Supplementary Materials for "Mapping Gene Expression QTL of Impure Tumor Samples"

Contents

A Supplementary Methods

A.1 Notation Table

The following table contains the notations used to develop the TReC and TReCASE models for an arbitrary gene and a candidate eQTL of this gene. Subscripts specifying the gene and eQTL are suppressed. The A allele and B allele are defined based on the genotype of the candidate eQTL.

TReC + ASE Quantities		
Value	**Dimension**	**Description**
$G(i)$	NA	The genotype of subject i at the specified eQTL. Can take values: AA – homozygous for A allele AB – heterozygous BB – homozygous for B allele
ρ_i	1×1	Estimate of the tumor purity for the tumor sample of subject i, defined as the proportion of cells that are tumor cells.
TReC Only Quantities		
Value	**Dimension**	**Description**
Y_i	1×1	Total read count at the given gene in the tumor sample of subject i.
μ_{iA}	1×1	The mean TReC for subject i at A allele.
μ_{iB}	1×1	The mean TReC for subject i at B allele.
μ_i	1×1	The mean TReC for subject i.
ϕ	1×1	The overdispersion parameter for the distribution of TReC.
$\mathbf{x}_i$	$P \times 1$	Vector of covariate values for subject i
$\boldsymbol{\beta}$	$P \times 1$	Vector of covariate impacts on log total read count.
d_i	1×1	Read depth of RNA-Seq experiment for subject i.
ASE Only Quantities		
Value	**Dimension**	**Description**
R_i	1×1	The total number of allele specific reads for subject i.
R_{iB}	1×1	The number of allele specific reads mapped to the B allele for subject i.
ψ	1×1	The overdispersion parameter for the distribution of the ASE.
eQTL Parameters		
Value	**Dimension**	**Description**
η	1×1	The eQTL effect in normal tissue: $\mu_{iB}^{(N)}/\mu_{iA}^{(N)}$.
γ	1×1	The eQTL effect in tumor tissue: $\mu_{iB}^{(T)}/\mu_{iA}^{(T)}$.
κ	1×1	An over-expression effect in the tumor for A allele: $\mu_{iA}^{(T)}/\mu_{iA}^{(N)}$.
ξ_i	1×1	The ratio of gene expression of B allele versus A allele for subject i, defined as μ_{iB}/μ_{iA}.

Table S1: Notation for defining the TReC and TReCASE models.

A.2 Optimization Algorithm

As mentioned in main text, the optimization routine for solving the TReC and TReCASE models uses a coordinate block ascent routine with the following steps.

(0) Select initial estimates for κ, η, and γ.

(1) Holding κ, η, γ, and ψ constant, use negative binomial regression to update $\boldsymbol{\beta}$ and ϕ.

(2) Holding $\boldsymbol{\beta}, \phi$ and ψ constant, use a Quasi-Newton method (LBFGS) to update κ, η, and γ.

(3) Holding $\boldsymbol{\beta}, \phi, \kappa, \eta$, and γ constant, update ψ using a Quasi-Newton method (LBFGS).

(4) Iterate steps (1)-(3) until convergence.

The algorithm above is specified for the TReCASE model. A similar algorithm is used for TReC model except that we need to remove step (3) and iterate steps (1) and (2) repeatedly (while removing ψ from estimation procedures) until convergence.

To fully define the algorithm above, a discussion of Step (0) is warranted. Under the null hypothesis $\eta = 1$, model fit proceeds following the above algorithm starting at position $\kappa = 1$ and $\gamma = 1$ and holding η fixed at 1. Under the null hypothesis $\gamma = 1$, model fit proceeds as above, starting at position $\kappa = 1$ and $\eta = 1$ and holding γ at 1 throughout. To fit the full model, we choose initial values for κ, η, and γ in accordance with the fit of the null hypothesis, either $\eta = 1$ or $\gamma = 1$, which gives larger likelihood value at its MLE. This initialization method ensures that the suggested likelihood ratio tests are well defined by avoiding situations where the likelihood of full model is less than the likelihood of a restricted model.

A.3 Mathematical Details for Optimization

Mathematical details for section (A.2) are presented in the following. Note that, as defined, κ, η and γ are strictly positive parameters. Thus, we estimate $\log(\eta)$, $\log(\gamma)$, and $\log(\kappa)$ in the optimization process to guarantee that κ, η and γ are all positive, and avoid constrained optimization when working directly with κ, η and γ.

A.3.1 Total Read Count (TReC) Model Component

To motivate the structure of the TReC model, consider the ratio of the mean expressions for alleles B versus allele A for subject i. Assume that the expression of each allele is a weighted sum of its expression in normal and tumor tissues, weighted by the proportional composition of the sample with respect to each type. One can then specify this ratio for subject i as:

$$\begin{aligned}
\xi_i = \frac{\mu_{iB}}{\mu_{iA}} &= \frac{(1-\rho_i)\mu_{iB}^{(N)} + \rho_i \mu_{iB}^{(T)}}{(1-\rho_i)\mu_{iA}^{(N)} + \rho_i \mu_{iA}^{(T)}} \\
&= \frac{(1-\rho_i)\left(\mu_{iB}^{(N)}/\mu_{iA}^{(N)}\right) + \rho_i \left(\mu_{iB}^{(T)}/\mu_{iA}^{(T)}\right)\left(\mu_{iA}^{(T)}/\mu_{iA}^{(N)}\right)}{1-\rho_i + \rho_i\left(\mu_{iA}^{(T)}/\mu_{iA}^{(N)}\right)} \\
&= \frac{(1-\rho_i)\eta + \rho_i\kappa\gamma}{1-\rho_i+\rho_i\kappa} = (1-c_i)\eta + c_i\gamma,
\end{aligned}$$

where $c_i = (\rho_i\kappa)/(1-\rho_i+\rho_i\kappa)$. Assuming now that the total expression for subject i is the sum of the expressions from each constituent allele and modeling $\mu_{i,AA}^{(N)} = \exp(x_i^T\boldsymbol{\beta})$, the above implies that our mean takes the following form:

$$\mu_i = \begin{cases} e^{x_i^T\boldsymbol{\beta}}(1-\rho_i+\rho_i\kappa), & \text{if } G(i) = AA \\ e^{x_i^T\boldsymbol{\beta}}(1-\rho_i+\rho_i\kappa)(1+\xi_i)/2, & \text{if } G(i) = AB \\ e^{x_i^T\boldsymbol{\beta}}(1-\rho_i+\rho_i\kappa)\xi_i, & \text{if } G(i) = BB \end{cases}$$

Under a negative binomial distribution, the likelihood component for the TReC model for a single subject is given by:

$$f(Y_i;\mu_i,\phi) = \frac{\Gamma(Y_i+1/\phi)}{Y_i!\Gamma(1/\phi)}\left(\frac{1}{1+\phi\mu_i}\right)^{1/\phi}\left(\frac{\phi\mu_i}{1+\phi\mu_i}\right)^{Y_i}.$$

Thus, the log-likelihood for this component takes the form:

$$\begin{aligned}\ell_{TReC} &= \sum_{i=1}^{N} \ell_{TReC}^{(i)} \\ &= \sum_{i=1}^{N} \left[\ln \left\{ \frac{\Gamma(y_i + 1/\phi)}{y_i! \Gamma(1/\phi)} \right\} - (1/\phi + y_i) \log(1 + \phi\mu_i) + y_i \log(\phi) + y_i \log(\mu_i) \right].\end{aligned}$$

Letting λ denote one of κ, η, or γ, we have:

$$\frac{\partial \ell_{TReC}}{\partial \log(\lambda)} = \sum_{i=1}^{N} \left(\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \lambda} \right) \left(\frac{\partial \lambda}{\partial \log(\lambda)} \right).$$

We derive each of these components in turn. First, consider $\partial \ell_{TReC}^{(i)} / \partial \mu_i$:

$$\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} = \frac{y_i}{\mu_i} - \frac{1 + \phi y_i}{1 + \phi \mu_i}.$$

For utility in later steps, lets consider derivatives of the form $\frac{\partial \xi_i}{\partial \lambda}$ and $\frac{\partial c_i}{\partial \kappa}$:

$$\frac{\partial \xi_i}{\partial \kappa} = (\gamma - \eta) \left(\frac{\partial c_i}{\partial \kappa} \right), \quad \frac{\partial c_i}{\partial \kappa} = \kappa^{-1} c_i (1 - c_i), \quad \frac{\partial \xi_i}{\partial \gamma} = c_i, \text{ and } \frac{\partial \xi_i}{\partial \eta} = 1 - c_i.$$

Next, consider $\partial \mu_i / \partial \lambda$. It is easiest to consider this component separately for each genotype. For $G(i) = AA$, μ_i is dependent on κ, but free of η and γ. Thus:

$$\frac{\partial \mu_i}{\partial \kappa} = e^{x_i^T \boldsymbol{\beta}} \rho_i, \text{ and } \frac{\partial \mu_i}{\partial \eta} = \frac{\partial \mu_i}{\partial \gamma} = 0.$$

For $G(i) = AB$, we have:

$$\begin{aligned}\frac{\partial \mu_i}{\partial \kappa} &= e^{x_i^T \boldsymbol{\beta}} \left[\rho_i \left(\frac{1 + \xi_i}{2} \right) + (1 - \rho_i + \rho_i \kappa)(1/2) \left(\frac{\partial \xi_i}{\partial \kappa} \right) \right] = e^{x_i^T \boldsymbol{\beta}} (\rho_i / 2)(1 + \gamma), \\ \frac{\partial \mu_i}{\partial \eta} &= e^{x_i^T \boldsymbol{\beta}} (1 - \rho_i + \rho_i \kappa)(1/2) \left(\frac{\partial \xi_i}{\partial \eta} \right) = e^{x_i^T \boldsymbol{\beta}} (1/2)(1 - \rho_i), \\ \frac{\partial \mu_i}{\partial \gamma} &= e^{x_i^T \boldsymbol{\beta}} (1 - \rho_i + \rho_i \kappa)(1/2) \left(\frac{\partial \xi_i}{\partial \gamma} \right) = e^{x_i^T \boldsymbol{\beta}} (\rho_i / 2) \kappa.\end{aligned}$$

Finally, for $G(i) = BB$, we have:

$$\frac{\partial \mu_i}{\partial \kappa} = e^{x_i^T \boldsymbol{\beta}} \left[\rho_i \xi_i + (1 - \rho_i + \rho_i \kappa) \left(\frac{\partial \xi_i}{\partial \kappa} \right) \right] = e^{x_i^T \boldsymbol{\beta}} \rho_i \gamma,$$
$$\frac{\partial \mu_i}{\partial \eta} = e^{x_i^T \boldsymbol{\beta}} (1 - \rho_i + \rho_i \kappa) \left(\frac{\partial \xi_i}{\partial \eta} \right) = e^{x_i^T \boldsymbol{\beta}} (1 - \rho_i),$$
$$\frac{\partial \mu_i}{\partial \gamma} = e^{x_i^T \boldsymbol{\beta}} (1 - \rho_i + \rho_i \kappa) \left(\frac{\partial \xi_i}{\partial \eta} \right) = e^{x_i^T \boldsymbol{\beta}} \rho_i \kappa.$$

While not used for the C++ implementation of the model, the R-version uses the Hessian matrix with respect to the κ, η, and γ variables. We derive it here for completeness. Let $\dot{\ell}_{TReC} = \frac{\partial \ell_{TReC}}{\partial log(\lambda)}$ where λ is one of κ, η, and γ. As specified above:

$$\dot{\ell}_{TReC} = \lambda \sum_{i=1}^{N} \left\{ \left(\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} \right) \left(\frac{\partial \mu_i}{\partial \lambda} \right) \right\}.$$

Then:

$$\begin{aligned}
\frac{\partial^2 \dot{\ell}_{TReC}}{\partial \log(\lambda)^2} &= \left(\frac{\partial \dot{\ell}_{TReC,\kappa}}{\partial \kappa} \right) \left(\frac{\partial \kappa}{\partial \log(\kappa)} \right) = \kappa \left(\frac{\partial \dot{\ell}_{TReC,\kappa}}{\partial \kappa} \right) \\
&= \dot{\ell}_{TReC,\kappa} + \kappa^2 \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i \partial \kappa} \right) \left(\frac{\partial \mu_i}{\partial \kappa} \right) + \kappa^2 \sum_{i=1}^{N} \left(\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} \right) \left(\frac{\partial^2 \mu_i}{\partial \kappa^2} \right) \\
&= \dot{\ell}_{TReC,\kappa} + \kappa^2 \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \kappa} \right)^2 + \kappa^2 \sum_{i=1}^{N} \left(\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} \right) \left(\frac{\partial^2 \mu_i}{\partial \kappa^2} \right) \\
&= \dot{\ell}_{TReC,\kappa} + \kappa^2 \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \kappa} \right)^2.
\end{aligned}$$

The last equality holds since $\frac{\partial^2 \mu_i}{\partial \kappa^2} = 0$ and we may plug in:

$$\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} = -\left(\frac{y_i}{\mu_i^2} \right) + \frac{\phi + \phi^2 y_i}{(1 + \phi \mu_i)^2}.$$

Similar results hold for η and γ and are given below:

$$\frac{\partial^2 \ell_{TReC}}{\partial \log(\eta)^2} = \dot{\ell}_{TReC,\eta} + \eta^2 \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \eta} \right)^2,$$
$$\frac{\partial^2 \ell_{TReC}}{\partial \log(\gamma)^2} = \dot{\ell}_{TReC,\gamma} + \gamma^2 \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \gamma} \right)^2.$$

To complete the Hessian, we compute the remaining results:

$$\frac{\partial^2 \ell_{TReC}}{\partial \log(\kappa) \partial \log(\eta)} = \eta\kappa \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \kappa} \right) \left(\frac{\partial \mu_i}{\partial \eta} \right),$$

$$\frac{\partial^2 \ell_{TReC}}{\partial \log(\kappa) \partial \log(\gamma)} = \gamma\kappa \sum_{i=1}^{N} \left[\left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \kappa} \right) \left(\frac{\partial \mu_i}{\partial \gamma} \right) + \left(\frac{\partial \ell_{TReC}^{(i)}}{\partial \mu_i} \right) \left(\frac{\partial^2 \mu_i}{\partial \kappa \partial \gamma} \right) \right],$$

$$\frac{\partial^2 \ell_{TReC}}{\partial \log(\eta) \partial \log(\gamma)} = \eta\gamma \sum_{i=1}^{N} \left(\frac{\partial^2 \ell_{TReC}^{(i)}}{\partial \mu_i^2} \right) \left(\frac{\partial \mu_i}{\partial \eta} \right) \left(\frac{\partial \mu_i}{\partial \gamma} \right),$$

where $\frac{\partial^2 \mu_i}{\partial \kappa \partial \eta} = \frac{\partial^2 \mu_i}{\partial \eta \partial \gamma} = 0$ and

$$\frac{\partial^2 \mu_i}{\partial \kappa \partial \gamma} = \begin{cases} 0, & \text{if } G(i) = AA \\ (1/2) e^{x_i^T \boldsymbol{\beta}} \rho_i, & \text{if } G(i) = AB \\ e^{x_i^T \boldsymbol{\beta}} \rho_i, & \text{if } G(i) = BB \end{cases}$$

A.3.2 Allele Specific Expression (ASE) Model Component

In the following, let μ_{i1} represent the number of reads that are expressed by allele 1 on average for subject i and μ_{i2} be its counterpoint for allele 2. Within a sample prepped for RNA-seq, the pool of reads for the given gene contains $\mu_{i1} + \mu_{i2}$ reads. The proportion of reads belonging to allele 1 on average is then given by:

$$\pi_i = \frac{\mu_{i1}}{\mu_{i1} + \mu_{i2}} = \frac{(\mu_{i1}/\mu_{i2})}{1 + \mu_{i1}/\mu_{i2}}.$$

Thus, viewing the RNA-Seq sampling procedure as drawing a group of reads at random and allowing for extra-binomial variation, we can model the data-generation mechanism via a beta-binomial distribution. Extra-binomial variation is often observed in genetic studies and in the case of ASE reads can in part be attributed to incorrectly genotyped alleles resulting from genotyping or imputation error.

In order to model a consistent eQTL effect within the TReC and ASE components of the model, define allele 1 as that containing the minor allele B for heterozygous subjects. In homozygous subjects, an arbitrary allele is selected as the expression between the two alleles is assumed to be equal on average. Thus, by the statement above and previous

definitions, we may model the average reads for allele 1 as:

$$\pi_i = \begin{cases} \xi_i/(1+\xi_i), & \text{if } G(i) = BB \\ (1/2), & \text{otherwise} \end{cases}$$

Thus, the likelihood for the ASE component of the model is given by:

$$f(r_{iB}; r_i, \pi_i, \psi) = \binom{r_i}{r_{iB}} \left[\frac{\Gamma\left(\psi^{-1}\right)}{\Gamma\left(\psi^{-1}\pi_i\right)\Gamma\left(\psi^{-1}(1-\pi_i)\right)} \right] \times \left[\frac{\Gamma\left(\psi^{-1}\pi_i + r_{iB}\right)\Gamma\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right)}{\Gamma\left(\psi^{-1} + r_i\right)} \right].$$

Define $\ell_{ASE}^{(i)}$ be the ASE likelihood from the $i-th$ sample. Then:

$$\ell_{ASE} = \sum_{i=1}^{n} \ell_{ASE}^{(i)} = \sum_{i=1}^{n} \log\left[f(r_{iB}; r_i, \pi_i, \psi)\right].$$

It can be seen that the gradient functions for π_i and ψ are given by:

$$\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} = \psi^{-1}\left[\Psi_0\left(\psi^{-1}\pi_i + r_{iB}\right) - \Psi_0\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right) - \Psi_0\left(\psi^{-1}\pi_i\right) + \Psi_0\left(\psi^{-1}(1-\pi_i)\right)\right],$$

$$\begin{aligned} \frac{\partial \ell_{ASE}}{\partial \psi} = &\sum_{i=1}^{n} -\psi^{-2}\pi_i\left[\Psi_0\left(\psi^{-1}\pi_i + r_{iB}\right) - \Psi_0\left(\psi^{-1}\pi_i\right)\right] - \\ &\sum_{i=1}^{n} \psi^{-2}(1-\pi_i)\left[\Psi_0\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right) - \Psi_0\left(\psi^{-1}(1-\pi_i)\right)\right] - \\ &\sum_{i=1}^{n} \psi^{-2}\left[\Psi_0\left(\psi^{-1}\right) - \Psi_0\left(\psi^{-1} + r_i\right)\right]. \end{aligned}$$

Before deriving the remaining components necessary for the gradient, we note that only individuals of heterozygous genotype contribute to the gradient of κ, η and γ, whereas all individuals contributed to the gradient of ψ. Thus, we have:

$$\begin{aligned} \frac{\partial \ell_{ASE}}{\partial \log(\lambda)} \equiv \dot{\ell}_{ASE,\lambda} &= \sum_{i;G(i)=AB} \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i}\right)\left(\frac{\partial \pi_i}{\partial \xi_i}\right)\left(\frac{\partial \xi_i}{\partial \lambda}\right)\left(\frac{\partial \lambda}{\partial \log(\lambda)}\right) \\ &= \lambda \sum_{i;G(i)=AB} \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i}\right)\left(\frac{\partial \pi_i}{\partial \xi_i}\right)\left(\frac{\partial \xi_i}{\partial \lambda}\right). \end{aligned}$$

To calculate the above quantity, we need:

$$\frac{\partial \pi_i}{\partial \xi_i} = (1+\xi_i)^{-2}, \quad \frac{\partial \xi_i}{\partial \eta} = 1 - c_i, \quad \frac{\partial \xi_i}{\partial \gamma} = c_i, \quad \frac{\partial \xi_i}{\partial \kappa} = (\gamma - \eta) c_i (1 - c_i) \kappa^{-1}.$$

As noted in the previous section, the C++ fit routine does not utilize the Hessian but we provide its derivation here for completeness. We will make repeated use of the following terms, so they are presented here for later reference.

$$\begin{aligned}
\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} &= \psi^{-2} \left[\Psi_1 \left(\psi^{-1} \pi_i + r_{iB} \right) + \Psi_1 \left(\psi^{-1} (1 - \pi_i) + r_i - r_{iB} \right) - \Psi_1 \left(\psi^{-1} \pi_i \right) - \Psi_1 \left(\psi^{-1} (1 - \pi_i) \right) \right] \\
\frac{\partial^2 \pi_i}{\partial \xi_i^2} &= -2(1+\xi_i)^{-3} \\
\frac{\partial^2 \xi_i}{\partial \kappa^2} &= (\gamma - \eta) \left[\kappa^{-1} (1 - 2c_i) \left(\frac{\partial c_i}{\partial \kappa} \right) - \kappa^{-2} c_i (1 - c_i) \right],
\end{aligned}$$

where

$$\Psi_0(x) = \frac{\partial \ln \Gamma(x)}{\partial x} \quad \text{and } \Psi_1(x) = \frac{\partial^2 \ln \Gamma(x)}{\partial x^2}.$$

We complete the derivation in the following.

$$\begin{aligned}
\frac{\partial^2 \ell_{ASE}}{\partial \log(\kappa)^2} &= \left(\frac{\partial \dot{\ell}_{ASE,\kappa}}{\partial \kappa} \right) \left(\frac{\partial \kappa}{\partial \log(\kappa)} \right) = \kappa \left(\frac{\partial \dot{\ell}_{ASE,\kappa}}{\partial \kappa} \right) \\
&= \dot{\ell}_{ASE,\kappa} + \\
&\quad \kappa^2 \sum_{i;G(i)=AB} \left\{ \left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 \left(\frac{\partial \xi_i}{\partial \kappa} \right)^2 + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \left(\frac{\partial \xi_i}{\partial \kappa} \right)^2 + \left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right) \left(\frac{\partial^2 \xi_i}{\partial \kappa^2} \right) \right\}.
\end{aligned}$$

Similarly for η and γ, we have:

$$\begin{aligned}
\frac{\partial^2 \ell_{ASE}}{\partial \log(\eta)^2} &= \dot{\ell}_{ASE,\eta} + \eta^2 \sum_{i;G(i)=AB} (1 - c_i)^2 \left[\left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \right], \\
\frac{\partial^2 \ell_{ASE}}{\partial \log(\gamma)^2} &= \dot{\ell}_{ASE,\gamma} + \gamma^2 \sum_{i;G(i)=AB} c_i^2 \left[\left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \right].
\end{aligned}$$

Finally, for the "mixed" second derivatives, we have:

$$\frac{\partial^2 \ell_{ASE}}{\partial \log(\eta)\partial \log(\kappa)} = \kappa\eta \sum_{i;G(i)=AB} \left[\left\{ \left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 - \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \right\} \left(\frac{\partial \xi_i}{\partial \kappa} \right) (1 - c_i) - \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right) \left(\frac{\partial c_i}{\partial \kappa} \right) \right],$$

$$\frac{\partial^2 \ell_{ASE}}{\partial \log(\gamma)\partial \log(\kappa)} = \kappa\gamma \sum_{i;G(i)=AB} \left[\left\{ \left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \right\} \left(\frac{\partial \xi_i}{\partial \kappa} \right) c_i + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right) \left(\frac{\partial c_i}{\partial \kappa} \right) \right],$$

$$\frac{\partial^2 \ell_{ASE}}{\partial \log(\gamma)\partial \log(\eta)} = \gamma\eta \sum_{i;G(i)=AB} c_i(1 - c_i) \left[\left(\frac{\partial^2 \ell_{ASE}^{(i)}}{\partial \pi_i^2} \right) \left(\frac{\partial \pi_i}{\partial \xi_i} \right)^2 + \left(\frac{\partial \ell_{ASE}^{(i)}}{\partial \pi_i} \right) \left(\frac{\partial^2 \pi_i}{\partial \xi_i^2} \right) \right].$$

A.4 Cis-Trans Score Test

Recall that eQTL come in two varieties: *cis*- and *trans*-eQTL. *cis*-eQTLs induce allelic imbalance of gene expression whereas *trans*-eQTLs affect the expression of two alleles to the same degree. Sun [2011] and Hu et al. [2015] have developed and refined a "Cis-Trans test" to identify whether eQTL act in a *cis*- or *trans*- fashion. Under the null hypothesis (*cis*-), the eQTL effect sizes are the same between TReC and ASE models. A small p-value using this test leads to rejection of the null hypothesis, and thus the conclusion that the given Gene-SNP pair behave in a *trans*-eQTL manner. In that case, only the TReC data should be used for eQTL mapping.

To develop this test for eQTL mapping in tumor tissues, we follow Hu et al. [2015] by extending the likelihood framework through the introduction of new parameters which allow eQTL effects to differ between TReC and ASE components. Specifically, we define:

$$\eta_{ASE} = \eta + \alpha_\eta, \text{ and } \gamma_{ASE} = \gamma + \alpha\gamma,$$

where η and γ are the TReC-specific eQTL effects in normal and tumor tissues, respectively; η_{ASE} and γ_{ASE} are the ASE-specific counterparts; α_η and α_γ are the discrepancies of eQTL effects between ASE and TReC components of the model in normal and tumor tissues, respectively. Then to test *cis*- versus *trans*-eQTL, we employ a score test for the two-dimensional hypothesis: $\alpha_\eta = \alpha_\gamma = 0$.

A.4.1 Structure of the Score Test

Define the following groups of parameters: $\epsilon = (\kappa, \eta, \gamma)^T$; $\alpha = (\alpha_\eta, \alpha_\gamma)^T$; and $\Theta = (\boldsymbol{\beta}^T, \epsilon^T, \alpha^T, \phi, \psi)$. Let $\ell = \ell_{TReC} + \ell_{ASE}$ be the full data log-likelihood, $\dot{\ell}$ be the first

derivative of the log-likelihood with respect to the parameters, and $I(\Theta)$ be the Fisher's Information Matrix. We may specify the Fisher's Information Matrix in the following way:

$$I(\Theta) = \begin{pmatrix} I_{\boldsymbol{\beta},\boldsymbol{\beta}} & I_{\boldsymbol{\beta},\epsilon} & I_{\boldsymbol{\beta},\phi} & I_{\boldsymbol{\beta},\psi} & I_{\boldsymbol{\beta},\alpha} \\ I_{\epsilon,\boldsymbol{\beta}} & I_{\epsilon,\epsilon} & I_{\epsilon,\phi} & I_{\epsilon,\psi} & I_{\epsilon,\alpha} \\ I_{\phi,\boldsymbol{\beta}} & I_{\phi,\epsilon} & I_{\phi,\phi} & I_{\phi,\psi} & I_{\phi,\alpha} \\ I_{\psi,\boldsymbol{\beta}} & I_{\psi,\epsilon} & I_{\psi,\phi} & I_{\psi,\psi} & I_{\psi,\alpha} \\ I_{\alpha,\boldsymbol{\beta}} & I_{\alpha,\epsilon} & I_{\alpha,\phi} & I_{\alpha,\psi} & I_{\alpha,\alpha} \end{pmatrix} = \begin{pmatrix} M_1 & M_2 \\ M_2^T & I_{\alpha,\alpha} \end{pmatrix},$$

where M_1 is the upper-left block of the Fisher's Information matrix through $I_{\psi,\psi}$ and M_2 is the remaining block excluding $I_{\alpha,\alpha}$.

Following the developments of Radhakrishna Rao [1948], we may compute the score test of $\alpha_\eta = \alpha_\gamma = 0$ in the following way:

$$\begin{aligned} SC &= \dot{\ell}\left(\hat{\Theta}\right)^T I\left(\hat{\Theta}\right)^{-1} \dot{\ell}\left(\hat{\Theta}\right) \\ &= \begin{pmatrix} \frac{\partial \ell}{\partial \alpha_\eta} & \frac{\partial \ell}{\partial \alpha_\gamma} \end{pmatrix} \left(I_{\alpha,\alpha} - M_2^T M_1^{-1} M_2\right)^{-1} \begin{pmatrix} \frac{\partial \ell}{\partial \alpha_\eta} \\ \frac{\partial \ell}{\partial \alpha_\gamma} \end{pmatrix} \Bigg|_{\Theta=\hat{\Theta}}, \end{aligned}$$

where $\hat{\theta}$ is the estimate of our parameters under the null. SC asymptotically follows a Chi-squared distribution with two degrees of freedom under the null.

A.4.2 TReC Derivatives

Preceding development of the gradients and Hessians of the TReC components in the following section, it will be helpful to compose a list of definitions and useful derivatives for later use. Recall that μ_i is the mean read count in the TReC component of the model, given by:

$$\mu_i = \begin{cases} e^{x_i^T \boldsymbol{\beta}} \left[1 - \rho_i + \rho_i \kappa\right], & \text{if } G(i) = AA, \\ e^{x_i^T \boldsymbol{\beta}} \left[1 - \rho_i + \rho_i \kappa\right] \left[\frac{1+\xi_i}{2}\right], & \text{if } G(i) = AB, \\ e^{x_i^T \boldsymbol{\beta}} \left[1 - \rho_i + \rho_i \kappa\right] \xi_i, & \text{if } G(i) = BB, \end{cases}$$

where $\xi_i = (1 - c_i)\eta + c_i\gamma$ and $c_i = (\rho_i\kappa)/(1 - \rho_i + \rho_i\kappa)$. It is clear that:

$$\begin{aligned} \frac{\partial c_i}{\partial \kappa} &= \kappa^{-1} c_i (1 - c_i) \\ \frac{\partial^2 c_i}{\partial \kappa^2} &= \kappa^{-1}(1 - 2c_i)\left(\frac{\partial c_i}{\partial \kappa}\right) - \kappa^{-2} c_i (1 - c_i) \end{aligned}$$

This allows us to compose the following derivatives for ξ_i:

$$\frac{\partial \xi_i}{\partial \kappa} = (\gamma - \eta)\left(\frac{\partial c_i}{\partial \kappa}\right)$$
$$\frac{\partial \xi_i}{\partial \eta} = (1 - c_i)$$
$$\frac{\partial \xi_i}{\partial \gamma} = c_i$$

The Hessian for ξ_i is provided by the following

$$\frac{\partial \xi_i}{\partial \epsilon \partial \epsilon^T} = \begin{pmatrix} (\gamma - \eta)\left(\frac{\partial^2 c_i}{\partial \kappa^2}\right) & -\frac{\partial c_i}{\partial \kappa} & \frac{\partial c_i}{\partial \kappa} \\ & 0 & 0 \\ & & 0 \end{pmatrix}$$

The gradient of μ_i with respect to ϵ is provided below:

$$\left.\frac{\partial \mu_i}{\partial \epsilon}\right|_{G(i)=AA} = e^{x_i^T \boldsymbol{\beta}} \begin{pmatrix} \rho_i \\ 0 \\ 0 \end{pmatrix}$$

$$\left.\frac{\partial \mu_i}{\partial \epsilon}\right|_{G(i)=AB} = e^{x_i^T \boldsymbol{\beta}} \begin{pmatrix} \left[\rho_i\left(\frac{1+\xi_i}{2}\right) + (1 - \rho_i + \rho_i\kappa)(1/2)\left(\frac{\partial \xi_i}{\partial \kappa}\right)\right] \\ \left[(1 - \rho_i + \rho_i\kappa)(1/2)\left(\frac{\partial \xi_i}{\partial \eta}\right)\right] \\ \left[(1 - \rho_i + \rho_i\kappa)(1/2)\left(\frac{\partial \xi_i}{\partial \gamma}\right)\right] \end{pmatrix}$$

$$\left.\frac{\partial \mu_i}{\partial \epsilon}\right|_{G(i)=BB} = e^{x_i^T \boldsymbol{\beta}} \begin{pmatrix} \left[\rho_i\xi_i + (1 - \rho_i + \rho_i\kappa)\left(\frac{\partial \xi_i}{\partial \kappa}\right)\right] \\ \left[(1 - \rho_i + \rho_i\kappa)\left(\frac{\partial \xi_i}{\partial \eta}\right)\right] \\ \left[(1 - \rho_i + \rho_i\kappa)\left(\frac{\partial \xi_i}{\partial \gamma}\right)\right] \end{pmatrix}$$

The Hessian for μ_i is identically 0 for genotype AA. However, for genotypes AB and BB, we have the following where we define $\delta_i = 1 - \rho_i + \rho_i\kappa$.

$$\frac{\partial^2 \mu_i}{\partial \epsilon \partial \epsilon^T} = e^{x_i^T \boldsymbol{\beta}} \begin{pmatrix} \left[\rho_i\left(\frac{\partial \xi_i}{\partial \kappa}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa^2}\right)\right] & (1/2)\left[\rho_i\left(\frac{\partial \xi_i}{\partial \eta}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \eta}\right)\right] & (1/2)\left[\rho_i\left(\frac{\partial \xi_i}{\partial \gamma}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \gamma}\right)\right] \\ (1/2)\left[\rho_i\left(\frac{\partial \xi_i}{\partial \eta}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \eta}\right)\right] & 0 & 0 \\ (1/2)\left[\rho_i\left(\frac{\partial \xi_i}{\partial \gamma}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \gamma}\right)\right] & 0 & 0 \end{pmatrix}$$

$$\frac{\partial^2 \mu_i}{\partial \epsilon \partial \epsilon^T} = e^{x_i^T \boldsymbol{\beta}} \begin{pmatrix} \left[2\rho_i\left(\frac{\partial \xi_i}{\partial \kappa}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa^2}\right)\right] & \left[\rho_i\left(\frac{\partial \xi_i}{\partial \eta}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \eta}\right)\right] & \left[\rho_i\left(\frac{\partial \xi_i}{\partial \gamma}\right) + \delta_i\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \gamma}\right)\right] \\ \left[\rho_i\left(\frac{\partial \xi_i}{\partial \eta}\right) + \delta_i(1/2)\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \eta}\right)\right] & 0 & 0 \\ \left[\rho_i\left(\frac{\partial \xi_i}{\partial \gamma}\right) + \delta_i(1/2)\left(\frac{\partial^2 \xi_i}{\partial \kappa \partial \gamma}\right)\right] & 0 & 0 \end{pmatrix}$$

To simplify the notation in our derivation, we define the following $n \times n$ diagonal matrices, Δ_1 through Δ_6. Elements on the diagonal are contained within the diag() notation below and are specified for a single subject.

$$\Delta_1 = \text{diag}\left(\frac{\mu_i}{Var[Y_i]}\right)$$
$$\Delta_2 = \text{diag}\left(\frac{\mu_i^2}{Var[Y_i]}\right)$$
$$\Delta_3 = \text{diag}\left(\frac{\mu_i^3(y_i - \mu_i)}{Var[Y_i]^2}\right)$$
$$\Delta_4 = \text{diag}\left(\frac{\mu_i^2(y_i - \mu_i)}{Var[Y_i]^2}\right)$$
$$\Delta_5 = \text{diag}\left(\frac{1}{Var[Y_i]}\right)$$
$$\Delta_6 = \text{diag}\left(\frac{(y_i - \mu_i)(1 + 2 * \phi\mu_i)}{Var[Y_i]^2}\right)$$

The log-likelihood for the TReC component is given by:

$$\ell_{TReC} = \sum_{i=1}^{N} \ln\Gamma\left(y_i + 1/\phi\right) - \ln\Gamma\left(1/\phi\right) - \ln\Gamma\left(y_i + 1\right) - \left[1/\phi + y_i\right]\ln\left(1 + \phi\mu_i\right) + y_i\left(\ln(\phi) + \ln(\mu_i)\right)$$

It can be shown that the following hold for derivatives involving $\boldsymbol{\beta}$:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\left(\frac{y_i - \mu_i}{1 + \phi\mu_i}\right)x_i = X^T\Delta_1(Y - \mu)$$
$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{N}\left[\frac{\mu_i}{1 + \phi\mu_i} + \frac{\phi\mu_i\left(y_i - \mu_i\right)}{(1 + \phi\mu_i)^2}\right]x_i x_i^T = -\left[X^T\Delta_2 X + \phi X^T\Delta_3 X\right]$$
$$\frac{\partial \ell}{\partial \boldsymbol{\beta}\partial \epsilon^T} = -\sum_{i=1}^{N}\left[\frac{1}{1 + \phi\mu_i} + \frac{\phi(y_i - \mu_i)}{(1 + \phi\mu_i)^2}\right]x_i\frac{\partial \mu_i}{\partial \epsilon}^T = -\left[X^T\Delta_1 D_\mu(\epsilon) + \phi X^T\Delta_4 D_\mu(\epsilon)\right]$$
$$\frac{\partial \ell}{\partial \boldsymbol{\beta}\partial \phi} = -\sum_{i=1}^{N}\left[\frac{(y_i - \mu_i)\mu_i}{(1 + \phi\mu_i)^2}\right]x_i = -X^T\Delta_3 J_N$$

Regarding derivatives involving ϵ, we have:

$$\frac{\partial \ell_{TReC}}{\partial \epsilon} = \sum_{i=1}^{N} \left[\frac{y_i - \mu_i}{\mu_i + \phi \mu_i^2} \right] \frac{\partial \mu_i}{\partial \epsilon} = D_\mu(\epsilon)^T \Delta_5 \left(Y - \mu \right),$$

$$\begin{aligned}
\frac{\partial \ell_{TReC}}{\partial \epsilon \partial \epsilon^T} &= \sum_{i=1}^{N} - \left[\frac{1}{\mu_i + \phi \mu_i^2} + \frac{(y_i - \mu_i)(1 + 2\phi\mu_i)}{(\mu_i + \phi \mu_i^2)^2} \right] \left(\frac{\partial \mu_i}{\partial \epsilon} \right) \left(\frac{\partial \mu_i}{\partial \epsilon} \right)^T + \left(\frac{y_i - \mu_i}{\mu_i + \phi \mu_i^2} \right) \left(\frac{\partial^2 \mu_i}{\partial \epsilon \partial \epsilon^T} \right) \\
&= - \left[D_\mu(\epsilon)^T \Delta_5 D_\mu(\epsilon) + D_\mu(\epsilon)^T \Delta_6 D_\mu(\epsilon) \right] + \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\mu_i + \phi \mu_i^2} \right) \left(\frac{\partial^2 \mu_i}{\partial \epsilon \partial \epsilon^T} \right),
\end{aligned}$$

$$\frac{\partial \ell}{\partial \epsilon \partial \phi} = - \sum_{i=1}^{N} \frac{(y_i - \mu_i)\mu_i^2}{(\mu_i + \phi \mu_i^2)^2} = -D_\mu(\epsilon)^T \Delta_4 J_N.$$

Finally, derivatives involving ϕ are provided below:

$$\frac{\partial \ell}{\partial \phi} = \sum_{i=1}^{N} -\phi^{-2} \left[\Psi_0(y_i + \phi^{-1}) - \Psi_0(\phi^{-1}) - \ln(1 + \phi\mu_i) \right] - (\phi^{-1} + y_i) \left[\frac{\mu_i}{1 + \phi\mu_i} \right] + \frac{y_i}{\phi}$$

$$\begin{aligned}
\frac{\partial \ell}{\partial \phi^2} = \sum_{i=1}^{N} & 2\phi^{-3} \left[\Psi_0(y_i + \phi^{-1}) - \Psi_0(\phi^{-1}) - \ln(1 + \phi\mu_i) \right] + \phi^{-4} \left[\Psi_1(y_i + \phi^{-1}) - \Psi_1(\phi^{-1}) \right] + 2\phi^{-2} \left[\frac{\mu_i}{1 + \phi\mu_i} \right] - \frac{y_i}{\phi^2} \\
& + (\phi^{-1} + y_i) \left[\frac{\mu_i^4}{V[Y_i]^2} \right]
\end{aligned}$$

A.4.3 ASE Derivatives

Preceding development of the gradients and Hessians of the ASE component in the following section, it will be helpful to compose a list of definitions and useful derivatives for later use. Recall the definitions of ξ_i^A and π_i:

$$\xi_i^A = (1-c_i)(\eta+\alpha_\eta) + c_i(\gamma+\alpha_\gamma)$$
$$\pi_i = \begin{cases} \xi_i^A/(1+\xi_i^A) & \text{, if } G(i) = AB \\ 0.5 & \text{, otherwise} \end{cases}$$

For genotypes AA and AB, π_i is independent of our parameters. Only genotype AB will be considered. Thus, consider the gradient of ξ_i^A with respect to our parameters.

$$\frac{\partial \xi_i^A}{\partial(\epsilon,\alpha)} = \begin{pmatrix} [(\gamma+\alpha_\gamma)-(\eta+\alpha_\eta)]\left(\frac{\partial c_i}{\partial \kappa}\right) \\ 1-c_i \\ c_i \\ 1-c_i \\ c_i \end{pmatrix}$$

The Hessian of ξ_i is presented below:

$$\frac{\partial \xi_i^A}{\partial(\epsilon,\alpha)\partial(\epsilon,\alpha)^T} = \begin{pmatrix} [(\gamma+\alpha_\gamma)-(\eta+\alpha_\eta)]\left(\frac{\partial^2 c_i}{\partial \kappa^2}\right) & -\frac{\partial c_i}{\partial \kappa} & \frac{\partial c_i}{\partial \kappa} & -\frac{\partial c_i}{\partial \kappa} & \frac{\partial c_i}{\partial \kappa} \\ -\frac{\partial c_i}{\partial \kappa} & 0 & 0 & 0 & 0 \\ \frac{\partial c_i}{\partial \kappa} & 0 & 0 & 0 & 0 \\ -\frac{\partial c_i}{\partial \kappa} & 0 & 0 & 0 & 0 \\ \frac{\partial c_i}{\partial \kappa} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Then for an arbitrary λ, we have:

$$\frac{\partial \pi_i}{\partial \lambda} = (1+\xi_i^A)^{-2}\left(\frac{\partial \xi_i^A}{\partial \lambda}\right)$$

$$\frac{\partial^2 \pi_i}{\partial \lambda_1 \partial \lambda_2} = -2(1+\xi_i^A)^{-3}\left(\frac{\partial \xi_i}{\partial \lambda_1}\right)\left(\frac{\partial \xi_i}{\partial \lambda_2}\right) + (1+\xi_i^A)^{-2}\left(\frac{\partial^2 \xi_i^A}{\partial \lambda_1 \partial \lambda_2}\right)$$

The log-likelihood for the ASE component of the data is given below:

$$\ell_{ASE} = \sum_{i=1}^{n} \ln\Gamma(r_i+1) - \ln\Gamma(r_{iB}+1) - \ln\Gamma(r_i - r_{iB}+1) + \ln\Gamma\left(\psi^{-1}\right) - \ln\Gamma\left(\psi^{-1}\pi_i\right) - \ln\Gamma\left(\psi^{-1}(1-\pi_i)\right) + \ln\Gamma\left(\psi^{-1}\pi_i + r_{iB}\right) + \ln\Gamma\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right) - \ln\Gamma\left(\psi^{-1}+r_i\right)$$

Let λ represent a single parameter from either ϵ or α. For such terms, it can be shown that:

$$\begin{aligned}
\frac{\partial \ell_{ASE}}{\partial \lambda} &= \sum_{i=1}^{n} \psi^{-1} B_i \left(\frac{\partial \pi_i}{\partial \lambda} \right) \\
\frac{\partial^2 \ell_{ASE}}{\partial \lambda_1 \partial \lambda_2} &= \sum_{i=1}^{n} \psi^{-1} B_i \left(\frac{\partial^2 \pi_i}{\partial \lambda_1 \partial \lambda_2} \right) + \psi^{-1} \left(\frac{\partial B_i}{\partial \pi_i} \right) \left(\frac{\partial \pi_i}{\partial \lambda_1} \right) \left(\frac{\partial \pi_i}{\partial \lambda_2} \right) \\
\frac{\partial \ell_{ASE}}{\partial \lambda \partial \psi} &= \sum_{i=1}^{n} -\psi^{-2} B_i \left(\frac{\partial \pi_i}{\partial \lambda} \right) + \psi^{-1} \left(\frac{\partial B_i}{\partial \psi} \right) \left(\frac{\partial \pi_i}{\partial \lambda} \right)
\end{aligned}$$

Where we define B_i and it's derivatives in the following way:

$$\begin{aligned}
B_i &= -\Psi_0\left(\psi^{-1}\pi_i\right) + \Psi_0\left(\psi^{-1}(1-\pi_i)\right) + \Psi_0\left(\psi^{-1}\pi_i + r_{iB}\right) - \Psi_0\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right) \\
\frac{\partial B_i}{\partial \pi_i} &= \psi^{-1}\left[-\Psi_1\left(\psi^{-1}\pi_i\right) - \Psi_1\left(\psi^{-1}(1-\pi_i)\right) + \Psi_1\left(\psi^{-1}\pi_i + r_{iB}\right) + \Psi_1\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right)\right] \\
\frac{\partial B_i}{\partial \psi} &= -\psi^{-2}\pi_i\left[-\Psi_1\left(\psi^{-1}\pi_i\right) + \Psi_1\left(\psi^{-1}\pi_i + r_{iB}\right)\right] + \\
&\quad - \psi^{-2}(1-\pi_i)\left[\Psi_1\left(\psi^{-1}(1-\pi_i)\right) - \Psi_1\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right)\right]
\end{aligned}$$

Derivatives involving ψ are specified below:

$$\begin{aligned}
\frac{\partial \ell_{ASE}}{\partial \psi} &= \sum_{i=1}^{N_{AS}} -\psi^{-2} A_i, \\
\frac{\partial^2 \ell_{ASE}}{\partial \psi^2} &= \sum_{i=1}^{N_{AS}} 2\psi^{-3} A_i - \psi^{-2}\left(\frac{\partial A_i}{\partial \psi}\right),
\end{aligned}$$

where A_i and its derivatives are specified by:

$$\begin{aligned}
A_i &= \pi_i\left[-\Psi_0\left(\psi^{-1}\pi_i\right) + \Psi_0\left(\psi^{-1}\pi_i + r_{iB}\right)\right] + \\
&\quad (1-\pi_i)\left[-\Psi_0\left(\psi^{-1}(1-\pi_i)\right) + \Psi_0\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right)\right] + \\
&\quad \left[\Psi_0\left(\psi^{-1}\right) - \Psi_0\left(\psi^{-1} + r_i\right)\right], \\
\frac{\partial A_i}{\partial \psi} &= -\psi^{-2}\pi_i^2\left[-\Psi_1\left(\psi^{-1}\pi_i\right) + \Psi_1\left(\psi^{-1}\pi_i + r_{iB}\right)\right] - \\
&\quad \psi^{-2}(1-\pi_i)^2\left[-\Psi_1\left(\psi^{-1}(1-\pi_i)\right) + \Psi_1\left(\psi^{-1}(1-\pi_i) + r_i - r_{iB}\right)\right] - \\
&\quad \psi^{-2}\left[\Psi_1\left(\psi^{-1}\right) - \Psi_1\left(\psi^{-1} + r_i\right)\right].
\end{aligned}$$

A.4.4 Fisher's Information: Observed or Expected

The traditional form of the score test involves use of the expected Fisher's Information Matrix. In the case where the expected value of the Fisher's Information Matrix is difficult to compute, the observed Fisher's Information Matrix is often used [Freedman, 2007]. In some situations, while using the observed Fisher's Information Matrix still provides a statistically valid test under the null, it can be unstable and produce inconsistent estimates of the variance matrix for MLEs [Freedman, 2007]. In the likelihood framework proposed by this paper, there is an inherent, stochastic dependence of R_i on Y_i. Namely, the value of R_i depends on the number of heterozygous SNPs present within the gene body and cannot exceed Y_i. This makes computing the expected Fisher's Information Matrix challenging as it becomes an infinite sum of finite sums containing the digamma and trigamma functions.

As such, we may compute an approximation to the expected Fisher's Information Matrix which assumes that Y_i and R_i are stochastically independent or we may use the observed Fisher's Information Matrix. The observed Fisher's Information Matrix can be computed as in the previous section using untransformed κ, η, γ, and ψ or the log-transformations of these quantities. The log transformation variant of the observed score test, termed Observed Score test (log), is slightly more stable than its untransformed competitor. A comparison of these three methods [observed, observed (log), expected] on simulated data is provided below (Supplementary Table S2). To evaluate Type I error of the Cis-Trans score test, simulations follow the structure provided for the power simulations. To evaluate power, $\xi_{i,ASE}$ is set to 1 for all subjects regardless of eQTL genotype and eQTL effect size. This behavior is designed to mimic trans-eQTL behavior. In the case of numerical instability for the observed information Cis-Trans score tests, the expected information variant is substituted.

As we can see from Supplementary Table S2, the observed information matrix variants of the Cis-Trans Score test display superior power to the expected information variant at the cost of an inflated type I error ($\sim$8%). In addition, we note that the numerical instability of the observed information variants leads to a high rate of computation failure for the Cis-Trans score test. Due to its superior stability and Type I error, we opt to use the approximated expected Fisher's Information matrix within the real data analysis.

γ Value	Observed Score Test Power	Observed Score Test Type I Error	Observed Score Test (log) Power	Observed Score Test (log) Type I Error	Expected Score Test Power	Expected Score Test Type I Error
1.0	–	8.4 (7)	–	8.9 (4)	–	6.6 (5)
1.2	25.8 (6)	8.0 (2)	25.3 (4)	8.0 (4)	15.5 (0)	5.3 (1)
1.4	66.8 (38)	9.5 (10)	63.5 (35)	8.5 (3)	48.8 (0)	3.5 (0)
1.6	89.0 (88)	9.3 (2)	86.3 (66)	8.5 (4)	82.3 (0)	4.3 (0)
1.8	99.0 (185)	8.5 (2)	98.5 (164)	10.3 (3)	97.3 (0)	3.8 (0)

Table S2: Summarizing the power and Type I error of the derived score tests. Number in parentheses represents the number of failures due to numerical instability.

B Supplementary Results for Real Data Analysis

B.1 Sample Size

Among these 728 patients, 178 were excluded from our analysis: 18 did not have genotype data (Affymetrix 6.0 array) from both tumor and paired normal samples, 35 failed Affymetrix genotype quality control (QC), 22 were male or of unknown gender, and 112 were non-Caucasian individuals, and 1 failed RNA-seq QC (Supplementary Figure S1).

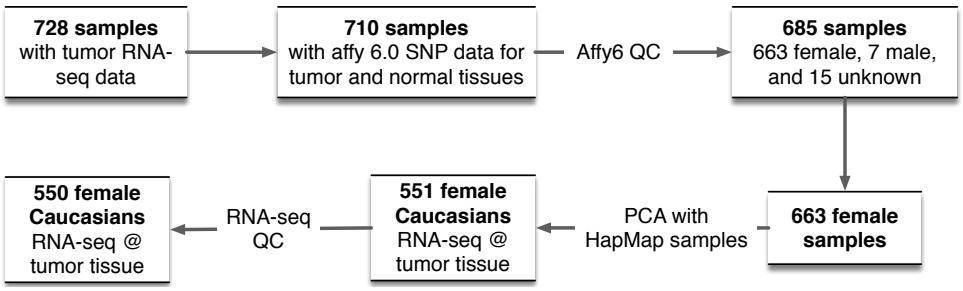

Figure S1: Sample size after each step of filtering.

B.2 Genotype Data Preparation

B.2.1 Genotype calling and quality control (QC)

We started our genotype data analysis with raw data in CEL files. After downloading all the CEL files of Affymetrix 6.0 arrays, we saved the file locations of these CEL files into file `cel_files_normal.txt` and ran the following APT (Affymetrix Power Tools) command to check genotype quality.

```
apt-geno-qc \
    --cdf-file  /path_to_lib_files/GenomeWideSNP_6.cdf \
    --qcc-file  /path_to_lib_files/GenomeWideSNP_6.r2.qcc \
```

```
--qca-file  /path_to_lib_files/GenomeWideSNP_6.r2.qca \
--cel-files /path_to_working_folder/cel_files_normal.txt \
--out-file  /path_to_working_folder/apt-geno-qc.txt
```

Low quality samples were determined via low contrast QC (contrast.qc ≤ 0.4) or low QC call rate (qc.call.rate.all ≤ 0.8) (Supplementary Figure S2).

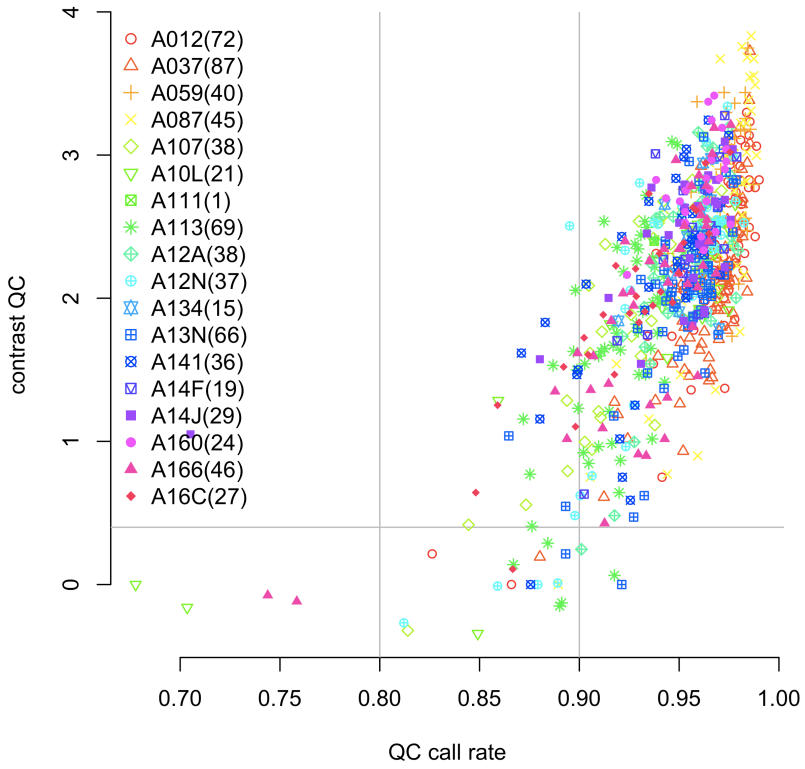

Figure S2: Results of genotype QC by APT. Each sample is labeled by the plate to which it belongs. The cutoff we use to select samples are QC call rate > 0.8 and contrast QC > 0.4.

After removing low quality samples, the new list of 685 remaining CEL files was recorded in file `cel_files_normal_after_qc.txt`. We called genotypes and genders for these 685 samples using birdseed-v2 implemented as part of APT.

```
apt-probeset-genotype \
  -o ../genotype_normal \
  -c  /path_to_lib_files/GenomeWideSNP_6.cdf \
```

```
--set-gender-method cn-probe-chrXY-ratio \
--chrX-probes /path_to_lib_files/GenomeWideSNP_6.chrXprobes \
--chrY-probes /path_to_lib_files/GenomeWideSNP_6.chrYprobes \
--special-snps /path_to_lib_files/GenomeWideSNP_6.specialSNPs \
--read-models-birdseed /path_to_lib_files/GenomeWideSNP_6.birdseed-v2.models \
-a birdseed-v2 \
--cel-files /path_to_working_folder/cel_files_normal_after_qc.txt
```

To determine sample ethnicity, we performed PCA using genotype from TCGA samples together with genotypes from HAPMAP CEU (Caucasian), YRI (African), and CHB (Asian) samples. The PC1 versus PC2 plot clearly separated CEU, YRI, and CHB samples, and the TCGA samples that were clustered with CEU samples in the PC1 versus PC2 plot were classified as Caucasian samples (Supplementary Figure S3).

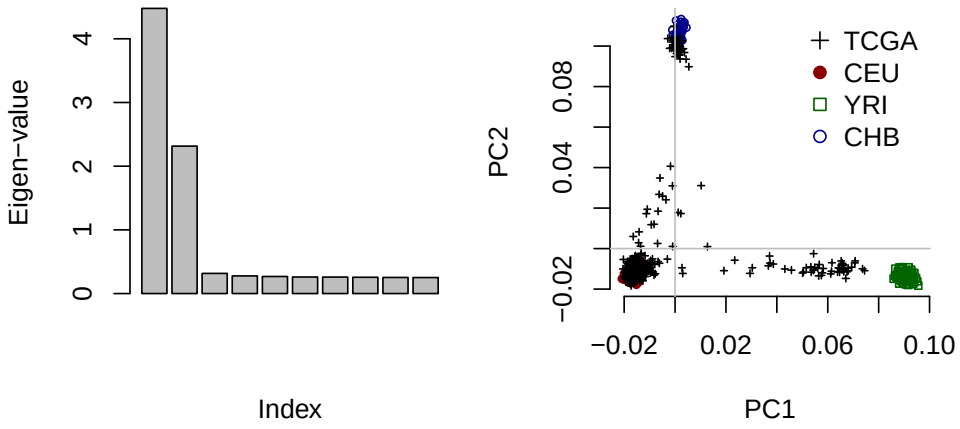

Figure S3: The left panel shows eigen-values of the PCA, and the right panel shows PC1 versus PC2 plot. Based on this plot, we choose the Caucasian samples as those with PC1 < 0 and PC2 < 0.

B.2.2 Genotype Imputation

We imputed genotype data for the 551 samples that passed all the genotype-related filters. The output of birdseed includes genotype calls for 909,622 SNPs. We removed those SNPs without chromosome location information or with more than 5% of missing values leaving 832,334 SNPs that passed these filters. We used MACH Li et al. [2010] (mach.1.0.18.Linux) to phase and impute the genotypes using the 1000 Genome Reference (∼36 million SNPs), which were downloaded from MACH website (`http://csg.sph.`

umich.edu/abecasis/MaCH/download/1000G.2012-02-14.html).

B.3 RNA-seq Data Preparation

We downloaded RNA-seq bam files from the TCGA data portal. First, we pre-processed these bam files using the R function `prepareBAM` of R package `asSeq` (`http://research.fhcrc.org/sun/en/software/asSeq.html`), to remove duplicated reads, or reads with average sequencing quality or mapping quality lower than 10. Next the expression of each gene in a sample is calculated as the number of RNA-seq reads that overlap with the exonic regions of this gene, obtained using R function `asSeq/countReads`. Annotations of exonic regions of each gene were obtained from Ensembl (version Homo_sapiens.GRCh37.66). Based on this version of gene annotation, we obtained read counts for 53,561 genes. Many of these genes have zero expression across most of the samples. We selected the 18,827 genes for which the 75 percentile of gene expression is equal or larger than 20. In other words, we remove those genes whose expression is less than 20 in more than 75% of samples.

To obtain allele-specific read counts for each sample, we first extracted all the heterozygous SNPs per sample, and then extracted those RNA-seq reads that overlap with at least one heterozygous SNP by R function `asSeq/extractAsReads`. Such RNA-seq reads were saved into three bam files, one for reads that match haplotype 1, one for those that match haplotype 2, and one for those with conflicts. For example, a conflicting read may overlap with more than one heterozygous SNPs, and its haplotype assignment is not consistent across these heterozygous SNPs. Usually the number of reads assigned to the conflict bam file is much smaller than the number of reads assigned to the two other bam files, otherwise it indicates errors in the data files or the data processing pipeline. Approximately 3.4% of the RNA-seq reads are classified as allele-specific reads (Supplementary Figure S4) across all 551 samples, with one apparent outlier (sample ID: A15R), which is la-

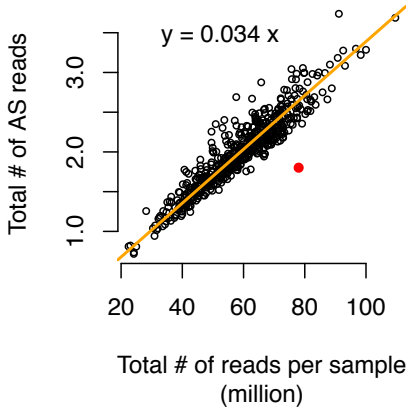

Figure S4: The total number of reads (across all genes) per sample versus the total number of allele-specific reads per sample. The red point indicates a sample (A15R) that has unexpected low proportion of allele-specific reads and it is excluded from further analysis.

beled as red in Supplementary Figure S4. We removed this sample in the following analysis.

For any association analysis using TReC per gene, one has to account for read-depth difference across samples. One way to quantify read-depth of a sample is to simply add up the total number of reads of this sample. Here we adopted a more robust approach, to quantify read-depth using 75 percentile of TReC across all the genes of a sample. In fact, in this data set, the two measurements of read depth are highly correlated (Supplementary Figure S5).

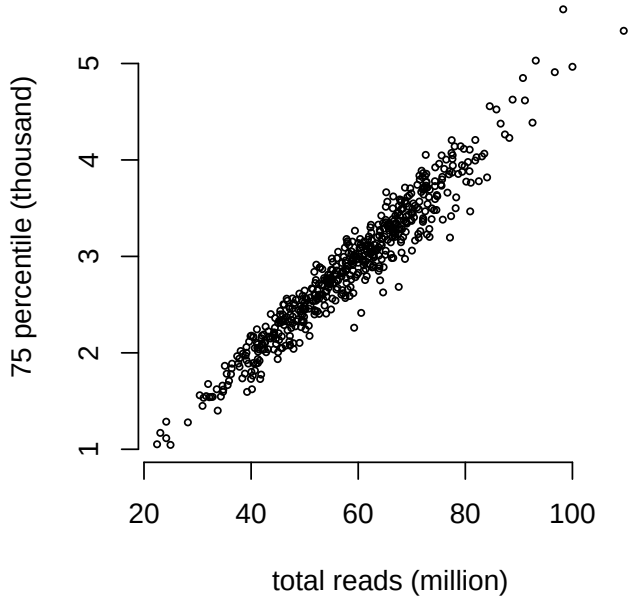

Figure S5: The total number of reads (across all genes) per sample versus the 75 percentile of the TReC of all the genes within a sample.

B.4 eQTL mapping results

We summarize the agreement and disagreement of each tested model for individual gene-SNP pairs and number of genes with at least 1 significant eQTL, respectively.

Gene-SNP Pairs				
P-value Cutoff	**Category**	**pTReC(ASE)**	**TReC(ASE)**	**pLR**
5×10^{-4}	# of gene-SNP pairs	133,599	436,021	48,717
	overlap/alternative	–	19.8%	69.4%
	overlap/pTReC(ASE)	–	64.6%	25.3%
5×10^{-6}	# of gene-SNP pairs	43,605	208,546	14,285
	overlap/alternative	–	16.8%	80.5%
	overlap/pTReC(ASE)	–	80.2%	26.4%
5×10^{-8}	# of gene-SNP pairs	19,867	131,795	6,593
	overlap/alternative	–	13.5%	78.5%
	overlap/pTReC(ASE)	–	89.8%	26.0%

Table S3: Summarizing the results of pTReC(ASE), TReC(ASE) and Westra models for TCGA data analysis. Here the notation pTReC(ASE) indicate that we use pTReCASE or pTReC model, depending on the results of Cis-Trans test. "overlap" represents the gene-SNP pairs identified by both pTReC(ASE) and an alternative method. "overlap/alternative" is the number of overlaps divided by the number of findings by the alternative method. "overlap/pTReC(ASE)" is the number of overlaps divided by the number of findings by pTReC(ASE). If we consider the results of pTReC(ASE) as true findings, then "overlap/alternative" is true discovery rate and "overlap/pTReC(ASE)" is sensitivity.

Genes				
P-value Cutoff	**Category**	**pTReC(ASE)**	**TReC(ASE)**	**pLR**
5×10^{-4}	# of Genes	4788	7793	2055
	overlap/alternative	–	42.7	70.2
	overlap/pTReC(ASE)	–	69.5	30.3
5×10^{-6}	# of Genes	1245	2982	268
	overlap/alternative	–	27.0	85.4
	overlap/pTReC(ASE)	–	64.7	18.4
5×10^{-8}	# of Genes	496	1612	110
	overlap/alternative	–	21.4	93.6
	overlap/pTReC(ASE)	–	69.6	20.8

Table S4: Summarizing the results of pTReC(ASE), TReC(ASE), the Westra models for TCGA data at gene level. The results are presented in the same format as Table S3, though the results are summarized at gene level instead of the level of SNP-gene pairs.

We have also compared the eQTLs identified by our study versus the eQTLs reported by another study of breast cancer patients [Li et al., 2013]. These two studies have used different list of genes and SNPs. We used gene expression measured by RNA-seq and filtered out genes with low expression. Li et al. [2013] used gene expression from microarray. We have used more SNPs but search for smaller region around each gene. Specifically, we used more than 6 million SNPs after imputation and filtering by MAF ≥ 0.02 and search for 100Kb around each gene, while Li et al. [2013] used around 800,000 SNPs but search 1Mb around each gene. Since two studies have used different gene and SNP list, it is not easy to make a precise comparison. Here we just assess how much the list of genes with significant local eQTLs overlap. Table S5 and S6 show the comparison results using TReCASE and pTReCASE, respectively. As expected, there are significant overlap in either case and the overlap is larger for TReCASE model because Li et al. [2013] did not account for tumor purity in their study.

		LI			
TReCASE	p_1 / p_2	$7.5e-5$	$5e-6$	$5e-8$	**# Genes T**
	$5e-6$	343 ($6.5e-20$)	179 ($4.6e-25$)	101 ($5.3e-25$)	2982
	$5e-8$	251 ($6.9e-33$)	148 ($4.2e-40$)	85 ($8.5e-34$)	1612
	# Genes Li	1325	513	219	

Table S5: Summary of the overlap of eGenes (i.e., the genes with at least one local eQTL) reported by our study and an earlier study by Li et al. (2013) Li et al. [2013]. For example, at p-value cutoff $5e-8$, TReCASE identified eQTLs for 1,612 genes and Li et al. (2013) Li et al. [2013] identified eQTLs for 219 genes. The overlap is 85 genes. This overlap is much larger than expected by chance (p-value 8.5e-34).

		LI			
pTReCASE	p_1 / p_2	$7.5e-5$	$5e-6$	$5e-8$	**# Genes T**
	$5e-6$	165 ($1.6e-14$)	102 ($5.0e-23$)	63 ($2.4e-23$)	1245
	$5e-8$	86 ($3.5e-14$)	61 ($1.9e-22$)	45 ($1.1e-26$)	496
	# Genes Li	1325	513	219	

Table S6: Similar to Table S5, but here the comparison is between the results of pTReCASE and Li et al. (2013) Li et al. [2013].

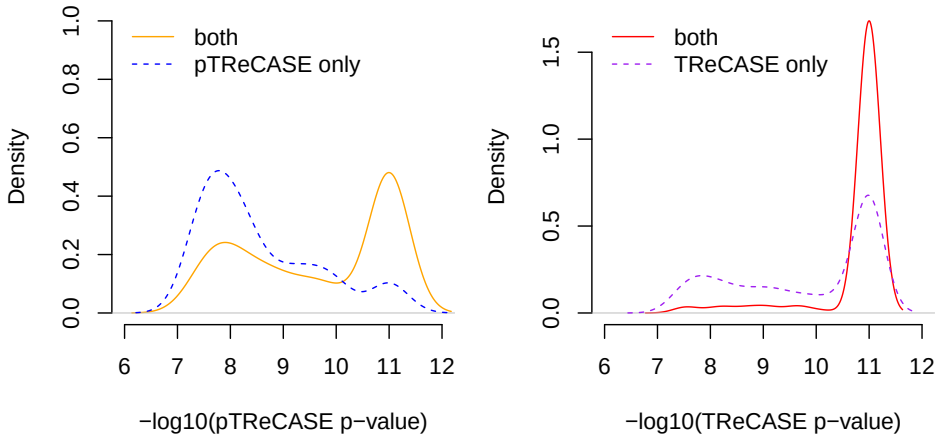

Figure S6: The left panel shows the distribution of -$\log_{10}$(p-value) from pTReCASE model for those 345 genes with eQTLs by both TReCASE and pTReCASE methods (p-value $< 5 \times 10^{-8}$, orange solid curve), and those 151 genes with pTReCASE p-value $< 5 \times 10^{-8}$ and TReCASE p-value $\geq 5 \times 10^{-8}$ (blue dotted line). The right panel shows the distribution of -$\log_{10}$(p-value) from TReCASE model for those 345 genes with eQTLs by both TReCASE and pTReCASE methods (p-value $< 5 \times 10^{-8}$, red solid curve), and those 1267 genes with TReCASE p-value $< 5 \times 10^{-8}$ and pTReCASE p-value $\geq 5 \times 10^{-8}$ (purple dotted line).

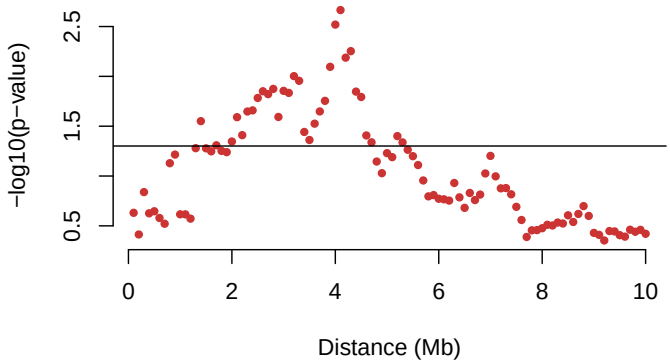

Figure S7: -$\log_{10}$(p-value) for Chi-squared test whether the three categories of eQTL SNPs (with eQTL p-value $< 5 \times 10^{-8}$ in TReCASE and/or pTReCASE model have equal probability to be located within certain distance (X-axis of this plot) of breast cancer GWAS hits. We downloaded the breast cancer GWAS results by Michailidou et al. [2017] from GWAS catalog (`https://www.ebi.ac.uk/gwas/`). Among all the 813 GWAS hits, 795 have location information and can be liftOver to hg19. We used 469 of these 795 GWAS hits with p-value $< 5 \times 10^{-7}$ for our test. Results are consistent when using 371 of the 795 GWAS hits with p-value $< 5 \times 10^{-8}$.

Gene Symbol	Gene Name	Entrez Id	Genome Location
ETV1	ets variant gene 1	2115	7:13895866-13989052
GAS7	growth arrest-specific 7	8522	17:9917228-10198390
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor for (CD32)	2213	1:161663242-161677553
HLF	hepatic leukemia factor	3131	17:55265486-55320880
SGK1	serum/glucocorticoid regulated kinase 1	6446	6:134170268-134174807
USP44	ubiquitin specific peptidase 44	84101	12:95518154-95534256
TCF7L2	transcription factor 7-like 2	6934	10:112950757-113165972

Table S7: Seven eGenes with eQTLs identified by pTReCASE but not by TReCASE (at p-value cutoff 5×10^{-8}).

Gene Symbol	Name	Entrez Id	Genome Location
PAX8	paired box gene 8	7849	2:113218533-113278394
CYP2C8	cytochrome P450 family 2 subfamily C member 8	1558	10:95037128-95069402
RET	ret proto-oncogene	5979	10:43077259-43128269
CNTNAP2	contactin associated protein like 2	26047	7:146116877-148415616

Table S8: Four eGenes with eQTLs identified by both pTReCASE and TReCASE (at p-value cutoff 5×10^{-8}).

Gene Symbol	Name	Entrez Id	Genome Location
FLT4	fms-related tyrosine kinase 4	2324	5:180608281-180649545
CLTCL1	clathrin, heavy polypeptide-like 1	8218	22:19180219-19291641
FSTL3	follistatin-like 3 (secreted glycoprotein)	10272	19:676425-681709
PTK6	protein tyrosine kinase 6	5753	20:63529536-63537314
NDRG1	N-myc downstream regulated 1	10397	8:133238878-133284311
POU2AF1	POU domain, class 2, associating factor 1 (OBF1)	5450	11:111354261-111379177
PDGFRB	platelet-derived growth factor receptor, beta polypeptide	5159	5:150115763-150137047
PRRX1	paired related homeobox 1	5396	1:170664220-170730332
CEP89	centrosomal protein 89kDa	84902	19:32879162-32971874
MYH11	myosin, heavy polypeptide 11, smooth muscle	4629	16:15708811-15838252
CBLC	Cas-Br-M (murine) ecotropic retroviral transforming sequence c	23624	19:44777932-44800443
NRG1	neuregulin 1	3084	8:32548727-32764402
H3F3A	H3 histone, family 3A	3020	1:226064352-226071479
CASP3	caspase 3	836	4:184629272-184638453
IL7R	interleukin 7 receptor	3575	5:35856978-35876486
MGMT	O-6-methylguanine-DNA methyltransferase	4255	10:129536253-129766997
RMI2	RecQ mediated genome instability 2	116028	16:11345472-11350790
CD28	CD28 molecule	940	2:203706697-203734912
PRF1	perforin 1 (pore forming protein)	5551	10:70598053-70600902
BCR	breakpoint cluster region	613	22:23180961-23315522
MITF	melanogenesis-associated transcription factor	4286	3:69936723-69965248
CARD11	caspase recruitment domain family, member 11	84433	7:2906638-2958506
POU5F1	POU domain, class 5, transcription factor 1	5460	6:31164601-31170620
HLA-A	major histocompatibility complex, class I, A	3105	6:29942554-29945455

Table S9: Twenty-four eGenes with eQTLs identified by TReCASE but not by pTReCASE (at p-value cutoff 5×10^{-8}).

B.5 Additional results for potential confounding due to copy number alteration or DNA methylation

Using the omic data prepared by Sun et al. [2018], we examined the correlation between gene expression before and after removing copy number effects. Such correlations are very high for most of the genes. For example, it is larger than 0.8 for 86% of 15,284 genes with both gene expression and copy number data.

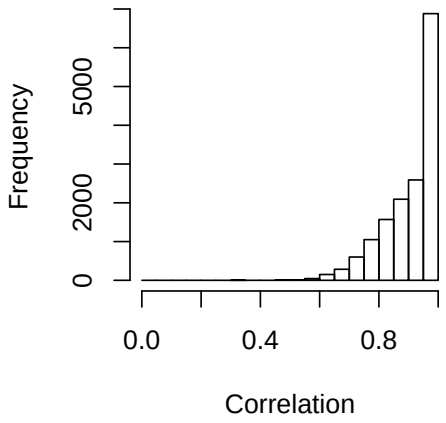

Figure S8: The distribution of correlations between gene expression before and after removing copy number effects using a linear regression. For the i-th gene, we have its expression in n samples before and after removing copy number effects. Denote these two vectors as x_{i1} and x_{i2}, we calculated the correlation between x_{i1} and x_{i2}. Then the histogram is generated from such correlations across all genes

We also checked whether copy number of DNA methylation may confound the eQTLs reported in Figure 2 of the main paper. The first example is about gene ENSG00000115525 (ST3GAL5). Its expression is not associated with its copy number (p-value 0.22, $R^2 = 0.039$), but is associated with the methylation level of two CpG's: cg10017626 (p-value 6.2e-05, $R^2 = 0.039$) and cg07214715 (p-value 2.8e-05, $R^2 = 0.043$) after correcting for tumor purity and cell type compositions [Sun et al., 2018]. The second example is about gene ENSG00000142794 (NBPF3). Its expression is not associated with DNA methylation but is associated with its copy number (p-value 1.3e-07, $R^2 = 0.067$). These associations are illustrated in Figure S10.

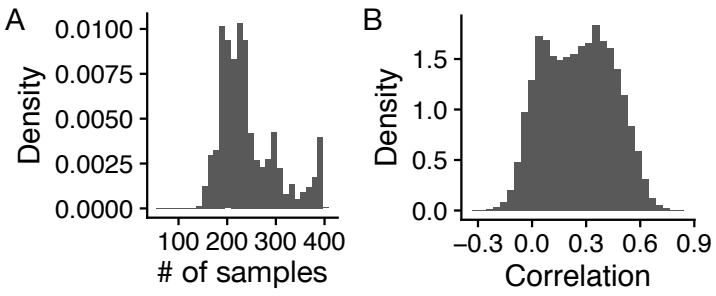

Figure S9: (A) The distribution of the number of samples with copy number events (i.e., with $|D_{ij}| > 0.5$ across all 18,134 genes. Note that $D_{ij} = C_{ij} - N_i$, where C_{ij} is the total copy number of gene j in sample i, and N_i is the ploidy of the i-th sample. (B) We measure the copy number changes using G_{ij}, which equals to -1, 0, or 1 if $D_{ij} < -0.5$, $|D_{ij}| \leq 0.5$, or $D_{ij} > 0.5$, respectively. This figure shows the distribution of the correlations between G_{ij} and relative gene expression summarized across all 18,134 genes.

Next we check whether the associations between eQTL SNP genotype and gene expression are affected after controlling DNA methylation of gene expression measurement (Figure S11). We conducted this analysis in 328 samples (a subset of the 550 samples in main analysis) with all the data needed: SNP genotype, copy number, gene expression, and DNA methylation. Using a simple linear regression of gene expression versus SNP genotype (without using allele-specific expression), the eQTL p-value for ST3GAL5 is 2.3e-4, and after controlling for methylation, the p-values remain similar (1.8e-4 for cg10017626 and 5.4e-4 for cg07214715). The eQTL p-value for NBPF3 is also similar before and after controlling for copy number (t-statistics being 9.309 and 9.295 before and after controlling for copy number and p-value $<$ 2e-16 in both cases).

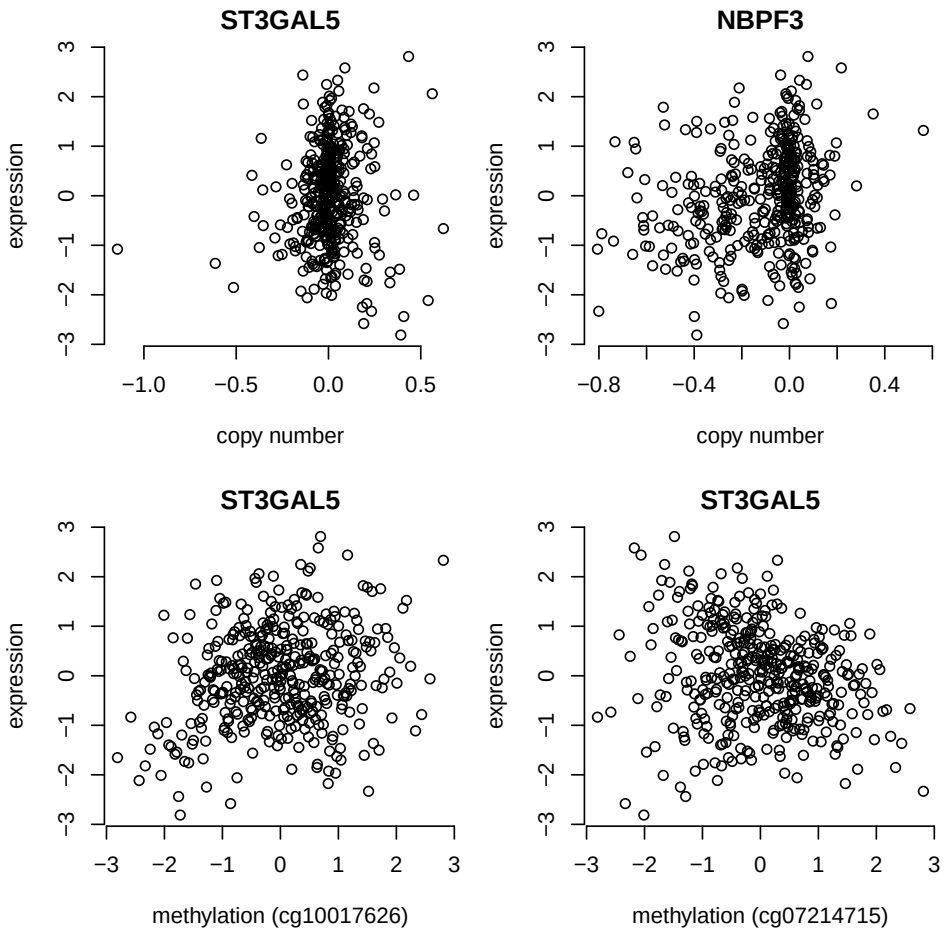

Figure S10: Scatter plots demonstrate the associations between gene expression and copy number of two genes ST3GAL5 and NBPF3 (upper panel), and the associations between gene expression of ST3GAL5 and DNA methylation of two CpG's.

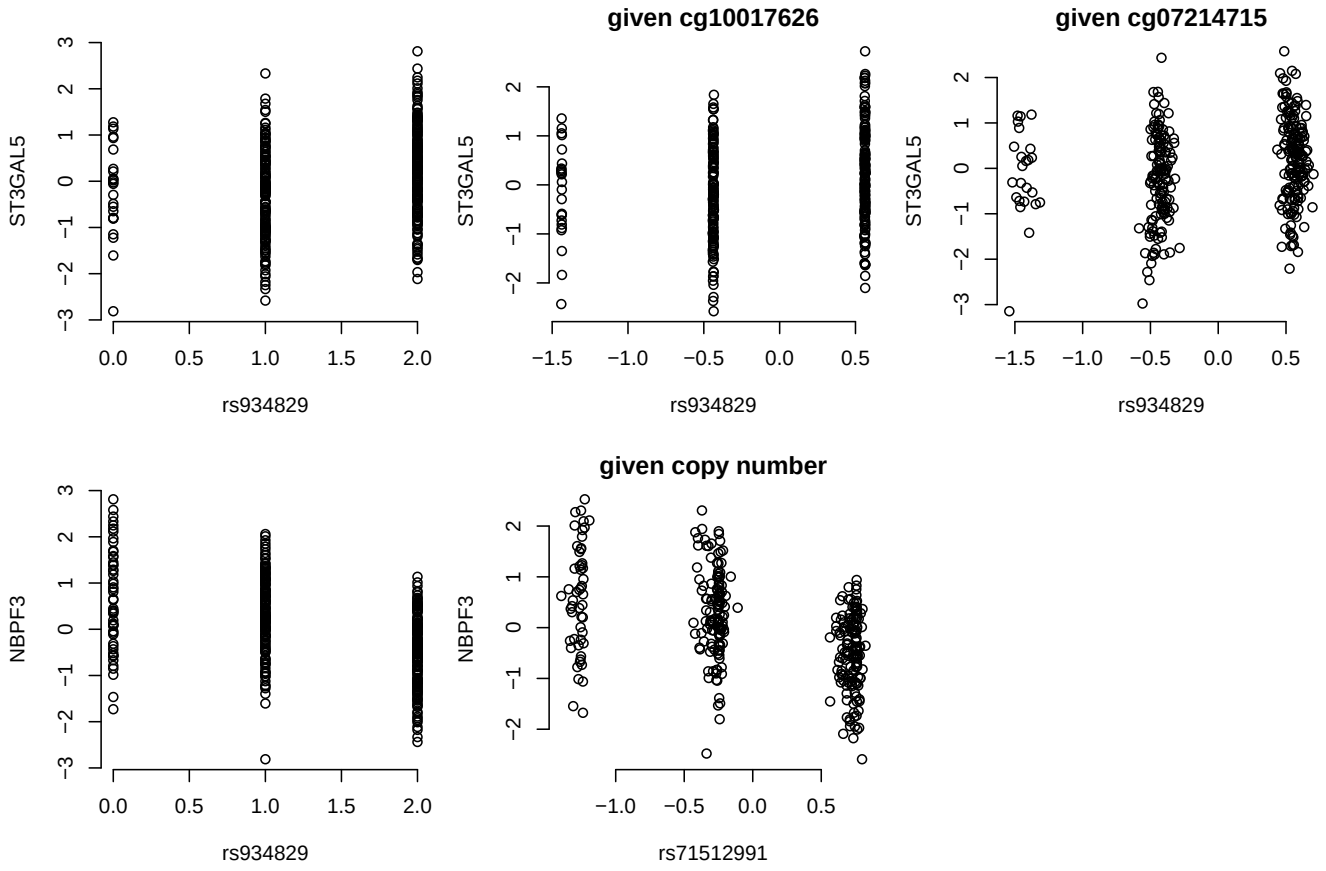

Figure S11: Scatter plots demonstrate the associations between eQTL and gene expression before or after conditioning on two CpG's for gene ST3GAL5 (upper panel), and associations between eQTL and gene expression before or after conditioning on copy number alteration for gene NBPF3 (lower panel).

References

David A. Freedman. How can the score test be inconsistent? *The American Statistician*, 61(4):291–295, 2007. ISSN 00031305. URL `http://www.jstor.org/stable/27643926`.

Yi-Juan Hu, Wei Sun, Jung-Ying Tzeng, and Charles M. Perou. Proper use of allele-specific expression improves statistical power for cis-eQTL mapping with RNA-seq data. *Journal of the American Statistical Association*, 110(511):962–974, 3 2015. doi: 10.1080/01621459.2015.1038449.

Qiyuan Li, Ji-Heui Seo, Barbara Stranger, Aaron McKenna, Itsik Pe'Er, Thomas LaFramboise, Myles Brown, Svitlana Tyekucheva, and Matthew L. Freedman. Integrative eQTL-based analyses reveal candidate causal genes and loci across five tumor types. *Cell*, 152 (3):633–641, 2013. doi: 10.1016/j.cell.2012.12.034.

Yun Li, Cristen J Willer, Jun Ding, Paul Scheet, and Gonçalo R Abecasis. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic epidemiology*, 34(8):816–834, 2010.

Kyriaki Michailidou, Sara Lindström, Joe Dennis, Jonathan Beesley, Shirley Hui, Siddhartha Kar, Audrey Lemaçon, Penny Soucy, Dylan Glubb, Asha Rostamianfar, et al. Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678):92, 2017.

C. Radhakrishna Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1):5057, 1948. doi: 10.1017/S0305004100023987.

Wei Sun. A statistical framework for eQTL mapping using RNA-seq data. *Biometrics*, 68 (1):1–11, 12 2011. doi: 10.1111/j.1541-0420.2011.01654.x.

Wei Sun, Paul Bunn, Chong Jin, Paul Little, Vasyl Zhabotynsky, Charles M Perou, David Neil Hayes, Mengjie Chen, and Dan-Yu Lin. The association between copy number aberration, dna methylation and gene expression in tumor samples. *Nucleic acids research*, 46(6):3009–3018, 2018.