

## Scotland's Rural College

Genome-wide association mapping of Hagberg falling number, protein content, test weight and grain yield in UK wheat

White, Jon; Sharma, Rajiv; Balding, David; Cockram, James; Mackay, Ian J.

*Published in:*  
Crop Science

*DOI:*  
[10.1002/csc2.20692](https://doi.org/10.1002/csc2.20692)

Print publication: 01/05/2022

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (APA):*

White, J., Sharma, R., Balding, D., Cockram, J., & Mackay, I. J. (2022). Genome-wide association mapping of Hagberg falling number, protein content, test weight and grain yield in UK wheat. *Crop Science*, 62(3), 965-981. <https://doi.org/10.1002/csc2.20692>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## ORIGINAL RESEARCH ARTICLE

## Crop Breeding &amp; Genetics

## Genome-wide association mapping of Hagberg falling number, protein content, test weight, and grain yield in U.K. wheat

Jon White<sup>1,2,#</sup> | Rajiv Sharma<sup>3,#</sup> | David Balding<sup>2,4</sup> | James Cockram<sup>1,†</sup>  | Ian J. Mackay<sup>3,†</sup><sup>1</sup>Genetics and Breeding Dep., NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, UK<sup>2</sup>Institute of Genetics, Univ. College London, London, WC1E 6BT, UK<sup>3</sup>Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh, EH9 3JG, UK<sup>4</sup>Current address: Melbourne Integrative Genomics, Univ. of Melbourne, Melbourne, Australia

## Correspondence

James Cockram, NIAB, 93 Lawrence Weaver Road, Cambridge, CB3 0LE, UK.  
Ian J. Mackay, Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh, EH9 3JG, UK.  
Email: [james.cockram@niab.com](mailto:james.cockram@niab.com);  
[ian.mackay@sruc.ac.uk](mailto:ian.mackay@sruc.ac.uk)

Assigned to Associate Editor Georgia Eizenga.

#Equal first author contribution.

†Equal last author contribution.

## Funding information

Biotechnology and Biological Sciences Research Council, Grant/Award Numbers: BB/J002542/1, BB/M011666/1

## Abstract

Association mapping using crop cultivars allows identification of genetic loci of direct relevance to breeding. Here, 150 U.K. wheat (*Triticum aestivum* L.) cultivars genotyped with 23,288 single nucleotide polymorphisms (SNPs) were used for genome-wide association studies (GWAS) using historical phenotypic data for grain protein content, Hagberg falling number (HFN), test weight, and grain yield. Power calculations indicated experimental design would enable detection of quantitative trait loci (QTL) explaining  $\geq 20\%$  of the variation (PVE) at a relatively high power of  $>80\%$ , falling to 40% for detection of a SNP with an  $R^2 \geq .5$  with the same QTL. Genome-wide association studies identified marker-trait associations for all four traits. For HFN ( $h^2 = .89$ ), six QTL were identified, including a major locus on chromosome 7B explaining 49% PVE and reducing HFN by 44 s. For protein content ( $h^2 = 0.86$ ), 10 QTL were found on chromosomes 1A, 2A, 2B, 3A, 3B, and 6B, together explaining 48.9% PVE. For test weight, five QTL were identified (one on 1B and four on 3B; 26.3% PVE). Finally, 14 loci were identified for grain yield ( $h^2 = 0.95$ ) on eight chromosomes (1A, 2A, 2B, 2D, 3A, 5B, 6A, 6B; 68.1% PVE), of which five were located within 16 Mbp of genetic regions previously identified as under breeder selection in European wheat. Our study demonstrates the utility of exploiting historical crop datasets, identifying genomic targets for independent validation, and ultimately for wheat genetic improvement.

**Abbreviations:** GWAS, genome-wide association study; HFN, Hagberg falling number; HGCA, Home Grown Cereals Authority; LD, linkage disequilibrium; MAF, minor allele frequency; PCA, principal component analysis; PHS, preharvest sprouting; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Crop Science* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

## 1 | INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is the world's most widely cultivated cereal crop, with global demand projected to increase by 60% by 2050. Breeding over the last 100 years has been very successful in increasing wheat yields (e.g., Mackay et al., 2021), and the most sustainable way to meet increased future demand is via faster genetic improvement.

Accordingly, the development of high yielding wheat cultivars with good bread-making quality is a major focus for wheat breeders (Reif et al., 2011). Over recent decades, breeding has resulted in steady genetic gains in European wheat (Mackay et al., 2011). With the aim of further increasing the rates of wheat genetic improvement, the incorporation of marker-assisted strategies into breeding pipelines is now common. In marker-assisted selection (MAS) approaches, genetic loci must first be identified, and linked or diagnostic markers developed (Adamski et al., 2020). Historically, biparental mapping populations have predominantly been used for genetic mapping in wheat. A disadvantage of such approaches is that such populations capture limited genetic variation, and that the effects of quantitative trait loci (QTL) can be overestimated (Cockram & Mackay, 2018). The use of these QTL in marker-assisted selection often does not give expected results, as the biparental populations investigated may not represent breeding germplasm or may not segregate for the targeted QTL (Cockram & Mackay, 2018; Kulwal et al., 2012). More recently, association mapping, first developed for human genetics (Bodmer, 1986; Wellcome Trust Case Control Consortium, 2007), has been used for genetic analysis in cereal crops (e.g., Bentley et al., 2014; Cockram et al., 2008, 2010; Corsi et al., 2020; Mellers et al., 2020; Waugh et al., 2010). These capture high levels of genetic diversity and genetic recombination (Cavanagh et al., 2008). Additionally, historical phenotypic data is often available (Mackay et al., 2011). A useful source of historical traits of relevance to growers and users are those collected during the process of cultivar registration (Jamali et al., 2019). In the United Kingdom, four such traits of agronomic or end-use importance are routinely measured during the process of wheat cultivar registration for the Agriculture and Horticulture Development Board Recommended List: grain yield, protein content, Hagberg falling number (HFN), and test weight (also known as 'specific weight').

Grain yield is the principle agronomic trait for wheat breeders and growers. Yield is often found to be of low to moderate heritability, although it shows clear evidence for genetic improvement—for example in U.K. wheat yields over a period of 30 years (Mackay et al., 2011). However, historical datasets however often cover a far broader range of cultivar yields, with each cultivar also tested over several years at many locations, with a consequential and substantial increase in heritability. However, the highly polygenic nature of yield means identification of QTL remains challenging. At a fundamental level, grain yield can be described using three major factors: grain size (e.g., grain length, width, weight, shape), the numbers of grains per ear, and the number of ears within a specific area (Gegas et al., 2010). These are in turn influenced by other traits, such as accumulation and transport of photosynthetic products, flag leaf size, plant height, biomass, and the rate of plant development across its lifecycle (most com-

### Core Ideas

- An association mapping panel of 150 genotyped U.K. winter wheats was assembled.
- Data for four grain quality traits was sourced from cultivar registration records.
- Genome-wide association studies (GWAS) identified at least five QTL for each trait.
- A major GWAS hit for Hagberg falling number was identified on chromosome 7A.
- A third of yield GWAS hits co-located with genetic loci under breeder selection.

monly determined by measurement of flowering time). Additionally, the postharvest processing of wheat grain for subsequent food use mean that traits that affect milling performance, such as the shape, size, density, and uniformity of the grain, are important for flour yield. In addition to yield, there are several grain quality characters that are important in determining end use, and thus the monetary value of the grain. In the United Kingdom, wheat cultivars are classified into four quality groups by the National Association of British and Irish Millers (Nabim, <http://www.nabim.org/>), based on grain protein content, HFN (a measure of  $\alpha$ -amylase enzymatic activity in sprout-damaged grain) and test weight (a measure of grain density, determined by the weight of a set volume of grain). Nabim Group 1 and Group 2 cultivars are high protein, hard grained cultivars suitable for bread making and attract a price premium. Group 3 represents cultivars with lower protein (~11%) and soft milling characteristics due to their soft grain phenotype and are typically used for biscuit making. Group 4 cultivars have low protein content, can be either hard or soft grained, and are used predominantly for animal feed (UK Flour Millers, 2017).

Protein content in wheat determines how flour performs during baking and is normally measured as percent of dry matter using near infrared spectrometry (NIR). For bread making (Nabim Groups 1–2), grain protein content values of  $\geq 13\%$  are required, whereas for biscuit making (Group 3), grain protein content levels are typically 11–11.5% (UK Flour Millers, 2017). In addition, protein content represents an important factor for the nutritional value of the grain or flour: annual wheat production is estimated to be ~730 Tg (FAO, 2016), which translates to ~73 Tg of protein for human and animal consumption. Protein is predominantly present in the form of grain storage proteins, including the high molecular weight glutenin subunits (HMW-GS), low molecular weight glutenin subunits (LMW-GS), and the gliadins. Grain quality is determined by the quantity and quality of the storage proteins

present, as well as the ratio of glutenin to gliadin (Jouanin et al., 2018).

The HFN (Hagberg, 1960) is an internationally standardized method to assess starch properties in wheat flour. The test measures the time in seconds a stirrer that is placed in a test tube of heated coarse meal-water mix takes to fall down through the mixture, thus determining the effects of  $\alpha$ -amylase enzymes on starch properties within the mixture (Mohler et al., 2014). The HFN is principally determined by two factors: preharvest sprouting (PHS), a loss of dormancy resulting in premature germination of the grain while still on the parent plant, and late-maturity amylase, the untimely synthesis of high levels of high isoelectric point isozymes of  $\alpha$ -amylase during the latter stages of grain development, without sprouting. In both cases,  $\alpha$ -amylase activity during the later stages of grain development lead to starch breakdown and reduced grain quality (Mares & Mrva, 2008), resulting in lower HFN values due to reduced viscosity of the mixtures used to assess HFN. In the United Kingdom, grain suitable for bread making must have a HFN  $\geq 250$  s (UK Flour Millers, 2017). Grain which doesn't meet this threshold result in breads with lower volume and a compact, sticky crumb structure (Gale & Lenton, 1987).

Test weight, also known as specific weight, is a measurement of the weight of a known volume of grain and is an important indicator of quality. Low test weight can indicate poorly filled or misshapen grains, and/or high moisture content. Wheat with a low test weight generally results in low flour extraction rates. For bread making quality, test weights of more than 76 kg ha<sup>-1</sup> are generally required (UK Flour Millers, 2017). Relatively few studies reporting QTL for test weight have been published (e.g., Narasimhamoorthy et al., 2006), one of the most recent of which identified eight loci were identified on chromosomes 1D, 2A, 2B, 2D, 3B, 3D, 4D, and 7A (Cabral et al., 2018). However, test weight is largely determined by grain weight, shape and volume, and so can be broken down into its constitutive components and investigated as separate yield components (e.g., Corsi et al., 2021).

Previous genome-wide association study (GWAS) analysis in U.K. wheat have used the available historical yield data to identify QTLs controlling grain yield, identifying QTL located on chromosomes 1A and 5A (Sharma et al., 2021). Here, the aim of our study was to use historical grain quality and yield datasets to identify genetic loci controlling these traits. To this end, a panel of 150 elite U.K. winter wheat cultivars genotyped with a 90,000-feature single nucleotide polymorphism (SNP) genotyping array was used to perform GWAS for the three grain quality traits measured during wheat registration in the United Kingdom: grain protein content, HFN, and test weight. Although our panel overlaps with to some extent with that recently used by Sharma et al. (2021) to undertake GWAS on grain yield, we also include yield in our analysis here, principally to help any direct compar-

isons with marker-trait association made for our analysis of grain protein content, HFN, and test weight. We discuss the genetic loci identified in the context of previously published QTL.

## 2 | MATERIALS AND METHODS

### 2.1 | Wheat germplasm

An association mapping panel of 150 U.K. winter wheat cultivars was assembled, consisting of lines released between 1965 and 2004, and that had undergone at least 3 yr of phenotypic evaluation under Home Grown Cereals Authority (HGCA; Since termed the 'Agriculture and Horticulture Development Board') National List trials between 1975 and 2007 (Supplemental Table S1). Seed of the cultivars was obtained from seed banks (John Innes Centre, UK; IPK Genebank, Germany; USDA National Small Grains Collection, USA), or from a reference seed collection held by NIAB on behalf of Defra, with the written permission of the relevant breeding companies.

### 2.2 | Historical phenotypic data, estimates of heritability, and trait correlations

Historical phenotypic data recorded during HGCA Recommended List trials for grain protein content (measured as percent dry matter), HFN (measured in seconds), and test weight (the weight in grams of seed that can be packed into a 1-L standard container) from HGCA trials between 1975 and 2007 were sourced, with permission, from HGCA Recommended List trial data (Table 1). Historical phenotyping followed standard HGCA National List protocols (<http://www.hgca.com/>). Estimates of cultivar effect for each trait were obtained from historical experimental data originating from UK National List and Recommended List trials (1950–2004). The variance structure of the phenotypic data in a mixed effects model was estimated using REML (Genstat Version v.11, VSN International). We estimated best linear unbiased estimates using the models summarized in Supplemental Table S2. Heritability ( $h^2$ ), more strictly defined as repeatability, was estimated from the random model compromising only main effects, the denominator of the error term being estimated from the ratio of stratum variance to variance component for cultivar. Phenotypic correlations were investigated using R-package "Hmisc" and the correlation values highlighted in Microsoft Excel 2016 using conditional formatting. Where possible, Nabim group (1, 2, 3, or 4) classification was sourced from Agriculture and Horticulture Development Board Recommended List data, for Year 2004 onwards, available online at <https://ahdb.org.uk/knowledge-library/recommended-lists-archive>. Prior to 2004, data was sourced from nondigitized NIAB Pocket

TABLE 1 The volume and structure of historical winter wheat trials data

Trait	No. samples	No. years	No. vars	No. trials	Averages				
					Trial size (no. vars.)	Trials per year	Replications per var. <sup>a</sup>	Var. years in trial	Var. replications per year
Yield	74,332	32	1,219	1,472	50.5	46.0	61.0	2.2	22.0
HFN	21,458	31	1,324	586	36.6	18.9	16.2	2.0	8.1
Protein	17,097	31	1,145	691	24.7	22.3	14.9	1.8	8.4
TW	18,452	31	1,211	546	33.8	17.6	15.2	1.7	9.1

Note. HFN, Hagberg falling number; Protein, protein content; TW, test weight; Var., variety.

<sup>a</sup>The average number of estimations of the trait for a cultivar in the database.

Handbooks, and for those cultivars which predated the introduction of formal Nabim group classification in the Recommended List datasets, group was estimated based on the available quality data.

### 2.3 | Genotypic data and principal component analysis

Genotypic data for the 150 cultivars, generated using the 90,000 feature Illumina SNP array (Wang et al., 2014), was sourced from <https://www.niab.com/pages/id/326/Resources>. Principal component analysis (PCA) was undertaken using 16,801 genetically mapped markers as well as a set of 463 skimmed markers parsed using TASSEL (Bradbury et al., 2007) and further thinned to a minimum distance between markers of 10 cM. Information on the presence or absence of the 1B/1R wheat/rye chromosomal translocation (which has been widely used in wheat breeding programs as a source of beneficial traits such as disease resistance; Rabinovich, 1998) in the varietal panel was overlaid onto PCA plots.

### 2.4 | GWAS, bioinformatics, and power analysis

The SNPs were parsed based on minor allele frequency (MAF)  $\geq 0.05$  and  $R^2 \geq .2$ . The GWAS was undertaken using a mixed linear model to fit a linear mixed model with cultivars treated as random and SNPs effects as fixed. To account for population substructure in the association mapping panel, centered identity by state (IBS) based kinship was used. For the kinship matrix generation, a subset of markers was selected based on linkage disequilibrium (LD) implemented in the R-package SNPRelate (Zheng et al., 2012). The following equation was fitted in the mixed linear model:  $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$ , where  $\mathbf{y}$  is a vector representing traits values;  $\beta$  is the fixed effect of the SNP under test;  $\mathbf{u}$  refers to a vector of random additive genetic effects for the 150 lines;  $\mathbf{e}$  is the vector of error terms; and  $\mathbf{X}$  and  $\mathbf{Z}$  represents known

design matrices. More details are provided by Yu et al. (2006). Analysis was performed using options optimum compression level and P3D variance component estimation as implemented using the mixed linear model (MLM) method within the software TASSEL version 5.2.54 (Bradbury et al., 2007). Additionally, MLM method with both population structure and kinship relationship (MLM-PCA-K) was employed. In the Manhattan plots, SNPs are arranged in genetic map order (Wang et al., 2014), with unmapped markers excluded from the plots. An arbitrary significance threshold of  $-\log_{10}P = 3$  was used to report marker-trait associations. This threshold was supported by our power calculations, which estimated a false positive rate of 0.001 (i.e.,  $-\log_{10} = 3$ ). Significant markers within a conservatively defined 15-cM interval of each other were treated as identifying the same QTL. For each trait, the percentage of variation explained (PVE) was determined in two ways: (a) via the outputs of the software TASSEL; and (b) specifically just for the QTL detected by GWAS, calculated in R (R Core Team, <https://www.rproject.org/contributors.html>) using the regression `lm` function without population structure adjustments. The positions of selected SNPs on the wheat reference genome of ‘Chinese Spring’ (genome assembly RefSeq v1.0; IWGSC, 2018) were identified using the Triticeae Toolbox database via <https://triticeaetoolbox.org/wheat/>, which anchors SNPs based on nucleotide basic local alignment search tool (BLASTn) homology. Where an SNP was anchored to a different homoeologous chromosome to that identified in the genetic map (Gardner et al., 2016), physical map location was based on manual BLASTn analysis using Ensembl Plants (Yates et al., 2020). For cross-comparison of our GWAS hits with published QTL, other genetic marker types were anchored in a similar way, after extraction of available sequence data from GrainGenes (<https://wheat.pw.usda.gov/GG3/>) or in the case of Diversity Array Technology (DArT) markers, from information downloaded from <https://www.diversityarrays.com/technology-and-resources/sequences/>. Power calculations were performed using the custom R-scripts detailed in Supplemental Text S1, which implemented the power calculation functions described by Wang and Xu (2019). For

TABLE 2 Variance components and trait heritability

Trait	Variance components	Mean square		$(\text{VarMS} - \text{ResMS})/V_g$	$V_g$	ResMS/ $n_v$	$V_g/(V_g + V_e)$	Trait $h^2$ using all data
	$V_g$	Cultivar	Residual					
Yield	0.393	10.265	0.400	24.89	0.393	0.020	0.950	
Protein	0.226	2.017	0.287	7.67	0.226	0.037	0.858	
HFN	1591	13,287	1,437	7.45	1591	192.89	0.892	
TW	2.580	20.150	2.020	7.03	2.580	0.287	0.899	

Note.  $h^2$ , heritability; HFN, Hagberg falling number;  $n_v$ , average replication per cultivar; Protein, protein content; ResMS, residual mean square; TW, test weight VarMS, varieties mean square;  $V_g$ , genetic variance;  $V_e$ , environmental variance.

this, the kinship matrix was decomposed into eigenvalues and their corresponding eigenvectors. The eigenvalues were subsequently used to compute power using functions provided in the power calculations scripts published by Wang and Xu (2019). Power was calculated for QTL heritability values ranging from .01 to .4 and for linkage disequilibrium ranging from  $R^2 = .011$  to  $R^2 = 1$ .

## 2.5 | Associations with Nabim classifications

To test whether the classification of cultivars into Nabim quality groups accounted for any population structure, we used multiple regression of cultivar loadings of each of the first two principal components onto the Nabim classifications, with chromosome 1B/1R translocation, protein content, test weight, HFN, and grain hardness traits included as additional covariates. The importance of the Nabim classification was estimated by statistical significance in analysis of variance (ANOVA), with the factor for Nabim fitted both before and after the other covariates.

## 3 | RESULTS

### 3.1 | Analysis of phenotypic and genotypic data

Phenotypic data for protein content, HFN, and test weight were extracted from historical HGCA National List datasets. Summary data for each trait (including the number of samples, years, cultivars, and trials, as well as mean values) are listed in Table 1. Estimation of the variance structure of the phenotypic data is shown in Table 2. Random main effects were fitted for years, sites within years, management regimes, and cultivars. Residual variance was estimated from a model fitting random main effects for years, sites within years, management regimes, and cultivars. Average replication per cultivar was taken from the table of stratum variances. Estimates of

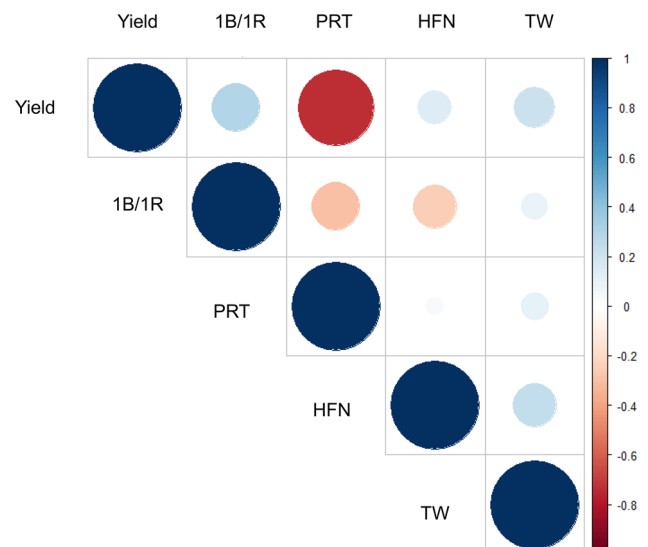
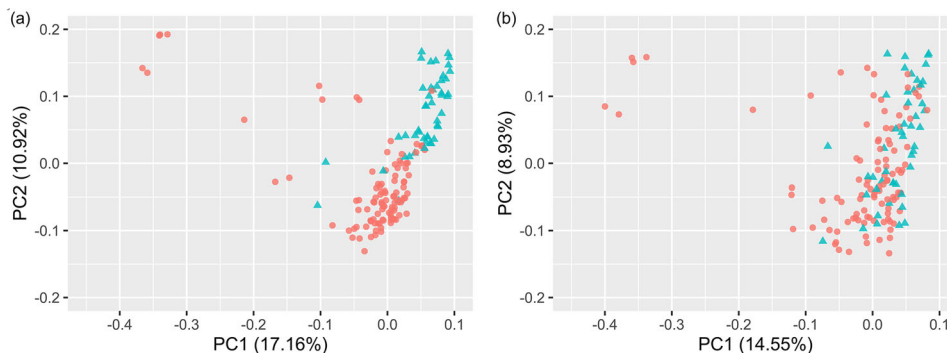


FIGURE 1 Pearson correlation between grain yield (Yield), the presence or absence of the wheat/rye chromosome 1B/1R translocation, grain protein content (PRT), Hagberg falling number (HFN), and test weight (TW)

$h^2$  for all traits was  $\geq .86$ , with grain yield showing the highest value (.95), followed by test weight (.90), HFN (.89), and protein content (.86). These high values are inevitable given the high average numbers of years of testing, locations, replication number, and high genetic variances. Explicitly for yield, in addition to far higher replication of locations and years in comparison to almost all contemporary experiments, the genetic variation is substantially inflated by inclusion of old and modern cultivars, spanning 40 years, during which period yield was increased substantially by genetics alone. Best linear unbiased estimates were estimated for all four traits and are listed in Supplemental Table S1. Traits showed broadly normal distributions, although yield was notable skewed towards higher values (Figure 1). Analysis of correlation between our three target traits, as well yield, found showed strong negative correlation between yield and protein content ( $P > .0001$ ), as well as significant negative correlations between the 1B/1R



**FIGURE 2** Population substructure in the association mapping panel of 150 accessions, based on (a) all markers ( $n = 16,801$ ), and (b) skimmed markers ( $n = 463$ ; linkage disequilibrium threshold  $R^2 = .2$ ). Markers with a minor allele frequency  $>0.05$  were used for both analyses. The presence or absence of the chromosome 1B/1R wheat/rye chromosomal substitution is indicated by red circles and blue triangles, respectively. PC, principal component

translocation and protein ( $P = .0003$ ) and HFN ( $P = .0183$ ) (Figure 1; Supplemental Table S3). Significant positive correlations were found between the 1B/1R translocation and yield ( $P = .0004$ ), and between test weight and both HFN ( $P = .0002$ ) and yield ( $P = .0124$ ).

### 3.2 | Population structure and power calculations

Genetic population substructure was investigated using a matrix of similarity between cultivars for PCA (Figure 2). The PCA using 16,801 polymorphic genetically mapped markers revealed a clear division in the first principal component. Overlaying data for the presence or absence of the 1B/1R wheat/rye chromosomal translocation indicated that the division based on Principal Component 2 was largely due to this chromosomal rearrangement (Figure 2). While PCA analysis after removal of one of each pair of highly correlated ( $R^2 > .98$ ) markers removed much of this distinction (Figure 2), the presence of further genetic substructure indicated statistical correction would likely be necessary for genome-wide association studies. To determine whether classification of cultivars into the four Nabim end-use grain quality groups could explain significant amounts of PCA space, we undertook analysis of variance. Nabim was found to explain a modest amount of the variation for each of the two PCA components (16% of PCA1,  $p = .0017^{**}$ ; 8% of PCA2,  $p = .0306^*$ ). However, Nabim group is a man-made ‘trait’, and is determined by a number of underlying quality traits. When such component traits (1B/1R translocation, protein content, test weight, HFN, grain hardness) were fitted before Nabim classification, then Nabim classification no longer identified significant variation in either PCA1 or PCA2 space.

To estimate the power of the experimental design to detect QTL, we undertook power analyses. These found that despite low number of accessions of this study ( $n = 150$ ) this association mapping panel was predicted to be able to detect QTL of reasonable effects (Supplemental Table S4). For example, the analysis predicts us to be able to identify QTL with a heritability of  $\geq 20\%$  with high power of  $>80\%$ , with power falling to 40% for a SNP in LD of  $R^2 \geq .5$  with a given QTL. With the relatively high density of markers in this panel, it is likely that most QTL will be in high LD with one or more markers (data not shown)—at least on the A and B subgenomes which have much higher marker density than the D subgenome (6,188, 8,488, and 2,125 genetically mapped markers, respectively).

### 3.3 | Genome-wide association studies

After removal of markers with  $MAF \leq 0.05$ , the final data-matrix for GWAS consisted of 150 cultivars and 20,921 SNPs (including 5,515 unmapped SNPs). After accounting for population substructure using a Kinship matrix, the results of GWAS identified significant ( $-\log_{10}P > 3$ ) marker-trait associations for all four of the phenotypes investigated. The GWAS results are displayed as Manhattan plots in Figure 3, histograms of the phenotypic trait expression are shown in Figure 4, and the genomic hits are summarized in Table 3, with the GWAS outputs for all markers used listed in Supplemental Table S5. The adoption of a significance threshold of  $-\log_{10}P = 3$  was arbitrary, but an excess of low  $P$ -values for all four traits is apparent from the Q-Q plots and histograms of  $P$ -values shown in Supplemental Figure S1. For HFN, 39 significant genetic markers with  $-\log_{10}P$ -values ranging from 3.10 to 3.00 were identified, which when modelled together explained 49.0% of the phenotypic variance.

**TABLE 3** Summary of the quantitative trait loci (QTL) identified by genome wide association studies (GWAS) for Hagberg falling number (HFN), protein content (PRT), test weight (TW), and grain yield (YLD)

Trait	QTL	Chr	Peak SNP	cM	bp	$-\log_{10}P$	Marker $R^2$ <sup>a</sup>	Effect <sup>b</sup>	Allele Freq
HFN	<i>HFN_1B.1</i>	1B	Jagger_c5878_119	0	1,254,498	3.10	0.079	28.42	0.58
HFN	<i>HFN_3A.1</i> <sup>c</sup>	3A	RAC875_c99055_69	283.735979	728,322,627	4.08	0.114	-30.65	0.47
HFN	<i>HFN_3B.1</i>	3B	BS00099738_51	227.3745367	759,168,681	3.04	0.077	-31.32	0.73
HFN	<i>HFN_6A.1</i> <sup>c</sup>	6A	Kukri_c29110_360	88.77758706	Un	3.83	0.103	-31.56	0.61
HFN	<i>HFN_6A.2</i> <sup>c</sup>	6A	BS00082104_51	192.8384162	581,841,966	3.69	0.098	32.95	0.38
HFN	<i>HFN_7B.1</i> <sup>c</sup>	7B	RAC875_c525_202	261.6310052	750,082,927	4.78	0.134	-44.19	0.19
PRT	<i>PRT_1A.1</i>	1A	BS00086680_51	71.0976257	281,657,026	3.26	0.087	0.39	0.52
PRT	<i>PRT_1A.2</i>	1A	Excalibur_c13489_867	96.6018344	452,020,371	3.03	0.079	-0.36	0.38
PRT	<i>PRT_1A.3</i>	1A	RAC875_c5882_307	231.6429733	589,054,742	3.01	0.079	-0.33	0.34
PRT	<i>PRT_2A.1</i>	2A	RAC875_rep_c69619_78	207.5522799	734,352,032	3.15	0.083	-0.33	0.56
PRT	<i>PRT_2B.1</i>	2B	BS00046165_51	260.7730538	697,510,384	3.04	0.080	0.44	0.19
PRT	<i>PRT_3A.1</i> <sup>c</sup>	3A	wsnp_Ku_c30545_40369365	107.2210875	363,458,708	3.34	0.090	-0.35	0.52
PRT	<i>PRT_3B.1</i> <sup>c</sup>	3B	wsnp_Ex_c20652_29734133	112.6476302	292,024,034	3.39	0.091	-0.44	0.79
PRT	<i>PRT_3B.2</i>	3B	RAC875_c58159_989	155.9566423	564,248,743	3.10	0.081	-0.46	0.83
PRT	<i>PRT_3B.3</i> <sup>c</sup>	3B	GENE_1618_780	279.0622821	820,894,420	3.02	0.079	-0.33	0.52
PRT	<i>PRT_6B.1</i> <sup>c</sup>	6B	BS00009795_51	12.20934913	4,876,473	3.36	0.090	-0.35	0.44
TW	<i>TW_1B.1</i> <sup>c</sup>	1B	RAC875_rep_c95069_54	81.49097047	336,647,904	3.12	0.080	-1.24	0.21
TW	<i>TW_3B.1</i> <sup>c</sup>	3B	GENE_1771_541	57.06175117	32,458,901	3.60	0.095	1.20	0.36
TW	<i>TW_3B.2</i> <sup>c</sup>	3B	RAC875_c58159_989	155.9566423	571,753,368	3.20	0.083	-1.43	0.86
TW	<i>TW_3B.3</i> <sup>c</sup>	3B	Excalibur_c33274_498	207.7302565	738,752,902	3.21	0.083	2.04	0.05
TW	<i>TW_3B.</i>	3B	BS00073480_51	269.7509427	812,721,677	3.41	0.089	-1.91	0.91
YLD	<i>YLD_1A.1</i> <sup>c</sup>	1A	RAC875_rep_c105092_114	72.6102779	304,040,711	4.03	0.109	0.62	0.50
YLD	<i>YLD_1A.2</i> <sup>c</sup>	1A	wsnp_CAP8_c4785_2322876	166.3032581	544,054,745	3.28	0.085	0.56	0.79
YLD	<i>YLD_2A.1</i> <sup>c</sup>	2A	RAC875_c48625_182	22.63785838	18,636,671	4.30	0.118	0.57	0.53
YLD	<i>YLD_2A.2</i> <sup>c</sup>	2A	RAC875_c16993_839	245.2533379	774,815,015	4.57	0.127	-0.73	0.38
YLD	<i>YLD_2B.1</i>	2B	BS00002660_51	38.40853424	16,025,160	3.09	0.079	0.50	0.57
YLD	<i>YLD_2B.2</i> <sup>c</sup>	2B	BS00091099_51	221.2304675	578,601,402	3.97	0.107	-0.55	0.54
YLD	<i>YLD_2B.3</i> <sup>c</sup>	2B	BS00046165_51	260.7730538	697,510,384	3.45	0.090	-0.70	0.19
YLD	<i>YLD_2B.4</i> <sup>c</sup>	2B	Kukri_c34553_188	319.3174469	766,234,466	5.56	0.161	-1.00	0.17
YLD	<i>YLD_2D.1</i> <sup>c</sup>	2D	wsnp_Ex_c1668_3169623	35.64284331	Un	3.94	0.106	0.59	0.70
YLD	<i>YLD_3A.1</i>	3A	wsnp_Ex_c8884_14841846	179.8202073	625,239,797	3.13	0.08027	-0.51	0.73
YLD	<i>YLD_5B.1</i> <sup>c</sup>	5B	CAP8_rep_c5825_165	3.286161131	15,054,913	3.27	0.0847	0.5	0.4
YLD	<i>YLD_6A.1</i> <sup>c</sup>	6A	BS00082812_51	0.502519793	639,383	4.22	0.11534	0.73	0.78
YLD	<i>YLD_6A.2</i> <sup>c</sup>	6A	wsnp_Ku_c3354_6228393	132.9525631	430,933,984	3.43	0.09203	-0.71	0.24
YLD	<i>YLD_6B.1</i> <sup>c</sup>	6B	Kukri_c21405_2131	15.73078008	166,814	3.8	0.10161	-0.73	0.19

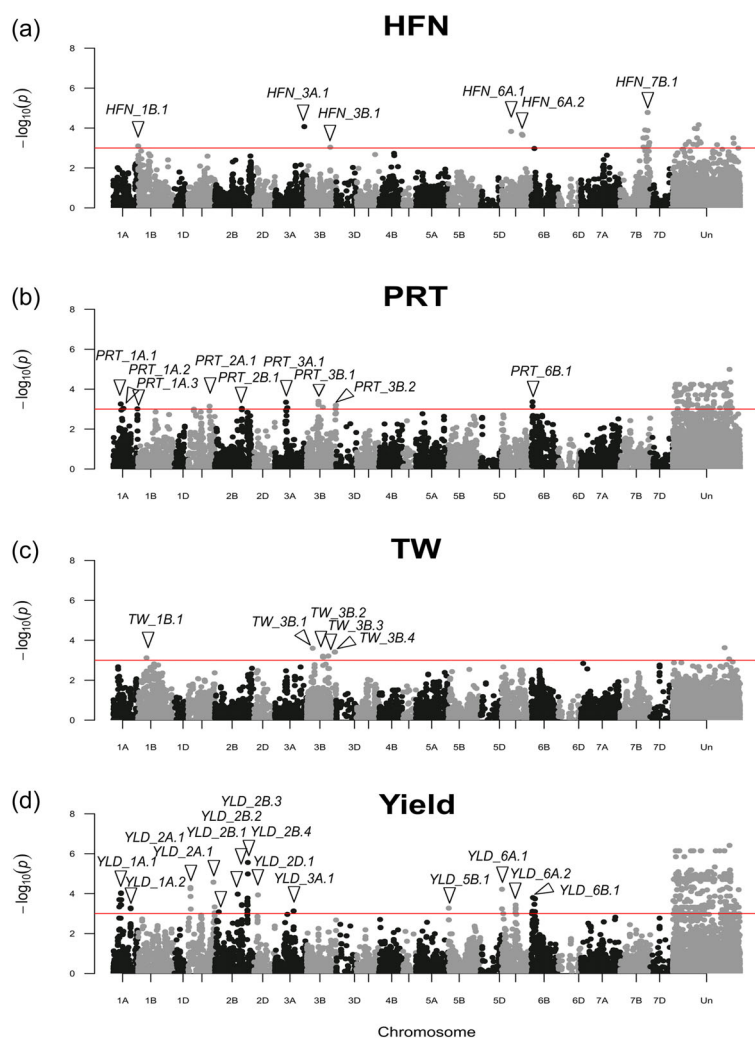
Note. The most significant marker at each QTL is listed. Genetic (Gardner et al., 2016) and physical (IWGSC, 2018) map positions for peak single nucleotide polymorphisms (SNPs) at each QTL are indicated. Where SNPs were anchored to physical map regions currently not allocated to a chromosome designation in the wheat reference genome assembly, the bp position is recorded as unknown (Un). SNP effect at each QTL were determined via two methods (defined below).

<sup>a</sup>SNP effects determined via the outputs of the software TASSELL.

<sup>b</sup>SNP effects determined via modeling the effects considering just those QTL identified for a specific trait.

<sup>c</sup>GWAS hits identified using both kinship and kinship+PCA to correct for population structure. The remaining hits were identified using kinship correction only.





**FIGURE 3** Manhattan plots of genome-wide association studies for Hagberg falling number (HFN; a), protein content (PRT; b), test weight (TW; c), and grain yield (Yield; d). Markers are shown in genetic map order, according to the genetic map published by Wang et al. (2014). Unmapped markers are not shown here but are listed in Supplemental Table S5. The significance threshold ( $-\log_{10}P = 3$ ) is indicated by the horizontal red line

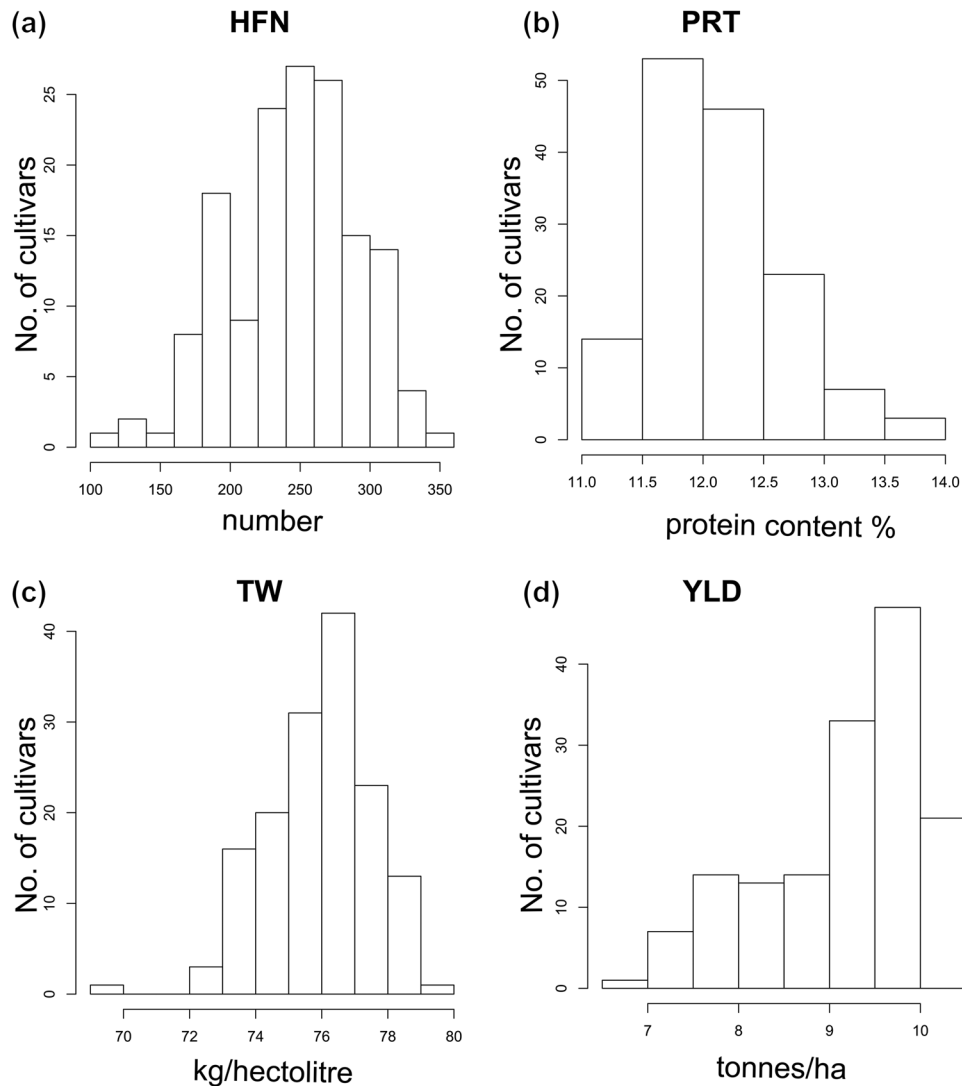
Of these, 24 were genetically mapped and defined six genetic loci on chromosomes 1B (locus *HFN\_1B.1*), 3A (*HFN\_3A.1*), 3B (*HFN\_3B.1*), 6A (*HFN\_6A.1*, *HFN\_6A.2*), and 7B (*HFN\_7B.1*). The GWAS for protein content identified 151 significant marker-trait associations ( $-\log_{10}P$  from 4.99 to 3.01) with the 41 genetically mapped markers, coalescing into nine genetic locations on chromosomes 1A (locus *PRT\_1A.1*, *PRT\_1A.2*, *PRT\_1A.3*), 2A (*PRT\_2A.1*), 2B (*PRT\_2B.1*), 3A (*PRT\_3A.1*), 3B (*PRT\_3B.1*, *PRT\_3B.2*, *PRT\_3B.3*), and 6B (*PRT\_6B.1*). Together, these QTL were modelled to explain 48.9% of the phenotypic variance. The GWAS for test weight identified 39 significant marker-trait associations ( $-\log_{10}P$  from 3.63 to 3.06), defining five genetic loci located on chromosomes 1B (*SW\_1B.1*) and 3B (*SW\_3B.1*, *SW\_3B.2*, *SW\_3B.3*, *SW\_3B.4*), plus two unmapped markers. When modelled together, these QTL explained 26.3% the phenotypic variance. Finally, GWAS for yield identified 533 significant SNPs ( $-\log_{10}P$  from 6.42 to 3.00). Of these, 351 were unmapped with the remaining 182 genetically mapped SNPs identifying 14 genetic loci on chromosomes 1A (*YLD\_1A.1*,

*YLD\_1A.2*), 2A (*YLD\_2A.1*, *YLD\_2A.2*), 2B (*YLD\_2B.1*, *YLD\_2B.2*, *YLD\_2B.3*, *YLD\_2B.4*), 2D (*YLD\_2D.1*), 3A (*YLD\_3A.1*), 5B (*YLD\_5B.1*), 6A (*YLD\_6A.1*, *YLD\_6A.2*), and 6B (*YLD\_6B.1*) which when modelled together explained 68.1% of the phenotypic variance. We also undertook GWAS after adjusting for population structure using kinship + PCA (Supplemental Table S6). Overall, the results remained much the same, although some hits were lost (as indicated in Table 6) and some were gained.

## 4 | DISCUSSION

### 4.1 | Exploiting historical phenotypic data

Recent advances in molecular marker technologies in Triticeae species, such as Diversity Array Technology (DArT) markers, SNP arrays (e.g., Close et al., 2009; Wang et al., 2014), and Kompetitive Allele Specific Polymerase chain reaction (KASP) assays (e.g., Cockram et al., 2012) have



**FIGURE 4** Histogram of Hagberg falling number (HFN; a), protein content (PRT; b), test weight (TW; c), grain yield (YLD; d) in the association mapping panel

meant that phenotyping capacity and capability is now widely recognized as having replaced genotyping as the major bottleneck for genetic studies (Furbank & Tester, 2011). A central feature of this study is that the yield and quality phenotypes investigated were obtained from historical databases. Such data have been accumulated in many countries as a normal part of cultivar testing, either by breeders or testing authorities under the internationally recognized protocols established by the Union for the Protection of plant Varieties (UPOV) and are increasingly used in association mapping studies (e.g., Cockram et al., 2010; Wang et al., 2012) and the development of molecular approaches for use in cultivar testing and registration (Cockram et al., 2009, 2012; Jones, Norris, Smith et al., 2013; Jones, Norris, Cockram, & Lee, 2013; Saccomanno et al., 2020). Historic analyses of trends

in cultivars over time have used regression models to infer year, site, and treatment effects to allow the genetic potential of cultivars to be compared, even when the two cultivars were never in the same field trial (Mackay et al., 2011; Silvey, 1978, 1981, 1986). While year-to-year variability (which includes the joint effect of differences in management, climate, and noise over the years) is accounted for in our models (Supplemental Table S2), here we focus on genetic effects, as used in our GWAS analyses. A key advantage of using historical phenotypic data collected as part of cultivar registration is that the resulting panel consists of elite material. This is of benefit as QTL identified are more likely to be of direct relevance to ongoing breeding programs. Additionally, elite panels are more useful than those consisting of exotic or diverse materials for traits strongly influenced by biotic

or abiotic stresses and differences in maturity (and therefore limiting meaningful phenotypic data collection; Kulwal et al., 2012).

## 4.2 | Characteristics of the association mapping panel used

Association mapping panels consisting of collections of cultivars, landraces, or breeder's lines are likely to have hidden kinship and population structures (White et al., 2008). Thus, kinship and population structure are likely to be strong predictors of many traits (e.g., Cockram et al., 2008). In this study, major genetic stratification was found due to the presence of the wheat/rye chromosome 1B/1R translocation. Parsing marker sets on MAF and  $R^2$  removed much of this substructure, with residual effects of relationships among cultivars removed through incorporation of a relationship matrix in the linear mixed model. The dataset analyzed encompasses all the trait data available to us. While a larger dataset would have been preferable, our power calculations taken in conjunction with the Q-Q plots of  $-\log_{10}P$  and histograms of  $P$ -values demonstrate that despite the relatively small sample size of  $n = 150$ , a substantial proportion of our significant results are genuine. The data we use is 'the best of the best' in the sense that the lines have already been selected by the breeders, leading to potential selection bias in QTL discovery. We do not believe this is a problem for two reasons. First, it should be accounted for by inclusion of the genomic relationship matrix in the GWAS. Second, the heritabilities of our line means are very high, so the opportunity for selection of cultivars to bias the results or QTL segregating among those cultivars is low. Within this study, there is potential for selection bias arising from selection of lines being advanced from assessment Year 1 to Year 2 and so on. This may result in some "regression to the mean" of cultivar estimates. Laidig et al. (2014) discussed this source of bias in relation to estimating genetic and agronomic trends, finding only a slight bias in trend estimation. However, we are not aware of discussion of the consequences of varietal selection bias on the effect of QTL estimation in GWAS. Our expectation is it will be small compared with the effect of selection bias resulting from the imposition of a significance threshold on QTL detection (the 'Beavis effect'; Beavis, 1994), since both increasing and decreasing alleles for QTL of small effect will be approximately randomized across lines, and the bias for QTL of large effect will be lower whatever their distribution across cultivars. However, we are not aware that this potential source of bias, which will affect almost all GWAS studies in crops, has been studied and merits further investigation.

## 4.3 | QTL detection, phenotypic correlations, and comparison with previous studies

While a relatively limited number of studies in investigating the genetics of HFN have been undertaken to date, HFN QTL have been detected on numerous wheat chromosomes (e.g., Börner et al., 2018; Fofana et al., 2009; Gooding et al., 2012; Guo et al., 2020; Kunert et al., 2007; Li et al., 2020; Martinez et al., 2018; Mohler et al., 2014; Tang et al., 2017; Zanetti et al., 2020; Zhang et al., 2014). Based on the interconnection between HFN, PHS, and late-maturity amylase phenotypes, these traits likely share QTL in common (Kulwal et al., 2010). The PHS QTL have been mapped to all 21 wheat chromosomes (reviewed by Kulwal et al., 2010), and the most commonly reported loci are on the Group 2 chromosomes (2B, 2D), the Group 3 chromosomes (3A, 3B, 3D, associated with the wheat *VIVIPAROUS1* gene *TaVP1*, and the red kernel color *R* genes), as well as *Pre-harvest sprouting 1* (*Phs1*, synonym: *Phs-A1*) on the long arm of chromosome 4A (Flintham, 2000; Groos et al., 2002; Jaiswal et al., 2012; Kulwal et al., 2012; Mares et al., 2005; Mori et al., 2005; Munkvold et al., 2009; Shorinola et al., 2016). Similarly, QTL for late-maturity  $\alpha$ -amylase have been found on chromosomes 3B, 6B, and 7B (Emebiri et al., 2010; Mrva & Mares, 2001; 2002). In our study, six genetic loci controlling HFN were identified, the majority of which were located within genetic intervals spanning previously identified QTL related to the HFN phenotype (Supplemental Table S6). The most significant locus we identified was on the long arm of chromosome 7B (located at 725–750 Mbp; peak at 750 Mbp), for which favorable alleles increased HFN by 44 s. The QTL for HFN have previously been identified around this region in at least four biparental populations, with high falling number alleles contributed by the wheat cultivars Rubens (QTL peak at 709 Mbp; Börner et al., 2018), Dream (no SNP sequence data available), W332-84 (744 Mbp), Format (744 Mbp; Mohler et al., 2014), and the winter spelt cultivar Oberkulmer (741 Mbp; Zanetti et al., 2020), resulting in an increase of between 18 and 75 s, depending on the population and test environment. Our 7B locus is in a similar genomic region to a QTL for high-isoelectric point  $\alpha$ -amylase content which is used as a measure of late maturity  $\alpha$ -amylase activity (Mrva & Mares, 2001; Emebiri et al., 2010) (Supplemental Table S6). As HFN and  $\alpha$ -amylase activity are genetically correlated and inversely related, it suggests that the same locus likely underlies the 7B QTL for these traits. Our chromosome 1B HFN locus (located at 1.254 Mbp) was associated with the 1BL/1RS wheat/rye chromosomal translocation, for which analysis of exome capture sequence coverage and nonreference allele frequency relative to the wheat genome assembly for Chinese Spring found to be located on chromosome 1B

between 0 and 220 Mbp (Scott et al., 2021). The inclusion of the 1B/1R translocation as a covariate in the GWAS analysis resulted in the markers at the 1B HFN locus becoming non-significant ( $-\log_{10}P = 2.01$ ), consistent with the HFN locus *HFN\_1B.1* being located on the translocation itself (although this result does not rule out *HFN\_1B.1* being in high linkage disequilibrium with the translocation). The indication that the 1B HFN locus may present on the translocation is in agreement with previous studies (Mohler et al., 2014; Tang et al., 2017) and reflects the strong negative correlation between HFN and the presence of the 1B/1R translocation identified in our association mapping panel (Figure 1). Therefore, in our analysis, the absence of the 1B/1R translocation was predicted to result in an HFN increase of 28 s. Previous analysis of near isogenic lines has shown that gibberellic insensitive semidwarf lines that lack the 1B/1R translocation require a cool temperature shock during grain filling for the expression of late maturity  $\alpha$ -amylase activity while the presence of the translocation results in constitutive expression (Mrva et al., 2008). In both classes of response, late maturity  $\alpha$ -amylase activity is due to the synthesis of high-isoelectric point  $\alpha$ -amylase encoded by genes at the  *$\alpha$ -Amylase1* ( *$\alpha$ -Amy1*) locus at 554–564 Mbp on the long arm of chromosome 6A, at which at least three genes are present (RefSeq v1.1 gene models *TraesCS6A02G319300*, *TraesCS6A02G334100*, *TraesCS6A02G334200*; Ju et al., 2019). Interestingly, we also identified a QTL for HFN on chromosome 6A at 581 Mbp (second only to the 7B locus in its effect on HFN), indicating this GWAS hit could be due to allelic variation at the *Amy1* locus. Indeed, a QTL at a similar location on chromosome 6A has been reported in German winter wheat (Mohler et al., 2014; Supplemental Table S6). Lastly, while our HFN loci on chromosomes 3A (728 Mbp) and 3B (759 Mbp) were identified at approximately homoeologous locations on the long arms of the Group 3 chromosomes, they were not located close to *VP-1A* (660 Mbp) or *VP-1B* (693 Mbp). Neither did our 3B QTL appear to correspond to the QTL *QFn.crc-3B* previously reported to be located between markers *barc77* and *wmc307* (3B: 78–430 Mbp; Fofana et al., 2009). However, they were located relatively close to the wheat *MYB* gene *TaMYB10-A1* located on chromosome 3A at 704 Mbp, and its homologue *TaMYB10-B1* on 3B at 571 Mbp. *TaMYB-A1* has previously been identified as a candidate gene for both PHS and grain color via GWAS in U.S. winter wheat (Lin et al., 2016). Similarly, a study of European winter wheat has recently identified genetic loci for PHS close to both *TaMYB10-A1* and *TaMYB-B1* (Scott et al., 2021). While the correlation between grain color and PHS has been often reported, how the *R* genes controlling grain color might pleiotropically influence PHS remains unknown (Lin et al., 2016) and could be due to linkage. While two genes in the HFN/PHS/late-maturity amylase phenotypic trait complex have been map-based cloned, *Phs1* on chromosome

4A (Torada et al., 2016) and *TaPHS1* on chromosome 3A (Liu et al., 2013), we did not find GWAS hits at these locations indicating that natural genetic variation at these genes does not play a major role in controlling HFN in U.K. wheat.

Protein content QTL have previously been discovered on all wheat chromosomes (reviewed by Kumar et al., 2018), with major QTL encoding for clusters of for high molecular weight glutenin genes located at the *Glu-A1*, *-B1*, and *-D1* loci on the long arms of the Group 1 chromosomes (e.g., Zhang et al., 2011). Although we identified three genetic loci for protein content were identified on chromosome 1A, none of them are predicted to be located near the *Glu-A1* locus (based on gene model *TraesCS1A02G317311* at 509 Mbp). The *GRAIN PROTEIN CONTENT-B1* (*GPC-B1*) locus on the short arm of chromosome 6B controlling protein content in tetraploid wheat has been shown to be encoded by a NAC transcription factor (*NAM-B1*; Uauy et al., 2006). While we found a locus on chromosome 6B controlling protein content, its position at 5 Mbp placed it far from *NAM-B1* (gene model *TraesCS6B02G207500LC* at 134 Mbp), supporting reports that most hexaploid bread wheat lines do not possess the wild-type functional form of the gene (Mellers et al., 2020; Uauy et al., 2006). Our finding that grain protein content and grain yield are strongly negatively correlated with each other is in agreement with many previous studies (e.g., Groos et al., 2003; Scott et al., 2021). Indeed, of the 10 genetic loci we identified for protein content, three were also found to putatively co-locate with hits for yield: *PRT\_1A.1/YLD\_1A.1* on chromosome 1A at 4–30 Mbp, *PRT\_2B.1/YLD\_2B.3* on chromosome 2B at 698 Mbp, and *PRT\_6B.1/YLD\_6B.1* on chromosome 6B at 1–4 Mbp. Interestingly, the *PRT\_1A.1/YLD\_1A.1* region has previously been identified as being located within a putative introgression (Sharma et al., 2021). For the *PRT\_2B.1/YLD\_2B.3* QTL pair controlling grain protein content and grain yield, previous studies have identified a multitrait QTL on chromosome 2B just 4 Mbp away at 694 Mbp (*Qmt.tamu.2B.1.1*, for grain yield, spikes per m<sup>2</sup>, 1,000-grain weight, and test weight. Peak marker: *RFL\_Contig3172\_1752*; Assanga et al., 2017; Supplemental Table S6). Additionally, two other protein content QTL, both on chromosome 3B, co-located with GWAS hits for test weight. The first was QTL pair *PRT\_3B.2* (at 564 Mbp)/ *TW\_3B.2* (at 567–574 Mbp). A relatively large effect QTL for test weight has previously been reported that spans this interval on chromosome 3B (Cabral et al., 2018; Supplemental Table S6). The second QTL pair was *PRT\_3B.3* (819–826 Mbp)/ *TW\_3B.4* (812–816 Mbp), located further towards the distal end of the long arm of chromosome 3B. Interestingly, beneficial alleles at the first of these pairs of loci increased both protein content and test weight. While no grain yield QTL were identified at this location, it is possible that increased test weight comes at a cost of yield, as fewer but

larger grain are formed with a higher protein content. However, our analysis of phenotypic correlations found test weight to be weakly positively correlated with protein content, but to have a stronger positive correlation with grain yield. Further studies are needed to determine whether favorable alleles at the *PRT\_3B.2/SW\_3B.2* locus can improve both protein and test weight without affecting yield. The remaining five grain protein GWAS hits were located on chromosomes 1A, 2A, 3A, and 3B. Previously identified grain protein content QTL have been identified on all of these chromosomes, with for example a robust QTL identified on the long arm of chromosome 1A in a biparental population of tetraploid wheat (Blanco et al., 2012) estimated to be in a similar location to our GWAS hit *PRT\_1A.3* (peak at 598 Mbp) based on anchoring the simple sequence repeat marker *wmc254* to the wheat physical map at 569 Mbp. However, the large genetic and physical intervals of the majority of these hinders further meaningful comparison. As discussed above, two of the five GWAS hits for test weight co-located with hits for grain protein content. Of the three remaining test weight QTL, *SW\_1B.1* does not appear to be present on the wheat/rye 1B/1R translocation, based on the interval determined by Scott et al. (2021). The remaining two GWAS hits, *SW\_3B.1* and *SW\_3B.3*, are both located on chromosome 3B, with the latter having the largest effect on test weight of all the genetic loci identified in our panel. Notably, we did not identify GWAS hits at the *Glu-D1* locus encoding high molecular weight glutenin genes. Cross-referencing the cultivars in our panel with those previously genotyped for *Glu-D1* (data sourced from CerealsDB, <https://www.cerealsdb.uk.net/cerealgenomics/>) found that of the 20 cultivars that overlapped, the high '5+10' *Glu-D1* allele was present in a quarter of these lines (data not shown). The wheat D subgenome has low genetic diversity, and this is reflected in the low D subgenome SNP coverage of the 90k array (Wang et al., 2014; Mellers et al., 2020). Accordingly, as allelic variation at the *Glu-D1* locus appears likely to be present at sufficient minor allele frequency to be detected via GWAS, we assume that the *Glu-D1* locus was not identified in our GWAS for protein content due to lack of SNPs in sufficiently close linkage disequilibrium.

Of the four traits investigated here, the largest number of significant genetic loci were identified for grain yield. Many studies have reported QTL have been identified for yield and yield components (e.g., Gegas et al., 2010; Corsi et al., 2021; Liu et al., 2020; Suliman et al., 2021), and to date just two yield component genes have been identified via map-based approaches: *Grain Number Increase 1 (GNI-A1)*; Sakuma et al., 2019) and *WHEAT ORTHOLOG OF APO1 (WAP01)*; Kuzay et al., 2019; Muqaddasi et al., 2019). While we did not identify GWAS hits for yield at either of these genes, *YLD\_6A.2* (defined within a 3-cM interval that corresponds to a large physical interval spanning the centromere on chromosome 6A of around 440 Mbp from 101 to 438 Mbp), spans

the wheat gene *GRAIN WIDTH 2*, termed *TaGW2-A*. Located at 238 Mbp, *TaGW2-A* and has previously been associated with the control of grain size characters via both forward (e.g., Zhai et al., 2018; Corsi et al., 2021) and reverse genetic studies (Simmonds et al., 2016; Wang et al., 2018; Zhang et al., 2018). The most significant yield QTL identified in our panel, *YLD\_2B.4* ( $-\log_{10}P = 5.56$ ), was located towards the end of the long arm of chromosome 2B at 766 Mbp. To our knowledge, no yield QTL has been reported in this location, and no wheat orthologues of yield or yield component map-based cloned genes from the related cereal species rice are located nearby. Recently, Fradgley et al. (2019) used gene-dropping simulations on a panel of genotyped northwestern European wheat cultivars that overlaps with the panel we investigate here, finding evidence of breeder selection at 15 genetic loci across the wheat genome. Interestingly, a third of these loci either co-locate or are  $\leq 16$  Mbp of GWAS hits for yield we identify here, indicating they have been under strong breeder selection: *YLD\_1A.2* (GWAS location = chromosome 1A at 544 Mbp, gene-dropping location = 1A at 557 Mbp based on SNP *BS00032825\_51*), *YLD\_2A.1* (GWAS = 2A at 12-24 Mbp, gene-dropping = 16 Mbp based on SNP *Excalibur\_rep\_c110303\_320*), *YLD\_2A.2* (GWAS = 2A at 763-774 Mbp, gene-dropping = 751 Mbp based on SNP *BS00023202\_5*), *YLD\_2B.1* (GWAS = 2B at 16 Mbp, gene-dropping = 25 Mbp based on SNP *BS00064706\_51*), and *YLD\_3A.1* (GWAS = 3A at 625-627 Mbp, gene-dropping = 643 Mbp based on SNP *BS00038663\_51*). Further investigation and independent validation would be required to further study the yield loci identified.

## 5 | CONCLUSIONS

Evaluation of seed quality traits from field grown samples is expensive and time consuming. Therefore, the identification of molecular markers linked to these traits in elite wheat cultivars provides valuable information for further research and ultimately for the development of appropriate genetic markers for marker-assisted breeding. A key limitation of our study is the relatively small size of the association mapping panel ( $n = 150$ ). However, our finding that (a) all four traits had relatively high heritability ( $h^2 \geq .86$ ); (b) power analysis showed we had reasonable power to detect most of the underlying major QTL and at least a proportion of the smaller-effect QTL; (c) we detected at least one previously described genetic locus per trait; and (d) co-localizing GWAS hits between related traits agreed with a priori assumptions about their genetic control (e.g., the *PRT\_2B.1/YLD\_2B.3* QTL pair for protein and yield) provides support for the validity of the results and approaches used. Finally, the demonstration of the utility of historic datasets collected during statutory varietal testing paves the way for similar analyses of for wheat and

other crops across the many Union for the Protection of plant Varieties signatory countries.

## ACKNOWLEDGMENTS

JW's PhD was funded by The NIAB Trust, Crop Evaluation Limited, and The Home Grown Cereals Authority (HGCA) R&D Committee. IM and JC's time was supported by Biotechnology and Biological Sciences Research Council (BBSRC) grants BB/J002542/1 and BB/M011666/1, respectively.

## AUTHOR CONTRIBUTIONS

Jon White: Data curation; Formal analysis; Investigation; Methodology; Writing – original draft. Rajiv Sharma: Formal analysis; Investigation; Methodology; Writing – review & editing. David Balding: Conceptualization; Funding acquisition; Supervision; Writing – review & editing. James Cockram: Formal analysis; Resources; Writing – original draft; Writing – review & editing. Ian J. Mackay: Conceptualization, Funding acquisition, Project administration, Supervision, Writing-review & editing.

## CONFLICT OF INTEREST

The authors report no conflicts of interest.

## ORCID

James Cockram  <https://orcid.org/0000-0002-1014-6463>

## REFERENCES

- Adamski, N., Borrill, P., Brinton, J., Harrington, S. A., Marchal, C., Bentley, A. R., Bovill, W. D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., Ford, B., Ghosh, S., Harwood, W., Hassani-Pak, K., Hayta, S., Hickey, L. T., Kanyuka, K., King, J., Maccaferri, M., ... Uauy, C. (2020). A roadmap for gene functional validation in large genome crops: Lessons from polyploid wheat. *eLIFE*, 9, e55646. <https://doi.org/10.7554/eLife.55646>
- Assanga, S. O., Fuentealba, M., Zhang, G., Tan, C., Dhakal, S., Rudd, J. C., Ibrahim, A. M. H., Xue, Q., Haley, S., Chen, J., Chao, S., Baker, J., Jessup, K., & Liu, S. (2017). Mapping of quantitative trait loci for grain yield and its components in a US popular winter wheat TAM 111 using 90K SNPs. *PLOS ONE*, 12, e0189669. <https://doi.org/10.1371/journal.pone.0189669>
- Beavis, W. D. (1994). The power and deceit of QTL experiments: lessons from comparative QTL studies. In *Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference* (pp. 250–266). American Seed Trade Association.
- Bentley, A. R., Scutari, M., Gosman, N., Faure, S., Bedford, F., Howell, P., Cockram, J., Rose, G. A., Barber, T., Irigoyen, J., Horsnell, R., Pumfrey, C., Winnie, E., Schacht, J., Beauchêne, K., Praud, S., Greenland, A., Balding, D., & Mackay, I. J. (2014). Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theoretical and Applied Genetics*, 127, 2619–2633. <https://doi.org/10.1007/s00122-014-2403-y>
- Blanco, A., Mangini, G., Giancaspro, A., Giove, S., Colasuonno, P., Simeone, R., Signorile, A., De Vita, P., Mastrangelo, A. M., Cattivelli, L., & Gadaleta, A. (2012). Relationships between grain protein content and grain yield components through quantitative trait locus analyses in a recombinant inbred line population derived from two elite durum wheat cultivars. *Molecular Breeding*, 30, 79–92. <https://doi.org/10.1007/s11032-011-9600-z>
- Bodmer, W. F. (1986). Human genetics: The molecular challenge. *Bioessays*, 7, 41–45. <https://doi.org/10.1002/bies.950070109>
- Börner, A., Nagel, M., Agacka-Moldoch, M., Gierke, P. U., Oberforster, M., Albrecht, T., & Mohler, V. (2018). QTL analysis of falling number and seed longevity in wheat (*Triticum aestivum* L.). *Journal of Applied Genetics*, 59, 35–42. <https://doi.org/10.1007/s13353-017-0422-5>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Cabral, A., Jordan, M. C., Larson, G., Somers, D. J., Humphreys, G., & McCartney, C. A. (2018). Relationship between QTL for grain shape, test weight, milling yield, and plant height in the spring wheat cross RL4452/AC Domain'. *PLOS ONE*, 13, e0190681. <https://doi.org/10.1371/journal.pone.0190681>
- Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). From mutations to MAGIC: Resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, 11, 215–221.
- Close, T. J., Bhat, P. R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., Druka, A., Stein, N., Svensson, J. T., Wanamaker, S., Bozdag, S., Roose, M. L., Moscou, M. J., Chao, S., Varshney, R. K., Szűcs, P., Sato, K., Hayes, P. M., Matthews, D. E., ... Waugh, R. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics*, 10, 582. <https://doi.org/10.1186/1471-2164-10-582>
- Cockram, J., White, J., Leigh, F. J., Lea, V. J., Chiapparino, E., Laurie, D. A., Mackay, I. J., Powell, W., & O'Sullivan, D. M. (2008). Association mapping of partitioning loci in barley (*Hordeum vulgare* ssp. *vulgare* L.). *BMC Genetics*, 9, 16. <https://doi.org/10.1186/1471-2156-9-16>
- Cockram, J., Norris, C., & O'Sullivan, D. M. (2009). PCR markers diagnostic for spring and winter seasonal growth habit in barley. *Crop Science*, 49, 403–410. <https://doi.org/10.2135/cropsci2008.07.0398>
- Cockram, J., White, J., Zuluaga, D. L., Smith, D., Comadran, J., Macaulay, M., Luo, Z., Kearsey, M. J., Werner, P., Harrap, D., Tapsell, C., Liu, H., Hedley, P. E., Stein, N., Schulte, D., Steuernagel, B., Marshall, D. F., Thomas, W. T. B., Ramsay, L., ... O'Sullivan, D. M. (2010). Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *Proceedings of the National Academy of Sciences U S A*, 107, 21611–21616. <https://doi.org/10.1073/pnas.1010179107>
- Cockram, J., Jones, H., Norris, C., & O'Sullivan, D. M. (2012). Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare* ssp. *vulgare* L.). *Theoretical and Applied Genetics*, 125, 1735–1749. <https://doi.org/10.1007/s00122-012-1950-3>
- Cockram, J., & Mackay, I. (2018). Genetic mapping populations for conducting high-resolution trait mapping in plants. In R. Varshney, M. Pandey, & A. Chitkineni (Eds.), *Plant genetics and molecular biology*. Advances in Biochemical Engineering/Biotechnology (Vol 164, pp. 109–138). Springer.
- Corsi, B., Percival-Alwyn, L., Downie, R. C., Venturini, L., Iagallo, E. M., Mantello, C. C., McCormick-Barnes, C., See, P. T., Oliver, R. P.,

- Mofat, C. S., & Cockram, J. (2020). Genetic analysis of wheat sensitivity to the ToxB fungal effector from *Pyrenophora tritici-repentis*, the causal agent of tan spot. *Theoretical and Applied Genetics*, *133*, 935–950. <https://doi.org/10.1007/s00122-019-03517-8>
- Corsi, B., Obinu, L., Zanella, C. M., Cutrupi, S., Day, R., Geyer, M., Lillemo, M., Lin, M., Mazza, L., Percival-Alwyn, L., Stadlmeier, M., Mohler, V., Hartl, L., & Cockram, J. (2021). Identification of eight QTL controlling multiple yield components in a German multiparental wheat population, including *Rht24*, *WAP0-A1*, *WAP0-B1* and genetic loci on chromosomes 5A and 6A. *Theoretical and Applied Genetics*, *134*, 1435–1454. <https://doi.org/10.1007/s00122-021-03781-7>
- Emebiri, L. C., Oliver, J. R., Mrva, K., & Mares, D. (2010). Association mapping of late maturity  $\alpha$ -amylase (LMA) activity in a collection of synthetic hexaploid wheat. *Molecular Breeding*, *26*, 39–49. <https://doi.org/10.1007/s11032-009-9375-7>
- FAO. (2016). *Food outlook*. <http://www.fao.org/giews/reports/food-outlook/en/>
- Flintham, J. E. (2000). Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. *Seed Science Research*, *10*, 43–50. <https://doi.org/10.1017/S0960258500000052>
- Fofana, B., Humphreys, D. G., Rasul, G., Cloutier, S., Brûlé-Babel, A., Woods, S., Lukow, O. M., & Somers, D. J. (2009). Mapping quantitative trait loci controlling pre-harvest sprouting resistance in a red x white seeded spring wheat cross. *Euphytica*, *165*, 509–521. <https://doi.org/10.1007/s10681-008-9766-6>
- Fradgley, N., Gardner, K. A., Cockram, J., Elderfield, J., Hickey, J. M., Howell, P., Jackson, R., & Mackay, I. (2019). A large-scale pedigree resource of wheat reveals evidence for adaptation and selection by breeders. *PLOS Biology*, *17*, e3000071. <https://doi.org/10.1371/journal.pbio.3000071>
- Furbank, R. T., & Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Sciences*, *16*, 635–644. <https://doi.org/10.1016/j.tplants.2011.09.005>
- Gale, M. D., & Lenton, J. R. (1987). Preharvest sprouting in wheat: A complex genetic and physiological problem affecting bread making quality in UK wheat. *Aspects of Applied Biology*, *15*, 115–124.
- Gardner, K. A., Wittern, L. M., & Mackay, I. J. (2016). A highly recombined, high-density, eight-founder wheat MAGIC map reveals extensive segregation distortion and genomic locations of introgression segments. *Plant Biotechnology Journal*, *14*, 1406–1417. <https://doi.org/10.1111/pbi.12504>
- Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., Sayers, L., Doonan, J. H., & Snape, J. W. (2010). A genetic framework for grain size and shape variation in wheat. *Plant Cell*, *22*, 1046–1056. <https://doi.org/10.1105/tpc.110.074153>
- Gooding, M. J., Uppal, R. K., Addisu, M., Uauy, C., Simmonds, J. R., & Murdoch, A. J. (2012). Reduced height alleles (*Rht*) and Hagberg falling number of wheat. *Journal of Cereal Science*, *55*, 305–311. <https://doi.org/10.1016/j.jcs.2012.01.003>
- Groos, C., Gay, G., Perretant, M. R., Gervais, L., Bernard, M., Dedryver, F., & Charmet, G. (2002). Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white-red grain bread wheat cross. *Theoretical and Applied Genetics*, *104*, 39–47. <https://doi.org/10.1007/s001220200004>
- Groos, C., Robert, N., Bervas, E., & Charmet, G. (2003). Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theoretical and Applied Genetics*, *106*, 1032–1040. <https://doi.org/10.1007/s00122-002-1111-1>
- Guo, Y., Zhang, G., Guo, B., Qu, C., Zhang, M., Kong, F., Zhao, Y., & Li, S. (2020). QTL mapping for quality traits using a high-density genetic map of wheat. *PLOS ONE*, *15*, e0230601. <https://doi.org/10.1371/journal.pone.0230601>
- Hagberg, S. (1960). A rapid method for determining alpha-amylase activity. *Cereal Chemistry*, *37*, 218–222.
- The International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C. J., Stein, N., Choulet, F., Distelfeld, A., Eversole, K., Poland, J., Rogers, J., Ronen, G., Sharpe, A. G., Pozniak, C., Ronen, G., Stein, N., ... Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*, eaar7191.
- Jaiswal, V., Mir, R. R., Mohan, A., Balyan, H. S., & Gupta, P. K. (2012). Association mapping for pre-harvest sprouting tolerance in common wheat (*Triticum aestivum* L.). *Euphytica*, *188*, 89–102. <https://doi.org/10.1007/s10681-012-0713-1>
- Jamali, S. H., Cockram, J., & Hickey, L. T. (2019). Insights into deployment of DNA markers in plant variety protection and registration. *Theoretical and Applied Genetics*, *132*, 1911–1929. <https://doi.org/10.1007/s00122-019-03348-7>
- Jones, H., Norris, C., Smith, D., Cockram, J., Lee, D., O'Sullivan, D. M., & Mackay, I. (2013). Evaluation of the use of high-density SNP genotyping to implement UPOV Model 2 for DUS testing in barley. *Theoretical and Applied Genetics*, *126*, 901–911. <https://doi.org/10.1007/s00122-012-2024-2>
- Jones, H., Norris, C., Cockram, J., & Lee, D. (2013). Variety protection and plant breeders' rights in the 'DNA era'. In T. Lübberstedt & R. K. Varshney (Eds.), *Diagnostics in Plant Breeding* (pp. pp369–402) Springer.
- Jouanin, A., Gilissen, L. J. W. J., Boyd, L. A., Cockram, J., Leigh, F. J., Wallington, E. J., van den Broeck, H. C., van der Meer, I. M., Schaart, J. G., Visser, R. G., & Smulders, M. J. M. (2018). Food processing and breeding strategies for coeliac-safe and healthy wheat products. *Food Research International*, *110*, 11–21. <https://doi.org/10.1016/j.foodres.2017.04.025>
- Ju, L., Deng, G., Liang, J., Zhang, H., Li, Q., Pan, Z., Yu, M., & Long, H. (2019). Structural organization and functional divergence of high isoelectric point  $\alpha$ -amylase genes in bread wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.). *BMC Genetics*, *20*, 25. <https://doi.org/10.1186/s12863-019-0732-1>
- Kulwal, P. L., Mir, R. R., Kumar, S., & Gupta, P. K. (2010). QTL analysis and molecular breeding for seed dormancy and pre-harvest sprouting tolerance in bread wheat. *Journal of Plant Biology*, *37*, 59–74.
- Kulwal, P., Ishikawa, G., Bensch, D., Feng, Z., Yu, L.-X., Jadhav, A., Mehete, S., & Sorrells, M. E. (2012). Association mapping for pre-harvest sprouting resistance in white winter wheat. *Theoretical and Applied Genetics*, *125*, 793–805. <https://doi.org/10.1007/s00122-012-1872-0>
- Kumar, A., Jain, S., Elias, E. M., Ibrahim, M., & Sharma, L. K. (2018). An overview of QTL identification and marker-assisted selection for grain protein content in wheat. In R. S. Sengar & A. Singh (Eds.), *Eco-friendly agro-biological techniques for enhancing crop productivity* (pp. 245–274). Springer.
- Kunert, A., Naz, A. A., Dedek, O., Pillen, K., & Léon, J. (2007). AB-QTL analysis in winter wheat: I. Synthetic hexaploid wheat (*T. turgidum* ssp. *dicoccoides* X *T. tauschii*) as a source of favourable alleles for milling and baking quality traits. *Theoretical and Applied Genetics*, *115*, 683–695. <https://doi.org/10.1007/s00122-007-0600-7>

- Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., Fraser, M., Anderson, J. A., Brown-Guedira, G., DeWitt, N., Haugrud, A. P., Faris, J. D., Akhunov, E., Bai, G., & Dubcovsky, J. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theoretical and Applied Genetics*, *132*, 2689–2705. <https://doi.org/10.1007/s00122-019-03382-5>
- Laidig, F., Piepho, H. P., Drobek, T., & Meyer, U. (2014). Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics*, *127*, 2599–2617. <https://doi.org/10.1007/s00122-014-2402-z>
- Li, Q., Pan, Z., Gao, Y., Li, T., Liang, J., Zhang, Z., Zhang, H., Deng, G., Long, H., & Yu, M. (2020). Quantitative trait locus (QTL) mapping for quality traits of wheat based on high density genetic map combined with bulked segregant analysis RNA-seq (BSR-Seq) indicates that the *Basic 7S Globulin* gene is related to falling number. *Frontiers in Plant Sciences*. <https://doi.org/10.3389/fpls.2020.600788>
- Lin, M., Zhang, D., Liu, S., Zhang, G., Yu, J., Ak, F., & Bai, G. (2016). Genome-wide association analysis on pre-harvest sprouting resistance and grain color in U.S winter wheat. *BMC Genomics*, *17*, 794. <https://doi.org/10.1186/s12864-016-3148-6>
- Liu, H., Mullan, D., Zhang, C., Zhao, S., Li, X., Zhang, A., Lu, Z., Wang, Y., & Yan, G. (2020). Major genomic regions responsible for wheat yield and its components as revealed by meta-QTL and genotype-phenotype association analyses. *Planta*, *252*, 65. <https://doi.org/10.1007/s00425-020-03466-3>
- Liu, S., Seghal, S. K., Meng, J. L., Lin, M., Trick, H. N., Yu, J., Gill, B. S., & Bai, G. (2013). Cloning and characterization of a critical regulator for preharvest sprouting in wheat. *Genetics*, *195*, 263–273. <https://doi.org/10.1534/genetics.113.152330>
- Mackay, I. J., Cockram, J., Howell, P., & Powell, W. (2021). Understanding the classics: The unifying concepts of transgressive segregation, inbreeding depression and heterosis and their central relevance for crop breeding. *Plant Biotech J*, *19*, 26–34. <https://doi.org/10.1111/pbi.13481>
- Mackay, I., Horwell, A., Garner, J., White, J., McKee, J., & Philpott, H. (2011). Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. *Theoretical and Applied Genetics*, *122*, 225–238. <https://doi.org/10.1007/s00122-010-1438-y>
- Mares, D., & Mrva, K. (2008). Late-maturity  $\alpha$ -amylase: Low falling number in wheat in the absence of preharvest sprouting. *Journal of Cereal Science*, *47*, 6–17. <https://doi.org/10.1016/j.jcs.2007.01.005>
- Mares, D., Mrva, K., Cheong, J., Williams, K., Watson, B., Storlie, E., Sutherland, M., & Zou, Y. (2005). A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. *Theoretical and Applied Genetics*, *111*, 1357–1364. <https://doi.org/10.1007/s00122-005-0065-5>
- Martinez, S. A., Godoy, J., Huang, M., Zhang, Z., Carter, A. H., Campbell, K. A. G., & Steber, C. M. (2018). Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. *Frontiers in Plant Science*, *9*, 141. <https://doi.org/10.3389/fpls.2018.00141>
- Mellers, G., Aguilera, J. G., Bird, N., Caierão, E., Consoli, L., Santana, F. M., Schreen, P. L., Uauy, C., Nicholson, P., & Cockram, J. (2020). Genetic characterization of a core collection of wheat germplasm relevant to Brazilian breeding using a high-density SNP array. *G3: Genes Genomes Genetics*, *10*, 2229–2239. <https://doi.org/10.1534/g3.120.401234>
- Mohler, V., Albrecht, T., Mrva, K., Schweizer, G., & Hartl, L. (2014). Genetic analysis of falling number in three bi-parental common winter wheat populations. *Plant Breeding*, *133*, 448–453. <https://doi.org/10.1111/pbr.12177>
- Mori, M., Uchino, N., Chono, M., Kato, K., & Miura, H. (2005). Mapping QTLs for grain dormancy on wheat chromosome 3A and the group 4 chromosomes, and their combined effect. *Theoretical and Applied Genetics*, *110*, 1315–1323. <https://doi.org/10.1007/s00122-005-1972-1>
- Mrva, K., & Mares, D. (2001). Quantitative trait locus analysis of late maturity  $\alpha$ -amylase in wheat using the doubled haploid population Cranbrook Halberd. *Australian Journal of Agricultural Research*, *52*, 1267–1273. <https://doi.org/10.1071/AR01047>
- Mrva, K., & Mares, D. (2002). Screening methods and identification of QTLs associated with late maturity  $\alpha$ -amylase in wheat. *Euphytica*, *126*, 55–59. <https://doi.org/10.1023/A:1019667521448>
- Mrva, K., Mares, D., & Cheong, J. (2008). Genetic mechanisms involved in late maturity  $\alpha$ -amylase in wheat. In R. Appels, R. Eastwood, E. Lagudah, P. Langridge, M. MacKay, L. McIntyre, & P. Sharp (Eds.), *Proceedings of the 11th International Wheat Genetics Symposium* (pp. 940–942). Sydney University Press.
- Muqaddasi, Q. H., Brassac, J., Koppolu, R., Plieske, J., Ganal, M. W., & Röder, M. S. (2019). *TaAPO-A1*, an ortholog of rice *ABERRANT PANICLE ORGANIZATION 1*, is associated with total spikelet number per spike in elite European hexaploid winter wheat (*Triticum aestivum* L.) varieties. *Scientific Reports*, *9*, 13853. <https://doi.org/10.1038/s41598-019-50331-9>
- Munkvold, J. D., Tanaka, J., Benscher, D., & Sorrells, M. E. (2009). Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. *Theoretical and Applied Genetics*, *119*, 1223–1235. <https://doi.org/10.1007/s00122-009-1123-1>
- UK Flour Millers. (2017). Nabim wheat guide. UK Flour Millers. <http://www.nabim.org.uk/>
- Narasimhamoorthy, B., Gill, B. S., Fritz, A. K., Nelson, J. C., & Brown-Guedira, G. L. (2006). Advanced backcross QTL analysis of a hard winter wheat  $\times$  synthetic wheat population. *Theoretical and Applied Genetics*, *112*, 787–796. <https://doi.org/10.1007/s00122-005-0159-0>
- Rabinovich, S. V. (1998). Importance of wheat-rye translocations for breeding modern cultivars of *Triticum aestivum* L. *Euphytica*, *100*, 323–340. <https://doi.org/10.1023/A:1018361819215>
- Reif, J. C., Gowda, M., Maurer, H. P., Longin, C. F. H., Korzun, V., Ebmeyer, E., Bothe, R., Pietsch, C., & Würschum, T. (2011). Association mapping for quality traits in soft winter wheat. *Theoretical and Applied Genetics*, *122*, 961–970. <https://doi.org/10.1007/s00122-010-1502-7>
- Saccomanno, B., Wallace, M., O'Sullivan, D. M., & Cockram, J. (2020). Use of genetic markers for the detection of off-types for DUS phenotypic traits in the inbreeding crop, barley. *Molecular Breeding*, *40*, 13. <https://doi.org/10.1007/s11032-019-1088-y>
- Sakuma, S., Golan, G., Guo, Z., Ogawa, T., Tagiri, A., Sugimoto, K., Bernhardt, N., Brassac, J., Mascher, M., Hensel, G., Ohnishi, S., Jinno, H., Yamashita, Y., Ayalon, I., Peleg, Z., Schnurbusch, T., & Komatsuda, T. (2019). Unleashing floret fertility in wheat through the mutation of a homeobox gene. *Proceedings of the National Academy*



- of *Sciences U S A*, 116, 5182–5187. <https://doi.org/10.1073/pnas.1815465116>
- Scott, M. F., Fradgley, N., Bentley, A. R., Brabbs, T., Corke, F., Gardner, K. A., Horsnell, R., Howell, P., Ladejobi, O., Mackay, I. J., Mott, R., & Cockram, J. (2021). Limited haplotype diversity underlies polygenic trait architecture across 70 years of wheat breeding. *Genome Biology*, 22, 137. <https://doi.org/10.1186/s13059-021-02354-7>
- Shorinola, O., Bird, N., Simmonds, J., Berry, S., Henriksson, T., Jack, P., Werner, P., Gerjets, T., Scholefield, D., Balcárkova, B., Valárik, M., Holdsworth, M. J., Flintham, J., & Uauy, C. (2016). The wheat *Phs-A1* pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the *PM19* genes in UK germplasm. *Journal of Experimental Botany*, 67, 4169–4178. <https://doi.org/10.1093/jxb/erw194>
- Sharma, R., Cockram, J., Gardner, K. A., Russell, J., Ramsay, L., Thomas, W. T. B., O'Sullivan, D. M., Powell, W., & Mackay, I. J. (2021). Trends of genetic changes uncovered by Env- and EigenGWAS in wheat and barley. *Theoretical and Applied Genetics*. <https://doi.org/10.1007/s00122-021-03991-z>
- Silvey, V. (1978). Contribution of new varieties to increasing cereal yield in England and Wales. *Journal of the National Institute of Agricultural Botany*, 14, 367–384.
- Silvey, V. (1981). The contribution of new wheat, barley and oat varieties to increasing yield in England and Wales. *Journal of the National Institute of Agricultural Botany*, 15, 399–412.
- Silvey, V. (1986). Contribution of new varieties to cereal yields in England and Wales between 1947 and 1983. *Journal of the National Institute of Agricultural Botany*, 17, 155–168.
- Simmonds, J., Scott, P., Brinton, J., Mestre, T. C., Bush, M., Blanco, A., Dubcovsky, J., & Uauy, C. (2016). A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theoretical and Applied Genetics*, 129, 1099–1112. <https://doi.org/10.1007/s00122-016-2686-2>
- Suliman, S., Alemu, A., Abdelmula, A. A., Badawi, G. H., Al-Abdallat, A., & Tadesse, W. (2021). Genome-wide association analysis uncovers stable QTLs for yield and quality traits of spring bread wheat (*Triticum aestivum*) across contrasting environments. *Plant Gene*, 25, 100269. <https://doi.org/10.1016/j.plgene.2020.100269>
- Tang, H., Wang, H., He, M., Zhang, M., Hu, Y., Li, Z., Ren, T., Ren, Z., & Yan, B. (2017). QTL analysis for wheat falling number in a recombinant inbred line population segregated with 1BL/1RS translocation in a rainfed agricultural area of China. *Euphytica*, 213, 235. <https://doi.org/10.1007/s10681-017-2028-8>
- Torada, A., Koike, M., Ogawa, T., Takenouchi, Y., Tadamura, K., Wu, J., Matsumoto, T., Kawaura, K., & Ogiwara, Y. (2016). A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase. *Current Biology*, 26, 782–787. <https://doi.org/10.1016/j.cub.2016.01.063>
- Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., & Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc and iron content in wheat. *Science*, 314, 1298–1301. <https://doi.org/10.1126/science.1133649>
- Wang, M., Jiang, N., Jia, T., Leach, L., Cockram, J., Comadran, J., Waugh, R., Thomas, B., & Luo, Z. (2012). Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theoretical and Applied Genetics*, 124, 233–246. <https://doi.org/10.1007/s00122-011-1697-2>
- Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., Akhunova, A., Trick, H. N., Uauy, C., & Akhunov, E. (2018). Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of *TaGW2* homoeologues to grain size and weight in wheat. *Theoretical and Applied Genetics*, 131, 2463–2475. <https://doi.org/10.1007/s00122-018-3166-7>
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., Maccaferri, M., Salvi, S., Milner, S. G., Cattivelli, L., Mastrangelo, A. M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., International Wheat Genome Sequencing Consortium, Lillemo, M., Mather, D., ... Akhunov, E. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnology Journal*, 12, 787–796. <https://doi.org/10.1111/pbi.12183>
- Wang, M., & Xu, S. (2019). Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity*, 123, 287–306. <https://doi.org/10.1038/s41437-019-0205-3>
- Waugh, R., Marshall, D., Thomas, B., Comadran, J., Russell, J., Close, T., Stein, N., Hayes, P., Muehlbauer, G., Cockram, J., O'Sullivan, D., Mackay, I., Flavell, A., AGOUEB, BarleyCAP, & Ramsay, L. (2010). Whole-genome association mapping in elite inbred crop varieties. *Genome*, 53, 967–972. <https://doi.org/10.1139/G10-078>
- The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661–678. <https://doi.org/10.1038/nature05911>
- White, J., Law, J. R., Mackay, I., Chalmers, K. J., Smith, J. S. C., Kilian, A., & Powell, W. (2008). The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. *Theoretical and Applied Genetics*, 116, 439–453. <https://doi.org/10.1007/s00122-007-0681-3>
- Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., ... Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, 48, D682–D688. <https://doi.org/10.1093/nar/gkz966>
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203–208. <https://doi.org/10.1038/ng1702>
- Zanetti, S., Winzeler, M., Keller, M., Keller, B., & Messmer, M. (2020). Genetic analysis of pre-harvest sprouting resistance in a wheat x spelt cross. *Crop Science*, 40, 1406–1417. <https://doi.org/10.2135/cropsci2000.4051406x>
- Zhai, H., Feng, Z., Du, X., Song, Y., Liu, X., Qi, Z., Song, L., Li, J., Li, L., Peng, H., Hu, Z., Yao, Y., Xin, M., Xiao, S., Sun, Q., & Ni, Z. (2018). A novel allele of *TaGW2-A1* is located in a finely mapped QTL that increases grain weight but decreases grain number in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 131, 539–553. <https://doi.org/10.1007/s00122-017-3017-y>
- Zhang, Y., Tang, J., Zhang, Y., Yan, J., Xiao, Y., Zhang, Y., Xia, X., & He, Z. (2011). QTL mapping for quantities of protein fractions in bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 122, 971–987. <https://doi.org/10.1007/s00122-010-1503-6>

- Zhang, J., Chen, J., Bowman, B. C., O'Brien, K., Marshall, J. M., & Bonman, J. M. (2014). Association mapping of Hagberg falling number in hard white spring wheat. *Crop Science*, *54*, 1243–1252. <https://doi.org/10.2135/cropsci2013.08.0551>
- Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao, C., & Wang, D. (2018). Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *Plant Journal*, *94*, 857–866. <https://doi.org/10.1111/tpj.13903>
- Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., & Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** White, J., Sharma, R., Balding, D., Cockram, J., & Mackay, I. J. (2022). Genome-wide association mapping of Hagberg falling number, protein content, test weight, and grain yield in UK wheat. *Crop Science*, 1–16. <https://doi.org/10.1002/csc2.20692>