

University of Nebraska - Lincoln

## DigitalCommons@University of Nebraska - Lincoln

---

Agronomy & Horticulture -- Faculty Publications

Agronomy and Horticulture Department

---

4-11-2020

### Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions

Waseem Hussain

*University of Nebraska-Lincoln*

Malachy T. Campbell

*Virginia Polytechnic Institute and State University, [campbell.malachy@gmail.com](mailto:campbell.malachy@gmail.com)*

Diego Jarquin

*University of Nebraska-Lincoln, [jhernandezjarquin2@unl.edu](mailto:jhernandezjarquin2@unl.edu)*

Harkamal Walia

*University of Nebraska-Lincoln, [hwalia2@unl.edu](mailto:hwalia2@unl.edu)*

Gota Morota

*Virginia Polytechnic Institute and State University, Blacksburg, [morota@vt.edu](mailto:morota@vt.edu)*

Follow this and additional works at: <https://digitalcommons.unl.edu/agronomyfacpub>



Part of the [Agricultural Science Commons](#), [Agriculture Commons](#), [Agronomy and Crop Sciences Commons](#), [Botany Commons](#), [Horticulture Commons](#), [Other Plant Sciences Commons](#), and the [Plant Biology Commons](#)

---

Hussain, Waseem; Campbell, Malachy T.; Jarquin, Diego; Walia, Harkamal; and Morota, Gota, "Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions" (2020). *Agronomy & Horticulture -- Faculty Publications*. 1362.  
<https://digitalcommons.unl.edu/agronomyfacpub/1362>

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

## ORIGINAL RESEARCH

# Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions

Waseem Hussain<sup>1</sup>  | Malachy T. Campbell<sup>2</sup>  | Diego Jarquin<sup>1</sup> | Harkamal Walia<sup>1</sup> | Gota Morota<sup>2</sup> 

<sup>1</sup>Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

<sup>2</sup>Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

## Correspondence

Waseem Hussain, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA.

Email: waseem.hussain@irri.org;

Gota Morota, Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, 175 West Campus Drive, Blacksburg, VA 24061, USA.

Email: morota@vt.edu

## Present address

Waseem Hussain, International Rice Research Institute, Los Banos, Philippines

## Funding information

National Science Foundation, Grant/Award Number: 1736192

## Abstract

Genome-wide association mapping identifies quantitative trait loci (QTL) that influence the mean differences between the marker genotypes for a given trait. While most loci influence the mean value of a trait, certain loci, known as variance heterogeneity QTL (vQTL) determine the variability of the trait instead of the mean trait value (mQTL). In the present study, we performed a variance heterogeneity genome-wide association study (vGWAS) for grain cadmium (Cd) concentration in bread wheat. We used double generalized linear model and hierarchical generalized linear model to identify vQTL associated with grain Cd. We identified novel vQTL regions on chromosomes 2A and 2B that contribute to the Cd variation and loci that affect both mean and variance heterogeneity (mvQTL) on chromosome 5A. In addition, our results demonstrated the presence of epistatic interactions between vQTL and mvQTL, which could explain variance heterogeneity. Overall, we provide novel insights into the genetic architecture of grain Cd concentration and report the first application of vGWAS in wheat. Moreover, our findings indicated that epistasis is an important mechanism underlying natural variation for grain Cd concentration.

## 1 | INTRODUCTION

Genome-wide association studies (GWAS) are routinely conducted to study the genetic basis of important traits in crops.

**Abbreviations:** ABC transporter, ATP-binding cassette transporter; DGLM, double generalized linear model; GLM, generalized linear model; GRM, genomic relationship matrix; HGLM, Hierarchical generalized linear model; HWW, hard-red winter wheat; mQTL, mean quantitative trait loci; mvQTL, mean-variance quantitative trait loci; QTL, quantitative trait loci; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; vQTL, variance heterogeneity quantitative trait loci; vGWAS, variance heterogeneity genome-wide association studies.

GWAS link phenotypic variation with dense genetic marker data using a linear modeling framework (e.g., Huang & Han, 2014; Ingvarsson & Street, 2011; Nordborg & Weigel, 2008; Xiao, Liu, Wu, Warburton, & Yan, 2017). Standard GWAS approaches seek to identify marker-trait associations that influence the mean phenotypic values. However, differences in the variance between genotypes are also under genetic control (Shen, Pettersson, Rönnegård, & Carlborg, 2012). As a result, several recent studies have identified loci associated with differences in variance between genotypes (Cao, Wei, Bailey, Kauwe, & Maxwell, 2014; Corty, Kumar, Tarantino, Takahashi, & Valdar, 2018). Such genetic variants that affect the variance heterogeneity of traits have been referred to as

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *The Plant Genome* published by Wiley Periodicals, Inc. on behalf of Crop Science Society of America

variance heterogeneity quantitative trait loci (vQTL) (Rönnegård & Valdar, 2011). vQTL can be detected by searching the difference in the variability between the groups of genotypes that carry alternative alleles at a particular locus (Forsberg & Carlborg, 2017). A simple example is genotypes of wheat with difference in plant height. One genotype group is homozygous for a certain allele and manifests greater variability (including both shorter and taller plants), while the second genotype group that is homozygous for the alternative allele involves plants that are similar or uniform in height. This contrast in plant height across two allelic groups leads to genetic variance heterogeneity. Note that the mean difference between the two groups does not have to be different for variance heterogeneity to arise (Figure 1).

Variance heterogeneity-based genome-wide association studies (vGWAS) have emerged as a new approach for identifying and mapping vQTL. vQTL contribute to variability, which is undetected through standard statistical mapping (bi-parental or association) procedures (Forsberg & Carlborg, 2017; Rönnegård & Valdar, 2011; Shen et al., 2012). It has been argued that variance heterogeneity between genotypes can be partially explained by epistasis or gene-by-environment interactions (Brown et al., 2014; Forsberg & Carlborg, 2017; Young, Wauthier, & Donnelly, 2018). Thus, vQTL can provide insights into epistasis or phenotypic plasticity (Nelson, Pettersson, Li, & Carlborg, 2013; Young et al., 2018). Moreover, these vGWAS frameworks can serve as tractable approaches to reduce the search space when assessing epistasis among markers (Brown et al., 2014; Wei et al., 2016). This is because we can limit the number of interacting marker pairs  $\binom{m}{2}$  to be investigated into  $m + \binom{k}{2}$ , where  $k$  is the number of markers ( $k < m$ ) associated with vQTL or mvQTL.

Numerous studies have reported vQTL associated with diverse phenotypes, including the tendency to left-right turning and bristles (Mackay & Lyman, 2005) and locomotor handedness (Ayroles et al., 2015) in *Drosophila*; coat color (Nachman, Hoekstra, & D'Agostino, 2003), circadian activity, and exploratory behavior (Corty et al., 2018) in mice; thermotolerance (Queitsch, Sangster, & Lindquist, 2002), flowering time (Salomé et al., 2011), and molybdenum concentration (Forsberg et al., 2015; Shen et al., 2012) in *Arabidopsis*; litter size in swine (Sell-Kubiak et al., 2015); urinary calcium excretion in rats (Perry et al., 2012); and body mass index (Yang et al., 2012; Young et al., 2018), sero-negative rheumatoid arthritis (Wei, Viatte, Merriman, Barton, & Worthington, 2017), and serum urate (Topless et al., 2015) in humans. In plants, vGWAS have been limited to few species, including *Arabidopsis* (Forsberg et al., 2015; Shen et al., 2012) and maize (Kusmec, Srinivasan, Nettleton, & Schnable, 2017).

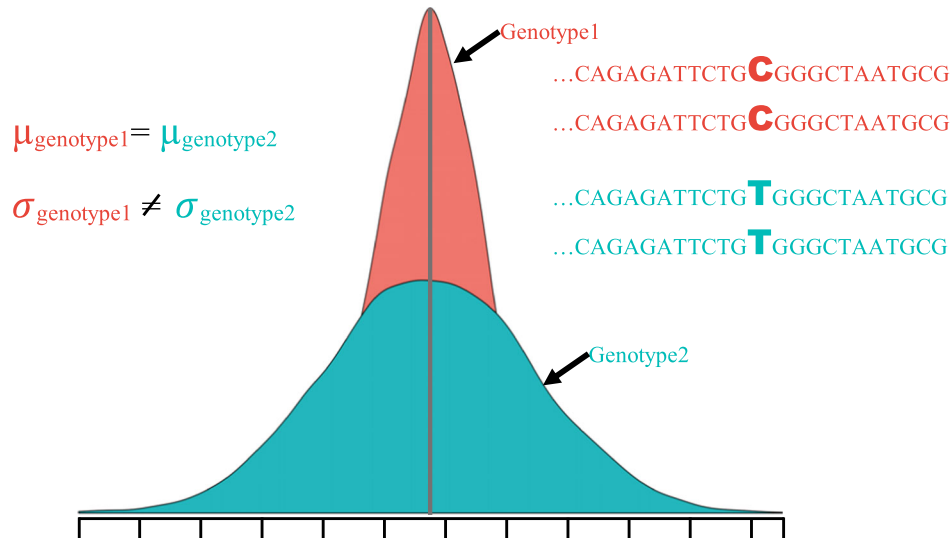
Methodologically, vQTL have been detected by performing statistical tests searching for unequal variance for a

### Core Ideas

- Variance heterogeneity mapping for grain Cadmium (Cd) concentration in bread wheat was performed.
- Novel variance heterogeneity loci were detected on chromosomes 2A and 2B.
- Loci influencing both mean and variance were identified on chromosome 5A.
- Identified variance heterogeneity loci were associated with epistatic interactions.
- Homoeology within the vQTL on chromosomes 2A and 2B was found.

quantitative trait between the marker genotypes (Rönnegård & Valdar, 2012). The most common statistical tests used to identify vQTL include Levene's test (Paré, Cook, Ridker, & Chasman, 2010), Brown-Forsythe test (Brown & Forsythe, 1974), squared residual value linear modeling (Struchalin, Amin, Eilers, van Duijn, & Aulchenko, 2012), and correlation least squares test (Brown et al., 2014). However, these methods have certain drawbacks when applied to genetic data. For example, Levene's and Brown-Forsythe tests are sensitive to deviations from normality of residuals and have an inherent inability to model continuous covariates (Dumitrascu, Darnell, Ayroles, & Engelhardt, 2019; Rönnegård & Valdar, 2011).

Double generalized linear model (DGLM) has emerged as an alternative approach to model the variance heterogeneity for genetic studies (Rönnegård & Valdar, 2011). In DGLM, sample means and residuals are modelled jointly. Here, generalized linear models (GLM) are fit by including only the fixed effects in the linear predictor(s) for the mean and then the squared residuals are used to estimate the dispersion effects. It is important to correct for population structure, which can otherwise lead to spurious associations in GWAS (Patterson, Price, & Reich, 2006). In DGLM, population structure can be corrected by incorporating the first few principal components of a genomic relationship matrix (GRM) (Patterson et al., 2006; Price, Zaitlen, Reich, & Patterson, 2010) as fixed covariates in the model. However, the first few principal components may not be sufficient to account for complex population structure or family relatedness (Hoffman, 2013; Sul, Martin, & Eskin, 2018). Alternatively, we can fit linear mixed models (LMM) to explicitly correct for population structure, where the whole GRM can be included to account for relationships among individuals and correct for background genotype effects. Hierarchical generalized linear model (HGLM) has been proposed as an extension of the DGLM to model random effects in the mean component (Rönnegård &



**FIGURE 1** Illustration of variance heterogeneity of two genotype groups at a biallelic locus affecting the variance not the mean. Genotypes with CC allelic combination present narrow variance, whereas genotypes with TT allelic combination show greater variability. The mean difference between two genotype groups is the same as shown by the solid vertical gray line

Valdar, 2012; Tan et al., 2014). In HGLM, the GRM can be used to model correlated random effects and account for population structure.

We applied a vGWAS framework to examine the genetic architecture of grain cadmium (Cd) accumulation in wheat. Cadmium is a heavy metal that is highly toxic to human health (Menke, Muntner, Silbergeld, Platz, & Guallar, 2009). Identifying genetic variants that control low-grain Cd concentration in wheat is necessary to understand the basis for phenotypic variation in grain Cd and can help accelerate the development of low Cd wheat varieties. A recent study assessed natural variation in bread wheat grain Cd by conducting GWAS (Guttieri et al., 2015a). However, only a fraction of phenotypic variation could be explained by the top marker associations, indicating that grain Cd concentration is a complex trait that is influenced by multiple loci and/or loci with non-additive effects (Guttieri et al., 2015a). Given the genetic complexity of Cd in wheat, we hypothesized that variation in grain Cd concentration in wheat is influenced by vQTL that are likely to be involved in epistatic interactions; this would allow us to capture additional variation that is not accounted for in a standard GWAS approach.

In this study, we sought to provide additional insights into natural variation in grain Cd concentration by extending the standard GWAS to vGWAS using a hard winter wheat association mapping panel. To achieve this, we used DGLM and HGLM to perform vGWAS. Previously, Guttieri et al. (2015a) conducted the standard GWAS using this association panel and identified a single mean effect QTL (mQTL) for grain Cd concentration on chromosome 5A. In addition, we aimed to understand the basis of vQTL by searching for pairwise epistatic interactions among vQTL and mQTL. To our

knowledge, the present study is the first to conduct vGWAS and identify vQTL associated with grain Cd concentration in wheat.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant materials and genotyping

We analyzed a publicly available dataset comprising of phenotypes for grain mineral concentration for  $n = 299$  genotyped hard-red winter wheat accessions (hereafter called as HWW association panel). The details of the study are discussed in Guttieri et al. (2015a, 2015b), and access to the data is available at <http://triticeaetoolbox.org/wheat/>. The data are also downloadable at <https://github.com/whussain2/vGWAS/tree/master/Data>. Here, we focused on grain Cd concentration (mg/kg) collected across two years (2012 and 2013) in one location (Oklahoma, USA). Briefly, the experiment was laid in an augmented incomplete block design with two replications and 15 blocks within each replication. We employed a two-stage approach because of its simplicity and computational efficiency. Least square means adjusted across the replications and blocks in each year were obtained for each genotype. In this study, we averaged the least square means for each genotype across two years because of non-significant genotype  $\times$  year interaction (Guttieri et al., 2015a). The association panel was genotyped using a 90K iSelect Infinium array (Wang et al., 2014b). We used a filtered marker data set consisting of  $m = 14,731$  single nucleotide polymorphism (SNP) markers from the 90K iSelect Infinium array as described by Guttieri et al. (2015a). All the SNP markers were physically

anchored on the new reference genome of hexaploid wheat RefSeq v1.0 (IWGSC et al., 2018).

## 2.2 | Statistical modeling

### 2.2.1 | Genome-wide association mapping

Standard GWAS or mQTL analysis based on mean differences between marker genotypes for grain Cd concentration was performed similar to Gutteri et al. (2015a) using the rrBLUP package (Endelman, 2011) in the R environment (R Core Team, 2018).

### 2.2.2 | Variance heterogeneity genome-wide association mapping

We used DGLM and HGLM to perform vGWAS and detect vQTL in the current study. The description of models used is given below.

## 2.3 | DGLM

DGLM is a parametric approach that can be used to jointly model the mean and dispersion using a GLM framework (Smyth, 1989). The DGLM works iteratively by first fitting a linear model to estimate the mean effects (mQTL). The squared residuals are used to estimate the dispersion effects (vQTL) using GLM with a gamma-distributed response and the log link function. This process is cycled until convergence. Here, we extended the DGLM to marker-based association analysis according to Rönnegård and Valdar (2011). The mean part of DGLM was as follows:

$$y = \mathbf{1}\mu_m + \mathbf{X}\beta + \mathbf{s}_j a_{mj} + \varepsilon \quad (1)$$

where  $y$  is the Cd concentration (mg/kg);  $\mathbf{1}$  is the column vector of 1;  $\mu_m$  is the intercept;  $\mathbf{X}$  is  $n \times 4$  covariate matrix of the top four principle components (PCs) obtained by performing principal component analysis (PCA) of marker data using the SNPRelate R package (Zheng et al., 2012);  $\beta$  is the regression coefficient for the covariates;  $\mathbf{s}_j \in (0,2)$  is the vector containing the number of reference allele at the marker  $j$ ;  $a_{mj}$  is the effect size or allele substitution effect of the  $j$ th marker; and  $\varepsilon$  is the residual. We assumed

$$\varepsilon_i \sim N(0, \mathbf{I}\sigma_{\varepsilon_i}^2)$$

$$\log(\sigma_{\varepsilon_i}^2) = \mu_v + \mathbf{s}_{ij} a_{vj}$$

where  $\sigma_{\varepsilon_i}^2$  is the residual variance for genotype  $i$ ;  $\mathbf{I}$  is an identity matrix; and  $\mu_v$  and  $a_v$  are the intercept and marker regression coefficients for the variance part of the model, respectively. While we fit separate effects for the mean using a standard linear model and for the variance using the squared residuals in gamma distributed GLM with a log link function, this is equivalent to modeling  $y \sim N(\mathbf{1}\mu_m + \mathbf{X}\beta + \mathbf{s}_j a_{mj}, \exp(\mathbf{1}\mu_v + \mathbf{s}_j a_{vj}))$  or  $\varepsilon \sim N[0, \exp(\mathbf{1}\mu_v + \mathbf{s}_j a_{vj})]$  in Equation (1).

The DGLM was fitted using the dglm package (<https://cran.r-project.org/web/packages/dglm/index.html>) in R. SNP markers were fitted one by one, and for each marker, the effect sizes, standard errors, and p-values were obtained for the mean and dispersion components. To account for multiple testing, we determined the effective number of independent tests ( $M_{\text{eff}}$ ) using the method described by Li and Ji (2005). Subsequently, a genome-wide significance threshold level ( $P < 1.44 \times 10^{-5}$ ) was determined using the following formula:

$$\alpha_p = 1 - (1 - \alpha_e)^{\frac{1}{M_{\text{eff}}}} \quad (2)$$

where  $\alpha_p$  is the genome-wide significance threshold level;  $\alpha_e$  is the desired level of significance (0.05); and  $M_{\text{eff}} = 3,495$ .

## 2.4 | HGLM

To explicitly account for population structure and kinship in GWAS, LMM have been proposed as alternative methods that allow the genetic relationships between individuals to be modeled as random effects. To perform vGWAS in the LMM framework and to identify genome-wide vQTL, we used a HGLM approach. HGLM (Lee & Nelder, 1996) is a class of GLM and is a direct extension of the DGLM that allows joint modelling of the mean and dispersion parts and introduces random effects as a linear predictor for the mean (Rönnegård & Carlborg, 2007). The mean part of HGLM was given as follows:

$$y = \mathbf{1}\mu_m + \mathbf{s}_j a_{mj} + \mathbf{Z}\mathbf{u} + \varepsilon \quad (3)$$

assuming that

$$\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$$

where  $\mathbf{Z}$  is the incident matrix of random effects of genotypes;  $\mathbf{u}$  is the vector of random effects with  $\text{Var}(\mathbf{u}) = \mathbf{G}\sigma_u^2$ ;  $\mathbf{G}$  is the GRM of VanRaden (2008); and  $\sigma_u^2$  is the additive genetic variance. A log-link function was used for the residual variance given by  $\exp(\mathbf{s}_j a_{vj})$ , which is equivalent to modeling  $y|a_{mj}, \mathbf{u}, a_{vj} \sim N[\mathbf{1}\mu_m + \mathbf{s}_j a_{mj} + \mathbf{Z}\mathbf{u}, \exp(\mathbf{1}\mu_v + \mathbf{s}_j a_{vj})]$ .

We fitted HGLM using the hglm R package (Rönnegård, Shen, & Alam, 2010b). We reformulated the term  $\mathbf{Z}\mathbf{u}$  as



$\mathbf{Z}^* \mathbf{u}^*$ , where  $\mathbf{u}^* \sim N(0, \mathbf{I}\sigma_u^2)$ ;  $\mathbf{Z}^* = \mathbf{Z}_0 \mathbf{L}$ ;  $\mathbf{L}$  is the Cholesky factorization of the  $\mathbf{G}$  matrix; and  $\mathbf{Z}_0$  is the identity matrix (Rönnegård, Felleki, Fikse, Mulder, & Strandberg, 2010a) because the hglm package does not accept the  $\mathbf{G}$  matrix. Markers treated as fixed effects were fit one by one, and for each marker, the effect sizes, standard errors, and p-values were obtained for the mean and dispersion components. The genome-wide significance threshold level was derived as described in the DGLM analysis. Circular Manhattan and quantile-quantile (QQ) plots were created using the CMplot R package (<https://github.com/YinLiLin/R-CMplot>).

## 2.5 | Epistasis analysis

We investigated the extent of epistasis that was manifested through variance heterogeneity. All the possible pairwise interaction analyses for markers that were associated with grain Cd concentration were performed using the following two markers at a time epistatic model:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\beta + \mathbf{s}_j a_j + \mathbf{s}_k a_k + (\mathbf{s}_j \mathbf{s}_k) v_{jk} + \varepsilon \quad (4)$$

where  $\mathbf{y}$  is the vector of Cd concentration (mg/kg);  $\mathbf{X}$  is the incident matrix for the first four PCs;  $\beta$  is the regression coefficients for the PCs;  $\mathbf{s}_j$  and  $\mathbf{s}_k$  are SNP codes for the  $j$ th and  $k$ th markers, respectively;  $a_j$  and  $a_k$  are the additive effects of the markers  $j$  and  $k$ , respectively; and  $v_{jk}$  is the additive  $\times$  additive epistatic effect of the  $j$ th and  $k$ th markers. We used Bonferroni correction to account for the multiple testing. The threshold of  $-\log_{10}(0.05/325) = -\log_{10}(1.54 \times 10^{-4}) = 3.8$  was used to declare the significance of interaction effects.

## 2.6 | Homoeology and candidate gene analysis

Homoeologous gene construction was performed as per procedure described by (Santantonio, Jannink, & Sorrells, 2019). Briefly, the annotated coding sequences within the 2A vQTL were aligned back onto themselves using the IWGSC RefSeq v.1.0 coupled with BLAST tool in Ensemble Plants browser (Bolser, Staines, Perry, & Kersey, 2017). For candidate gene identification for the SNP markers associated with variance heterogeneity, we used Ensembl Plants browser to retrieve the candidate genes and functional annotations ([http://plants.ensembl.org/Triticum\\_aestivum/Info/Index](http://plants.ensembl.org/Triticum_aestivum/Info/Index)) and the wheat RefSeq v1.0 annotations (IWGSC et al., 2018; <https://wheaturgi.versailles.inra.fr/Seq-Repository/Annotations>). For candidate gene analysis, we first determined the positions of significant SNP markers, and the interval was defined as the physical distance between the lowest and highest positions of SNP markers for each of the vQTL regions. For example, if the position of the lowest SNP and highest SNP was

715,333,165 bp and 717,146,211 bp in the vQTL region on chromosome 2A, we defined 2A as the 715,333,165 to 717,146,211 interval for candidate gene identification. After defining the interval for the 2A (2A: 715,333,165 to 717,146,211) and 2B (2B: 691,780,716 to 701,097,263 bp) regions, we explored the intervals using Ensembl Plants browser and extracted the Gene IDs within these intervals. The Gene IDs within the defined interval on chromosomes 2A and 2B were analyzed using the IWGSC RefSeq v.1.0 (IWGSC et al., 2018) integrated genome annotations to obtain the predicted genes and functional annotations.

## 3 | RESULTS

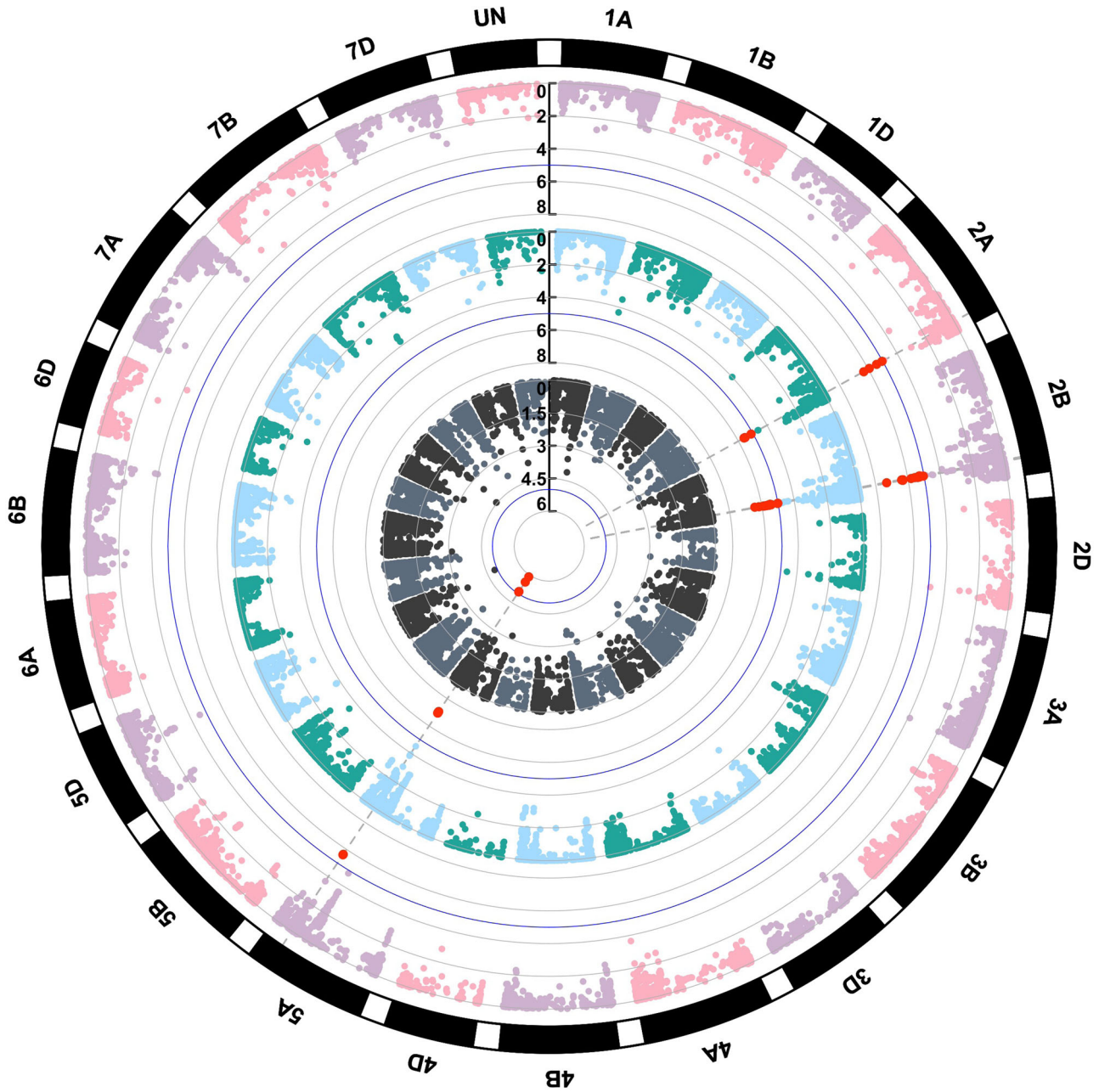
### 3.1 | Variance heterogeneity GWAS provide additional insight into natural variation in grain Cd

Although grain Cd concentration is a highly heritable trait, recent GWAS revealed that significant loci can only explain a fraction of the variation for this trait (Guttieri et al., 2015a). We found the single genomic region on chromosome 5A affecting the grain Cd concentration (Figure 2) from the standard GWAS analysis confirming the results of Guttieri et al. (2015a). The DGLM and HGLM approaches were used to detect vQTL while controlling for population structure. The population structure based on PCA of the HWW association panel is given in Supplemental Figure S1. Supplemental Figure S2 shows QQ plots from DGLM and HGLM.

We classified the QTL into the following categories: mQTL, which contributes to difference in the means between marker genotypes; vQTL, which influences the variability between the genotypes; and mean-variance QTL (mvQTL), which contributes to differences in both the mean and variance between the genotypes.

Based on the DGLM, we identified two vQTL associated with the variance heterogeneity of Cd concentration. One vQTL on 2A contained four SNP markers, and one vQTL on 2B contained 17 SNP markers (Figure 2; Supplemental Table S1). The four SNP markers associated with the vQTL region on the chromosome 2A region spanned the physical distance of 1.81 Mb; among the four SNP markers, three of them within the 666 bp region formed a single linkage disequilibrium (LD) block (Supplemental Figure S3). The vQTL region on 2B associated with 17 SNP markers spanned the physical distance of 9.32 Mb, and the SNP markers were located within four LD blocks of sizes < 1 kb, 1 kb, 1 kb, and 204 kb (Supplemental Figure S4).

In addition, we identified a single mvQTL (containing four SNP markers) associated with both mean and variance heterogeneity on chromosome 5A (Figure 2; Supplemental



**FIGURE 2** Circular Manhattan plot of standard genome-wide association studies (GWAS) based on mean differences (inner), and variance GWAS using double generalized linear model (middle) and hierarchical generalized linear model (outer) for grain cadmium concentration in the hard-red winter wheat association panel. The red dots represent the significant markers associated with either mean or variance heterogeneity quantitative trait loci. The blue line in each circular plot shows the cutoff for the statistical significance. The P-values in  $-\log_{10}$  scale are given in black vertical line

Table S2). The markers associated with mvQTL on chromosome 5A were identical to those obtained in the original GWAS analysis according to Guttieri et al. (2015a), indicating that this region affects both the mean and the variance heterogeneity (Supplemental Figure S5). Moreover, these results showed that DGLM serves as an accurate framework to jointly detect mean and variance QTL and provides additional insights into phenotypic variation that would otherwise not be captured by standard GWAS.

The HGLM analysis revealed the same results as those obtained using DGLM and showed identical vQTL on chromosomes 2A and 2B and mvQTL on chromosome 5A associated with variance heterogeneity of Cd concentration (Figure 2; Supplemental Table S1). Further, we observed a potential vQTL region on 2D from the DGLM and HGLM analyses. This region was slightly below the significance threshold level but may have an implication on Cd variation given that the allopolyploid nature of wheat and the role of

homoeologous gene sets on phenotypic variation (Borrill, Harrington, & Uauy, 2019).

### 3.2 | Variance heterogeneity loci can be partially explained by epistasis

We investigated all significant markers (25 markers) associated with mvQTL on chromosome 5A and vQTL on chromosomes 2A and 2B and explored all possible pairwise additive  $\times$  additive epistatic interactions. We detected significant additive  $\times$  additive interactions between the markers (Figure 3). The interaction was more evident between mvQTL on chromosome 5A and vQTL on chromosomes 2A and 2B. Specifically, all the markers associated with the 5A mvQTL region revealed highly significant interactions with all the markers associated with the 2A and 2B vQTL regions. Interactions between vQTL on chromosomes 2A and 2B were also observed; however, the interactions were less evident, and only a few markers within these regions showed statistically significant interactions. Taken together, these results suggested that the vQTL and mvQTL may be manifested because of pairwise epistatic interactions.

### 3.3 | Homoeology and candidate genes

Homoeology analysis between the defined regions on chromosomes 2A and 2B resulted in 22 homoeologous gene sets, consisting of 21 triplicates and only one duplicate gene set. Additional details on the homoeologous gene sets can be found in Supplemental Materials (excel file). As compared to the total number of candidate genes equal to 39 within the 1.18 Mb 2A region, 22 (58%) were homoeologous across the three genomes. Based on the annotations for the 22 homoeologous gene sets, a few of the genes encoded homeobox-leucine zipper family protein, plant peroxidase, and glycosyltransferase, which have been associated with the genetic regulation of minerals in plants (Whitt et al., 2018). For example, homeodomain-leucine zipper family protein has been functionally associated with Cd tolerance by regulating the expression of metal transporters *OsHMA2* and *OsHMA3* in rice (Ding et al., 2018; Yu et al., 2019). These genes have been found to play important roles in loading Cd onto the xylem and root-to-shoot translocation of Cd in rice. In plants, response to heavy metals involves the accumulation of reactive oxygen species (ROS) that damage DNA and cellular machinery (Kumari, Taylor, & Deyholos, 2008; Rascio & Navari-Izzo, 2011). In *Arabidopsis*, the peroxidase genes *At2g35380*, *PER20*, and *At2g18150* have been found to be associated with Cd responses by affecting the lignin biosynthesis in root cells under high Cd stress (Chen & Kao, 1995; van de Mortel et al., 2008). Full list of candidate genes within

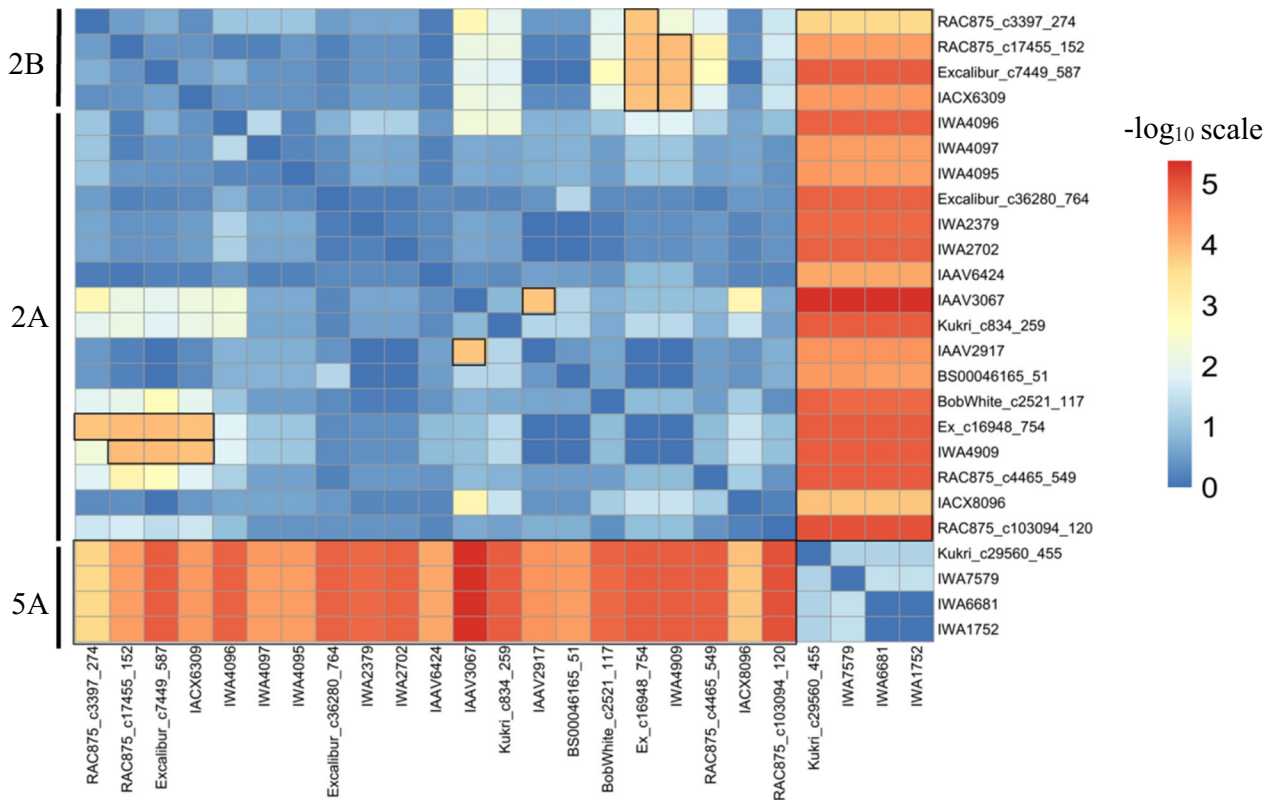
the 2A and 2B region, and within the homoeologous gene sets is in Supplemental Materials (excel file). These results clearly indicate that most of the genes with vQTL regions are redundant across the genomes and may have significant role in the genetic regulation of grain Cd concentration in wheat. However, we contend that further investigation of these regions using dense markers and increased sample size is necessary to fine-map the QTL and validate potential candidate genes underlying these loci and also the role of gene redundancy in generating phenotypic variation.

## 4 | DISCUSSION

In the present study, we explored the genetic variants affecting variance heterogeneity of Cd. Given the complexity of genetic regulation of Cd in wheat (Guttieri et al., 2015a) and the influence of epistatic interactions, we anticipated that partial genetic regulation of Cd in wheat can be detected using methods that have been developed to identify vQTL. As reported by Rönnegård and Valdar (2012), a potential explanation for variance-controlling QTL is epistatic interactions that are unspecified in the model. Herein, we utilized two approaches, namely, DGLM and HGLM, to detect vQTL and mvQTL associated with grain Cd concentration in wheat.

The DGLM framework is a powerful approach for vGWAS analysis. However, in DGLM, GLM is fit by including only the fixed effects in the linear predictor of mean and dispersion. Therefore, by using the DGLM approach, population structure can only be accounted for by using the first few PCs obtained from the SNP matrix; however, this may not completely account for complex population structure and family relationships (Price et al., 2010). We hypothesized that the use of random effects to model the mean component can better account for population structure and reduce spurious associations. In this approach, a random additive genetic effect is introduced to the mean component of the model that accounts for population structure and cryptic relatedness between accessions. Therefore, we performed vGWAS analysis using HGLM. Interestingly, both DGLM and HGLM approaches were effective in identifying the genetic variants controlling variability of Cd, suggesting that the loci detected with the DGLM approach are likely to be true QTL rather than artifacts from population structure. The impact of population structure on the power of DGLM and HGLM remains to be explored; further examination is warranted. We also tested a double hierarchical generalized linear model (DHGLM), which can account for random effects in both mean and variance parts. However, we did not include DHGLM in the current study because we encountered model convergence issues for many SNP markers. This is likely because of fitting a complex model to a small data set.





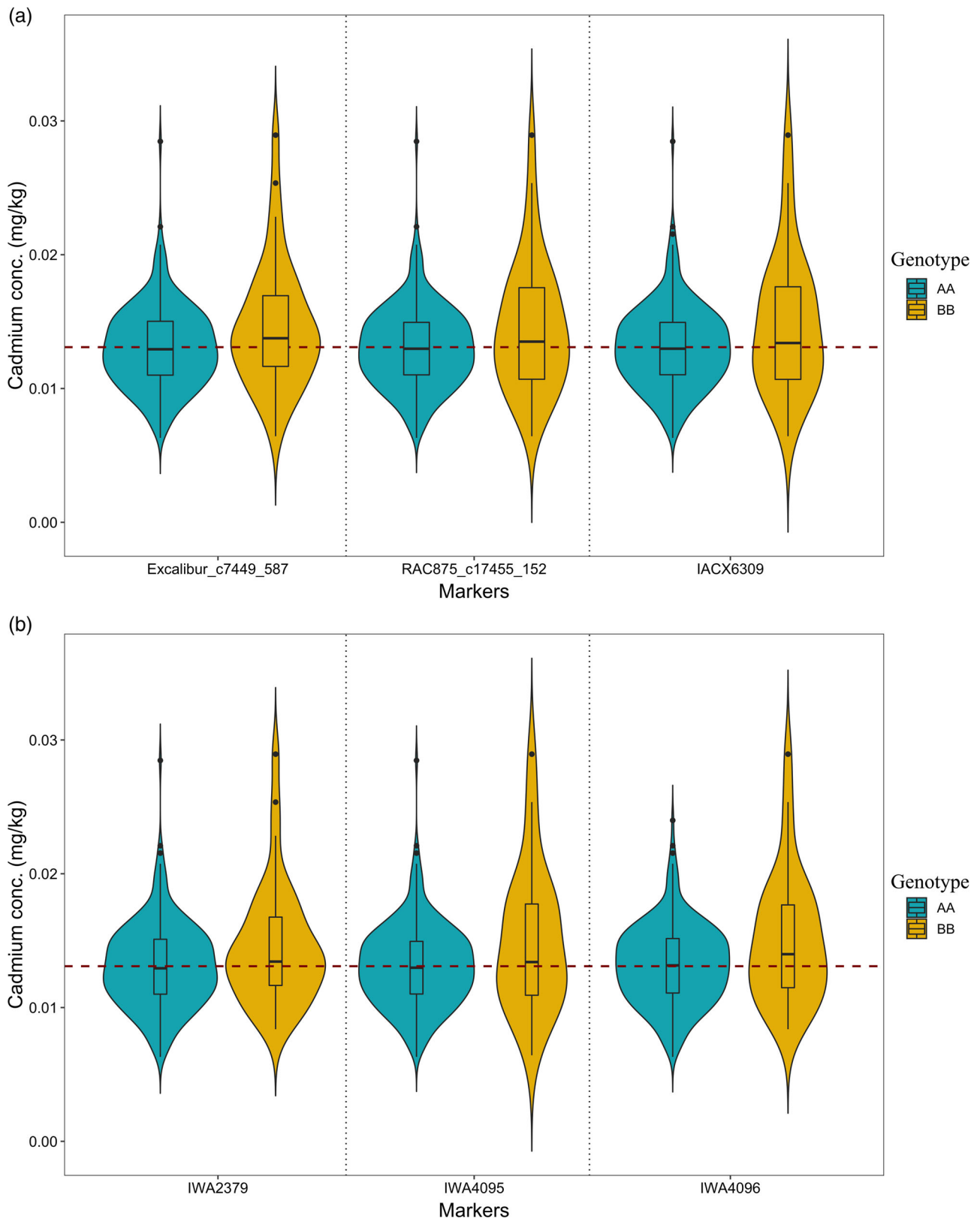
**FIGURE 3** Heat map showing all possible pairwise epistatic interactions between the markers associated with vQTL on chromosomes 2A and 2B or mvQTL on chromosome 5A. Chromosome information of each marker is given on the left side. The heat map is sorted and color coded based on  $-\log_{10}$  (p-value) scale with the legend given on right side. Interactions that are significant ( $-\log_{10} > 3.8$ ) are color coded as red or orange in color and outlined in black box

In the literature, it has been argued that variance heterogeneity can also arise by a simple mean variance relationship, which does not have biological significance (Young et al., 2018). To rule out the role of the mean-variance function in generating variance heterogeneity, we plotted the estimated effects of the top three significant associated vQTL markers at the alternate genotypes and observed that the means of all the markers were the same (Figure 4), indicating that the effect of SNP on variance heterogeneity was not due to the consequences of mean-variance function but likely due to the genetic effects (Yang et al., 2012).

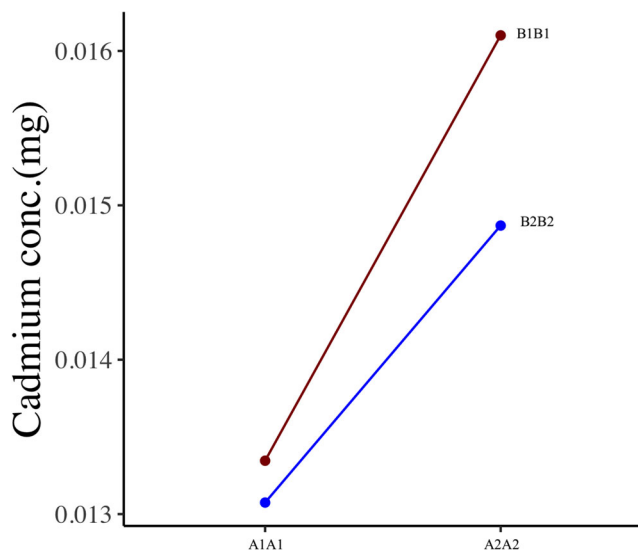
Further, variance heterogeneity can also be observed in a population when two or more alleles having different effects on the phenotype are in high LD (Cao et al., 2014; Forsberg & Carlborg, 2017; Wang, Yang, Brinkmeyer-Langford, & Cai, 2014a). To rule out the possibility of LD as a source for variance heterogeneity in grain Cd in this population, we suggest the use of high-density markers and larger sample size to identify the actual functional alleles associated with Cd, their LD patterns, and their effects on the Cd phenotype (Forsberg & Carlborg, 2017; Struchalin, Dehghan, Wittman, van Duijn, & Aulchenko, 2010).

In QTL studies, variance heterogeneity arises because of various underlying mechanisms, such as epistatic interactions

(Nelson et al., 2013; Shen et al., 2012; Struchalin et al., 2010). Epistasis gives rise to variance heterogeneity when the different allele combinations at one locus change the effect of the other loci in the genome, as shown in one pair of interacting markers (Figure 5). Hence, identifying the loci affecting variance heterogeneity through vGWAS means that the loci are likely to be involved in epistatic interactions. To validate this assumption and investigate whether epistasis can explain the identified vQTL and mvQTL in this study, we analyzed all possible pairwise interactions between the associated markers. We detected significant epistatic interactions between the associated markers (Figure 2), which can explain the existence of variance heterogeneity in the genotypes. Additionally, identifying vQTL through vGWAS serves as an effective way to restrict the search space when detecting epistatic QTL. Thus, with the vGWAS approach, many of the requirements necessary for conventional epistasis mapping can be avoided (e.g., large sample size and extensive multiple testing corrections that reduce power). However, Forsberg and Carlborg (2017) empirically showed that the presence of variance heterogeneity does not always guarantee the presence of epistatic interactions that contribute to the total variation of the trait; therefore, the results should be interpreted carefully when multi-locus interactions are involved.



**FIGURE 4** Violin plot showing the differences in the mean and variance of grain cadmium concentration with alternative marker genotype groups coded as AA and BB for the top three significant markers associated with vQTL on (a) chromosome 2A and (b) chromosome 2B. The mean of marker genotypes AA and BB are connected by red dotted line



**FIGURE 5** Epistatic interaction plot between the marker pair IAAV3067 (A1A1 and A2A2) on chromosome 2B (vQTL) and IWA7579 (B1B1 and B2B2) on chromosome 5A (mvQTL). The y-axis shows the phenotypic value of cadmium concentration (mg). Plotted points are phenotypic means

The genomic regions on chromosomes 2A and 2B associated with variance heterogeneity revealed homoeologous gene sets with 58% genes revealing the gene redundancy mostly present as three functional homoeologous copies (triplicated). This also indicates that genetic complexity of Cd phenotype is not only controlled by multiple genes but may be affected by the multiple homoeologs of the individual genes which warrants further investigation. Presence of multiple copies of homoeologous genes (Glover, Redestig, & Dessimoz, 2016) may have consequence on phenotypic variation due to dosage effects and or functional redundancy (Borrill et al., 2019). Dosage effect, in which the phenotypic variation is amplified by the addition of each gene copies can act additively (e.g., genes controlling grain protein content (Avni et al., 2014) and grain size (Wang et al., 2018, 2019) or non-additively (e.g., genes controlling amylopectin content in wheat (Kim, Johnson, Graybosch, & Gaines, 2003)). Non-additive variation between homoeologous gene has been shown to be an important source of variation in wheat. However, its relative contribution across the wheat genome as compared to non-syntenic regions was proportionately less (Santantonio et al., 2019). This is in agreement with our results because we observed interactions among the homoeologous genomic regions on chromosomes 2A and 2B. However, this homoeologous gene interactions was less evident as compared to two-way interactions found between non-syntenic vQTL regions on 2A and 2B with the mvQTL region on 5A. The nature and functional role of homoeologous gene sets within the vQTL region on 2A and 2B is not clear. However, it is increasingly feasible in wheat to examine the effects of gene redundancy

and explore the contribution of homoeologous genes in generating phenotypic variation (Wang et al., 2018).

## 5 | CONCLUSION

We showed the potential of vGWAS for dissecting the genetic architecture of complex traits and identifying novel genomic regions influencing variance heterogeneity in wheat. We provided evidence that the vQTL contribute to natural variation in grain Cd concentration through non-additive genetic effects. This is particularly evidenced by epistatic interactions between mvQTL on chromosome 5A and vQTL on chromosomes 2A and 2B.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grant Number 1736192 to H.W. and G.M. Data analysis was performed using the Holland Computing Center computational resources at the University of Nebraska-Lincoln.

## DATA AVAILABILITY

The wheat phenotypic and genotypic data can be downloaded from (<http://triticeaetoolbox.org/wheat/>) and also available on the GitHub repository <https://github.com/whussain2/vGWAS>. The R code used for the analysis is available on the GitHub repository <https://github.com/whussain2/vGWAS>.

## AUTHOR CONTRIBUTIONS

W.H. and G.M. conceived the study. W.H. performed the data analysis and drafted the manuscript. D.J. helped the data analysis. M.C., D.J., H.W., and G.M. revised the manuscript. G.M. supervised and directed the study. All authors read and approved the manuscript.

## CONFLICT OF INTEREST

The authors declare there are no competing interests.

## ORCID

Waseem Hussain  <https://orcid.org/0000-0002-6861-0193>

Malachy T. Campbell  <https://orcid.org/0000-0002-8257-3595>

<https://orcid.org/0000-0002-8257-3595>

Gota Morota  <https://orcid.org/0000-0002-3567-6911>

## REFERENCES

- Avni, R., Zhao, R., Pearce, S., Jun, Y., Uauy, C., Tabbita, F., ... Distelfeld, A. (2014). Functional characterization of GPC-1 genes in hexaploid wheat. *Planta*, 239, 313–324. <https://doi.org/10.1007/s00425-013-1977-y>
- Ayroles, J. F., Buchanan, S. M., O'Leary, C., Skutt-Kakaria, K., Grenier, J. K., Clark, A. G., ... de Bivort, B. L. (2015). Behavioral

- idiosyncrasy reveals genetic control of phenotypic variability. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 6706–6711. <https://doi.org/10.1073/pnas.1503830112>
- Bolser, D. M., Staines, D. M., Perry, E., & Kersey, P. J. (2017). Ensembl Plants: Integrating tools for visualizing, mining, and analyzing plant genomic data. In A. D. J. van Dijk (Ed.), *Plant bioinformatics* (pp. 115–140). New York, NY: Humana Press.
- Borrill, P., Harrington, S. A., & Uauy, C. (2019). Applying the latest advances in genomics and phenomics for trait discovery in polyploid wheat. *The Plant Journal*, *97*, 56–72. <https://doi.org/10.1111/tpj.14150>
- Brown, A. A., Buil, A., Viñuela, A., Lappalainen, T., Zheng, H.-F., Richards, J. B., ... Durbin, R. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *eLife*, *3*, e01381. <https://doi.org/10.7554/eLife.01381>
- Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, *16*, 129–132. <https://doi.org/10.2307/1267501>
- Cao, Y., Wei, P., Bailey, M., Kauwe, J. S. K., & Maxwell, T. J. (2014). A versatile omnibus test for detecting mean and variance heterogeneity. *Genetic Epidemiology*, *38*, 51–59
- Chen, S. L., & Kao, C. H. (1995). Cd induced changes in proline level and peroxidase activity in roots of rice seedlings. *Plant Growth Regulation*, *17*, 67–71. <https://doi.org/10.1007/BF00024497>
- Corty, R. W., Kumar, V., Tarantino, L. M., Takahashi, J. S., & Valdar, W. (2018). Mean-variance QTL mapping identifies novel QTL for circadian activity and exploratory behavior in mice. *G3: Genes, Genomes, Genetics*, *8*. <https://doi.org/10.1534/g3.118.200194>
- Ding, Y., Gong, S., Wang, Y., Wang, F., Bao, H., Sun, J., ... Zhu, C. (2018). MicroRNA166 modulates cadmium tolerance and accumulation in rice. *Plant Physiology*, *177*, 1691–1703. <https://doi.org/10.1104/pp.18.00485>
- Dumitrascu, B., Darnell, G., Ayroles, J., & Engelhardt, B. E. (2019). Statistical tests for detecting variance effects in quantitative trait studies. *Bioinformatics*, *35*, 200–210. <https://doi.org/10.1093/bioinformatics/bty565>
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome*, *4*, 250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Forsberg, S. K. G., Andreatta, M. E., Huang, X.-Y., Danku, J., Salt, D. E., & Carlborg, Ö. (2015). The multi-allelic genetic architecture of a variance-heterogeneity locus for molybdenum concentration in leaves acts as a source of unexplained additive genetic variance. *PLOS Genetics*, *11*, e1005648. <https://doi.org/10.1371/journal.pgen.1005648>
- Forsberg, S. K. G., & Carlborg, Ö. (2017). On the relationship between epistasis and genetic variance heterogeneity. *Journal of Experimental Botany*, *68*, 5431–5438. <https://doi.org/10.1093/jxb/erx283>
- Glover, N. M., Redestig, H., & Dessimoz, C. (2016). Homoeologs: What are they and how do we infer them? *Trends in Plant Science*, *21*, 609–621. <https://doi.org/10.1016/j.tplants.2016.02.005>
- Guttieri, M. J., Baenziger, P. S., Frels, K., Carver, B., Arnall, B., Wang, S., ... Waters, B. M. (2015a). Prospects for selecting wheat with increased zinc and decreased cadmium concentration in drain. *Crop Science*, *55*, 1712–1728. <https://doi.org/10.2135/cropsci2014.08.0559>
- Guttieri, M. J., Baenziger, P. S., Frels, K., Carver, B., Arnall, B., & Waters, B. M. (2015b). Variation for grain mineral concentration in a diversity panel of current and historical Great Plains hard winter wheat germplasm. *Crop Science*, *55*, 1035–1052. <https://doi.org/10.2135/cropsci2014.07.0506>
- Hoffman, G. E. (2013). Correcting for population structure and kinship using the linear mixed model: Theory and extensions. *Plos One*, *8*, e75707. <https://doi.org/10.1371/journal.pone.0075707>
- Huang, X., & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, *65*, 531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Ingvarsson, P. K., & Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytologist*, *189*, 909–922. <https://doi.org/10.1111/j.1469-8137.2010.03593.x>
- International Wheat Genome Sequencing Consortium (IWGSC), Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N. et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, *361*. <https://doi.org/10.1126/science.aar7191>
- Kim, W., Johnson, J. W., Graybosch, R. A., & Gaines, C. S. (2003). Physicochemical properties and end-use quality of wheat starch as a function of waxy protein alleles. *Journal of Cereal Science*, *37*, 195–204. <https://doi.org/10.1006/jcrs.2002.0494>
- Kumari, M., Taylor, G. J., & Deyholos, M. K. (2008). Transcriptional responses to aluminum stress in roots of *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, *279*, 339. <https://doi.org/10.1007/s00438-007-0316-z>
- Kusmec, A., Srinivasan, S., Nettleton, D., & Schnable, P. S. (2017). Distinct genetic architectures for phenotype means and plasticities in *Zea mays*. *Nature Plants*, *3*, 715–723. <https://doi.org/10.1038/s41477-017-0007-7>
- Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 619–678
- Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*, *95*, 221–227. <https://doi.org/10.1038/sj.hdy.6800717>
- Mackay, T. F. C., & Lyman, R. F. (2005). Drosophila bristles and the nature of quantitative genetic variation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *360*, 1513–1527. <https://doi.org/10.1098/rstb.2005.1672>
- Menke, A., Muntner, P., Silbergeld, E. K., Platz, E. A., & Guallar, E. (2009). Cadmium levels in urine and mortality among U.S. adults. *Environmental Health Perspectives*, *117*, 190–196. <https://doi.org/10.1289/ehp.11236>
- Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 5268–5273. <https://doi.org/10.1073/pnas.0431157100>
- Nelson, R. M., Pettersson, M. E., Li, X., & Carlborg, Ö. (2013). Variance heterogeneity in *Saccharomyces cerevisiae* expression data: Trans-regulation and epistasis. *Plos One*, *8*, e79507. <https://doi.org/10.1371/journal.pone.0079507>
- Nordborg, M., & Weigel, D. (2008). Next-generation genetics in plants. *Nature*, *456*, 720–723. <https://doi.org/10.1038/nature07629>
- Paré, G., Cook, N. R., Ridker, P. M., & Chasman, D. I. (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: A report from the Women's Genome Health Study. *PLoS Genetics*, *6*, e1000981. <https://doi.org/10.1371/journal.pgen.1000981>



- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *Plos Genetics*, 2, e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Perry, G. M. L., Nehrke, K. W., Bushinsky, D. A., Reid, R., Lewandowski, K. L., Hueber, P., & Scheinman, S. J. (2012). Sex modifies genetic effects on residual variance in urinary calcium excretion in rat (*Rattus norvegicus*). *Genetics*, 191, 1003–1013. <https://doi.org/10.1534/genetics.112.138909>
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11, 459–463. <https://doi.org/10.1038/nrg2813>
- Queitsch, C., Sangster, T. A., & Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. *Nature*, 417, 618–624. <https://doi.org/10.1038/nature749>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rascio, N., & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, 180, 169–181. <https://doi.org/10.1016/j.plantsci.2010.08.016>
- Rönnegård, L., & Carlborg, Ö. (2007). Separation of base allele and sampling term effects gives new insights in variance component QTL analysis. *BMC Genetics*, 8, 1. <https://doi.org/10.1186/1471-2156-8-1>
- Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010a). Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, 42, 8. <https://doi.org/10.1186/1297-9686-42-8>
- Rönnegård, L., Shen, X., & Alam, M. (2010b). hglm: A package for fitting hierarchical generalized linear models. *The R Journal*, 2, 20. Retrieved from <https://doi.org/10.32614/RJ-2010-009>
- Rönnegård, L., & Valdar, W. (2011). Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics*, 188, 435–447. <https://doi.org/10.1534/genetics.111.127068>
- Rönnegård, L., & Valdar, W. (2012). Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genetics*, 13, 63. <https://doi.org/10.1186/1471-2156-13-63>
- Salomé, P. A., Bomblies, K., Laitinen, R. A. E., Yant, L., Mott, R., & Weigel, D. (2011). Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. *Genetics*, 188, 421–433. <https://doi.org/10.1534/genetics.111.126607>
- Santantonio, N., Jannink, J.-L., & Sorrells, M. (2019). Homeologous epistasis in wheat: The search for an immortal hybrid. *Genetics*, 211, 1105–1122. <https://doi.org/10.1534/genetics.118.301851>
- Sell-Kubiak, E., Duijvesteijn, N., Lopes, M. S., Janss, L. L. G., Knol, E. F., Bijma, P., & Mulder, H. A. (2015). Genome-wide association study reveals novel loci for litter size and its variability in a large white pig population. *Bmc Genomics [Electronic Resource]*, 16, 1049. <https://doi.org/10.1186/s12864-015-2273-y>
- Shen, X., Pettersson, M., Rönnegård, L., & Carlborg, Ö. (2012). Inheritance beyond plain heritability: Variance-controlling genes in *Arabidopsis thaliana*. *PLOS Genetics*, 8, e1002839. <https://doi.org/10.1371/journal.pgen.1002839>
- Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51, 47–60.
- Struchalin, M. V., Amin, N., Eilers, P. H., van Duijn, C. M., & Aulchenko, Y. S. (2012). An R package “VariABEL” for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity. *BMC Genetics*, 13, 4. <https://doi.org/10.1186/1471-2156-13-4>
- Struchalin, M. V., Dehghan, A., Wittelman, J. C., van Duijn, C., & Aulchenko, Y. S. (2010). Variance heterogeneity analysis for detection of potentially interacting genetic loci: Method and its limitations. *BMC Genetics*, 11, 92. <https://doi.org/10.1186/1471-2156-11-92>
- Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *Plos Genetics*, 14, e1007309. <https://doi.org/10.1371/journal.pgen.1007309>
- Tan, Q., Hjelmborg, J. V. B., Thomassen, M., Jensen, A. K., Christiansen, L., Christensen, K., ... Kruse, T. A. (2014). Hierarchical linear modeling of longitudinal pedigree data for genetic association analysis. *BioMed Central*, 8.
- Topless, R. K., Flynn, T. J., Cadzow, M., Stamp, L. K., Dalbeth, N., Black, M. A., & Merriman, T. R. (2015). Association of SLC2A9 genotype with phenotypic variability of serum urate in pre-menopausal women. *Frontiers in Genetics*, 6. <https://doi.org/10.3389/fgene.2015.00313>
- van de Mortel, J. E., Schat, H., Moerland, P. D., Ver Loren van Themaat, E., van der Ent, S., Blankestijn, H., ... Aarts, M. G. M. (2008). Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in *Arabidopsis thaliana* and the related Zn/Cd-hyperaccumulator *Thlaspi caerulescens*. *Plant, Cell and Environment*, 31, 301–324. <https://doi.org/10.1111/j.1365-3040.2007.01764.x>
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wang, G., Yang, E., Brinkmeyer-Langford, C. L., & Cai, J. J. (2014a). Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics*, 196, 413–425. <https://doi.org/10.1534/genetics.113.157503>
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E. et al. (2014b). 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, W., Pan, Q., Tian, B., He, F., Chen, Y., Bai, G., ... Akhunov, E. (2019). Gene editing of the wheat homologs of TONNEAU1-recruiting motif encoding gene affects grain shape and weight in wheat. *The Plant Journal*, 100, 251–264. <https://doi.org/10.1111/tpj.14440>
- Wang, W., Simmonds, J., Pan, Q., Davidson, D., He, F., Battal, A., ... 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>
- Wei, W.-H., Bowes, J., Plant, D., Viatte, S., Yarwood, A., Massey, J., ... Eyre, S. (2016). Major histocompatibility complex harbors widespread genotypic variability of non-additive risk of rheumatoid arthritis including epistasis. *Scientific Reports*, 6, 25014. <https://doi.org/10.1038/srep25014>
- Wei, W.-H., Viatte, S., Merriman, T. R., Barton, A., & Worthington, J. (2017). Genotypic variability based association identifies novel non-additive loci DHCR7 and IRF4 in seronegative rheumatoid arthritis. *Scientific Reports*, 7, 5261. <https://doi.org/10.1038/s41598-017-05447-1>

- Whitt, L., Ricachenevsky, F. K., Ziegler, G., Clemens, S., Walker, E., Maathuis, F., ... Baxter, I. (2018). A curated list of genes that control elemental accumulation in plants. *bioRxiv* 456384. <https://doi.org/10.1101/456384>
- Xiao, Y., Liu, H., Wu, L., Warburton, M., & Yan, J. (2017). Genome-wide association studies in maize: Praise and stargaze. *Molecular Plant*, *10*, 359–374. <https://doi.org/10.1016/j.molp.2016.12.008>
- Yang, J., Loos, R. J. F., Powell, J. E., Medland, S. E., Speliotes, E. K., Chasman, D. I. et al. (2012). *FTO* genotype is associated with phenotypic variability of body mass index. *Nature*, *490*, 267–272. <https://doi.org/10.1038/nature11401>
- Young, A. I., Wauthier, F. L., & Donnelly, P. (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nature Genetics*, *50*, 1608. <https://doi.org/10.1038/s41588-018-0225-6>
- Yu, J., Wu, L., Fu, L., Shen, Q., Kuang, L., Wu, D., & Zhang, G. (2019). Genotypic difference of cadmium tolerance and the associated microRNAs in wild and cultivated barley. *Plant Growth Regulation*, *87*, 389–401. <https://doi.org/10.1007/s10725-019-00479-1>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (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 online in the Supporting Information section at the end of the article.

**How to cite this article:** Hussain W, Campbell MT, Jarquin D, Walia H, Morota G. Variance heterogeneity genome-wide mapping for cadmium in bread wheat reveals novel genomic loci and epistatic interactions. *Plant Genome*. 2020;13:e20011. <https://doi.org/10.1002/tpg2.20011>