

Navigating the GTEx eQTL database and using TaqMan Assays to verify eQTL links in disease research

Highlights

In this white paper, we:

- Describe the importance of expression quantitative trait loci (eQTLs)
- Demonstrate ways to use the Genotype-Tissue Expression (GTEx) database to find eQTLs affecting the expression of various genes in various tissues
- Illustrate how to find Applied Biosystems™ TaqMan® Assays that can be used to verify and extend genome-wide association study (GWAS) and eQTL analysis results

Introduction

In the past 15 years, tremendous progress has been made using GWAS to link genetic variation with phenotypes and pathologies [1,2]. According to the most recent data, over 76,000 genetic variants have been associated with over 2,700 traits or diseases [3]. In spite of this success, it has

been estimated that as many as 90% of single-nucleotide polymorphisms (SNPs) identified in GWAS map to noncoding regions (see “Example 4” below), complicating the mechanistic interpretation of the results. It is thought that these noncoding SNPs fall into regulatory regions and influence the expression of a gene or genes. The study of eQTLs provides a method for understanding the link between genetic variants and altered gene expression, and could potentially provide new insights connecting GWAS results to molecular mechanisms of various diseases.

eQTLs are identified by linking variations in transcript abundance with variations in genotypes (Figure 1A). Typically, this involves collecting transcriptomic and genomic variant data from samples in a large cohort, and correlating the difference in gene mRNA levels with SNP genotypes in individuals. eQTLs can fall into the following two categories: *cis*-eQTLs are SNPs that are proximal to a target locus (Figure 1B). Proximity is

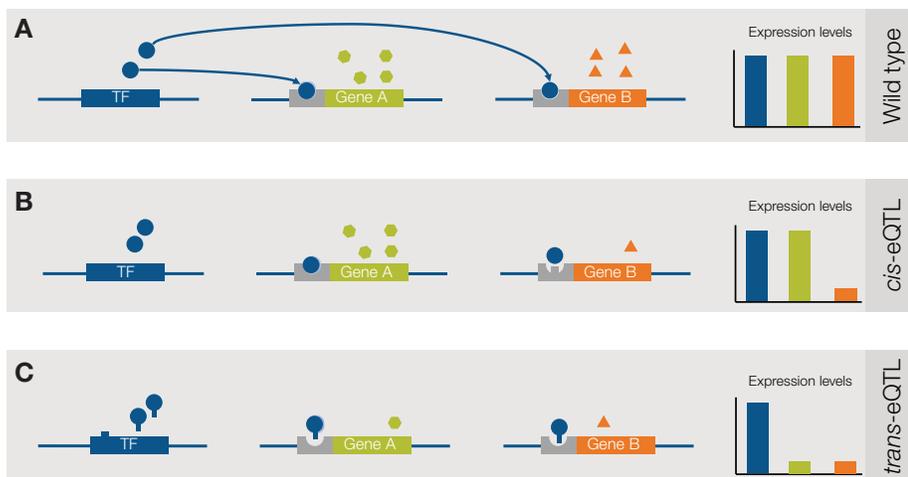


Figure 1. How eQTLs affect expression levels of genes. (A) In a wild type individual, a transcription factor (TF) binds to the controlling region of two genes, resulting in normal transcript levels. **(B)** An individual with a *cis*-eQTL can have a SNP in the region bound by the transcription factor that reduces (but may not eliminate) transcript levels of gene B while leaving levels of gene A unaffected. **(C)** In an individual with a *trans*-eQTL, a SNP in the DNA-binding motif of the transcription factor reduces (but may not eliminate) its affinity to all binding sites, resulting in reduced expression of multiple genes.

variably defined, but typically *cis*-eQTLs fall within 200 kb–1 Mb of the transcriptional start site of a gene. Thus, these are usually thought of as SNPs that fall within the promoter, enhancer, or other transcriptional regulatory sequences. The other type of eQTL, *trans*-eQTLs, are usually genetically unlinked and are thought to regulate expression of a locus from a distance. For example, a *trans*-eQTL could be a SNP in a transcriptional activating protein that lowers, but does not eliminate, the DNA-binding

affinity of the protein to all binding sites (Figure 1C). eQTLs can also be tissue specific—a SNP that affects the expression of a gene in one tissue may have no effect in a different tissue. eQTL data are often deposited into publicly accessible databases. One of these, arising from the GTEx project [5], is rapidly becoming the gold-standard repository for eQTL data.

To be useful in contributing to disease research, SNPs that are eQTLs need to be linked to GWAS of phenotypes or pathologies. Several tools have been developed that look for overlaps of eQTLs and GWAS-identified SNPs. For example, the Sherlock software tool provides a statistical basis for finding SNPs common to GWAS and eQTL studies [6]. A link between the GWAS-identified SNP locus and eQTL suggests that differential expression of the gene affected by the eQTL could contribute to the trait. Once a putative link has been established, verification of the link is needed. This could be done by screening the transcriptome and genome of even more samples; however, this could become unwieldy and expensive. Alternatively, an orthogonal technology, such

as qPCR, could be used to confirm the association of the SNP genotype and mRNA levels of the affected gene in a targeted manner. This method has the potential to deliver information on targeted associations in a cost-effective manner.

In this document, we provide short descriptions of how to navigate the GTEx database. We use the GTEx database to introduce some example case studies that illustrate how qPCR using TaqMan Assays can be used to verify GWAS and eQTL analysis results.

Navigating the GTEx database

The GTEx database is accessed through the GTEx portal URL (gtexportal.org, Figure 2). From this web page, information about the tissues collected and transcriptomes obtained from the tissues can be accessed. SNP information can be found through the center panel “Genetic Association”. Five principle methods to query the database for SNP functions are accessible here and are exemplified below. For full details, please see the tutorial links and supporting pages on the GTEx portal page.

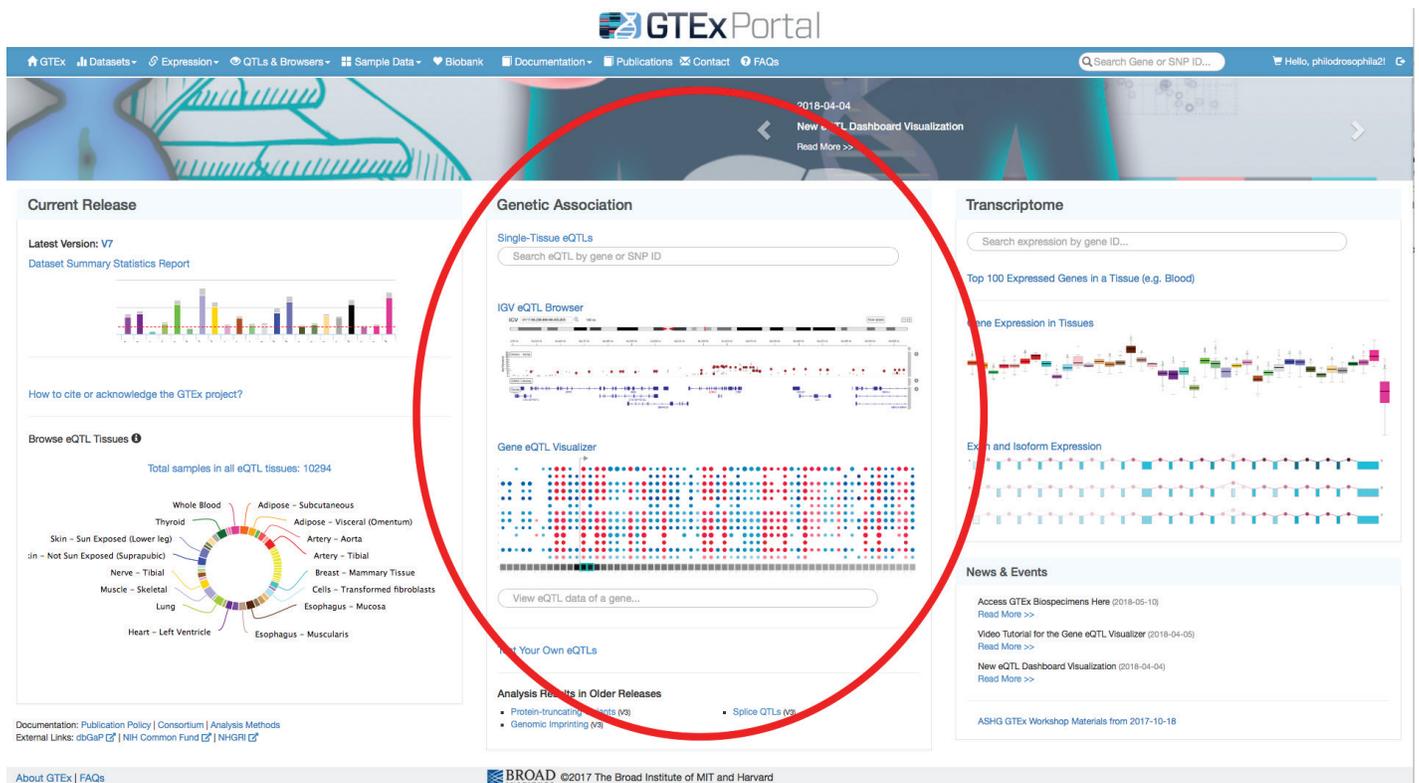


Figure 2. The GTEx website. The portal web page (gtexportal.org/home/) provides links to query the database in three general areas: details on donor samples and usage is provided on the leftmost side (“Current Release”), information on tissue-specific expression of genes and alternate transcripts is provided on the rightmost side (“Transcriptome”, not covered in this white paper), and the information on eQTLs is provided in the center of the page (“Genetic Association”). This white paper details the usage of the links in the “Genetic Association” section.

Example 1: Finding eQTLs associated with a gene

First, eQTLs can be queried by entering a gene name in the “Single-Tissue eQTLs” field at the top of the panel. For example, entering *XRRRA1* (for the X-ray radiation resistance associated-1 gene) returns the page shown in Figure 3. Here, all the eQTLs associated with this gene in all tissues can be accessed. However, note that this is a particularly eQTL-rich gene, with over 1,500 pages of information. The information can be filtered to a specific tissue (or tissues) through the “Change Tissue” button at the top of the page. The reference SNP cluster ID (rsID) associated with each eQTL is shown in the SNP column and links to the dbSNP database. The *P* value and normalized effect size (NES) are displayed in their respective columns. In addition, links to graphical displays (box plot, the Integrative Genomics Viewer (IGV) browser described below, and a representation across multiple tissues) are shown.

You can find the Applied Biosystems™ TaqMan® SNP Genotyping Assay that corresponds to each SNP from the “Search Applied Biosystems TaqMan Assays” page. Simply choose from the “Search Applied Biosystems TaqMan

Assays” page (thermofisher.com/taqman), simply choose “SNP Genotyping” from the assay type box, enter the rsID number for the SNP of interest, and click the magnifying glass icon. A page with the TaqMan Assay will appear; from here, the assay can be added to the cart. Multiple assays for different eQTLs can also be easily ordered: From the GTEx page, click the “CSV” button to export the eQTL table in readable Microsoft™ Excel™ format. The CSV document can then be opened in Excel software and the column with all the rsIDs can be copied and pasted in the rsID entry box on the TaqMan ordering page. The search results will appear as a list of assays that can then be added individually or as a group to the cart. The TaqMan SNP Genotyping Assays might display as mapping to different genes than the GTEx database; this does not mean that the SNP associated with the rsID is not an eQTL for your gene of interest. In instances where no existing TaqMan Assay exists for a SNP, a custom genotyping assay can be designed (go to the Custom TaqMan Assay Design Tool at thermofisher.com/cadt). Once obtained, these TaqMan Assays can be used to confirm or determine the genotypes of other samples.

The screenshot shows the GTEx website interface. At the top, it says "Data Source: GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2)". Below that is a search bar for "All eQTL Tissues (n >= 70)" and a "Change Tissue" button. The "Tissue Information" section shows a table with the following data:

	Counts
Total samples	11614
Total samples with donor genotype	10294
Number of significant eGenes	341316 (31403 unique genes)

The "Single-Tissue eQTLs in" section has a search bar containing "rs9787897" and a "Search significant eQTL By ID" button. Below this is a table of eQTLs for rs9787897. The table has columns for Gene Symbol, Variant Id, SNP, P-Value, NES, and Tissue. The first few rows are:

Gene Symbol	Variant Id	SNP	P-Value	NES	Tissue
XRRRA1	11_74659302_A_T_b37	rs9787897	9.0e-173	1.4	Muscle - Skeletal
XRRRA1	11_74659302_A_T_b37	rs9787897	3.8e-167	1.3	Thyroid
XRRRA1	11_74659302_A_T_b37	rs9787897	3.4e-147	1.4	Skin - Sun Exposed (Lower leg)

An orange oval highlights the "SNP" column. An arrow points from the "rs9787897" entry in the table to a search bar on the "Search Applied Biosystems TaqMan Assays" page. The search bar contains "rs9787897" and a magnifying glass icon. Below the search bar, there is a dropdown menu for "TaqMan Assay" with "SNP Genotyping" selected. The search results show a table with columns for "Functionality tested", "Gene expression analysis", "SNP ID", "Chr", "Position", "Transition Substitution", "Functionality", "Size", "Availability", and "Price". The first row shows:

Functionality tested	Gene expression analysis	SNP ID	Chr	Position	Transition Substitution	Functionality	Size	Availability	Price
✓	✓	C_3240139_10	XRRRA1	Chr11: 7468888 on GPO38	Transition Substitution	Introm, Intragenic	300 nms	Made to order	255.00

Figure 3. Identifying eQTLs associated with a gene. Entering a gene name in the “Single-Tissue eQTLs” field on the landing page returns a list of SNPs that are eQTLs in the database. eQTLs for *XRRRA1* are shown in this example. For more details on navigating the other links on the page, see the text or the GTEx tutorial. To find the TaqMan Assays for these SNPs, simply enter the rsID number into the TaqMan query form (thermofisher.com/taqman).

Example 2: Identifying genes that are associated with a putative eQTL

Second, a SNP identified in a GWAS or other study can be analyzed whether or not it is an eQTL, and linked genes for which the SNP has eQTL data can be found. For example, assume that a study identifies the SNP rs9787897 as genetically linked to a trait. The eQTL database can be queried by entering the rsID (rs9787897) in the “Single-Tissue eQTLs” field at the top of the panel as described above. A page similar to Figure 3 will appear, highlighting the tissue(s) where the SNP has been shown to be an eQTL, the *P* value, and the NES. In addition, genes with which the eQTL has been shown to have a significant association are also shown in the “Gene Symbol” column. As above, clicking on any of the hyperlinks will provide more visual information or link out to other pages. Unfortunately, the GTEx database does not support batch queries of SNPs at this time.

However, if the SNP has not been mapped to an eQTL, then the page will state “No significant eQTLs were found for SNP rsXXXXX in tissue.” This does not mean the SNP could not be an eQTL. It only means that an eQTL has not been found to act as one in the tissues that are in the GTEx database.

To find the right TaqMan genotyping assays corresponding to the SNP, simply enter the rsID in the TaqMan Assay ordering page as described above. Additionally, the goal of the study might be to examine expression levels of the gene(s) associated with the eQTL. Once the genes that will be tested have been identified, ordering the Applied Biosystem™ TaqMan® Gene Expression Assays is very easy: enter the gene symbol in the query box, choose “Gene Expression” from the TaqMan Assays box, and click the magnifying glass. Note that to order a TaqMan Gene Expression Assay, the gene symbol and not the rsID must be used. A page will appear that lists all available gene expression assays for the gene, with the assay that has the best coverage appearing at the top of the list. From here, the assays can be added to the cart and ordered when convenient. These assays can be used to further analyze mRNA expression characteristics of a gene that might be linked to an eQTL.

Example 3: Making specific association queries

The GTEx portal allows queries for specific SNP, gene, and tissue associations to be addressed. To make these queries, click on the “GTEx eQTL Calculator” link at the

bottom of the “Genetic Association” section of the portal. The resulting page will have a box that will allow you to enter an rsID, gene name/symbol, and tissue type. For example, entering “rs9787897, *XRR1*, Whole_Blood” will confirm that the SNP is indeed an eQTL for *XRR1* gene in whole blood with a *P* value of 9.2×10^{-118} , a very significant result (Figure 4A). On the other hand, entering “rs9787897, *RPS26*, Whole_Blood” returns data, meaning that there is information for the SNP, but the *P* value is 0.41, and the NES score is very low (Figure 4B). Thus, rs9787897 is not a significant eQTL for *RPS26* in blood. If the SNP is not in the database, the page will clearly state this. Note that, in order to make these types of queries, the tissue name entered in the box must match exactly to the text in the database, as outlined in “allowable tissue names”. Once a positive association is made, studies can be extended by analyzing more samples using TaqMan SNP Genotyping and Gene Expression Assays.

Example 4: Visualizing all the eQTLs associated with a gene

Sometimes it is useful to examine all known eQTLs associated with a gene. The GTEx portal provides a way to do this, using the “GTEx Gene-eQTL Visualizer” section of the portal. To use this feature, enter the gene name in the query box. The page returned is a display rich with visual information. The top of the page shows basic information on the gene and various filters that help trim the results (Figure 5). Below that, a high-level representation of the gene is shown, where red dots/bubbles indicate the position of eQTLs that have increased expression relative to the reference, and blue dots have decreased expression. The size of the dot and the shade of the color indicate the significance and magnitude of the SNP.

For more details, the area contained within the gray box in the high-level diagram is magnified and shown below it (Figure 5). In this view, the y-axis shows the different tissue types for which eQTL information exists. On the x-axis, the eQTL in GTEx ID format is shown. Note that in this depiction, the distance between SNPs is not measured to scale. The gray box can be dragged along the gene to highlight different regions. Similar to the high-level view, the size and shade of the dot in the magnified view indicate the significance and magnitude of the eQTL, respectively. Hovering the mouse over a dot (Figure 5, yellow circle) will bring up specific information, including the rsID, tissue, and NES in that tissue. Clicking on a dot will bring up a box plot of the expression in different genotypes in that

tissue. Along the bottom of the figure, between the dots and the alphanumeric eQTL labels, a gray shaded square indicates the relative distance to the transcription start site (TSS). The darker the gray shade, the closer the eQTL is to the TSS. Hovering the mouse over one of these boxes (bottom of Figure 5, yellow box) will overlay more information for the SNP, including the rsID, position relative to TSS in nucleotides, and genomic coordinates. The TSS is also shown at the top of the bubble diagram, with an arrow indicating the direction of transcription (if the TSS is included in the gray magnification box).

Studies further examining the link between SNPs and gene expression can be performed using the appropriate TaqMan SNP Genotyping and Gene Expression Assays. It is important to note that the TaqMan SNP Genotyping Assays must be searched and ordered using the rsID (rs190222032) and not the GTEx ID (e.g., 11_73806478_G_A_b37), as displayed in the main view.

Example 5: Visualizing the eQTLs in the context of the genome

Finally, the GTEx portal can also provide a view of eQTLs in a genomic context. This interface makes use of some of the features in the IGV developed at the Broad Institute [7]. Clicking on “GTEx IGV eQTL Browser” will bring up a default view of the first chromosome’s first 500,000 nucleotides. Immediately below the drawing of the chromosome are light green boxes showing the location of SNPs that have been associated with GWAS (GWAS

Catalog space). Clicking on one of these boxes will bring up information on the strongest SNP-risk allele, disease, *P* value, odds ratio, and PubMed ID for the study. Below the GWAS Catalog space are intron/exon diagrams of genes covered by the view. A track showing eQTLs can be added by choosing a tissue (or tissues) via “Click Here To Select Tissue Tracks” at the top left of the viewer. The view can be refined by zooming in and out, or by entering genomic coordinates or specific gene names in the search box. For example, entering “chr11:74,000,000–75,000,000” in the search box brings up the region around the *XRR1* locus (Figure 6). eQTLs that have been mapped to the region are shown as gray dots, giving a comprehensive view of the region. Clicking on any of the individual dots will bring up information about that eQTL. Note that in this region, there are data from only one GWAS—mapping a SNP to the *POLD3* gene. If the search is done by entering a gene symbol or rsID, then eQTLs associated with the gene or SNP are shown as red dots instead of gray dots.

As described in the other examples, studies further examining the link between SNPs of interest and expression of affected genes can be performed using the appropriate TaqMan SNP Genotyping Assays and Gene Expression Assays. From this view, it is easy to see the genes that may be of interest for further examination. rsIDs can be obtained by clicking on the dot of interest. Again, the TaqMan SNP Genotyping Assays can only be searched and ordered using the rsID (rs190222032), not the GTEx ID (e.g., 11_73806478_G_A_b37).

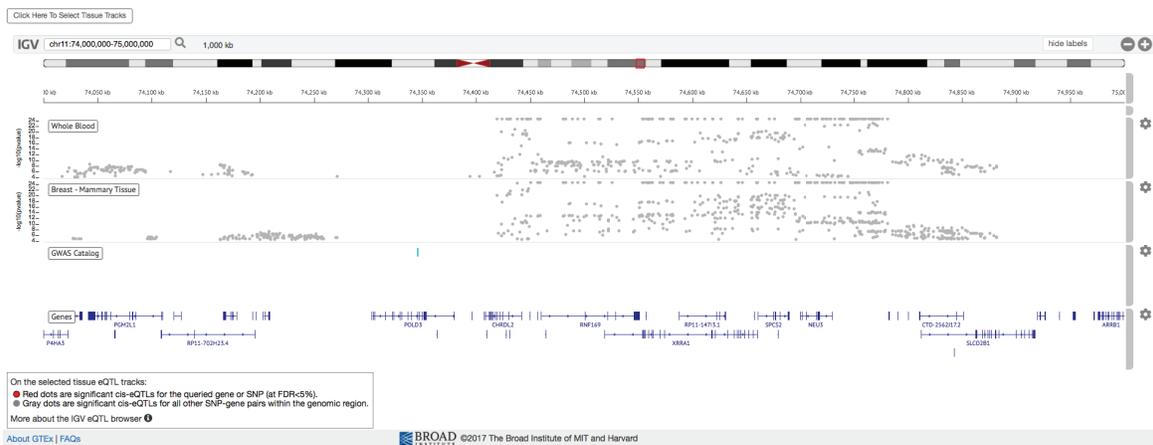


Figure 6. View of eQTLs using the IGV browser interface. The location of eQTLs for particular tissues can also be visualized using the IGV interface. Tracks showing eQTLs for specific tissues are chosen by clicking on the “Click Here to Select Tissue Tracks” button and selecting the desired tissue type. Shown here are eQTLs in blood and mammary tissue that fall within the genomic coordinates entered (ch11:74,000,000–75,000,000). Clicking on a gray dot (or red if a specific gene is entered in the query box) displays details about that SNP. In addition to gene structures that map to this region (genes track), data from GWAS studies that fall into the same region are shown by light green bars on the “GWAS Catalog” track.

Case example—*SERPINH1* and breast cancer

We illustrate how these tools might be employed by making use of a hypothetical example. Browsing using the IGV interface around the region of chr11:75,198,101–75,358,828, we can see that there are GWAS results that map to the *SERPINH1* gene (Figure 7A). In addition, looking at eQTLs in whole blood and breast mammary tissue, we see there are three SNPs that are eQTLs and they fall in the same region. Note that these three eQTLs do not appear to affect expression in whole blood. Entering *SERPINH1* into the query box and choosing whole blood and breast tissue, we find there are indeed three eQTLs for this gene in breast mammary tissue (Figure 7B). Clicking on each of the three

dots brings up box plots, showing the effect of different genotypes on the expression: one of them increases expression relative to the reference, and two decrease expression (Figure 7C). A literature search suggests that changes in *SERPINH1* (also known as *HSP47*) can influence cancer progression by increasing extracellular matrix deposition [8]. Thermo Fisher Scientific has TaqMan Assays available for *SERPINH1* gene expression and for two of the SNPs that are eQTLs for this gene in breast mammary tissue (Figure 7D). A custom TaqMan SNP Genotyping Assay can be designed and ordered for the third target.

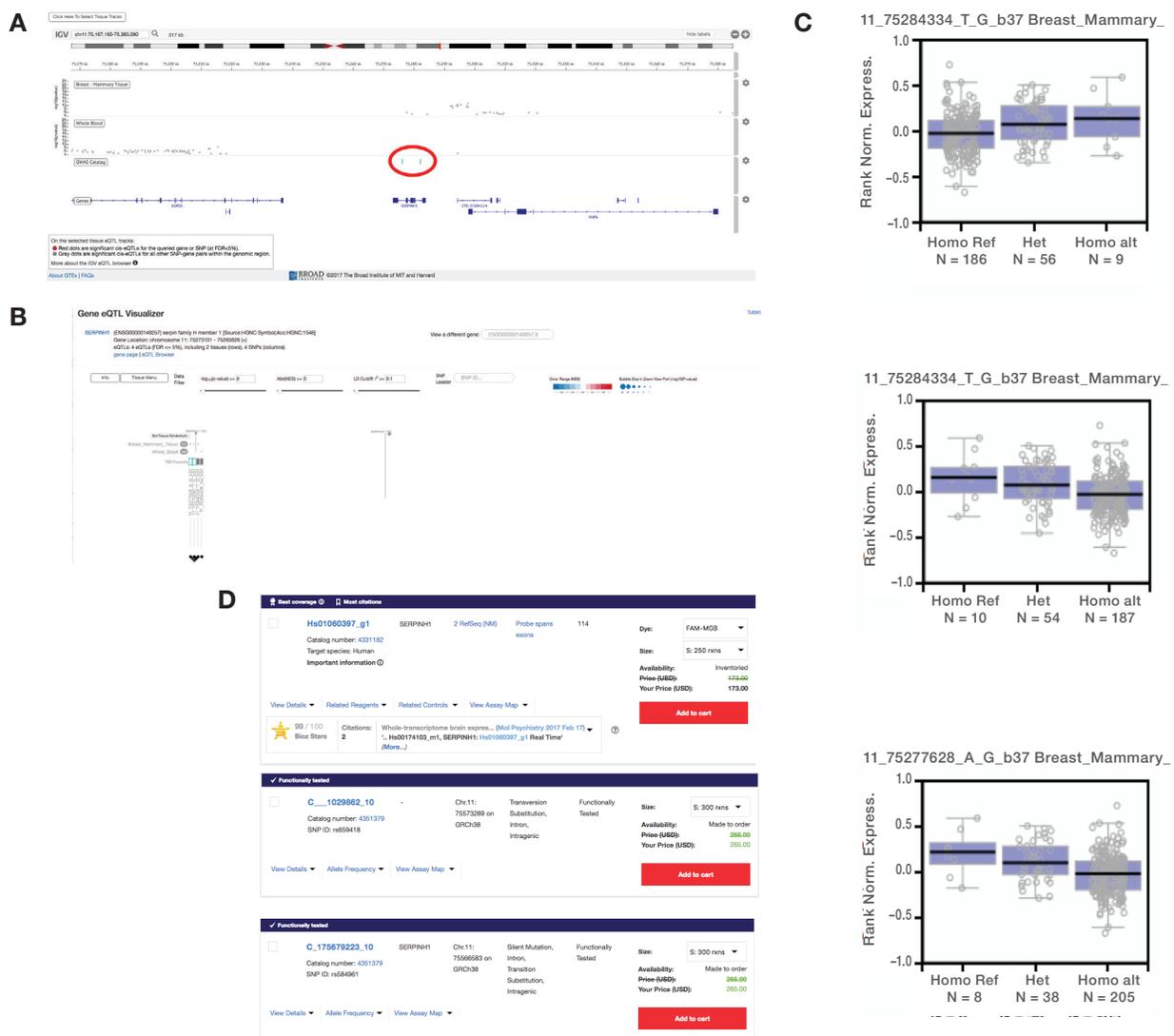


Figure 7. Example of how to use the GTEx database to link possible eQTLs to GWAS. (A) The IGV browser indicates two locations that are significant in GWAS studies (circled) that fall within the *SERPINH1* gene. There is potential overlap with three eQTLs in mammary tissue. **(B)** Entering *SERPINH1* in the “Gene-eQTL Visualizer” confirms that three eQTLs have been mapped to mammary tissue, but not to whole blood. **(C)** Clicking on the bubbles in (B) brings up box plots showing the relationship between genotype and expression levels. **(D)** TaqMan Assays can be used to confirm and extend these results. TaqMan Gene Expression Assays can measure the expression levels of *SERPINH1*, and can be used to identify SNP genotypes of individuals at the putative eQTL.

Thus, an investigator might use these reagents to analyze more samples to reinforce the link between these genotypes and *SERPINH1* expression in breast cancer. The investigator could use the SNP assays to identify individuals that might be more at risk because they have a genotype resulting in higher *SERPINH1* expression. Studies could be developed to determine whether modulating *SERPINH1* expression reduces the pathology of breast cancer samples. Thus, using these tools together may provide insights into how genotypic variation driving gene expression differences can influence pathologies.

Conclusions

GWAS have provided extraordinarily useful information on the association between genotypes and phenotypes or pathologies. Although SNPs that fall within coding regions of genes provide clear avenues for further investigations, the role of SNPs falling outside of coding regions have been less obvious. The ability of eQTLs to control gene expression is significant for understanding how these SNPs might contribute to phenotypes. Combining the GTEx database queries with TaqMan Assay verification is an easy way to solidify these links. Since the vast fraction of SNPs identified in GWAS fall outside of protein-coding regions, the analysis of eQTLs in the context of GWAS provides many opportunities to advance our understanding of the genotype–phenotype association.

The GTEx project was supported by the Common Fund of the Office of the Director of the National Institutes of Health (NIH), and by National Cancer Institute (NCI), National Human Genome Research Institute (NHGRI), National Heart, Lung, and Blood Institute (NHLBI), National Institute on Drug Abuse (NIDA), National Institute of Mental Health (NIMH), and National Institute of Neurological Disorders and Stroke (NINDS).

References

1. Visscher PM et al. (2017) 10 years of GWAS discovery: biology, function and translation. *Am J Hum Gen* 101:5-22.
2. Pal LR et al. (2015) Insights from GWAS: emerging landscape of mechanisms underlying complex trait loci. *BMC Genomics* 16 Suppl 8:S4.
3. MacArthur J et al. (2017) The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* 45:D896-D901.
4. Joehanes et al. (2017) Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* 18:16.
5. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) Project. *Nat Genet* 45:580-585.
6. He X et al. (2013) Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet* 92:667-680.
7. Thorvaldsdóttir H et al. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178-192.
8. Zhu et al. (2015) Chaperone Hsp47 drives malignant growth and invasion by modulating an ECM gene network. *Cancer Res* 75:1580-1591.

Find out more at thermofisher.com/qpcr

ThermoFisher
SCIENTIFIC