The genetics of gene expression in a *C. elegans* multi parental recombinant inbred line population.

Basten L. Snoek[1,2,+], Mark G. Sterken[1], Harm Nijveen[3], Rita J.M. Volkers[1], Joost Riksen[1], Philip C. Rosenstiel[4,5], Hinrich Schulenburg[6,7,+], Jan E. Kammenga[1,+].

[1] Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

[2] Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

[3] Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, NL-6708 PB Wageningen, The Netherlands

[4] Institute for Clinical Molecular Biology, University of Kiel, 24098 Kiel, Germany

[5] Competence Centre for Genomic Analysis (CCGA) Kiel, University of Kiel, 24098 Kiel, Germany

[6] Zoological Institute, University of Kiel, 24098 Kiel, Germany

[7] Max Planck Institute for Evolutionary Biology, August-Thienemann-Str. 2, 24306 Ploen, Germany

[+] Corresponding authors

Keywords: Multi parental RILs, expression QTL, eQTL, SNPs, *C. elegans*

l.b.snoek@uu.nl; 0000-0001-5321-2996
mark.sterken@wur.nl; 0000-0001-7119-6213
harm.nijveen@wur.nl; 0000-0002-9167-4945
rvolkers@hotmail.com;
joost.riksen@wur.nl;
p.rosenstiel@mucosa.de;
hschulenburg@zoologie.uni-kiel.de; 0000-0002-1413-913X
jan.kammenga@wur.nl; 0000-0003-4822-4436

Abstract

Studying genetic variation of gene expression provides a powerful way to unravel the molecular components underlying complex traits. Expression QTL studies have been performed in several different model species, yet most of these linkage studies have been based on genetic segregation of two parental alleles. Recently we developed a multi-parental segregating population of 200 recombinant inbred lines (mpRILs) derived from four wild isolates (JU1511, JU1926, JU1931 and JU1941) in the nematode *Caenorhabditis elegans*. We used RNA-seq to investigate how multiple alleles affect gene expression in these mpRILs. We found 1,789 genes differentially expressed between the parental lines. Transgression, expression beyond any of the parental lines in the mpRILs, was found for 7,896 genes. For expression QTL mapping almost 9,000 SNPs were available. By combining these SNPs and the RNA-seq profiles of the mpRILs, we detected almost 6,800 eQTLs. Most *trans*-eQTLs (63%) co-locate in six newly identified *trans*-bands. The *trans*-eQTLs found in previous 2-parental allele eQTL experiments and this study showed some overlap (17.5%-46.8%), highlighting on the one hand that a large group of genes is affected by polymorphic regulators across populations and conditions, on the other hand it shows that the mpRIL population allows identification of novel gene expression regulatory loci. Taken together, the analysis of our mpRIL population provides a more refined insight into *C. elegans* complex trait genetics and eQTLs in general, as well as a starting point to further test and develop advanced statistical models for detection of multi-allelic eQTLs and systems genetics studying the genotype-phenotype relationship.

Introduction

Investigation of the genotype-phenotype relationship is at the heart of genetic research. The detection and description of allelic variants and genetic mechanisms have been a demanding task due to the quantitative nature of most phenotypic variation. Quantitative trait locus (QTL) mapping has been one of the methods of choice for finding the loci on which these allelic variants can be found. Many functional polymorphisms in plants and animals, including many model species such as model nematode *C. elegans*, have been discovered using QTL mapping [1-25]. Over the last decade molecular phenotypes such as transcript levels, protein levels and metabolites have also been used in QTL mapping [26-32]. Heritable variation in these molecular phenotypes often plays a role in heritable phenotypic variation [10, 30, 33]. Mapping expression QTLs (eQTLs) can provide insight into the transcriptional architecture of complex traits and have been conducted in model species such as *Arabidopsis thaliana* and *C. elegans* as well as several other taxa [26, 28-31, 34-41].

Most eQTL studies have been done on populations of recombinant inbred lines (RILs) originating from a cross between two different parental genotypes [26, 28-31, 34-40]. Inclusion of more than two parents can capture more genetic variation, increasing the number of detected QTLs, potentially allowing more precise mapping and therefore reducing the number of potential candidate causal genes to be verified [42]. Such a strategy was first used for *Arabidopsis* by developing a Multiparent Advanced Generation Inter-Cross (MAGIC) lines population consisting of 527 RILs developed from 19 different parental accessions [43]. Several other MAGIC populations have been developed since then for a range of species, including *C. elegans* [44-46].

Recently multi parental RIL (mpRILs) populations have been developed in *C. elegans* [45, 46]. These populations have been created using other strains than the most frequently used N2 strain and the Hawaiian CB4856 strain [26-31, 37, 47-61]. In this study we used the population of 200 mpRILs, derived from an advanced cross between four wild-types: JU1511 and JU1941 isolated from Orsay (France) and JU1926 and JU1931 isolated from Santeuil (France) (kindly provided by MA Félix, Paris, France; [45, 62]). In a previous study, the RNA-seq data of these mpRILs was used to obtain almost 9,000 SNPs variable

between the four parental genotypes and used to identify QTLs for life-history traits [45]. The RNA was sampled from the mpRILs grown under standardised conditions (24°C, OP50, 48h after bleaching) and obtained from animals from two 6-cm dishes, with one RNA-seq replicate per mpRIL and two per parental isolate. To investigate the effect of multiple genetic backgrounds on gene expression, we used the RNA-seq data to associate gene expression levels to genetic variants present in the population. We compared the gene expression level differences between the parental wild isolates, calculated transgression, as well as heritability and mapped eQTLs. We identified six *trans*-bands, hotspots at which many *trans*-eQTLs co-locate, which we further studied by gene ontology enrichment. Lastly, we compared the eQTLs found in this study to the eQTLs found in previous eQTL studies in *C. elegans* [26, 28, 30, 31, 37, 39]. Together these results present the first insights into the genetic architecture of gene expression in a *C. elegans* multi parental RIL population.

Results

Gene expression differences between the parental lines

To study the effect of genetic variation on gene expression we used RNA-seq on a population of 200 multi parental recombinant inbred lines (mpRILs) [45], made from a cross between four parental lines isolated from Orsay, France (JU1511, JU1941) and Santeuil, France (JU1926, JU1931) [62]. The animals used were grown on two 6-cm dishes (24°C, OP50, 48h after bleaching) per sample pooled for RNA isolation, with one RNA-seq replicate per mpRIL and two per parental isolate. First, we determined the expression differences between the parental lines (**Supplement table 1**). Of the 12,029 detected transcripts we found 1,789 genes differently expressed between at least one parental pair (TukeyHSD $p < 0.001$; FDR < 0.05; **Figure 1**). Of the four strains, JU1926 was most different when compared to the other lines, with 409 genes being differently expressed between JU1926 and the other three lines. Thereafter, JU1941 was most different from the remaining two lines. These differences in gene expression between the parental lines are likely genotype dependent.

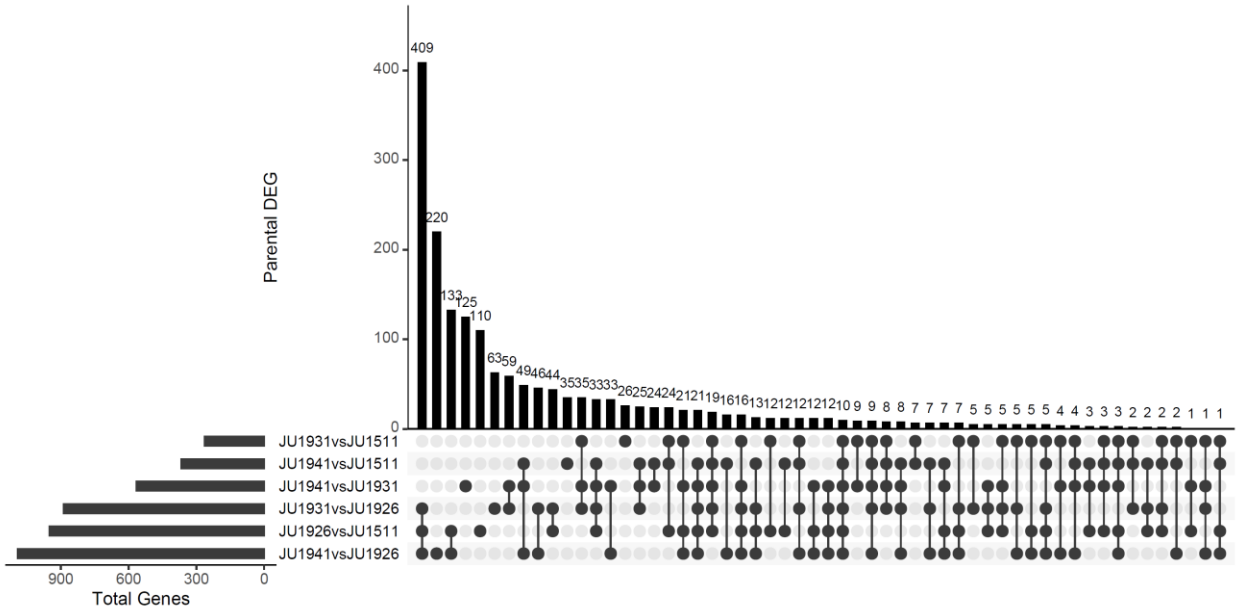

Figure 1: Gene expression differences between the four mpRIL parental lines. Upset plot shows the pairwise comparisons and the overlap between the pairs (TukeyHSD; $p<0.001$; FDR=0.05). The horizontal bar plot shows the number of differentially expressed genes per parental pair, while the vertical bar plot indicates the number of shared differentially expressed genes per comparison. For example, an overlap of 409 genes was found between the three comparisons that include the JU1926 parental line, which shows that JU1926 differed most from all other lines.

Transgression and Heritability

To explore the variation in gene expression between the different parental and mpRIL genotypes we applied principal component analysis on the $\log_2$ gene expression ratios (**Figure2A**). Here we can see that the expression variation in many of the mpRILs extends beyond the parental expression variation, which suggests transgression. We quantified this and found transgression for 7,896 genes (FDR = 0.08; **Figure2B-C**, **Supplement table 2**). Notably, most transgression was one-sided, showing increased expression level beyond the highest expression level found in the parental lines. This suggests that multiple segregated loci are involved in regulating the transcription in the mpRILs. Transgression often indicates that the trait variation, in this case gene expression levels, is heritable. We calculated the narrow sense heritability (h^2) and found significant h^2 for expression variation of 9,500 genes (FDR = 0.05; **Figure 2D**, **Supplement table 2**). Most gene expression variation showed a h^2 below 0.5, indicating that part of the variation is caused by other factors than additive genetic effects. These factors could be technical, environmental but also more complex genetic interactions such as epistasis.

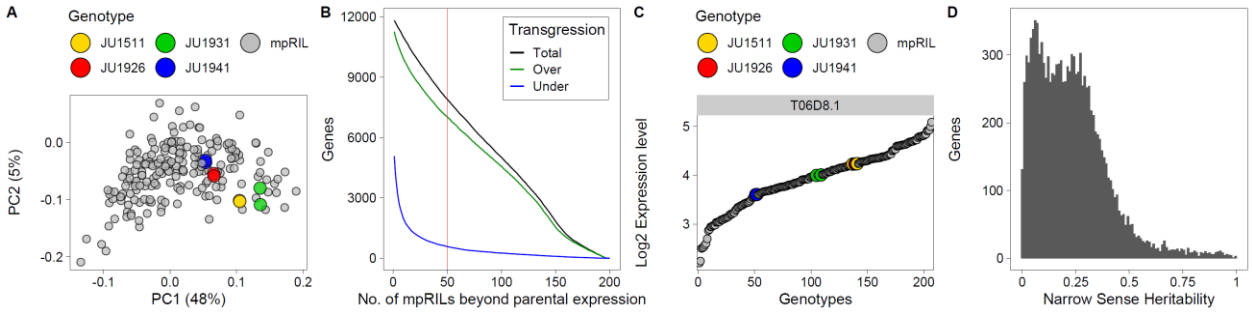

Figure 2: Gene expression variation in the mpRILs and parental genotypes. A) Principal component analysis (PCoA) of the $\log_2$ ratios, mpRILs shown in grey, parental lines shown in colour. **B)** Transgression: number of mpRILs beyond the parental expression level (x-axis) against the number of genes (y-axis). The mpRILs below (under) the lowest parental expression level in blue, mpRILs over the highest parental expression level in green and the sum of these (total) in black. **C)** Example of two-sided transgression for expression levels of gene T06D8.1. **D)** Genes with significant narrow sense heritability (h^2) and the distribution of heritable variation of gene expression variation at FDR = 0.05.

Expression QTLs

To find the loci involved in gene expression variation between the mpRILs we used a single marker QTL model. We found 6,784 eQTLs (one eQTL per gene, $-\log_{10}(p) > 5.35$; FDR = 0.01), of which 929 were *cis*-

and 5,855 *trans*-eQTLs (**Table 1**; **Figure 3; Supplement table 2**). Most *cis*-eQTLs were found on chromosome V and most *trans*-eQTLs on chromosomes I and X. For both *cis*- and *trans*-eQTLs, fewest where found on chromosome II and IV. The SNP Distribution Pattern (SDP) groups SNPs with the same distribution in the parental lines. When the SDP is considered, many of the *cis*-eQTLs were found to have an effect where either the JU1511 or JU1941 allele was different from the three other parental genotypes. For the *trans*-eQTLs the largest groups also show this allelic difference or those SNPs that distinguish JU1511/JU1941 from JU1926/JU1931. A substantial group was found for the JU1931 allele, whereas hardly any were found for the JU1926 specific SNPs. The lack of JU1926 is somewhat surprising as it had the most differentially expressed genes (DEG) in the comparison of the parental lines, however we found much more genes with eQTLs than being DEG in the parental comparison. These are much more likely to be caused by new allelic combinations present in the mpRILs. Overall, the majority of the eQTLs are found on a few major effect loci with a specific SDP linkage (**Figure 3**). Moreover, comparing the h^2 to the eQTLs showed that genes with an eQTL have a much higher h^2 than those without an eQTL, where genes with an $h^2 > 0.25$ almost all have an eQTL (**Figure 4**). Comparing *cis*- and *trans*-eQTLs showed that genes with a *cis*-eQTL have a higher h^2 on average, yet the h^2 distributions of *cis*- and *trans*-eQTLs are overlapping.

Table 1: eQTLs per type (*cis*/*trans*) per chromosome per SNP Distribution Pattern (SDP).

	Cis							*Trans*						
SDP	I	II	III	IV	V	X	Tot	I	II	III	IV	V	X	Tot
12 JU1511 & JU1926 vs. JU1931 & JU1941	14	2	17	2	23	13	71	35	0	1	1	3	67	107
13 JU1511 & JU1931 vs. JU1926 & JU1941	6	1	14	39	41	3	104	0	2	106	13	15	27	163
14 JU1511 & JU1941 vs. JU1926 & JU1931	12	0	19	0	53	11	95	1373	0	119	5	103	44	1644
JU1511	37	32	61	14	18	81	243	457	28	211	20	9	430	1155
JU1926	0	32	4	59	5	1	101	5	44	5	26	10	5	95
JU1931	8	0	15	3	81	1	108	31	0	12	21	919	2	985
JU1941	76	0	66	5	38	22	207	150	1	155	35	94	1271	1706
Total	153	67	196	122	259	132	**929**	2051	75	609	121	1153	1846	**5855**

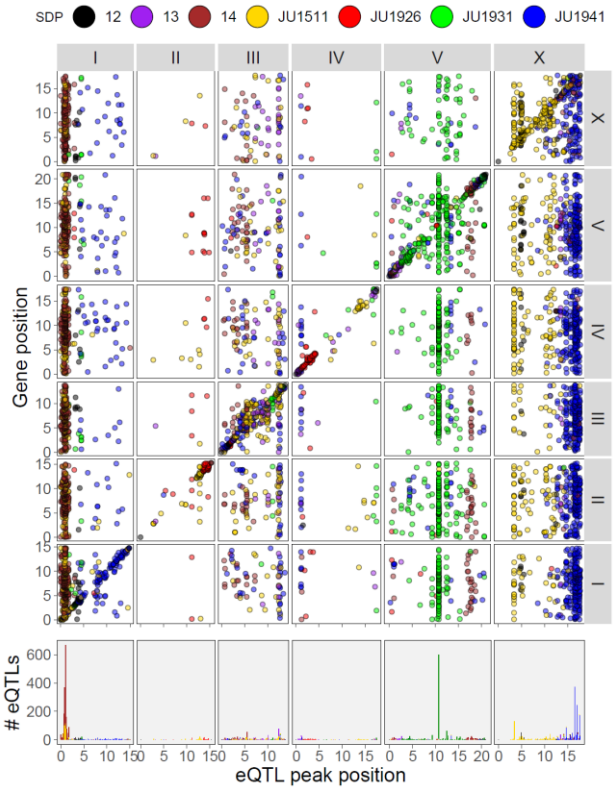

Figure 3: *Cis*/*Trans* plot of the identified eQTLs. eQTL position shown on the x-axis, gene position shown on the y-axis (upper plot) or number of eQTLs (bottom plot). SDP shown in colour, chromosomes shown in the grey strips on top and on the right of the panels.

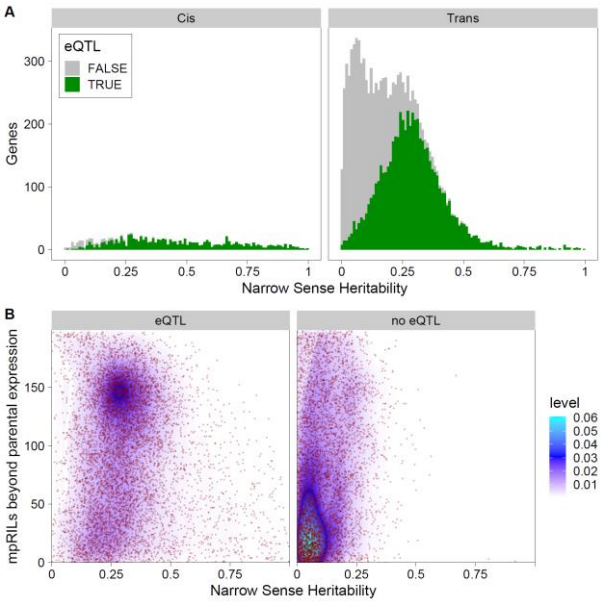

Figure 4: Relation between eQTLs, transgression and Narrow Sense Heritability (h^2). A) Narrow Sense Heritability (h^2; x-axis), distribution in genes (y-axis) with *cis*- and *trans*-eQTLs, significance of the eQTLs is TRUE (green) when > 5.35 and FALSE (grey) otherwise. **B)** Relation between Narrow Sense Heritability (h^2; x-axis) and transgression (y-axis) for genes with and without a significant eQTL, individual datapoints shown in red, colour gradient indicates datapoint density.

Trans-bands

A large majority of the *trans*-eQTLs (3,704; 63% of all *trans*-eQTLs) were found in six hotspots, so called *trans*-bands (TBs) (number of *trans*-eQTLs > 100, window 1Mbp to both sides; **Table 2**; **Figure 3**). Two TBs were found on chromosome I, one on chromosome V and three on chromosome X. The two TBs on chromosome I co-located but were linked to different SDP: the SDP 14 (JU1511/JU1941 vs JU1926/JU1931) and SDP JU1511 (*vs*. the rest). The TB on chromosome V was linked to SDP JU1931 and the three TBs found on chromosome X were linked to SPD JU1511 and JU1941.

Table 2: Descriptive overview of the 6 identified trans-bands. SNP Distribution Pattern (SDP), Chromosome, Peak position and left and right borders in Mega-base pairs. Selection of enriched GO terms from supplement table 3 and overlap with phenotypic QTLs found in Snoek *et al* 2019 [45].

	SDP	CHR	Peak (Mbp)	Left (Mbp)	Right (Mbp)	eQTLs	**GO Enrichment (selection from enrichment table)**	**Phenotypic QTL** (in Snoek *et al.* 2019 [45])
TB1	14 (JU1511 & JU1941 vs JU1926 & JU1931)	I	1.03	0.03	2.03	1339	thermosensory behaviour, negative regulation of engulfment of apoptotic cell, DNA replication, embryonic body morphogenesis, establishment or maintenance of actin cytoskeleton polarity, muscle fiber development, epidermis development, response to unfolded protein and, molting cycle, collagen and cuticulin-based cuticle,	Population growth on *Erwinia* and on *B. thuringiensis*
TB2	JU1511	I	0.83	0	1.83	443	regulation of protein stability, regulation of vulval development, DNA replication, anaphase-promoting complex and, microtubule polymerization	NA
TB3	JU1931	V	10.74	9.74	11.74	607	hemidesmosome assembly, external side of plasma membrane and, negative regulation of response to oxidative stress	NA
TB4	JU1511	X	3.40	2.40	4.40	133	few	heat-shock sensitivity
TB5	JU1941	X	14.69	13.69	15.64	225	few	population growth on *B. thuringiensis*
TB6	JU1941	X	16.60	15.65	17.6	957	embryonic body morphogenesis, DNA replication, integral component of peroxisomal membrane, anaphase-promoting complex, endosome, phagocytic vesicle membrane, neuronal signal transduction, response to anoxia, cuticle pattern formation, cell fate commitment, hemidesmosome associated protein complex and, response to lipid	sensitivity to oxidative stress

GO enrichment

To study the effect of TBs on biological function we used GO term enrichment (**Table 2**, **Supplement table 3**). Each of the TBs was linked to mostly different sets of GO terms, suggesting an effect on different parts of *C. elegans* biology. The genes mapping to TB1 on chromosome I were enriched for behaviour and muscle and epidermis development GO categories. The genes mapping to TB2 on chromosome I were enriched for the GO term "vulval development", among others. The genes with a *trans*-eQTL on TB3 on chromosome V were enriched for GO terms associated with oxidative stress. The genes mapping to TB4 and TB5 on chromosome X only showed a few GO terms to be enriched and the genes mapping to TB6 on chromosome X were enriched for the GO term "response to anoxia" and many more. This shows that these TBs can be involved in several developmental processes and in the interaction with the environment.

Overlap with other eQTL experiments

To investigate if the genes with eQTLs found in the present mpRIL study also had eQTLs in other studies, we compared them with the studies found in WormQTL2 (**Table 3**; [26, 28, 30, 31, 37, 39, 56]). In general, we found that a substantial group of genes with a trans eQTL in any of the studies had an eQTL in our mpRIL experiment (26.5% - 36.9%). The groups of genes with *trans*-eQTLs show much higher overlap than the genes with a *cis*-eQTL in any of the experiments (10.2% - 20.0%). Around a third of the genes with a *trans*-eQTL in Vinuela & Snoek *et al.* 2010 and Snoek & Sterken *et al.* 2017 also showed a *trans*-eQTL in the mpRILs, with numbers almost equal between developmental stages and treatments. Slightly fewer overlapping genes with eQTLs were found with Rockman *et al.* 2010 and Sterken *et al.* 2017. Comparing the experiments performed with the same N2 x CB4856 in the same lab, Li *et al* 2006, Vinuela & Snoek *et al.* 2010, Snoek & Sterken *et al* 2017, shows that environmental conditions and developmental stage only have a small effect on the global overlap and difference between *cis*-and *trans*-eQTLs. As the genetic backgrounds of the mpRILs are different from the N2 x CB4856 populations used in the other experiments, the low percentage of overlapping *cis*-eQTLs could be expected. The large group of genes with a *trans*-eQTL in both experiments shows that the expression levels of a substantial group of genes are more prone

to be affected by genetic variation independent of environment or developmental stage, while the loci involved are most likely different in each experiment/condition [28, 30, 31].

Table 3: **Overlapping eQTLs** between this mpRIL experiment and the RIL experiments available in WormQTL2 [63]. Percentages indicate the percentage of eQTLs found in the indicated experiment that are also found in the mpRILs eQTLs. Threshold used for the eQTL experiments in this table: -log10(p) > 3.5; *Cis*-eQTLs were called if the peak of the eQTL was within 1Mbp of the gene start, otherwise it was called a *trans*-eQTL.

eQTL experiment	**Total *Cis***	***Cis* Overlap(%)**	**Total *Trans***	***Trans* Overlap(%)**
Li *et al.* 2006 16°C [37]	240	14.6	817	31.6
Li *et al.* 2006 24°C [37]	337	12.2	998	30.5
Li *et al.* 2010 [26]	752	14.5	3544	28.7
Rockman *et al.* 2010 [39]	1958	12.0	2792	28.8
Snoek & Sterken *et al.* 2017 control [28]	961	17.1	1481	36.1
Snoek & Sterken *et al.* 2017 heat-shock [28]	976	20.0	2776	36.9
Snoek & Sterken *et al.* 2017 recovery [28]	992	16.1	1519	33.4
Sterken *et al.* 2017 [30]	719	10.2	1116	26.5
Vinuela & Snoek *et al.* 2010 juvenile [31]	303	11.9	2206	33.4
Vinuela & Snoek *et al.* 2010 old [31]	220	15.0	1790	34.9
Vinuela & Snoek *et al.* 2010 reproductive [31]	348	13.2	2010	32.7

Discussion

In this experiment we used a population of multi-parental RILs (mpRILs) and RNA-seq to find 6,784 expression quantitative trait loci (eQTLs), of which 929 were *cis*-eQTLs and 5,855 were *trans*-eQTLs. A large proportion (63%) of the *trans*-eQTLs were found in six *trans*-bands. The total number of eQTLs found in this mpRIL study (6,784) is at the high end of what was previously found in other experiments (mean: 2,560; 653 – 6,518) [28, 30, 31, 37, 39]. This number is hard to compare as the number of identified eQTLs depend on many factors, such as population size, number of recombinations, statistical model, and RNA measurement technology used, which are nearly all different between this and the other eQTL studies in *C. elegans* [28, 30, 31, 37, 39]. Nevertheless, it seems that a combination of RNA-seq and multiple genetic backgrounds increased the number of detected eQTLs. A very clear increase was found for *trans*-eQTLs

(5,855) compared to the numbers found in previous studies, even at a much lower significance threshold. For example, the study of Rockman *et al*. 2010 used a comparable number of recombinant inbred advanced intercross lines (RIAILs) as the number of mpRILs in this study (~200), yet found fewer *trans*-eQTLs, however the different conditions and technologies used prevent any definitive conclusions. With respect to *trans*-eQTLs we do know that they depend on environmental conditions or a response to changing conditions. It could be that with a background of four parental genotypes the mpRILs perceive the ambient environment in a broader range than the RIAILs with a background of two parental genotypes used by Rockman *et al*. 2010, and the RILs in the other studies. For example, the mpRILs could have inherited parts of four different sets of environmental preferences as opposed to two in the RIAILs and RILs, potentially extending the accompanying gene expression patterns and eQTLs. Yet, the most likely reason for the increased number of *trans*-eQTLs is the use of RNA-seq in this study compared to micro-arrays in the other studies. Another reason for finding more *trans*-eQTLs could be due to the generally genome-wide equal allelic distributions in this population [45]. Namely, a similar *trans*-band as the chromosome I *trans*-band at 1 Mb (TB1) related to development has been spotted before, but has been spurious due to the *peel-1 zeel-1* incompatibility near that location [16, 28, 39]. Another advantage of using RNA-seq is that the genotype and gene-expression levels can be obtained from the same sample, preventing mis-labelling errors and the need for "reGenotyping" [64]. In summary, as has been shown for yeast [65], the combination of generally smaller effect of *trans*-eQTLs and higher dynamic range of RNA-seq would at least increase the possibility to pick-up *trans*-eQTLs in *C. elegans* and in general.

We previously found QTLs for several different phenotypes, such as population growth on different bacteria, sensitivity to heat-shock and oxidative stress [45]. Four *trans*-bands were found to co-locate with the previously found phenotypic QTLs (**Table 2**). Population growth on *Erwinia* and on *B. thuringiensis* DSM was found to co-locate with TB1, which was enriched for GO terms related to muscle, epidermis, and moulting. This could indicate a difference in these structures that can affect the interaction with different types of bacteria or could indicate that there is a difference in developmental speed through which differences in the expression, and subsequent eQTLs, of moulting related genes are picked up. A QTL for

heat-shock sensitivity was inferred to co-locate with TB4, however no indication for a link with this phenotype was found in the annotation of the genes with an eQTL at this position. The same was observed for TB5 and the overlap with population growth on *B. thuringiensis*, where GO enrichment also did not provide any leads to a potential mechanistic link. The overlap between the QTL for sensitivity to oxidative stress and TB6 however did show some clues from GO enrichment as genes involved in the peroxisome as well as DNA replication and cuticle formation could be involved in dealing with oxidative stress.

We expect to have only found a fraction of the eQTLs, as we only used a simple additive mapping model, a conservative score of one eQTL per gene, and standard lab conditions with only one time point for RNA isolation. Both the number of eQTLs and genes with one or more eQTLs are expected to increase when more complex models are applied to this data and/or different experimental conditions and time points are considered. Moreover, we use a SNP-based method for eQTL mapping, which has a binary option for each marker and therefore does not consider the genetic origin (parent) of the SNP. Using the genetic origin of the SNPs could reveal the more complex genetic interactions that could underly the differences in transcript levels between the mpRILs. These complex genetic interactions are suggested to be present in this mpRIL population, by the heritability and transgression found. A model in which each marker has the four parental options might indicate loci with more than two alleles affecting gene expression. Furthermore, some (relatively small) genetic loci might have been missed all together as our investigations are based on the N2 reference genome and wild-isolates can have vastly divergent regions of which sequences reads fail to align to the N2 reference genome with conventional methods [49].

This study provides a more detailed insight into the genetic architecture of heritable gene expression variation in a multiparent recombinant inbred population. The use of RNA-seq data in combination with more than two alleles allows for a more precise detection of QTLs and incorporates a wider band of standing genetic variation, resulting in a substantial increase in eQTLs especially *trans*-eQTLs. Comparison to bi-allelic studies supports the position of eQTLs and may be used to detect a more detailed pattern of associated loci. We expect this study, data, and results to provide new insights into *C. elegans* genetics and eQTLs in

264 general as well as to be a starting point to further test and develop advanced statistical models for detection

265 of eQTLs and systems genetics studying the genotype-phenotype relationship.

Methods

Nematode strains and culturing, RNA-sequencing, Construction of the genetic map

The *C. elegans* strains and culturing condition, RNA-sequencing and construction of the genetic map can all be found in Snoek *et al*. 2019. RILs, Genetic map and eQTL profiles can found on WormQTL2 [66] (http://www.bioinformatics.nl/EleQTL; Snoek & Sterken *et al.* 2020 [56])

SNP calling and gene expression levels

The paired end reads were mapped against the N2 reference genome (WS220) using Tophat [67], allowing for 4 read mismatches, and a read edit distance of 4. SNPs were called using samtools [66], mpileup with bcftools and vcfutils also described in Snoek *et al.* 2019 [45]. Expression levels were determined using the tuxedo pipeline [68]. Transcripts were assembled from the mapped reads using cufflinks [68]. Raw RNA-seq data can be found in the Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) with ID PRJNA495983. Normalized read-counts can be found on WormQTL2 (http://www.bioinformatics.nl/EleQTL; [56])

Heritability and Transgression

Heritability of gene expression levels was calculated using the heritability package in "R". A narrow-sense heritability was calculated using the function *marker_h2* [69]. The required kinship matrix was calculated using the *emma.kinship* function from the EMMA package [70]. To determine significance, we used a permutation approach where we shuffled the expression levels per transcript. After 100 permutations, the 95th highest value was taken as the 0.05 false-discovery rate [69, 71, 72]. Transgression was determined by counting the number of mpRILs with an expression level beyond the mean + 2SD of the most extreme parental lines. SD was calculated on the within variation of the parental samples. False discovery rate (FDR) was determined by permutations, randomly assigning the parental labels to gene-expression values.

Transgression was evaluated at an arbitrary 50 mpRILs (25% of all lines; FDR = 0.08) beyond the most extreme parental lines.

eQTL mapping and FDR

For eQTL mapping we first selected the genes with consistently found transcripts, meaning those expressed in at least 20 samples with a mean $\log_2$ expression (fpkm) > -5. eQTLs were mapped by a linear model using a single maker model explaining gene expression (as $\log_2$ ratio with the mean) by one SNP-marker at the time for the whole genome. False Discovery Rate (FDR) was determined by one round of permutations where for each transcript the counts were randomly distributed over the RILs before eQTL mapping. The -$\log_{10}$(p) value when number of false positives divided by the number of true positives was < 0.01 (-$\log_{10}$(p) > 5.35). Genome wide eQTL significance profiles (-log10(p)) can be found on WormQTL2 (http://www.bioinformatics.nl/EleQTL; [56])

Enrichment analysis and figures

Enrichment of GO terms was done using the hyper-geometric test in "R" [73]. GO term genes associations were download from Wormbase (www.wormbase.org) version WS276. Only genes that passed the filtering step for eQTL mapping where used as background genes. For significant enrichment, a p-value < $1e^{-5}$ was used and a geneset size per GO term > 3. Most figures were made using the R package ggplot2 [74] except figure 1 which was made using the UpSetR library.

eQTL comparison between experiments/studies

To compare how many genes with an eQTL overlapped between the different studies [26, 28, 30, 31, 37, 39, 56] available in WormQTL2 [56], we downloaded the eQTL profiles and markers used per experiment and listed the genes with a *cis* or a *trans* eQTL. For eQTL determination, the most significant marker per gene was taken as the peak. A -$\log_{10}$(p) > 3.5 was use as threshold for calling the eQTL. An eQTL was determined *cis* when the peak position was within 1Mbp of the start position of the gene. These lists were

316 compared with the genes having an eQTL in this study. The percentage overlap was calculated against the
317 original study.
318
319 **Acknowledgements**
320 We acknowledge financial support from the Deutsche Forschungsgemeinschaft to HS (grant number SCHU
321 1415/11 and project A1 within the CRC 1182), and to PCR (Competence Centre for Genomic Analysis
322 (CCGA) No. 07495230). JK was funded by NIH grant 1R01AA 026658. Furthermore, financial support
323 from the NWO-ALW (project 855.01.151) to RJMV was given. MGS was supported by NWO domain
324 Applied and Engineering Sciences VENI grant (17282). The funders had no role in study design, data
325 collection and analysis, decision to publish, or preparation of the manuscript.
326
327 **Author contributions**
328 LBS, HS and JEK conceived the study, RJMV and JR performed the experiments, PCR coordinated and
329 supervised transcriptome sequencing, LBS, MGS and HN analysed the data, LBS wrote the paper with
330 contributions from all authors.

References

1. Noble LM, Chang AS, McNelis D, Kramer M, Yen M, Nicodemus JP, Riccardi DD, Ammerman P, Phillips M, Islam T *et al*: **Natural Variation in plep-1 Causes Male-Male Copulatory Behavior in C. elegans**. *Curr Biol* 2015, **25**(20):2730-2737.
2. Andersen EC, Bloom JS, Gerke JP, Kruglyak L: **A variant in the neuropeptide receptor npr-1 is a major determinant of Caenorhabditis elegans growth and physiology**. *PLoS Genet* 2014, **10**(2):e1004156.
3. Ghosh R, Andersen EC, Shapiro JA, Gerke JP, Kruglyak L: **Natural variation in a chloride channel subunit confers avermectin resistance in C. elegans**. *Science* 2012, **335**(6068):574-578.
4. Kammenga JE, Doroszuk A, Riksen JA, Hazendonk E, Spiridon L, Petrescu AJ, Tijsterman M, Plasterk RH, Bakker J: **A Caenorhabditis elegans wild type defies the temperature-size rule owing to a single nucleotide polymorphism in tra-3**. *PLoS Genet* 2007, **3**(3):e34.
5. Large EE, Xu W, Zhao Y, Brady SC, Long L, Butcher RA, Andersen EC, McGrath PT: **Selection on a Subunit of the NURF Chromatin Remodeler Modifies Life History Traits in a Domesticated Strain of Caenorhabditis elegans**. *PLoS genetics* 2016, **12**(7):e1006219.
6. Bendesky A, Pitts J, Rockman MV, Chen WC, Tan MW, Kruglyak L, Bargmann CI: **Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in Caenorhabditis elegans**. *PLoS Genet* 2012, **8**(12):e1003157.
7. Bendesky A, Tsunozaki M, Rockman MV, Kruglyak L, Bargmann CI: **Catecholamine receptor polymorphisms affect decision-making in C. elegans**. *Nature* 2011, **472**(7343):313-318.
8. O'Donnell MP, Chao PH, Kammenga JE, Sengupta P: **Rictor/TORC2 mediates gut-to-brain signaling in the regulation of phenotypic plasticity in C. elegans**. *PLoS Genet* 2018, **14**(2):e1007213.
9. Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga JE *et al*: **The Genetic Basis of Natural Variation in Caenorhabditis elegans Telomere Length**. *Genetics* 2016, **204**(1):371-383.
10. Schmid T, Snoek LB, Frohli E, van der Bent ML, Kammenga J, Hajnal A: **Systemic Regulation of RAS/MAPK Signaling by the Serotonin Metabolite 5-HIAA**. *PLoS Genet* 2015, **11**(5):e1005236.
11. Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L: **A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans**. *PLoS Biol* 2011, **9**(7):e1001115.
12. McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, Bargmann CI: **Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors**. *Neuron* 2009, **61**(5):692-699.
13. Reddy KC, Andersen EC, Kruglyak L, Kim DH: **A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans**. *Science* 2009, **323**(5912):382-384.
14. Rogers C, Persson A, Cheung B, de Bono M: **Behavioral motifs and neural pathways coordinating O2 responses and aggregation in C. elegans**. *Current biology : CB* 2006, **16**(7):649-659.
15. Gloria-Soria A, Azevedo RB: **npr-1 Regulates foraging and dispersal strategies in Caenorhabditis elegans**. *Current biology : CB* 2008, **18**(21):1694-1699.
16. Seidel HS, Rockman MV, Kruglyak L: **Widespread genetic incompatibility in C. elegans maintained by balancing selection**. *Science* 2008, **319**(5863):589-594.
17. Tijsterman M, Okihara KL, Thijssen K, Plasterk RH: **PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of C. elegans**. *Current biology : CB* 2002, **12**(17):1535-1540.

18. Palopoli MF, Rockman MV, TinMaung A, Ramsay C, Curwen S, Aduna A, Laurita J, Kruglyak L: **Molecular basis of the copulatory plug polymorphism in Caenorhabditis elegans**. *Nature* 2008, **454**(7207):1019-1022.
19. Reiner DJ, Ailion M, Thomas JH, Meyer BJ: **C. elegans anaplastic lymphoma kinase ortholog SCD-2 controls dauer formation by modulating TGF-beta signaling**. *Current biology : CB* 2008, **18**(15):1101-1109.
20. Zdraljevic S, Fox BW, Strand C, Panda O, Tenjo FJ, Brady SC, Crombie TA, Doench JG, Schroeder FC, Andersen EC: **Natural variation in C. elegans arsenic toxicity is explained by differences in branched chain amino acid metabolism**. *Elife* 2019, **8**.
21. Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC: **Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles**. *PLoS pathogens* 2018, **14**(10):e1007226.
22. Zdraljevic S, Strand C, Seidel HS, Cook DE, Doench JG, Andersen EC: **Natural variation in a single amino acid substitution underlies physiological responses to topoisomerase II poisons**. *PLoS genetics* 2017, **13**(7):e1006891.
23. Ben-David E, Burga A, Kruglyak L: **A maternal-effect selfish genetic element in Caenorhabditis elegans**. *Science* 2017, **356**(6342):1051-1055.
24. Greene JS, Brown M, Dobosiewicz M, Ishida IG, Macosko EZ, Zhang X, Butcher RA, Cline DJ, McGrath PT, Bargmann CI: **Balancing selection shapes density-dependent foraging behaviour**. *Nature* 2016, **539**(7628):254-258.
25. Brady SC, Zdraljevic S, Bisaga KW, Tanny RE, Cook DE, Lee D, Wang Y, Andersen EC: **A Novel Gene Underlies Bleomycin-Response Variation in Caenorhabditis elegans**. *Genetics* 2019.
26. Li Y, Breitling R, Snoek LB, van der Velde KJ, Swertz MA, Riksen J, Jansen RC, Kammenga JE: **Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans**. *Genetics* 2010, **186**(1):405-410.
27. Singh KD, Roschitzki B, Snoek LB, Grossmann J, Zheng X, Elvin M, Kamkina P, Schrimpf SP, Poulin GB, Kammenga JE *et al*: **Natural Genetic Variation Influences Protein Abundances in C. elegans Developmental Signalling Pathways**. *PLoS One* 2016, **11**(3):e0149418.
28. Snoek BL, Sterken MG, Bevers RPJ, Volkers RJM, Van't Hof A, Brenchley R, Riksen JAG, Cossins A, Kammenga JE: **Contribution of trans regulatory eQTL to cryptic genetic variation in C. elegans**. *BMC Genomics* 2017, **18**(1):500.
29. Sterken MG, Bevers RPJ, Volkers RJM, Riksen JAG, Kammenga JE, Snoek BL: **Dissecting the eQTL micro-architecture in Caenorhabditis elegans**. *BioRxiv* 2019.
30. Sterken MG, van Bemmelen van der Plaat L, Riksen JAG, Rodriguez M, Schmid T, Hajnal A, Kammenga JE, Snoek BL: **Ras/MAPK Modifier Loci Revealed by eQTL in Caenorhabditis elegans**. *G3 (Bethesda)* 2017, **7**(9):3185-3193.
31. Vinuela A, Snoek LB, Riksen JA, Kammenga JE: **Genome-wide gene expression regulation as a function of genotype and age in C. elegans**. *Genome Res* 2010, **20**(7):929-937.
32. Gao AW, Sterken MG, Uit de Bos J, van Creij J, Kamble R, Snoek BL, Kammenga JE, Houtkooper RH: **Natural genetic variation in C. elegans identified genomic loci controlling metabolite levels**. *Genome Res* 2018, **28**(9):1296-1308.
33. Jimenez-Gomez JM, Wallace AD, Maloof JN: **Network analysis identifies ELF3 as a QTL for the shade avoidance response in Arabidopsis**. *PLoS Genet* 2010, **6**(9):e1001100.
34. Hartanto M, Joosen RVL, Snoek BL, Willems LAJ, Sterken MG, de Ridder D, Hilhorst HWM, Ligterink W, Nijveen H: **Network analysis prioritizes DEWAX and ICE1 as the candidate genes for two major eQTL hotspots in seed germination**. *BioRxiv* 2020.
35. Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in Arabidopsis by**

using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci U S A* 2007, **104**(5):1708-1713.

36. Snoek LB, Terpstra IR, Dekter R, Van den Ackerveken G, Peeters AJ: **Genetical Genomics Reveals Large Scale Genotype-By-Environment Interactions in Arabidopsis thaliana**. *Front Genet* 2012, **3**:317.

37. Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC *et al*: **Mapping determinants of gene expression plasticity by genetical genomics in C. elegans**. *PLoS genetics* 2006, **2**(12):e222.

38. Cubillos FA, Stegle O, Grondin C, Canut M, Tisne S, Gy I, Loudet O: **Extensive cis-regulatory variation robust to environmental perturbation in Arabidopsis**. *Plant Cell* 2014, **26**(11):4298-4310.

39. Rockman MV, Skrovanek SS, Kruglyak L: **Selection at linked sites shapes heritable phenotypic variation in C. elegans**. *Science* 2010, **330**(6002):372-376.

40. West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA: **Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis**. *Genetics* 2007, **175**(3):1441-1450.

41. Ranjan A, Budke JM, Rowland SD, Chitwood DH, Kumar R, Carriedo L, Ichihashi Y, Zumstein K, Maloof JN, Sinha NR: **eQTL Regulating Transcript Levels Associated with Diverse Biological Processes in Tomato**. *Plant Physiol* 2016, **172**(1):328-340.

42. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ: **Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource**. *Genome research* 2012, **22**(8):1558-1566.

43. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R: **A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana**. *PLoS genetics* 2009, **5**(7):e1000551.

44. de Koning DJ, McIntyre LM: **Back to the Future: Multiparent Populations Provide the Key to Unlocking the Genetic Basis of Complex Traits**. *Genetics* 2017, **206**(2):527-529.

45. Snoek BL, Volkers RJM, Nijveen H, Petersen C, Dirksen P, Sterken MG, Nakad R, Riksen JAG, Rosenstiel P, Stastna JJ *et al*: **A multi-parent recombinant inbred line population of C. elegans allows identification of novel QTLs for complex life history traits**. *BMC Biol* 2019, **17**(1):24.

46. Noble LM, Chelo I, Guzella T, Afonso B, Riccardi DD, Ammerman P, Dayarian A, Carvalho S, Crist A, Pino-Querido A *et al*: **Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the Caenorhabditis elegans Multiparental Experimental Evolution (CeMEE) Panel**. *Genetics* 2017, **207**(4):1663-1685.

47. Rockman MV, Kruglyak L: **Recombinational landscape and population genomics of Caenorhabditis elegans**. *PLoS Genet* 2009, **5**(3):e1000419.

48. Sterken MG, Snoek LB, Kammenga JE, Andersen EC: **The laboratory domestication of Caenorhabditis elegans**. *Trends Genet* 2015, **31**(5):224-231.

49. Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkers RJ, Brenchley R, Van't Hof A, Bevers RP, Cossins AR, Yanai I *et al*: **Remarkably Divergent Regions Punctuate the Genome Assembly of the Caenorhabditis elegans Hawaiian Strain CB4856**. *Genetics* 2015, **200**(3):975-989.

50. Doroszuk A, Snoek LB, Fradin E, Riksen J, Kammenga J: **A genome-wide library of CB4856/N2 introgression lines of Caenorhabditis elegans**. *Nucleic Acids Res* 2009, **37**(16):e110.

51. Vinuela A, Snoek LB, Riksen JA, Kammenga JE: **Aging Uncouples Heritability and Expression-QTL in Caenorhabditis elegans**. *G3 (Bethesda)* 2012, **2**(5):597-605.

52. Jovic K, Sterken MG, Grilli J, Bevers RPJ, Rodriguez M, Riksen JAG, Allesina S, Kammenga JE, Snoek LB: **Temporal dynamics of gene expression in heat-stressed Caenorhabditis elegans**. *PLoS One* 2017, **12**(12):e0189445.

53. Kamkina P, Snoek LB, Grossmann J, Volkers RJ, Sterken MG, Daube M, Roschitzki B, Fortes C, Schlapbach R, Roth A *et al*: **Natural Genetic Variation Differentially Affects the Proteome and Transcriptome in Caenorhabditis elegans**. *Mol Cell Proteomics* 2016, **15**(5):1670-1680.
54. Nakad R, Snoek LB, Yang W, Ellendt S, Schneider F, Mohr TG, Rosingh L, Masche AC, Rosenstiel PC, Dierking K *et al*: **Contrasting invertebrate immune defense behaviors caused by a single gene, the Caenorhabditis elegans neuropeptide receptor gene npr-1**. *BMC Genomics* 2016, **17**:280.
55. Rodriguez M, Snoek LB, Riksen JA, Bevers RP, Kammenga JE: **Genetic variation for stress-response hormesis in C. elegans lifespan**. *Exp Gerontol* 2012, **47**(8):581-587.
56. Snoek BL, Sterken MG, Hartanto M, van Zuilichem AJ, Kammenga JE, de Ridder D, Nijveen H: **WormQTL2: an interactive platform for systems genetics in Caenorhabditis elegans**. *Database (Oxford)* 2020, **2020**.
57. Snoek LB, Joeri van der Velde K, Li Y, Jansen RC, Swertz MA, Kammenga JE: **Worm variation made accessible: Take your shopping cart to store, link, and investigate!** *Worm* 2014, **3**(1):e28357.
58. Snoek LB, Orbidans HE, Stastna JJ, Aartse A, Rodriguez M, Riksen JA, Kammenga JE, Harvey SC: **Widespread genomic incompatibilities in Caenorhabditis elegans**. *G3 (Bethesda)* 2014, **4**(10):1813-1823.
59. Snoek LB, Van der Velde KJ, Arends D, Li Y, Beyer A, Elvin M, Fisher J, Hajnal A, Hengartner MO, Poulin GB *et al*: **WormQTL--public archive and analysis web portal for natural variation data in Caenorhabditis spp**. *Nucleic Acids Res* 2013, **41**(Database issue):D738-743.
60. Stastna JJ, Snoek LB, Kammenga JE, Harvey SC: **Genotype-dependent lifespan effects in peptone deprived Caenorhabditis elegans**. *Sci Rep* 2015, **5**:16259.
61. Jovic K, Grilli J, Sterken MG, Snoek BL, Riksen JAG, Allesina S, Kammenga JE: **Transcriptome resilience predicts thermotolerance in Caenorhabditis elegans**. *BMC Biol* 2019, **17**(1):102.
62. Volkers RJ, Snoek LB, Hubar CJ, Coopman R, Chen W, Yang W, Sterken MG, Schulenburg H, Braeckman BP, Kammenga JE: **Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild Caenorhabditis elegans populations**. *BMC Biol* 2013, **11**:93.
63. Nijveen H, Ligterink W, Keurentjes JJ, Loudet O, Long J, Sterken MG, Prins P, Hilhorst HW, de Ridder D, Kammenga JE *et al*: **AraQTL - workbench and archive for systems genetics in Arabidopsis thaliana**. *Plant J* 2017, **89**(6):1225-1235.
64. Zych K, Snoek BL, Elvin M, Rodriguez M, Van der Velde KJ, Arends D, Westra HJ, Swertz MA, Poulin G, Kammenga JE *et al*: **reGenotyper: Detecting mislabeled samples in genetic data**. *PLoS One* 2017, **12**(2):e0171324.
65. Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L: **Genetics of trans-regulatory variation in gene expression**. *Elife* 2018, **7**.
66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
67. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq**. *Bioinformatics* 2009, **25**(9):1105-1111.
68. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**. *Nature protocols* 2012, **7**(3):562-578.
69. Kruijer W, Boer MP, Malosetti M, Flood PJ, Engel B, Kooke R, Keurentjes JJ, van Eeuwijk FA: **Marker-based estimation of heritability in immortal populations**. *Genetics* 2015, **199**(2):379-398.

70. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E: **Efficient control of population structure in model organism association mapping**. *Genetics* 2008, **178**(3):1709-1723.
71. Gilmouret al. Gilmour; A.R. RTBRC: **Average InformationREML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models**. *Biometrics* 1995, **51**(4):1440-1450.
72. Speed D, Hemani G, Johnson MR, Balding DJ: **Improved heritability estimation from genome-wide SNPs**. *American journal of human genetics* 2012, **91**(6):1011-1021.
73. R-Core-Team: **R: A Language and Environment for Statistical Computing**. *R Foundation for Statistical Computing* 2017, **Vienna, Austria**(https://www.R-project.org/).
74. Wickham H: **GGplot2: elegant graphics for data analysis.** 2009.