Dietary fat-dependent transcriptional architecture and copy number alterations associated with modifiers of mammary cancer metastasis

Ryan R Gordon[1,7], **Michele La Merrill**[5], **Kent W Hunter**[6], **Peter Sørensen**[8], **David W Threadgill**[2,4,7], and **Daniel Pomp**[1,2,3,4]

[1]Department of Nutrition, University of North Carolina Chapel Hill, North Carolina 27599
[2]Department of Genetics, University of North Carolina Chapel Hill, North Carolina 27599
[3]Department of Cell and Molecular Physiology, University of North Carolina Chapel Hill, North Carolina 27599 [4]Lineberger Comprehensive Cancer Center, University of North Carolina Chapel Hill, North Carolina 27599 [5]Department of Preventive Medicine, Mount Sinai School of Medicine, New York, NY 10029 [6]Laboratory of Cancer Biology & Genetics, NIH/NCI, Bethesda, Maryland [7]Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695 [8]Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University

Abstract

Breast cancer is a complex disease resulting from a combination of genetic and environmental factors. Among environmental factors, body composition and intake of specific dietary components like total fat are associated with increased incidence of breast cancer and metastasis. We previously showed that mice fed a high-fat diet have shorter mammary cancer latency, increased tumor growth and more pulmonary metastases than mice fed a standard diet. Subsequent genetic analysis identified several modifiers of metastatic mammary cancer along with widespread interactions between cancer modifiers and dietary fat. To elucidate diet-dependent genetic modifiers of mammary cancer and metastasis risk, global gene expression profiles and copy number alterations from mammary cancers were measured and expression quantitative trait loci (eQTL) identified. Functional candidate genes that colocalized with previously detected metastasis modifiers were identified. Additional analyses, such as eQTL by dietary fat interaction analysis, causality and database evaluations, helped to further refine the candidate loci to produce an enriched list of genes potentially involved in the pathogenesis of metastatic mammary cancer.

Keywords

Breast Cancer; Causality; eQTL; High-fat diet; Tumors

Corresponding Author: Ryan Gordon, Department of Genetics, North Carolina State University, 100 Derieux Pl, Gardner Hall 3510, Raleigh NC 27695, Phone: 919-515-0250, Fax: 919-515-3355, rrgordon@ncsu.edu.

Introduction

Breast cancer is the second most common cancer type and the third-leading cause of cancer-related deaths of women living in the United States. It has been estimated that in 2009, nearly 200,000 women in the United States were diagnosed with some form of breast cancer and that over 40,000 died of this disease, typically from secondary metastasis. Breast cancer is a complex disease resulting from a combination of environmental and genetic factors. Although genetic factors altering predisposition to mammary cancer are starting to be elucidated with the identification of several familial risk factors [1–3], the overall genetic architecture remains largely unknown. This deficit is particularly evident at the intersection of genes and environmental factors like diet, which likely have the greatest impact on risk for metastatic mammary cancer.

Previously we generated an F_2 intercross [4] between the M16i polygenic mouse model of obesity [5] and the FVB/NJ-TgN(MMTV-PyMT)634Mul (PyMT) mouse model of metastatic mammary cancer [6]. We found that animals fed a high-fat diet not only had decreased mammary cancer latency, but also increased tumor growth and pulmonary metastases relative to mice fed a standard mouse chow. Several modifier loci for metastatic mammary cancer were detected along with widespread modifier by dietary fat interactions [7]. Subsequently, subsets of mammary tumors from mice on standard and high fat diets were evaluated for differences in gene expression. Surprisingly, only five genes were differentially expressed between mice on the two dietary fat treatments [8]. Nonetheless, dietary fat was found to alter pulmonary metastasis of mammary cancers through both cancer autonomous and non-autonomous changes in gene expression [8].

In this report, the M16i x PyMT F_2 population was analyzed to detect cancer expression quantitative trait loci (eQTL) and copy number alterations (CNA) regulating transcriptional networks associated with metastatic mammary cancer modifiers. With the aid of causality analysis and database evaluations, the candidate genes were further refined to produce an enriched list of genes potentially involved in the pathogenesis of metastatic mammary cancer.

Materials and Methods

Population development

An F_2 population (n = 615) was generated by crossing M16i, a polygenic obesity line [9], and FVB/NJ-TgN(MMTV-PyMT)634Mul (PyMT), a line transgenic for the Polyoma Middle T Oncoprotein that induces development of mammary cancers and subsequent pulmonary metastasis [6] (see [4] for full details of population development). F_2 mice were randomly assigned, within litter, gender, and genotype (PyMT or no PyMT), to receive one of two synthetic purified diets at four weeks of age, either a high-fat diet (HFD; Research Diets D12451) containing 45% of total calories from fat, 20% from protein, and 35% from carbohydrates, or a matched-control-fat diet (MCD; Research Diets D12450B) containing 10% of total calories from fat, 20% from protein, and 70% from carbohydrates. Mice were evaluated for body weight, body composition and metastatic mammary cancer phenotypes

as previously described [4, 7]. Axillary mammary tumors were harvested from 131 F_2 female PyMT carriers (HFD = 64 and MCD = 67) for microarray expression analysis.

Microarray analysis

For complete details refer to La Merrill et al [8]. Briefly, total RNA was isolated from axillary tumors using TRIzol (Invitrogen, Carlsbad, CA) and preprocessed for array hybridization using the Illumina® TotalPrep RNA Amplification kit (Ambion, Austin, TX). Expression profiles were generated using the Illumina Mouse 6 Sentrix arrays [10] (Illumina, San Diego, CA) and the resulting data were log transformed and normalized using the R-based evaluation program for Illumina expression data Lumi [11] prior to filtering at an Illumina detection score of 0.95 and above. Log transformed normalized data for all samples were run at a false discovery rate q- value < 0.05 in SAM software using two class-unpaired analyses with 1000 permutations [12, 13].

Correlation analysis

Pearson correlations were generated between all significantly expressed genes and the metastatic cancer phenotypes previously measured in the F_2 population [7]. The correlations were generated as three separate data sets: one for the whole population and the other two for animals on either the HFD or the MCD. All resulting p-values were adjusted with the FDR multiple comparisons test. Transcripts were then sorted by their strength of correlation within each metastasis phenotype.

Expression QTL (eQTL) analysis

eQTL were identified utilizing a customized R package, GaRT, for eQTL mapping in F2 populations [http://gbi.agrsci.dk/~pso/GaRT]. The eQTL models were fit with the following effects: additive, dominance, breeding-replicate and diet. The resulting eQTL were classified in one of two categories: "*cis*" if they mapped within 10 cM of the physical location of the gene they represent, and "*trans*" if they mapped elsewhere [14]. The significance threshold for all eQTL was set at a likelihood ratio statistic of 16.1 (LOD: 3.5). All eQTL classified as *cis*-acting were evaluated for the presence of eQTL x diet interactions by running an additional model which included all the aforementioned effects in addition to an eQTL x diet interaction term. The sum-of-squares error and the degrees of freedom for the peak position for each eQTL in both the interaction and non-interaction models were then calculated and used to compute an F statistic.

Pathway evaluation

All eQTL were evaluated using Ingenuity Pathway Analysis (IPA; Ingenuity Systems Inc., Redwood City, CA). Using the IPA core analysis function, reports containing information regarding the function, regulation, known mutations, tissue expression patterns, cellular location, and disease implication were generated for all significant genes. IPA was also used to provide statistical rankings of biological functions based upon the number of genes from the analysis set that participate in particular function.

Causality evaluation

Causality relationships between *cis*-acting eQTL and cancer phenotypes were evaluated using the R-based package eQTL-TF [15]. When estimating the relationships between *cis*-acting eQTL and cancer phenotypes three models were considered. The causal model is where genetic alterations (G) result in changes in the expression of a *cis*-acting eQTL (C), which in turn result in modification of the phenotype (P); a reactive model is where variation in G directly impacts P, resulting in altered gene expression of C; and a independent model that evaluates whether variation in G can result in changes in both P and C independently.

Oncomine evaluation

All *cis*-eQTL colocalizing with previously detected metastatic QTL [7] were entered into the Oncomine database [16] to determine if the genes they represent had been previously linked to metastatic cancer in humans. Specifically, all human breast cancer prognosis datasets in Oncomine, which best represents the possibility of having a metastatic cancer phenotype, were evaluated at a p-value threshold of 0.01 for each gene. If a gene was identified in any of the datasets in the prognosis category, it was further evaluated to determine if its expression is positively (no metastasis, alive, no disease) or negatively (metastasis, dead, relapse) associated with the clinical phenotype.

Evaluation of CNA

Copy number alterations were evaluated using the NimbleGen mouse 385K whole genome tiling array (NimbleGen Systems, Inc. Madison, WI) with a median probe spacing of 5.7 kb, in a subset of the F_2 females selected to represent the largest spectrum of cancer metastasis phenotype. This spectrum was achieved by selecting within each diet the 17 individuals with the highest number of observed pulmonary metastases and the 17 with the lowest number of observed pulmonary metastasis for a total of 68 individuals. RNA-free genomic DNA was extracted using the Puregene Tissue Core Kit (Qiagen, Minneapolis, MN) from the axillary mammary tumors (test) and tails (reference) of all 68 individuals. DNA samples were fragmented and fluorescently labeled with either cy-3 (tumor DNA) or cy-5 (tail DNA). The labeled tumor samples were pooled with their reference and co-hybridized to an array. Arrays were evaluated for relative fluorescence to determine the copy number profile across the genome for each individual using the NimbleScan software v2.3 (NimbleGen Systems, Inc. Madison, WI).

Resulting data were evaluated using SignalMap software (NimbleGen Systems, Inc. Madison, WI), to determine if patterns of CNA could be identified within the population. To provide common settings across the population, the $\log_2$ ratio scale was set at −2 and 2, while the track height was set at 120. Segments that deviated from 0.0 by ± 0.2–0.4 indicated a single amplification or deletion. A segment that deviated from 0 by ± 0.4–0.6 indicated a double amplification or deletion. If CNA were identified on chromosomes where cancer metastasis modifiers were previously detected (i.e. Chr 11), Proc Mixed in SAS (SAS Institute, Cary, NC) was used to determine whether significant associations were present between the copy number change and the phenotype used to map the modifier locus. The model evaluated included the fixed effects of diet and CNA, the interaction of diet x CNA, and the random effect of breeding replicate. If an association between the phenotype and

copy number change was detected, then *cis*-eQTL on the chromosome in question were also evaluated using the aforementioned Proc mixed model to determine if the gene expression levels were altered by the amplification or deletion.

Results

Correlation analysis

Correlations amongst gene expression and metastatic phenotypes were analyzed in a multi-step process, beginning with the entire population regardless of diet. While many genes were weakly correlated with the metastatic phenotypes, no genes surpassed the significance threshold of p < 0.05 after adjustment for multiple testing. Given the substantial influence of the HFD upon the metastatic phenotype [7], we hypothesized that the correlations between gene expression and metastatic traits were obscured by the dietary treatment. As such, the data were separated according to diet group and reevaluated. No significant correlations between gene expression and metastatic virulence were detected in the animals only fed the HFD. However, when the correlations were evaluated in animals fed the MCD, 14 genes were found to have significant correlations with the metastatic phenotype (Table 1), including several that have been previously reported to be associated with metastatic progression (i.e. *Col1a1* and *Col5a1* [17]). IPA identified direct and indirect connections between several of the genes within this set and *Trp53*.

eQTL evaluation

Totals of 220 *cis*-acting eQTL and 890 *trans*-acting eQTL were detected (Figure 1 A and B). The *cis*-acting eQTL were found across all chromosomes. Chr 11 and 17 harboring the largest number, while Chr 2 and 16 had the fewest. *Trans*-acting eQTL were likewise distributed across all chromosomes with Chr 3, 8 and 19 containing the largest numbers. When averaged, the likelihood ratio (LRT) score for all *cis*-acting eQTL was 45.9 with individual LRT scores ranging from 16.2 to 197.1. The average LRT score for all *trans*-acting eQTL was 23.9. If the *trans*-acting eQTL were evaluated as two separate groups, those mapping to the same chromosome (but not within 10 cM) as the actual gene they represent and those mapping to different chromosomes, the average LRT scores were 37.5 and 19.0, respectively.

Comparing the locations of *cis*-acting eQTL with phenotypic cancer QTL previously reported for this population detected 76 candidates for the previously mapped metastasis modifiers and 95 candidates for tumor growth and latency modifiers. Candidate genes were further refined based on their proximity to the cancer modifiers. If genes only within 15 cM of the modifier peaks are considered, the number of candidates is reduced from 76 to 44 and from 95 to 33 for metastatic and tumor growth/latency modifiers, respectively (Tables 2 and 3; Figures 2 and 3).

To investigate the functions of *cis* and *trans*-eQTL, the two separate gene lists were entered into IPA and evaluated with the core analysis function. Of the 220 *cis*-eQTL, 206 were found in the IPA database and of those, 117 were eligible for network and bio-function analysis (Table 4). The top bio-functions as indicated by IPA in the *cis*-acting dataset were

cellular movement and cancer, with genes in the cancer category occurring in many subgroups including apoptosis, tumor growth, migration, proliferation and invasion, along with categories representing a wide spectrum of specific cancer types. When the cellular movement group was expanded, the top functions were migration, invasion and localization. Additionally, genes were partitioned into networks based on the known or predicted interactions from the IPA database and ranked based on the number of candidate genes that appear within the dataset. Assessing the top functions of the highest ranked datasets revealed that many of these genes were involved in metabolism, cellular proliferation/death, and multiple processes of cancer.

Out of the 890 *trans*-acting eQTL, 808 were identified in the IPA database, and of those 515 were network and bio-function eligible (Table 4). The top bio-functions represented in this dataset included carbohydrate metabolism, small molecule biochemistry and cancer. When the cancer bio-function category was expanded, the top functions of genes within this category were revealed to be involved in tumorigenesis, apoptosis and cell death. The expansion of the carbohydrate metabolism set identified genes involved in the metabolism of fructose-6-phosphate, glycosaminoglycan and UDP-N-acetylglucosamine. Networks with the highest scores in the *trans*-eQTL data shared similar functions to the top networks in the *cis*-eQTL, such as the processes involved in metabolism, small molecule biochemistry and cancer. DNA replication and repair, gene expression, and endocrine disorders were represented as well.

Diet interactions

When this population was previously evaluated for cancer modifiers [7], we found that the majority of QTL exhibited interactions with dietary fat levels. For example, the metastasis modifier previously identified on Chr 8 resulted from the presence of a significant QTL in mice fed the HFD but not the MCD [7]. As such it was possible that the *cis*- acting eQTL would also show dietary interactions and be useful as a filter to further narrow candidates within the cancer modifier confidence intervals. We found that ~7% of the *cis*-eQTL had diet interactions (Table 5). Of the interactions detected, four resulted from differential allelic effects within the two diets (*Slc43a3* on Chr 2, *Basp1* on Chr 15, *E4f1* on Chr 17 and *Frmd8* on Chr 19). All other eQTL by diet interactions resulted from the detection of a significant effect for animals fed one diet but not the other; the majority of significant effects were found in animals fed MCD.

Causality evaluation

To further refine the candidate gene list, the nature of the relationships between *cis*-eQTL and observed metastasis modifier loci was evaluated. If eQTL have a causal relationship with a metastasis modifier then they may in turn represent the actual underlying genetic modifier of the phenotype of interest [18]. The results revealed that few significant relationships existed between *cis*-eQTL and metastasis modifiers. In a several situations significant but weak independent associations were detected between eQTL and cancer modifiers, such as those observed for the *Riken, Gtbp3*, and *Gcnt1* on Chr 1, 8 and 19, respectively (Table 6). One additional weak causal relationship was detected between *H2afv* and the metastasis modifier on Chr 11 (Table 6).

Oncomine evaluation

Evaluation of the Oncomine database revealed that 22 of the 44 candidates for the metastasis modifiers have been previously reported in human breast cancer prognosis datasets (Table 7). Of these, five were found to have an association between both increased and decreased gene expression and the clinical phenotype among the different datasets. Six of the candidates were found to have an increase in their expression associated with a better clinical outcome. In this category candidates whose orthologs had the greatest number of human studies supporting their association were *Dusp4* on Chr 8 and *Cxcl14* on Chr 13 with nine and five studies providing evidence, respectively. The final 11 candidates were found to have an increase in their expression that was associated with a poorer clinical prognosis. The candidate with the most compelling evidence in this group was *H2afv* on Chr 11 with seven studies supporting this link.

CNA evaluation

Although copy number alterations were detected in the mammary tumors (Figure 4), only four segments were shared among the tumors tested. The first was a 0.5 Mb segment on Chr 8 at approximately 20 Mb. The second was a 9 Mb duplication on the distal end of Chr 11 that was detected in 58% of the tumors tested. The third was on Chr 13 where a 1 Mb segment starting at 65 Mb was detected; 11% of the tumors having a duplication and 17% a deletion. The fourth region detected was a 0.25 Mb segment starting at 9.6 Mb on Chr 19 with 8% of tumors having a duplication and 20% having a deletion. Additionally, the partitioning of samples based on dietary treatment did not reveal any significant associations between the diet consumed and CNA detection.

When copy number changes were evaluated to determine if they were associated with metastasis development, no relationship was identified within the regions on Chr 13 and 19, but significant associations were detected within the regions on Chr 8 (Figure 5 A) and 11 (Figure 5 B). Individuals with a deletion of the Chr 8 region had significantly fewer pulmonary metastases compared to individuals with no CNA ($p < .05$), while individuals with a duplication of the region had almost a threefold increase in the number of metastasis detected when compared to individuals with no CNA ($p < .05$). Individuals with a duplication on Chr 11 had a significant reduction in metastasis development ($p < .05$). This evaluation was expanded to investigate whether the expression levels of the *cis*-eQTL on these two chromosomes were associated with the detected CNAs. One *cis*-eQTL, *Ascc2* located on Chr 11, was found to be associated with the CNA. A small but significant increase in expression of this gene was detected in animals that had a duplication on the distal end of Chr 11 ($p < .05$) (Figure 5 C).

Discussion

Estimates indicate that 60–70% of patients presenting with breast cancer have progressed to metastatic disease by the time of their diagnosis [19]. Consequently, the elucidation of the genetic and environmental factors influencing cancer progression to metastasis is essential for decreasing cancer mortality. Utilizing an F_2 mouse population generated by crossing mice genetically predisposed to obesity and metastatic mammary cancer [4], several primary

tumor phenotypes were characterized such as tumor latency and growth and development of pulmonary metastasis [4]. In addition, we found that the consumption of a high fat diet not only contributes to increased tumor growth, but increased metastasis when compared to mice fed a normal level of dietary fat. These cancer phenotypes along with the genotypic data obtained from this F_2 population facilitated the detection of multiple metastatic mammary cancer modifier loci, several of which exhibited modifier by diet interactions [4]. Unexpectedly, subsequent gene expression analyses within F_2 tumors revealed that few genes were differentially expressed between the two dietary treatments [8]. However, when the individuals were segregated based on the level of metastasis, many more differentially expressed genes were identified [8]. Given the wide range of diet-dependent metastatic phenotypes and the clinical importance of the metastatic process, the current investigation into the genetic and transcriptional architecture of metastatic mammary cancer was undertaken.

A previous eQTL study limited to a select subset of genes that had been previously predicted to be important for metastasis was successful in identifying a pathway altered in metastatic mammary cancer [20]. Our whole genome eQTL analysis supported detection of *cis* and *trans* regulated eQTL present on all chromosomes. Several of the *cis*-acting eQTL mapped in close proximity to previously detected metastasis modifiers. *cis*-eQTL colocalizing with metastatic mammary cancer modifiers represent logical candidate genes [21, 22], such as reported for *Rrp1b*, a *cis*-eQTL that colocalized with a previously detected metastasis modifier, and was later predicted to be clinically important in metastatic disease [17]. Filtering the *cis*-eQTL based on their proximity to previously detected phenotypic QTL enabled the detection of 44 candidates colocalizing with metastatic modifier loci.

Given that modifier by diet interactions were previously identified in this F_2 population [7] we anticipated that these interactions would be reflected in eQTL as well. However, only a small number of interactions were detected in the transcriptome mapping population. Additionally, the eQTL for which we detected diet interactions colocalizing with a cancer phenotype modifier, such as *Gdi3* on Chr 13, and *Frmd8* and *Ms4a6c* on Chr 19, resulted from significant eQTL effects in mice fed the MCD. The aforementioned results were not consistent with the interactions detected for the metastasis modifiers colocalizing with these eQTL, which resulted from significant eQTL effects in mice fed the HFD. Whereas the limited number of interactions was initially surprising given our previous findings, this result seems consistent with the lack of differentially expressed genes between the two dietary treatments [8].

The limited number of interactions detected may indicate that indirect effects of diet upon the cancers may have been responsible for the effects on cancer growth and metastasis previously observed. These effects may be mediated through transcriptional changes in another tissue such as the liver, known to respond metabolically to lipid consumption [23]. Links between hepatic expression of genes involved in the regulation of estrogenic compounds and breast cancer pathogenesis have been previously reported [24]. This appears to be supported by our findings that 36 of the 211 genes differentially expressed in the liver between the two dietary treatments were involved in cancer processes [8].

Using causality testing, we detected one *cis*-eQTL, *H2afv* on Chr 11 coding for a protein that is a member of the histone H2a family, that was causally associated with a metastasis modifier. Histone modifications have been detected in human cancers [25] and can be used to provide insight into clinical outcomes [26]. Currently, very little information regarding the function of *H2afv* exists, but other members of the histone family H2A (*H2AX*) [27] and proteins they encode (Buforin IIb) [28] have been shown to be potent inhibitors of cancer. While a casual relationship between *H2afv* and a metastasis modifier on Chr 11 was detected, the overall limited number of observed associations (casual, reactive and independent) may be highlighting the complexity of mammary cancer metastasis.

Evaluation of the Oncomine database provided additional support for several of the candidates being involved in the development of metastatic mammary cancer. Analysis of this database yielded results that appear to support the link between *H2afv* and metastatic mammary cancer. Not only did we implicate *H2afv* as a candidate for the metastasis modifier on Chr 11 through causality analysis, the Oncomine evaluation supported a link between increased expression of this gene and a poorer clinical prognosis in humans. Another candidate, *Dusp4*, was implicated by Oncomine as a modifier of metastatic mammary cancer in humans as well. *Dusp4*, a member of the dual-specificity phosphatase family, has been previously implicated as a potential tumor suppressor in a variety of cancer types [29, 30] and may be potentially a strong candidate for the metastasis modifier identified on Chr 8.

One gene not identified in the Oncomine evaluation, which has been previously associated with metastatic mammary cancer in humans, is butyrophilin (*Btn1a1*) [31]. *Btn1a1* was previously reported as being expressed significantly lower in primary cancers associated with metastatic versus those without metastasis [8]. Subsequent eQTL analysis provided evidence that *Btn1a1* is a potential modifier of the metastasis modifier on Chr 13 [8].

We also analyzed mammary cancers for CNA to evaluate their impact on both metastasis development and the detection of *cis*-eQTL. All but one of the *cis*-eQTL acted independently of detected CNA. The expression of *Ascc2*, a gene physically located outside the genomic boundaries of a CNA on Chr 11, was associated with the adjacent CNA. It is possible that the CNA interval detected on Chr 11 contains an enhancer element that impacts the expression *Ascc2* leading to the effect of the metastasis modifier on Chr 11. Ultimately, this highlights the importance of accounting for CNA in transcriptional studies, epically given the recent evidence suggesting that differences in CNA can alter the ability to detect eQTL [32].

In conclusion, this work is the first global eQTL analysis for metastatic mammary cancer, and demonstrates the utility of using transcriptome mapping to identify candidates for previously detected cancer modifiers. The use of causality testing provided insight into the relationship between the *cis*-eQTL detected for *H2afv* and the metastasis modifier on Chr 11. The potential roles of *H2afv* and *Dusp4* in metastatic mammary cancer were further supported by our evaluation of the Oncomine database. Aditionally, utilizing CNA analysis we detected a somatic alteration on chromosome 11 that was associated with increased expression of *Ascc2*, a gene for which a *cis*-eQTL was detected as a potential candidate for a

metastasis modifier on Chr 11. These findings provide additional insight into the genetic control of metastatic mammary cancer and possible targets for therapeutic interventions.

Acknowledgments

Abbreviations

AMD	Average metastasis density
C	*Cis*-acting eQTL
CAN	Copy number alteration
CNC	Copy number change
Chr	Chromosome
eQTL	Expression quantitative trait loci
G	Genetic alteration
HFD	High fat diet
IPA	Ingenuity pathway analysis
LRT	Likeihood ratio statistic
MCD	Matched control diet
MET	Metastasis detected at sacrifice
P	Phenotype
PyMT	Polyoma middle t oncoprotein
QTL	Quantitative trait loci

References

1. Rohan TE, Li SQ, Hartwick R, et al. p53 Alterations and protein accumulation in benign breast tissue and breast cancer risk: a cohort study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2006; 15(7):1316–1323.

2. Song CG, Hu Z, Wu J, et al. The prevalence of BRCA1 and BRCA2 mutations in eastern Chinese women with breast cancer. Journal of cancer research and clinical oncology. 2006; 132(10):617–626. [PubMed: 16835750]

3. Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA : the journal of the American Medical Association. 2006; 295(12):1379–1388.

4. Gordon RR, Hunter KW, Sorensen P, et al. Genotype X diet interactions in mice predisposed to mammary cancer I. Body weight and fat. Mamm Genome. 2008; 19(3):163–178. [PubMed: 18286334]

5. Allan MF, Eisen EJ, Pomp D. The M16 mouse: an outbred animal model of early onset polygenic obesity and diabesity. Obes Res. 2004; 12(9):1397–1407. [PubMed: 15483204]

6. Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Molecular and cellular biology. 1992; 12(3):954–961. [PubMed: 1312220]

7. Gordon RR, Hunter KW, La Merrill M, et al. Genotype X diet interactions in mice predisposed to mammary cancer: II. Tumors and metastasis. Mamm Genome. 2008; 19(3):179–189. [PubMed: 18288525]

8. La Merrill M, Gordon RR, Hunter KW, et al. Dietary fat alters pulmonary metastasis of mammary cancers through cancer autonomous and non-autonomous changes in gene expression. Clin Exp Metastasis. 27(2):107–116. [PubMed: 20151316]

9. Allan MF, Eisen EJ, Pomp D. Genomic mapping of direct and correlated responses to long-term selection for rapid growth rate in mice. Genetics. 2005; 170(4):1863–1877. [PubMed: 15944354]

10. Kuhn K, Baker SC, Chudin E, et al. A novel, high-performance random array platform for quantitative gene expression profiling. Genome research. 2004; 14(11):2347–2356. [PubMed: 15520296]

11. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008; 24(13):1547–1548. [PubMed: 18467348]

12. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98(9):5116–5121. [PubMed: 11309499]

13. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc Ser B. 2002; 64:479–498.

14. Doss S, Schadt EE, Drake TA, et al. Cis-acting expression quantitative trait loci in mice. Genome research. 2005; 15(5):681–691. [PubMed: 15837804]

15. Sun W, Yu T, Li KC. Detection of eQTL modules mediated by activity levels of transcription factors. Bioinformatics. 2007; 23(17):2290–2297. [PubMed: 17599927]

16. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004; 6(1):1–6. [PubMed: 15068665]

17. Crawford NP, Qian X, Ziogas A, et al. Rrp1b, a new candidate susceptibility gene for breast cancer progression and metastasis. PLoS Genet. 2007; 3(11):e214. [PubMed: 18081427]

18. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. Nature genetics. 2005; 37(7):710–717. [PubMed: 15965475]

19. Eccles SA, Box G, Court W, et al. Preclinical models for the evaluation of targeted therapies of metastatic disease. Cell biophysics. 1994; 24–25:279–291.

20. Crawford NP, Walker RC, Lukes L, et al. The Diasporin Pathway: a tumor progression-related transcriptional network that predicts breast cancer survival. Clin Exp Metastasis. 2008; 25(4):357–369. [PubMed: 18301994]

21. Yamashita S, Wakazono K, Nomoto T, et al. Expression quantitative trait loci analysis of 13 genes in the rat prostate. Genetics. 2005; 171(3):1231–1238. [PubMed: 16079240]

22. Wang SS, Schadt EE, Wang H, et al. Identification of pathways for atherosclerosis in mice: integration of quantitative trait locus analysis and global gene expression data. Circ Res. 2007; 101(3):e11–30. [PubMed: 17641228]

23. Morgan K, Uyuni A, Nandgiri G, et al. Altered expression of transcription factors and genes regulating lipogenesis in liver and adipose tissue of mice with high fat diet-induced obesity and nonalcoholic fatty liver disease. Eur J Gastroenterol Hepatol. 2008; 20(9):843–854. [PubMed: 18794597]

24. Gong H, Guo P, Zhai Y, et al. Estrogen deprivation and inhibition of breast cancer growth in vivo through activation of the orphan nuclear receptor liver X receptor. Molecular endocrinology (Baltimore, Md. 2007; 21(8):1781–1790.

25. Fraga MF, Ballestar E, Villar-Garea A, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005; 37(4):391–400. [PubMed: 15765097]

26. Seligson DB, Horvath S, Shi T, et al. Global histone modification patterns predict risk of prostate cancer recurrence. Nature. 2005; 435(7046):1262–1266. [PubMed: 15988529]

27. Liu Y, Tseng M, Perdreau SA, et al. Histone H2AX is a mediator of gastrointestinal stromal tumor cell apoptosis following treatment with imatinib mesylate. Cancer Res. 2007; 67(6):2685–2692. [PubMed: 17363589]

28. Lee HS, Park CB, Kim JM, et al. Mechanism of anticancer activity of buforin IIb, a histone H2A–derived peptide. Cancer Lett. 2008; 271(1):47–55. [PubMed: 18617323]

29. Sieben NL, Oosting J, Flanagan AM, et al. Differential gene expression in ovarian tumors reveals Dusp 4 and Serpina 5 as key regulators for benign behavior of serous borderline tumors. J Clin Oncol. 2005; 23(29):7257–7264. [PubMed: 16087957]

30. Chitale D, Gong Y, Taylor BS, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene. 2009; 28(31):2773–2783. [PubMed: 19525976]

31. Woelfle U, Cloos J, Sauter G, et al. Molecular signature associated with bone marrow micrometastasis in human breast cancer. Cancer Res. 2003; 63(18):5679–5684. [PubMed: 14522883]

32. Williams, Rt; Lim, JE.; Harr, B., et al. A common and unstable copy number variant is associated with differences in Glo1 expression and anxiety-like behavior. PLoS ONE. 2009; 4(3):e4649. [PubMed: 19266052]

eQTL detected at LRT 16.1
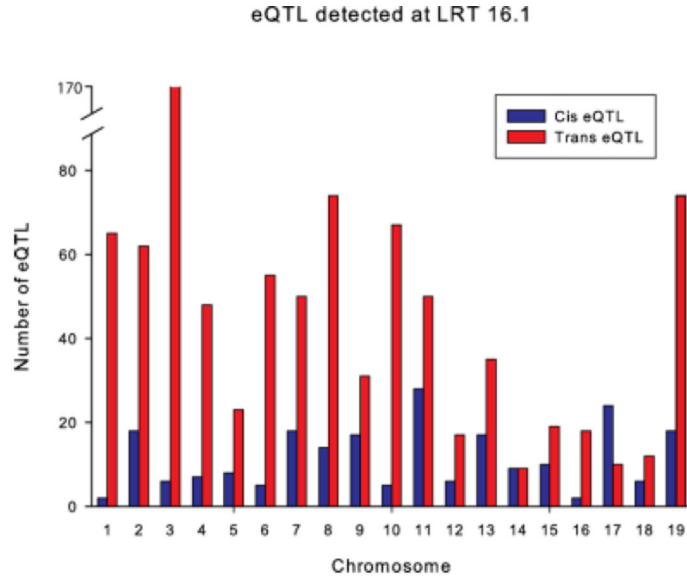
b.
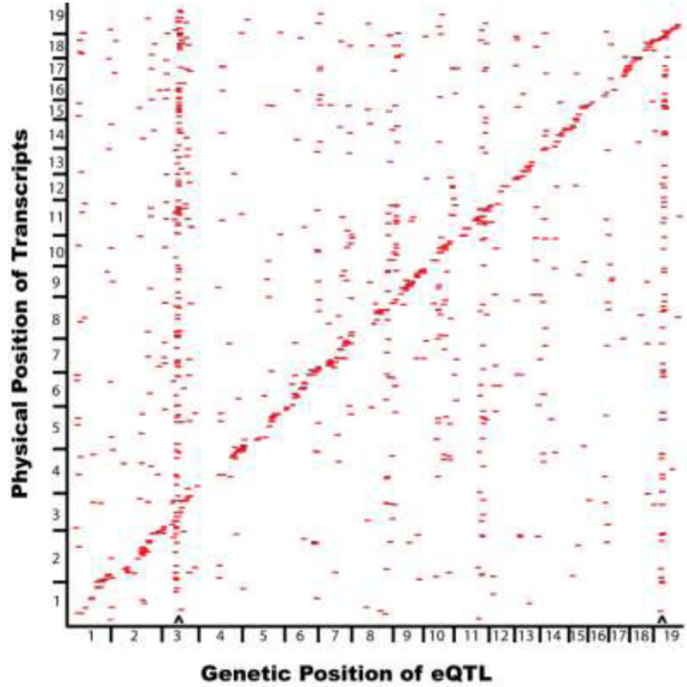
Figure 1.
The eQTL mapping results. A: The number of *cis and trans*-eQTL detected on each chromosome at the significance LRT threshold of 16.1. B: Mapping of expression QTL (eQTL). Physical gene positions on the microarray are plotted along the y-axis and the genetic locations of the QTL are plotted along the x-axis. eQTL along the cis diagonal map within 10 cM of the transcript that they represent. eQTL acting in trans map to a different chromosome than the transcript they represent. ^Master regulators are eQTL acting in cis that map to a region associated with many trans-acting eQTL.

a.

b.

c.
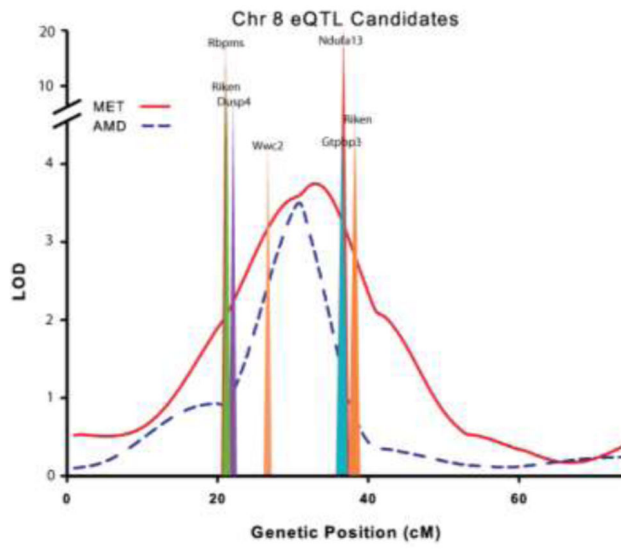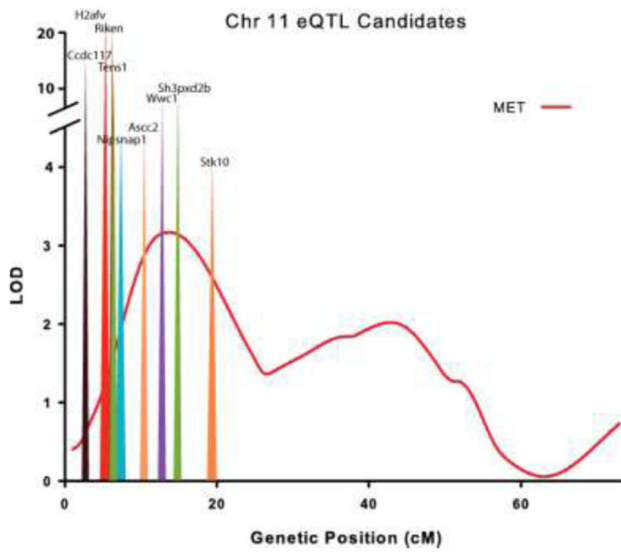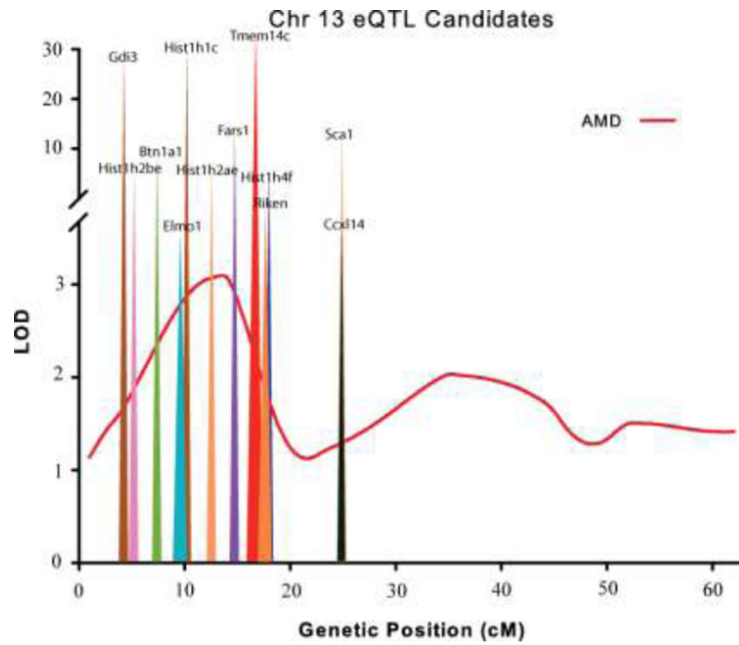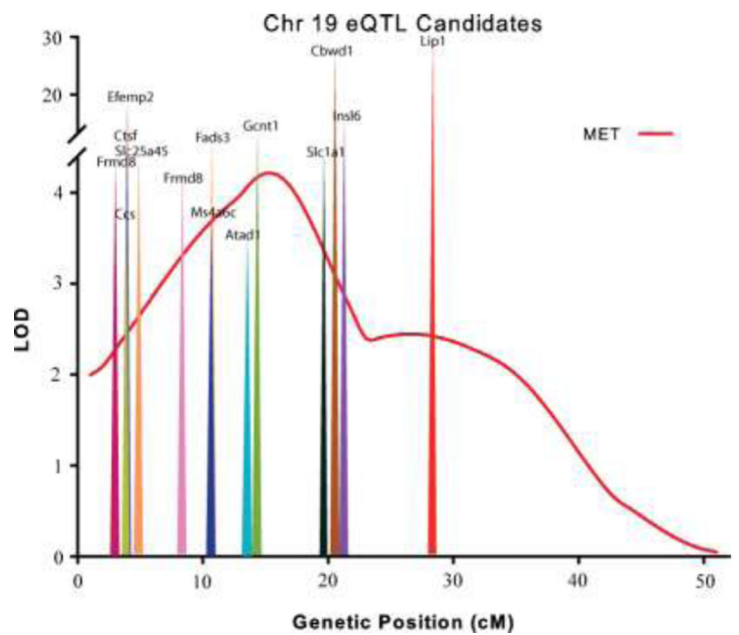
Figure 2.
Metastatic QTL with colocalizing candidates. A: *Cis*-eQTL colocalizing with the metastatic locus on Chr 8. B: *Cis*-eQTL colocalizing with the metastatic locus on Chr 11. C: *Cis*-eQTL colocalizing with the metastatic locus on Chr 13. D: *Cis*-eQTL colocalizing with the metastatic locus on Chr 19.
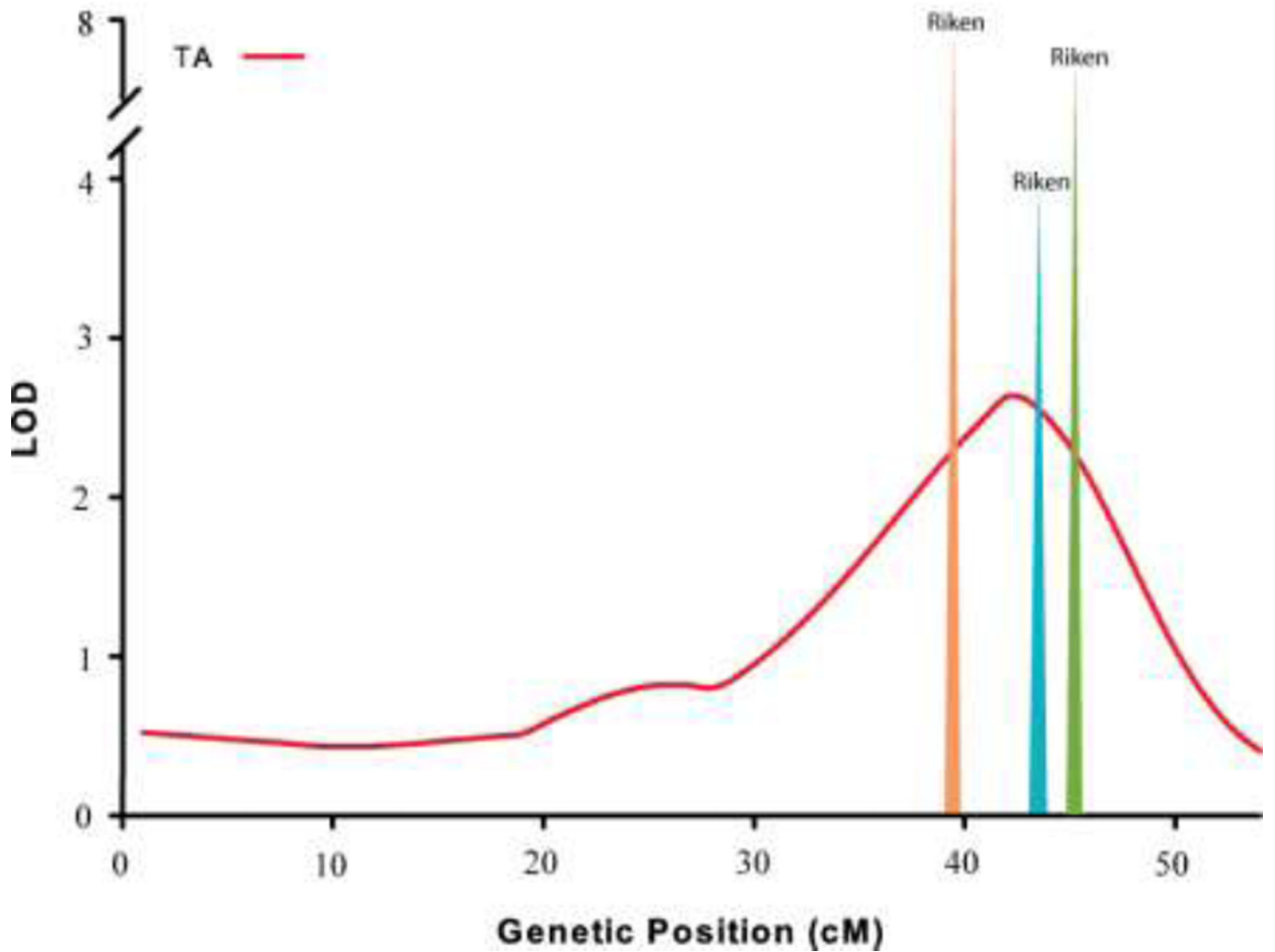
a.

Chr 5 eQTL Candidates

b.

Chr 7 eQTL Candidates

c.

Chr 9 eQTL Candidates

d.

Chr 13 eQTL Candidates

e.

Chr 14 eQTL Candidates

f.
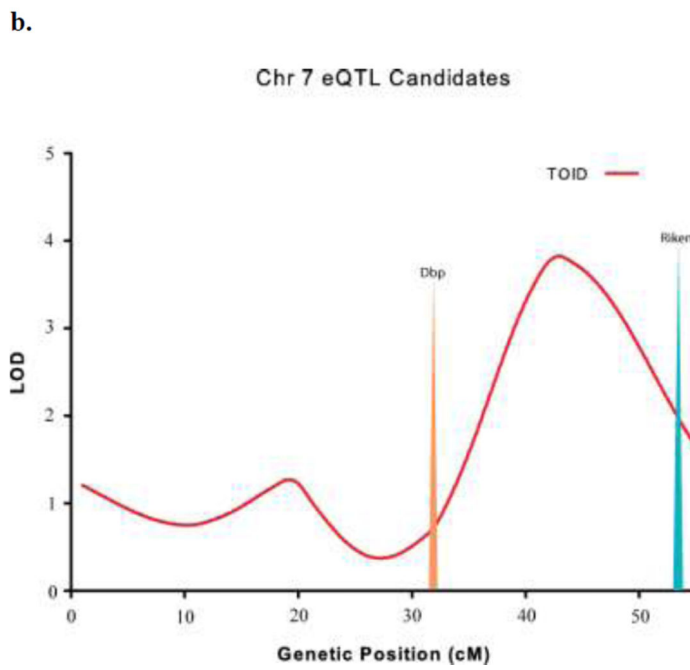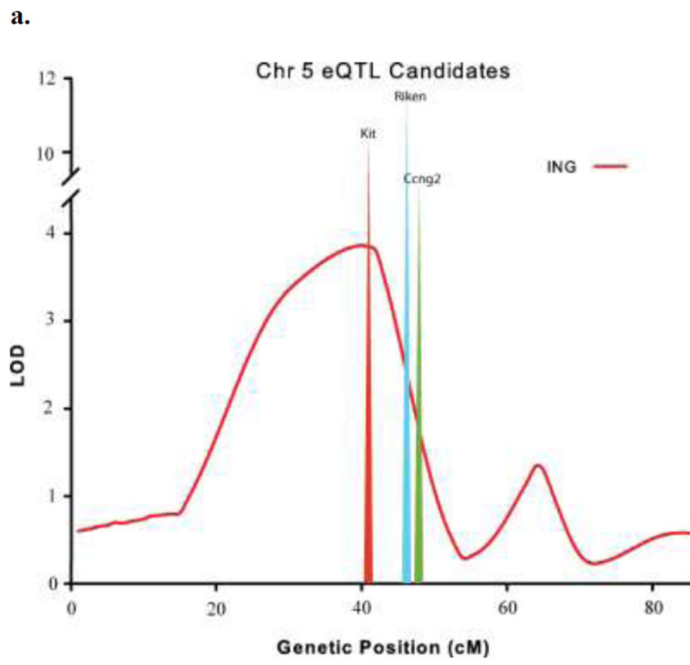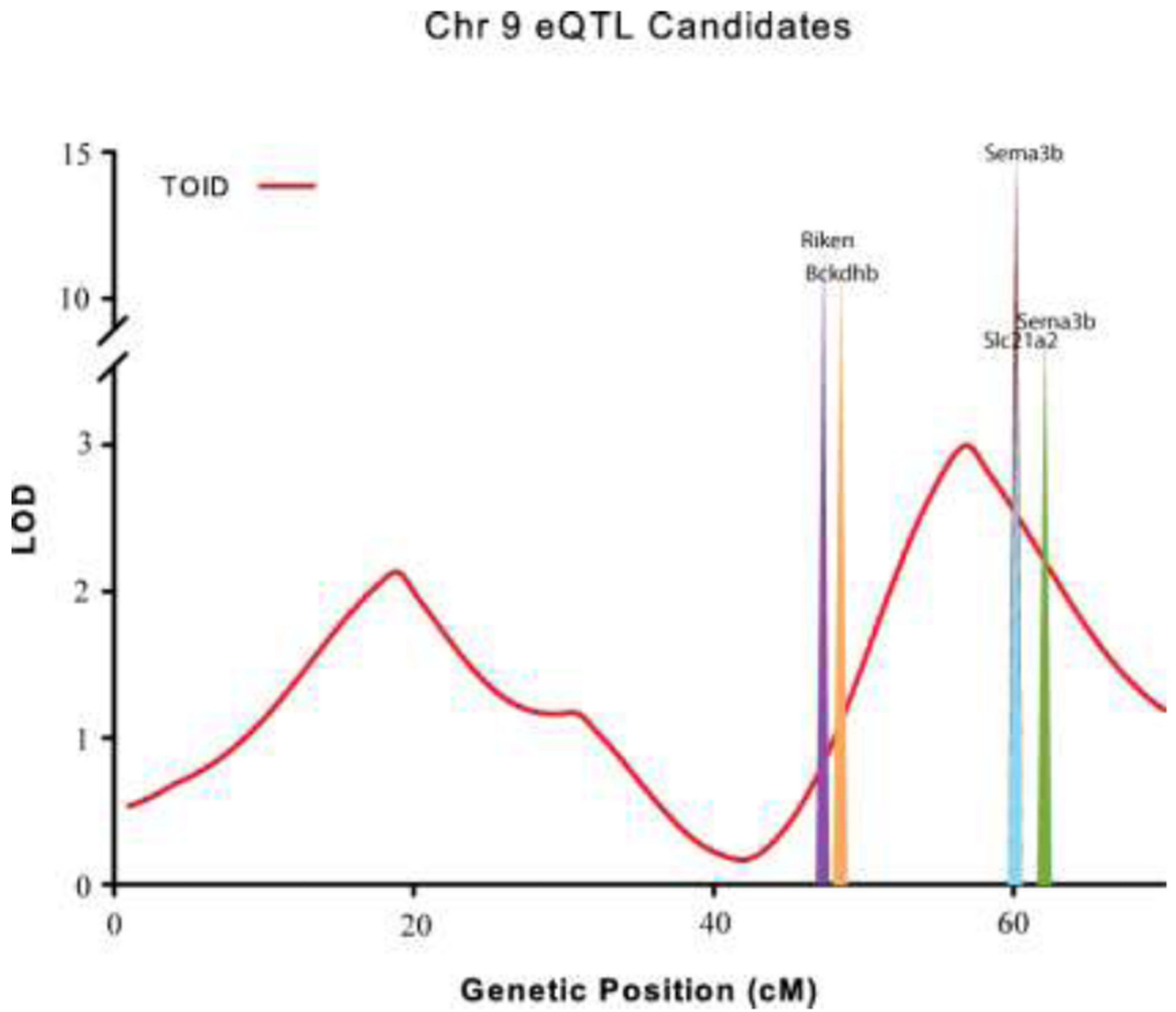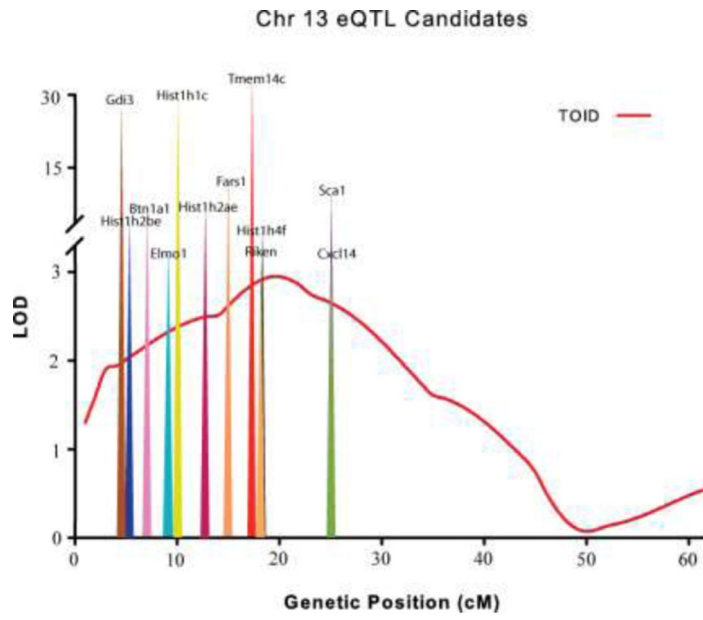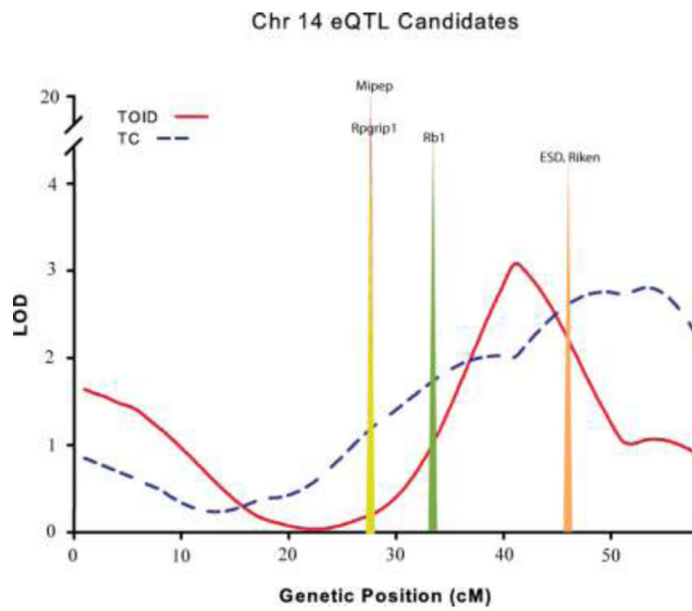
Figure 3.
Primary Tumor QTL with colocalizing candidates. A: *Cis*-eQTL colocalizing with the inguinal tumor growth locus on Chr 5. B: *Cis*-eQTL colocalizing with the tumor latency locus on Chr 7. C: *Cis*-eQTL colocalizing with the tumor latency locus on Chr 9. D: *Cis*-eQTL colocalizing with the tumor latency locus on Chr 13. E: *Cis*-eQTL colocalizing with the tumor latency loci and total number of tumors at sacrifice loci on Chr 14. B: *Cis*-eQTL colocalizing with the axillary tumor growth locus on Chr 17.
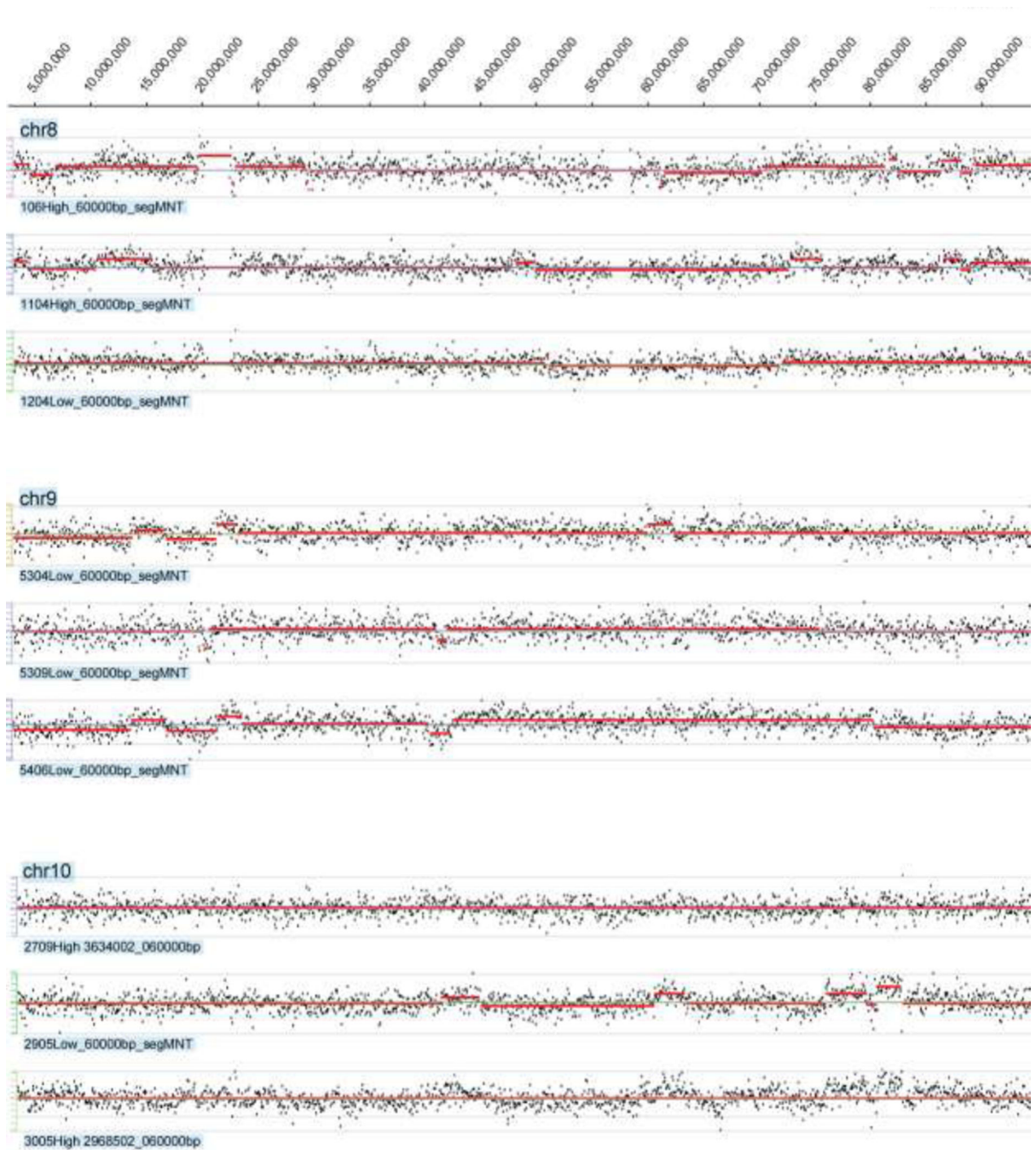
Figure 4.
Examples of Copy Number Alteration (CNA) found in the F_2 Population. A snapshot of the CNA found within the F_2 population. Examples from three separate chromosomes (8, 9 and 10) are shown.
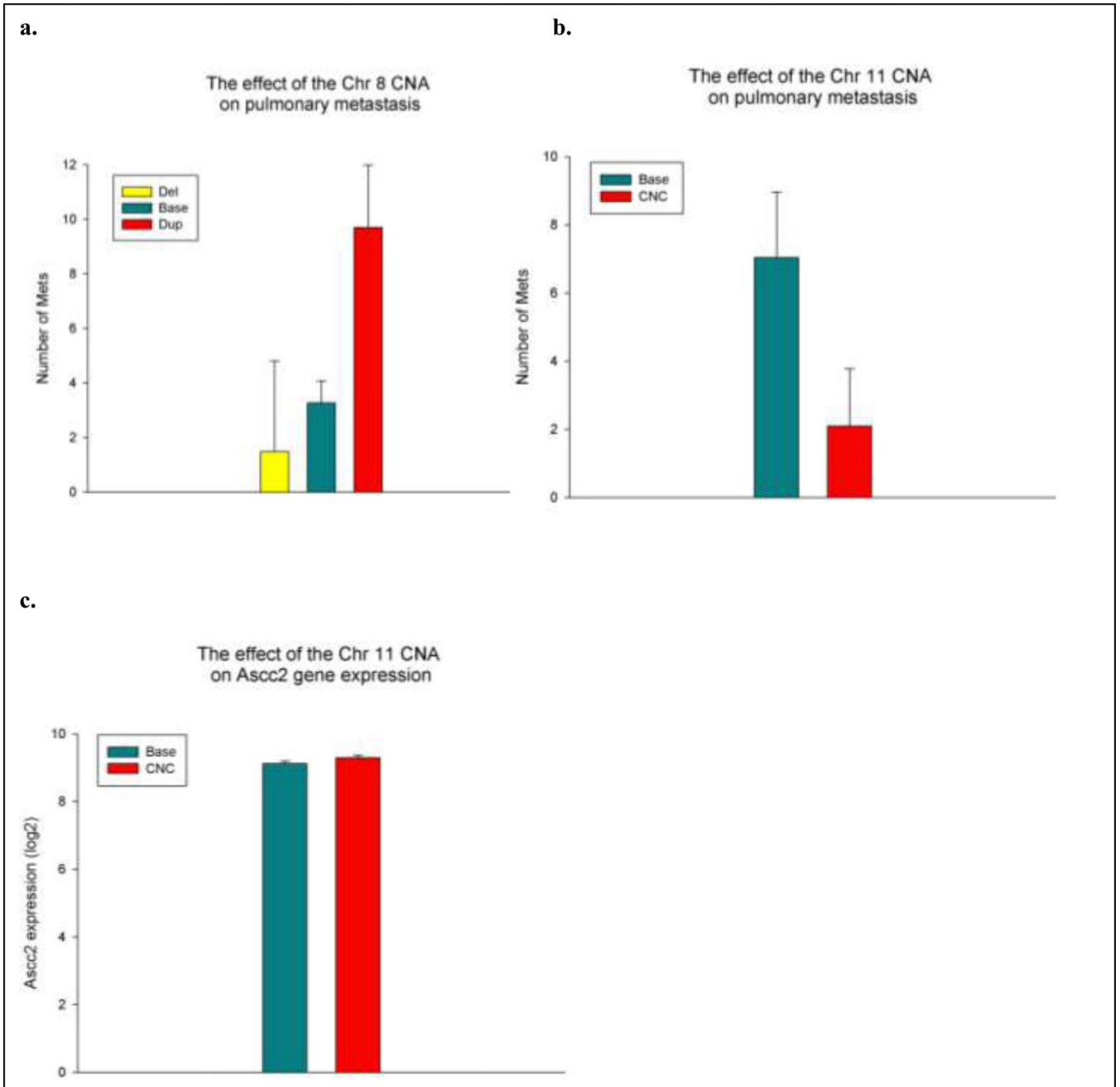
Figure 5.
Evaluation of Copy Number Alteration (CNA). A: The impact of the Chr 8 CNA on the development of pulmonary metastasis. A duplication of this region was associated with a significant increase metastatic formation. B: The impact of the Chr 11 CNA on pulmonary metastasis development. The presence of an additional copy in this region was associated with a significant reduction in metastatic development. C: The impact of the Chr 11 CNA upon the expression of *Ascc2*. The presence of an additional copy in this region was

associated with a significant increase in the expression of *Ascc2*. *CNC: Copy number change. **All comparisons significant at $p < .05$.

Table 1

Significant correlations between gene expression and metastatic phenotype after adjustment for multiple test comparisons in animals fed the MCD

Gene	Chr	Correlation	Raw P-value	Adjusted	Function
Ubc	5	−0.53	0.000005	0.041	Unknown
S100a3	3	0.50	0.000014	0.041	calcium ion binding, protein binding
4933428G20Rik	11	−0.50	0.000017	0.041	Unknown
Arhgap17	7	−0.49	0.000023	0.041	protein binding
Nrm	17	−0.49	0.000029	0.041	Unknown
Bscl2	19	−0.49	0.000031	0.041	Unknown
Vps53	11	−0.49	0.000031	0.041	Unknown
Psmd3	11	−0.48	0.000033	0.041	protein binding , enzyme regulation
Slc25a39	11	−0.48	0.000039	0.041	Binding
Sdhc	1	−0.48	0.000040	0.041	succinate dehydrogenase activity
Ctsk	3	−0.47	0.000051	0.042	protein binding , peptidase activity
Col1a1	11	−0.47	0.000056	0.042	ECM structure, protein binding
Col5a1	2	−0.47	0.000056	0.042	ECM structure, integrin binding
Sfpq	4	−0.47	0.000058	0.042	nucleotide binding, protein binding

Table 2

Candidate eQTL based on proximity to metastatic QTL

Accession	Symbol	Chr	LRT	LOD	eQTL peak (cM)[a]	eQTL to Gene (cM)[b]	eQTL to QTL (cM)[c]
NM_183019.1	9330140K16Rik	1	23.3	5.1	19.9	2.0	12.9
NM_133791.3	Wwc2	8	20.4	4.4	28.2	1.1	2.8
NM_023312.2	Ndufa13	8	91.6	19.9	36.2	7.0	5.2
NM_032544.2	Gtpbp3	8	24.9	5.4	36.2	8.0	5.2
NM_028993.2	9130404D08Rik	8	24.4	5.3	38.2	5.0	7.2
NM_176933.3	Dusp4	8	33.1	7.2	22.2	1.2	8.8
	9530006C21Rik	8	45.6	9.9	21.2	0.7	9.8
NM_019733	Rbpms	8	77.6	16.9	21.2	0.8	9.8
XM_109956	Wwc1	11	34.6	7.5	13.9	8.4	0.1
NM_177364	Sh3pxd2b	11	47.4	10.3	15.9	4.3	1.9
NM_029291.1	Ascc2	11	23.9	5.2	9.9	7.0	4.1
NM_009288.1	Stk10	11	18.4	4.0	19.9	0.4	5.9
NM_008698.1	Nipsnap1	11	22.4	4.9	7.9	4.9	6.1
XM_109868	Tens1	11	77.1	16.8	6.9	1.6	7.1
XM_488586	2210015D19Rik	11	93.6	20.3	6.9	3.4	7.1
XM_126043.3	H2afv	11	110.0	23.9	5.9	1.9	8.1
NM_134033.1	Ccdc117	11	67.6	14.7	3.9	0.5	10.1
NM_178187.2	Hist1h2ae	13	40.3	8.8	13	1.4	0
NM_024274.1	Fars1	13	61.4	13.3	15	7.5	2
NM_015786	Hist1h1c	13	134.9	29.3	10	4.5	3
NM_198093.2	Elmo1	13	16.7	3.6	9	3.6	4
NM_025387.1	Tmem14c	13	155.0	33.7	17	8.4	4
NM_026947.2	1810022C23Rik	13	18.6	4.0	18	3.5	5
NM_175655.1	**Hist1h4f	13	28.2	6.1	18	3.6	5
NM_013483.1	**Btn1a1	13	33.8	7.4	7	7.3	6
NM_178194.2	Hist1h2be	13	31.3	6.8	5	9.4	8

Accession	Symbol	Chr	LRT	LOD	eQTL peak (cM)[a]	eQTL to Gene (cM)[b]	eQTL to QTL (cM)[c]
NM_008112.2	Gdi3	13	129.4	28.1	4	1.8	9
NM_009124.2	Sca1	13	49.0	10.6	25	3.2	12
NM_019568	Cxcl14	13	16.5	3.6	25	9.3	12
NM_173442.1	Gcnt1	19	45.0	9.8	14.5	4.2	1
NM_026487.2	Atad1	19	16.2	3.5	13.5	6.5	2
NM_009199.1	Slc1a1	19	27.9	6.1	19.5	1.9	4
NM_028595	Ms4a6c	19	16.3	3.5	10.5	3.8	5
NM_021890	Fads3	19	39.0	8.5	10.5	4.7	5
NM_146097.1	Cbwd1	19	133.6	29.0	20.5	5.4	5
NM_013754.1	Insl6	19	55.7	12.1	21.5	3.6	6
NM_026169.3	Frmd8	19	20.3	4.4	8.5	5.0	7
NM_134154.1	Slc25a45	19	30.6	6.6	5.5	2.0	10
NM_021474.2	Efemp2	19	75.9	16.5	4.5	1.2	11
NM_019861.1	Ctsf	19	45.8	10.0	4.5	1.6	11
NM_016892.2	Ccs	19	16.8	3.7	4.5	1.6	11
AK032179	Saps3	19	23.5	5.1	3.5	1.5	12
NM_021460.1	Lip1	19	141.4	30.7	28.5	7.4	13

[a] eQTL peak: The estimated peak position of the eQTL in cM

[b] eQTL to gene: The distance in cM from the peak eQTL position to the position of the gene it represents.

[c] eQTL to QTL: The distance in cM from the peak eQTL position to the nearest metastatic QTL

** Genes previously reported as differentially expressed between tumors of varying metastatic tendencies.

Table 3

Candidate eQTL based on proximity to tumor growth and latency QTL

Accession	Symbol	Chr	LRT	LOD	eQTL peak (cM)[a]	eQTL to Gene (cM)[b]	eQTL to QTL (cM)[c]
NM_021099	Kit	5	48.7	10.6	43.6	3.0	4.1
NM_175270.2	5730467H21Rik	5	54.2	11.8	48.6	9.1	9.1
NM_007635.2	Ccng2	5	40.0	8.7	49.6	8.3	10.1
NM_016974.1	Dbp	7	16.7	3.6	32	7.1	10.0
AK030267	4933439J20Rik	7	18.4	4.0	54	6.1	12.0
XM_135023.2	2610018I03Rik	9	55.0	12.0	45.2	9.0	10.8
AK004616	Slc21a2	9	32.6	7.1	60.2	4.2	4.2
NM_009275.2	Srprb	9	33.5	7.3	62.2	2.3	6.2
NM_009153.1	Sema3b	9	69.3	15.1	60.2	7.1	4.2
NM_199195.1	Bckdhb	9	54.0	11.7	46.2	6.6	9.8
NM_008112.2	Gdi3	13	129.4	28.1	4	1.8	15.0
NM_178194.2	Hist1h2be	13	31.3	6.8	5	9.4	14.0
NM_013483.1	Btn1a1	13	33.8	7.4	7	7.3	12.0
NM_198093.2	Elmo1	13	16.7	3.6	9	3.6	10.0
NM_015786	Hist1h1c	13	134.9	29.3	10	4.5	9.0
NM_178187.2	Hist1h2ae	13	40.3	8.8	13	1.4	6.0
AK021333	Btn1a1	13	21.1	4.6	14	0.3	5.0
NM_024274.1	Fars1	13	61.4	13.3	15	7.5	4.0
NM_025387.1	Tmem14c	13	61.5	13.4	17	8.4	2.0
NM_025387.1	Tmem14c	13	155.0	33.7	17	8.4	2.0
NM_175655.1	Hist1h4f	13	28.2	6.1	18	3.6	1.0
NM_026947.2	1810022C23Rik	13	18.6	4.0	18	3.5	1.0
NM_009124.2	Sca1	13	49.0	10.6	25	3.2	6.0
NM_019568	Cxcl14	13	16.5	3.6	25	9.3	6.0
NM_023879.1	Rpgrip1	14	38.7	8.4	27.9	1.6	12.1
NM_027436.1	Mipep	14	91.4	19.9	27.9	6.7	12.1

Accession	Symbol	Chr	LRT	LOD	eQTL peak (cM)[a]	eQTL to Gene (cM)[b]	eQTL to QTL (cM)[c]
NM_009029.1	*Rb1*	14	24.6	5.4	33.9	8.3	6.1
NM_016903.2	*Esd*	14	20.7	4.5	45.9	2.7	5.9
320628	*A130038J17Rik*	14	20.9	4.5	45.9	0.1	5.9
NM_008549.1	*Man2a1*	17	31.7	6.9	29.2	9.8	12.8
78592	*A330106M24Rik*	17	35.8	7.8	39.2	5.9	2.8
NM_152817.2	*2610511O17Rik*	17	18.0	3.9	44.2	1.2	2.2
NM_144802.2	*2810036L13Rik*	17	29.3	6.4	46.2	2.5	4.2

[a] eQTL peak: The estimated peak position of the eQTL in cM

[b] eQTL to gene: The distance in cM from the peak eQTL position to the position of the gene it represents.

[c] eQTL to QTL: The distance in cM from the peak eQTL position to the nearest tumor growth/latency QTL

Table 4

IPA Evaluation for *cis/trans*-eQTL

	Mapped Id's[a]	Eligible eQTL[b]	Bio functions[c]	Networks[d]
cis-acting eQTL	206	117	Cancer (25)	Cell death, Lipid metabolism, Small molecule biochemistry (22)
			Cellular movement (10)	Lipid metabolism, Small molecule biochemistry, cell morphology (20)
			Gastrointestinal Disease (6)	Genetic disorder, Neurological disease, Ophthalmic disease (14)
			Cell cycle (9)	Amino acid metabolism, Cancer, Carbohydrate metabolism (14)
trans-acting eQTL	808	515	Carbohydrate metabolism (14)	Gene expression, DNA replication and repair, Endocrine system disorders (32)
			Small molecule biochemistry (54)	Drug metabolism, Cancer, Lipid metabolism (31)
			Gene expression (21)	Carbohydrate metabolism, Small molecule biochemistry, immune response (31)
			Cancer (174)	Gene expression, Cellular development, Nervous system development/function (29)

[a]The number of candidates that were found in the IPA database

[b]The number of candidates that had corresponding bio-function and network information in the IPA database

[c]The top bio-functions as indicated by IPA . The number in () indicates how many of the candidates from the eligible list are involved in the function

[d]The functions of the top networks as indicated by IPA. The number in () indicates how many of the candidates are in the particular network

Table 5

Significant cis-eQTL by diet interactions

Symbol	Chr	Interaction[a]	LRT	eQTL peak[b]
E030013I19Rik	2	MCD	18.48	13.8
Dnajc1	2	MCD	23.40	7.8
Slc43a3	2	BD	42.13	46.8
Accs	2	HFD	22.93	55.8
Wfdc3	2	MCD	47.08	94.8
Prdx2	8	MCD	16.70	52.2
Bckdhb	9	MCD	54.01	46.2
Gdi3	13	MCD	129.35	4
D14Ertd449e	14	HFD	85.90	15.9
Basp1	15	BD	23.16	16.4
E4f1	17	BD	18.25	13.2
Notch4	17	MCD	26.02	14.2
Frmd8	19	BD	20.31	8.5
Ms4a6c	19	MCD	16.26	10.5
Pkd2l1	19	MCD	24.48	34.5

[a]Interaction: (HFD) Significant effect in high-fat diet only; (MCD) Significant effect in the control diet only; (BD) Differential effects in high-fat and control diets

[b]eQTL peak position in cM

Table 6

Causality results between cis-eQTL and metastatic phenotype

Marker (cM)[a]	Chr	eQTL	Phenotype	Relationship[b]	P-value
C1L10 (34.1)	1	3222401M22Rik	Met	Independent	0.04
C8L3 (20.5)	8	Afg3\l	Met	Independent	0.05
C8L6 (30.4)	8	Gtbp3	Met	Independent	0.05
C11L2 (2.9)	11	H2afv	Met	Causal	0.05
C19L6 (12.8)	19	6430407L02Rik	Met	Independent	0.04
C19L6 (12.8)	19	AW491445	Met	Independent	0.04
C19L6 (12.8)	19	1200004M23Rik	Met	Independent	0.04
C19L6 (12.8)	19	Gcnt1	Met	Independent	0.05
C19L9 (23.3)	19	1200004M23Rik	Met	Independent	0.05

[a]Marker: The genetic marker [4] used as the anchor for the analysis

[b]Relationship: Testing of the association between the marker, eQTL and phenotype

Table 7

Results from Oncomine database evaluation

Symbol	Chr	NF [a]	EE positive [b]	EE negative [c]
9330140K16Rik	1	x		
Wwc2	8			1
Ndufa13	8		2	2
Gtpbp3	8	x		
9130404D08Rik	8	x		
Dusp4	8		9	
9530006C21Rik	8	x		
Rbpms	8		4	
Wwc1	11			1
Sh3pxd2b	11	x		
Ascc2	11	x		
Stk10	11		1	4
Nipsnap1	11		1	1
Tens1	11			1
2210015D19Rik	11	x		
H2afv	11			7
Ccdc117	11	x		
Hist1h2ae	13			4
Btn1a1	13	x		
Fars1	13			1
Hist1h1c	13			4
Elmo1	13	x		
Tmem14c	13	x		
1810022C23Rik	13	x		
Hist1h4f	13	x		
Btn1a1	13	x		
Hist1h2be	13			3
Gdi3	13			1
Sca1	13		4	
Cxcl14	13		5	
Gcnt1	19		1	2
Atad1	19	x		
Slc1a1	19		4	
Ms4a6c	19	x		
Fads3	19	x		
Cbwd1	19		1	

Symbol	Chr	NF [a]	EE positive [b]	EE negative [c]
Insl6	19	x		
Frmd8	19	x		
Slc25a45	19			1
Efemp2	19			1
Ctsf	19	x		
Ccs	19	x		
Saps3	19	x		
Lip1	19	x		

The 44 candidates for the Metastatic QTL were evaluated in the Oncomine database to determine if a potential association between their expression and formation of metastasis has been previously reported in humans.

[a]NF: An x indicates that the particular candidate was not found in the Oncomine human breast cancer prognosis datasets

[b]EE positive: Indicates the number of studies reporting that increased expression is significantly associated with a positive (no metastasis, alive, disease free) clinical outcome.

[c]EE negative: Indicates the number of studies reporting that increased expression is significantly associated with a negative (metastasis, dead, relapse) clinical outcome.