

# Linking Metabolic QTLs with Network and *cis*-eQTLs Controlling Biosynthetic Pathways

Adam M. Wentzell<sup>1,2</sup>, Heather C. Rowe<sup>1,2</sup>, Bjarne Gram Hansen<sup>3</sup>, Carla Ticconi<sup>1</sup>, Barbara Ann Halkier<sup>3</sup>, Daniel J. Kliebenstein<sup>1,2\*</sup>

**1** Department of Plant Sciences, University of California Davis, Davis, California, United States of America, **2** Genetics Graduate Group, University of California Davis, Davis, California, United States of America, **3** Plant Biochemistry Laboratory, Department of Plant Biology, Faculty of Life Sciences, University of Copenhagen, Copenhagen, Denmark

**Phenotypic variation between individuals of a species is often under quantitative genetic control. Genomic analysis of gene expression polymorphisms between individuals is rapidly gaining popularity as a way to query the underlying mechanistic causes of variation between individuals. However, there is little direct evidence of a linkage between global gene expression polymorphisms and phenotypic consequences. In this report, we have mapped quantitative trait loci (QTLs)—controlling glucosinolate content in a population of 403 *Arabidopsis* Bay × Sha recombinant inbred lines, 211 of which were previously used to identify expression QTLs controlling the transcript levels of biosynthetic genes. In a comparative study, we have directly tested two plant biosynthetic pathways for association between polymorphisms controlling biosynthetic gene transcripts and the resulting metabolites within the *Arabidopsis* Bay × Sha recombinant inbred line population. In this analysis, all loci controlling expression variation also affected the accumulation of the resulting metabolites. In addition, epistasis was detected more frequently for metabolic traits compared to transcript traits, even when both traits showed similar distributions. An analysis of candidate genes for QTL-controlling networks of transcripts and metabolites suggested that the controlling factors are a mix of enzymes and regulatory factors. This analysis showed that regulatory connections can feedback from metabolism to transcripts. Surprisingly, the most likely major regulator of both transcript level for nearly the entire pathway and aliphatic glucosinolate accumulation is variation in the last enzyme in the biosynthetic pathway, *AOP2*. This suggests that natural variation in transcripts may significantly impact phenotypic variation, but that natural variation in metabolites or their enzymatic loci can feed back to affect the transcripts.**

Citation: Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, et al. (2007) Linking metabolic QTL with network and *cis*-eQTLs controlling biosynthetic pathways. *PLoS Genet* 3(9): e162. doi:10.1371/journal.pgen.0030162

## Introduction

A longstanding goal in genetics is to unravel the molecular and genetic bases of complex traits such as disease resistance, growth, and development. While phenotypic variation in natural populations is largely quantitative and polygenic, understanding this variation is complicated by the interaction of environmental and genetic factors [1,2]. Quantitative trait mapping, the most common approach to analyze complex traits, measures the association of genetic markers with phenotypic variation, delineating quantitative trait loci (QTLs) [1,3]. Advances in statistical models, improvements in marker technology, and expanding genomic resources have lead to increasingly refined QTL maps for a wide array of traits, ranging from development and morphology to metabolism and disease resistance [4–10]. In spite of these considerable efforts, the molecular basis of many quantitative traits remains unknown.

Recently, our understanding of quantitative traits has been enhanced by genomic approaches that use microarray technology to measure global transcript levels in mapping populations and map expression QTL (eQTL) [11–13]. Whole-genome eQTL analysis in yeast, mice, and humans has revealed that gene expression traits are highly heritable, and can have surprisingly complex underlying genetic

architecture [13–15]. Recently, similar global analysis of gene expression was conducted in two independent *A. thaliana* recombinant inbred line (RIL) mapping populations [16,17]. These studies revealed large numbers of both *cis*- and *trans*-acting eQTL, with evidence of nonadditive genetic variation and transgressive segregation, consistent with results from animal systems. In addition, network eQTL analysis in the Bay × Sha RIL population showed that transcript variation was controlled by variation in specific biological networks including both biosynthetic and signal transduction pathways

**Editor:** Vivian G. Cheung, University of Pennsylvania, United States of America

**Received** May 31, 2007; **Accepted** August 1, 2007; **Published** September 14, 2007

A previous version of this article appeared as an Early Online Release on August 1, 2007 (doi:10.1371/journal.pgen.0030162.eor).

**Copyright:** © 2007 Wentzell et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Abbreviations:** 4-MSO, 4-methylsulfinylbutyl; 4-MT, 4-methylthiobutyl; ANOVA, analysis of variance; AOP, alkenyl/hydroxypropyl locus; eQTL, expression QTL; GSL, glucosinolate locus; HPLC, high-performance liquid chromatography; Ler, Landsberg *erecta*; QTL, quantitative trait locus; RIL, recombinant inbred line

\* To whom correspondence should be addressed. E-mail: kliebenstein@ucdavis.edu

© These authors contributed equally to this work.

## Author Summary

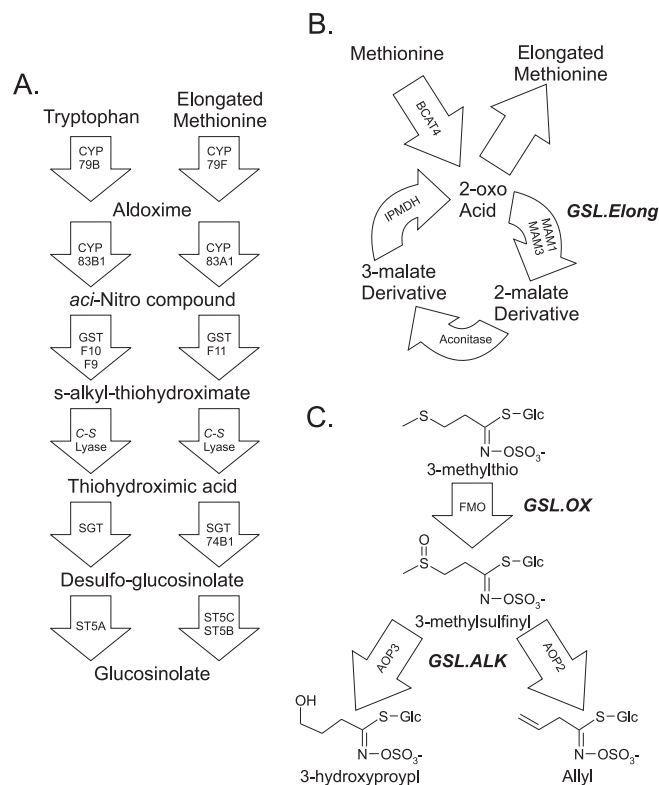
Natural genetic variation and the resulting phenotypic variation between individuals within a species have been of longstanding interest in wide-ranging fields. However, the molecular underpinnings of this phenotypic variation are relatively uncharted. Recently, genomics methodologies have been applied to understanding natural genetic variation in global gene expression. This, however, did not resolve the connection between variation in gene expression and the resulting physiological phenotype. We used two metabolic pathways within the model plant *Arabidopsis* to show that it is possible to connect genomic analysis of genetic variation to the resulting phenotype. This analysis showed that the connections between gene expression and metabolite variation were complex. Finally, the major regulators of gene expression variation for these pathways are two biosynthetic enzymes rather than traditional transcription factors. This analysis provides insights into how to connect transcriptomic and metabolomic datasets using natural genetic variation.

[18]. These studies present a detailed picture of variation in gene expression and its underlying genetic architecture, but the relationship between transcript levels and the resultant phenotypic variation remains poorly understood.

Testing the connection between eQTLs and downstream phenotypic variation requires a phenotype with detailed molecular and quantitative genetic information. Metabolic phenotypes are ideal for these studies, because these traits are highly variable and can be accurately measured using high-throughput techniques [5,19,20]. Knowledge of biochemical pathways enables comparison between the transcript level of a biosynthetic gene and downstream metabolic phenotypes. This engenders detailed hypotheses about the basis of metabolic variation that incorporate biochemical relationships, flux concepts, and transcriptional regulation [21]. Derived traits generated from the raw metabolite accumulation data can provide unique insights into the metabolic network [22,23]. These derived traits can include the sum of related metabolites, providing information about the whole pathway, or the ratio of two metabolites related as precursor and product can be used to query variation in a specific enzymatic process.

We test our ability to link eQTLs with phenotypic variation using two secondary metabolite pathways responsible for the synthesis of aliphatic and indolic glucosinolates within *A. thaliana*. These metabolites play an important role in plant defense against herbivory, and have chemopreventive activity in the human diet. An improved understanding of the genetic basis of glucosinolate variation thus affects evolution and ecology as well as nutrition and agriculture [24,25]. Glucosinolates are synthesized by a well-studied biosynthetic pathway [24,26,27], with known transcription factors [28,29] and cloned QTLs controlling structural diversity and content within *Arabidopsis* [19,30] (Figure 1).

Aliphatic and indolic glucosinolates, derived from elongated methionine derivatives and tryptophan, respectively, are synthesized and subsequently modified by two independent yet parallel pathways (Figure 1A). These biosynthetic pathways possess distinct enzymes and divergent regulation [27]. Production of aliphatic glucosinolates is controlled by three cloned QTLs controlling specific biosynthetic enzymes: *GSL.Elong*, *GSL.ALK*, and *GSL.OX* (GSL = GlucSinoLate;



**Figure 1. Glucosinolate Biosynthesis**

Arrows show the known and predicted steps for glucosinolate biosynthesis with the gene name for each biochemical reaction within the arrow. For compounds that are undetected intermediates, chemical names only are provided. For detected compounds, both the structure and chemical name are provided. The position of known genetic loci controlling biosynthetic variation is shown in italics.

(A) The pathway and genes responsible for the production of the core glucosinolate structure from tryptophan (indolic glucosinolates) and methionine (aliphatic glucosinolates).

(B) The chain elongation cycle for aliphatic glucosinolate production. Each cycle of these reactions adds a single carbon to a 2-oxo-acid, which is then transaminated to generate homo-methionine for aliphatic glucosinolate biosynthesis. The *GSL.Elong* QTL alters this cycle through variation at the *MAM1*, *MAM2*, and *MAM3* genes that leads to differential glucosinolate structure and content [32].

(C) The enzymes and genetic loci controlling aliphatic glucosinolate side chain modification within the Bay-0 × Sha RIL population. Side-chain modification is controlled by variation at the *GSL.ALK* QTL via *cis*-eQTLs at the *AOP2* and *AOP3* genes. The Cvi and Sha accessions express *AOP2* to produce alkenyl glucosinolates. In contrast, the Ler and Bay-0 accessions express *AOP3* to produce hydroxyl glucosinolates. Col-0 is null for both *AOP2* and *AOP3*, producing only the precursor methylsulfinyl glucosinolates [31]. The *GSL.OX* QTL appears to be controlled by *cis*-eQTLs regulating flavin-monoxygenase enzymes that oxygenate a methylthio to methylsulfinyl glucosinolate [33].

doi:10.1371/journal.pgen.0030162.g001

Figure 1B and 1C) [31–33]. Additional QTLs have been identified which, according to current knowledge, are not associated with known biosynthetic genes [30,34]. As such, the aliphatic and indolic glucosinolate metabolic pathways provide a useful model system to link phenotypic QTLs with eQTLs.

To compare phenotypic QTLs and eQTLs, we measured the accumulation of aliphatic and indolic glucosinolates in the Bay × Sha RIL population. In addition to 14 and 11 metabolic QTL for the indolic and aliphatic metabolites, respectively, several epistatic interactions were detected. Using the same seeds and developmental stages as the

**Table 1.** Variation in Aliphatic Glucosinolates within Bay-0, Sha, and the Bay-0 × Sha RILs

Pathway	Glucosinolate	Abbreviation	Bay-0		Sha		RIL Percentile	
			Mean	SE	Mean	SE	95th	5th
Aliphatic	3-hydroxypropyl	3-OHP	2.90	0.40	0.00	-	2.24	0.00
Aliphatic	4-methylsulfinylbutyl	4-MSO	0.00	-	0.00	-	0.79	0.00
Aliphatic	Allyl	Allyl	0.00	-	2.52	0.38	2.58	0.00
Aliphatic	But-3-enyl	But-3-enyl	0.00	-	6.38	1.08	4.92	0.00
Aliphatic	3-methylthiopropyl	3-MT	0.19	0.03	0.08	0.02	0.11	0.00
Aliphatic	4-methylthiobutyl	4-MT	0.00	-	0.00	-	0.36	0.00
Aliphatic	8-methylsulfinyloctyl	8-MSO	1.48	0.29	0.51	0.16	1.25	0.05
Aliphatic	8-methylthiooctyl	8-MT	0.19	0.02	0.08	0.01	0.22	0.02
Aliphatic	Total aliphatic		4.71	0.69	9.56	1.53	7.03	0.77
Indolic	Indol-3-ylmethyl	I3M	0.64	0.10	0.69	0.10	0.52	0.10
Indolic	4-methoxy-indol-3-ylmethyl	4MO-I3M	0.26	0.03	0.08	0.01	0.20	0.05
Indolic	N-methoxy-indol-3-ylmethyl	NMO-I3M	0.05	0.01	0.04	0.01	0.05	0.01
Indolic	Total indolic		0.95	0.13	0.81	0.11	0.65	0.22

Mean glucosinolate content in  $\mu\text{mol cm}^{-2}$  is shown for Bay-0 and Sha grown concurrently with RILs, as well as the 95th and 5th percentile values for the RILs. SE shows the standard error of the measurement. Pathway indicates whether the glucosinolate is derived from methionine (aliphatic) or tryptophan (indolic). Total aliphatic and total indolic are the sum of all aliphatic and indolic glucosinolates, respectively.  
doi:10.1371/journal.pgen.0030162.t001

previous eQTL mapping in the Bay × Sha RIL population allowed us to compare metabolite QTL and eQTL locations [17]. Comparing metabolite QTLs with network eQTLs controlling the expression of aliphatic and indolic glucosinolate biosynthetic pathways indicates that all network eQTLs colocalize with metabolic QTLs, but the inverse statement is not true, as some QTLs are detected only for metabolite traits. We obtained evidence that variation in biosynthetic enzymes and possibly transcription factors can control natural variation for transcript levels of glucosinolate biosynthetic genes. The heritability of metabolite traits was on average lower than that for transcript levels, suggesting that metabolite accumulation may be more susceptible to environmental factors. This detailed picture of glucosinolate accumulation and modification shows that eQTLs can be associated with changes in the resulting phenotype, allowing us to generate testable mechanistic hypotheses regarding the interplay between expression variation and downstream phenotypic variation.

## Results

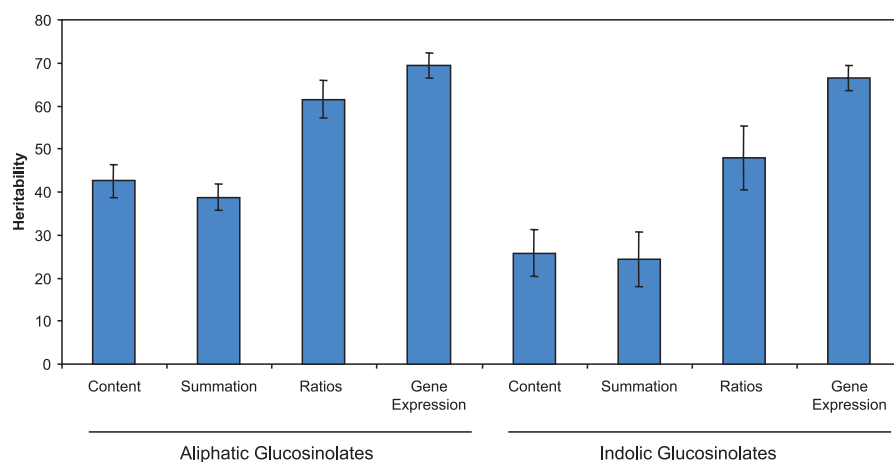
### Metabolite Trait Distributions

We measured glucosinolate production by the *A. thaliana* accessions, Bayreuth (Bay-0) and Shahdara (Sha), the parents of the Bay × Sha RIL population. The Bay-0 and Sha accessions differed in both the quantity of glucosinolates accumulated and the specific structures synthesized, verifying that the Bay-0 × Sha RIL population is potentially informative for analyzing the relationship of eQTLs to metabolic variation. The glucosinolate profile of Sha is similar to that previously published for the Cape Verdi Islands (Cvi-1) accession, which forms predominantly three carbon (C3) and four carbon (C4) alkenyl glucosinolates, with high total aliphatic glucosinolate content (Tables 1, S1, and S2) [19,30]. In contrast, Bay-0 resembles Landsberg *erecta* (*Ler*), which contains mostly C3 hydroxy glucosinolates, with lower total aliphatic glucosinolate content (Table 1) [19,30]. The parental accessions also differed in partitioning of indolic glucosino-

lates into different structures, with Bay-0 producing significantly more 4-methoxy-indol-3-ylmethyl glucosinolate (Table 1).

We measured the average glucosinolate content within the Bay-0 × Sha RILs and compared the trait distribution among 403 RILs to the Bay-0 and Sha parental means (Table 1). Some RILs accumulated two aliphatic glucosinolates (4-methylsulfinylbutyl [4-MSO] and 4-methylthiobutyl [4-MT]) that are not found in the parental accessions. Transgressive segregation for this biosynthetic capability was previously observed in the *Ler* × *Cvi* RIL population and shown to result from epistasis between the *GSLAOP* and *GSL.Elong* loci [6,30–32,35–37]. In the Sha parent, the AOP2 enzyme fully converts all 4-MSO into but-3-enyl glucosinolate, preventing the detectable accumulation of 4-MSO within Sha. In Bay-0, 4-MSO does not accumulate due to the Bay-0 allele at *Elong*, preventing the formation of 4C glucosinolates. Plants containing the *GSLAOP<sup>OHP</sup>* (*AOP3*) allele from the Bay-0 parent and the *GSL.Elong<sup>CA</sup>* (*MAM1*) allele from Sha accumulate 4-MSO and 4-MT because the AOP3 enzyme expressed by the *GSLAOP<sup>OHP</sup>* allele from Bay-0 can not convert the 4-MSO precursor to hydroxyl glucosinolates (Figure 1) [31].

In addition to transgressive segregation for biosynthetic potential, there is transgressive segregation for glucosinolate levels. For this population, the transgressive segregation for the quantity of glucosinolates produced is almost entirely negative, as the RIL population includes numerous lines producing less total aliphatic glucosinolate than either parental accession, but no lines that accumulate average levels higher than the Sha parent (Table 1). This is especially striking for total indolic glucosinolate content, where all of the RILs were significantly lower than both parents. Because the Bay-0 and Sha parents were grown concurrently with the RILs, this is not due to environmental effects. This observation of negative transgressive segregation contrasts with the *Ler* × *Cvi* RIL population, where both positive and negative transgressive segregation was observed, with RILs producing both greater and lesser quantities of glucosinolates than either parental accession [30]. Given that the *GSLAOP* and



**Figure 2.** Estimated Heritability for Different Glucosinolate Traits

Heritability of glucosinolate traits for both the aliphatic and indolic glucosinolate pathways as estimated from 403 Bay-0 × Sha RILs. These include the Content = content of individual glucosinolates, Summation = the summation of various glucosinolates, Ratio = the ratio of glucosinolates to each other, and Gene Expression = the expression of glucosinolate biosynthetic genes. Heritability for Gene Expression was estimated using the RMA obtained expression values. Error bars show standard error of the mean. doi:10.1371/journal.pgen.0030162.g002

*GSL.ELONG* loci are in common between both populations, this difference in transgressive segregation suggests the influence of additional QTLs that are not variable in both populations.

### Heritability

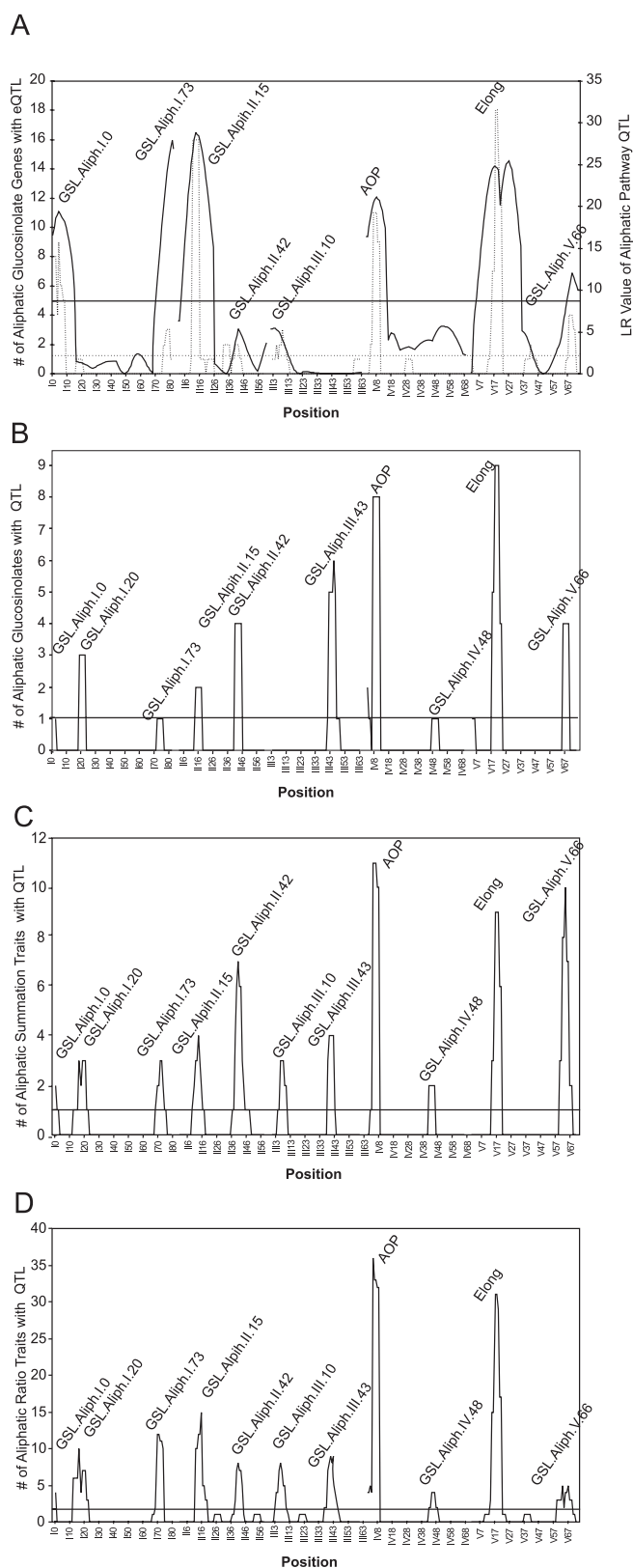
To compare the underlying genetics controlling metabolite variation with variation in gene expression, we estimated the heritability of individual glucosinolate metabolic traits and total glucosinolate content traits, as well as transcript levels for glucosinolate biosynthetic genes (Figure 2 and Tables S1 and S2). Heritability estimates for glucosinolate content traits were significantly lower than the heritability of RMA (robust multichip analysis) estimated transcript levels of their respective biosynthetic genes. This was true for both indolic and aliphatic glucosinolates (Figure 2). Differences in heritability of metabolite and expression traits could arise from differences in population size used for these estimates (403 lines for metabolites and 211 lines for transcript levels). However, recalculating heritabilities for the glucosinolate metabolites using the same 211 RILs measured in the eQTL analysis did not significantly change the values (unpublished data).

### Aliphatic Glucosinolate QTLs

Analysis of aliphatic glucosinolate variation among the RILs identified 11 QTLs that control either aliphatic glucosinolate content or the partitioning of aliphatic glucosinolates into particular structures (Figure 3). The QTLs affecting the largest number of aliphatic glucosinolate traits and causing the largest phenotypic differences were the previously identified *GSL.AOP* and *GSL.Elong* loci (Tables S4 and S5 and Figure 1). Polymorphism at these QTLs altered aliphatic glucosinolate content as well as derived ratio and summation traits. The *GSL.ALIPH.II.42* and *GSL.ALIPH.V.66* QTLs altered individual glucosinolate content and summations but did not have as dramatic an effect on the glucosinolate ratios (Figures 3 and 4). In contrast, *GSL.ALIPH.I.0*, *GSL.ALIPH.III.10*, and *GSL.ALIPH.IV.48* QTLs were

more specific to the derived summation and ratio traits than to the raw glucosinolate metabolite accumulation (Figure 3 and Tables S4 and S5). For example, the *GSL.ALIPH.IV.48* QTL influences only 3C aliphatic glucosinolate accumulation, and the *GSL.ALIPH.I.0* QTL controls ratio traits involving 4-MT and 8C glucosinolate (Figure 4 and Tables S4 and S5). These QTLs highlight the ability of derived traits to provide unique insights into metabolic variation.

Mapping eQTLs controlling transcript levels for the known aliphatic glucosinolate biosynthetic genes identified eight eQTL clusters that all coincided with aliphatic glucosinolate metabolite QTLs (*GSL.ALIPH.I.0*, *GSL.ALIPH.I.73*, *GSL.ALIPH.II.15*, *GSL.ALIPH.II.42*, *GSL.ALIPH.III.10*, *GSL.AOP*, *GSL.Elong*, and *GSL.ALIPH.V.66*; Figures 3 and 4 and Table S6). Two of these eQTLs, *GSL.AOP* and *GSL.Elong*, are known to be *cis*-eQTLs controlling the expression of four biosynthetic genes [31]. However, these loci appear to modify in *trans* the transcript levels for a broad set of aliphatic glucosinolate biosynthetic genes. Conducting a network expression analysis using an updated gene list for the aliphatic biosynthetic pathway showed that six of these eQTL clusters (including *GSL.AOP* and *GSL.Elong*) were also detected using the network mean z-score for the aliphatic glucosinolate biosynthetic gene network. The network mean z-score is a derived trait obtained by averaging across the aliphatic glucosinolate transcripts within a RIL (Table S7) [18]. As the aliphatic glucosinolate transcripts are believed to participate in a highly coregulated network, this derived trait is potentially informative regarding network control [18,38,39]. These six network eQTLs appear to control transcript levels of the pathway in *trans*, but whether *trans* functionality occurs via metabolite feedback or transcriptional mechanisms remains to be elucidated (Figure 4). The eQTLs detected in the network analysis predominantly controlled biosynthetic enzymes acting early in the aliphatic glucosinolate pathway—in the elongation and core biosynthetic stages—and not the secondary modification stages (Figures 4 and S1). Of the other three metabolite QTLs that did not show a network eQTL, the *GSL.ALIPH.I.20* QTL overlapped with a *cis*-eQTL



**Figure 3. QTL Summary for Aliphatic Glucosinolates**

QTL position for aliphatic glucosinolate biosynthetic gene expression estimates and aliphatic glucosinolate metabolite accumulation is shown on the x-axis with the Roman numeral representing the chromosome number (I–V), while the arabic numeral shows the cM position on that chromosome. All five *Arabidopsis* chromosomes are represented contiguously. Names of the QTL positions that were shown by ANOVA

to be statistically significant are included within the figure for reference. (A) The left-hand y-axis and the dotted line show the number of genes in the aliphatic glucosinolate biosynthetic pathway (22 total) that have an eQTL at a given position as determined within QTL Cartographer. The right-hand y-axis and solid line shows the likelihood ratio (LR) trace for direct QTL analysis of the average transcript level across the genes for the aliphatic glucosinolate biosynthetic pathway as estimated using the mean z-score approach. The dashed and solid horizontal lines show significance thresholds ( $\alpha = 0.05$ ) as estimated by 1,000 permutations. Breaks in lines show the end of chromosomes.

(B) The QTLs for the nine aliphatic glucosinolate metabolites were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

(C) The QTLs for the 11 derived summation aliphatic glucosinolate metabolite traits were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

(D) The QTLs for the 39 derived ratio aliphatic glucosinolate metabolite traits were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

doi:10.1371/journal.pgen.0030162.g003

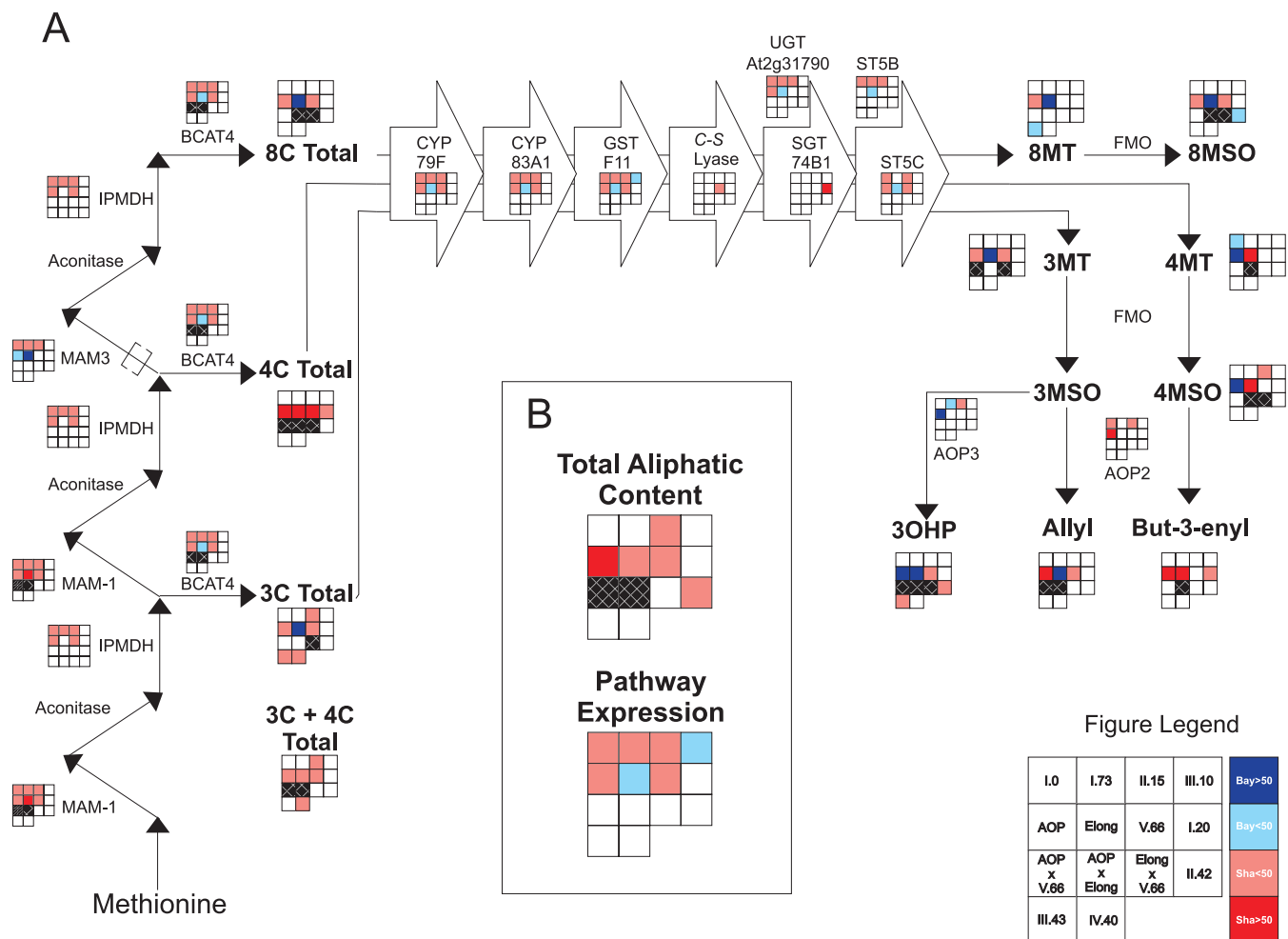
for *UGT74B1*, and the other two did not coincide with loci affecting the expression of any known biosynthetic genes.

### Enzyme-Encoding Genes and Network eQTLs

Variation at the side-chain-modifying *GSLAOP* and *GSL.Elong* loci controlled the accumulation of most metabolites and transcripts (Figures 3 and 4). Neither locus had been previously identified as impacting transcript accumulation for the whole glucosinolate biosynthetic network. Both *GSLAOP* and *GSL.Elong* are controlled by *cis*-eQTLs leading to differential enzyme expression. As in *Cvi*, the Sha *GSLAOP* locus expresses the *AOP2* enzyme, leading to alkenyl glucosinolate production, higher glucosinolate content, and elevated transcript levels for most aliphatic glucosinolate genes (Figure 4 and 5). We used Col-0 (which is null for *AOP2* and *AOP3*) transformed with a functional *AOP2* gene from *Brassica oleracea* to test if the presence of functional *AOP2* transcript can affect both metabolite and transcript levels [40]. *AOP2* from *B. oleracea* conducts the same reaction as *AOP2* from Sha and is also associated with elevated aliphatic glucosinolate content in *Brassica*, allowing us to test the conservation of this locus across the two species [40,41].

Introduction of a transcript encoding functional *AOP2* results in the production of alkenyl glucosinolates and a statistically significant doubling of total aliphatic glucosinolate content, as is the case with the presence of a functional *AOP2* transcript contributed by the Sha allele at the *GSLAOP* QTL (Figure 5). The introduction of the *AOP2* transcript also leads to induction of 17 of 22 aliphatic glucosinolate biosynthetic genes and three of seven regulatory genes (Figure 5). This supports the hypothesis that the Sha allele at the *GSLAOP* QTL controls metabolite and transcript levels for aliphatic glucosinolates due to increased expression of the *AOP2* gene. This suggests the presence of a previously unrecognized regulatory effect of *AOP2*, whereby it controls transcript accumulation for most biosynthetic genes potentially through transcription factors. While we could not detect any micro-RNA signatures within the *AOP2* transcript or gene, it remains to be shown whether the metabolite and transcript effect is due to the enzymatic activity of *AOP2* or





**Figure 4.** Pathway Summary of Aliphatic Glucosinolate QTLs

Results of ANOVA testing of all identified aliphatic glucosinolate QTLs for significant impact upon the accumulation of individual glucosinolate metabolites, transcript level of all biosynthetic genes, total aliphatic glucosinolate content, and the average expression of the aliphatic glucosinolate biosynthetic pathway is presented graphically. The genes and metabolites are shown with respect to the currently theorized biosynthetic pathway. Trait abbreviations are as listed in Table S1. Gene names are as listed in Table S3, and TAIR locus identifiers are used for gene families where there is no settled naming system.

Cells within boxes represent aliphatic glucosinolate QTLs. The legend at the bottom right contains the QTL name. Cells representing QTLs significantly controlling the represented trait are colored to show the directionality of the allele substitution effect; a positive effect of the Bay-0 allele is blue, and a positive effect of the Sha allele is red. Dark red and dark blue show that the allele substitution at the given QTL led to greater than 50% phenotypic change in the trait, while the lighter colors represent QTLs of smaller phenotypic effect. Significant epistasis between the *GSL.Aliph.AOP*, *GSL.Aliph.Elong*, and *GSL.Aliph.V.66* QTLs are shown by black cross-hatching within the respectively labeled cell. For example, but-3-enyl is controlled by QTL at *GSL.AOP*, *GSL.Elong*, and *GSL.I.20* with epistasis between the *GSL.AOP* and *GSL.Elong* loci. QTLs for gene expression are shown in smaller font with a smaller ANOVA box, while QTLs for metabolites are shown in bold larger font with a larger box.

(A) QTLs for the whole pathway broken down into individual metabolites and transcripts.  
(B) QTLs for total aliphatic glucosinolate content and the mean z-score for the biosynthetic pathway.

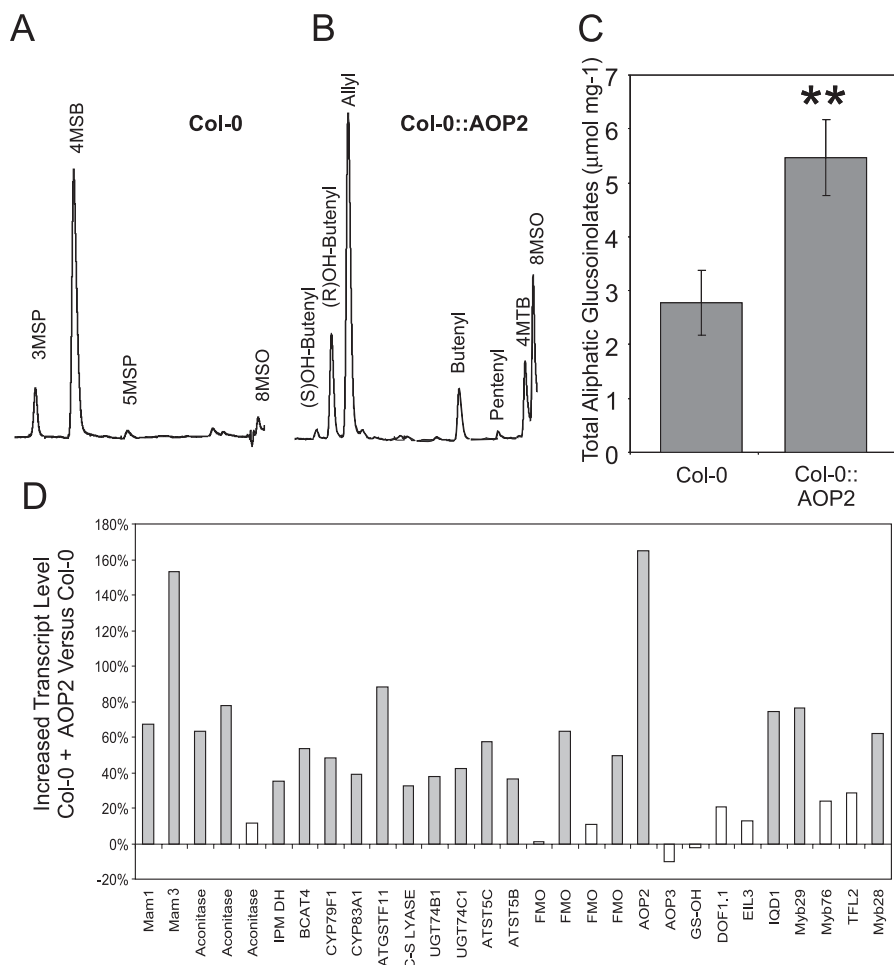
doi:10.1371/journal.pgen.0030162.g004

some other regulatory signature within the *AOP2* transcript. The association of both *GSL.AOP* and *GSL.Elong* with eQTLs for the majority of aliphatic glucosinolate biosynthetic genes and metabolites (Figure 4) suggests a regulatory interplay between the metabolites directly synthesized by these enzymes and transcript levels for the aliphatic glucosinolate biosynthetic genes.

### Epistasis for Transcripts versus Metabolites

We tested the identified QTLs for pairwise epistatic interactions controlling metabolite accumulation, partitioning, or transcript levels. This analysis identified at least one pairwise epistatic QTL interaction for all metabolites, with

variation in at least half of the metabolites controlled by interactions between *GSL.AOP*, *GSL.Elong*, and *GSL.V.66* (Figure 4). The most common pairwise interaction was detected between *GSL.AOP* and *GSL.Elong*, controlling 7 of 9 aliphatic glucosinolate metabolites (Figure 4). In contrast to the metabolites, most expression traits did not identify epistatic eQTL interactions. The few transcripts traits that identified epistatic eQTL interactions encode biosynthetic enzymes functioning in the early steps of the elongation cycle: *MAM1*, *BCAT4*, and an *Aconitase*. This suggests that genes in the elongation cycle may be regulated differently from the rest of the aliphatic glucosinolate pathway genes (Figures 1, 4, and S1).



**Figure 5.** AOP2 Transcript Alters Metabolic Profiles, Content, and Gene Expression

Wild-type Col-0 that is null for *AOP2* and *AOP3* was modified through the introduction of a functional *AOP2* transcript from *B. oleracea*. All glucosinolate abbreviations are as described in Table S1.

(A) HPLC profile of aliphatic glucosinolates detected in foliar tissue of wild-type Col-0.

(B) HPLC profile of aliphatic glucosinolates detected in foliar tissue of Col-0 containing the functional *AOP2* transcript.

(C) Average total foliar aliphatic glucosinolate content in Col-0 and Col-0::*AOP2* is shown with standard error bars. Six plants per genotype were measured for total aliphatic glucosinolate content within an experiment, and the experiment was conducted twice to provide 32 total measurements. \*\* $p < 0.0001$  as determined by ANOVA.

(D) Percentage increase in transcript levels in Col-0::*AOP2* as compared with Col-0 is presented. RNA from 3 plants per genotype were individually hybridized to ATH1 Affymetrix arrays to obtain transcript levels for the aliphatic glucosinolate genes. ANOVA was used to test for significant differences between the two genotypes for the glucosinolate biosynthetic and regulatory genes via ANOVA with a false discovery rate of 0.05. Gray bars show transcripts significantly increased by the introduction of the *AOP2* transcript. Nonsignificant changes are shown in white bars. doi:10.1371/journal.pgen.0030162.g005

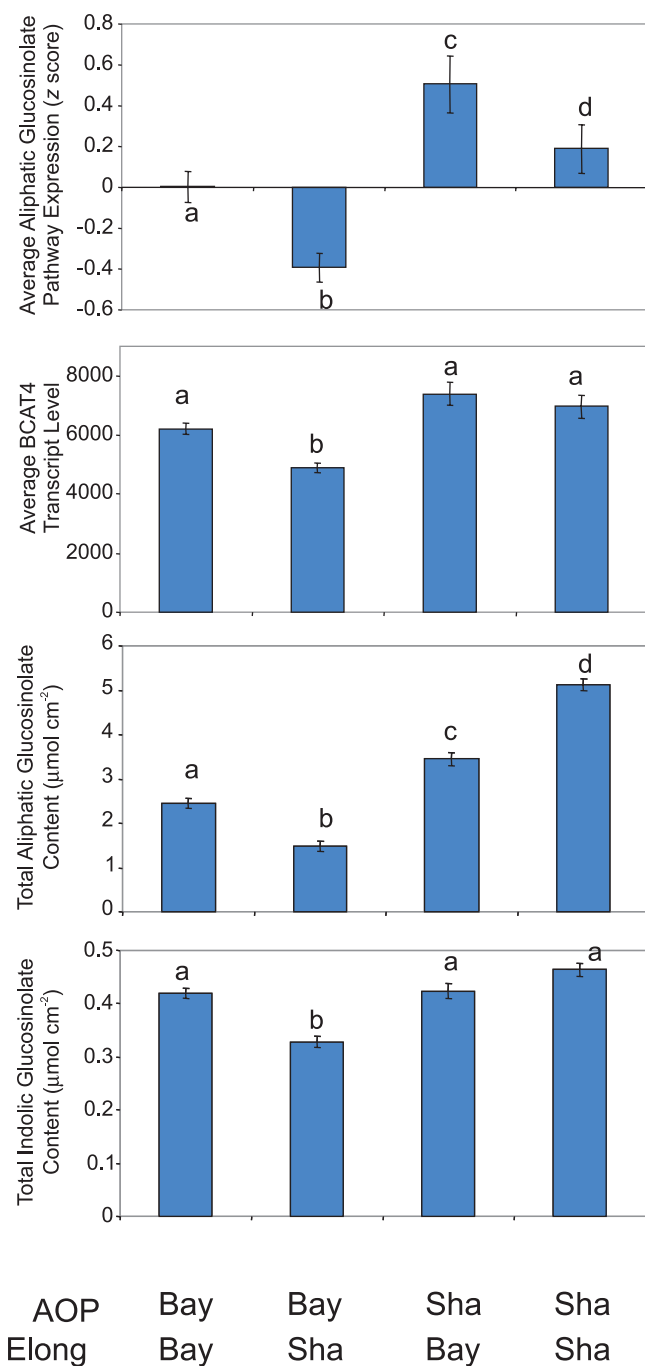
To investigate the nature of the epistatic interaction between *GSLAOP* and *GSL.Elong*, we calculated mean phenotypic values for the RILs containing each of the allelic combinations at these two loci (Figure 6). *GSLAOP* × *GSL.Elong* interaction had a negative epistatic effect on the total content of both aliphatic and indolic glucosinolates, with lines possessing the nonparental allelic combination of *GSLAOP*<sup>OHP</sup> from Bay-0 and *GSL.Elong*<sup>CA</sup> from Sha exhibiting significantly lower glucosinolate content than either parent (Figure 6). Lines possessing Sha alleles at both loci had the highest glucosinolate accumulation.

The *GSLAOP* and *GSL.Elong* loci also control the network expression mean z-score for the aliphatic biosynthetic gene network, but did not exhibit a pairwise epistasis for this trait (Figure 4). In contrast, the genes in the methionine elongation cycle did identify an epistatic interaction between

*GSLAOP* and *GSL.Elong* (Figure 6; *BCAT4* is shown as an example). This lack of an epistatic effect on aliphatic glucosinolate network expression is not likely a statistical artifact, as the pattern of the group means shows a striking difference between aliphatic glucosinolate accumulation and network expression of the aliphatic biosynthetic genes (Figure 6). Substitution of the Bay-0 *GSL.Elong*<sup>C3</sup> allele for the Sha allele in a background containing the Sha *GSLAOP*<sup>Alk</sup> allele leads to increased accumulation of aliphatic glucosinolate biosynthetic transcripts but lower aliphatic glucosinolate content (Figure 6). This suggests that these two loci regulate both transcript and metabolite accumulation via distinct mechanisms.

#### Indolic Glucosinolate QTLs

We analyzed the indolic glucosinolate pathway as a second test of our ability to link eQTLs altering transcript levels for



**Figure 6.** AOP and Elong Epistasis

The Bay-0 × Sha RILs were grouped by their *GSLAOP* and *GSL.Elong* genotypes. The mean ( $\pm$  standard error) for each group is presented for: (A) Mean z-score for the aliphatic glucosinolate biosynthetic pathway. (B) Transcript level for the *BCAT4* gene. (C) Total aliphatic glucosinolate content. (D) Total indolic glucosinolate content. doi:10.1371/journal.pgen.0030162.g006

biosynthetic genes with metabolite accumulation QTLs. A total of 13 QTLs were identified that control the accumulation of indolic glucosinolates, their partitioning into particular structures, or both (Figure 7 and Table S8). These loci affect one or more of the indolic glucosinolate traits in this population, with the directionality of allelic effects mixed, so that Bay-0 alleles at some loci increase trait values

while the Bay-0 alleles at other loci decrease the metabolite trait values (Table S8). Interestingly, the *GSL.INDOLIC.IV.8* and *GSL.INDOLIC.V.20* QTLs map to the same genomic locations as the previously known *GSLAOP* and *GSL.ELONG* loci, which control aliphatic glucosinolate variation. Transgenic analysis confirmed that simulating the Sha *GSLAOP* allele by introducing the *AOP2* gene into a null background increases total indolic glucosinolate content by about 30% ( $p = 0.035$ ). This shows that variation at *GSLAOP* affects indolic glucosinolate metabolism and that there is cross-talk between the pathways for indolic and aliphatic glucosinolate production.

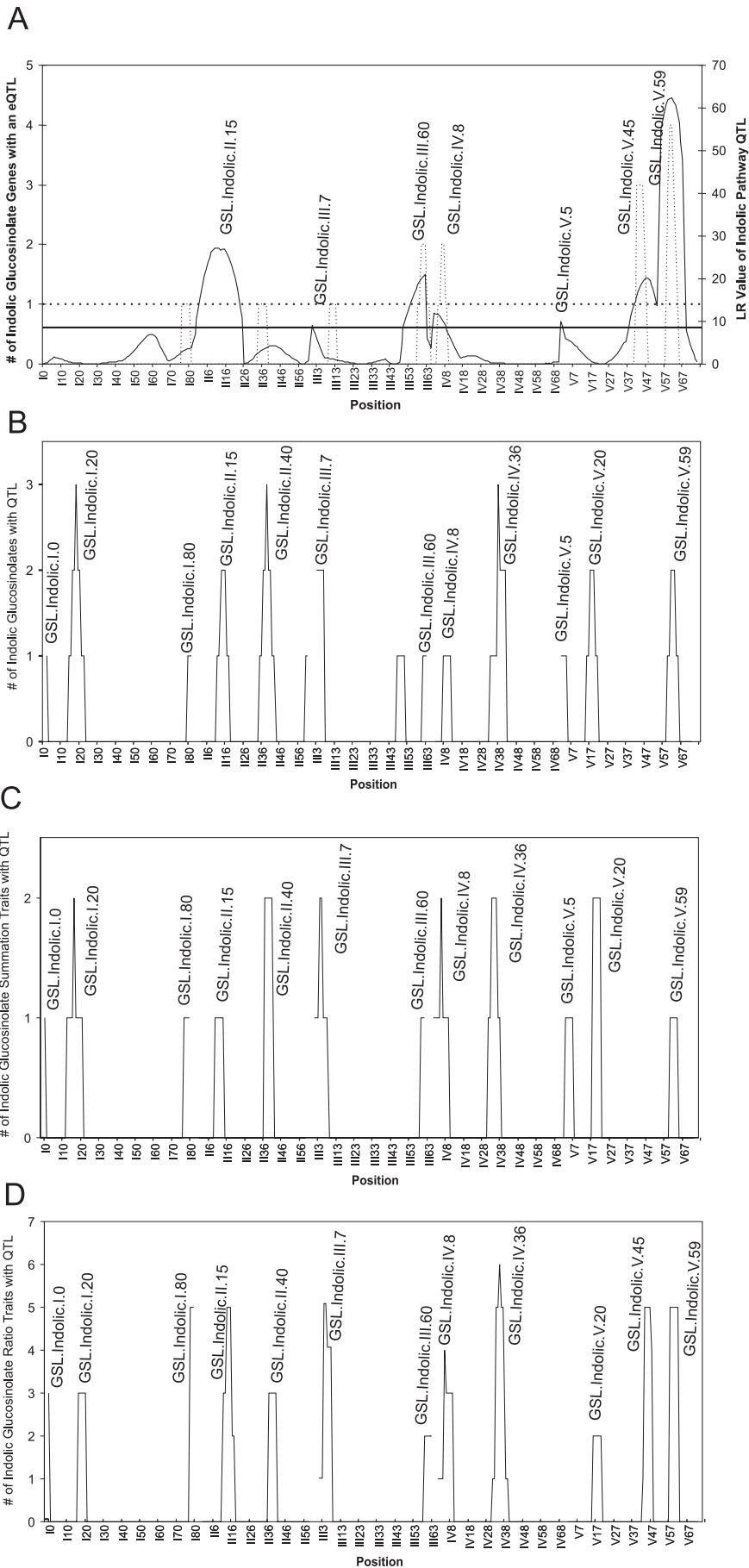
Testing the identified indolic metabolite QTLs for pairwise epistatic interactions identified numerous epistatic interactions affecting indolic glucosinolate accumulation. Notably, the *GSL.INDOLIC.II.15* QTL was detected as epistatically interacting with five other QTLs (*GSL.INDOLIC.III.7*, *GSL.INDOLIC.III.3.60*, *GSL.INDOLIC.IV.4.36*, *GSL.INDOLIC.V.45*, and *GSL.INDOLIC.V.59*; Table S8). These epistatic interactions affect the partitioning of the indolic glucosinolates into two distinct methoxylated derivatives without significantly altering the total indolic glucosinolate content. *GSL.INDOLIC.II.15* might encode or regulate enzymes responsible for this methoxylation (Table S8). A regulation hypothesis is supported by the fact that the *GSL.INDOLIC.II.15* region contains a massive *trans*-acting eQTL that influences transcript levels for more than 5,000 genes [17].

Analysis of eQTLs controlling transcript levels of individual indolic glucosinolate biosynthetic genes identified four eQTL clusters (*GSL.INDOLIC.III.60*, *GSL.INDOLIC.IV.8*, *GSL.INDOLIC.V.45*, and *GSL.INDOLIC.V.59*; Figure 7 and Tables S6 and S7). Because the indolic glucosinolate biosynthetic genes are also believed to be coregulated, we estimated pathway expression mean z-value to map network QTLs. This identified the same four loci, plus three additional QTLs affecting expression of the indolic gene network (*GSL.INDOLIC.II.15*, *GSL.INDOLIC.III.7*, and *GSL.INDOLIC.V.5*). All three of these network-specific eQTLs colocalized with metabolite QTLs, supporting the ability of the z-scale network approach to derive biological information. All seven eQTLs colocalize with loci that control either the accumulation or partitioning of the indolic glucosinolates produced by these genes. This suggests that the eQTLs controlling transcript levels for the indolic glucosinolate biosynthetic genes also affect variation in their metabolite products. Again, it remains to be tested whether the QTLs primarily affect transcript levels, causing downstream metabolite effects, or if the polymorphisms first affect the metabolites, influencing transcript levels through some form of feedback. There are also indolic metabolite QTLs, such as *GSL.INDOLIC.IV.36*, with no detectable impact on gene expression traits.

## Discussion

We used the *Arabidopsis* aliphatic and indolic glucosinolate biosynthetic pathways to test our ability to link QTLs controlling metabolic and expression polymorphisms. These metabolic pathways are well characterized with near complete identification of most biosynthetic enzymes and genes, as well as detailed quantitative genetic analysis [25,27]. A direct comparison of metabolite QTL and eQTL maps reveals





**Figure 7.** QTL Summary for Indolic Glucosinolates

QTL positions for indolic glucosinolate biosynthetic gene expression estimates and indolic glucosinolate metabolite accumulation were determined within QTL Cartographer. The position is shown on the x-axis, with the Roman numeral representing the chromosome number (I–V) while the Arabic numeral shows the cM position on that chromosome. All five chromosomes are represented contiguously. Names of the QTL positions that were shown by ANOVA to be statistically significant are shown within the figure for reference.

(A) The left-hand y-axis and the dotted line show the number of genes in the indolic glucosinolate biosynthetic pathway (six total) that have an eQTL at a given position as determined within QTL Cartographer. The right-hand y-axis and solid line shows the likelihood ratio (LR) trace for direct QTL analysis of the average transcript level across the genes for the indolic glucosinolate biosynthetic pathway as estimated using the z-score approach. The dashed and solid horizontal lines show significance thresholds ( $\alpha = 0.05$ ) as estimated by 1,000 permutations for eQTL hotspots and the pathway expression QTLs, respectively.

(B) The QTLs for the three indolic glucosinolate metabolites were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

(C) The QTLs for the two derived summation indolic glucosinolate metabolite traits were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

(D) The QTLs for the six derived ratio indolic glucosinolate metabolite traits were mapped independently using the average across all experiments as well as the average within each experiment. QTL positions for all indolic glucosinolate traits were then summed across experiments to identify metabolite QTL “hotspots.”

doi:10.1371/journal.pgen.0030162.g007

that for both pathways, all eQTLs collocate with metabolite QTLs (Table 2, Figures 2 and 7). This suggests that expression differences for genes comprising these two pathways are predictive of phenotypic differences. However, the predictive capacity of eQTLs is not perfect, as indicated by the presence of metabolite-specific QTL clusters that do not associate with eQTL clusters.

Metabolite-specific QTL clusters cannot be explained by the superior power of the metabolite QTL analysis, which used 403 RILs versus the 211 phenotyped for transcript level, as restriction of the metabolite analysis to the set of 211 RILs included in the expression QTL analysis did not alter this result (unpublished data). In addition, higher heritability of transcript levels versus metabolite traits creates a higher statistical likelihood of failure to detect metabolite QTLs than eQTLs (Figure 2). Although the eQTL and metabolite QTL analysis were done at different times, all experiments

used seed from the same mothers, and were grown in the same growth chamber under similar growth conditions. As such, environmental variance was minimized and is unlikely to explain this result.

We therefore feel that biological processes, rather than statistical or experimental considerations, explain these metabolite-specific QTLs. Metabolite-specific QTLs could be caused by metabolic feedback mechanisms, whereby a product is responsible for fine-tuning the activity of a rate-limiting enzyme in the biosynthetic pathway. Alternatively, metabolite-specific QTLs may identify loci controlling flux of substrates into the pathway. In addition, polymorphisms that affect post-translational regulation or structural variation in enzymes could lead to metabolite QTLs independent of loci controlling transcript levels. A similar result would be observed if the metabolite QTLs were caused by variation in other biosynthetic pathways that affect substrate availability, or by eQTLs affecting an unknown biosynthetic gene. Segregating variation in pathways responsible for the catabolic turnover of glucosinolates would also affect metabolite accumulation without affecting biosynthetic gene expression. The presence of QTLs that exclusively affect metabolites and not expression traits illustrates the complexity involved in extrapolating variation in gene expression to changes in metabolite content.

**Network eQTL Identity**

The three most consistently detected QTLs for aliphatic glucosinolate content and transcript variation were *GSL.AOP*, *GSL.Elong*, and *GSL.V.66*. Two of these network QTLs, *AOP* and *Elong*, contain *cis*-eQTLs controlling variation in transcript levels for enzymes located at the beginning and end of the aliphatic glucosinolate pathway. Neither *AOP* nor *Elong* has previously been associated with regulation of other glucosinolate biosynthetic or regulatory genes. This suggests that the accumulation of aliphatic glucosinolates is regulated by multiple mechanisms functioning such that enzyme variation can feed back to alter transcript accumulation (Figure 1). The *GSL.AOP* QTL is likely controlling glucosinolate content and expression through natural variation in the *AOP2* gene (Figure 5). Altering the expression of the *AOP2* transcript, the enzyme at the end of the aliphatic glucosinolate pathway might pull carbon into aliphatic glucosinolate production (Figures 4–6). In addition, introducing the tran-

**Table 2.** Shared and Unique QTLs

QTL	Aliphatic		Indolic	
	Met	Exp	Met	Exp
I.0	Yes	Yes	Yes	–
I.20	Yes	–	Yes	–
I.73	Yes	Yes	–	–
I.80	–	–	Yes	–
II.15	Yes	Yes	Yes	Yes
II.40	Yes	–	Yes	–
III.10	Yes	Yes	Yes	Yes
III.43	Yes	–	–	–
III.60	–	–	Yes	Yes
AOP	Yes	Yes	Yes	Yes
IV.36	–	–	Yes	–
IV.48	Yes	–	–	–
V.5	–	–	Yes	Yes
Elong	Yes	Yes	Yes	–
V.45	–	–	Yes	Yes
V.59	–	–	Yes	Yes
V.66	Yes	Yes	–	–

The identified and named QTL are listed in chromosomal order.

A “Yes” indicates if a metabolic QTL (Met) or network eQTL or eQTL cluster (Exp) was identified for the given glucosinolate class. A dash indicates no evidence for a QTL for this glucosinolate class in this region. AOP and Elong refer to *GSL.Indolic.IV.8* and *GSL.Indolic.V.20*, respectively, as these map to the same location.

doi:10.1371/journal.pgen.0030162.t002

script for *AOP2* increases transcript level for the aliphatic biosynthetic pathway, suggesting that regulation also occurs by control of gene expression and potentially metabolic fluxes through the beginning of the aliphatic glucosinolate pathway (Figures 5 and 6). It is possible that a metabolite produced by or used by the *AOP2* enzyme may have the capacity to regulate transcript accumulation for the rest of the biosynthetic network through feedback. Direct regulation of transcript accumulation by metabolites has been noted for a variety of riboswitches [42–45].

For the *GSL.Elong* network QTLs, the comparison of eQTLs to metabolite QTLs suggests that variation in transcript levels for at least two biosynthetic genes (*MAMI* and *MAM3*) at *GSL.Elong* directly causes metabolic variation, and that this biosynthetic variation may alter transcript regulation for nearly the entire biosynthetic pathway. Specifically, lines that contain 4C glucosinolates have lower transcript levels for almost all biosynthetic genes. One possibility is that a metabolic intermediate produced in the *GSL.Elong*<sup>4C</sup> background negatively regulates gene expression. Alternatively, both the *GSLAOP* and *GSL.Elong* loci could possess tightly linked second loci that cause the observed transcriptional polymorphisms and also interact epistatically. Testing the interaction between biosynthetic loci and transcript levels will require careful genetic manipulation with full transcriptome analysis.

Comparing the remaining network eQTLs to candidate glucosinolate transcriptional regulators shows that the *GSL.ALIPH.V.66* QTL overlaps the physical position of the *MYB28* transcription factor, which has a large-effect *cis*-eQTL [17,39]. Likewise, the *GSL.INDOLIC.V.45* and *GSL.INDOLIC.V.59* QTLs overlap the physical position of the indolic transcription factors, *ATR1* and *ATR2*, and both genes have *cis*-eQTLs [17,29,46]. This suggests that these three QTLs may be explained by variation in the expression of these known regulatory genes. In addition, the *GSL.ALIPH.I.0* and *GSL.ALIPH.III.10* QTLs overlap with the physical position for the aliphatic glucosinolate transcriptional regulators *AtDOF1.1* and *IQD1*, respectively [28,47]. Neither gene exhibits a *cis*-eQTL, suggesting that if these genes are responsible for *GSL.ALIPH.I.0* and *GSL.ALIPH.III.10* QTLs, it is potentially due to an activity polymorphism. Alternatively, small changes in transcript levels for these transcription factors might lead to large changes in network regulation (Figure S1).

The *GSL.ALIPH.II.15* QTL, a major network eQTL for transcript levels for both aliphatic and indolic glucosinolate pathways, does not overlap any known biosynthetic or regulatory genes. This locus does colocalize with a massive eQTL cluster controlling the expression of several thousand genes [17]. This suggests that this region may be highly pleiotropic and its effects on glucosinolate content may be indirect. These results suggest that eQTLs can control metabolite production through a variety of direct and indirect regulatory mechanisms.

### *cis*-eQTLs and QTLs for Specific Metabolites

While most eQTLs co-located with QTL clusters controlling several metabolites, there were also instances in which a statistically significant association between expression and metabolic phenotypes was limited to one or few metabolic traits. For example, *SGT74B1* catalyzes glycosylation of the characteristic glucosinolate backbone structure, and its

expression is controlled by a single eQTL at *GSL.ALIPH.II.20* [48]. This large-effect *cis*-eQTL maps to a 100-kb interval that includes the physical position of *SGT74B1*. While we might predict that variation in the expression of *SGT74B1* would influence production of multiple aliphatic glucosinolates, this locus was only identified as controlling one metabolic trait, the accumulation of but-3-enyl glucosinolate. This gene is not specific to the synthesis of but-3-enyl glucosinolate, as previous work has demonstrated that *SGT74B1* has a broad biochemical capacity to glucosylate glucosinolates [48]. Instead, the lack of detected effects in this study for this eQTL on the accumulation of other glucosinolates is likely because genotypes accumulating but-3-enyl glucosinolate also exhibit the highest level of total aliphatic glucosinolates within this population (Table 1). This suggests that the *SGT74B1* expression polymorphism is only limiting when flux across the biosynthetic pathway is maximized. Consequently, unexpected factors directly related to biochemical pathway connectivity and flux can interfere with our ability to directly associate eQTLs for biosynthetic genes with specific metabolite QTLs. As such, eQTLs are not predictive in all contexts.

### QTL Causality

This comparative analysis of eQTLs to metabolic QTLs has provided novel insights, including the identification of the enzymes *AOP2* and *MAMI* as well as the transcription factor *MYB28* as potential regulators of transcript accumulation for the complete aliphatic glucosinolate biosynthetic pathway. In addition, differential expression of the *ATR1* and *ATR2* transcription factors have been implicated as the underlying cause of QTLs controlling the indolic glucosinolate pathway. Data presented in this study validate the potential of the enzyme *AOP2* to control aliphatic glucosinolate gene expression. The other regulatory roles hypothesized above remain to be verified. Identification and validation of the molecular causes of the 17 different QTLs identified in this study will require a complex mixture of experiments ranging from transgenic complementation to validate the gene, promoter swaps to validate the difference in gene expression, and recombination mapping to more precisely identify the causal polymorphism [49].

### Trait Heritability and Epistasis

For glucosinolates and their biosynthetic genes, we observed significant differences between the estimated heritability of metabolite accumulation and transcript levels, respectively (Figure 2). Expression traits had consistently higher broad sense heritability than the accumulation of individual metabolites or metabolite summation traits. Because genetically identical individuals were used for both experiments, there is no difference in the amount of genotype variation available to transcript and metabolite traits. Thus, environmental inputs may affect metabolic traits more strongly than transcripts. The mechanistic link between sequence polymorphism and variation in transcript levels may have fewer intervening processes than a causal sequence polymorphism and metabolic variation. The presence of additional regulatory processes, both metabolic and post-transcriptional, may allow greater environmental heterogeneity effects on metabolite accumulation. It is also possible that the results obtained for the glucosinolate pathway are not indicative of typical transcript or metabolite heritability.

There is evidence that glucosinolate production is under diverse selection pressures, favoring high levels of plasticity in glucosinolate accumulation mediated by environmental stimuli such as nutrient and water availability and wounding [50–52]. This would require that metabolite traits be influenced by subtle environmental heterogeneity, leading to reduced estimates of their heritability. It is therefore important to expand this analysis to other metabolic pathways to determine the extent to which these conclusions are generalizable.

We also observed that metabolic traits identified significantly more epistatic interactions than the corresponding transcripts. Regulatory processes that occur between transcript accumulation and metabolite accumulation, e.g., metabolic feedback, post-transcriptional regulation, and enzyme activity regulation, may increase regulatory interactions between loci, leading to metabolite traits showing higher levels of epistasis. The constant adjustment of metabolite flux through complex networks may also enhance the potential for epistasis. This finding may be specific to the glucosinolate system. A broader metabolomics analysis in comparison to eQTLs for all known enzymatic loci will be required to test whether these differences in heritability and epistasis are a general feature of transcriptomic and metabolomic networks. If this is the case, a detailed modeling approach may contribute to understanding differences in genetic architecture between metabolic and transcript networks.

## Conclusion

In this report, we show that it is possible to relate natural variation at the transcript and metabolite levels for two glucosinolate biosynthetic networks. Furthermore, this analysis shows that the comparison of eQTLs to metabolite QTLs within an a priori-defined framework can identify complex regulatory mechanisms whereby variation in enzymes or metabolites may feed back to alter transcript accumulation. For aliphatic glucosinolates, the beginning and end of the biosynthetic pathway interact to control the whole pathway. These feedback associations can lead to the rapid generation of new hypotheses about the regulation of biosynthetic networks, but also show that the *de novo* reconstruction of biosynthetic relationships from metabolite data will require great care. In all cases, variation in gene expression also affected the resultant metabolites, although extrapolating the effects of gene expression on metabolism requires caution due to interplay of biochemical mechanisms. Combining different genomics datasets will greatly expand our ability to understand both the regulation of metabolism and the relationship between transcription and metabolism.

## Materials and Methods

**Mapping populations.** The Bay × Sha population of 403 *A. thaliana* RILs [53] was used to map QTLs controlling individual and total glucosinolate content for both the aliphatic and indolic glucosinolates. The QTL on Chromosome V for total content of aliphatic glucosinolates in the August 2005 experiment is also presented elsewhere for clarity (Sonderby, Hansen, Halkier, and Kliebenstein, unpublished data). Further, 211 of these lines have been analyzed for variation in gene expression and used to map QTLs controlling transcript levels [17,18], allowing comparison of QTLs controlling metabolite accumulation with transcript levels for the underlying biosynthetic genes.

**Plant growth conditions.** Seeds were imbibed and cold-stratified at 4 °C for 3 d to break dormancy. Two complete plantings were grown

simultaneously in neighboring growth chambers to provide independent biological replicates for each experiment. The full experiment was replicated three times between March of 2004 and August of 2005, providing six glucosinolate measurements for most lines, totaling nearly 2,600 measurements. The replicates were labeled May 2004, May 2005, and August 2005. For the May 2005 and August 2005 experiments, plants were grown in flats with 36 cells per flat, and maintained under short-day conditions in controlled environment growth chambers. For the May 2004 experiment, plants were grown in flats with 96 cells per flat, and maintained under short-day conditions in controlled environment growth chambers. Using the same growth chambers, similar growth conditions, and assaying glucosinolate content at the same developmental stage analyzed in the eQTL mapping experiment maximizes our ability to compare the metabolic QTL results with eQTLs for the biosynthetic genes [17,18]. At 35 days after germination, a fully expanded mature leaf was harvested, digitally photographed, and analyzed for glucosinolate content as described below at the same age as the plants used for eQTL analysis [17,18].

**Analysis of glucosinolate content.** The glucosinolate content of excised leaves was measured using a previously described high-throughput analytical system [30,54]. Briefly, one leaf was removed from each plant, digitally photographed, and placed in a 96-well microtiter plate with 500  $\mu$ L of 90% methanol and one 3.8-mm stainless steel ball-bearing. Tissue was homogenized for 5 min in a paint shaker, centrifuged, and the supernatant was then transferred to a 96-well filter plate with 50  $\mu$ L of DEAE sephadex. The sephadex-bound glucosinolates were eluted by incubation with sulfatase. Individual desulfo-glucosinolates within each sample were separated and detected by high-performance liquid chromatography (HPLC)–diode-array detection, and identified and quantified by comparison to purified standards. Tissue area for each leaf was digitally measured using Image J with scale objects included in each digital image [55]. The glucosinolate traits are reported per square centimeter of leaf area. There was no significant variation detected for leaf density within these lines (unpublished data).

In addition to the content of individual glucosinolates, we developed a series of summation and ratio traits based on prior knowledge of the glucosinolate pathways (Table S1) [56]. For instance, the content of 3-MT, 3-MSO, 3-OHP, and allyl glucosinolates were summed (sum3C) to provide an estimate of the content of 3C aliphatic glucosinolates within these lines (Figure 1C and Table S1). This enables the detection of QTLs that specifically alter 3C glucosinolate accumulation irrespective of specific side-chain modification. The ratio traits were created to measure the efficiency of partitioning a class of glucosinolates into particular structures. For example, the ratio allyl glucosinolate to total 3C aliphatic glucosinolates (all\_r3) allows discrimination of the efficiency of production of 3C alkenyl glucosinolates independent of the accumulation of 3C glucosinolates (Figure 1C and Table S1). These ratios and summation traits allow us to isolate the effects of variation at individual steps of glucosinolate biosynthesis from variation affecting the rest of the biosynthetic pathway [56].

For each glucosinolate trait, we determined the average value per RIL per experiment for QTL mapping. Because there was no significant difference in the variance of the traits between the experiments, we also calculated the average value per RIL across all three experiments for all traits (Table S2). Heritability of each glucosinolate trait was estimated using the general linear model procedure within SAS (<http://www.sas.com>) where broad sense heritability was defined as  $\sigma_g^2/\sigma_p^2$  (Table S1), where  $\sigma_g^2$  is the estimated genetic variance for the metabolite among different genotypes in this sample of RILs, and  $\sigma_p^2$  is the estimated phenotypic variance for the metabolite [2].

**Analysis of gene expression QTL.** We used previously published biochemical and coexpression data to identify all known or predicted genes encoding glucosinolate biosynthetic enzymes [27,31,38,57,58]. For these purposes, the indolic and aliphatic glucosinolate pathways are considered to be independent biosynthetic processes. This appears to reflect the biological reality, as the two pathways use different genes and amino acid precursors [25,27]. Gene families were separated into genes involved in aliphatic or indolic glucosinolate pathways based on biochemical or phenotypic data where possible [59–61]. Where this was not possible, coexpression with the known indolic or aliphatic glucosinolate genes was combined with published biochemistry to separate gene family members into their respective pathways [38]. This generated a list of genes involved in aliphatic and indolic glucosinolate biosynthesis (Table S3).

Heritability, eQTL position, eQTL effect, and transcript levels for individual transcripts were obtained using the RMA estimated

expression values from the previously published analysis of gene expression in the Bay  $\times$  Sha population [17]. To conduct network expression analysis, transcript levels for each biosynthetic gene within each RIL were standardized as z-scores. This is done by taking the transcript level for a gene within a RIL, subtracting the mean transcript level for that gene among the RILs, then dividing the resulting value by the standard deviation for that transcript among the RILs [18]. The mean z-score for the aliphatic and indolic glucosinolate biosynthetic genes was calculated within each RIL for each replicate [18]. This pathway mean z-score per RIL per replicate was then used to estimate heritability of the aliphatic and indolic glucosinolate pathway gene expression as described for the metabolites. The mean z-score per RIL across replicates was calculated and used to map QTLs controlling transcript levels of the aliphatic and indolic glucosinolate biosynthetic pathways (Table S4). Because these global transcription studies were conducted in the same mapping population grown under the similar conditions and in the same growth chambers, it was possible to compare the expression and metabolite data.

**QTL analysis.** The Bay  $\times$  Sha RIL population has been previously genotyped [53], and additional markers were obtained from the expression QTL analysis [62] as well as markers specific for the *GSLAOP* and *GSL.Elong* loci [6,30]. To maximize our ability to detect all possible QTLs, we used the averages from each experiment, May 2004, May 2005, and August 2005, as well as the average across all experiments to conduct four QTL mapping tests. This was done for all individual glucosinolate traits, summation traits, and ratio traits (Table S1).

The four averages for each trait were independently used for QTL mapping within Windows QTL Cartographer v2.5 (<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>) [63–65]. Composite interval mapping was implemented using Zmap (Model 6 within Windows QTL Cartographer v2.5) with a 10-cM window and an interval mapping increment of 2 cM. Forward regression was used to identify five cofactors per quantitative trait. The declaration of statistically significant QTLs is based on permutation-derived empirical thresholds using 1,000 permutations for each trait mapped [66,67]. The Eqtl module of QTL Cartographer was used to automatically identify the location of each significant QTL for each trait from each experiment and the whole experiment average (Tables S5–S7) [65]. Composite interval mapping with permutations to assign significance using the underlying trait distribution is fairly robust at handling normal or near-normal traits as found for most of our traits [68]. In addition, all data from the three different experiments were used in the multi-trait composite interval algorithm within QTL Cartographer v2.5.

QTL clusters were identified by using the QTL summation approach, where the position of each QTL for each trait for each experiment is indicated by a 1, and the number of traits controlled by a QTL at a given position is totaled [18]. This summation was conducted using four groups: Group I, all aliphatic glucosinolate metabolite traits; Group II, all eQTLs for aliphatic glucosinolate biosynthetic genes; Group III, all indolic glucosinolate metabolite traits; and Group IV, all eQTLs for indolic glucosinolate biosynthetic genes. These QTL clusters identified a set of defined genetic positions that were then named respective to their position and whether they affected aliphatic or indolic glucosinolate content (Tables S5–S7). The QTLs at the previously characterized and cloned *AOP* and *Elong* loci were named as such [30–32,69].

To further validate each QTL identified and query for potential epistasis, we conducted an analysis of variance (ANOVA) using all experiments. For each trait, the markers most closely associated with each significant main-effect QTL for that trait were used as main effect cofactors. In addition, experiment (May 2004, May 2005, and August 2005) was used as a main effect cofactor. An automated SAS script was then developed to directly test all main effects as well as all possible pairwise interactions, including experiment  $\times$  marker(QTL) and marker(QTL)  $\times$  marker(QTL) interactions. Significance values were corrected for multiple testing within a model using false discovery rate ( $< 0.05$ ) in the automated script. The script returned all significance values as well as QTL main-effect estimates in terms of allelic substitution values (Tables S5–S7). No significant three-way interactions were identified.

**AOP2 analysis.** Two independent homozygous lines containing functionally expressed *AOP2* transcript from *B. oleracea* expressed from a 35S promoter were obtained in the Col-0 background that is null for *AOP2* and *AOP3* [31,40]. These two lines were grown in a randomized block design with wild-type Col-0 and tested by HPLC for glucosinolate content and by ATH1 Affymetrix microarrays (<http://www.affymetrix.com>) for altered transcript levels [17,18]. For total aliphatic glucosinolate content, six individual plants per line

were measured and ANOVA used to test for altered glucosinolate accumulation. The complete experiment was conducted twice. As there was no difference between the independent transgenic lines, this factor was removed from the model. From each experiment, two independent RNA samples from Col-0 and two independent RNA samples from Col-0::AOP2 were obtained and hybridized with ATH1 Affymetrix microarrays as described [17,18]. Transcript levels for the genes involved in aliphatic glucosinolate biosynthesis (Table S3) were obtained and used in a targeted ANOVA testing the effect of the *AOP2* transgene. *p*-values were tested for significance against a false discovery rate of 0.05 using this subset of genes [70].

## Supporting Information

**Figure S1.** Gene Family and Transcription Factor eQTLs for the Aliphatic Glucosinolates

QTL position information for the FMOs, aconitases, and putative glucosinolate regulatory factors are presented as described in Figure 3.

Found at doi:10.1371/journal.pgen.0030162.sg001 (1.2 M TIF).

**Table S1.** Heritability of Glucosinolate Variation within the Bay  $\times$  Sha RILs

For all traits, ANOVA was conducted to test for the effect of variation between the RILs (Line), Experiment, Flat within an Experiment (Rep(Exp)) and an interaction between Line and Experiment. The *p*-values, sums of squares, and percent of total variance are presented.

Found at doi:10.1371/journal.pgen.0030162.st001 (38 KB XLS).

**Table S2.** Mean Values for Glucosinolate Traits in Each RIL

The mean value and standard deviation for each trait as measured within the described experiments is provided. Trait abbreviations are as listed in Table S1. All glucosinolate values are in  $\mu\text{mol mg}^{-1}$ . Line represents the specific Bay  $\times$  Sha RIL and N is the number of measurements obtained for this RIL.

Found at doi:10.1371/journal.pgen.0030162.st002 (841 KB XLS).

**Table S3.** Gene lists for Glucosinolate Pathway Genes

The genes defined as being involved in glucosinolate biosynthesis are presented along with their pathway. Functional Proof indicates whether the gene has been experimentally validated as playing a role in glucosinolate production or if the function is predicted.

Found at doi:10.1371/journal.pgen.0030162.st003 (29 KB XLS).

**Table S4.** Metabolic QTLs for Aliphatic Glucosinolate Traits

The average trait value per RIL across experiments along with the trait value per RIL for each individual experiment was used for QTL mapping; the presence of a QTL in a specific experiment is shown under “QTL Detection in Given Experiment.” The MT QTL column shows the complete results from the multitrait composite interval mapping within QTL Cartographer. M means that the QTL was identified as a main effect, while I means that it had an experiment specific interaction, and MI shows that it was found with both main and interaction effects using multitrait composite interval mapping. The data from each experiment were used to conduct ANOVA to test each QTL using the nearest marker as the main effect, with significance presented under “Main Effect” column. For each marker/QTL, the model tested for a QTL  $\times$  Experiment interaction, “Exp Interaction.” The model tested all pairwise marker  $\times$  marker interactions for evidence of epistasis. The significance of each pairwise test is presented under the columns marked “Pairwise assessment of Epistasis.” Finally, the model was used to estimate the allele substitution effect of each locus as shown. NS represents nonsignificant results within this model.

Found at doi:10.1371/journal.pgen.0030162.st004 (53 KB XLS).

**Table S5.** QTLs for Aliphatic Glucosinolate Summation Traits

QTLs detected for each aliphatic summation glucosinolate trait are presented as a separate table. The average trait value per RIL across experiments along with the trait value per RIL for each individual experiment was used for QTL mapping; the presence of a QTL in a specific experiment is shown under “QTL Detection in Given Experiment.” The MT QTL column shows the complete results from the multitrait composite interval mapping within QTL Cartographer. M means that the QTL was identified as a main effect, while I means that it had an experiment specific interaction, and MI shows that it was found with both main and interaction effects using multitrait composite interval mapping. The data from each experiment were



used to test each QTL using the nearest marker as main effect, with significance determined by this ANOVA presented under the “Main Effect” column. For each marker/QTL, the same model tested for a QTL  $\times$  Experiment interaction, “Exp Interaction.” This model also tested all pairwise marker  $\times$  marker interactions for evidence of epistasis. The significance of each pairwise test is presented under the columns marked “Pairwise assessment of Epistasis.” Finally, the model estimated the allele substitution effect of each locus. For all cells, NS represents nonsignificance within the model.

Found at doi:10.1371/journal.pgen.0030162.st005 (50 KB XLS).

#### Table S6. eQTL for Glucosinolate Biosynthetic Genes

Significant eQTLs previously detected for each glucosinolate biosynthetic gene. The genetic position in chromosome and cM is presented along with the genes' identities. In addition, the additive effect of the Bay-0 allele at each eQTL is provided.

Found at doi:10.1371/journal.pgen.0030162.st006 (31 KB XLS).

#### Table S7. Average Gene Network Expression (z-Score)

The mean z-score for each RIL for the aliphatic glucosinolate and indolic glucosinolate biosynthetic pathways are presented for the 211 RILs where there are available data.

Found at doi:10.1371/journal.pgen.0030162.st007 (27 KB XLS).

#### Table S8. Metabolic QTLs for Indolic Glucosinolate Traits

QTLs detected for each indolic glucosinolate trait are presented as a separate table. The average trait value per RIL across experiments along with the trait value per RIL for each individual experiment was used for QTL mapping, and the presence of a QTL in a specific experiment is shown under “QTL Detection in Given Experiment.” The MT QTL column shows the complete results from the multitrait composite interval mapping within QTL Cartographer. M means that the QTL was identified as a main effect, while I means that it had an experiment specific interaction, and MI shows that it was found with both main and interaction effects using multitrait composite interval mapping. The data from each experiment were used within an

ANOVA to test each QTL using the nearest marker as main effect; significance is presented under the “Main Effect” column. For each marker/QTL, the same model tested for a QTL  $\times$  Experiment interaction, “Exp Interaction.” Finally, the same model tested all pairwise marker  $\times$  marker interactions for evidence of epistasis. The significance of each pairwise test is presented under the columns marked “Pairwise assessment of Epistasis.” Finally, the model was used to estimate the allele substitution effect of each locus as shown. For all cells, NS represents nonsignificance within the model.

Found at doi:10.1371/journal.pgen.0030162.st008 (32 KB XLS).

#### Accession Numbers

The microarray dataset used in this study has been deposited at European Bioinformatics Institute ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) under numbers E-TABM-126 and E-TABM-224. All gene identifiers are listed in Table S3.

#### Acknowledgments

We would like to thank Steffen Abel, Katherine Denby, Dina St. Clair, and Olivier Loudet for useful discussions and manuscript reviews. The Danish National Research Foundation is acknowledged for its support to PlaCe (Center for Molecular Plant Physiology). Forsker-skole for Bioteknologi (FOBI) graduate school is thanked for the PhD stipend to BGH.

**Author contributions.** HCR and DJK conceived and designed the experiments. HCR and DJK performed the experiments. HCR, BGH, BAH, and DJK analyzed the data. BGH, CT, BAH, and DJK contributed reagents/materials/analysis tools. AMW, HCR, BGH, BAH, and DJK wrote the paper.

**Funding.** Funding for this project was obtained by National Science Foundation grants DBI 0642481 and MCB 0323759 to DJK.

**Competing interests.** The authors have declared that no competing interests exist.

#### References

- Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sunderland (Massachusetts): Sinauer Associates, Inc. 980 p.
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Essex (United Kingdom): Longman, Harlow. 340 p.
- Liu BH (1998) Statistical genetics: Linkage, mapping and QTL analysis. Boca Raton (Florida): CRC Press. 611 p.
- Symonds VV, Godoy AV, Alconada T, Botto JF, Juenger TE, et al. (2005) Mapping quantitative trait loci in multiple populations of *Arabidopsis thaliana* identifies natural allelic variation for trichome density. *Genetics* 169: 1649–1658.
- Keurentjes JB, Fu J, de Vos CH, Lommen A, Hall RD, et al. (2006) The genetics of plant metabolism. *Nat Genet* 38: 842–849.
- Kliebenstein DJ, Pedersen D, Mitchell-Olds T (2002) Comparative analysis of insect resistance QTL and QTL controlling the myrosinase/glucosinolate system in *Arabidopsis thaliana*. *Genetics* 161: 325–332.
- Yagil C, Barkalifa R, Sapojnikov M, Wechsler A, Ben-Dor D, et al. (2007) Metabolic and genomic dissection of diabetes in the Cohen rat. *Physiol Genomics* 29: 181–192.
- Anderson JR, Schneider JR, Grimstad PR, Severson DW (2006) Identification of quantitative trait loci for larval morphological traits in interspecific hybrids of *Ochlerotatus triseriatus* and *Ochlerotatus hendersoni* (Diptera: Culicidae). *Genetica* 127: 163–175.
- Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH (2005) Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* 169: 2225–2239.
- Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: A critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* 129: 133–147.
- Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* 3: 43–52.
- Jansen RC, Nap JP (2001) Genetical genomics: The added value from segregation. *Trends Genet* 17: 388–391.
- Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297–302.
- Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752–755.
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430: 743–747.
- Keurentjes JB, Fu JY, Terpstra IR, Garcia JM, van den Ackerveken G, et al. (2007) Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A* 104: 1708–1713.
- West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, et al. (2007) Global eQTL mapping reveals the complex genetic architecture of transcript level variation in *Arabidopsis*. *Genetics* 175: 1441–1450.
- Kliebenstein D, West M, van Leeuwen H, Loudet O, Doerge R, et al. (2006) Identification of QTLs controlling gene expression networks defined a priori. *BMC Bioinformatics* 7: 308.
- Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, et al. (2001) Genetic control of natural variation in *Arabidopsis thaliana* glucosinolate accumulation. *Plant Physiol* 126: 811–825.
- McMullen MD, Byrne PF, Snook ME, Wiseman BR, Lee EA, et al. (1998) Quantitative trait loci and metabolic pathways. *Proc Natl Acad Sci U S A* 95: 1996–2000.
- Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics* 2: 155–168.
- Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci U S A* 101: 7809–7814.
- Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, et al. (1999) Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J Bacteriol* 181: 6679–6688.
- Wittstock U, Halkier BA (2002) Glucosinolate research in the *Arabidopsis* era. *Trends Plant Sci* 7: 263–270.
- Kliebenstein DJ, Kroymann J, Mitchell-Olds T (2005) The glucosinolate-myrosinase system in an ecological and evolutionary context. *Curr Opin Plant Biol* 8: 264–271.
- Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. *Trends Plant Sci* 11: 89–100.
- Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Ann Rev Plant Biol* 57: 303–333.
- Levy M, Wang QM, Kaspi R, Parrella MP, Abel S (2005) *Arabidopsis* IQD1, a novel calmodulin-binding nuclear protein, stimulates glucosinolate accumulation and plant defense. *Plant J* 43: 79–96.
- Celenza JL, Quiel JA, Smolen GA, Merrick H, Silvestro AR, et al. (2005) The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. *Plant Physiol* 137: 253–262.
- Kliebenstein DJ, Gershenzon J, Mitchell-Olds T (2001) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* 159: 359–370.
- Kliebenstein D, Lambrix V, Reichelt M, Gershenzon J, Mitchell-Olds T

- (2001) Gene duplication and the diversification of secondary metabolism: Side chain modification of glucosinolates in *Arabidopsis thaliana*. *Plant Cell* 13: 681–693.
32. Kroymann J, Donnerhacke S, Schnabelrauch D, Mitchell-Olds T (2003) Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc Natl Acad Sci U S A* 100: 14587–14592.
  33. Hansen BG, Kliebenstein DJ, Halkier BA (2007) Identification of a flavin-monooxygenase as the S-oxygenating enzyme in aliphatic glucosinolate biosynthesis in *Arabidopsis*. *Plant J* 50: 902–910.
  34. Kliebenstein DJ, Figuth A, Mitchell-Olds T (2002) Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics* 161: 1685–1696.
  35. Textor S, Bartram S, Kroymann J, Falk KL, Hick A, et al. (2004) Biosynthesis of methionine-derived glucosinolates in *Arabidopsis thaliana*: Recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. *Planta* 218: 1026–1035.
  36. de Quiros HC, Magrath R, McCallum D, Kroymann J, Schnabelrauch D, et al. (2000)  $\alpha$ -Keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. *Theor Appl Genet* 101: 429–437.
  37. Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, et al. (2006) The genetics of plant metabolism. *Nat Genet* 38: 842–849.
  38. Hirai MY, Klein M, Fujikawa Y, Yano M, Goodenowe DB, et al. (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* 280: 25590–25595.
  39. Hirai M, Sugiyama K, Sawada Y, Tohge T, Obayashi T, et al. (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci U S A* 104: 6478–6483.
  40. Li G, Quiros CF (2003) In planta side-chain glucosinolate modification in *Arabidopsis* by introduction of dioxygenase *Brassica* homolog BoGSL-ALK. *Theor Appl Genet* 106: 1116–1121.
  41. Mithen R, Clarke J, Lister C, Dean C (1995) Genetics of aliphatic glucosinolates. III. Side-chain structure of aliphatic glucosinolates in *Arabidopsis thaliana*. *Heredity* 74: 210–215.
  42. Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. *Trends Biochem Sci* 29: 11–17.
  43. Montange RK, Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441: 1172–1175.
  44. Nudler E (2006) Flipping riboswitches. *Cell* 126: 19–22.
  45. Cheah MT, Wachter A, Sudarsan N, Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447: 497–500.
  46. Smolen GA, Pawlowski L, Wilensky SE, Bender J (2002) Dominant alleles of the basic helix-loop-helix transcription factor ATR2 activate stress-responsive genes in *Arabidopsis*. *Genetics* 161: 1235–1246.
  47. Skirycz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, et al. (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. *Plant J* 47: 10–24.
  48. Grubb CD, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, et al. (2004) *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J* 40: 893–908.
  49. Mackay TFC (2001) The genetic architecture of quantitative traits. *Ann Rev Genet* 35: 303–339.
  50. Clauss MJ, Dietel S, Schubert G, Mitchell-Olds T (2006) Glucosinolate and trichome defenses in a natural *Arabidopsis lyrata* population. *J Chem Ecol* 32: 2351–2373.
  51. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, et al. (2006) Positive selection driving diversification in plant secondary metabolism. *Proc Natl Acad Sci U S A* 103: 9118–9123.
  52. Windsor AJ, Reichelt M, Figuth A, Svatos A, Kroymann J, et al. (2005) Geographic and evolutionary diversification of glucosinolates among near relatives of *Arabidopsis thaliana* (Brassicaceae). *Phytochem* 66: 1321–1333.
  53. Loudet O, Chaillou S, Camilleri C, Bouchez D, Daniel-Vedele F (2002) Bay-0  $\times$  Shahdara recombinant inbred line population: A powerful tool for the genetic dissection of complex traits in *Arabidopsis*. *Theor Appl Genet* 104: 1173–1184.
  54. Kliebenstein DJ, Rowe HC, Denby KJ (2005) Secondary metabolites influence *Arabidopsis/Botrytis* interactions: Variation in host production and pathogen sensitivity. *Plant J* 44: 25–36.
  55. Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image processing with ImageJ. *Biophotonics Intl* 11: 36–42.
  56. Kliebenstein DJ (2007) Metabolomics and plant quantitative trait locus analysis—The optimum genetical genomics platform? In: Nikolau BJ, Wurtele ES, editors. *Concepts in plant metabolomics*. Dordrecht (the Netherlands): Springer. pp. 29–45.
  57. Schuster J, Knill T, Reichelt M, Gershenzon J, Binder S (2006) BRANCHED-CHAIN AMINOTRANSFERASE4 is part of the chain elongation pathway in the biosynthesis of methionine-derived glucosinolates in *Arabidopsis*. *Plant Cell* 18: 2664–2679.
  58. Gachon CMM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005) Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: Functional and evolutionary implications. *Plant Mol Biol* 58: 229–245.
  59. Klein M, Reichelt M, Gershenzon J, Papenbrock J (2006) The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS J* 273: 122–136.
  60. Hansen CH, Du LC, Naur P, Olsen CE, Axelsen KB, et al. (2001) CYP83B1 is the oxime-metabolizing enzyme in the glucosinolate pathway in *Arabidopsis*. *J Biol Chem* 276: 24790–24796.
  61. Bak S, Nielsen HL, Halkier BA (1998) The presence of CYP79 homologues in glucosinolate-producing plants shows evolutionary conservation of the enzymes in the conversion of amino acid to aldoxime in the biosynthesis of cyanogenic glucosides and glucosinolates. *Plant Mol Biol* 38: 725–734.
  62. West MAL, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, et al. (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. *Genome Res* 16: 787–795.
  63. Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. *Gen Res* 75: 345–355.
  64. Basten CJ, Weir BS, Zeng ZB (1999) QTL Cartographer, version 1.13. Department of Statistics, North Carolina State University, Raleigh, NC [computer program]. Available at: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Accessed 10 January 2001.
  65. Wang S, Basten CJ, Zeng ZB (2006) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC Available at: <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Accessed 10 January 2006.
  66. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
  67. Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142: 285–294.
  68. Rebai A (1997) Comparison of methods for regression interval mapping in QTL analysis with non-normal traits. *Genet Rec* 69: 69–74.
  69. Magrath R, Bano F, Morgner M, Parkin I, Sharpe A, et al. (1994) Genetics of aliphatic glucosinolates. I. Side chain elongation in *Brassica napus* and *Arabidopsis thaliana*. *Heredity* 72: 290–299.
  70. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—A practical and powerful approach to multiple testing. *J Royal Stat Soc Series B Methodol* 57: 289–300.