1   **FIND-IT: Ultrafast mining of genome diversity.**

2   Søren Knudsen†[1], Toni Wendt†[1], Christoph Dockter†[1]*, Hanne C. Thomsen†[1], Magnus Rasmussen[1],

3   Morten Egevang Jørgensen[1], Qiongxian Lu[1], Cynthia Voss[1], Emiko Murozuka[1], Jeppe Thulin

4   Østerberg[1], Jesper Harholt[1], Ilka Braumann[1], Jose A. Cuesta-Seijo[1], Sabrina Bodevin[1], Lise T.

5   Petersen[1], Massimiliano Carciofi[1], Pai Rosager Pedas[1], Jeppe Opstrup Husum[1], Martin Toft

6   Simmelsgaard Nielsen[1], Kasper Nielsen[1], Mikkel K. Jensen[1], Lillian A. Møller[1], Zoran Gojkovic[1],

7   Alexander Striebeck[1], Klaus Lengeler[1], Ross T. Fennessy[1], Michael Katz[1], Rosa Garcia Sanchez[1],

8   Natalia Solodovnikova[1], Jochen Förster[1], Ole Olsen[1], Birger Lindberg Møller[2], Geoffrey B. Fincher[3] &

9   Birgitte Skadhauge[1]*

10   *To whom correspondence should be addressed (Christoph.Dockter@carlsberg.com,

11   Birgitte.Skadhauge@carlsberg.com)

12   **Affiliations**

13   *¹Carlsberg Research Laboratory, J.C. Jacobsens Gade 4, DK-1799 Copenhagen V, Denmark*

14   *²Plant Biochemistry Laboratory, Centre for Synthetic Biology, Department of Plant and Environmental*

15   *Sciences, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark*

16   *³Australian Research Council Centre of Excellence in Plant Cell Walls, School of Agriculture, Food and*

17   *Wine, University of Adelaide, Waite Campus, Glen Osmond, SA 5064, Australia.*

18

19   Novel crop improvement methodologies, including the exploitation of natural genetic variation,

20   are urgently required to feed our rapidly growing human population in the context of global

21   climate change. Here we describe a 'Fast Identification of Nucleotide variants by DigITal PCR'

22   (FIND-IT) method for the rapid identification of pre-targeted genetic variants or rare alleles in

23   large genomic populations. Libraries of 500,000 individuals can be screened and desired variants

24   isolated within two weeks. FIND-IT is widely applicable for mining valuable diversity in any

25   genomic population, including elite breeding and wild germplasm collections. The method

26   provides single nucleotide resolution that has been validated by identifying and isolating knockout

27   lines, non-synonymous codon changes and variants of miRNA and transcription factor binding

28   sites in the agronomically important crop barley. In contrast to existing methods, FIND-IT does not

29   require transformation, cloning or enzymatic steps, and is exempt from GMO regulations. Thus,

30   FIND-IT can be applied immediately to elite crop cultivars and can be tailored to minimize or

31   eliminate time-consuming backcrossing requirements.

32   Introduction

33   Domestication and subsequent improvement of cereal crops was initiated by human selection of

34   traits for enhanced crop productivity and edibility some 10,000 years ago. Early Neolithic hunter-

35   gatherers, peasants and later farmers, breeders, and scientists, using ever evolving technologies,

36   have strived to improve annual average crop yields and quality. At a global scale, societies are now

37   challenged by crop productivity stagnation across many environments, aggravated by climate

38   change and soil nutrient depletion[1,2]. Current yield increase rates will not accommodate the

39   projected food requirements of 9-10 billion people by 2050[3], demanding a boost of sustainable crop

40   improvement through new, targeted genome engineering technologies. This can be achieved

41   through harnessing natural or induced genetic variation in crop species or their wild relatives and

42   acceleration of the application of variants in elite breeding stock programmes[3].

2

43   Despite advances in plant genome editing by CRISPR-Cas9 technologies, which hold great promise

44   for the long term future of crop adaptation[4], their use in commercial plant breeding is not yet widely

45   applied. This is because *i*) they require elaborate transformation protocols and tissue culture steps

46   that can adversely affect genome integrity[5,6], *ii*) they are therefore non-compatible with most

47   existing breeding pipelines, i.e. not directly applicable to elite, high-yielding, climate-resilient and

48   regionally adapted breeding stock, and *iii*) they are considered GM-technologies in many countries.

49   Large, natural variant populations (e.g. crop fields and germplasm collections), with low

50   spontaneous mutation densities have been the source of natural gene variation used in human

51   endeavours to domesticate and adapt crops to superior performance, independent of changing

52   environments and novel agricultural practices[7]. Induced mutagenesis increases variant densities in

53   populations and has been investigated and utilized for crop development for nearly 100 years[8],

54   notably in the genetic improvements underlying the green revolution crop traits[9]. In particular, the

55   more recently developed chemical treatments, which primarily induce single nucleotide

56   substitutions[10], can be used to obtain large variant populations and eventually single nucleotide

57   resolution, where a substitution of any genome nucleotide is present with high probability.

58   Traditional, phenotype-based methods to screen natural or induced variant populations for

59   desirable genetic variants are painstakingly slow and labour intensive. Targeted screening methods

60   like TILLING (Targeting Induced Local Lesions IN Genomes)[11] or EcoTILLING[12], provide genome-wide

61   scanning of induced and natural variant populations directly at the genotype level. However,

62   technical hurdles and low detection sensitivity usually restrict population sizes to fewer than 10,000

63   individuals. This prevents the achievement of single nucleotide resolution or, alternatively, demands

64   very high mutation rates per genome[13], which result in high off-target background mutation loads

65   that are undesirable in downstream breeding applications[14].

66   Here, we present a simple, agile, and high-throughput strategy to screen extremely large, low

67   mutation-density variant populations, scalable to single nucleotide resolution, for targeted

68   identification of desired traits that are immediately applicable in crop or organism improvement

69   strategies. In this 'Fast Identification of Nucleotide variants by DigITal PCR' (FIND-IT) method, we

70   induce, identify and isolate specific targeted nucleotide variants by combining systematic sample

71   pooling-and-splitting with high-sensitivity genotyping. FIND-IT *i*) does not require transformation or

72   tissue culture protocols, *ii*) is applicable to any living organism that can be grown in the field or in

73   culture (i.e. elite plant crop varieties, bacteria, yeasts), and *iii*) is exempt from GMO regulation. Our

74   pooling strategy reduces the time required for DNA isolation up to 300-fold and, using cultivated

75   barley as an example, we show that it is feasible to screen, identify and isolate pre-targeted variants

76   within two weeks, in large variant libraries (>500,000 individuals) with low individual mutation loads.

77   This reduces or eliminates extensive backcrossing requirements and thereby greatly reduces the

78   time from variant detection to field evaluation. By targeting specified single-nucleotide changes in

79   genes that modify barley crop performance, grain morphology and grain quality properties, we

80   demonstrate how FIND-IT circumvents limitations of comparable techniques and provides a novel

81   methodology for ultrafast variant mining and improvement of today's germplasm, through precision

82   breeding.

83   # Results

84   **Designing libraries for targeted variant identification.**

85   The mutational load and the type of genetic variance introduced by mutagenesis form the basis for

86   successful identification of the desired genetic variant. Chemical mutagenesis has a major advantage

87   over radiation-induced mutagenesis by predominantly causing single nucleotide changes and,

88   depending on the respective library size, facilitating single nucleotide resolution without additional

89   changes associated with base/nucleotide excision repair[14]. We generated large variant libraries in

90   barley (Fig. 1a), wheat, rapeseed, oat, yeast and bacteria, using either sodium azide (NaN$_3$) or ethyl

91   methanesulfonate (EMS) as the mutagen (Supplementary Notes 1, Section 1). The spring barley

92   libraries used here to demonstrate and validate the FIND-IT methodology were designed with the

4

93   aim of obtaining a large population of 500,000 individual plants with 1,000 – 5,000 induced

94   nucleotide substitutions per genome. The recent publication of the barley pan genome identified

95   large structural chromosome rearrangements and approximately 1.5 million presence/absence

96   variations (PAV) in the range of 50 basepairs to 1 million bases (Mb) between 20 pan genome

97   assemblies[7]. Our summary statistics of genome variation in whole genome sequencing data from the

98   200 domesticated and 100 wild barley lines published with the barley pan genome revealed more

99   than 223 million naturally occurring nucleotide substitutions, with 299 of 300 lines having more than

100   5 million nucleotide substitutions when compared to the barley reference genome Morex_V2

101   (Supplementary notes 1, Section 2). Thus, our aim of increasing the natural variation by 1,000 - 5,000

102   induced nucleotide substitutions per plant is extremely low in the context of existing genome

103   variation and novel genome variation created by crossing events in traditional breeding.

104   **Determining mutation load of variant libraries.**

105   To define library mutation density, we applied two levels of characterization (Supplementary Notes

106   1, Section 3). We first Sanger sequenced four individual gene fragments in some 5,000 barley plants

107   from the library treated with low NaN_3 concentration (0.3 mM) (Table Supplementary Notes 1.3a).

108   Here, mutation densities were in the range of 0.29 to 1.45 induced variants per Mb per plant. Given

109   an average barley genome size of approx. 4.2 gigabases[7], this equals approx. 1,200 to 6,000 induced

110   variations per plant and is within the desired mutation density range.

111   In a complementary approach, Genotyping-By-Sequencing (GBS) of 40 barley individuals from low

112   and high (0.3 and 1.67 mM NaN_3) mutagen treated libraries (20 individuals/library) verified low

113   mutation rates after 0.3 mM NaN_3 treatment (0.46 to 0.9 induced variants/1Mb, accounting for

114   approx. 2,000 to 3,750 induced variations per plant) and indicated a tendency for higher mutation

115   rates after treatment with the higher NaN_3 concentration (Table Supplementary Notes 1.3b). It

116   should be noted that in 16 and 11 genotyped individuals chosen from both libraries no mutation was

117   detected, respectively (Supplementary Notes 1, Section 3).

118 Thus, libraries can be designed to have the desired mutation load by controlling treatment intensity

119 and population size.

120 **FIND-IT employs a split-and-pool strategy coupled with high-sensitivity droplet digital PCR**

121 **(ddPCR).**

122 To facilitate high-throughput screening of >500,000 individuals, we integrated a systematic and cost-

123 effective split-and-pool strategy. The ddPCR technology is 1,000 fold more sensitive than

124 conventional PCR[15], and we confirmed its sensitivity by spiking pools of 1,000 wild-type barley grains

125 with 4, 2, 1, or a half grain homozygous for a specific nucleotide exchange in the *LIPOXYGENASE-1*

126 (*Lox-1*) gene. When targeted in total DNA of such dilution pools, the *Lox-1* variant signal achieved by

127 ddPCR was clearly distinguishable from the background signal at all dilutions tested (Supplementary

128 Notes 1, Section 4). We based the FIND-IT libraries on DNA pools representing grains from up to

129 1,200 individual M2 plants. For practical reasons, we harvested all material in pools of up to 300

130 plants, and only after DNA extraction we combined DNA samples to represent an estimate of 1,200

131 plants per pool (see Online Methods)[16]. The initial pooling strategy of up to 300 plants reduced the

132 time required for DNA isolation by up to 300-fold, compared with screening strategies that use DNA

133 from individual plants in the library. This removes a key technological bottleneck inherent in

134 screening methods involving single plant sequencing or genotyping.

135 In each pool, multiple grains will originate from the same plant and hence a given induced variant

136 will be represented several times in a pool. Thus, pools can be split into a 25% sample for initial

137 large-scale genotyping while the remaining grain pool (75%) is stored for subsequent specific variant

138 identification and isolation (Fig. 1b).

139 In Phase 1 of the FIND-IT procedure, DNA representing approximately 100,000 plants is screened on

140 a single 96-well PCR plate. Here, 94 DNA pools (representing up to 1,200 plants each), plus two

141 negative controls, are genotyped by ddPCR for an induced nucleotide target using a single set of

142   primers and two competitive hydrolysis probes that detect either wild-type or variant alleles (Fig.

143   1b).

144   Positive pools containing the targeted variant are further analyzed in FIND-IT's Phase 2 (Fig. 1b).

145   Here, a new 96-well plate where each well contains DNA from 10 grains from the Phase 1 identified

146   grain pool (75% of original grain pool, and likely to contain additional grains of the targeted variant)

147   are ddPCR genotyped to identify a positive sub-pool with the targeted variant. The DNA is extracted

148   from flour sampled by non-destructively drilling into the endosperm of individual seeds[17]. Thus, by

149   combining a split-and-pool strategy with ddPCR genotyping, we can quickly narrow down the

150   number of candidate grains carrying the targeted gene variant from 500,000 plants to 10 grains.

151   Total DNA from drilled grains or drilled grains from sub-pools can be stored safely at -20°C. DNA

152   from drilled grain sub-pools can be re-screened multiple times for other desired gene variants if

153   identified in FIND-IT's Phase 1, further reducing sub-pool generating workload. Finally (Phase 3; Fig.

154   1b), the 10 grains from the positive sub-pool are germinated and genotyped by ddPCR and Sanger

155   sequenced for identification and isolation of the plant carrying the desired induced gene variant. If

156   present in the library, the FIND-IT workflow enables the identification and isolation of a desired

157   variant within days and weeks, respectively. During Phase 2 and 3, Phase 1 can be re-initiated for

158   screening of new targets and Phase 3-targets can be accumulated for germination and validation

159   (Supplementary Notes 1, Section 5), further accelerating the FIND-IT workflow.

160   **Modulating plant- and grain-morphology by targeting loss-of-function variants.**

161   To demonstrate FIND-IT's high capacity to identify knockout variants, we identified 100 induced

162   premature stop codons in specified barley genes (Fig. 2a, Supplementary Table 1 and Supplementary

163   Notes 2). Of these, we isolated and field characterized five primary knockout variants of genes that

164   mediate well-documented changes in phenotypic characteristics of the barley grain (Fig. 2b-h).

165   The *Nud* gene (Supplementary Notes 3, Section 1) encodes a transcription factor of the ethylene

166   response factor (ERF) family that controls the covered/naked caryopsis trait in barley and knockout

7

167   of this gene (Fig. 2b) converts hulled to hull-less grain[18] (Fig. 2c). Using the FIND-IT technology, we

168   isolated a novel premature stop allele of the *Nud* gene (*nud^{W16*}*; Fig. 2b). Field trial evaluation of

169   *nud^{W16*}* in an elite spring barley cultivar (RGT Planet) background showed the expected loosely

170   attached hull and accompanying reduced grain length, width, and thousand grain weight (TGW) after

171   threshing and hull removal (Fig. 2b-e and Fig. Supplementary Notes 3.8a-e). This known

172   domestication trait renders any wild, locally adapted or climate-resilient barley, or any elite hulled

173   barley cultivar, suitable for human consumption and the ability to directly proceed to agronomic

174   evaluation allows precise trait market value evaluation by comparison to elite hulled barley varieties.

175   The *JMJ706/VRS3* locus (Supplementary Notes 3, Section 1) encodes a putative Jumonji C-type

176   H3K9me2/me3 histone demethylase that is involved in defining the barley inflorescence (spike) row-

177   type, a yield-related domestication trait controlled by several genes[19]. We identified a novel

178   homozygous loss-of-function variant in the *JMJ706* gene in cultivar Planet (*jmj706^{W313*}*; Fig. 2b) that

179   has the reported barley spikelet phenotype, important for grain size uniformity in six-rowed barley[19]

180   (Fig. 2c). Using FIND-IT, novel alleles of row-type genes, as shown here for *JMJ706/VRS3*, can be

181   directly isolated from elite two- and six-rowed barley libraries to quickly optimize agronomic

182   performance[19].

183   The *Cellulose synthase like* gene *CslF6* (Supplementary Notes 3, Section 2) is a key determinant

184   controlling the biosynthesis and structure of (1,3;1,4)-β-glucan in the barley grain, making it a prime

185   target for lowering (1,3;1,4)-β-glucan[20], a preferred grain trait for the distilling and brewing

186   industries. Using FIND-IT we identified and isolated a novel *CslF6* knockout variant (*cslF6^{W676*}*; Fig.

187   2b) in cultivar Planet with undetectable levels of (1,3;1,4)-β-glucan content in the grain (Fig. 2c,f).

188   Field-testing for agronomic performance of *cslF6^{W676*}* confirmed previously published greenhouse

189   experiments on *CslF6* knockouts, namely that TGW and grain width were reduced and changed,

190   respectively[21] (Fig. 2d-f; Fig. Supplementary Notes 3.17a-f). Here, we show for the first time that

191   total grain yield of *CslF6* knockout variants grown in the field is reduced by more than 30% (Fig. 2e),

192   limiting the applicability of *CslF6* knockout varieties.

193   Granule-Bound Starch Synthase I (GBSS1) (Supplementary Notes 3, Section 1) is the enzyme for

194   synthesizing amylose in barley[22]. We identified a variant carrying a novel loss-of-function allele

195   (*gbss1^{W126*}*; Fig. 2b) that features the characteristic waxy barley grain phenotype and the underlying

196   reduction of amylose content in the grain starch (Fig. 2c,g), with no significant yield and TGW change

197   (Fig 2d,e). Waxy starches are of high value in many processed foods as the amylose content directly

198   influences the texture of cooked starches[23]. Now waxy variants of common cereals can be isolated

199   from elite, high yielding varieties directly for market potential evaluation.

200   To illustrate the translation of agronomic traits from one cereal to another we used the FIND-IT

201   methodology to induce a yield related loss-of-function trait *Grain Width 2* (*GW2*) (Supplementary

202   Notes 3, Section 1), a modulator of grain morphology known in rice, into barley[24] by isolating a novel

203   loss-of-function allele in the elite malting barley Paustian (*gw2^{W233*}*; Fig. 2b). To date, no knockout of

204   barley *GW2* has been identified or studied[25]. Initial field yield trials (Denmark 2020) showed that

205   *gw2^{W233*}* has increased TGW, caused by wider grain size as shown in rice, but that total yield per

206   area was not increased (Fig. 2c-e, h). Knockout variants of *OsGW2* improve grain yield in rice via

207   acceleration of the grain milk filling rate, a trait of potential importance for cereals grown in

208   suboptimal growth conditions such as drought and heat stress associated with shortened grain filling

209   periods. The novel barley *gw2^{W233*}* variant can now be field-tested in diverse climatic environments

210   to identify its potential for yield improvement and climate adaptation in barley and other cereals.

211   **Targeting non-synonymous single-nucleotide variants and variants that modulate transcript**

212   **abundance.**

213   The ability to finely regulate gene expression, transcript abundance and protein function is a highly

214   desirable objective of genome engineering tools for precision breeding. FIND-IT enables the

215   identification and isolation of such finely tuned gene variants. Here, 55 genetic variants with single,

9

216   non-synonymous base-changes were targeted, either for changing transcript levels or by altering the

217   amino acid sequences of encoded proteins (Fig. 2a, Supplementary Table 1, Supplementary Notes 2).

218   In one example we targeted the GROWTH-REGULATING FACTOR 4 (GRF4) transcription factor and

219   the DELLA protein (SLENDER1/SLN1) that promote and repress growth, respectively, as well as

220   influencing nitrogen assimilation and carbon fixation[26] (Supplementary Notes 3, Section 2). In rice,

221   enhanced nitrogen-use efficiency and grain yield are achieved by increasing *GRF4* transcript

222   abundance and by reducing DELLA activity[26]. We identified 31 variants of *SLN1/DELLA* with specific

223   amino acid exchanges (Fig. 3a, Supplementary Table 1) and characterized one variant *sln1^{A277T}* (Fig.

224   3b-d), showing normal plant height, increased seed length and increased TGW but no change in yield

225   (Fig. Supplementary Notes 3.4). Next, we identified three variants targeting the *miRNA396*-binding

226   site of *GRF4* (Fig. 3e) with the aim of increasing GRF4 activity by reducing miRNA-mediated transcript

227   repression. The genetic variants of *GRF4* and *SLN1/DELLA* identified will enable field trials aimed at

228   generating new elite barley varieties with enhanced nitrogen-use efficiency[26].

229   In a second example, we targeted a promoter variant of a barley α-amylase gene to enhance gene

230   expression and seed germination vigour in Planet barley (Fig. 3f and Supplementary Notes 3, Section

231   2). Expression of the *Amy1_2* gene in the germinated Planet control grain exceeded the expression

232   of genes in the *Amy1_1* or *Amy2* clusters at 72 hours after germination[27] (Fig. 3g). Inspired by natural

233   variation in the tandem repeat W-box/O2S binding domain in α-amylase promoters, to relieve the

234   repression by WRKY38[28], we identified and isolated a novel allele of *Amy1_1* (*Amy1_1-var*) with a

235   single nucleotide substitution in its promoter that changes one TGAC core motif to TGAT (Fig. 3f).

236   We then demonstrated that the overall *Amy1_1* transcript level in the novel Planet variant *Amy1_1-*

237   *var* increased 72h after initiation of germination by close to 100% relative to control (Fig. 3h),

238   indicating that changing the TGAC to TGAT released repression of expression. This result shows the

239   function of the tandem repeat W-box/O2S binding domain in α-amylase promoters *in vivo* and

240   highlights the possibilities of modifying gene transcript levels by targeting promoter domains using

241   FIND-IT, both for gene function analyses and for specific breeding purposes.

242   To isolate variants with an intermediate (1,3;1,4)-β-glucan content phenotype, to overcome the

243   adverse yield and grain morphology phenotypes of the knockout line (Fig.2 c-e), we identified and

244   isolated three additional *CslF6* variants with amino acid exchanges in close proximity to the TED

245   motif that controls (1,3;1,4)-β-glucan fine structure (T709I)[29], or within CslF6 transmembrane helices

246   (G748D and G847E)[20] (Fig. 3i and Supplementary Notes 3, Section 2). These novel *CslF6* variants have

247   a 50% reduction in grain (1,3;1,4)-β-glucan levels but maintain elite barley yields and for G748D and

248   G847E also TGW (Fig. 3j,l and Supplementary Notes 3, Section 2). While grains of variants *cslF6^G847E*

249   and *cslF6^T709I* are susceptible to breakage in hard mechanical threshing (Fig. 3k), variant *cslF6^G748D* is

250   proving valuable in malting and brewing processes, where high (1,3;1,4)-β-glucan causes filtration

251   difficulties and undesirable hazes in the final product[30].

252   **FIND-IT beyond barley.**

253   The broad applicability of FIND-IT makes it possible to screen large variant libraries of any living

254   organism that can be grown in the field or in culture, as highlighted here by additional examples in

255   wheat, yeast (*Saccharomyces cerevisiae)* and bacteria (*Lactobacillus pasteurii)* (Supplementary Notes

256   3, Section 3). Diverse libraries of rapeseed and oat are also ready for variant targeting

257   (Supplementary Notes 1, Section 1).

258   # Discussion

259   Here we have described and validated the FIND-IT method as a new species- and variety-agnostic

260   method for untargeted generation and targeted isolation of single nucleotide variants in genome

261   pools of plants and other biological systems. FIND-IT relies on three crucial components, *i)* the

262   century old method of induced mutagenesis, especially through chemical treatment, *ii)* a sample

263   pool-and-split strategy , and *iii)* high-sensitivity ddPCR technology[15]. Based on the integrated use of

264   these three components, FIND-IT allows the size of routinely screen-able variant libraries to be

11

265   expanded beyond 500,000 individuals and enables variant pool generation towards single nucleotide

266   resolution. The size of the library and its mutation load are inversely correlated with respect to the

267   probability of finding the variant of interest. FIND-IT allows each of these two parameters to be

268   independently adjusted to meet the intended application of the library (e.g. for direct incorporation

269   into a breeding pipeline or for functional analysis of individual genes).

270   In the present study, we chose low or high mutagen concentrations in combination with large library

271   sizes to demonstrate the versatility of the method and to optimize the fast inclusion of the identified

272   variant of interest into breeding programs with no or little backcrossing required. We show that very

273   large libraries can be efficiently handled using the demonstrated protocol and that the low overall

274   library mutation load achieved here nevertheless provides a high probability of finding *i)* a knockout

275   of a specified target gene, *ii)* a non-synonymous variant in a single target codon and *iii)* a single

276   nucleotide variant in a promoter element or a miRNA binding site. The total number of more than

277   150 targeted and successfully identified single nucleotide variants show the high-capacity of the

278   FIND-IT method. We demonstrate the strength of FIND-IT to quickly evolve elite germplasms in

279   accelerated commercial precision breeding through isolation, detailed analysis, and direct field

280   assessment of multiple variants. These are either novel or mimicked from existing natural cereal

281   variants and improve important agronomical or grain quality traits.

282   Precision breeding by CRISPR-Cas9 mediated genome editing of crop plants is rapidly developing

283   despite unwanted off-target edits[4,31], which are challenging to identify[32]. However, full applicability

284   to commercial breeding pipelines awaits further improvements, such as solving the need for time-

285   consuming and potentially chromosome rearranging transformation and/or tissue culture steps that

286   limit the technology to the relatively small number of plant varieties that are transformable and will

287   grow in culture[33], or where *de novo* meristem induction can be used[34]. Across multiple jurisdictions,

288   lines generated by CRISPR-Cas9 currently fall under GMO guidelines[35], and are subject to the

289   regulatory barriers associated with GMO methodologies. These intrinsic drawbacks and

290   uncertainties in the accompanying legislation of CRISPR-Cas9 driven gene editing currently precludes

291   its widespread utilization for accelerated development of crops with desired traits. Other traditional

292   breeding methods dependent on phenotypic or targeted selection from natural or induced genome

293   pools are either slow and imprecise or limited in population size and therefore restricted in their

294   resolution. FIND-IT releases the bottlenecks of phenotypic selection and TILLING, and stands

295   together with evolving CRISPR-Cas9 technologies for the efficient improvement of today's

296   germplasms.

297   We have successfully initiated FIND-IT based targeting of single nucleotide variants in elite-varieties

298   of barley, wheat, and in industrial production strains of yeast and bacteria. This demonstrates the

299   species-independent potential of FIND-IT for precision breeding.

300   FIND-IT will prove invaluable in new approaches to crop improvement by providing a direct means to

301   select and implement: *i*) domestication traits into undomesticated or semi-domesticated plants such

302   as perennials or wild crop relatives (e.g. quinoa[36]), *ii*) potentially valuable gene alleles retained in

303   wild relatives that were replaced during domestication[37], *iii*) gene alleles found in variomes (e.g.

304   exome collections and pan genomes[7,38–40]) and *iv*) novel improved variants identified *in vitro*.

305   In conclusion, FIND-IT constitutes a powerful and ready-to-use, technological innovation-platform

306   for trait development that allows fast and simple precision breeding of the high-yielding, nutrient-

307   use-efficient and stress-tolerant crops that have become essential for food security in a world faced

308   with a changing climate and increasing population.

Acknowledgements

312

13

Author Contributions

S.K., T.W., C.D. and H.C.T. contributed equally. S.K., T.W. and B.S. conceived the project, S.K., H.C.T., C.V., T.W., M.R., C.D. and P.R.P. were responsible for the mutagenesis and variant propagation, L.T.P., C.V., S.K., H.C.T., T.W., M.C., M.R., A.S., L.M., M.K.J. and J.A.C.S. generated plant libraries, Q.L., C.D., M.E.J., J.T.Ø., K.N., J.O.H. validated library mutation load, H.C.T., L.T.P., C.V., M.C., M.R. and T.W. performed the ddPCR experiments, C.D., J.A.C.S., I.B., S.B. and S.K. designed barley variant targets, C.D., S.B., J.A.C.S., P.R.P., E. M., J.H., H.C.T., M.T.S.N. and S.K. analysed barley variants, K.L., M.K., R.G.S. and N.S. prepared and screened the yeast library, R.T.F. and T.W. prepared and screened the bacterial library, Z.G., J.F., O.O. and B.L.M. assisted with experimental design, and C.D., M.E.J., H.C.T., M. R., B.L.M., G.B.F. and B.S. were responsible for the preparation of the manuscript, with assistance from the other authors.

Competing interests

S.K., T.W., H.C.T, M.R., M.C., A.S., O.O. and B.S. are listed as inventors in a patent application WO2018/001884, and S.K., T.W., M.R., J.Ø and B.S. are listed as inventors in patent application PCT/EP2020/078321, both relating to the methodology and filed by the Carlsberg A/S. Furthermore, S.K., S.B, O.O., H.T., T.W. and J.H. are listed as inventors in a patent application WO2019129736A1, and C.D., P.R.P., S.K., O.O., T.W., M.C., H.T. and M.R. are listed as inventors in a patent application WO2019129739A1, both relating to traits and filed by the Carlsberg A/S. All authors are previous or current Carlsberg A/S employees. B.L.M. is a Distinguished Professor at the Carlsberg Research Laboratory and serves as a consultant for the laboratory. G.B.F. is a Senior Scientific Consultant and Distinguished Professor of Raw Materials at the Carlsberg Research Laboratory. The methodology presented in the manuscript is applied commercially by Traitomic, a business unit at Carlsberg A/S.

References

1. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **3**, 1293 (2012).

338   2.   Voss-Fels, K. P., Stahl, A. & Hickey, L. T. Q&A: Modern crop breeding for future food security.
339   *BMC Biol.* **17**, 1–7 (2019).

340   3.   Hickey, L. T. *et al.* Breeding crops to feed 10 billion. *Nat. Biotechnol.* **37**, 744–754 (2019).

341   4.   Zhang, Y., Pribil, M., Palmgren, M. & Gao, C. A CRISPR way for accelerating improvement of
342   food crops. *Nat. Food* **1**, 200–205 (2020).

343   5.   Zhang, D. *et al.* Tissue Culture-Induced Heritable Genomic Variation in Rice , and Their
344   Phenotypic Implications. **9**, 5–7 (2014).

345   6.   Phillips, R. L., Kaeppler, S. M. & Olhoft, P. Genetic instability of plant tissue cultures:
346   breakdown of normal controls. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5222–5226 (1994).

347   7.   Jayakodi, M. *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding.
348   (2020) doi:10.1038/s41586-020-2947-8.

349   8.   Bradshaw, John E: Edited by Q. Y. Shu, B. P. F. and H. N. Plant Mutation Breeding and
350   Biotechnology. *Exp. Agric.* **49**, 476 (2013).

351   9.   Peng, J. *et al.* 'Green revolution' genes encode mutant gibberellin response modulators.
352   *Nature* **400**, 256–261 (1999).

353   10.   Henry, I. M. *et al.* Efficient genome-wide detection and cataloging of EMS-induced mutations
354   using exome capture and next-generation sequencing. *Plant Cell* **26**, 1382–1397 (2014).

355   11.   Till, B. J. *et al.* High-throughput TILLING for functional genomics. in *Plant functional genomics*
356   205–220 (Springer, 2003).

357   12.   Comai, L. *et al.* Efficient discovery of DNA polymorphisms in natural populations by Ecotilling.
358   *Plant J.* **37**, 778–786 (2004).

359   13.   Schreiber, M. *et al.* A highly mutagenised barley (cv. Golden Promise) TILLING population
360   coupled with strategies for screening-by-sequencing. *Plant Methods* **15**, 99 (2019).

361   14.   Li, G. *et al.* The sequences of 1504 mutants in the model rice variety kitaake facilitate rapid
362   functional genomic studies. *Plant Cell* **29**, 1218–1231 (2017).

363   15.   Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of
364   DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).

365   16.   Page, D. R. & Grossniklaus, U. The art and design of genetic screens: Arabidopsis thaliana.
366   *Nat. Rev. Genet.* **3**, 124–136 (2002).

367   17.   von Post, R. *et al.* A high-throughput DNA extraction method for barley seed. *Euphytica* **130**,
368   255–260 (2003).

369   18.   Taketa, S. *et al.* Barley grain with adhering hulls is controlled by an ERF family transcription
370   factor gene regulating a lipid biosynthesis pathway. *Proc. Natl. Acad. Sci.* **105**, 4062–4067
371   (2008).

372   19.   Bull, H. *et al.* Barley SIX-ROWED SPIKE3 encodes a putative Jumonji C-type H3K9me2/me3
373   demethylase that represses lateral spikelet fertility. *Nat. Commun.* **8**, 1–9 (2017).

374   20.   Hu, G., Burton, C., Hong, Z. & Jackson, E. A mutation of the cellulose-synthase-like (CslF6)
375   gene in barley (Hordeum vulgare L.) partially affects the β-glucan content in grains. *J. Cereal*
376   *Sci.* **59**, 189–195 (2014).

377   21.   Garcia-Gimenez, G. *et al.* Targeted mutation of barley (1,3;1,4)-β-glucan synthases reveals

378 complex relationships between the storage and cell wall polysaccharide content. *Plant J.* **104**,
379 1009–1022 (2020).

380 22. Patron, N. J. *et al.* The altered pattern of amylose accumulation in the endosperm of low-
381 amylose barley cultivars is attributable to a single mutant allele of granule-bound starch
382 synthase I with a deletion in the 5'-non-coding region. *Plant Physiol.* **130**, 190–198 (2002).

383 23. Denyer, K. a. y., Johnson, P., Zeeman, S. & Smith, A. M. The control of amylose synthesis. *J.*
384 *Plant Physiol.* **158**, 479–487 (2001).

385 24. Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. & Lin, H.-X. A QTL for rice grain width and weight
386 encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–630 (2007).

387 25. Xu, X. *et al.* Genome-Wide Association Analysis of Grain Yield-Associated Traits in a Pan-
388 European Barley Cultivar Collection. *Plant Genome* **11**, 170073 (2018).

389 26. Li, S. *et al.* Modulating plant growth–metabolism coordination for sustainable agriculture.
390 *Nature* **560**, 595–600 (2018).

391 27. Betts, N. S. *et al.* Transcriptional and biochemical analyses of gibberellin expression and
392 content in germinated barley grain. *J. Exp. Bot.* **71**, 1870–1884 (2020).

393 28. Zou, X., Neuman, D. & Shen, Q. J. Interactions of two transcriptional repressors and two
394 transcriptional activators in modulating gibberellin signaling in aleurone cells. *Plant Physiol.*
395 **148**, 176–186 (2008).

396 29. Dimitroff, G. *et al.* (1,3;1,4)-β-Glucan Biosynthesis by the CSLF6 Enzyme: Position and
397 Flexibility of Catalytic Residues Influence Product Fine Structure. *Biochemistry* **55**, 2054–2061
398 (2016).

399 30. Bamforth, C. W. *Beer: Tap Into the Art and Science of Brewing*. (Oxford University Press,
400 2009).

401 31. Tang, X. *et al.* Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Mol. Plant* **13**,
402 667–670 (2020).

403 32. Blondal, T. *et al.* Verification of CRISPR editing and finding transgenic inserts by Xdrop indirect
404 sequence capture followed by short- and long-read sequencing. *Methods* (2021)
405 doi:10.1016/j.ymeth.2021.02.003.

406 33. Jansing, J., Schiermeyer, A., Schillberg, S., Fischer, R. & Bortesi, L. Genome Editing in
407 Agriculture : Technical and Practical Considerations. (2019).

408 34. Maher, M. F. *et al.* Plant gene editing through de novo induction of meristems. *Nat.*
409 *Biotechnol.* **38**, (2020).

410 35. Callaway, E. CRISPR plants now subject to tough GM laws in European Union. *Nature* **560**, 16–
411 17 (2018).

412 36. López-Marqués, R. L. *et al.* Prospects for the accelerated improvement of the resilient crop
413 quinoa. *J. Exp. Bot.* **71**, 5333–5347 (2020).

414 37. Langridge, P. & Waugh, R. Harnessing the potential of germplasm collections. *Nat. Genet.* **51**,
415 200–201 (2019).

416 38. Russell, J. *et al.* Exome sequencing of geographically diverse barley landraces and wild
417 relatives gives insights into environmental adaptation. (2016) doi:10.1038/ng.3612.

418 39. Walkowiak, S. *et al.* Multiple wheat genomes reveal global variation in modern breeding. *Nat.*

419   *2020* 1–7 (2020) doi:10.1038/s41586-020-2961-x.

420   40.   Zhao, Q. *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated
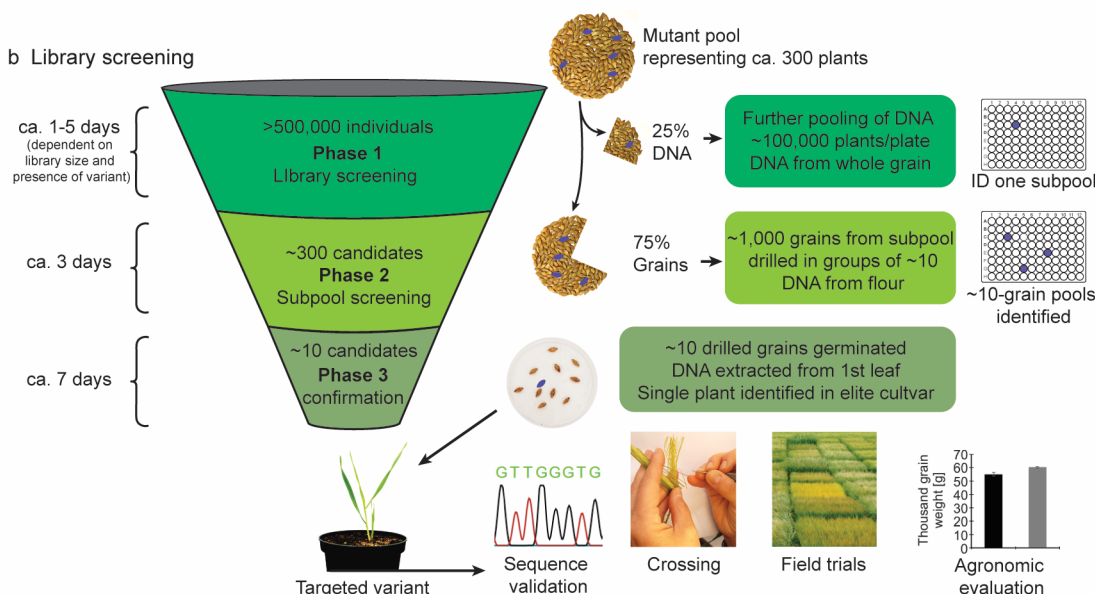421   and wild rice. *Nat. Genet.* **50**, 278–284 (2018).

422

423

Figures
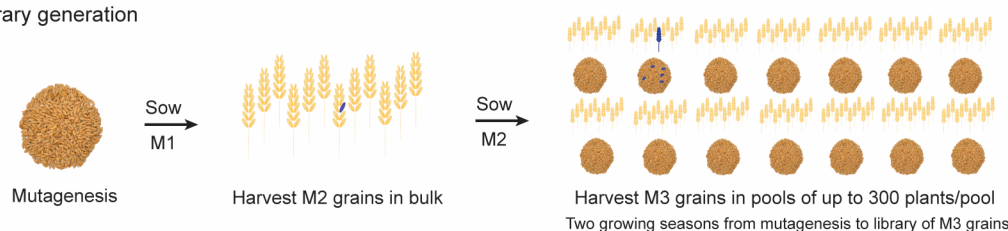
Fig. 1. FIND-IT screening procedure. a, Barley library generation: Mutagenized barley plants are grown densely

to maturity in the field. M3 grains are harvested in pools of approximately 300 plants. **b**, Library screening

consists of three phases that narrow down the search of a target variant from >500,000 to eventually a single

plant within days. See detailed description of the method in Main Text, Online Methods and Supplementary

Notes 3, Section 1. Screening Phase 1: The harvested grain pools are each split into two fractions; one fraction

(25%) is utilized for total DNA extraction, the other (75%) is stored for further analysis. DNA from the 25%

pools is further pooled and one 96-well library plate contains DNA from >100,000 individuals. The following

ddPCR screening (Fig. Supplementary Notes 3.2a), using competing fluorescently labelled probes to target the

nucleotide variants of interest, identifies one well, representing ~300 candidates, for further analysis.

Screening Phase 2: the corresponding 75% grain pool of the identified DNA pool in Phase 1 is selected for

further analysis. DNA from approximately 1,000 grains is sampled non-destructively in pools of 10 and

rescreened using ddPCR (Fig. Supplementary Notes 3.2b), and several wells are potentially identified to

contain the targeted variant. Screening Phase 3: the grains from the identified wells in Phase 2 are germinated

18

439   and total DNA is extracted from the first leaf of each individual plant and re-screened by ddPCR (Fig.

440   Supplementary Notes 3.2c). From this, a single plant is identified containing the targeted variant which can be

441   sequence-validated (Fig. Supplementary Notes 3.2d). Identified variants of interest are immediately available

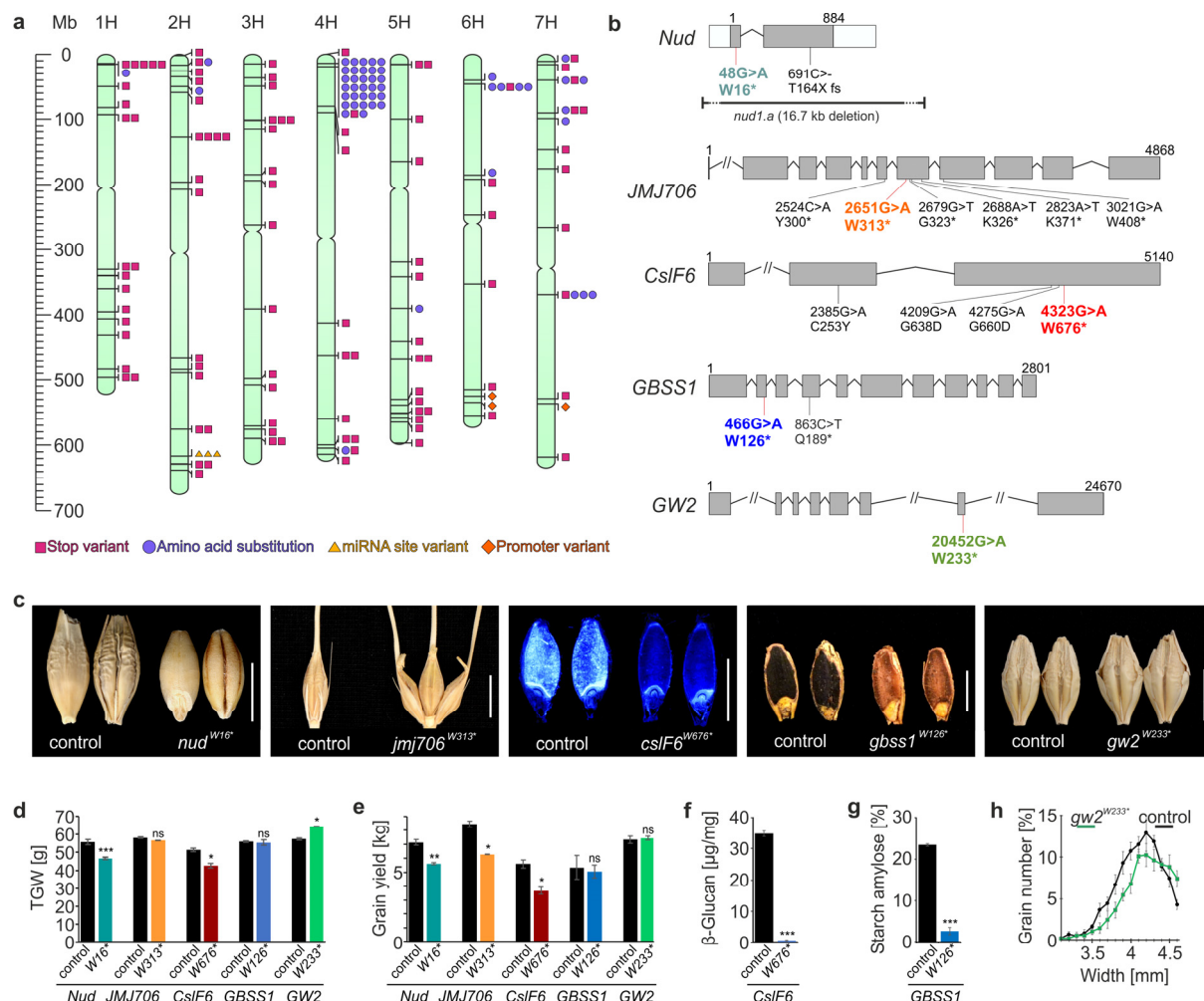442   for field trials, agronomic analyses, and incorporation into routine breeding backcrossing programs.

443

Fig. 2. Barley knockout library with selected variants isolated and agronomically evaluated. a, Barley physical reference genome map with chromosomes 1H – 7H with more than 150 identified stop and other SNP mutations indicated by number (Supplementary Table 1). **b**, Gene models of selected barley reference targets with known loss-of-function mutations and novel, isolated stop variants (in colour). **c**, Loss-of-function phenotypes of novel variants (hulless grain, nud^{W16*}; six-rowed spikelets, $jmj706^{W313*}$; (1,3;1,4)-β-glucan Calcofluor staining, $cslF6^{W676*}$; Starch amylose Lugol staining, $gbss1^{W126*}$; grain width, $gw2^{W233*}$). **d**, TGW and **e**, grain yield of field-grown novel barley variants. **f**, Grain (1,3;1,4)-β-glucan content of $cslF6^{W676*}$ grain. **g**, Starch amylose content of $gbss1^{W126*}$ grain. **h**, Grain width distribution of $gw2^{W233*}$ grain. Scale bars in c, 5 mm. Error bars are standard diviation, two-tailed t-test was performed to obtain P values (*P<0.05, **P<0.01, ***P<0.001, ns not statistically significant). See Supplementary Table 3 for statistical test details and number of replicates.
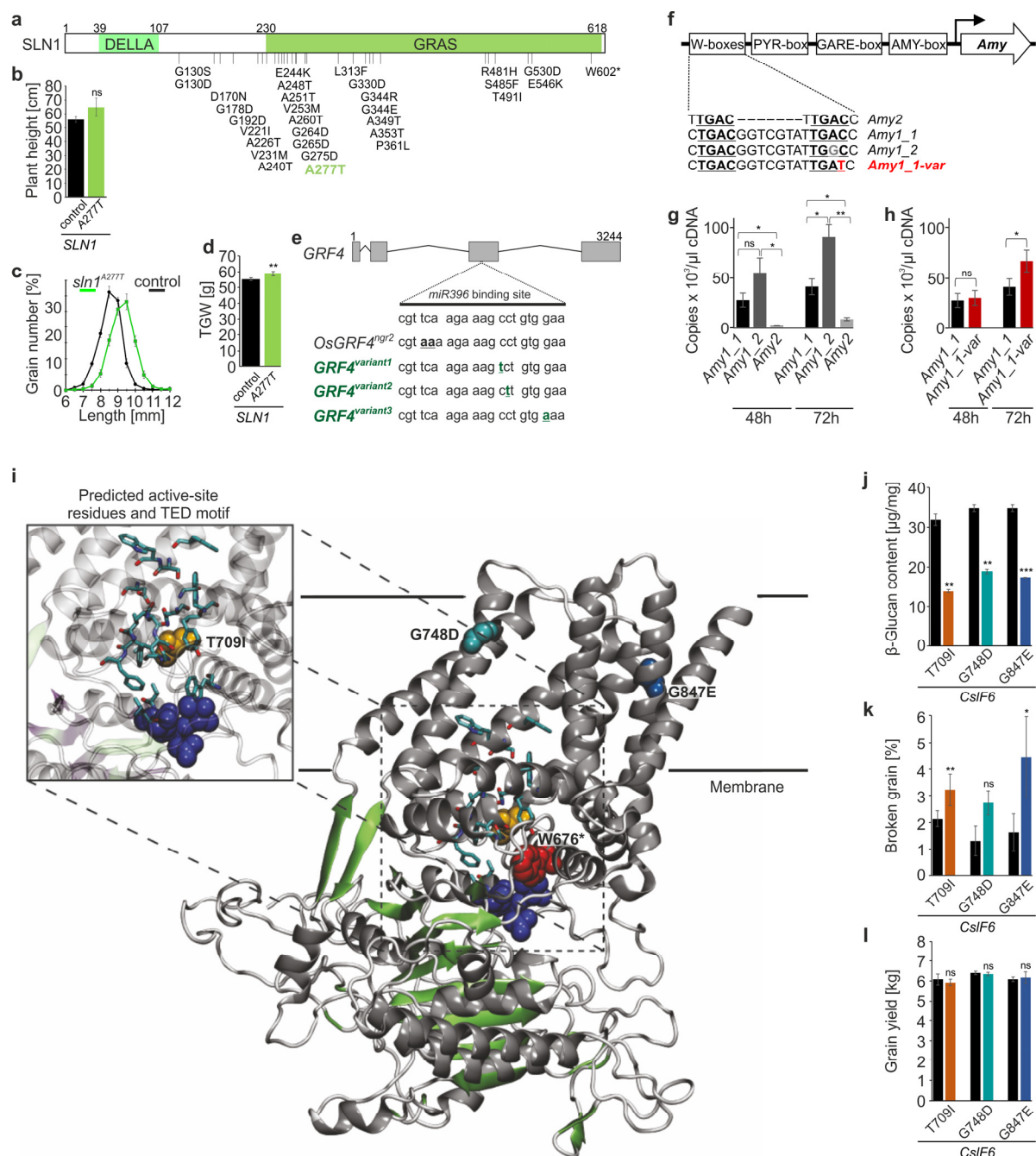
20

Fig. 3. FIND-IT libraries with distinct non-synonymous nucleotide variants and variants to modulate transcript abundance for precision breeding. a, Identified and isolated variants (in black and green, respectively) of the SLN1/DELLA protein. **b,** Plant height, **c,** grain length distribution and **d,** TGW of *sln1^A277T*. **e,** Gene model of barley *GRF4* showing the 21-bp *miR396* binding site on exon 3 including a known double substitution in rice (*OsGRF4^ngr2*) and three novel barley variants (in green). **f,** Known cis-acting elements in α-amylase promoters with natural and induced W-box sequence variation. TGAC core motifs in W-boxes are underlined and in bold, natural variation in the *Amy1_2* gene in grey and a novel, induced variation in an

464   *Amy1_1* gene in red. **g**, Expression of α-amylase genes in Planet grain 48h and 72h after initiation of

465   germination. **h**, Expression of *Amy1_1* genes in Planet grain (black) and Planet variant *Amy1_1-var* (red) 48h

466   and 72h after initiation of germination. **i**, Protein homology model of CslF6 based on poplar cellulose synthase

467   isoform 8 (PDB# 6wlb.1.A)**.** Amino acid positions at residue W676, T709, G748 and G847 are shown as spheres.

468   Insert shows UDP- glucose binding sites - predicted by alignment of CslF6 and poplar cellulose synthase

469   isoform 8 - as light blue sticks and the TED motif is shown as dark blue sticks. The insert is shown at a slightly

470   tilted angle to visually optimize the position of the residues discussed. **j**, Grain (1,3;1,4)-β-glucan content of

471   controls (black bars) and *CslF6* variants. **k**, Percentage of broken grain after threshing in controls (black bars)

472   and *CslF6* variants. **l**, Field grain yield of controls (black bars) and *CslF6* variants. Error bars are standard

473   diviation, two-tailed t-test was performed to obtain P values (*P<0.05, **P<0.01, ***P<0.001, ns not

474   statistically significant). See Supplementary Table 3 for statistical test details and number of replicates.