the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2024) **117**, 713–728 doi: 10.1111/tpj.16524

Integrated multi-omics analyses and genome-wide association studies reveal prime candidate genes of metabolic and vegetative growth variation in canola

Dominic Knoch[1,*], Rhonda C. Meyer[1], Marc C. Heuermann[1], David Riewe[1,2], Fritz F. Peleke[1], Jędrzej Szymański[1,3], Amine Abbadi[4,5], Rod J. Snowdon[6] and Thomas Altmann[1]

[1]*Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466, Corrensstrasse 3, Seeland OT, Gatersleben, Germany,*

[2]*Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Institute for Ecological Chemistry, Plant Analysis and Stored Product Protection, 14195 Berlin, Germany,*

[3]*Institute of Bio- and Geosciences IBG-4: Bioinformatics, Forschungszentrum Jülich, 52428 Jülich, Germany,*

[4]*NPZ Innovation GmbH, Hohenlieth, 24363 Holtsee, Germany,*

[5]*Norddeutsche Pflanzenzucht Hans-Georg Lembke KG, Hohenlieth, 24363 Holtsee, Germany, and*

[6]*Department of Plant Breeding, Research Centre for Biosystems, Land Use and Nutrition (iFZ), Justus-Liebig-University Giessen, 35392 Giessen, Germany*

Received 1 February 2023; revised 17 October 2023; accepted 23 October 2023; published online 15 November 2023.
*For correspondence (e-mail knochd@ipk-gatersleben.de).

SUMMARY

Genome-wide association studies (GWAS) identified thousands of genetic loci associated with complex plant traits, including many traits of agronomical importance. However, functional interpretation of GWAS results remains challenging because of large candidate regions due to linkage disequilibrium. High-throughput omics technologies, such as genomics, transcriptomics, proteomics and metabolomics open new avenues for integrative systems biological analyses and help to nominate systems information supported (prime) candidate genes. In the present study, we capitalise on a diverse canola population with 477 spring-type lines which was previously analysed by high-throughput phenotyping of growth-related traits and by RNA sequencing and metabolite profiling for multi-omics-based hybrid performance prediction. We deepened the phenotypic data analysis, now providing 123 time-resolved image-based traits, to gain insight into the complex relations during early vegetative growth and reanalysed the transcriptome data based on the latest Darmor-*bzh* v10 genome assembly. Genome-wide association testing revealed 61 298 robust quantitative trait loci (QTL) including 187 metabolite QTL, 56814 expression QTL and 4297 phenotypic QTL, many clustered in pronounced hotspots. Combining information about QTL colocalisation across omics layers and correlations between omics features allowed us to discover prime candidate genes for metabolic and vegetative growth variation. Prioritised candidate genes for early biomass accumulation include A06p05760.1_BnaDAR (PIAL1), A10p16280.1_BnaDAR, C07p48260.1_BnaDAR (PRL1) and C07p48510.1_BnaDAR (CLPR4). Moreover, we observed unequal effects of the *Brassica* A and C subgenomes on early biomass production.

Keywords: ***Brassica napus*, biomass, GWAS, high-throughput phenotyping, metabolomics, transcriptomics.**

INTRODUCTION

In recent years, technological advances in high-throughput phenotyping (HTP) platforms enabled high-throughput non-invasive quantification of complex traits in model and crop plants over time at population scale (Chen et al., 2014; Flood et al., 2016; Junker et al., 2015; Langstroff et al., 2022; Scharr et al., 2016; Watt et al., 2020). In combination with ever-increasing genome sequencing data, phenomics enabled efficient identification of genetic determinants of multiple traits that define crop quality and performance (Crossa et al., 2021; Mir et al., 2019; Yang et al., 2020). However, despite the ever-increasing throughput and depth of phenomic analyses, the molecular mechanisms leading to the phenotype emergence remain

mostly unknown. Thus, one of the major goals of modern systems genetics is identification of such mechanisms via use of omics technologies and data science (Choi, 2019; Li & Yan, 2020; Shaw et al., 2021; Shen et al., 2022). In our recent study, we showed the power of HTP in *Brassica napus* (Knoch et al., 2020). While we were able to identify genetic loci associated with biomass and growth-related *B. napus* phenotypes, the molecular mechanisms linking that genetic variation with specific traits remained to be characterised. Here, we re-analysed the previously generated multi-omics data and performed integrative analyses across the multiple data layers to link observed genetic associations with specific genes, metabolites and functional ontology groups thus paving the way towards mechanistic explanation of genotype–phenotype associations.

Early growth and biomass formation are crucial traits for plant productivity and yield. In winter-type (Basunanda et al., 2010) and semi-winter rapeseed (Zhao, Wang, et al., 2016) rapeseed, early biomass and biomass heterosis correlate with seed yield. Vigorous early seedling growth provides efficient ground coverage and avoids competition with weeds in the field. Quantitative traits like seed yield, biomass production or plant height are under complex genetic control and are strongly influenced by the environment (Shi et al., 2009; Zhao, Wang, et al., 2016). Dissecting the genetic basis of such traits is of high relevance to fundamental research and to crop improvement strategies. The ultimate goal for breeders is to identify favourable lines in breeding populations according to their genotypes and, ideally, to stack multiple beneficial alleles for different genes to generate lines with superior performance. Previous studies applied linkage QTL mapping and genome-wide association analyses to identify QTL for plant mineral nutrients and trace elements (Bus et al., 2014), biomass-related traits (Knoch et al., 2020; Körber et al., 2015; Yong et al., 2015), seed yield (Luo et al., 2017; Radoev et al., 2008) and yield-related traits (Cai et al., 2016; Chen et al., 2007; Dong et al., 2018; Yang et al., 2012) in *B. napus*. In some cases, genes underlying these QTL were identified (Zeng et al., 2011; Liu et al., 2015; Li, Jeong, et al., 2018); however, such examples remain the minority. Most previous GWAS studies in rapeseed focused on a limited number of phenotypic traits and/or only a single omics level. Some recent studies integrated at least two omics levels, such as Li et al. (2023) who studied the seed oil content in rapeseed using a marker metabolite-based multi-omics analysis, Li, Yao, et al. (2018) who integrated QTL and eQTL mapping to study fatty acid content, flowering time and growth-related traits, Yu et al. (2018) who coupled QTL and eQTL mapping focusing on the apetalous characteristic or Zhang et al. (2022) who dissected the genetic architecture of seed coat content in *B. napus* using a multi-omics analysis. Schaefer et al. (2018) demonstrated in maize that integrating co-expression networks with GWAS can be a powerful approach to prioritise candidate genes. Only very few studies such as Szymański et al. (2020) or Zhao et al. (2022) analysed larger populations and made use of the potential of the interconnected multiple omics layers.

Here, we investigated a genetically diverse population of 477 spring-type canola lines. We re-analysed and integrated previously generated multi-omics data sets (Knoch et al., 2020, 2021), including image-derived traits (i-traits) based on daily high-throughput phenotyping, gene expression and metabolite profiles from an early vegetative growth stage and used them to perform extensive genome-wide association studies. We identified genomic regions associated with phenotypic variation at the early phase of vegetative plant growth and related traits, studied links between the different omics strata by correlation and colocalisation analyses, selected genetic modules underlying early biomass by a weighted gene correlation network analysis and nominated prime candidate genes for the traits of interest after combining the results of the aforementioned approaches.

RESULTS

High-throughput phenotyping and generation of omics data

High-throughput imaging data from 6 to 27 days after sowing (DAS) were obtained from Knoch et al. (2020). All raw images were subjected to a deepened image analysis and a core set of 123 i-traits, complementary to the four traits analysed in Knoch et al. (2020), was selected for subsequent analyses (Figure 1). Three general i-trait heritability patterns over time could be distinguished (Figure S3a). First, traits displaying overall high heritability like 'projected leaf area', 'estimated biovolume' and 'compactness', second, traits displaying high heritability at early stages and low heritability at later stages, for example 'hull fill grade' or 'brown to green ratio' and third, traits with the inverse pattern, for example 'branch point count' and 'leaf width'. At 14 DAS, when most plants had observable epicotyls, shoot material was sampled for molecular/biochemical analyses. Metabolite profiles were obtained from Knoch et al. (2021), where shoot material was analysed by gas chromatography–mass spectrometry and 154 metabolites, 64 of known and 90 of unknown chemical structure, were quantified. The estimated genomic heritabilities for metabolites were rather low, ranging between 0 and 0.43 (Figure S3b). RNA sequencing reads were also obtained by Knoch et al. (2021). Sequencing was performed using aliquots of the same material as for metabolite profiling. Overall, 83% of the reads could be aligned to the Darmor-bzh v10 reference genome, 63% of them uniquely. In total, 89.172 genes (82% of the 108 190 annotated genes) were detected as expressed (>0 counts in at least one sample). Low-expressed genes were removed and transcripts of 41 380 genes (38%) were used for subsequent

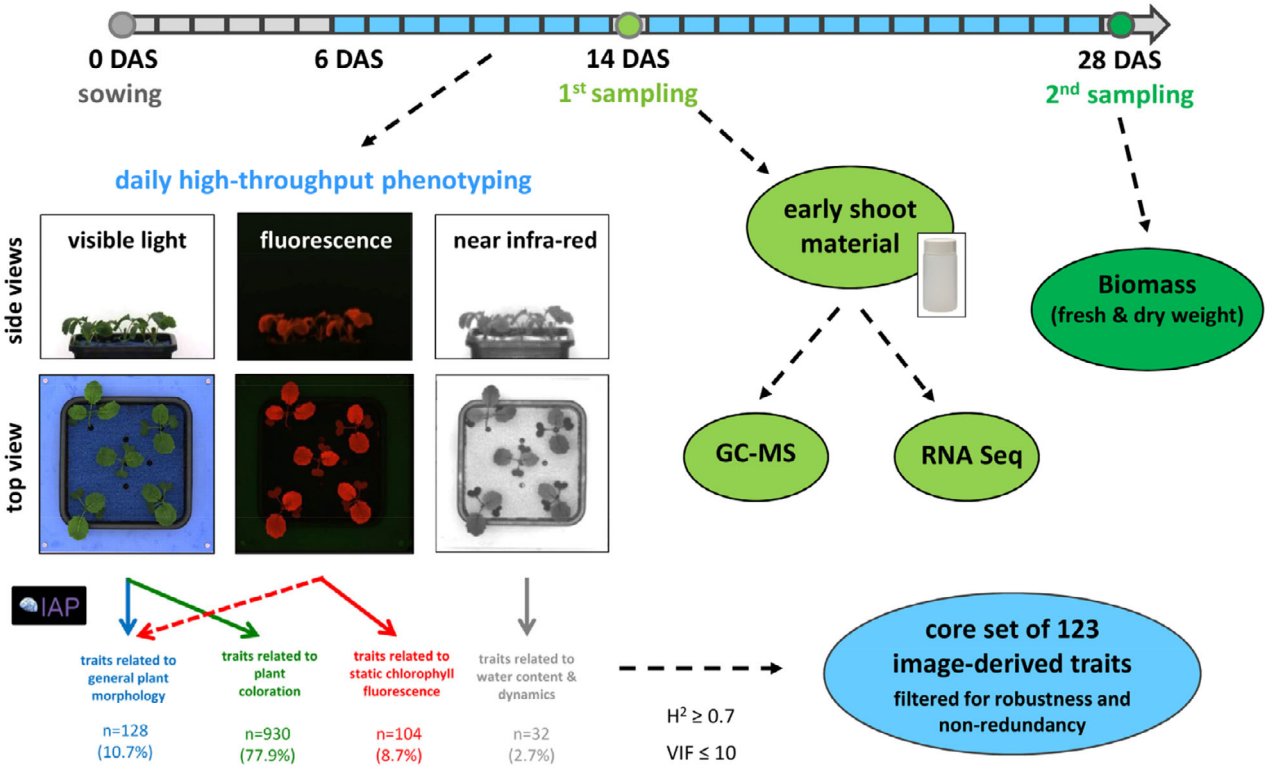

Figure 1. Experimental workflow to generate the phenotyping and omics data.
High-throughput phenotyping was performed daily in the IPK phenotyping platform for large plants. Images between 6 to 27 days after sowing (DAS) were obtained from three different camera systems (VIS, FLUO and NIR) from top and side views with different angles and analysed using IAP Version 2.0.7 (Klukas et al., 2014) and a customised pipeline. A core set of 123 robust image-derived traits (i-traits) was selected by filtering for heritability ($H^2 \geq 0.7$) and multicollinearity using variance inflation factors (VIF ≤10). At 14 DAS four of the nine plants grown within each pot were sampled and deep frozen in liquid nitrogen for metabolite profiling and RNA sequencing analyses. At the end of each experiment at 28 DAS, the remaining five plants per pot were sampled for biomass analysis.

analyses. The estimated genomic heritabilities for transcripts ranged from 0 to 0.76 (Figure S3c). Predictability of phenotypic traits by transcript levels was assessed using random forest (Data S2). The best average predictive performance of 0.39 was observed for the pigmentation-related i-trait 'top.intensity.vis.hsv.s.histogram.v_avg.bin.05.051_063' at 13 DAS. Fresh and dry weight yielded R^2 values of 0.36 and 0.33, respectively.

Correlations between omics features

Pairwise correlations were calculated between all omics features (Data S3). In total, 1385 significant correlations (P-values $_{FDR} \leq 0.05$, $|r| \geq 0.4$) were observed between transcripts and phenotypic traits, 479 between metabolites and transcripts and 22 between metabolites and phenotypic traits (Figure S4). Overall, transcripts and metabolites showed very similar numbers of positive and negative correlations. More negative than positive correlations were found between metabolites and phenotypic traits, while transcripts and phenotypic traits displayed more positive than negative correlations. Focusing on early plant biomass (fresh and dry weight), only weak correlations ($|r| \sim 0.3$) with metabolites were detected, whereas correlations between other phenotypic traits and metabolites were moderate. The highest correlation ($r \sim 0.44$) was found between indole-3-acetonitrile and plant stature (compactness).

Fifteen transcripts showed correlation coefficients $|r| > 0.4$ with fresh weight (Table 1). Among the highest were C06p39650.1_BnaDAR ($r = 0.46$), annotated as 'HAD-superfamily hydrolase' and C06p42580.1_BnaDAR ($r = 0.45$), annotated as 'SRP72 RNA-binding domain-containing protein'. In a complementary approach we identified 480 transcripts with relations to fresh weight using the Boruta algorithm (Data S4). Random forest regression on the top 50 transcripts with the highest mean importance values achieved a mean prediction accuracy of 0.49 for fresh weight (0.53 using all 480 transcripts). Notably, 14 of the 15 highest correlated transcripts were among these top 50.

We grouped a subset of i-traits into four categories combining related traits from different image modalities and days: 'plant height related', 'plant volume-related', 'projected leaf area-related' and 'compactness-related' (Table S1). Phenotypic traits showed substantially higher correlations ($|r|$) with transcripts than with metabolites. A03p39940.1_BnaDAR, annotated as 'Ethylene-responsive transcription factor', was correlated with

Table 1 Highest correlations between transcripts and early plant biomass

Phenotypic trait	Transcripts ($r \geq$ |0.4|)	Pearson's r	Boruta$_{mean\ Imp.}$	Annotation[a]
Biomass (fresh weight)	A02p29820.1_BnaDAR	0.4181	11.62	3-oxoacyl-[acyl-carrier-protein] synthase
	A02p33440.1_BnaDAR	0.4465	12.00	Ubiquitin carboxyl-terminal hydrolase 4/hypothetical protein[b]
	A04p10830.1_BnaDAR	0.4028	14.02	**Membrane-anchored ubiquitin-fold protein**
	A05p28550.1_BnaDAR	−0.4152	16.64	**Ubiquitin carboxyl-terminal hydrolase 12/TRAF-like family protein[b]**
	A06p19110.1_BnaDAR	0.4105	9.54	ATP-dependent DNA helicase
	BNapus_Darmor_BZH_scaffold_38p02360.1_BnaDAR	−0.4106	7.58	Elongation factor 1-beta 1
	C02p03750.1_BnaDAR	−0.4098	12.32	NAC domain protein
	C02p06120.1_BnaDAR	−0.4147	10.74	Peptidylprolyl isomerase
	C03p78480.1_BnaDAR	0.4433	12.34	Protein DETOXIFICATION
	C04p59470.1_BnaDAR	0.434	9.74	**Nucleolin 1**
	C05p43970.1_BnaDAR	−0.4049	13.55	Ubiquitin carboxyl-terminal hydrolase 12/TRAF-like family protein[b]
	C06p29830.1_BnaDAR	0.4291	15.12	Phosphoglycerate kinase
	C06p32710.1_BnaDAR	0.4107	16.57	DNA-3-methyladenine glycosylase, putative
	C06p39650.1_BnaDAR	0.4622	17.05	**T-complex protein 1 subunit alpha/HAD-superfamily hydrolase[b]**
	C06p42580.1_BnaDAR	0.4529	15.50	Signal recognition particle subunit SRP72

Underlined and **bold** text indicates transcripts correlated with multiple biomass-related traits, ($n \geq 2$ and ≥ 3, respectively; see Table S1).
[a]Functional annotation according to Vollrath et al. (2021).
[b]Description obtained by BLASTX to *A. thaliana* – Brassicaceae Database (BRAD; Chen et al., 2022; http://brassicadb.cn/#/Annotations/).

'top.intensity.vis.hsv.s.histogram.v_avg.bin.05.051_063' at 13DAS ($r = 0.50$), 'early biovolume' at 9DAS ($r = 0.43$) and other biomass-related i-traits. Plant height at multiple days was correlated ($r \geq 0.4$) with the expression of several genes annotated as 'xyloglucan endotransglucosylase/hydrolase', 'peroxidase' or 'Laccase' (Table S1).

The highest correlations between transcripts and metabolites of known chemical structure were detected between C01p00680.1_BnaDAR encoding an 'alanine:glyoxylate aminotransferase 2' homolog and beta-alanine ($r = -0.57$), C07p60730.1_BnaDAR encoding an arginine decarboxylase and putrescine ($r = 0.53$), C08p22140.1_BnaDAR, a putative trehalose-6-phosphate synthase and sucrose ($r = -0.5$), A05p41690.1_BnaDAR, annotated as 'dehydroquinate dehydratase, putative / shikimate dehydrogenase' and quinic acid ($r = 0.47$), and C01p48140.1_BnaDAR, annotated as 'Malonate—CoA ligase' and malonic acid ($r = -0.31$). Another interesting candidate is C03p47770.1_BnaDAR, which is substantially negatively correlated with sucrose (Figure S5a; $r = -0.57$). The gene encodes a protein of unknown function with homology to the Arabidopsis *AT3G15630* gene and was annotated with the GO terms 'response to sucrose stimulus' and 'response to fructose stimulus'. Notably, nine lines share a large deletion including of C03p47770.1_BnaDAR on chromosome C03 (Figure S5b). These lines display particularly low of C03p47770.1_BnaDAR expression levels and significantly higher values of glucose, sucrose and fructose than the population average (Figure S5c-e).

Genome-wide association studies and identification of QTL hotspots

The extensive omics data sets provided the opportunity to study the genetic basis of trait variation at different omics levels. 2496 phenotypic traits (123 i-traits at 21 days, growth rates and biomass), 154 metabolites and expression data of 41 380 genes were subjected to GWAS and the output filtered using three criteria: (1) P-value$_{FDR} \leq 0.05$, (2) PVE% $\geq$ 2% and (3) stability of associations using different numbers of PCs as cofactors. In total, 61 298 robust QTL matched these criteria (Table 2; Data S5), 4297 QTL for phenotypic traits, 187 mQTL detected for metabolites and 56 814 eQTL for gene transcript levels. For 1912 (77%) of the phenotypic traits, 29 027 (70%) of the transcripts and 89 (58%) of the metabolites, at least one association was detected.

Table 2 QTL and co-localisations between omics layers

Data set	Number of traits	MTAs filtered for *P*-value $_{FDR} \leq 0.05$	MTAs filtered for *P*-value $_{FDR} \leq 0.05$, $\geq$ 2 PVE%[d] and robustness[e]
Metabolites (M)	n = 154	n = 544	n = 187
Transcripts (T)[a]	n = 41 380	n = 302 226	n = 56 814
Phenotypic traits (P)[b]	n = 2496	n = 28 758	n = 4297
		Number of co-localisations	**Same genetic marker associated**
Co-localisations (T, M)		n = 110	n = 131
Co-localisations (T, P)		n = 366	n = 1484
Co-localisations (M, P)		n = 42	n = 28
Co-localisations (T, M, P)		n = 42	n = 24
Permutation threshold[c]			n = 8

[a]Low-expressed genes were filtered (see Material and Methods).
[b]Including image-derived traits for 21 time points (6–27 DAS), biomass and relative growth / absolute change rates.
[c]Estimated number of random co-localisations obtained by 10 000 permutations.
[d]Estimated percentage of phenotypic variance explained by the genetic marker.
[e]GWAS was repeated with different PC covariates and associations found in just one model were omitted.

Trans-eQTL (n = 282 932; $n_{filtered}$ = 40 117) were much more frequent than cis-eQTL (n = 19 294; $n_{filtered}$ = 16 697), but cis-eQTL explained on average a higher proportion of PVE (Figure 2). The highest number of eQTL was detected on chromosome C03 (n = 5949), the lowest number on chromosome A08 (n = 1757). For transcripts with multiple eQTL, the most significant one often corresponds to the gene locus itself, as illustrated by C06p42580.1_BnaDAR (Figure 2c). Binning eQTL in overlapping intervals (±500 kb) showed an unequal distribution across the 19 chromosomes. Some regions were depleted of eQTL while others showed accumulations (Figure 3; Figure S6a). In total, 96 hotspots with more than 200 eQTL were detected. Significant GO term enrichments were observed for 49 hotspots (Data S6). The biggest eQTL hotspot was detected on chromosome A05 (33–34 Mb; Figure 3). This hotspot mainly comprised trans-eQTL, suggesting pleiotropic effects on the expression of multiple genes rather than a clustering caused by high gene density. The significant GO term enrichment related to RNA methylation, ribosome and mitochondrial respiratory chain provides support for a major (transcriptional) regulator. Besides multiple ribosomal protein genes, the region contains A05p31230.1_BnaDAR, annotated as elongation factor.

Hotspots were also detected for metabolite QTL and phenotypic QTL, but positions differed. For phenotypic traits, a hotspot of 190 QTL was observed on chromosome C02 (Figure S6b), while mQTL hotspots were observed on chromosomes A01, C06 and C08 (Figure S6c).

QTL colocalisation across the omics layers

Under the hypothesis that functionally related features of different omics layers should be affected by common loci, QTL colocalisation was investigated. Using a stringent approach, 24 associations with identical markers were detected across all three omics layers, substantially more than expected by chance (n = 8; Table 2; Figure S7). In a second interval-based approach, we identified a total of 110 regions with eQTL-mQTL co-localisations. Twenty-two of these regions harboured features with Pearson correlations ($r \geq |0.3|$) and 16 of them involved metabolites of known chemical structure (Data S7).

For transcripts and phenotypic traits, 366 colocalisation regions were detected, 114 of them with correlated features (Data S7). Notably, 15 regions overlap with eQTL hotspots and contain correlated features ($r \geq |0.4|$). Focusing on vegetative biomass production as the main trait of interest, colocalised regions were further prioritised. QTL were filtered for phenotypic traits significantly correlated with biomass ($|r| \geq 0.3$), eQTL were restricted to cis-eQTL only and filtered by transcript-phenotype correlations ($|r| \geq 0.4$) to detect the most promising candidates. In total, 11 regions passed these filters (Table S2).

For metabolites and phenotypic traits, 42 co-localisations were detected, which all also colocalised with at least one eQTL (Data S7). However, no promising links between metabolites and phenotypic traits were identified. Some co-localisations are associated with deletions and several dozens of eQTL, for example candidate region 'coloc_n3_eQTL_mQTL_QTL_17' on chromosome A09 with 82 eQTL, candidate region 'coloc_n3_eQTL_mQTL_QTL_27' on chromosome C03 with 97 eQTL or candidate region 'coloc_n3_eQTL_mQTL_QTL_36' on chromosome C05 with 129 eQTL. The latter region colocalises with an mQTL for quinic acid, a quinic acid derivative and twelve phenotypic QTL. While there are no substantial correlations with phenotypic traits, the expression of two genes, A05p41690.1_BnaDAR (r = 0.47) and its homeologue, C05p60900.1_BnaDAR (r = 0.46), was correlated with quinic acid. Both genes are annotated as 'dehydroquinate dehydratase, putative / shikimate dehydrogenase'. For A05p41690.1_BnaDAR, a trans-eQTL was detected on C05 (marker: Bn-scaff_23186_1-

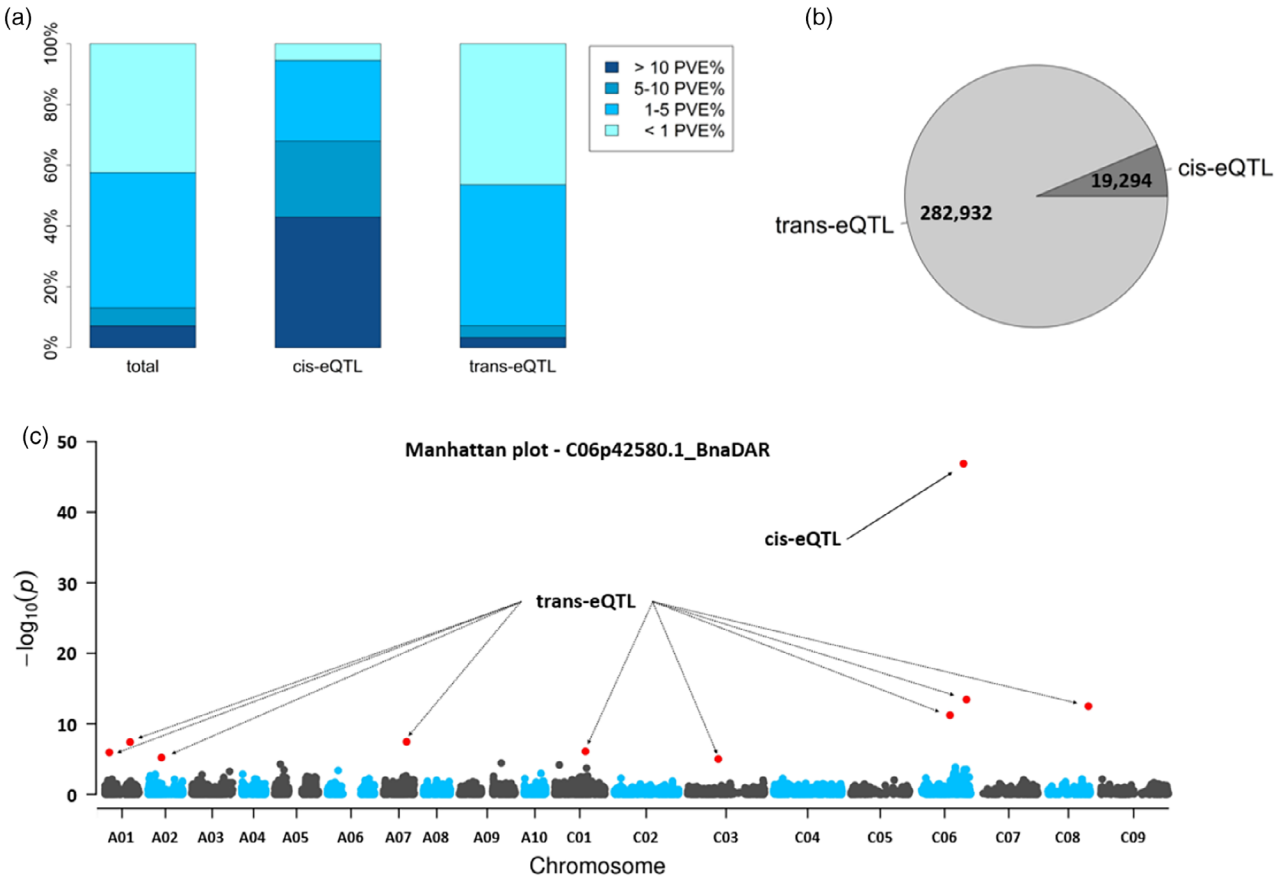

Figure 2. Number and explained phenotypic variance of cis- and trans-eQTL.
Panel (a) shows an overview of detected expression QTL (eQTL) and explained phenotypic variance (PVE%). The leftmost bar shows the distribution of PVE% for all eQTL with a *P*-value $_{FDR}$ ≤0.05. The blue colour code refers to eQTL grouped by: > 10%, 5–10%, 1–5% and <1% PVE. The two other bars group the eQTL in cis- and trans-eQTL, defined by a ± 500 kb interval around the transcription start site of the respective gene. QTL without explainable PVE% were omitted for this representation.
Panel (b) shows the number and proportion of eQTL classified as cis-eQTL and trans-eQTL. Initially, 19 294 cis-eQTL and 282 932 trans-eQTL were identified, of which 16 697 cis-eQTL and 40 117 trans-eQTL with PVE% ≥ 2 were considered for further analyses.
Panel (c) shows exemplarily the Manhattan plot of C06p42580.1_BnaDAR annotated as 'Signal recognition particle subunit SRP72'. Ten significant marker-trait associations (MTAs) on chromosomes A01, A02, A07, C01, C03, C06 and C08 are indicated by red dots (*P*-value $_{FDR}$ ≤0.05).

p121500; at 54.3 Mb) as part of candidate region 'coloc_n3_eQTL_mQTL_QTL_36' (Figure 4). The cis-eQTL of C05p60900.1_BnaDAR (marker: Bn-scaff_20270_1-p1323200_del; at 54.7 Mb) was not considered colocalised because the associated markers were not in LD.

Subgenome-specific expression and biomass accumulation

A PCA evaluated the transcriptome contributions to biomass on the population level. A clustering of lines in the PC3 and PC9 corresponds to breeding pools (Figure 5a). Two partially overlapping clusters, with higher or lower biomass were separated by PC3 (Figure 5b). To further explore this observation, the 5% quantile of top positive (direction of higher biomass; $n = 2069$) and negative (direction of lower biomass; $n = 2069$) ranked loadings of PC3 were separately subjected to GO term enrichment (Data S6). 'chloroplast stroma' (GO:0009570) in CC, and 'DNA helicase activity' (GO:0003678) in MF were the most significant terms for the positive loadings. For the negative loadings, cytosolic large ribosomal subunit' (GO:0022625) in CC, 'structural constituent of ribosome' (GO:0003735) in MF and 'RNA methylation' (GO:0001510) and 'SRP-dependent cotranslational protein targeting to membrane' (GO:0006614) in BP were the most significantly enriched terms. Furthermore, subgenome contributions differed between the top negative and positive loadings. For negative loadings, 814 transcripts were contributed from the A subgenome, 1206 transcripts from the C subgenome and 49 from unplaced scaffolds, respectively. In contrast, the top positive loadings contained 1084 transcripts from the A subgenome, 937 transcripts from the C subgenome and 48 from unplaced scaffolds, respectively. For both negative and positive loadings, these numbers differ significantly from the expected ratio of the quantified transcripts (A subgenome: C subgenome = 0.48: 0.52; Fisher's exact test *P*-value = 5.14e-12 and 1.48e-06, respectively).

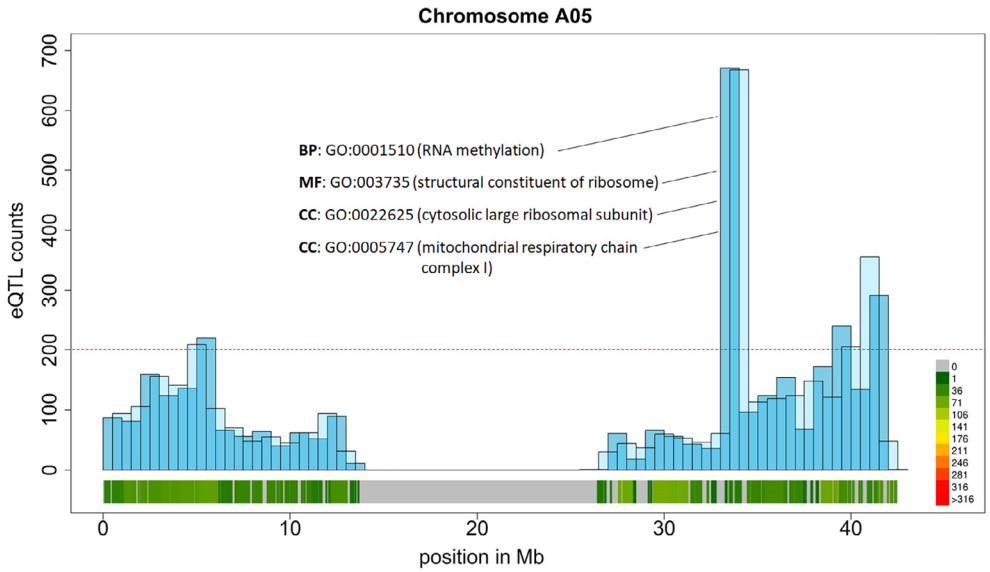

Figure 3. eQTL hotspots detected on the *Brassica napus* chromosome A05.
Expression QTL (eQTL) distribution on the *B. napus* chromosomes A05. The eQTL were binned into overlapping 1 Mb intervals (light blue and dark blue colour, for example 0–1 Mb and 0.5–1.5 Mb, respectively) for representation. The number of QTL per bin is indicated on the vertical axis, the chromosomal position in Mb is shown on the horizontal axis. The bottom section shows the marker distribution, SNPs (n = 911) and CNVs (n = 273), across the chromosome: markers were binned in 1 Mb intervals and marker density is indicated by the colour (green to red). Grey colour indicates regions without genetic markers. Four eQTL hotspots with more than 200 eQTL (surpassing the dashed red line) were detected on A05. Only for one of them (33–34 Mb / 33.5–34.5 Mb, respectively) the genes affected by the eQTL significantly enriched GO terms (BP: biological process, MF: molecular function, CC: cellular component).

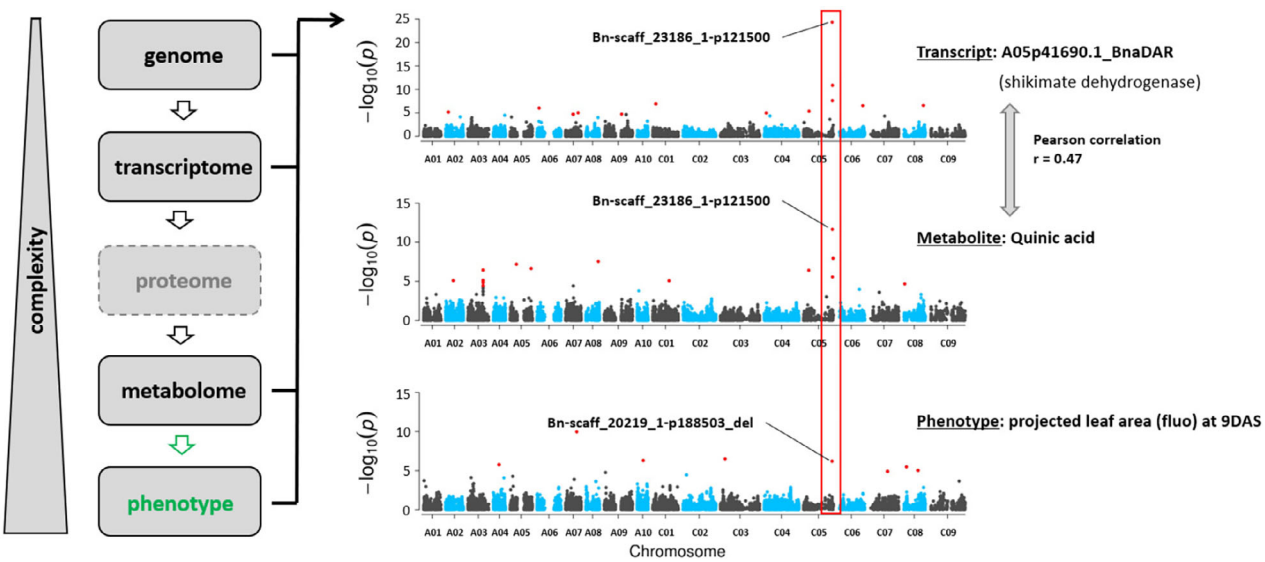

Figure 4. QTL co-localisations across multiple omics layers.
Shown is 'candidate region 36' as an example for QTL colocalisation across multiple omics layers, including the transcriptome, metabolome and phenome. The region on chromosome C05 spans approx. 800 kb and contains 4 mQTL, for quinic acid, 5-caffeoyl-trans-quinic acid and a metabolite of unknown chemical structure, 129 eQTL and 12 phenotypic QTL. Manhattan plots for the transcript of A05p41690.1_BnaDAR annotated as putative shikimate dehydrogenase (top), the metabolite quinic acid (middle) and the i-trait projected leaf area (fluo) at 9 DAS (bottom) are shown. A05p41690.1_BnaDAR and quinic acid were found to be associated with the same genetic marker (Bn-scaff_23186_1-p121500; at 54.3 Mb). Notably, a Pearson correlation of r = 0.47 was detected between the two traits. The i-trait was associated with the deletion marker Bn-scaff_20219_1-p188503_del (at 53.7 Mb; LD >0.8 between the two markers).

In a complementary approach, a weighted gene co-expression network analysis (*WGCNA*) detected 92 modules and pinpointed potential regulatory genes associated with early biomass. The 'red module' (Figure S8a), comprising 621 genes, was significantly correlated (r = 0.50) to fresh weight. An overwhelming majority of 89.5% (n = 574) genes were shared with the top PC3 loadings. A GO term enrichment indicated CC: 'proton-

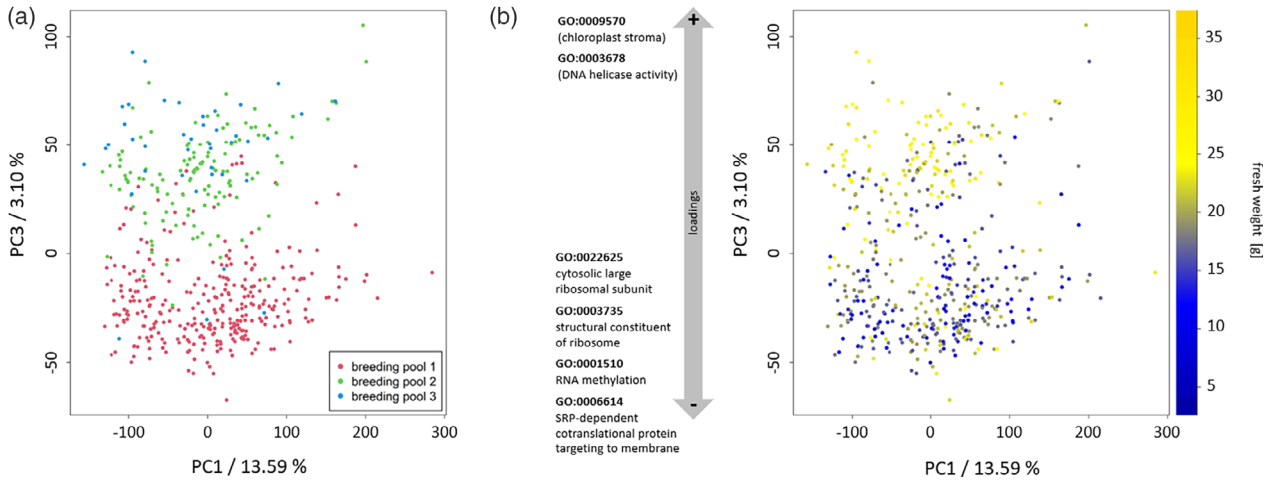

Figure 5. Transcript profiles separate lines according to biomass and breeding pools.
Principal component analysis was performed on filtered transcript (tpm) data for all 477 rapeseed lines after removing low-expressed features. Transcript data were centred and scaled (z-scores). The PCA calculation was done by singular value decomposition (svd) of the data matrix. The first four PCs explained 13.59%, 9.30%, 3.10% and 2.70% of variance, respectively.
(a) Scatter plot of PC1 and PC3 with samples coloured according to their affiliation to one of the breeding pools. (b) The same PCA plot with genotypes coloured according to their biomass (fresh weight BLUEs) using a gradual scale (colour gradient blue, low biomass to yellow, high biomass).

transporting V-type ATPase, V0 domain' (GO:0033179), MF: 'DNA helicase activity' (GO:0003678) and BP: 'telomere maintenance' (GO:0000723) as the most significantly enriched terms (Data S6). Several of these genes, including C06p39650.1_
BnaDAR, C06p42580.1_BnaDAR, A02p33440.1_BnaDAR, C03p78480.1_BnaDAR and C04p59470.1_BnaDAR were also found among the transcripts with the highest correlations to biomass and related i-traits. Some of the highest ranked genes of the 'red module' display a breeding pool specific expression pattern (Figure S8b). In contrast, the 'grey module', containing 10 376 genes which have not been clustered in any module by the WGCNA analysis, was negatively correlated to fresh weight ($r = -0.49$). For this module, many GO terms were found to be significantly enriched (Data S6), among the highest: CC: 'integral component of membrane' (GO:0016021), MF: 'ubiquitin-protein transferase activity' (GO:0004842), BP: 'respiratory burst involved in defense response' (GO:0002679) and 'protein ubiquitination' (GO:0016567).

Comparison of high biomass and low biomass lines

While genotypes in the main RNA-Seq experiment were analysed as pools, four genotypes with differing biomass were additionally sequenced with three biological replicates (Figure S9). In total, 1672 genes (1.55%) were higher expressed in the high biomass line 'Pol 419', 907 derived from the A subgenome (0.84%), 730 from the C subgenome (0.67%) and 35 genes (0.03%) from unplaced scaffolds, respectively (Data S8). In the low biomass line 'Pol 229', 1925 genes were higher expressed with 831 attributed to the A subgenome (0.77%), 1032 to the C subgenome (0.95%) and 62 (0.06%) genes from unplaced scaffolds, respectively. Thus, 'Pol 229' displayed an approximately even distribution between the subgenomes, more similar to the overall ratio of transcripts encoded by the A and C subgenomes, while 'Pol 419', the high biomass line, had substantially more genes from the A than the C subgenome significantly higher expressed.

GO term enrichment was assessed for the 3597 DEGs between the contrasting lines. Both lines shared GO terms, for example BP: 'translation' (GO:0006412), 'rRNA processing' (GO:0006364), 'glucose catabolic process' (GO:0006007), '-pentose-phosphate shunt' (GO:0006098), CC: 'chloroplast stroma' (GO:0009570), 'chloroplast envelope' (GO:0009941), 'cell wall' (GO:0005618), 'chloroplast thylakoid membrane' (GO:0009535) and 'cytosolic ribosome' (GO:0022626), but also displayed enrichment for separate terms. For Pol 229 the GO terms MF: 'structural constituent of ribosome' (GO:0003735), BP: 'RNA methylation' (GO:0001510), 'cell wall modification' (GO:0042545), CC: 'nucleolus' (GO:0005730), 'cytosolic large ribosomal subunit' (GO:0022625) and 'cytosolic small ribosomal subunit' (GO:0022627) were most significantly enriched. For Pol 419: MF: 'pullulanase activity' (GO:0051060), BP: 'starch biosynthetic process' (GO:0019252) and CC: 'apoplast' (GO:0048046) were identified as most significantly enriched (Data S6).

DISCUSSION

A diverse breeding population of 477 canola lines was analysed by high-throughput phenotyping and by measuring transcript and metabolite profiles. The central goal was to gain insight into the genetic factors controlling early biomass accumulation, a crucial trait for plant productivity

(Basunanda et al., 2010; Zhao, Jiang, et al., 2016) and to use multi-omics analyses to discover prime candidate genes for metabolic and vegetative growth variation. The determined i-traits displayed varying patterns of heritability over time as also observed by Chen et al. (2014) and Flood et al. (2016). Temporal fluctuations of heritability may be a consequence of changes in the magnitude of G and E effects (Visscher et al., 2008) or result from technical and/or environmental bias or difficulties to correctly estimate certain parameters, for instance leaf number at very early stages.

Our workflow of data integration has been summarised in the flowchart (Data S9). As a first step, we calculated pairwise correlations between molecular features and i-traits. The majority of metabolites showed little correlation with biomass and growth-related traits, but indole-3-acetonitrile (IAN), an inactive precursor of the phytohormone auxin (IAA) (Korasick et al., 2013), was highly correlated with compactness-related i-traits. Production of IAA from IAN by nitrilases (Shaw et al., 2022) might affect leaf expansion growth and thus compactness. C04p59470.1_BnaDAR, annotated as 'Nucleolin 1' involved in pre-rRNA processing and ribosome assembly (Kojima et al., 2007; Petricka & Nelson, 2007), BNapus_Darmor_BZH_scaffold_38p02360.1_BnaDAR annotated as 'Elongation factor 1-beta 1' and C06p42580.1_BnaDAR annotated as 'Signal recognition particle subunit SRP72', involved in translation and targeting of proteins to the endoplasmic reticulum, showed substantial correlations with fresh weight, indicating a potential contribution to differential growth. Several ubiquitin carboxyl-terminal hydrolases were correlated with biomass. These enzymes hydrolyse the peptide bond at the C-terminal Gly of ubiquitin and are involved in protein deubiquitylation (Hayama et al., 2019). Two of the genes, A05p28550.1_BnaDAR and C05p43970.1_BnaDAR show homology to TRAF-like family proteins (Qi et al., 2022; Teaster et al., 2012). The Arabidopsis homologue AT3G20370 is a potential floral repressor (Schmid et al., 2003), which might affect biomass by delaying flowering. Correlations with plant height identified A01p07850.1_BnaDAR, A03p59730.1_BnaDAR, C01p08760.1_BnaDAR and C07p57790.1_BnaDAR, all annotated as putative 'xyloglucan endotransglucosylase/hydrolase'. Overexpression of the Arabidopsis homologue ATXTH20 affects growth and cell wall mechanics (Miedes et al., 2013). Also, C07p07020.1_BnaDAR, a 'peroxidase', C09p71310.1_BnaDAR, a 'Laccase' and C03p28650.1_BnaDAR, a homologue of the Arabidopsis EXTENSIN 21, are promising candidates affecting cell differentiation, cell wall assembly, growth and lignification (Barros et al., 2015; Yi Chou et al., 2018).

In a second step, the extensive omics data were used for GWAS and colocalisation analyses across omics layers. 96 eQTL hotspots were detected of which 49 show Go term enrichments. A hotspot on chromosome A04 was associated with systemic acquired resistance and salicylic acid mediated signalling. Within this confidence region, A04p31530.1_BnaDAR, annotated as NPR (NON EXPRESSER OF PATHOGENESIS RELATED) regulatory protein, involved in salicylic acid perception (Wang et al., 2020), was identified as promising candidate with cis-eQTL. Another hotspot on chromosome C03 (20–21 Mb), was linked to defence responses. This region harbours the cis-eQTL of C03p35430.1_BnaDAR, annotated with function in disease resistance signalling (Walsh et al., 2006). As in our canola population confidence intervals contain dozens to hundreds of genes due to large blocks of conserved linkage disequilibrium, we utilised correlations between features to prioritise particularly promising candidates for metabolic and vegetative growth variation. Although this approach may fail in cases where causal transcripts are expressed at very low levels, display non-linear relationships or if contrasting effects mask QTL, it was successful, for both mQTL and phenotypic QTL. Correlated candidate genes were identified for 22 of the 110 co-localisations between mQTL and eQTL, and for 114 of the 366 co-localisations between phenotypic QTL and eQTL. We filtered for genetic linkage between associated markers to reduce the number of potential false positive results. However, this may have led to some false negative results, for example the markers Bn-A01-p4188629 and Bn-A01-p4164843, positioned 25.5 kb apart on chromosome A01, were associated with tyramine abundance and expression levels of a putative tyrosine decarboxylase (A01p09260.1_BnaDAR), respectively. A01p09260.1_BnaDAR emerged as candidate from our correlation analysis ($r = 0.31$), but the genetic associations were not considered colocalised due to low LD between markers. Similarly, we nominated C03p47770.1_BnaDAR, encoding an uncharacterised protein as a promising candidate gene for sucrose abundance ($r = -0.57$). However, no cis-eQTL or colocalised mQTL was detected. Nevertheless, we could link a deletion on chromosome C03 shared by a subset of nine closely related lines to severely reduced C03p47770.1_BnaDAR transcript levels as well as significantly higher sucrose, glucose and fructose levels compared to the population average.

The eQTL-mQTL colocalisation and transcript-metabolite correlations also reveal gene functions in primary metabolism causing metabolic variation. For instance, marker Bn-A01-p23602361 links an mQTL for malonic acid to an eQTL for C01p48140.1_BnaDAR, annotated as malonyl-CoA synthetase. Similarly, Bn-A06-p9119114 links an mQTL for putrescine to an eQTL for C07p60730.1_BnaDAR, encoding a putative arginine decarboxylase. Notably, C01p48140.1_BnaDAR was significantly negatively correlated with malonic acid. Other co-localisations were detected for Bn-A02-p2817281 associated with sucrose and C03p39680.1_BnaDAR annotated as 'hypersensitive-induced response protein 1' or for the

marker Bn-scaff_19244_1-p313887, corresponding to an eQTL hotspot on chromosome C01. The confidence region includes an mQTL for β-Alanine and 18 eQTL. The highest correlated transcript was C01p00680.1_BnaDAR ($r = -0.574$), annotated as alanine-glyoxylate aminotransferase.

In the third step, focusing on biomass and related traits, candidate genes were prioritised in eleven colocalisation regions (Table S2). These genes include A02p00340.1_BnaDAR, a homolog of the Arabidopsis flowering locus C (FLC). Transgenic tobacco lines expressing the Arabidopsis FLC show increased biomass and delayed flowering (Salehi et al., 2005). A05p06910.1_BnaDAR, another candidate, shows homology to the Arabidopsis photosynthetic NDH subunit of lumenal location 1 (PNSL1), which is part of the photosystem II oxygen evolving complex and has recently been identified as non-additive protein in a maize heterosis study (Wang et al., 2021). C07p48260.1_BnaDAR shows homology to the Arabidopsis 'pleiotropic regulatory locus1' (PRL1). *prl1* mutations result in transcriptional de-repression of many sucrose-regulated genes, arrests root elongation, alters leaf development and inhibits cell elongation (Farrás et al., 2001). PRL1 also appears to coordinate isoprenoid metabolism with sugar, hormone and stress responses (Flores-Pérez et al., 2010). A06p05760.1_BnaDAR (PIAL1) encodes an E3 SUMO-protein ligase. In Arabidopsis, Atpial1 and Atpial2 mutants displayed better growth compared to wild type under salinity and osmotic stress and exhibited altered sulphur metabolism (Tomanov et al., 2014). C07p48510.1_BnaDAR (CLPR4) is a component of the ClpPR protease complex. Shortage of CLPR4 causes decreases in PSI and PSII core proteins (Kim et al., 2009). Null alleles for CLPR4 caused delayed embryogenesis and albino embryos, with seedling development arrested in the cotyledon stage, demonstrating a central role in chloroplast biogenesis and protein homeostasis (Kim et al., 2009). C08p38440.1_BnaDAR, annotated as polyketide cyclase/dehydrase and lipid transport superfamily protein gene, was nominated as candidate for leaf lamina shape in poplar (Drost et al., 2015). C02p16750.1_BnaDAR, a homolog of the Arabidopsis *IAA Leucine Resistant* (ILR3) gene encodes a basic helix–loop–helix (bHLH) transcription factor, which regulates iron homeostasis, modulates auxin-conjugates hydrolysis (Selote et al., 2014; Zhang et al., 2015) and affects various stress responses (Rasheed et al., 2016; Samira et al., 2018) and photoprotection (Akmakjian et al., 2021). C06p47020.1_BnaDAR encoding casein kinase 1-like protein 2 (CKL2) regulates actin filament stability and stomatal closure in Arabidopsis, which is crucial for plant photosynthesis and transpiration (S. Zhao, Jiang, et al., 2016). Another candidate, C03p62930.1_BnaDAR, shows similarity to the Arabidopsis gamma-type carbonic anhydrase-like 1 (CAL1), which is part of the mitochondrial NADH dehydrogenase complex I and is of particular importance for the respiratory chain in mitochondria and for ATP generation (Fromm et al., 2016; Klodmann et al., 2010). Complex I is essential for development and plays a central role in photomorphogenesis and cellular energy metabolism (Wang et al., 2012). BNapus_Darmor_BZH_scaffold_38p02360.1_BnaDAR, annotated as 'Elongation factor 1-beta 1', while not located within the eleven regions, is another very promising candidate. Expression levels are significantly negatively correlated with biomass and projected leaf area. Moreover, the eQTL on chromosome C02 is colocalised with multiple QTL for biomass and biomass-related i-traits. The associated marker Bn-scaff_16804_1-p178142 was previously identified by Knoch et al. (2020) as dynamic QTL for projected leaf area and as one of five candidate regions with effects on multiple biomass-related traits.

To gain evidence for biomass and growth-related candidate genes even beyond the multi-omics QTL co-localisations and feature correlations, we investigated differences in biomass using three further complementary approaches: (1) by transcriptome PCA, (2) by analysis of differentially expressed genes between contrasting lines and (3) by weighted gene co-expression network analysis. In the PCA, a partial separation of lines with high/low biomass was observed for PC3. This pattern overlaps with the breeding pools of our population: lines of pools 2 and 3 display on average higher biomass compared to lines of pool 1 and the population mean. GO term enrichment analyses for the transcripts with the highest positive and negative loadings of PC3 revealed for low biomass lines a reduction in transcript abundance of genes related to ribosome, RNA methylation and cotranslational protein targeting, indicating a potential reduction in protein biosynthesis in these lines. In contrast, high biomass lines displayed an enrichment of genes related to chloroplast functions. Previous studies in Arabidopsis indicated that growth is associated with the ribosome number and polysome loading (Czedik-Eysenberg et al., 2016; Pal et al., 2013; Piques et al., 2009) and found growth rates negatively correlated with protein turnover (Ishihara et al., 2017). Interestingly, we observed differences in the subgenome-origin, with more transcripts from the A subgenome contributing to positive loadings, while negative loadings contain more transcripts from the C subgenome. The analysis of DEGs between 'Pol 229' (low biomass) and 'Pol 419' (high biomass) supported the subgenome-specificity, and the GO term enrichment analysis yielded enriched terms as for PC3.

In a final step, we used WGCNA to identify modules related to biomass. In particular the 'red module' displayed substantial correlation to both, biomass and growth-related traits. Notably, four genes, A06p05760.1_BnaDAR (PIAL1), A10p16280.1_BnaDAR (uncharacterised), C07p48260.1_BnaDAR (PRL1) and C07p48510.1_BnaDAR (CLPR4) were

detected by all four analyses: They are positioned within the eleven colocalisation regions, found among the top 100 positive PC3 loadings, differentially expressed between 'Pol 229' and 'Pol 419' and are part of the 'red module'. These genes thus represent prime candidates for further functional investigations. Validation might be performed by overexpression or gene silencing, targeted gene knockout using CRISPR/Cas9 or characterisation in a heterologous system, such as Arabidopsis. Transient assays, for example virus-induced gene silencing (VIGS) or RNAi might be used to validate mQTL candidate genes. Sequence variation information of the identified and prioritised genes will be useful for prediction and improvement of juvenile vigour in spring-type *B. napus*. Fertiliser, growth regulator and pesticide input depends on juvenile biomass and are critical for resource-efficient production. Also, juvenile vigour alters the competitive advantage over weeds and escape mechanisms for pest management. Efficient establishment of the canola crop thus contributes directly to economic success. Plants need to be strengthened in this phase, as they are increasingly threatened by climate change and restricted pesticide use.

Data resources generated and analyses performed in this study comprehensively addressed the major issue in QTL studies of gene candidate prioritisation and functional annotation of polymorphism effects. While several previous studies showed the value of such multi-omic approaches across different plant species [Arabidopsis: Brotman et al., 2011; Szymanski et al., 2014; Luzarowska et al., 2023, maize: De Abreu E Lima et al., 2018, tomato: Tieman et al., 2017; Szymański et al., 2020; Zhu et al., 2018], all these cases related to characterisation of molecular mechanisms responsible for emergence of relatively simple, monogenic phenotypes. In contrast, in this study we quantified functional relations of multiple omics layers to complex multigenic phenotypes related to growth and plant architecture. Our data enable us to link the phenotypic effects of genetic variation to molecular features that are either directly or indirectly associated with them. This link significantly eases the interpretation of QTL for complex phenotypes by providing ways of gene prioritisation and thus enabling to generate a plausible hypothesis about molecular elements responsible for emergence of reported phenotypes.

Such multi-omic approaches fill the knowledge gap originating from incomplete genomic information. While detection of structural variation by long read-based genome sequencing identified mechanistic links between change of a gene sequence and its downstream phenotypic effect (Alonge et al., 2020), the SNP data used for GWAS analysis rarely provide insights into the molecular effects of each polymorphism. Expression and metabolic QTL data provide such insights in an unbiased way and is more likely to point out new genes and gene functions in a QTL than a knowledge-based prioritisation of, for example transcription factors and enzymes (Brotman et al., 2011). In a longer perspective, we believe our data will contribute to development of gene-function based models for plant-performance prediction (Weckwerth et al., 2020) and to the efficient utilisation of the vast increases in (genome) sequence information that are expected to emerge from the ever-increasing throughput and depth of sequencing technologies (Belser et al., 2018).

EXPERIMENTAL PROCEDURES

Population and genotyping

As a basis for genome-wide association studies we re-analysed previously acquired genotype data of 477 spring-type *Brassica napus* (canola) lines from a hybrid breeding programme (Jan et al., 2016; Knoch et al., 2020, 2021) generated using the *Brassica* Infinium™ 60 k genotyping array (Clarke et al., 2016; Mason et al., 2017). Starting from the raw data (*.idat files), we called single nucleotide polymorphism (SNP) and copy-number variation (CNV) markers using the 'gsrc' R pipeline (Grandke et al., 2017). Positions were derived by anchoring the 50 bp oligonucleotide probes on the Darmor-bzh v10 reference genome (Rousseau-Gueutin et al., 2020) using BLASTN (parameters: -perc_identity 90 -evalue 10 -word_size 11). Alignments were filtered for 49–51 bp length, percent identity ≥98, allowing only one gap or mismatch. 'Best unique hits', the alignment with the highest bitscore and only one hit in the genome with the selected criteria, were retained. SNP calls were filtered <10% missing data, <25% heterozygous calls and a minor allele frequency (MAF) ≥ 0.01 to include 'low frequency' but exclude 'rare' variants. CNVs were filtered to be present in at least 5 lines. Filtering resulted in 25 000 SNPs and 6098 CNVs (6039 deletions and 59 duplications). Missing SNP calls were imputed using the BEAGLE v.4.1 implementation in the 'synbreed' R package (Wimmer et al., 2012).

Plant cultivation and high-throughput phenotyping

Plant cultivation and phenotyping (image acquisition) was performed in a previous study described in Knoch et al. (2020). Briefly, plants were cultivated and phenotyped in the IPK phenotyping facility for large plants (Junker et al., 2015, Figure S1) under controlled spring-like environmental conditions in an incomplete randomised block design with three replicates. Each replicate consisted of a pot with nine plants. Plants were imaged daily using visible light (VIS), static fluorescence (FLUO) and near-infrared (NIR) camera systems for the period between 6 and 27 days after sowing (DAS), acquiring top and sideview images. Shoot material of four inner plants was sampled at 14 DAS to generate transcriptome and metabolome profiles. Shoot fresh and dry weights of the remaining five plants were determined at 28 DAS.

Image analysis and extraction of i-traits

Phenotypic traits (i-traits) were derived from high-throughput image analyses performed on approximately 420 000 images using IAP v2.1.0 (Klukas et al., 2014). The raw image data, obtained from a previous study (Knoch et al., 2020) that focused on four growth-related phenotypic traits, were subjected to a deepened image analysis. We customised a pipeline with several pre-processing, segmentation and feature extraction steps (Data S1). 1194 i-traits were obtained, including 128 (10.7%)

geometric traits giving insights into plant morphology, 930 (77.9%) traits related to plant pigmentation, 104 (8.7%) traits related to static chlorophyll fluorescence and 32 (2.7%) traits related to water content and water dynamics. To cope with environmental differences between experiments and potential G x E interactions, best linear unbiased estimators (BLUEs) and broad-sense heritabilities (H^2) were calculated across experiments (Knoch et al., 2020). To define a core set, after outlier correction, i-traits were filtered for $H^2 > 0.7$ on at least one day and stepwise variable selection using variance inflation factors (VIF $\leq$10) was applied to minimise multicollinearity (Chen et al., 2014). 123 i-traits, including a subset of 32 manually selected biomass-related traits, were retained for subsequent analyses. Relative growth rates for 'estimated biovolume', 'projected leaf area' and 'early plant height', and final biomass values were obtained from Knoch et al. (2020).

Metabolite profiling

The second data set used for the analyses performed here consisted of relative abundancies of 154 metabolites, 64 of known and 90 of unknown chemical structure. The generation of this data set by GC-MC-based metabolite profiling, quality controls and data normalisation are described in Knoch et al. (2021). Samples were extracted in MeOH/$CHCl_3$/H_2O (15.0 $\pm$ 1.5 mg FW), dried and in-line derivatised (MPS2 autosampler, Gerstel, Mülheim an der Ruhr, Germany) prior to GC–MS analysis (Agilent, Waldbronn, Germany/Leco, Mönchengladbach, Germany) as described by Riewe et al. (2012, 2016). Polar metabolites were identified using ChromaTOF software (LECO) and the Golm Metabolome Database mass spectra library (GMD; http://gmd.mpimp-golm.mpg.de/download/). Peak intensities were determined using the R package 'TargetSearch' (Cuadros-Inostroza *et al.*, 2009), normalised for fresh weight and detector response variation, outlier-corrected and Box-Cox power transformed.

Transcriptome analysis

As a third data layer, RNA-Seq reads were obtained from Knoch et al. (2021). Sequencing was performed using 100 bp single end (SE) reads on a HiSeq 2500 platform (Illumina, Berlin, Germany), using aliquots of the same material as for metabolite profiling. Lines were covered on average with 9.5 million reads. Reads were trimmed using Trimmomatic v0.36 (Bolger et al., 2014) with the following options: SE, HEADCROP:6, LEADING:20, TRAILING:20, SLIDINGWINDOW:4:15 and MINLEN:50 and aligned to the Darmor-bzh v10 reference using Hisat2 v2.0.4 (Kim et al., 2015). Features were counted using HTSeq v0.6.1p1 (Anders et al., 2015) and normalised for sequencing depths and transcript length using the 'tpm' procedure (Wagner et al., 2012) in R (www.r-project.org). Low-expressed features with less than 10 raw counts in more than 90% of the samples were removed as they tend to reflect noise. Data were centred and scaled (z-scores) prior to principal component analysis (PCA) using the 'pcaMethods' R package (Stacklies et al., 2007). Functional proteins annotations using the 'Automatic assignment of Human Readable Descriptions' (AHRD) package were obtained from Vollrath et al. (2021). *Arabidopsis thaliana* homologs (best BLASTX hit) were derived from the Brassicaceae Database (BRAD; Chen et al., 2022, http://brassicadb.cn/#/Annotations/).

Estimation of genomic heritability

Genomic heritabilities of metabolite and transcript levels (SNP-based heritabilities, Yang et al., 2017) were estimated with the 'BGLR' R package (Pérez & De Los Campos, 2014) in a five-fold cross-validation with 20 cross-validation rounds. Masked values were predicted using the BGLR (Bayesian Generalised Linear Regression) function, with parameters set to nIter = 5000, burnIn = 1000 and ETA = list (list (K = G, model = 'RKHS')), whereby G is the genomic relationship matrix among individuals calculated according to VanRaden (2008). Genomic heritabilities were estimated as the squared Pearson correlation between predicted and observed values in the test set (Data S2).

Genome-wide association studies using multi-omics data

To be able to compare and connect genetic associations across all three omics layers, GWAS was performed using the 'FarmCPU' R package and filtered SNP and CNV markers. Fixed and random model circulating probability unification (FarmCPU) is a multi-locus GWAS method that divides the Multi-Locus Mixed Model (MLMM) into two parts: the fixed effect model and the random-effect model and uses them iteratively (Liu et al., 2016). To eliminate the confounding between kinship in a mixed model (MLM) and genes underlying a trait of interest, the kinship is substituted with a restricted kinship matrix derived from the SUPER algorithm (Wang et al., 2014). The set of associated markers are fitted as covariates (pseudo-QTNs) in the fixed effect model for testing markers. To avoid model overfitting, the set of covariate markers are optimised using restricted maximum likelihood (REML) in the random-effect model. The method was shown to provide increased computational efficiency and to control effectively for false positives and false negatives (Kaler et al., 2019; Kumar et al., 2022; Merrick et al., 2021; Miao et al., 2019; Tibbs Cortes et al., 2021). PCA was performed on centred genotype data and the ten first principal components (PCs) were calculated. As suggested by the authors of the FarmCPU package (Liu et al., 2016), the first four PCs were fitted as covariates in the GWAS model to improve statistical power and to correct for population stratification. The maxLoop parameter was increased from the default 10 to 100 and the two parameters p.threshold and QTN.threshold were set to 0.00001 (estimated by the FarmCPU.P.Threshold function) and 0.01, respectively. After multiple testing correction, associations with P-values $_{(FDR)} \leq 0.05$ were considered statistically significant. To test whether systematic inflation occurred due to population stratification, we calculated medians of P-values and genomic inflation factors lambda (λGC) for each trait using the respective vector of P-values and the inflation function of the 'ARTP2' R package (Figure S2). Explained phenotypic variances (PVE%) of significant markers were estimated as described by Knoch et al. (2020). As covariates can substantially influence results, we additionally performed GWAS without and with an increasing number of PCs and used these results as another filter. Only 'robust' associations identified in at least one additional GWAS run were retained. Expression QTL (eQTL) were classified as cis-eQTL if associated markers were within $\pm$500 kb of the transcription start site of respective genes.

Correlations and QTL co-localisations

Pearson and Spearman correlations between trait values were performed to identify potential links between the omics layers. Correlations were calculated using the cor.test function of the 'stats' R package. For each omics layer, QTL were binned in overlapping 1 Mb intervals to detect hotspots. In a stringent approach, associations were regarded as colocalised if the same genetic marker was detected. To estimate the frequency of random colocalisation, permutations (n = 10 000) were performed, distributing associations randomly, but keeping the number of associations per genetic marker/omics layer constant, as suggested by Breitling et al. (2008). The number of co-localisations across all three omics

layers per iteration was recorded and the 95% quantile of this distribution compared to the actual number of detected co-localisations. In a less stringent approach, all associations within a 1 Mb window (± 500 kb around an associated marker) and in LD ≥0.6 were regarded as colocalised. Overlapping or adjacent intervals were further collapsed into regions of interest.

Feature selection using the Boruta algorithm

To select relevant transcripts with effects on early plant biomass, we used the Boruta algorithm implemented in R (Kursa & Rudnicki, 2010) with the following parameters: *P*-Value = 0.01, maxRuns = 1000, ntree = 10 001. Confirmed features were used to train a random forest regression model using the 'randomForest' R package (Liaw & Wiener, 2002), implementing a five-fold cross-validation procedure with 100 cross-validation rounds and a test to training set split of 0.2–0.8. Parameters were adjusted to ntree = 1001 and mtry = [p/3]. Mean prediction accuracies were obtained from squared Pearson correlations of predicted and observed values in the test sets of each round.

Prediction of phenotypic traits using random forest

To evaluate the predictability of phenotypic traits by transcript levels, traits were predicted using random forest (RF) models implemented in the scikit-learn package in python. RF models were implemented with n_estimators = 1000. We used a shuffled cross-validation scheme in which for every round 20% of the data was randomly set aside for testing and 80% for fitting the models. Prediction performance for each run was estimated on the left out test set using the R^2 metric.

Analysis of differentially expressed genes

Because the transcriptome data of the entire population of 477 lines were generated without replication (one data set per line), we selected four contrasting lines, 'Pol 229' (low biomass), 'Pol 396' (medium biomass), 'Pol 467' (medium biomass) and 'Pol 419' (high biomass) for RNA-Seq analyses with three biological replicates each. Always four plants from one pot/phenotyping experiment were pooled to constitute one biological replicate. Differentially expressed genes (DEGs) were determined using the 'edgeR' R package (Robinson et al., 2010). All expressed genes (counts >0) were used as input. Fold-changes were calculated in pairwise comparisons and *P*-values corrected using Bonferroni multiple testing correction. The significance of DEGs was determined based on an alpha threshold ≤0.05 and $|\log_2 FC| \geq 1$.

Weighted gene co-expression network analysis

To detect co-expression modules and potential key regulatory genes associated with traits of interest, we generated a co-expression network using the 'WGCNA' R package (Langfelder & Horvath, 2008). log2 transformed counts per million (CMP) were calculated using the cpm function of the 'edgeR' R package (Robinson et al., 2010). The soft thresholding power ($\beta = 6$) was determined using the pickSoftThreshold function to satisfy the scale-free topology assumption. Automatic, one-step network construction and module detection was performed using the blockwiseModules function (power = 6; TOM-type = 'unsigned'; miniModuleSize = 20; mergeCutHeight = 0.25; maxBlockSize = 45 000).

Gene ontology (GO) term enrichment

Gene Ontology (GO) term enrichment analyses were conducted for eQTL hotspots, the top positive and negative loadings of PC3 (separating lines according to biomass), the 'red module' obtained from WGCNA and differentially expressed genes. Gene ontology terms were obtained by GOMAP-singularity v1.3.8 using the Darmor-*bzh* v10 protein sequences. Enrichment analyses were performed using the 'topGO' R package (Alexa & Rahnenfuhrer, 2023). 'topGOdata' objects were built with the annFUN.gene2GO function and nodeSize = 5. Ontology levels, 'biological processes' (BP), 'molecular function' (MF) and 'cellular component' (CC), were analysed separately, using the 'weight01' algorithm and Fisher's exact tests. *P*-values were corrected for multiple testing using the *P* adjust function of the 'stats' R package with the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995).

AUTHOR CONTRIBUTIONS

TA, RCM and DK conceived the study. AA provided seed material. RJS provided genotyping data. DR and DK performed metabolite profiling. DK performed transcriptome data analysis. MCH, DR, FFP and DK analysed data. TA, JS and RCM advised on interpretation and evaluation of results. TA, RJS and RCM supervised the project and obtained the funding. DK wrote the manuscript draft. All authors read and edited the manuscript.

ACKNOWLEDGMENTS

We thank Andrea Apelt, Sibille Bettermann, Sandra Driesslein, Iris Fischer, Monika Gottowik, Beatrice Knüpfer, Marion Michaelis, Ingo Mücke, Alexandra Rech and Gunda Wehrstedt for excellent technical assistance. We thank Axel Himmelbach and the IPK sequencing service, Anne Fiebig for her help uploading the data to the European Nucleotide Archive (ENA) and Daniel Arend for his help in preparing the phenotyping data in ISA-Tab format and the upload to e!DAL. NPZ Innovation GmbH (NPZi) and Deutsche Saatveredelung AG (DSV) provided seed material. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This research was supported by the ´Deutsche Forschungsgemeinschaft´ (DFG). 'PREDICT: Omics-based models for prediction of hybrid performance in oilseed rape' (project number 234585441). Costs for open access publishing were partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, grant 491250510).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

High-throughput imaging data were obtained from Knoch et al. (2020) and subjected to a deepened image analysis, extracting substantially more than the initially described four phenotypic traits. The produced phenotypic data set was uploaded to the e!DAL repository in ISA-Tab format (https://doi.org/10.5447/ipk/2023/19) according to the MIAPPE standard. Genetic marker data used for the GWAS were initially provided by the Lab of Rod Snowdon. SNPs and CNVs were newly called using the 'gsrc' pipeline and the latest Darmor-*bzh* v10 reference genome assembly. Raw data files are available at ArrayExpress (E-MTAB-

13142) and marker calls have been uploaded alongside the R code used for the analyses (Data S10). RNA sequencing and metabolite profiling data were obtained from Knoch et al. (2021), whereby the RNA-Seq data was re-analyzed using the Darmor-bzh v10 reference genome assembly. Transcriptome data are available at ENA (PRJEB63226) and metabolite data were uploaded to MetaboLights (MTBLS8056).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Overview of the IPK's Phenotyping facility for large plants.

Figure S2. Genomic control for confounding effects of population stratification.

Figure S3. Heritabilities of i-traits and genomic heritability of molecular features.

Figure S4. Correlation analyses between the omics data sets.

Figure S5. Differences in sugar content in lines with deletion on chromosome C03.

Figure S6. eQTL, mQTL and phenotypic QTL distribution across the *B. napus* genome.

Figure S7. Multi-omics co-localisations and permutation analyses.

Figure S8. Weighted gene co-expression network analysis (WGCNA).

Figure S9. Comparison of representative high-, medium- and low biomass lines.

Table S1. Highest correlations between transcripts and biomass-related traits.

Table S2. Candidate regions associated with biomass and biomass-related traits.

Data S1. Customised IAP image analysis pipeline.

Data S2. Predictability of i-traits using random forest and genomic heritabilities.

Data S3. Pearson correlations between the omics data sets.

Data S4. Selected features for early biomass by Boruta.

Data S5. Detected marker-trait associations.

Data S6. GO term enrichment analyses.

Data S7. List of marker-trait colocalisation between omics layers.

Data S8. List of differentially expressed genes between lines.

Data S9. Data integration.

Data S10.

REFERENCES

Akmakjian, G.Z., **Riaz, N.** & **Guerinot, M.L.** (2021) Photoprotection during iron deficiency is mediated by the bHLH transcription factors PYE and ILR3. *Proceedings. National Academy of Sciences. United States of America*, **118**, e2024918118.

Alexa, A. and **Rahnenfuhrer, J.** (2023) *topGO: enrichment analysis for gene ontology*. doi:10.18129/B9.bioc.topGO, R package version 2.54.0. Available from: https://bioconductor.org/packages/topGO [Accessed 9th November 2023].

Alonge, M., **Wang, X.**, **Benoit, M.**, **Soyk, S.**, **Pereira, L.**, **Zhang, L.** ***et al.*** (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, **182**, 145–161.e23.

Anders, S., **Pyl, P.T.** & **Huber, W.** (2015) HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Barros, J., **Serk, H.**, **Granlund, I.** & **Pesquet, E.** (2015) The cell biology of lignification in higher plants. *Annals of Botany*, **115**, 1053–1074.

Basunanda, P., **Radoev, M.**, **Ecke, W.**, **Friedt, W.**, **Becker, H.C.** & **Snowdon, R.J.** (2010) Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*, **120**, 271–281.

Belser, C., **Istace, B.**, **Denis, E.**, **Dubarry, M.**, **Baurens, F.C.**, **Falentin, C.** ***et al.*** (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*, **4**, 879–887.

Benjamini, Y. & **Hochberg, Y.** (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

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

Breitling, R., **Li, Y.**, **Tesson, B.M.**, **Fu, J.**, **Wu, C.**, **Wiltshire, T.** ***et al.*** (2008) Genetical genomics: spotlight on QTL hotspots. *PLoS Genetics*, **4**, e1000232.

Brotman, Y., **Riewe, D.**, **Lisec, J.**, **Meyer, R.C.**, **Willmitzer, L.** & **Altmann, T.** (2011) Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in Arabidopsis. *Journal of Plant Physiology*, **168**, 1387–1394.

Bus, A., **Körber, N.**, **Parkin, I.A.P.**, **Samans, B.**, **Snowdon, R.J.**, **Li, J.** ***et al.*** (2014) Species- and genome-wide dissection of the shoot ionome in Brassica napus and its relationship to seedling development. *Frontiers in Plant Science*, **5**, 485.

Cai, G., **Yang, Q.**, **Chen, H.**, **Yang, Q.**, **Zhang, C.**, **Fan, C.** ***et al.*** (2016) Genetic dissection of plant architecture and yield-related traits in *Brassica napus*. *Scientific Reports*, **6**, 21625.

Chen, D., **Neumann, K.**, **Friedel, S.**, **Kilian, B.**, **Chen, M.**, **Altmann, T.** ***et al.*** (2014) Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell*, **26**, 4636–4655.

Chen, H., **Wang, T.**, **He, X.**, **Cai, X.**, **Lin, R.**, **Liang, J.** ***et al.*** (2022) BRAD V3.0: an upgraded Brassicaceae database. *Nucleic Acids Research*, **50**, D1432–D1441.

Chen, W., **Zhang, Y.**, **Liu, X.**, **Chen, B.**, **Tu, J.** & **Tingdong, F.** (2007) Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F2 populations. *Theoretical and Applied Genetics*, **115**, 849–858.

Choi, H.-K. (2019) Translational genomics and multi-omics integrated approaches as a useful strategy for crop breeding. *Genes Genom*, **41**, 133–146.

Clarke, W.E., **Higgins, E.E.**, **Plieske, J.**, **Wieseke, R.**, **Sidebottom, C.**, **Khedikar, Y.** ***et al.*** (2016) A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theoretical and Applied Genetics*, **129**, 1887–1899.

Crossa, J., **Fritsche-Neto, R.**, **Montesinos-Lopez, O.A.**, **Costa-Neto, G.**, **Dreisigacker, S.**, **Montesinos-Lopez, A.** ***et al.*** (2021) The modern plant breeding triangle: optimizing the use of genomics, Phenomics, and Enviromics data. *Frontiers in Plant Science*, **12**, 651480.

Cuadros-Inostroza, Á., **Caldana, C.**, **Redestig, H.**, **Kusano, M.**, **Lisec, J.**, **Peña-Cortés, H.** ***et al.*** (2009) TargetSearch - a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics*, **10**, 428.

Czedik-Eysenberg, A., **Arrivault, S.**, **Lohse, M.A.**, **Feil, R.**, **Krohn, N.**, **Encke, B.** ***et al.*** (2016) The interplay between carbon availability and growth in different zones of the growing maize leaf. *Plant Physiology*, **172**, 943–967.

De Abreu E Lima, F., **Li, K.**, **Wen, W.**, **Yan, J.**, **Nikoloski, Z.**, **Willmitzer, L.** ***et al.*** (2018) Unraveling lipid metabolism in maize with time-resolved multi-omics data. *The Plant Journal*, **93**, 1102–1115.

Dong, H., **Tan, C.**, **Li, Y.**, **He, Y.**, **Wei, S.**, **Cui, Y.** ***et al.*** (2018) Genome-wide association study reveals both overlapping and independent genetic loci to control seed weight and silique length in *Brassica napus*. *Frontiers in Plant Science*, **9**, 921.

Drost, D.R., **Puranik, S.**, **Novaes, E.**, **Novaes, C.R.D.B.**, **Dervinis, C.**, **Gailing, O.** ***et al.*** (2015) Genetical genomics of Populus leaf shape variation. *BMC Plant Biology*, **15**, 166.

Farrás, R., **Ferrando, A.**, **Jásik, J.**, **Kleinow, T.**, **Okrész, L.**, **Tiburcio, A.** ***et al.*** (2001) SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *The EMBO Journal*, **20**, 2742–2756.

Flood, P.J., **Kruijer, W.**, **Schnabel, S.K.**, **Van Der Schoor, R.**, **Jalink, H.**, **Snel, J.F.H.** ***et al.*** (2016) Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals circadian and long-term fluctuations in heritability. *Plant Methods*, **12**, 14.

Flores-Pérez, U., **Pérez-Gil, J.**, **Closa, M.**, **Wright, L.P.**, **Botella-Pavía, P.**, **Phillips, M.A.** ***et al.*** (2010) Pleiotropic regulatory locus 1 (PRL1) integrates the regulation of sugar responses with isoprenoid metabolism in Arabidopsis. *Molecular Plant*, **3**, 101–112.

Fromm, S., **Senkler, J.**, **Eubel, H.**, **Peterhänsel, C.** & **Braun, H.-P.** (2016) Life without complex I: proteome analyses of an Arabidopsis mutant lacking the mitochondrial NADH dehydrogenase complex. *Journal of Experimental Botany*, **67**, 3079–3093.

Grandke, F., **Snowdon, R.** & **Samans, B.** (2017) Gsrc: an R package for genome structure rearrangement calling. *Bioinformatics*, **33**, 545–546.

Hayama, R., **Yang, P.**, **Valverde, F.**, **Mizoguchi, T.**, **Furutani-Hayama, I.**, **Vierstra, R.D.** ***et al.*** (2019) Ubiquitin carboxyl-terminal hydrolases are required for period maintenance of the circadian clock at high temperature in Arabidopsis. *Scientific Reports*, **9**, 17030.

Ishihara, H., **Moraes, T.A.**, **Pyl, E.-T.**, **Schulze, W.X.**, **Obata, T.**, **Scheffel, A.** ***et al.*** (2017) Growth rate correlates negatively with protein turnover in Arabidopsis accessions. *The Plant Journal*, **91**, 416–429.

Jan, H.U., **Abbadi, A.**, **Lücke, S.**, **Nichols, R.A.** & **Snowdon, R.J.** (2016) Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS One*, **11**, e0147769.

Junker, A., **Muraya, M.M.**, **Weigelt-Fischer, K.**, **Arana-Ceballos, F.**, **Klukas, C.**, **Melchinger, A.E.** ***et al.*** (2015) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Frontiers in Plant Science*, **5**, 770.

Kaler, A.S., **Gillman, J.D.**, **Beissinger, T.** & **Purcell, L.C.** (2019) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Frontiers in Plant Science*, **10**, 1794.

Kim, D., **Langmead, B.** & **Salzberg, S.L.** (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

Kim, J., **Rudella, A.**, **Ramirez Rodriguez, V.**, **Zybailov, B.**, **Olinares, P.D.B.** & **Van Wijk, K.J.** (2009) Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in *Arabidopsis*. *The Plant Cell*, **21**, 1669–1692.

Klodmann, J., **Sunderhaus, S.**, **Nimtz, M.**, **Jänsch, L.** & **Braun, H.-P.** (2010) Internal architecture of mitochondrial complex I from Arabidopsis thaliana. *Plant Cell*, **22**, 797–810.

Klukas, C., **Chen, D.** & **Pape, J.-M.** (2014) Integrated analysis platform: an open-source information system for high-throughput plant phenotyping. *Plant Physiology*, **165**, 506–518.

Knoch, D., **Abbadi, A.**, **Grandke, F.**, **Meyer, R.C.**, **Samans, B.**, **Werner, C.R.** ***et al.*** (2020) Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnology Journal*, **18**, 68–82.

Knoch, D., **Werner, C.R.**, **Meyer, R.C.**, **Riewe, D.**, **Abbadi, A.**, **Lücke, S.** ***et al.*** (2021) Multi-omics-based prediction of hybrid performance in canola. *Theoretical and Applied Genetics*, **134**, 1147–1165.

Kojima, H., **Suzuki, T.**, **Kato, T.**, **Enomoto, K.-i.**, **Sato, S.**, **Kato, T.** ***et al.*** (2007) Sugar-inducible expression of the nucleolin-1 gene of Arabidopsis thaliana and its role in ribosome synthesis, growth and development. *The Plant Journal*, **49**, 1053–1063.

Korasick, D.A., **Enders, T.A.** & **Strader, L.C.** (2013) Auxin biosynthesis and storage forms. *Journal of Experimental Botany*, **64**, 2541–2555.

Körber, N., **Bus, A.**, **Li, J.**, **Higgins, J.**, **Bancroft, I.**, **Higgins, E.E.** ***et al.*** (2015) Seedling development traits in *Brassica napus* examined by gene expression analysis and association mapping. *BMC Plant Biology*, **15**, 136.

Kumar, K., **Anjoy, P.**, **Sahu, S.**, **Durgesh, K.**, **das, A.**, **Tribhuvan, K.U.** ***et al.*** (2022) Single trait versus principal component based association analysis for flowering related traits in pigeonpea. *Scientific Reports*, **12**, 10453.

Kursa, M.B. & **Rudnicki, W.R.** (2010) Feature selection with the Boruta package. *Journal of Statistical Software*, **36**, 1–13. Available from: https://doi.org/10.18637/jss.v036.i11

Langfelder, P. & **Horvath, S.** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Langstroff, A., **Heuermann, M.C.**, **Stahl, A.** & **Junker, A.** (2022) Opportunities and limits of controlled-environment plant phenotyping for climate response traits. *Theoretical and Applied Genetics*, **135**, 1–16.

Li, K., **Yao, Y.**, **Xiao, L.**, **Zhao, Z.**, **Guo, S.**, **Fu, Z.** ***et al.*** (2018) Fine mapping of the *Brassica napus* Bnsdt1 gene associated with determinate growth habit. *Theoretical and Applied Genetics*, **131**, 193–208.

Li, L., **Tian, Z.**, **Chen, J.**, **Tan, Z.**, **Zhang, Y.**, **Zhao, H.** ***et al.*** (2023) Characterization of novel loci controlling seed oil content in Brassica napus by marker metabolite-based multi-omics analysis. *Genome Biology*, **24**, 141.

Li, Q. & **Yan, J.** (2020) Sustainable agriculture in the era of omics: knowledge-driven crop breeding. *Genome Biology*, **21**, 154 s13059-020-02073–5.

Li, R., **Jeong, K.**, **Davis, J.T.**, **Kim, S.**, **Lee, S.**, **Michelmore, R.W.** ***et al.*** (2018) Integrated QTL and eQTL mapping provides insights and candidate genes for fatty acid composition, flowering time, and growth traits in a F2 population of a novel synthetic allopolyploid *Brassica napus*. *Frontiers in Plant Science*, **9**, 1632.

Liaw, A. & **Wiener, M.** (2002) Classification and regression by randomForest. *R news*, **2**, 18–22.

Liu, J., **Hua, W.**, **Hu, Z.**, **Yang, H.**, **Zhang, L.**, **Li, R.** ***et al.*** (2015) Natural variation in ARF18 gene simultaneously affects seed weight and silique length in polyploid rapeseed. *Proceedings. National Academy of Sciences. United States of America*, **112**, E5123–E5132.

Liu, X., **Huang, M.**, **Fan, B.**, **Buckler, E.S.** & **Zhang, Z.** (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, **12**, e1005767.

Luo, Z., **Wang, M.**, **Long, Y.**, **Huang, Y.**, **Shi, L.**, **Zhang, C.** ***et al.*** (2017) Incorporating pleiotropic quantitative trait loci in dissection of complex traits: seed yield in rapeseed as an example. *Theoretical and Applied Genetics*, **130**, 1569–1585.

Luzarowska, U., **Ruß, A.-K.**, **Joubés, J.**, **Batsale, M.**, **Szymański, J.**, **Thirumalaikumar, V.P.** ***et al.*** (2023) Hello darkness, my old friend: 3-KETOACYL-COENZYME A SYNTHASE4 is a branch point in the regulation of triacylglycerol synthesis in Arabidopsis thaliana. *Plant Cell*, **35**, 1984–2005.

Mason, A.S., **Higgins, E.E.**, **Snowdon, R.J.**, **Batley, J.**, **Stein, A.**, **Werner, C.** ***et al.*** (2017) A user guide to the brassica 60K Illumina Infinium™ SNP genotyping array. *Theoretical and Applied Genetics*, **130**, 621–633.

Merrick, L.F., **Burke, A.B.**, **Zhang, Z.** & **Carter, A.H.** (2021) Comparison of single-trait and multi-trait genome-wide association models and inclusion of correlated traits in the dissection of the genetic architecture of a complex trait in a breeding program. *Frontiers in Plant Science*, **12**, 772907.

Miao, C., **Yang, J.** & **Schnable, J.C.** (2019) Optimising the identification of causal variants across varying genetic architectures in crops. *Plant Biotechnology Journal*, **17**, 893–905.

Miedes, E., **Suslov, D.**, **Vandenbussche, F.**, **Kenobi, K.**, **Ivakov, A.**, **van der Straeten, D.** ***et al.*** (2013) Xyloglucan endotransglucosylase/hydrolase (XTH) overexpression affects growth and cell wall mechanics in etiolated Arabidopsis hypocotyls. *Journal of Experimental Botany*, **64**, 2481–2497.

Mir, R.R., **Reynolds, M.**, **Pinto, F.**, **Khan, M.A.** & **Bhat, M.A.** (2019) High-throughput phenotyping for crop improvement in the genomics era. *Plant Science*, **282**, 60–72.

Pal, S.K., **Liput, M.**, **Piques, M.**, **Ishihara, H.**, **Obata, T.**, **Martins, M.C.M.** ***et al.*** (2013) Diurnal changes of polysome loading track sucrose content in the rosette of wild-type arabidopsis and the starchless pgm mutant. *Plant Physiology*, **162**, 1246–1265.

Pérez, P. & **De Los Campos, G.** (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, **198**, 483–495.

Petricka, J.J. & **Nelson, T.M.** (2007) Arabidopsis nucleolin affects plant development and patterning. *Plant Physiology*, **144**, 173–186.

Piques, M., **Schulze, W.X.**, **Höhne, M.**, **Usadel, B.**, **Gibon, Y.**, **Rohwer, J.** ***et al.*** (2009) Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Molecular Systems Biology*, **5**, 314.

Qi, H., **Xia, F.-N.**, **Xiao, S.** & **Li, J.** (2022) TRAF proteins as key regulators of plant development and stress responses. *Journal of Integrative Plant Biology*, **64**, 431–448.

Radoev, M., **Becker, H.C.** & **Ecke, W.** (2008) Genetic analysis of Heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics*, **179**, 1547–1558.

Rasheed, S., **Bashir, K.**, **Matsui, A.**, **Tanaka, M.** & **Seki, M.** (2016) Transcriptomic analysis of soil-grown Arabidopsis thaliana roots and shoots in response to a drought stress. *Frontiers in Plant Science*, **7**, 180.

Riewe, D., **Jeon, H.-J.**, **Lisec, J.**, **Heuermann, M.C.**, **Schmeichel, J.**, **Seyfarth, M.** ***et al.*** (2016) A naturally occurring promoter polymorphism of the Arabidopsis FUM2 gene causes expression variation, and is associated with metabolic and growth traits. *The Plant Journal*, **88**, 826–838.

Riewe, D., **Koohi, M.**, **Lisec, J.**, **Pfeiffer, M.**, **Lippmann, R.**, **Schmeichel, J. *et al.*** (2012) A tyrosine aminotransferase involved in tocopherol synthesis in Arabidopsis. *The Plant Journal*, **71**, 850–859.

Robinson, M.D., **McCarthy, D.J.** & **Smyth, G.K.** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rousseau-Gueutin, M., **Belser, C.**, **Da Silva, C.**, **Richard, G.**, **Istace, B.**, **Cruaud, C. *et al.*** (2020) Long-read assembly of the *Brassica napus* reference genome Darmor-bzh. *GigaScience*, **9**, giaa137.

Salehi, H., **Ransom, C.B.**, **Oraby, H.F.**, **Seddighi, Z.** & **Sticklen, M.B.** (2005) Delay in flowering and increase in biomass of transgenic tobacco expressing the Arabidopsis floral repressor gene FLOWERING LOCUS C. *Journal of Plant Physiology*, **162**, 711–717.

Samira, R., **Li, B.**, **Kliebenstein, D.**, **Li, C.**, **Davis, E.**, **Gillikin, J.W. *et al.*** (2018) The bHLH transcription factor ILR3 modulates multiple stress responses in Arabidopsis. *Plant Molecular Biology*, **97**, 297–309.

Schaefer, R.J., **Michno, J.-M.**, **Jeffers, J.**, **Hoekenga, O.**, **Dilkes, B.**, **Baxter, I. *et al.*** (2018) Integrating Coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell*, **30**, 2922–2942.

Scharr, H., **Dee, H.**, **French, A.P.** & **Tsaftaris, S.A.** (2016) Special issue on computer vision and image analysis in plant phenotyping. *Machine Vision and Applications*, **27**, 607–609.

Schmid, M., **Uhlenhaut, N.H.**, **Godard, F.**, **Demar, M.**, **Bressan, R.**, **Weigel, D. *et al.*** (2003) Dissection of floral induction pathways using global expression analysis. *Development*, **130**, 6001–6012.

Selote, D., **Samira, R.**, **Matthiadis, A.**, **Gillikin, J.W.** & **Long, T.A.** (2014) Iron-binding E3 ligase mediates iron response in plants by targeting basic helix-loop-helix transcription factors. *Plant Physiology*, **167**, 273–286.

Shaw, R.K., **Shen, Y.**, **Wang, J.**, **Sheng, X.**, **Zhao, Z.**, **Yu, H. *et al.*** (2021) Advances in multi-omics approaches for molecular breeding of black rot resistance in Brassica oleracea L. *Frontiers in Plant Science*, **12**, 742553.

Shaw, R.K., **Shen, Y.**, **Yu, H.**, **Sheng, X.**, **Wang, J.** & **Gu, H.** (2022) Multi-omics approaches to improve Clubroot resistance in brassica with a special focus on Brassica oleracea L. *International Journal of Molecular Sciences*, **23**, 9280.

Shen, Y., **Zhou, G.**, **Liang, C.** & **Tian, Z.** (2022) Omics-based interdisciplinarity is accelerating plant breeding. *Current Opinion in Plant Biology*, **66**, 102167.

Shi, J., **Li, R.**, **Qiu, D.**, **Jiang, C.**, **Long, Y.**, **Morgan, C. *et al.*** (2009) Unraveling the complex trait of crop yield with quantitative trait loci mapping in *Brassica napus*. *Genetics*, **182**, 851–861.

Stacklies, W., **Redestig, H.**, **Scholz, M.**, **Walther, D.** & **Selbig, J.** (2007) pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, **23**, 1164–1167.

Szymański, J., **Bocobza, S.**, **Panda, S.**, **Sonawane, P.**, **Cárdenas, P.D.**, **Lashbrooke, J. *et al.*** (2020) Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nature Genetics*, **52**, 1111–1121.

Szymanski, J., **Brotman, Y.**, **Willmitzer, L.** & **Cuadros-Inostroza, Á.** (2014) Linking gene expression and membrane lipid composition of Arabidopsis. *Plant Cell*, **26**, 915–928.

Teaster, N.D., **Keereetaweep, J.**, **Kilaru, A.**, **Wang, Y.-S.**, **Tang, Y.**, **Tran, C.N.-Q. *et al.*** (2012) Overexpression of fatty acid amide hydrolase induces early flowering in Arabidopsis thaliana. *Frontiers in Plant Science*, **3**, 32.

Tibbs Cortes, L., **Zhang, Z.** & **Yu, J.** (2021) Status and prospects of genome-wide association studies in plants. *Plant Genome*, **14**, e20077.

Tieman, D., **Zhu, G.**, **Resende, M.F.R. *et al.*** (2017) A chemical genetic roadmap to improved tomato flavor. *Science*, **355**, 391–394.

Tomanov, K., **Zeschmann, A.**, **Hermkes, R.**, **Eifler, K.**, **Ziba, I.**, **Grieco, M. *et al.*** (2014) Arabidopsis PIAL1 and 2 promote SUMO chain formation as E4-type SUMO ligases and are involved in stress responses and sulfur metabolism. *Plant Cell*, **26**, 4547–4560.

VanRaden, P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science*, **91**, 4414–4423.

Visscher, P.M., **Hill, W.G.** & **Wray, N.R.** (2008) Heritability in the genomics era—concepts and misconceptions. *Nature Reviews. Genetics*, **9**, 255–266.

Vollrath, P., **Chawla, H.S.**, **Alnajar, D.**, **Gabur, I.**, **Lee, H.T.**, **Weber, S. *et al.*** (2021) Dissection of quantitative blackleg resistance reveals novel variants of resistance gene Rlm9 in elite Brassica napus. *Frontiers in Plant Science*, **12**, 749491.

Wagner, G.P., **Kin, K.** & **Lynch, V.J.** (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences*, **131**, 281–285.

Walsh, T.A., **Neal, R.**, **Merlo, A.O.**, **Honma, M.**, **Hicks, G.R.**, **Wolff, K. *et al.*** (2006) Mutations in an auxin receptor homolog AFB5 and in SGT1b confer resistance to synthetic Picolinate auxins and not to 2,4-Dichlorophenoxyacetic acid or Indole-3-acetic acid in Arabidopsis. *Plant Physiology*, **142**, 542–552.

Wang, D., **Mu, Y.**, **Hu, X.**, **Ma, B.**, **Wang, Z.**, **Zhu, L. *et al.*** (2021) Comparative proteomic analysis reveals that the Heterosis of two maize hybrids is related to enhancement of stress response and photosynthesis respectively. *BMC Plant Biology*, **21**, 34.

Wang, Q., **Fristedt, R.**, **Yu, X.**, **Chen, Z.**, **Liu, H.**, **Lee, Y. *et al.*** (2012) The γ-carbonic anhydrase subcomplex of mitochondrial complex I is essential for development and important for photomorphogenesis of Arabidopsis. *Plant Physiology*, **160**, 1373–1383.

Wang, Q., **Tian, F.**, **Pan, Y.**, **Buckler, E.S.** & **Zhang, Z.** (2014) A SUPER powerful method for genome wide association study. *PLoS One*, **9**, e107684.

Wang, W., **Withers, J.**, **Li, H.**, **Zwack, P.J.**, **Rusnac, D.V.**, **Shi, H. *et al.*** (2020) Structural basis of salicylic acid perception by Arabidopsis NPR proteins. *Nature*, **586**, 311–316.

Watt, M., **Fiorani, F.**, **Usadel, B.**, **Rascher, U.**, **Muller, O.** & **Schurr, U.** (2020) Phenotyping: new windows into the Plant for Breeders. *Annual Review of Plant Biology*, **71**, 689–712.

Weckwerth, W., **Ghatak, A.**, **Bellaire, A.**, **Chaturvedi, P.** & **Varshney, R.K.** (2020) PANOMICS meets germplasm. *Plant Biotechnology Journal*, **18**, 1507–1525.

Wimmer, V., **Albrecht, T.**, **Auinger, H.-J.** & **Schön, C.-C.** (2012) Synbreed: a framework for the analysis of genomic prediction data using R. *Bioinformatics*, **28**, 2086–2087.

Yang, J., **Zeng, J.**, **Goddard, M.E.**, **Wray, N.R.** & **Visscher, P.M.** (2017) Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, **49**, 1304–1310.

Yang, P., **Shu, C.**, **Chen, L.**, **Xu, J.**, **Wu, J.** & **Liu, K.** (2012) Identification of a major QTL for silique length and seed weight in oilseed rape (*Brassica napus* L.). *Theoretical and Applied Genetics*, **125**, 285–296.

Yang, W., **Feng, H.**, **Zhang, X.**, **Zhang, J.**, **Doonan, J.H.**, **Batchelor, W.D. *et al.*** (2020) Crop Phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Molecular Plant*, **13**, 187–214.

Yi Chou, E., **Schuetz, M.**, **Hoffmann, N.**, **Watanabe, Y.**, **Sibout, R.** & **Samuels, A.L.** (2018) Distribution, mobility, and anchoring of lignin-related oxidative enzymes in Arabidopsis secondary cell walls. *Journal of Experimental Botany*, **69**, 1849–1859.

Yong, H.-Y., **Wang, C.**, **Bancroft, I.**, **Li, F.**, **Wu, X.**, **Kitashiba, H. *et al.*** (2015) Identification of a gene controlling variation in the salt tolerance of rapeseed (*Brassica napus* L.). *Planta*, **242**, 313–326.

Yu, K., **Wang, X.**, **Chen, F.**, **Peng, Q.**, **Chen, S.**, **Li, H. *et al.*** (2018) Quantitative trait transcripts mapping coupled with expression quantitative trait loci mapping reveal the molecular network regulating the Apetalous characteristic in *Brassica napus* L. *Frontiers in Plant Science*, **9**, 89.

Zeng, X., **Zhu, L.**, **Chen, Y.**, **Qi, L.**, **Pu, Y.**, **Wen, J. *et al.*** (2011) Identification, fine mapping and characterisation of a dwarf mutant (bnaC.Dwf) in *Brassica napus*. *Theoretical and Applied Genetics*, **122**, 421–428.

Zhang, J., **Liu, B.**, **Li, M.**, **Feng, D.**, **Jin, H.**, **Wang, P. *et al.*** (2015) The bHLH transcription factor bHLH104 interacts with IAA-LEUCINE RESISTANT3 and modulates iron homeostasis in Arabidopsis. *Plant Cell*, **27**, 787–805.

Zhang, Y., **Zhang, H.**, **Zhao, H.**, **Xia, Y.**, **Zheng, X.**, **Fan, R. *et al.*** (2022) Multi-omics analysis dissects the genetic architecture of seed coat content in Brassica napus. *Genome Biology*, **23**, 86.

Zhao, H., **Shang, G.**, **Yin, N.**, **Chen, S.**, **Shen, S.**, **Jiang, H. *et al.*** (2022) Multi-omics analysis reveals the mechanism of seed coat color formation in Brassica rapa L. *Theoretical and Applied Genetics*, **135**, 2083–2099.

Zhao, S., **Jiang, Y.**, **Zhao, Y.**, **Huang, S.**, **Yuan, M.**, **Zhao, Y. *et al.*** (2016) Casein KINASE1-LIKE PROTEIN2 regulates Actin filament stability and stomatal closure via phosphorylation of Actin depolymerizing factor. *Plant Cell*, **28**, 1422–1439.

Zhao, W., **Wang, X.**, **Wang, H.**, **Tian, J.**, **Li, B.**, **Chen, L. *et al.*** (2016) Genome-wide identification of QTL for seed yield and yield-related traits and construction of a high-density consensus map for QTL comparison in *Brassica napus*. *Frontiers in Plant Science*, **7**, 17.

Zhu, G., **Wang, S.**, **Huang, Z.**, **Zhang, S.**, **Liao, Q.**, **Zhang, C. *et al.*** (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e12.