS analyses**

358 Population structure was generated using both relaxed (all markers with a minor allele
359 frequency (MAF) > 0.05) and stringent criteria using STRUCTURE [43] (burn-in10000, MCMC
360 10000, 10 iterations). For the stringent criteria, SNPs were required to be biallelic, with a minor
361 allele frequency (MAF) > 0.05 and a minimum distance of 500-bp between markers.

362 STRUCTURE HARVESTER [44] was used to determine the optimal K value. The Q matrix
363 used in GWAS analysis was calculated using CLUMPP [45].

364 TASSEL [46] version 5.0 was used to select the most appropriate model for each trait based
365 on QQ plots. Generalised linear models (GLM), with correction for population structure using
366 the Q matrix or PCA (5 PCs) were used to look for associations. For GWAS analysis only SNP
367 markers with an allele frequency > 0.05 were used. To gauge the extent of linkage
368 disequilibrium, the mean pairwise r^2 was calculated using the SlidingWindow function within
369 TASSEL, with a linkage disequilibrium window of 50. TASSEL was used to construct
370 phylogenetic trees, using the Neighbour Joining method and all SNPs with MAF > 0.05 . Trees
371 were graphed in R using the package ggtree [47].

372 Gene expression marker (GEM) associations were calculated by an in-house script in R
373 Version 3.6.3 using a fixed effect linear model with RPKM values, excluding markers with an
374 average expression below 0.5 RPKM. Linear regression was performed using RPKM as a
375 predictor value to predict a quantitative outcome of the trait value. Both SNP and GEM outputs
376 were plotted as Manhattan Plots created using an in-house R script. All scripts are available
377 at <https://github.com/JIC-CSB/Boleracea-AssociativeTranscriptomics>. Statistical significance
378 for both GWAS and GEM association was determined by the false discovery rate (FDR) [48]
379 calculated using the QValue package [49] in R.

380

381 **DNA Extraction**

382 Genomic DNA of accessions used in bait capture sequencing was prepared from young leaf
383 tissue of plants grown in a glasshouse (16h LED supplementary light, 21/18 °C day/night).
384 Light was excluded for 48 h prior to harvesting. Nuclei were extracted from ~3 g of tissue prior
385 to CTAB based DNA extraction. Extracts were treated with RNase T1, RNaseA and

386 Proteinase K to remove RNA and protein contamination, respectively. DNA was resuspended
387 in 50 µl dH₂O and checked for quality. DNA was quantified by and stored at -20 °C.

388

389 **Targeted Sequence Enrichment analysis**

390 A bait library for targeted sequence enrichment for a specific subset of genes was developed
391 and synthesized with Arbor Biosciences (<https://arborbiosci.com/>). Samples were 4 plexed
392 and run on the NovaSeq S4, PE150, 1Gbp/library. Reads from individual accessions were
393 mapped to the reference sequence of *B. napus* cv. Darmor-*bzh* [29] using BWA [50] version
394 0.7.17-r1188 using aln/sampe and standard parameters. Mapped reads were sorted and
395 indexed using SAMTOOLS [51] version 1.10 sort and index, and subsequently visualized with
396 Integrative Genomics Viewer (IGV) [52].

397

398 **References**

399

- 400 1. Wichmann MC, Alexander MJ, Hails RS, Bullock JM. Historical distribution and regional
401 dynamics of two Brassica species. *Ecography (Cop)*. John Wiley & Sons, Ltd; 2008;31:673–
402 84.
- 403 2. Chouard P. Vernalization and its Relations to Dormancy. *Annu Rev Plant Physiol. Annual*
404 *Reviews* 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA;
405 1960;11:191–238.
- 406 3. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review.
407 2013.
- 408 4. Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, et al. Associative
409 transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat Biotechnol. Nature*

410 Publishing Group; 2012;30:798–802.

411 5. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. Genome-wide association study of
412 flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet.
413 Nature Publishing Group; 2012;44:32–9.

414 6. Raman H, Raman R, Qiu Y, Yadav AS, Sureshkumar S, Borg L, et al. GWAS hints at
415 pleiotropic roles for *FLOWERING LOCUS T* in flowering time and yield-related traits in
416 canola. BMC Genomics. 2019;20.

417 7. Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, et al. A
418 study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat Genet.
419 2017;49:476–80.

420 8. Miller CN, Harper AL, Trick M, Werner P, Waldron K, Bancroft I. Elucidation of the genetic
421 basis of variation for stem strength characteristics in bread wheat by Associative
422 Transcriptomics. BMC Genomics. BioMed Central; 2016;17:500.

423 9. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, et al. Genome-wide
424 association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat
425 Commun. Nature Publishing Group; 2011;2:467.

426 10. Cockram J, White J, Zuluaga DL, Smith D, Comadran J, MacAulay M, et al. Genome-
427 wide association mapping to candidate polymorphism resolution in the unsequenced barley
428 genome. Proc Natl Acad Sci U S A. National Academy of Sciences; 2010;107:21611–6.

429 11. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. Curr
430 Opin Biotechnol. Elsevier Current Trends; 2006;17:155–60.

431 12. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of Linkage Disequilibrium in
432 Plants. Annu Rev Plant Biol. 2003;54:357–74.

433 13. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using

434 the software structure: a simulation study. Mol Ecol. John Wiley & Sons, Ltd; 2005;14:2611–
435 20.

436 14. Walley PG, Teakle GR, Moore JD, Allender CJ, Pink DA, Buchanan-Wollaston V, et al.
437 Developing genetic resources for pre-breeding in *Brassica oleracea* L.: An overview of the
438 UK perspective. J. Plant Biotechnol. 2012. p. 62–8.

439 15. Putterill J, Laurie R, Macknight R. It's time to flower: the genetic control of flowering time.
440 BioEssays. Wiley Subscription Services, Inc., A Wiley Company; 2004;26:363–73.

441 16. Rosen A, Hasan Y, Briggs W, Uptmoor R. Genome-Based Prediction of Time to Curd
442 Induction in Cauliflower. Front Plant Sci. Frontiers; 2018;9:78.

443 17. Xu L, Hu K, Zhang Z, Guan C, Chen S, Hua W, et al. Genome-wide association study
444 reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.). DNA Res.
445 Oxford Academic; 2015;23:43–52.

446 18. Chen J, Zheng H, Bei J-X, Sun L, Jia W, Li T, et al. Genetic Structure of the Han
447 Chinese Population Revealed by Genome-wide SNP Variation. Am J Hum Genet. Cell
448 Press; 2009;85:775–85.

449 19. Breria CM, Hsieh CH, Yen J-Y, Nair R, Lin C-Y, Huang S-M, et al. Population Structure
450 of the World Vegetable Center Mungbean Mini Core Collection and Genome-Wide
451 Association Mapping of Loci Associated with Variation of Seed Coat Luster. Trop Plant Biol.
452 Springer; 2020;13:1–12.

453 20. Lu G, Harper AL, Trick M, Morgan C, Fraser F, O'Neill C, et al. Associative
454 transcriptomics study dissects the genetic architecture of seed glucosinolate content in
455 *brassica napus*. DNA Res. Narnia; 2014;21:613–25.

456 21. Prom LK, Ahn · Ezekiel, Thomas Isakeit ·, Magill · Clint. GWAS analysis of sorghum
457 association panel lines identifies SNPs associated with disease response to Texas isolates

458 of *Colletotrichum sublineola*. Theor Appl Genet. 2019;132:1389–96.

459 22. Cullingham CI, Miller JM, Peery RM, Dupuis JR, Malenfant RM, Gorrell JC, et al.
460 Confidently identifying the correct K value using the ΔK method: When does K = 2? Mol
461 Ecol. Blackwell Publishing Ltd; 2020;29:862–9.

462 23. Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, et al. The K =
463 2 conundrum. Mol Ecol. Blackwell Publishing Ltd; 2017;26:3594–602.

464 24. Labana KS, Gupta ML. Importance and Origin. Springer, Berlin, Heidelberg; 1993. p. 1–
465 7.

466 25. Maggioni L, von Bothmer R, Poulsen G, Branca F. Origin and domestication of cole
467 crops (*Brassica oleracea* L.): Linguistic and literary considerations. Econ Bot. Springer;
468 2010;64:109–23.

469 26. Poethig RS. Small RNAs and developmental timing in plants. Curr Opin Genet Dev.
470 Elsevier Current Trends; 2009;19:374–8.

471 27. Aukerman MJ, Sakai H. The Plant Cell Regulation of Flowering Time and Floral Organ
472 Identity by a MicroRNA and Its *APETALA2-Like* Target Genes. Plant Cell. 2003;15:2730–41.

473 28. Geraldo N, Bäurle I, Kidou SI, Hu X, Dean C. FRIGIDA delays flowering in Arabidopsis
474 via a cotranscriptional mechanism involving direct interaction with the nuclear cap-binding
475 complex. Plant Physiol. American Society of Plant Biologists; 2009;150:1611–8.

476 29. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, et al. Erratum: Early
477 allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome (Science (2014)
478 345:6199 (950-953)). Science (80-.). 2014. p. 6199.

479 30. Irwin JA, Soumpourou E, Lister C, Lighthart JD, Kennedy S, Dean C. Nucleotide
480 polymorphism affecting *FLC* expression underpins heading date variation in horticultural
481 brassicas. Plant J. 2016;87:597–605.

- 482 31. Wurr DCE, Fellows JR, Sutherland RA, Elphinstone ED. A model of cauliflower curd
483 growth to predict when curds reach a specified size. *J Hort Sci. Taylor & Francis*;
484 2016;65:555–64.
- 485 32. Hand DJ, Atherton JG. Curd initiation in the cauliflower: I. Juvenility. *J Exp Bot. Narnia*;
486 1987;38:2050–8.
- 487 33. Teotia S, Tang G. To bloom or not to bloom: Role of micornas in plant flowering. *Mol.*
488 *Plant.* 2015. p. 359–77.
- 489 34. Jung J-H, Seo Y-H, Seo PJ, Reyes JL, Yun J, Chua N-H, et al. The GIGANTEA-
490 Regulated MicroRNA172 Mediates Photoperiodic Flowering Independent of *CONSTANS* in
491 *Arabidopsis* W OA.
- 492 35. Wu G, Park MY, Conway SR, Wang J-W, Weigel D, Poethig RS. The Sequential Action
493 of miR156 and miR172 Regulates Developmental Timing in *Arabidopsis*. *Cell. Cell Press*;
494 2009;138:750–9.
- 495 36. Shivaraj S, Dhakate P, Mayee P, Negi M, Singh A. Natural genetic variation in MIR172
496 isolated from *Brassica* species. *Biol Plant.* 2014;58:627–40.
- 497 37. Ridge S, Brown PH, Hecht V, Driessen RG, Weller JL. The role of *BoFLC2* in cauliflower
498 (*Brassica oleracea* var. *botrytis* L.) reproductive development. *J Exp Bot. Oxford University*
499 *Press*; 2015;66:125–35.
- 500 38. Meier U. Growth stages of mono-and dicotyledonous plants BBCH Monograph Federal
501 Biological Research Centre for Agriculture and Forestry. 2001.
- 502 39. He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, et al. Extensive
503 homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based
504 visualization. *Plant Biotechnol J.* 2017;
- 505 40. Golicz AA, Bayer PE, Barker GC, Edger PP, Kim HR, Martinez PA, et al. The

506 pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun.
507 Nature Publishing Group; 2016;7:13390.

508 41. Li, H., Ruan J, Durbin R. Maq: Mapping and assembly with qualities. Version 0.6, 3.
509 2008.

510 42. Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, et al. Dissecting the
511 genome of the polyploid crop oilseed rape by transcriptome sequencing. Nat Biotechnol.
512 Nature Publishing Group; 2011;29:762–6.

513 43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus
514 genotype data. Genetics. 2000;155:945–59.

515 44. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for
516 visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet
517 Resour. Springer; 2012;4:359–61.

518 45. Jakobsson M, Rosenberg NA. Genetics and population analysis CLUMPP: a cluster
519 matching and permutation program for dealing with label switching and multimodality in
520 analysis of population structure. 2007;23:1801–6.

521 46. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL:
522 software for association mapping of complex traits in diverse samples. Bioinformatics.
523 Oxford Academic; 2007;23:2633–5.

524 47. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an r package for visualization and
525 annotation of phylogenetic trees with their covariates and other associated data. Methods
526 Ecol Evol. British Ecological Society; 2017;8:28–36.

527 48. Benjamini Y, Yekutieli D, Edwards D, Popper Shaffer J. False Discovery Rate-Adjusted
528 Multiple Confidence Intervals for Selected Para. J. Am. Stat. Assoc. 2005.

529 49. Storey J, Bass A, Dabney A, Robinson D. Qvalue [Internet]. qvalue Q-value Estim. false

530 Discov. rate Control. R Packag. version 2.18.0. 2019 [cited 2020 Mar 20]. Available from:
531 <http://github.com/jdstorey/qvalue>

532 50. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.
533 Bioinformatics. Oxford University Press; 2009;25:1754–60.

534 51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
535 Alignment/Map format and SAMtools. Bioinformatics. Oxford Academic; 2009;25:2078–9.

536 52. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
537 Integrative genomics viewer. Nat Biotechnol. NIH Public Access; 2011;29:24–6.

538 53. Woodhouse S. JIC-CSB/Boleracea-AssociativeTranscriptomics: Brassica oleracea
539 Associative Transcriptomics [Internet]. 2021 [cited 2021 Feb 12]. Available from:
540 <https://zenodo.org/record/4529809#.YCZ9ry2l1ao>

541 54. Steuernagel B, Woodhouse S, He Z, Hepworth J, Tidy A, Siles-Suarez L, et al. BRAVO
542 target sequence capture V3 [Internet]. 2021 [cited 2021 Feb 12]. Available from:
543 <https://zenodo.org/record/4473283#.YCZ91S2l1ap>

544

545 **Figure Legends**

546

547 **Figure 1:** Flowering time traits exhibit a varied response to different environmental
548 conditions within the population. Examples of opposing phenotypic response to different
549 vernalisation temperatures can be observed in A) Brussels Sprout, Cavolo Di Bruxelles
550 Precoce (GT120168) and B) Broccoli, Mar DH (GT110244). Variation across the population
551 for C) DTB post vernalisation per treatment, per line. D) DTF post vernalisation per
552 treatment, per line. Day 0 represents the end of vernalisation, negative values represent
553 heading or flowering during the pre-growth or vernalisation.

554

555 **Figure 2:** The choice of SNP pruning rules can significantly change the inferred population
556 structure. Density plots representing A) DTB, C) DTF for the accessions within the five
557 subpopulation clusters. Density plots representing B) DTB, D) DTF for the accessions within
558 the four subpopulation clusters. E) Population structure generated from SNPs with MAF >
559 0.05 F) Population structure generated from more stringent SNP pruning (Biallelic only, MAF
560 > 0.05, > 500-bp apart, one per gene).

561

562 **Figure 3:** The developed pipeline identifies associations with flowering traits. Distribution of
563 mapped markers associating with A) Number of DTF under non-vernalising conditions B)
564 DTB after a six-week pre-growth, twelve weeks vernalisation 10 °C C) The difference in DTB
565 between six and twelve weeks of vernalisation at 15 °C, after exposure to a ten-week pre-
566 growth D) The DTF after exposure to six-week pre-growth, twelve weeks vernalisation 10 °C.
567 Sixty-nine accessions of *B. oleracea* were phenotyped for DTB and DTF and marker
568 associations were calculated using a generalized linear model, implemented in TASSEL to
569 incorporate population structure. Log₁₀ (P values) were plotted against the nine *B. oleracea*
570 chromosomes in SNP order. Blue line FDR threshold, P< 0.05, FDR threshold was not met
571 for A) and D).

572

573 **Figure 4:** A significant phenotypic difference was found for individuals exhibiting SNP
574 variants for the associations pointing to miR172D as a candidate. Boxplots represent the
575 trait data, DTB or DTF for each of the significant markers alongside the different alleles
576 present across the population for each marker. The box represents interquartile range,
577 outliers are represented by black dots.

578

579 **Figure 5:** GEM analysis identifies *FLC* expression on chromosome C2 as a candidate for
580 flowering traits under non-vernalising conditions. Distribution of gene expression markers
581 associating with A) DTB after exposure to non-vernalising conditions B) DTF after exposure
582 to non-vernalising conditions. Log_{10} (P values) were plotted against the nine *B. oleracea*
583 chromosomes in SNP order. Blue line FDR threshold, $P < 0.05$.

584

585 **Figure 6:** A strong positive correlation can be seen between lines at the phenotypic
586 extremes and their *BoFLC.C2* expression levels. Colours represent the subpopulation of
587 each line, as determined by population structure analysis.

588

589 **Table 1:** Significant SNP associations with vernalisation response in diverse *B. oleracea*
590 accessions, detected across the genome (FDR < 0.05), including model information.

591

592 **Declarations**

593

594 **Ethics approval and consent to participate**

595 The plant material within this paper was obtained under material transfer agreement (MTA)
596 from Warwick Germplasm Research Unit (GRU), part of the European Cooperative Program
597 for Plant Genetic Resources (ECPFR). As such, it complies with relevant institutional,
598 national, and international guidelines and legislation. The appropriate permissions and/or
599 licences for collection of plant or seed specimens have been observed by Warwick GRU for
600 their collections and by the authors under MTA for their subsequent use.

601

602 **Consent for publication**

603 N/A

604

605 **Availability of data and materials**

606 Sequence data from this article can be found in the SRA data library under accession
607 number PRJNA309368, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA309368>. The R
608 scripts used to carry out GEM analysis and to generate the corresponding Manhattan plots
609 for both GEM and GWAS analysis are available on GitHub in the JIC_CSB/Boleracea-
610 AssociativeTranscriptomics repository, DOI 10.5281/zenodo.4529809 [53]. Raw data for
611 targeted sequence capture experiments has been deposited at EBI, under study number
612 PRJEB43076, <https://www.ebi.ac.uk/ena/browser/view/PRJEB43076>, and the bait library is
613 available at DOI 10.5281/zenodo.4473283 [54].

614

615 **Competing interests**

616 The authors declare that they have no conflicts of interest.

617

618 **Funding**

619 SW was supported by the UK Biotechnology and Biological Sciences Research Council
620 (BBSRC) NRPDTP PhD Studentship Programme. JI and RW acknowledge funding from
621 BBSRC Institute Strategic Programme (BB/P013511/1), RM, JI and RW acknowledge
622 support from the BBSRC sLoLa ‘Brassica Rapeseed and Vegetable Optimisation’
623 (BB/P003095/1) and IB and ZH acknowledge funding from BBSRC (BB/L002124/1). RM

624 acknowledges support from EU-Horizon2020 ERC Synergy Grant ‘PLAMORF’ (ID: 810131).
625 Additional funding was provided by BBSRC sLoLa ‘Renewable Industrial Products from
626 Rapeseed’ (BB/L002124/1).

627

628 **Authors’ contributions**

629 JI, RM, RW and SW designed the experiments that were carried out by SW with support
630 from JI and RW. The SNP calling was carried out by ZH under guidance of IB. SW
631 performed the phenotyping of material, all analyses and produced all figures. RW, IB and
632 WH provided genomics and bioinformatics advice. HW provided programming support and
633 guidance. BS designed and constructed the bait library for targeted sequence enrichment
634 and carried out subsequent sequence mapping, which was analysed by SW. SW drafted the
635 manuscript which was planned and refined by SW, RW and all authors contributed to writing
636 the manuscript.

637

638 **Acknowledgements**

639 We thank Profs Lars Ostergaard and Steve Penfield (JIC) and Drs Andrea Harper and
640 Lenka Havlickova (York) for discussion and critical comments on the manuscript. We thank
641 Dr Alex Calderwood (JIC) for guidance and advice. We also thank University of Warwick
642 Germplasm Resources Unit for use of the *B. oleracea* Diversity Fixed Foundation Set [14].

Figure 1

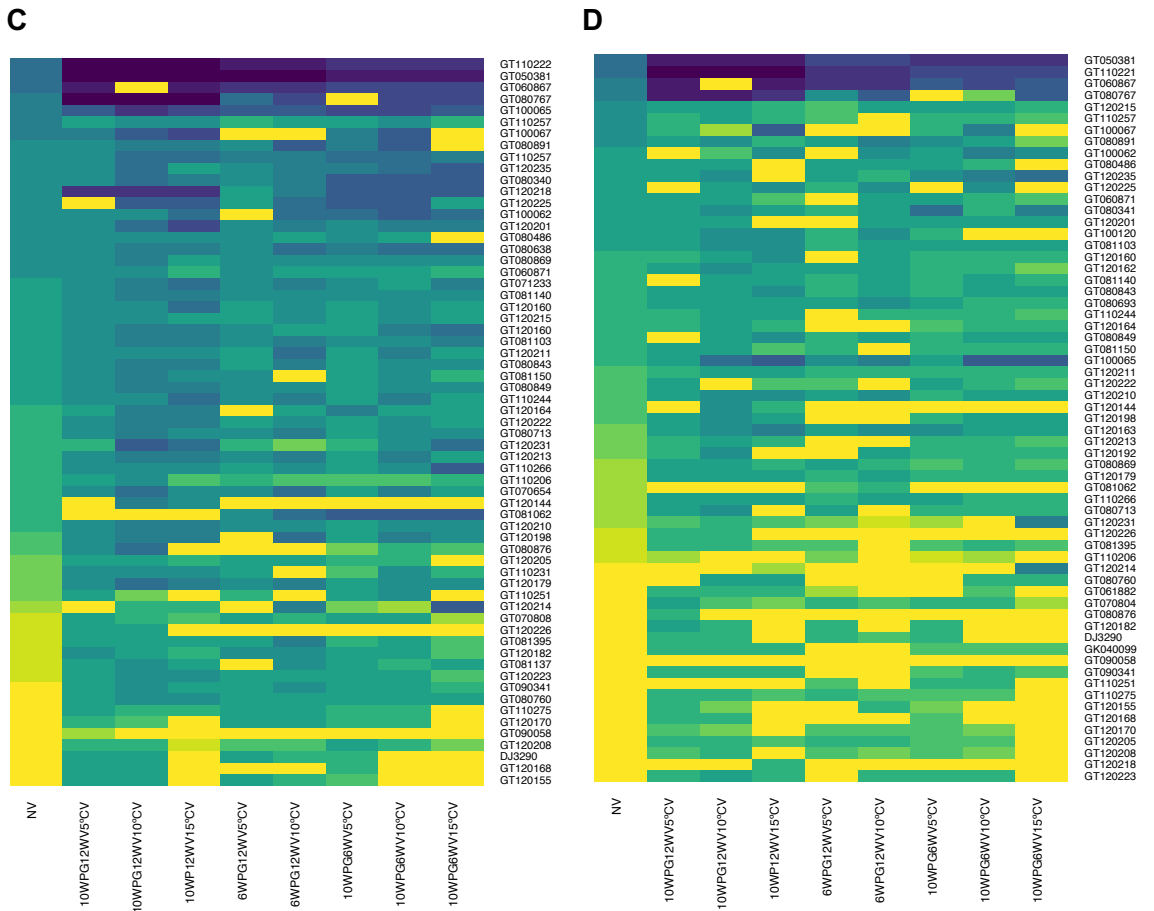
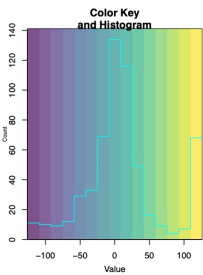
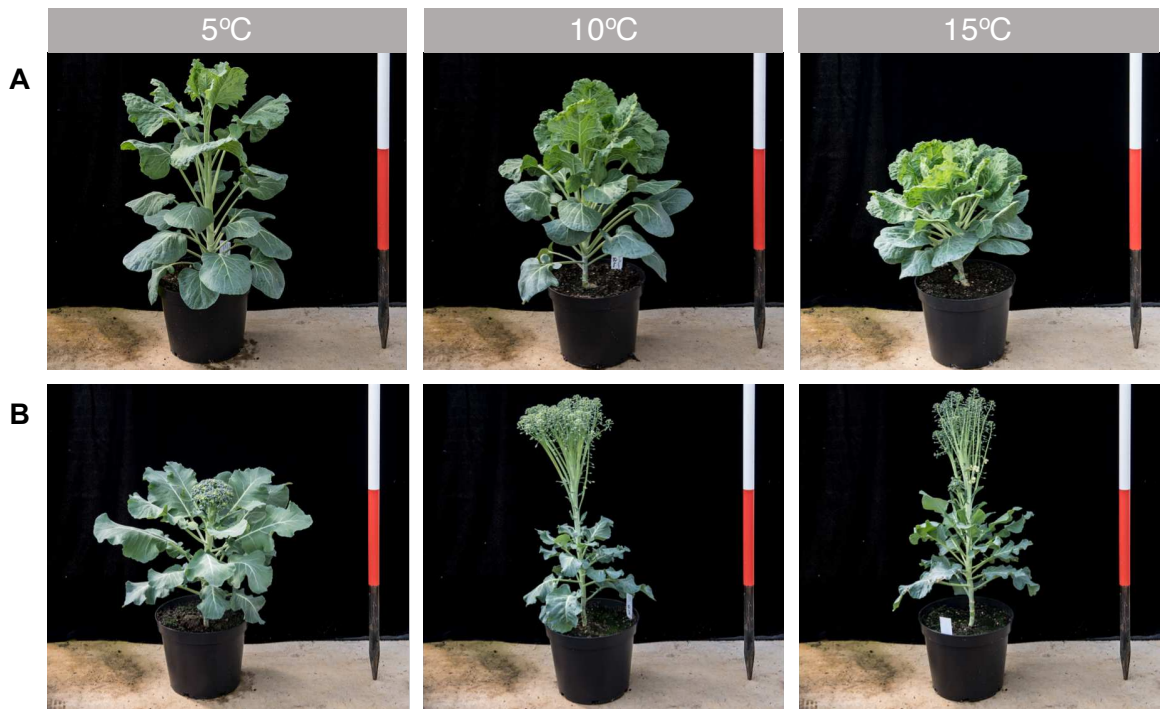


Figure 2

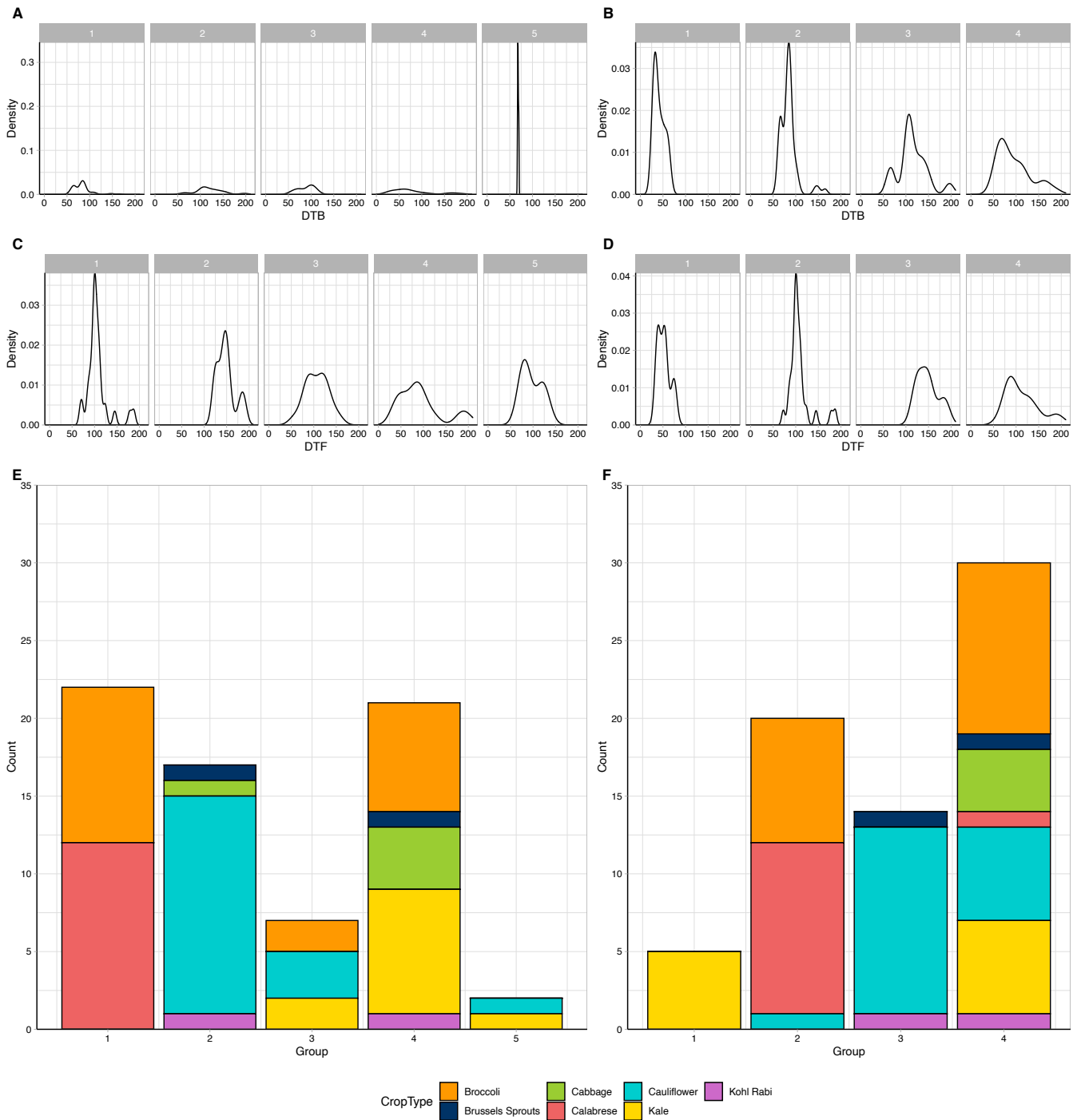


Figure3

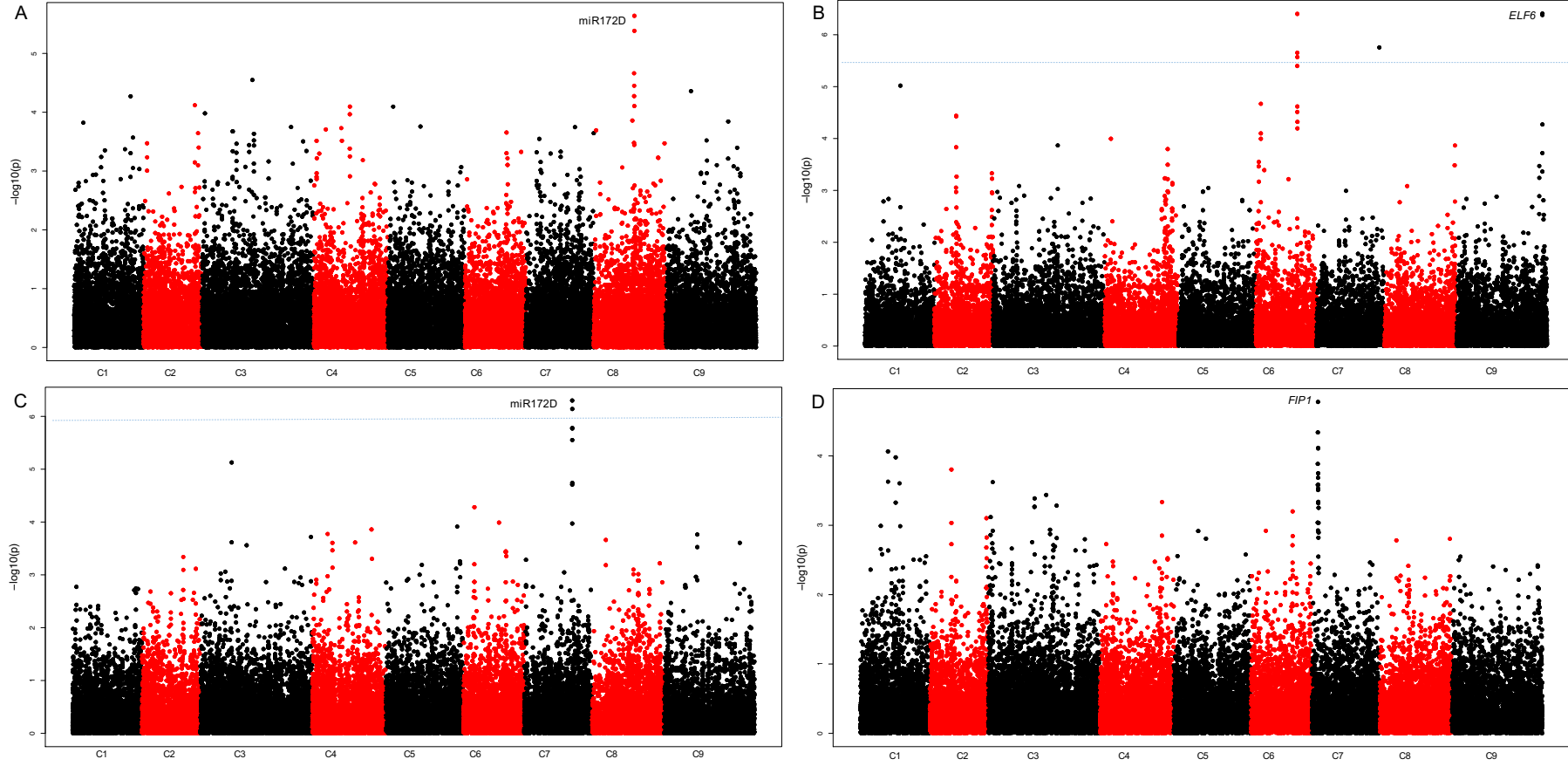


Figure4

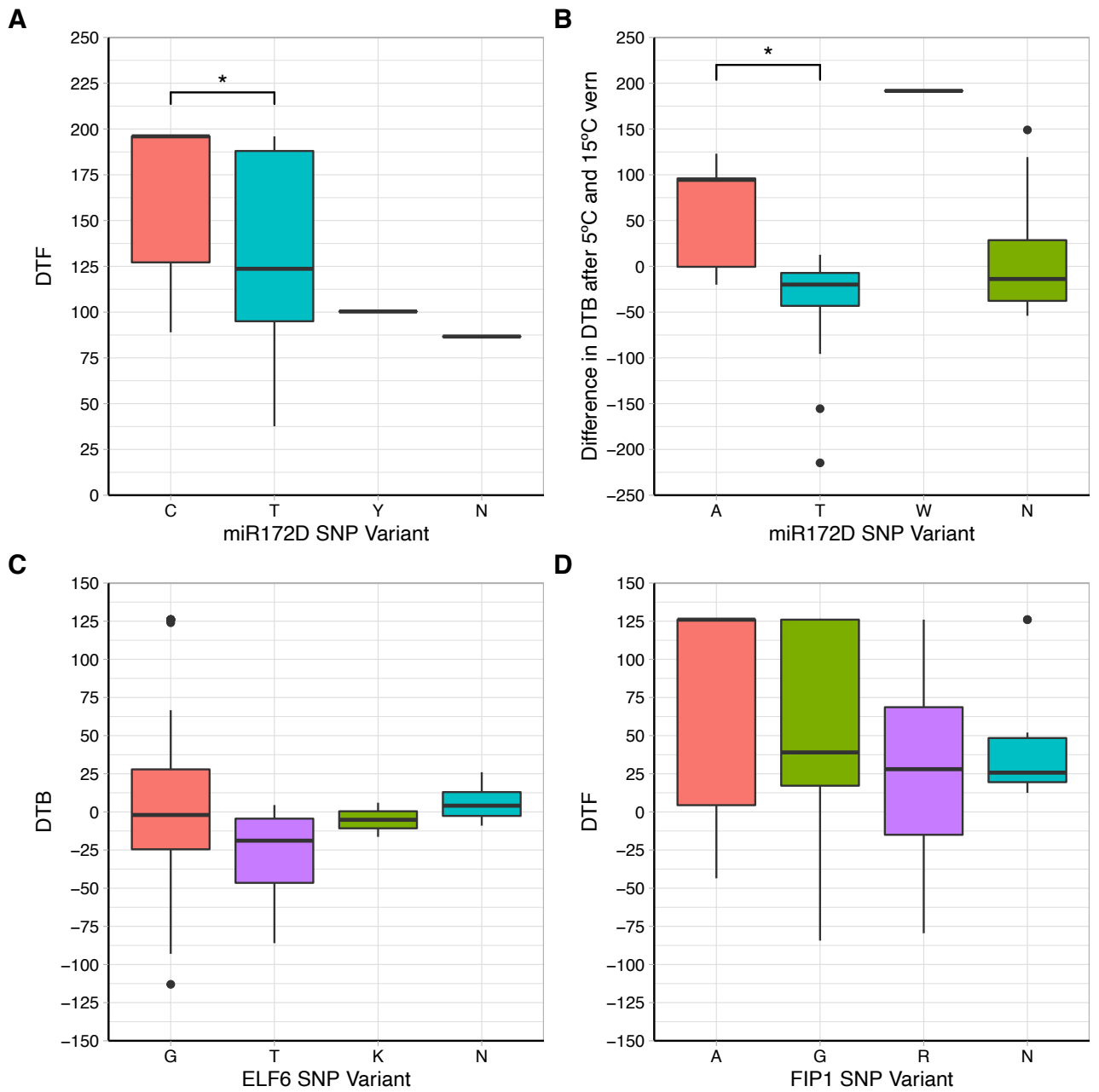


Figure5

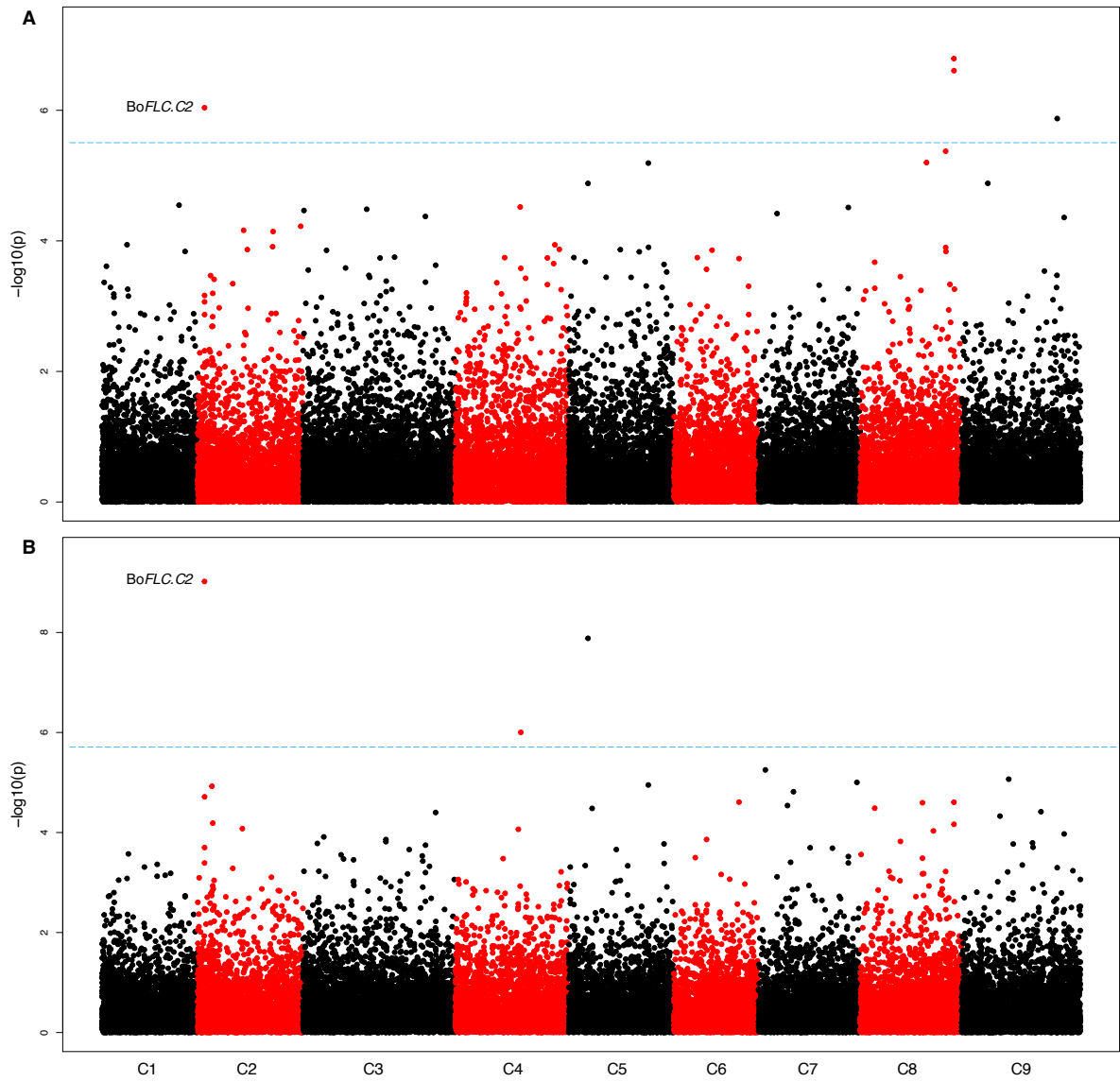
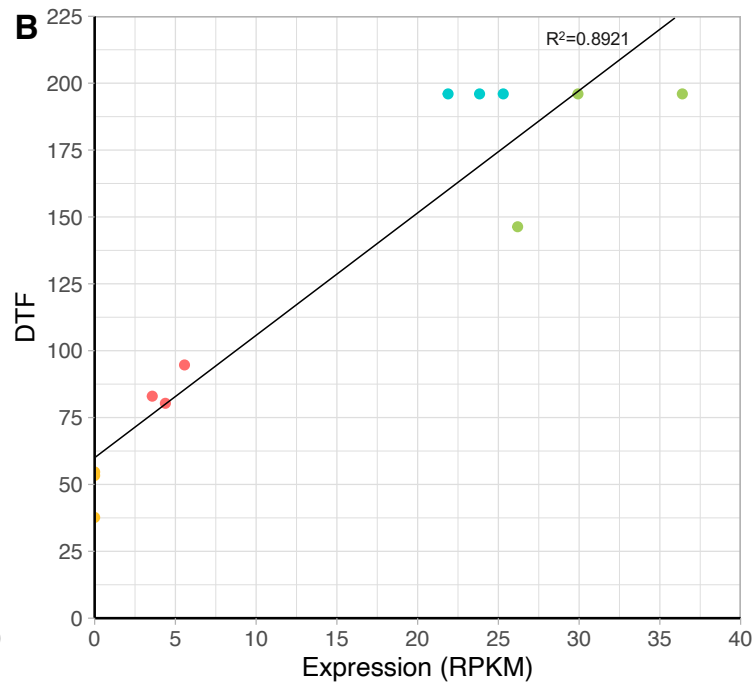
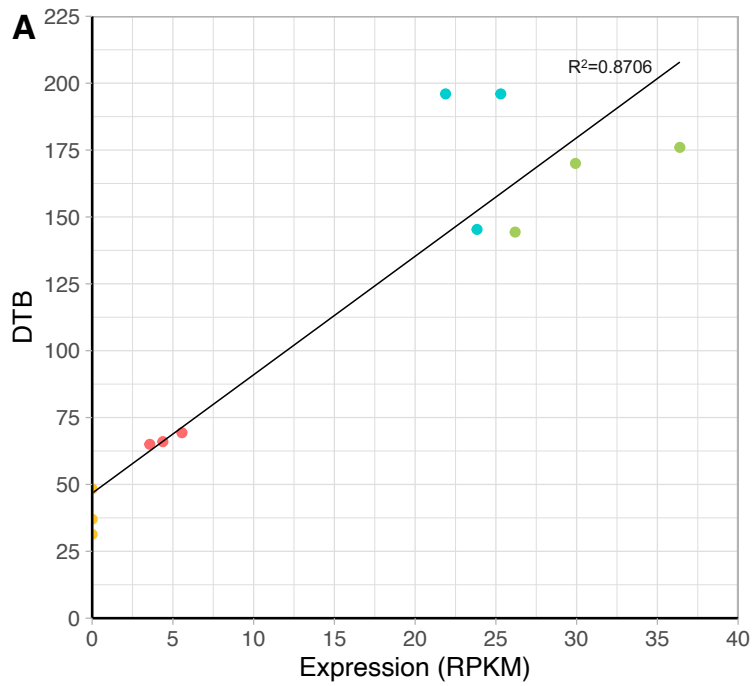


Figure6



Population Cluster

- 1
- 2
- 3
- 4

Table1

Marker Information			Association Information					Model Information	
Marker	Chromosome	Alleles	-Log10(p)	Marker R ²	Traits	Arabidopsis ID	Orthologue	Model	Population Structure Correction
Bo6g103650.1:2010:T	C06	C/T/Y	6.4017787	0.39231	6P 12V 10 °C DTB	AT1G67140.3	<i>SWEETIE</i>	GLM	Q-Matrix
Bo9g179000.1:2589:G	C09	G/T/K	6.4077566	0.39662	6P 12V 10 °C DTB	AT5G04240.1	<i>ELF6</i>	GLM	Q-Matrix
Bo1g011280.1:786:A	C01	A/T/W	6.0844894	0.44220	10P 12V 5 °C DTF	AT4G31490.1	Coatomer, beta subunit	GLM	Q-Matrix
Bo7g026810.1:124:G	C07	A/G/R	4.7781947	0.36476	6P 12V 10 °C DTF	AT2G05790.1	O-Glycosyl hydrolases family 17 protein	GLM	PCA
Bo7g104810.1:204:T	C07	A/T/W	5.9788107	0.41678	10P 6V 15 - 5 °C DTB	AT3G55512	mir172D	GLM	Q-Matrix
Bo2g009460.1:894:T	C02	C/T	7.6880767	0.40565	10P 6V 5 °C DTF - DTB	AT5G10140.4	<i>FLC.C2</i>	GLM	Q-Matrix