The effect of developmental variation on expression QTLs in a multi parental *C. elegans* population

Abraham L. van Eijnatten[1#], Mark G. Sterken[2], Jan E. Kammenga[2], Harm Nijveen[3], Basten L. Snoek[1#]

1 Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8 3584 CH UTRECHT The Netherlands
2 Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands
3 Laboratory of Bioinformatics, Wageningen University, Droevendaalsesteeg 1, Radix West, Building 107, 6708 PB Wageningen the Netherlands
Corresponding Authors: a.l.vaneijnatten@uu.nl, l.b.snoek@uu.nl

Keywords: C. elegans, genetics, eQTL mapping, developmental age

Abstract

Regulation of gene expression plays a crucial role in developmental processes and adaptation to changing environments. eQTL mapping is a technique used to study the genetic regulation of gene expression using the transcriptomes of recombinant inbred lines. Typically, the age of the inbred lines at the time of RNA sampling is carefully controlled. This is necessary because the developmental process causes changes in gene expression, complicating the interpretation of eQTL mapping experiments. However, due to genetics and variation in ambient micro-environments, organisms can differ in their "developmental age", even if they are of the same chronological age. As a result, eQTL patterns are affected by developmental variation in gene expression. The model organism *C. elegans* is particularly suited for studying the effect of developmental variation on eQTL mapping patterns. In a span of days, *C. elegans* transitions from embryo through four larval stages to adult while undergoing massive changes to its transcriptome. Here we use *C. elegans* to investigate the effect of developmental age variation on eQTL patterns and present a normalization procedure. We used dynamical eQTL mapping, which includes the developmental age as a co-factor, to separate the variation in development from genotypic variation and explain variation in gene expression levels. We compare classical single marker eQTL mapping and dynamical eQTL mapping using RNA-seq data of ~200 multi-parental recombinant inbred lines of *C. elegans*. The results show i) that many eQTLs are caused

by developmental variation, ii) most trans-bands are developmental QTLs and iii) dynamical eQTL mapping detects additional eQTLs not found with classical eQTL mapping. We recommend that correction for variation in developmental age should be strongly considered in eQTL mapping studies given the large impact of processes like development on the transcriptome.

Introduction

Regulation of gene expression is a key mechanism underlying the huge diversity of phenotypes, adaptations, and developmental stages within and across species. Understanding gene expression regulation provides insights into how organisms develop and adapt to changing environments(L. B. Snoek et al. 2014; Francesconi and Lehner 2014; Jovic et al. 2017; Spencer et al. 2011; Araya et al. 2014; Boeck et al. 2016; Reinke et al. 2000). The genetic regulation of gene expression can be investigated through expression quantitative trait locus (eQTL) mapping, also called genetical genomics(Jansen and Nap 2001). In this statistical procedure, polymorphic regions (eQTLs) are associated with variation in gene expression, pinpointing potential regulatory regions. The nematode *Caenorhabditis elegans* is often used in eQTL studies and has a small, well annotated genome in addition to well documented genetic variation(L. B. Snoek et al. 2013; B. L. Snoek et al. 2020; Evans et al. 2021; Andersen and Rockman 2022; Sterken et al. 2015). Its tolerance of cryo-preservation, large brood size and self-fertilizing ability allow for the construction of perpetual recombinant inbred lines (RILs) derived from genetically diverse *C. elegans* isolates(Gaertner and Phillips 2010; Sterken et al. 2015; Evans et al. 2021). RILs are homozygous for almost all loci and constitute a genetic mosaic of the parental genotypes, thereby increasing the resolution and power of methods for eQTL detection(Andersen and Rockman 2022). The above qualities make *C. elegans* an exceptional model system for genetical genomics.

The developmental process in *C. elegans* is associated with massive and rapid changes in gene expression(L. B. Snoek et al. 2014; Hashimshony et al. 2015; Levin et al. 2016). Many genes are expressed in a developmental stage-specific manner(McCarroll et al. 2004; Reinke et al. 2000; Boeck et al. 2016; Viñuela et al. 2010; 2012). Others are up- and down-regulated cyclically as *C. elegans* transitions through the four larval stages(Hendriks et al. 2014; Meeuse et al. 2020; Kim, Grün, and van Oudenaarden 2013). The speed of development, as well as the process of aging, varies between individual *C. elegans* strains due to stochastic and micro-environmental factors, maternal age and variation in genetic background(Viñuela et al. 2010; 2012; Francesconi and Lehner 2014; L. B. Snoek et al. 2014; Bulteau and Francesconi 2022; B. L. Snoek et al. 2019; Volkers et al. 2013; Perez et al. 2017). Due to the impact of the developmental process on gene expression combined with the interaction with the genetic

background, eQTL patterns can be affected by uncontrolled developmental variation(Viñuela et al. 2010; Francesconi and Lehner 2014; L. B. Snoek et al. 2014). To conceptualize developmental variation, it is useful to distinguish between the chronological age and the developmental age. The chronological age is simply the measured age of the organism, whereas the developmental age represents how far the organism has progressed along the developmental process. Because variation in developmental speed and thus developmental age depends partly on genetic factors, it is intrinsic to populations used for genetic mapping. Careful synchronization of worms at a particular developmental checkpoint can help to reduce but does not eliminate developmental age variation(Ben-David et al. 2021). As a result, the eQTL mapping procedure could attribute expression variation resulting from variation in developmental age to a genetic polymorphism. Conversely, developmental age variation could obscure the effects of the genetic background on gene expression.

One way to deal with the confounding effect of the developmental process in eQTL mapping is to apply a normalization procedure(Viñuela et al. 2012; Jovic et al. 2019). Another approach is to make the effects of developmental variation on gene expression explicit by including the developmental age as a co-factor in the statistical model for eQTL mapping (dynamical eQTL mapping)(Francesconi and Lehner 2014). The latter approach is more informative because it explicitly models how eQTLs affect the dynamics of gene expression, rather than just the mean over the developmental ages of the inbred lines. A seminal paper by Francesconi and Lehner(Francesconi and Lehner 2014) showed that dynamical eQTL mapping using the quantified developmental age can be leveraged to detect additional eQTLs. By relating the magnitude of gene expression to the developmental age the authors showed that eQTLs can affect the expression dynamics over the course of the developmental process.

In this study we aimed to quantify the difference between the classical eQTL mapping approach, which does not explicitly consider developmental variation, and dynamical eQTL mapping. To this end, we used RNA-seq data of ~200 *C. elegans* multi-parental recombinant inbred lines (mpRIL)(B. L. Snoek et al. 2019; 2021). RNA-seq samples were obtained 48 hours after synchronization by egg bleaching, ensuring there is no variation in chronological age within the population. At the time of RNA sampling the mpRILs were in the L4 larval stage. We started by quantifying the developmental age directly from the gene expression data with a straightforward approach involving principal component analysis (PCA). Next, we performed eQTL mapping using linear models, both with and without the developmental age included as a co-factor. We compared these models qualitatively, by the distribution of eQTLs over the genome, and quantitatively, by the number of eQTLs detected and the variance in gene expression attributed to the eQTLs. We show that most, but not all, *trans*-bands (regulatory hotspots) result from a shared association with the developmental age between the SNP marker and transcript levels of genes

involved in developmental processes. Moreover, we present evidence that such hotspot loci affect gene expression through their effect on the developmental speed. We also show how comparing the results from models with and without the developmental age can help to distinguish between loci linked to the developmental age and other regulators of gene expression.

Methods

All analyses were performed using R version 4.1.3(R Core Team (2022) 2022). Plots were generated using the ggplot2 package(Martin, Schbath, and Hennequet-Antier 2022). Pearson correlations were calculated using the cor() function from base R.

Data and pre-processing

We used data of 199 multi-parental recombinant inbred lines (mpRILs)(B. L. Snoek et al. 2019; 2021; Volkers et al. 2013). The mpRILs were derived from an advanced cross between four parental lines which were collected from the French regions of Orsay (2 lines) and Santeuil (2 lines). The mpRILs were grown for 48 hours after bleaching at 24 degrees. We verified that our population was still in L4 by calculating the correlation of the transcriptome of the mpRILs with two L4 reference genomes and two young adult reference genomes from the N2 strain (L4, L4b, YA and N2Yad-1 from (Boeck et al. 2016)) (**Figure S14**). The parental lines were excluded from the eQTL mapping procedure and all subsequent analysis unless explicitly mentioned. The data consists of fpkm values for 38322 transcripts obtained using RNA-seq and a genetic map of 8933 bi-allelic SNPs. Due to the SNPs being bi-allelic the genotype at each marker was coded as 0 or 1 for homozygous individuals and 0.5 for heterozygous individuals. Before analysis we filtered the gene expression data to retain the 12029 transcripts with more than 20 non-zero fpkm values and a log2(mean fpkm) value higher than -5. Out of the 12029 filtered transcripts 2341 are polycistronic, for a total of 15224 genes in the filtered dataset. In all analysis we use the center log ratio of gene expression (CLR), which for gene j and mpRIL i is calculated according to the following equation:

$$CLR_{i,j} = \log_2\left(\frac{\left(fpkm_{i,j}+1\right)}{\left(\frac{1}{n}\left(\sum_i^n fpkm_{i,j}\right)\ +1\right)}\right)$$

Developmental age

We quantified the developmental age by performing PCA on the filtered and normalized gene expression matrix (with the mpRILs as columns and the transcripts as rows, such that mpRILs have loadings on PC1 and transcripts have scores) using the prcomp() function from the stats package. The loadings of the

mpRILs on PC1 were taken as a proxy for the developmental age in subsequent analysis. For figure 1, S2, S3, S4 and S12 we calculated the Z-score of the center log ratio of the fpkm values by mean centering on 0 and scaling such that the expression of each gene has a standard deviation of 1 using the scale() function from base R.

We calculated developmental age estimates using the ae() function from the RAPToR package in R. We tried both the Cel_larval and Cel_larv_ya reference sets, obtaining a more even spread in distribution using the former. The ae() function allows for the possibility of specifying a prior but this did not provide a narrower distribution of the estimated ages, even when using a low standard deviation (1 hour). We therefore estimated the ages of the mpRILs with the Cel_larval reference and no prior.

eQTL mapping with linear models

We ran linear models using the FastLm() function from the RcppAramadillo package. We applied a single marker model (SMM, gene expression ~ marker genotype), additive age model (AAM, gene expression ~ marker genotype + PC1) and interaction model (IM, gene expression ~ marker genotype (IMM term) + PC1 + marker genotype * PC1 (IMI term)) to perform pairwise tests between the filtered transcripts and all markers. P-values were calculated for model terms using t-tests. For the interaction model, both a significant IMM term and a significant IMI term were considered an eQTL. Per gene, we called a maximum of one eQTL per chromosome. We called a *cis*-eQTL if the distance between the middle of the gene and the marker was < 2Mb, and a *trans*-eQTL otherwise. For all models, we also generated p-values by randomly distributing the gene expression values over the mpRILs for each gene. We used the p-values obtained by permutation to constrain the type I error and obtain significance thresholds. For thresholding the marker and interaction terms we used the lowest p-value per gene per chromosome. For the developmental age variable, we used the p-values of the models with the lowest marker p-value per transcript per chromosome. All results are obtained using an FDR of 0.05 unless explicitly mentioned in the text. P-value thresholds are in **Table S1**.

Transcripts affected by development

Transcripts were considered significantly affected by the developmental process if they had a significant p-value for the development variable using the AAM, in all the six models (one per chromosome) that resulted in the lowest marker p-value.

Calculating partial eta squared

We performed an anova on the model with the lowest marker p-value using the aov() function from the stats package and extracted from the effects vector the residual sum of squares (RSS) and the sum of

squares explained by the marker (SSM). Note that the marker with the lowest p-value can differ between the SMM and the AAM. We then calculated the partial eta squared according to the following equation:

$$\frac{SSM}{SSM + RSS}$$

GO enrichment

We performed GO-term enrichment on genes that map to specific hotspots. We tested for enrichment of GO terms from the subcategories biological process, molecular function and cellular component using a hypergeometric test. As the total gene set we used the genes on the 12029 transcripts used in the eQTL mapping. We called significant enrichment if p-value < 0.001 and at least three genes were associated with this GO-term.

Results

Developmental age estimation by principal component analysis

We quantified the developmental age of the multi-parental recombinant inbred lines (mpRIL) by performing a principal component analysis (PCA) on the normalized gene expression counts. Estimating relative differences in developmental age using a PCA was shown to be effective in a previous study(Jovic et al. 2019). Principal component (PC) 1 explained ~48% of the gene expression variation and PC2 explained ~5% (**Figure 1A**). Under the assumption that the developmental age would be the largest contributor to the variation in gene expression, we investigated whether PC1 is a good proxy for the developmental age of the mpRILs. First, we looked at the expression of the yolk protein *vit-2* gene, since it is part of a cluster of 53 genes (developmental indicator genes) that show a robust linear increase in expression during the L4 stage (cluster 1 in Snoek *et al.*, 2014, see figure S4 for gene names). The expression of *vit-2* increases as the projection of the mpRILs on PC1 increases (Pearson correlation ~ 0.91) (**Figure 1A**). Second, we investigated whether the other developmental indicator genes also show a positive correlation with PC1. We found a strong positive correlation between the projection on PC1 and the mean expression of developmental indicator genes (Pearson correlation ~ 0.97) (**Figure 1B**). The association between the developmental indicator genes and PC1 is further verified by high projections of these genes on PC1 (**Figure S1**). We validated the developmental indicator gene approach by taking a larger set of 2050 genes shown to be monotonically rising in a different study (Hendriks *et al.*, 2014). The mpRILs ordered by their projection on PC1 sort the expression of these genes well (**Figure S2**). Third, we verified the estimated developmental ages of the mpRILs by using the RAPToR package (Bulteau and Francesconi 2022). The RAPToR package estimates the age of an organism based on its expression

profile, by using an appropriate transcriptomic time series as a reference. The high correlation between the RAPToR ages and PC1 (Pearson correlation ~0.91) confirms that PC1 can be used as a proxy for the developmental age. However, the mpRILs ordered by the RAPToR age estimates provided a sorting that was less consistent with the known monotonically increasing expression of the developmental indicator genes (**Figure S3**) compared to PC1 (**Figure S4**). Hence, we used the projections of the mpRILs on PC1 as our developmental age estimates in the subsequent analysis. Note that PC1 appears to be associated with the genetic background, given the nearly identical loadings of the parental duplicates on PC1 **(Figure S5)**. We conclude that PC1 is strongly associated with the developmental process and can be used to approximate the developmental age of the mpRILs.

The effect of developmental age variation on eQTL mapping

To study the effect of variation in developmental age on eQTL mapping we compared a linear single marker model (SMM) with two models that include the developmental age as a co-factor: i) the linear additive age model (AAM) and ii) the linear interaction model (IM). For the IM, we distinguished eQTLs due to a significant marker effect from eQTLs due to a significant interaction between the marker and the developmental age. While the IM, marker (IMM) term detects additive differences in gene expression, a significant IM, interaction (IMI) term indicates an eQTL that influences the slope of gene expression over the developmental age. While the eQTLs detected by the models partially overlap, each model also detects a subset of unique eQTLs (**Figure S6**). We focus our initial investigation mostly on the SMM and AAM. We refer to SMM-only eQTLs, AAM-only eQTLs or SMM+AAM eQTLs to distinguish between eQTLs found by only the SMM, AAM or both respectively. We first show the difference between the models in terms of the distribution of eQTLs over the genome. Next, we show the quantitative effect of dynamical eQTL mapping by comparing the SMM with the AAM in terms of the number of eQTLs, p-values, strength of the marker effects and the heritability of transcript levels. Finally, we discuss the application of the IM for dynamical eQTL mapping.

Developmental age variation affects the manifestation of eQTL hotspots.

We investigated whether the AAM shows an altered distribution of eQTLs over the genome compared to the SMM. The SMM yielded several hotspots enriched in *trans*-eQTLs, whereas *cis*-eQTLs are distributed evenly over the genome (**Figure 2A, Figure S7A**). The AAM yielded fewer eQTLs, with a particularly strong decrease in the number of large hotspots **(Figure 2B, Figure S7B)** and *trans*-eQTLs in general **(Table 1**). On the other hand, the number of (unique) markers with an eQTL increased (**Table S2**). The largest hotspot found by the SMM, located at position 3.4 Mb on chromosome X, significantly affects the expression of 1196 transcripts. In contrast, the AAM yielded only a single eQTL at this

location. Replacing PC1 with the mean developmental indicator gene expression as co-factor in the AAM gave a very similar eQTL distribution **(Figure S8)**, with the same disappearing hotspots. These results indicate that most hotspots are caused by developmental variation.

We hypothesized that the SMM-only hotspots are developmental QTLs. Under this hypothesis we would expect (1) that transcripts involved in developmental processes are enriched at the hotspots that disappear when using the developmental age as a co-factor, (2) that the genotype at the hotspot position is linked to the developmental age and (3) that SMM-only eQTLs primarily map to such hotspots.

First, we investigated whether genes mapping to the largest hotspot were enriched for developmental GO-terms. The genes having an eQTL at the hotspot position were enriched for developmental processes such as *oocyte maturation*, *polar body extrusion after meiotic division*, *P granule*, *eggshell formation*, *pseudopodium*, *amoeboid sperm motility*, *male meiosis chromosome segregation*, *mitotic spindle pole* and *structural constituent of cuticle*. This supports the involvement of this hotspot locus in development. Other hotspots were similarly enriched for genes associated with developmental processes (**File S1**). Apart from functional categories hotspots also share many of the same genes. As an example, more than 80% (896 out of 1115) of the transcripts that map to the hotspot at position 12.8 Mb of chromosome II also map to the hotspot at position 3.4 Mb of chromosome X. This paints a picture of genes involved in developmental processes indiscriminately mapping to hotspots due to a shared association with the developmental age.

Second, we investigated whether the markers at the hotspots that are only found with the SMM are linked to the developmental age by comparing the distributions of the developmental age between the alleles of three major hotspot loci (indicated by red arrows in **Figure 2**) (**Figure 3**). As expected, the distribution of developmental ages was clearly different between the genotypes at these hotspots. These loci could be linked to the developmental age because they regulate or affect the developmental speed. In this view, SMM-only hotspots affect gene expression through their effect on the developmental process itself. To investigate this, we ran single marker models with the developmental age as trait. In line with our hypothesis the marker of the largest hotspot was also the most predictive of the developmental age of the mpRILs (**Figure S9**).

Third, we investigated the distribution of SMM-only eQTLs (**Figure S10**). In line with our hypothesis, this subset of eQTLs maps primarily to the disappearing hotspots. Together these three results suggest that most hotspots can be attributed to genes whose expression changes based on the developmental age mapping to loci that influence the developmental speed (**Figure 4**).

Developmental age as cofactor strongly affects the quantitative eQTL landscape

Having observed large differences between the models in the distribution of eQTLs over the genome, we further investigated the quantitative effect of adding the developmental age as a co-factor on the number of eQTLs and the p-values and allelic effects of individual eQTLs. First, we examined the impact of the developmental process on the transcriptome using the developmental age term of the AAM. We find 9473 transcripts (out of 12029 polycistronic transcripts) with expression levels significantly affected by the developmental age (-log(p) > 1.79 (FDR = 0.05) (**Table 1**). This number is in line with previous reports on the N2 reference strain(L. B. Snoek et al. 2014) and emphasizes the substantial changes in overall gene expression during development.

Next, we compared the p-values of the marker effects between the SMM and AAM (**Figure 5A**). For a subset of eQTLs (151), the p-value was more than 10 orders of magnitude lower in the AAM compared to the SMM (*Strong developmental effect eQTLs*, above top black line in (**Figure 5A**)). The transcripts with such eQTLs all displayed a gene expression pattern over the developmental age that has a clear linearly increasing trend and little within allele variation compared to between allele variation (**Figure 6**). Most strong developmental effect eQTLs are detected by both the SMM and AAM. Examining the changes in p-value around the thresholds (**Figure 5B, Table S1**) revealed that many eQTLs were only detected by one of the models. eQTLs that are differentially detected between the models are of special interest as in these cases the qualitative result of the eQTL mapping procedure is impacted, potentially obfuscating or revealing relevant biology.

The total number of eQTLs decreased by ~53% between the SMM (18605) and the AAM (8666) (**Table 1**). The AAM detected 1221 AAM-only eQTLs that would not be detected by the SMM (**Figure 5B cyan color, Table S2**). On the other hand, 11160 SMM-only eQTLs were no longer significantly associated with gene expression when the AAM is applied (**Figure 5B, orange color**). In conclusion, adding the developmental age as a co-factor to the model changed the detection outcome for more than 12000 potential eQTLs.

Dynamical eQTL mapping can reveal additional regulatory effects by correcting for developmental age

To understand why some eQTLs are differentially detected between the SMM and AAM we looked at individual eQTLs. For the SMM-only eQTLs the observed significant difference in gene expression between the alleles (**Figure 7A, boxplot**) can be explained primarily by the difference in developmental age (**Figure 7A; line plot, Figure 3A**). The blue genotype is on average developmentally younger than the red genotype, causing gene expression differences. For this subset of eQTLs, accounting for developmental variation in the model shows that the locus is not significantly associated with gene

expression when stratified for developmental age (**Figure 7A; line plot**). For the AAM-only eQTLs there is no significant mean difference in gene expression between the alleles (**Figure 7B; boxplot)**. However, the AAM reveals a significant association, over the entire range of developmental ages of the mpRILs, between this locus and gene expression (**Figure 7B, line plot**). In both cases of differentially detected eQTLs the SMM is being confounded by an association between the marker and the developmental age. These examples show how dynamical eQTL mapping can control for developmental variation in the mapping population **(Figure 7A)** and uncover regulatory regions obscured by this developmental variation **(Figure 7B**) on a gene-by-gene basis.

Developmental age variation impacts the allelic effect of eQTLs

The estimated allelic effects are expected to differ between the SMM and AAM given the large number of transcripts affected by development. In the previous section we showed that for SMM-only eQTLs, variation in mean gene expression between alleles does not always imply gene expression differences at a specific point in development (**Figure 7A**). In such cases we expect the allelic effect on gene expression estimated by the SMM, which considers only mean differences between alleles, to be higher than the allelic effect estimated by the AAM, which accounts for the developmental age. Conversely, for AAM-only eQTLs the allelic effect can be obscured unless viewed over a developmental axis **(Figure 7B, Figure S11).** In this scenario we expect the allelic effect estimated by the AAM to be higher than the SMM estimation. To quantify the strength of the allelic effect, we calculated the (partial) eta squared of the marker variable for SMM-only and AAM-only eQTLs. Indeed, the SMM tends to infer a stronger allelic effect for SMM-only eQTLs (**Figure 8A**), whereas the AAM tends to infer a stronger allelic effect for AAM-only eQTLs (**Figure 8B**). We also investigated the effect of adding developmental age as a cofactor in models of the narrow sense heritability (h_2). We find that with developmental age as cofactor the h_2 of gene expression changes by more than 0.05 for ~47% of genes, and that the direction of change depends on the type of eQTL(s) (SMM-eQTL and/or AAM-eQTL) mapping to that gene **(Text S1)**. These findings indicate that the allelic effect of eQTLs and the h_2 of gene expression can be either obscured or exaggerated by hidden developmental variation.

Interactions between the genomic background and developmental age affect the dynamics of gene expression

Genetic loci can affect, apart from the magnitude and timing of gene expression, also the shape of the dynamical pattern(Francesconi and Lehner 2014). To investigate the frequency of interactions between marker effects and the developmental age we applied a linear interaction model for eQTL mapping. A significant interaction term in this model indicates different dynamics between the marker alleles (**Figure 9**). Using the interaction model results in a ~20% decrease in the number of marker effects detected (6886) compared to the AAM. On top of the marker effects the interaction model detects a significant interaction between the marker and developmental age for 1161 transcripts (FDR = 0.05 **Table 1; Figure 9; FDR = 0.1 shown in Figure S12)**. Investigating the distribution of interaction eQTLs over the genome shows that these also cluster in hotspots. The hotspot with the most eQTLs (20.440891 Mb, chr V) affects the expression of almost 200 transcripts. Despite the correction for the developmental age inherent to the model, the transcripts mapping to this hotspot are significantly enriched in GO-terms associated with the cell cycle, meiosis, mitosis and other developmental processes (**File S2**). This locus could therefore be an important determinant in the developmental speed during the range of developmental ages spanned by the mpRILs by affecting the rate of up- or down regulation of many transcripts involved in developmental processes (**Figure S13 A-C**). The genotypes corresponding to this marker also differ substantially in their distribution over the developmental age (**Figure S13D**), suggesting that its effect on the developmental speed was also present at younger developmental ages. This shows that the interaction model can detect loci that have large effects on the transcriptome by influencing the rate at which transcript levels change for many transcripts.

Discussion

We have shown that eQTL hotspots that disappear after controlling for the developmental age can correspond to markers influencing the developmental speed. For such hotspots, there is likely no direct regulatory relationship between the hotspot and the genes with an eQTL mapping to it. After all, at any specific developmental age there is no significant difference in expression between the alleles. Rather, one of the genotypes is on average further developed than the other, causing gene expression differences. Because the genes mapping to such hotspots are enriched for developmental functions, they are likely to have non-constant expression over the developmental age range of the mapping population. A non-zero slope of gene expression over the developmental age can cause differences in the overall magnitude of gene expression between marker alleles, even in the absence of gene expression differences between

alleles at any specific developmental age, if developmental ages are not distributed equally between the alleles.

The mechanism through which a polymorphic SNP influences the developmental speed can be difficult to entangle because gene expression both drives and is the result of the developmental speed. Therefore, an eQTL might influence the developmental speed by affecting the expression of some genes, as a result changing the expression of many more genes. For some SNPs in hotspots the association with the developmental age can be understood from the genomic context. For example, the large hotspot at the beginning of chromosome X maps close to the *vit-5* gene encoding for the Vitogellin-5 protein which is a major yolk component. Inhibition of *vit-5* by means of RNAi has been shown to result in slower rates of post-embryonic growth(Kamath et al. 2003). Interestingly, the maternal expression of vitellogenin genes appears to be a major determinant of the developmental speed(Perez et al. 2017). As a second example, the hotspot at 10.1 MB on chromosome III is located in the *kel-10* gene. The human ortholog (*KLHL10*) of this gene is involved in spermatogenesis, and *kel-10* is affected by the *daf-2* gene, a well-known regulator of lifespan in *C. elegans*(Kenyon et al. 1993).

Hotspots are often assumed to be global regulators of gene expression, although it is well known that hotspots can also result from genes with correlated expression because of uncontrolled latent factors(Breitling et al. 2008; Michaelson, Loguercio, and Beyer 2009). We show that developmental variation can be a major cause of such correlations, to the extent that most prominent hotspots are no longer present after correcting for the developmental age. Therefore, we posit that the default interpretation for hotspots should be an association between the hotspot locus and the developmental age.

Loci that influence the developmental speed need not be seen as merely a confounder in eQTL mapping. Instead, they are crucial determinants of the expression state of an organism. Investigating the eQTLs detected by a non-dynamical but not a dynamical model is a convenient approach to search for loci that are causative for developmental age variation. This would allow a distinction between loci that cause gene expression differences even if organisms are at the same developmental age (eQTLs detected by AAM), and loci that affect gene expression by causing developmental age differences (SMM-only eQTLs). This approach is not limited to the developmental process but can be applied to any process which leaves a sufficiently strong signal in transcriptomics data. More generally, including latent factors in eQTL mapping can allow one to distinguish between loci that influence gene expression through their effect on generic processes and other effects of the genetic background on gene expression.

Because eQTLs can also cause allele dependent non-linear expression patterns over the developmental age(Francesconi and Lehner 2014), we performed the eQTL mapping using a natural spline of the

developmental age (**Text S2**). Using the interaction term of this model, we hoped to identify eQTLs responsible for complex dynamical patterns. However, constraining the FDR of the natural spline model using the same permutation strategy we used for the linear models was not possible, as the p-values obtained on the actual dataset were barely distinguishable from the p-values obtained on the randomized dataset. We suspect this is in part due to a lack of statistical power, resulting from a combination of the size of the population (199 lines) and the dense genetic map (8933 SNPs) used for the mapping. An additional factor could be that our population has a narrow distribution of developmental ages. A study by Francesconi and Lehner detected many non-linear patterns using natural splines(Francesconi and Lehner 2014). The population used in their study contained worms in the L3, L4 and young adult developmental stages. The mpRILs used in this study were all in the L4 larval stage (**Figure S14**). Possibly, non-linear dynamics only occur over a wider range of developmental ages than is covered by the mpRILs in our study or when multiple distinct stages are present. To investigate we clustered 2346 transcripts whose expression oscillates with a period of approximately eight hours(Hendriks et al. 2014) using k-means. The clustering revealed mostly monotonic or constant gene expression patterns for the clusters (**Figure S15**). This indicates that the developmental range spanned by the mpRILs is not broad enough to clearly observe the oscillations to which a large part of the *C. elegans* transcriptome is subject(Kim, Grün, and van Oudenaarden 2013; Meeuse et al. 2020; Hendriks et al. 2014). Therefore, the natural spline model may not be suitable, or at least superfluous, for mapping eQTLs in our experiment.

In this paper we used a relatively simple method (PCA) to quantify the developmental age. Various more complex methods that extract developmental ages from transcriptomics data have already been developed. An example is the RAPToR framework, which projects the samples on an interpolated reference time series(Bulteau and Francesconi 2022). Because we wanted a method that was generally applicable, we did not want to be completely dependent on a reference (time) series. Our approach enables detection of other (semi-)linear processes affecting gene expression on a genome-wide scale and can be applied to other published eQTL studies in *C. elegans*(Sterken et al. 2020; B. L. Snoek et al. 2017; Li et al. 2006; 2010; Sterken et al. 2017; van Wijk et al. 2023) and beyond(Ben-David et al. 2021). Furthermore, our method, despite being simpler, provided an ordering of the expression of developmental indicator genes that was more consistent with the monotonic linear increase described in previous literature.

We showed that PC1 accounted for most of the developmental variation in our population. The study by Francesconi and Lehner showed that the developmental signature was distributed between PC1 and PC2, which were assigned the interpretation of oogenesis and spermatogenesis respectively (Francesconi and Lehner 2014). Together PC1 and PC2 explained ~50% of the gene expression variation, which is close to

our PC1 (~48%). In accordance with two developmental axes, the projected samples formed an almost circular trajectory on PC1 and PC2. The authors showed that, while the distribution of developmental ages is centered on the L4 larval stage, the population also includes individuals from late L3 and early adult. It is possible that transitions between developmental stages cause developmental variation in gene expression to be distributed over more than one PC axis.

A previous study scored the time to first egg phenotype of the mpRILs(B. L. Snoek et al. 2019). Because time to first egg is a clear indicator of the developmental speed, we hypothesized that this phenotype would correlate with PC1. Surprisingly, there was almost no correlation between PC1 and the time to first egg, or indeed any of the other phenotypes scored in this study (**Figure S16**). This can likely be partially attributed to between experiment variation. A second relevant factor could be the reported inverse relationship between the duration of larval development and the time between the first adult molt and the development of the first embryo(Perez et al. 2017). Thirdly, the developmental speed could differ between developmental stages. Development is a complex process with crucial developmental events likely being regulated by independent timers(Filina et al. 2022). Accordingly, correlations between the durations of two larval stages in experiments with genetically identical worms are low when temperature is strictly controlled(Mata-Cabana et al. 2022). On top of this, the genetic background can have stage-specific effects on the developmental speed(Mata-Cabana et al. 2022; Filina et al. 2022). Stage-specific developmental variation would de-correlate developmental ages over the course of chronological time. Verifying this third explanation would require an experiment in which samples are obtained within and between time points. Such an experiment would allow us to measure the degree to which developmental variation is maintained on various timescales. More importantly, this experiment would allow us to compare gene expression over developmental age with gene expression over chronological age. In this way we could definitively show whether gene expression variation over PC1 is due to developmental variation. Furthermore, it would allow for an investigation into the extent to which eQTL patterns over developmental time translate to eQTL patterns over chronological time.

In conclusion, we have shown that performing a PCA and including principal components into the model for eQTL mapping leads to an improved understanding of eQTL patterns by separating the effects of generic processes, such as development, from other regulatory effects of the genetic background. Furthermore, the effect of these generic processes on gene expression and their interaction with the genetic background can be understood by comparing models with and without principal components. Within an experimental condition, principal components can correspond to the developmental process, technical noise or other latent variables unrelated to the research question. Including principal components as representations of such latent variables in the mapping procedure is a convenient way to

control for or come to understand sources of variance confounding the effect of interest(O'Duibhir et al. 2014). We recommend that this should be standard practice in eQTL mapping studies given the large effect of processes like development on the transcriptional state of the cell.

Data availability statement

All code used in the study is available at github.com/SnoekLab/van-Eijnatten-etal-2023. The genetic map used for eQTL mapping can be found in additional file 1 of the original paper on the multi-parental population(B. L. Snoek et al. 2019). Raw RNA-sequencing data used for constructing the genetic map and eQTL mapping is available from the NCBI Sequence Read Archive database (SRA; www.ncbi.nlm.nih.gov/bioproject/PRJNA495983/). Normalized read counts and p-value matrixes of the eQTL mapping for each of the models can be found on WormQTL2 (B. L. Snoek et al. 2020) as Eijnatten_etal_2023 SMM/AAM/IMM/IMI.

Author contributions

BLS conceived the study, BLS and ALvE designed and performed the data analysis steps. ALvE and BLS wrote the paper assisted by HN, JEK and MGS.

Acknowledgements

This research was in part supported by a grant from the NWO domain Applied and Engineering Sciences VENI grant.

Funding

M.G.S. was supported by NWO domain Applied and Engineering Sciences VENI grant (17282).

Competing interests

Authors declare there are no conflicts of interest.

References

Andersen, Erik C., and Matthew v. Rockman. 2022. 'Natural Genetic Variation as a Tool for Discovery in Caenorhabditis Nematodes'. *Genetics* 220 (1). https://doi.org/10.1093/GENETICS/IYAB156.

Araya, Carlos L., Trupti Kawli, Anshul Kundaje, Lixia Jiang, Beijing Wu, Dionne Vafeados, Robert Terrell, et al. 2014. 'Regulatory Analysis of the C. Elegans Genome with Spatiotemporal Resolution'. *Nature* 512 (7515). https://doi.org/10.1038/nature13497.

Ben-David, Eyal, James Boocock, Longhua Guo, Stefan Zdraljevic, Joshua S. Bloom, and Leonid Kruglyak. 2021. 'Whole-Organism Eqtl Mapping at Cellular Resolution with Single-Cell Sequencing'. *ELife* 10. https://doi.org/10.7554/eLife.65857.

Boeck, Max E., Chau Huynh, Lou Gevirtzman, Owen A. Thompson, Guilin Wang, Dionna M. Kasper, Valerie Reinke, Ladeana W. Hillier, and Robert H. Waterston. 2016. 'The Time-Resolved Transcriptome of C. Elegans'. *Genome Research* 26 (10): 1441–50. https://doi.org/10.1101/GR.202663.115/-/DC1.

Breitling, Rainer, Yang Li, Bruno M. Tesson, Jingyuan Fu, Chunlei Wu, Tim Wiltshire, Alice Gerrits, et al. 2008. 'Genetical Genomics: Spotlight on QTL Hotspots'. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1000232.

Bulteau, Romain, and Mirko Francesconi. 2022. 'Real Age Prediction from the Transcriptome with RAPToR'. *Nature Methods 2022 19:8* 19 (8): 969–75. https://doi.org/10.1038/s41592-022-01540-0.

Evans, Kathryn S., Marijke H. van Wijk, Patrick T. McGrath, Erik C. Andersen, and Mark G. Sterken. 2021. 'From QTL to Gene: C. Elegans Facilitates Discoveries of the Genetic Mechanisms Underlying Natural Variation'. *Trends in Genetics : TIG* 37 (10): 933. https://doi.org/10.1016/J.TIG.2021.06.005.

Filina, Olga, Burak Demirbas, Rik Haagmans, and Jeroen S. van Zon. 2022. 'Temporal Scaling in C. Elegans Larval Development'. *Proceedings of the National Academy of Sciences of the United States of America* 119 (11). https://doi.org/10.1073/pnas.2123110119.

Francesconi, Mirko, and Ben Lehner. 2014. 'The Effects of Genetic Variation on Gene Expression Dynamics during Development'. *Nature 2014 505:7482* 505 (7482): 208–11. https://doi.org/10.1038/nature12772.

Gaertner, Bryn E, and Patrick C. Phillips. 2010. 'Caenorhabditis Elegans as a Platform for Molecular Quantitative Genetics and the Systems Biology of Natural Variation'. *Center for Ecology and Evolutionary Biology*. https://doi.org/10.1017/S0016672310000601.

Hashimshony, Tamar, Martin Feder, Michal Levin, Brian K. Hall, and Itai Yanai. 2015. 'Spatiotemporal Transcriptomics Reveals the Evolutionary History of the Endoderm Germ Layer'. *Nature*. https://doi.org/10.1038/nature13996.

Hendriks, Gert Jan, Dimos Gaidatzis, Florian Aeschimann, and Helge Großhans. 2014. 'Extensive Oscillatory Gene Expression during C. Elegans Larval Development'. *Molecular Cell* 53 (3): 380–92. https://doi.org/10.1016/J.MOLCEL.2013.12.013.

Jansen, Ritsert C., and Jan Peter Nap. 2001. 'Genetical Genomics: The Added Value from Segregation'. *Trends in Genetics* 17 (7): 388–91. https://doi.org/10.1016/S0168-9525(01)02310-1.

Jovic, Katharina, Jacopo Grilli, Mark G. Sterken, Basten L. Snoek, Joost A.G. Riksen, Stefano Allesina, and Jan E. Kammenga. 2019. 'Transcriptome Resilience Predicts Thermotolerance in Caenorhabditis Elegans'. *BMC Biology* 17 (1): 1–12. https://doi.org/10.1186/S12915-019-0725-6/FIGURES/4.

Jovic, Katharina, Mark G. Sterken, Jacopo Grilli, Roel P.J. Bevers, Miriam Rodriguez, Joost A.G. Riksen, Stefano Allesina, Jan E. Kammenga, and L. Basten Snoek. 2017. 'Temporal Dynamics of Gene Expression in Heat-Stressed Caenorhabditis Elegans'. *PLOS ONE* 12 (12): e0189445. https://doi.org/10.1371/JOURNAL.PONE.0189445.

Kamath, Ravi S., Andrew G. Fraser, Yan Dong, Gino Poulin, Richard Durbin, Monica Gotta, Alexander Kanapin, et al. 2003. 'Systematic Functional Analysis of the Caenorhabditis Elegans Genome Using RNAi'. *Nature 2003 421:6920* 421 (6920): 231–37. https://doi.org/10.1038/nature01278.

Kenyon, Cynthia, Jean Chang, Erin Gensch, Adam Rudner, and Ramon Tabtiang. 1993. 'A C. Elegans Mutant That Lives Twice as Long as Wild Type'. *Nature* 366 (6454). https://doi.org/10.1038/366461a0.

Kim, Dong Hyun, Dominic Grün, and Alexander van Oudenaarden. 2013. 'Dampening of Expression Oscillations by Synchronous Regulation of a MicroRNA and Its Target'. *Nature Genetics* 45 (11). https://doi.org/10.1038/ng.2763.

Levin, Michal, Leon Anavy, Alison G. Cole, Eitan Winter, Natalia Mostov, Sally Khair, Naftalie Senderovich, et al. 2016. 'The Mid-Developmental Transition and the Evolution of Animal Body Plans'. *Nature* 531 (7596): 637–41. https://doi.org/10.1038/NATURE16994.

Li, Yang, Olga Alda Álvarez, Evert W. Gutteling, Marcel Tijsterman, Jingyuan Fu, Joost A.G. Riksen, Esther Hazendonk, et al. 2006. 'Mapping Determinants of Gene Expression Plasticity by Genetical Genomics in C. Elegans'. *PLoS Genetics* 2 (12). https://doi.org/10.1371/journal.pgen.0020222.

Li, Yang, Rainer Breitling, L. Basten Snoek, K. Joeri Van Der Velde, Morris A. Swertz, Joost Riksen, Ritsert C. Jansen, and Jan E. Kammenga. 2010. 'Global Genetic Robustness of the Alternative Splicing Machinery in Caenorhabditis Elegans'. *Genetics* 186 (1). https://doi.org/10.1534/genetics.110.119677.

Martin, Véronique, Sophie Schbath, and Christelle Hennequet-Antier. 2022. 'R Graphics with Ggplot2'.

Mata-Cabana, Alejandro, Francisco Javier Romero-Expósito, Mirjam Geibel, Francine Amaral Piubeli, Martha Merrow, and María Olmedo. 2022. 'Deviations from Temporal Scaling Support a Stage-Specific Regulation for C. Elegans Postembryonic Development'. *BMC Biology* 20 (1): 94. https://doi.org/10.1186/s12915-022-01295-2.

McCarroll, Steven A, Coleen T Murphy, Sige Zou, Scott D Pletcher, Chen-Shan Chin, Yuh Nung Jan, Cynthia Kenyon, Cornelia I Bargmann, and Hao Li. 2004. 'Comparing Genomic Expression Patterns across Species Shared Transcriptional Profile in Aging'. *NATURE GENETICS* 9 (10). https://doi.org/10.1038/ng1291.

Meeuse, Milou WM, Yannick P Hauser, Lucas J Morales Moya, Gert-Jan Hendriks, Jan Eglinger, Guy Bogaarts, Charisios Tsiairis, and Helge Großhans. 2020. ' Developmental Function and State Transitions of a Gene Expression Oscillator in Caenorhabditis Elegans '. *Molecular Systems Biology* 16 (7). https://doi.org/10.15252/msb.20209498.

Michaelson, Jacob J., Salvatore Loguercio, and Andreas Beyer. 2009. 'Detection and Interpretation of Expression Quantitative Trait Loci (EQTL)'. *Methods*. https://doi.org/10.1016/j.ymeth.2009.03.004.

O'Duibhir, Eoghan, Philip Lijnzaad, Joris J Benschop, Tineke L Lenstra, Dik Leenen, Marian JA Groot Koerkamp, Thanasis Margaritis, Mariel O Brok, Patrick Kemmeren, and Frank CP Holstege. 2014. 'Cell Cycle Population Effects in Perturbation Studies'. *Molecular Systems Biology* 10 (6). https://doi.org/10.15252/msb.20145172.

Perez, Marcos Francisco, Mirko Francesconi, Cristina Hidalgo-Carcedo, and Ben Lehner. 2017. 'Maternal Age Generates Phenotypic Variation in Caenorhabditis Elegans'. *Nature* 552 (7683). https://doi.org/10.1038/nature25012.

R Core Team (2022). 2022. 'R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.' Vienna, Austria.

Reinke, Valerie, Harold E. Smith, Jeremy Nance, John Wang, Carrie Van Doren, Rebecca Begley, Steven J.M. Jones, et al. 2000. 'A Global Profile of Germline Gene Expression in C. Elegans'. *Molecular Cell* 6 (3): 605–16. https://doi.org/10.1016/S1097-2765(00)00059-9.

Snoek, Basten L., Mark G. Sterken, Roel P.J. Bevers, Rita J.M. Volkers, Arjen van't Hof, Rachel Brenchley, Joost A.G. Riksen, Andrew Cossins, and Jan E. Kammenga. 2017. 'Contribution of Trans Regulatory EQTL to Cryptic Genetic Variation in C. Elegans'. *BMC Genomics* 18 (1). https://doi.org/10.1186/s12864-017-3899-8.

Snoek, Basten L., Mark G. Sterken, Margi Hartanto, Albert Jan Van Zuilichem, Jan E. Kammenga, Dick De Ridder, and Harm Nijveen. 2020. 'WormQTL2: An Interactive Platform for Systems Genetics in Caenorhabditis Elegans'. *Database: The Journal of Biological Databases and Curation* 2020: 149. https://doi.org/10.1093/DATABASE/BAZ149.

Snoek, Basten L., Mark G. Sterken, Harm Nijveen, Rita J.M. Volkers, Joost Riksen, Philip C. Rosenstiel, Hinrich Schulenburg, and Jan E. Kammenga. 2021. 'The Genetics of Gene Expression in a Caenorhabditis Elegans Multiparental Recombinant Inbred Line Population'. *G3 Genes|Genomes|Genetics* 11 (10). https://doi.org/10.1093/G3JOURNAL/JKAB258.

Snoek, Basten L., Rita J.M. Volkers, Harm Nijveen, Carola Petersen, Philipp Dirksen, Mark G. Sterken, Rania Nakad, et al. 2019. 'A Multi-Parent Recombinant Inbred Line Population of C. Elegans Allows Identification of Novel QTLs for Complex Life History Traits'. *BMC Biology* 17 (1). https://doi.org/10.1186/S12915-019-0642-8.

Snoek, L. Basten, Mark G. Sterken, Rita J.M. Volkers, Mirre Klatter, Kobus J. Bosman, Roel P.J. Bevers, Joost A.G. Riksen, Geert Smant, Andrew R. Cossins, and Jan E. Kammenga. 2014. 'A Rapid and Massive Gene Expression Shift Marking Adolescent Transition in C. Elegans'. *Scientific Reports 2014 4:1* 4 (1): 1–5. https://doi.org/10.1038/srep03912.

Snoek, L. Basten, K. Joeri Van Der Velde, Danny Arends, Yang Li, Antje Beyer, Mark Elvin, Jasmin Fisher, et al. 2013. 'WormQTL—Public Archive and Analysis Web Portal for Natural Variation Data in Caenorhabditis Spp'. *Nucleic Acids Research* 41 (Database issue): D738. https://doi.org/10.1093/NAR/GKS1124.

Spencer, W. Clay, Georg Zeller, Joseph D. Watson, Stefan R. Henz, Kathie L. Watkins, Rebecca D. McWhirter, Sarah Petersen, et al. 2011. 'A Spatial and Temporal Map of C. Elegans Gene Expression'. *Genome Research* 21 (2). https://doi.org/10.1101/gr.114595.110.

Sterken, Mark G., Roel P.J. Bevers, Rita J.M. Volkers, Joost A.G. Riksen, Jan E. Kammenga, and Basten L. Snoek. 2020. 'Dissecting the EQTL Micro-Architecture in Caenorhabditis Elegans'. *Frontiers in Genetics* 11. https://doi.org/10.3389/fgene.2020.501376.

Sterken, Mark G., Linda van Bemmelen van der Plaat, Joost A.G. Riksen, Miriam Rodriguez, Tobias Schmid, Alex Hajnal, Jan E. Kammenga, and Basten L. Snoek. 2017. 'Ras/MAPK Modifier Loci Revealed by EQTL in Caenorhabditis Elegans'. *G3: Genes, Genomes, Genetics* 7 (9). https://doi.org/10.1534/g3.117.1120.

Sterken, Mark G., L. Basten Snoek, Jan E. Kammenga, and Erik C. Andersen. 2015. 'The Laboratory Domestication of Caenorhabditis Elegans'. *Trends in Genetics* 31 (5): 224–31. https://doi.org/10.1016/J.TIG.2015.02.009.

Viñuela, Ana, L. Basten Snoek, Joost A.G. Riksen, and Jan E. Kammenga. 2012. 'Aging Uncouples Heritability and Expression-Qtl in Caenorhabditis Elegans'. *G3: Genes, Genomes, Genetics* 2 (5): 597–605. https://doi.org/10.1534/G3.112.002212/-/DC1/FIGURES5.PDF.

Viñuela, Ana, L. Basten Snoek, Joost A.G. Riksen, and Jan E. Kammenga. 2010. 'Genome-Wide Gene Expression Regulation as a Function of Genotype and Age in C. Elegans'. *Genome Research* 20 (7): 929–37. https://doi.org/10.1101/GR.102160.109.

Volkers, Rita J.M., L. B. Snoek, Caspara J.van Hellenberg Hubar, Renata Coopman, Wei Chen, Wentao Yang, Mark G. Sterken, Hinrich Schulenburg, Bart P. Braeckman, and Jan E. Kammenga. 2013. 'Gene-Environment and Protein-Degradation Signatures Characterize Genomic and Phenotypic Diversity in Wild Caenorhabditis Elegans Populations'. *BMC Biology* 11 (1): 1–13. https://doi.org/10.1186/1741-7007-11-93/FIGURES/4.

Wijk, Marijke H van, Joost A G Riksen, Mark Elvin, Gino B Poulin, Muhammad I Maulana, Jan E Kammenga, Basten L Snoek, and Mark G Sterken. 2023. ' Cryptic Genetic Variation of EQTL Architecture Revealed by Genetic Perturbation in C. Elegans '. *G3 Genes|Genomes|Genetics*. https://doi.org/10.1093/g3journal/jkad050.

Table 1: Significant effects detected by the single marker, additive age and interaction models. An FDR of 0.05 is used unless otherwise specified. * Applies to additive developmental age term of interaction model.

	Single marker model (SMM)	Additive age model (AAM)	Interaction model, marker term (IMM)	Interaction model, interaction term (IMI)
eQTLs detected (FDR <0.05)	18605	8666	6886	1161
Cis/Local	2136	1954	1814	92
Trans/Distant	16469	6712	5072	1069
Cis/*Trans* ratio	0.13	0.29	0.36	0.08
Transcripts w/ eQTL	8875	6133	5180	904
eQTLs detected (FDR	27445	13369	10938	4276

<0.1)				
Transcripts affected by developmental process	-	9473	*8806	*8806

Figure 1: Developmental age estimation using PCA. A) PCA results. Points are the projections of the mpRILs on PC1 (x-axis) and PC2 (y-axis). Color indicates the Z-score of the center log ratio of fpkm values of *vit-2*, a developmental indicator gene(L. B. Snoek et al. 2014). **B)** Gene expression of the 53 developmental indicator genes from Snoek *et al*., 2014 with expression in this dataset. The x-axis shows the projections of the mpRILs on PC1. The y-axis shows the Z-score of the center log ratio of fpkm values. Colored trendlines correspond to a generalized additive model fit to the expression values of each developmental indicator gene.

Figure 2: Distribution of eQTLs over the chromosome according to the different models. X-axis shows the position on the genome. Y-axis shows the number of eQTLs that map to this position. **A)** SMM. Red arrows indicate the three largest hotspots. **B)** AAM. **C)** IM. Cyan and red show eQTLs according to the IMM and IMI respectively. Counts are in terms of number of transcripts mapping to the location.

Figure 3: Distribution of developmental ages per allele at the three largest hotspots. Hotspots are indicated in figure 7 by red arrows. P-values are obtained with single marker linear model of PC1 ~ marker genotype (see **Figure S9**).

Figure 4: Schematic of how some hotspots cause gene expression variation. Some hotspots influence the developmental speed, causing the transcriptional program of the genotypes to be at distinct developmental ages at a given moment in chronological time. This in turn causes expression variation in the large number of transcripts whose expression level depends on the developmental age. The developmental age is the mediator through which these loci affect gene expression.

Figure 5: Comparison p-values obtained with the SMM and AAM. A) -log10(P) of SMM versus AAM, i.e. higher value corresponds to lower p-value. The x-axis shows the lowest marker p-value per chromosome per transcript obtained with the SMM (72174 total plotted p-values). The y-axis shows the same for the AAM. Since we call a maximum of one eQTL per transcript per chromosome, all eQTLs are represented in the plot. The red lines show the subsection of the plot depicted in B. The lower black line is the identity line, such that eQTLs above the line have a lower p-value according to the AAM, whereas eQTLs below the line have a lower p-value according to the SMM. eQTLs above the top black line are strong developmental effect eQTLs (p-value ten orders of magnitude lower with AAM as compared to SMM). **B)** Zoomed-in subsection of A. Points are colored by whether the eQTL was detected with none (brown), both (purple) or one of the models (orange for SMM-only eQTLs and cyan for AAM-only eQTLs).

Figure 6: Gene expression of three example strong developmental effect eQTLs. Line plots show center log ratio of fpkm values (y-axis) over the developmental age (x-axis) for three examples of strong developmental effect

eQTLs. Colors correspond to the genotype at the eQTL position. Boxplots show the magnitude of gene expression without the context of the developmental age.

Figure 7: Examples of differentially detected eQTLs. Dashed line and dashed boxplot indicate no significant difference between alleles. **A)** SMM-only eQTL affecting the nspd-3 gene. The SMM only considers mean gene expression differences between alleles (solid boxplot). For SMM-only eQTLs such mean differences occur because one genotype has on average a lower developmental age than the other. In this case the blue genotype is on average developmentally younger. The alleles shown correspond to the hotspot locus on ~3.4 Mb on chromosome X whose developmental age distribution is depicted in **Figure 3A.** For SMM-only eQTLs correcting for the developmental age reveals there is no significant difference in gene expression at any point along the developmental trajectory of the mpRILs (dashed lines). **B)** AAM-only eQTL affecting the sams-5 gene. For the AAM-only eQTLs there is no significant mean gene expression difference between the alleles (dashed boxplot). However, when considering the context of developmental age differences, a clear difference between the genotypes is revealed (solid lines).

Figure 8: Density plot of the partial eta squared of marker effects for subsets of eQTLs according to the SMM and AAM. For the eQTLs in the subset we selected the marker with the lowest p-value according to each of the models and calculated the partial eta squared of the marker effect using the respective model. **A)** SMM-only eQTLs. **B)** AAM-only eQTLs.

Figure 9: A selection of representative eQTLs detected with IMI term. Colors correspond to alleles. Lines are the best fit of the IM model (grey area is 95% CI) to gene expression over developmental age. Boxplots show the magnitude of gene expression without the context of the developmental age. **A)** eQTL that, depending on genotype, causes either increasing gene expression as development progresses or a constant expression level. **B)** eQTL that causes the slope of gene expression over developmental age to be higher for one genotype compared to the other. **C)** eQTL for which one genotype exhibits negative regulation of gene expression while the other genotype shows constant gene expression.

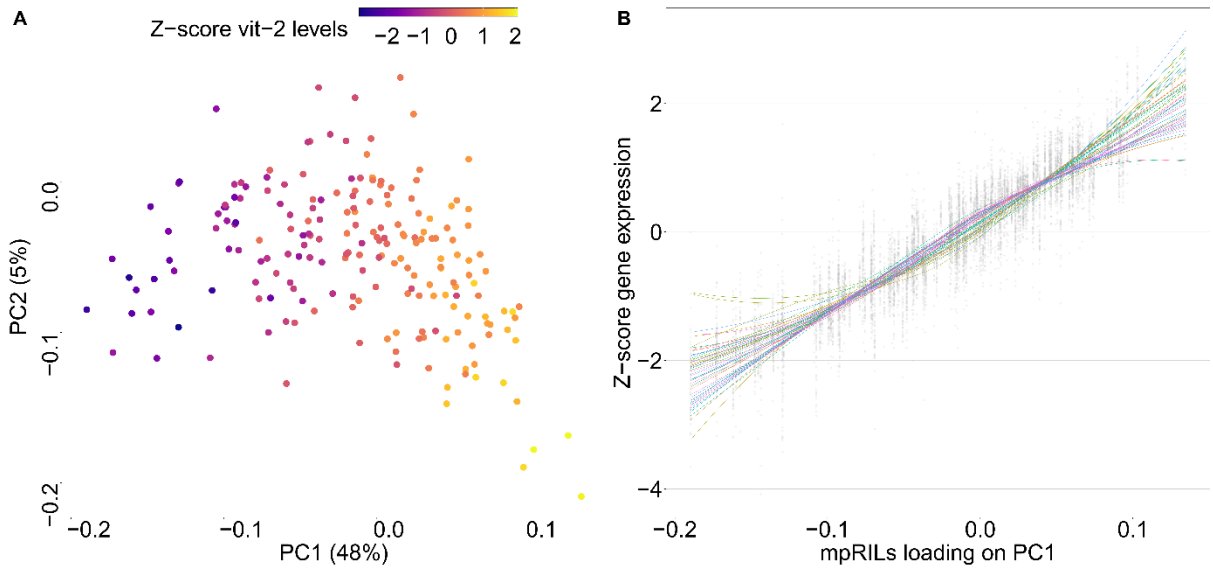

Figure 1
159x75 mm (x DPI)

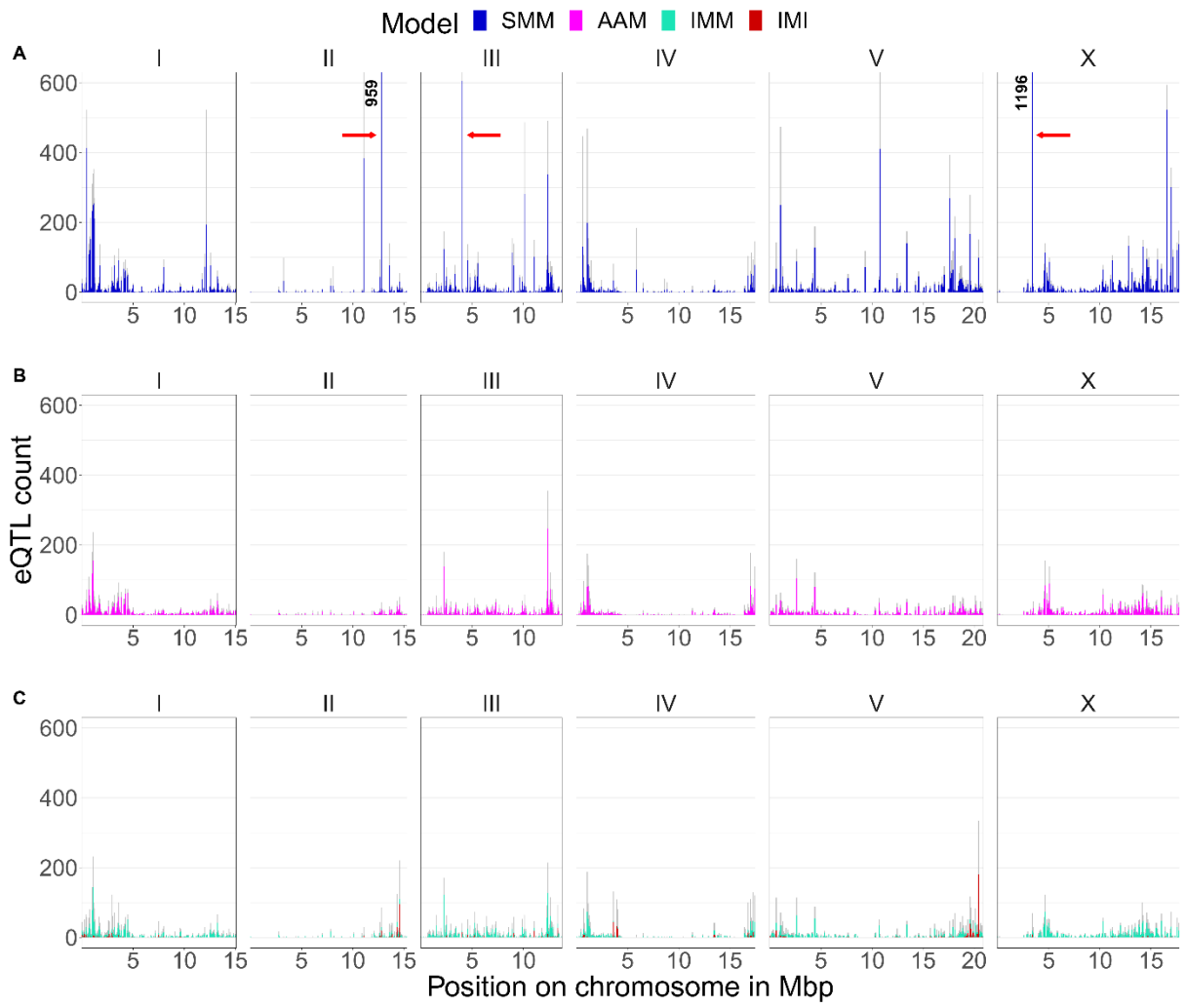

Figure 2
159x136 mm (x DPI)

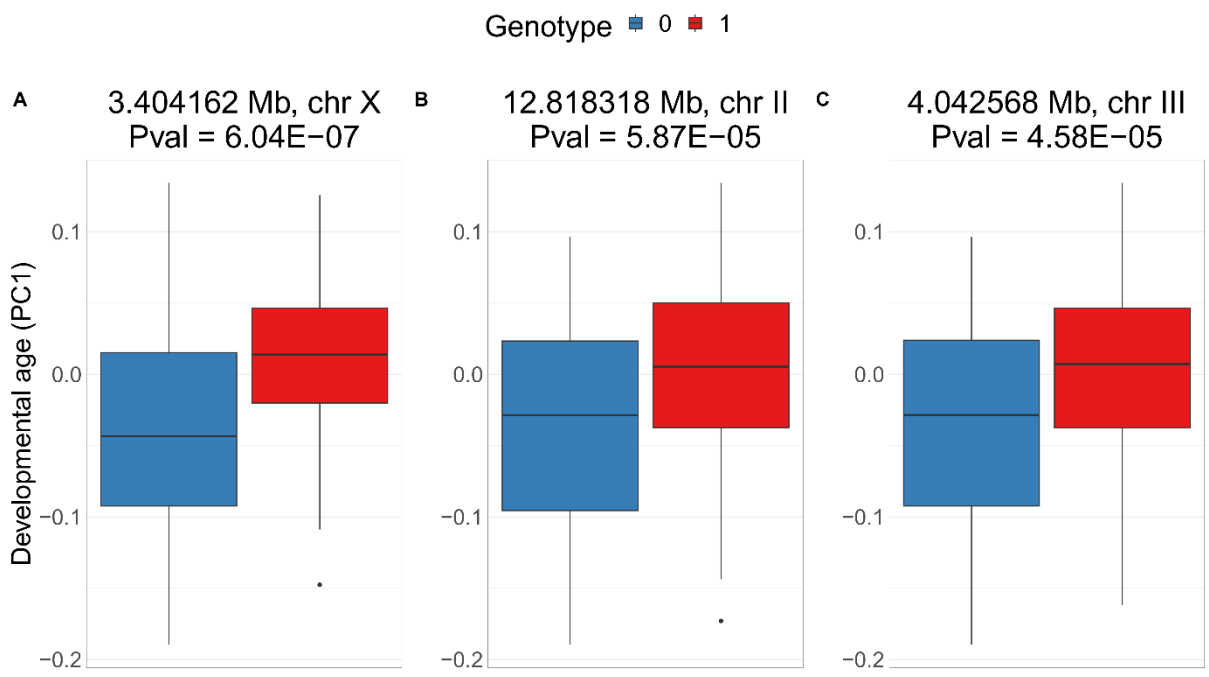

Figure 3
159x89 mm (x DPI)

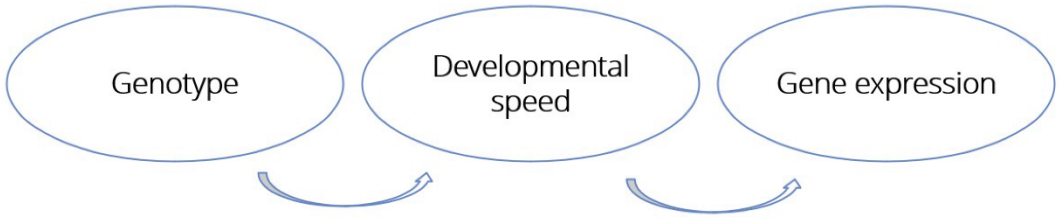

Figure 4
159x97 mm (x DPI)

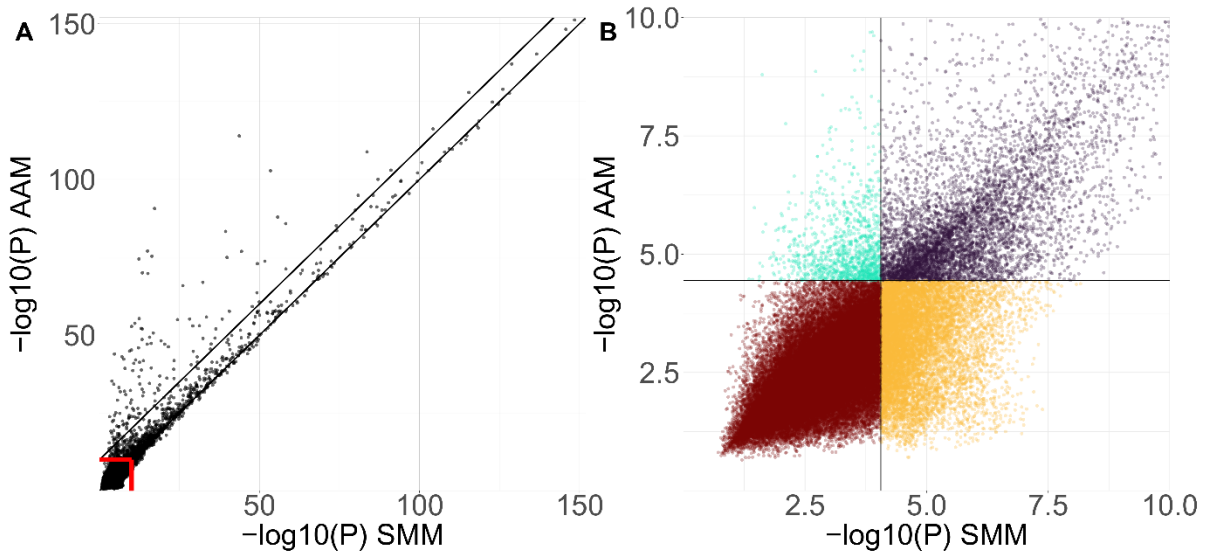

Figure 5
159x74 mm (x DPI)

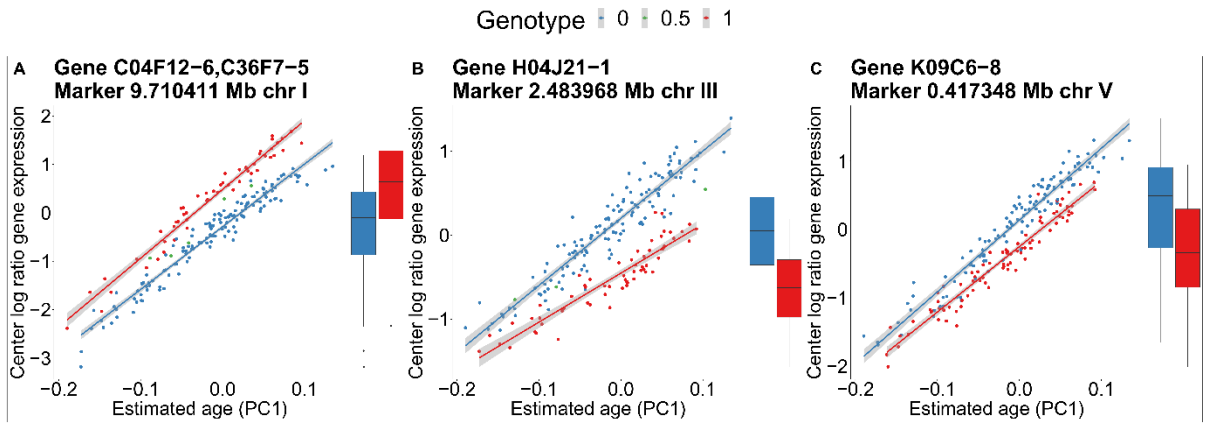

Figure 6
159x56 mm (x DPI)

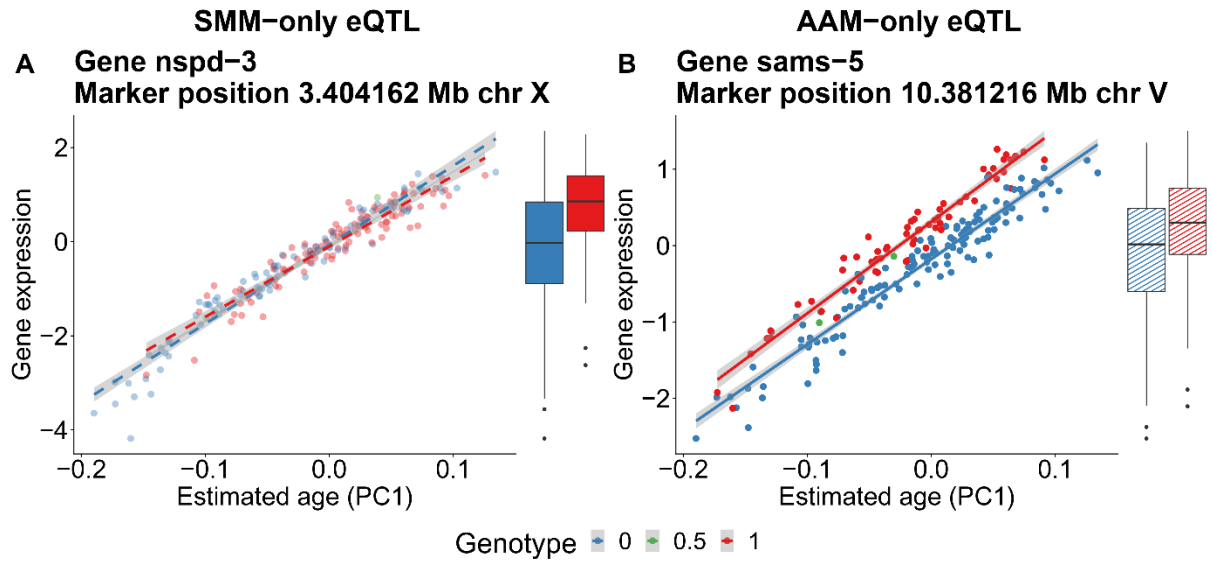

Figure 7
159x75 mm (x DPI)

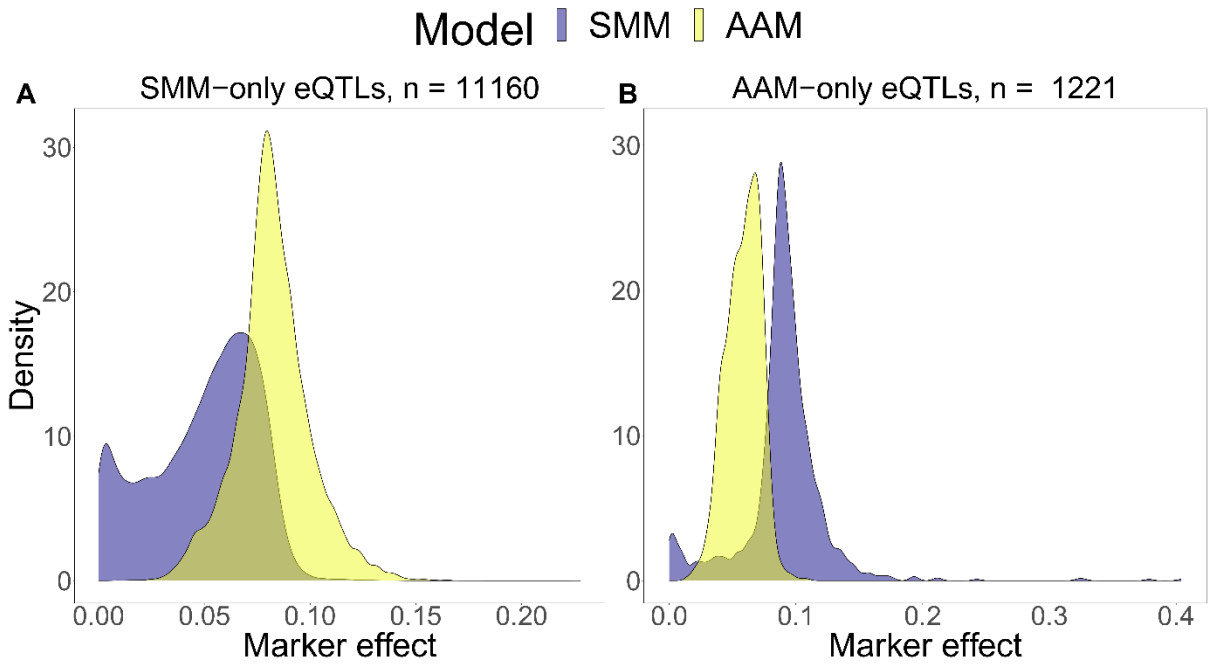

Figure 8
159x89 mm (x DPI)

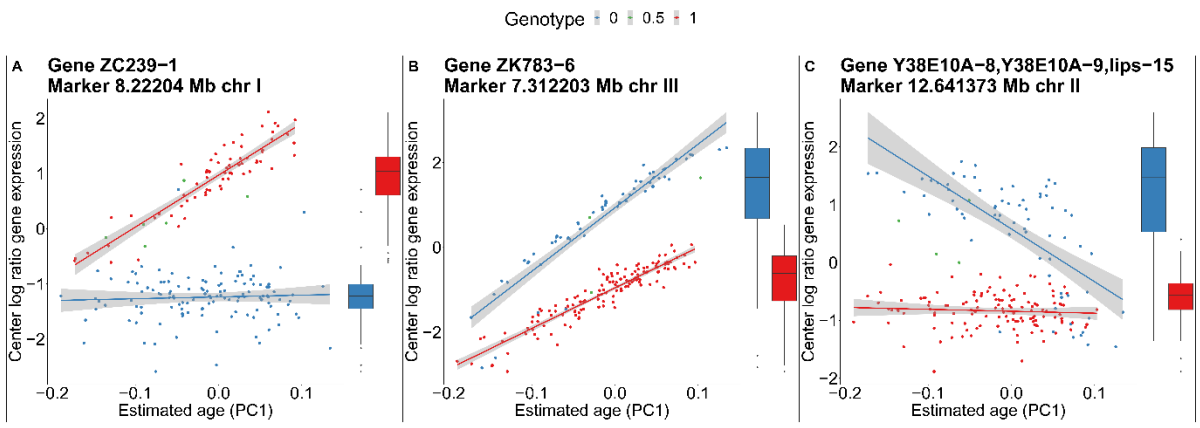

Figure 9
159x55 mm (x DPI)