Using Taqman™ assays to verify eQTL links arising from GWAS Studies

Steve Jackson, Harita Veereshligam, Kamini Varma. Thermo Fisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA 94080

ABSTRACT

Tremendous progress has been made using genome-wide association studies (GWAS) to link genetic variation with phenotypes and pathologies. In spite of these success, it has been estimated that as many as 90% of SNPs identified in GWAS studies map to non-coding regions, complicating the mechanistic interpretation of the results. It is thought that these non-coding SNPs fall into regulatory regions and influence the expression of a gene or genes. The study of expression quantitative trait loci (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. To illustrate how eQTLs can be found and verified, we generated transcriptomic information from various tumor samples using Applied Biosystems™' Clariom™ D microarrays. To find putative eQTLs, we compared the SNPs genotypes and the gene expression levels in these samples to the GTEx database. Potential eQTLs in this set of samples were verified using Applied Biosystems' Taqman™ SNP Genotyping and Taqman Gene Expression assays. We therefore illustrate a workflow where candidate eQTLs can be confirmed and screened in larger cohorts using the more economical qPCR reagents available from Applied Biosystems.

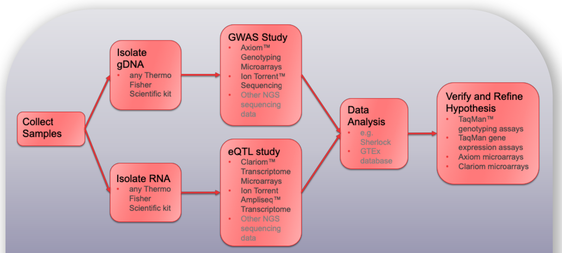


FIGURE 1. WORKFLOW FOR IDENTIFYING eQTLS

eQTLs are emerging as important markers of gene function, and may be useful as biomarkers for certain traits or pathologies. eQTLs are typically identified in a two-step process. To do this, a collection of samples from different individuals are needed. First, as part of a GWAS, genomic SNP variants among the samples may be identified by NGS or arrays. Next, transcriptomic analysis from the same samples is performed. RNA is collected from different tissues, treatments, tumor vs normal, etc, and subjected to microarray or RNAseq analysis. Transcripts that are differentially regulated in the different groups are then identified. The SNP genotypes and gene expression changes are then correlated. Thermo Fisher Scientific provides all the reagents needed for identifying eQTLs and verifying them as biomarkers.

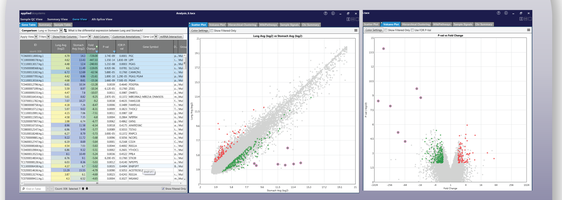


FIGURE 3. TISSUE SPECIFIC EXPRESSION CHANGES

Table of expression values and scatter plot showing relative expression of genes between FFPE stomach tumor and lung tumor samples. TAC 4.0 Software was used to examine pairwise relative expression of genes. The highlighted genes at the top of the table (blue rows) and in the plot (purple bubbles) were overexpressed in stomach tumor tissue relative to lung tumor tissue, consistent with previously published data. The same data can also be viewed as a volcano plot, with the same genes highlighted by purple bubbles.

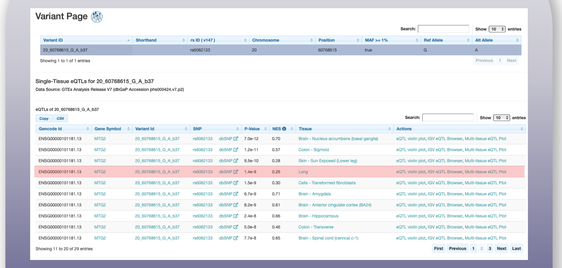


FIGURE 4. eQTL QUERIES USING GTEx DATABASE

The GTEx database collects and displays information about eQTLs. The database can be queried by gene, tissue, splice isoform, and SNP ID. In the example shown above, information about the SNP rs6062133 in MTG2 is displayed.

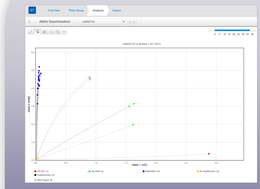
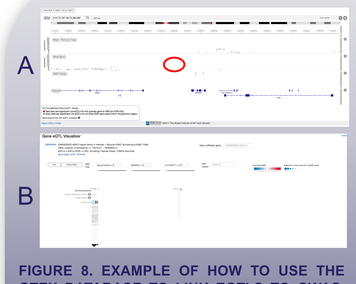
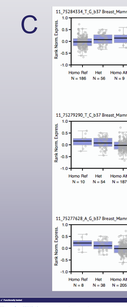


FIGURE 7. CONFIRMATION OF eQTL IN MTG2

Grouping expression levels of MTG2 relative to rs6062133 genotype reveals a slight increase in expression in GA heterozygotes relative to GG homozygotes (bar graph). These results are very similar to those in the GTEx database (violin plot). Note that none of the other SNPs (Figure 6) had diverse genotypes to confirm they were eQTLs.





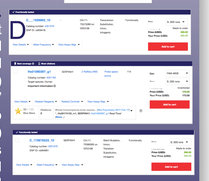


FIGURE 8. EXAMPLE OF HOW TO USE THE GTEX DATABASE TO LINK EQTLS TO GWAS.

(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.

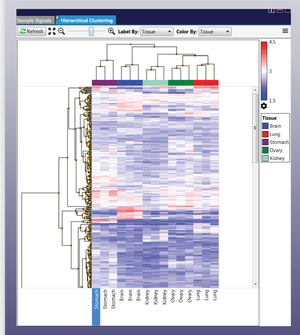
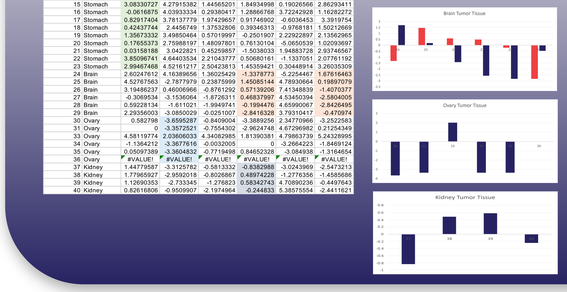


FIGURE 2. TRANSCRIPTOME ANALYSIS USING CLARIOM D ARRAYS

RNA was extracted from FFPE samples from brain, kidney, lung, ovary, and stomach tumor tissues and analyzed using Clariom D Microarrays. The clustering and heat map was generated using Transcriptome Analysis Console software 4.0 (TAC 4.0) and represents cluster analysis of 2,742 DEGs, filtered for those where confidences of differences are highly significant (P <0.001). DEGs from these samples are clustered by tissue type. Each row represents a single gene and each column represents a single sample. Colors represent log2 of relative expression levels (blue: low relative expression, red: high relative expression).



Sample		Stomach SLC12A2	Ovary CAMK2N1	Lung MTG2	K/Brain PLIN2	Lung PTPRZ1	Brain DAAM1
1	Lung	0.46413422	3.2626767	1.87433434	0.83977509	-0.291832	1.57363701
2	Lung	-0.2820187	3.45168495	1.38321449	1.43518592	2.77840424	1.58252296
3	Lung	0.80507851	4.18506622	1.27103996	0.84654427	-2.7000675	2.44141197
4	Lung	0.28075409	2.82642937	0.15559006	-0.0966663	1.77354431	1.38771439
5	Lung	-3.0376663	4.5014782	3.90583038	3.10616302	0.8693161	4.49963188
6	Lung	-0.602356	3.81539078	0.68644714	0.5343399	3.16471672	3.09586716
7	Lung	0.21424103	3.05561447	2.29639244	0.92623329	1.65219879	0.85616302
8	Lung	1.39889526	3.86637878	2.04773712	2.0897619	-4.829193	2.43449373
9	Lung	1.65423203	4.10580063	1.28172493	-1.5231819	-3.4114437	0.79397392
10	Lung	2.23487091	4.28033638	1.13050461	-0.2646523	-1.8409615	2.06763077
11	Lung	3.77326491	3.43093109	0.50865936	-2.7358894	-0.3498898	2.60133934
12	Lung	3.88363647	5.02052116	2.26078415	1.37465286	-0.3819046	3.92825127
13	Lung	1.59883118	4.10240936	1.89588682	0.30029915	-0.7115593	1.7939583
14	Stomach	2.57478714	4.64039993	1.60745239	0.60906982	-0.467783	1.42722893
15	Stomach	3.08530727	4.27915382	1.44565201	1.84534998	0.19026566	2.86293411
16	Stomach	-0.0616875	4.03933334	0.29380417	1.28666768	3.72242928	1.16282272
17	Stomach	0.82917404	3.76137779	1.97429857	0.91746902	-0.6026453	3.3919754
18	Stomach	0.42437744	2.4456749	1.37532806	0.39346313	-0.9768181	1.50212669
19	Stomach	1.35673332	3.49850464	0.57019997	-0.2501907	2.22922897	2.13562965
20	Stomach	0.17655373	2.75988197	1.48097801	0.76130104	-5.0656539	1.02093697
21	Stomach	0.03158188	3.0422821	0.45259857	-1.5038033	1.94883728	2.93746567
22	Stomach	3.85098741	4.64403534	2.21043777	0.50680161	-1.1337051	2.07761192
23	Stomach	2.99467468	4.52161217	2.50423813	1.45359421	0.30448914	3.26035309
24	Brain	2.60247612	4.16389856	1.36025429	-1.3378773	6.2254467	1.67616463
25	Brain	4.52767563	-2.7877979	0.23875999	1.45085144	4.78930664	0.19897079
26	Brain	3.19486237	0.46009966	-0.8761292	0.57139208	7.41348839	-1.4070377
27	Brain	-0.3069534	-3.1536064	-1.8726311	0.46837997	4.53450394	-2.5804005
28	Brain	0.59228134	-1.611021	-1.5949741	-0.1994476	4.65990067	-2.8426495
29	Brain	2.29356003	-3.0850029	-0.0251007	-2.8416328	3.79310417	-0.470974
30	Ovary	0.582798	-3.6595287	-0.8409004	-0.3889256	2.34770966	-3.2522583
31	Ovary	0	-3.3572521	-0.7554302	-2.9624748	4.67296992	0.21254349
32	Ovary	4.58119774	2.03606033	4.34082985	1.81390381	4.79863739	5.24328995
33	Ovary	-1.1364212	-3.3677816	-0.0032005	0	-3.2664223	-1.8469124
34	Ovary	0.05097389	-3.3604832	-0.7719498	0.84652328	-3.084938	-1.3164654
35	Ovary	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
36	Ovary	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!	#VALUE!
37	Kidney	1.44779587	-3.3125782	-0.5813332	-0.8382988	-3.0243969	-2.5473213
38	Kidney	1.77965927	-2.9592018	-0.8026867	0.48974228	-1.2776356	-1.4585686
39	Kidney	1.12690353	-2.733345	-1.276823	0.58342743	4.70899236	-0.4497643
40	Kidney	0.82616806	-0.9509907	-2.1974964	-0.244833	5.38575554	-2.4411621

FIGURE 5. TAQMAN VERIFICATION OF GENE EXPRESSION DIFFERENCES

Five genes with previously-defined eQTLs were chosen for further analysis. DNA and RNA were isolated from a second cohort of tumor samples using the Recoverall™ FFPE kit. RNA from these samples was analyzed using predesigned Taqman probes (best coverage) for the genes SLC12A2, CAMK2N1, MTG2, PLIN2, PTPRZ1 and DAAM1. These genes were chosen because the GTEx database showed there were known eQTLs that have been mapped these genes. None of these genes are hypothesized to be related to the tumor phenotype; they were chosen for eQTL illustration. The table shows the ΔΔCt of the gene relative to the OAZ1 reference transcript. Colored values highlight genes and tissues that have GTEx data for tissue-specific eQTLs. Note that one of the ovarian samples failed to produce any data. The accompanying graphs highlight the sample variation in ΔΔCt values for each tissue.

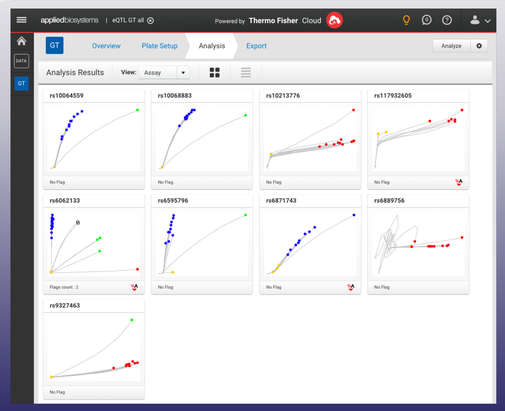


FIGURE 6. GENOTYPING USING TAQMAN SNP ASSAYS

DNA the same set of samples (Figure 5) was analyzed using predesigned Taqman SNP Genotyping assays. The database of Taqman assays was queried for SNPs that covered previously-mapped eQTLs proximal to the transcription start site (TSS) of the genes above. Genotypes were analyzed using the Thermo Fisher Connect™ (cloud-based) Genotyper (GT) application. Each sample is a different dot on each rsID plot; blue and red are called homozygous by the software, green dots are heterozygous, and yellow dots indicate no amplification in that sample.

CONCLUSIONS

- eQTLs can contribute to understanding link between genotype and phenotype
  - Influencing expression levels of a gene can influence a trait or pathology
  - May facilitate understanding how GWAS loci that fall outside coding regions act
  - Software tools exist that link GWAS results with eQTL studies
- TaqMan assays can be used to advance eQTL links to GWAS studies
  - TaqMan Gene Expression assays can analyze transcript levels
  - TaqMan Genotyping assays can analyze genotypes at specific SNPs
  - Other Thermo Fisher Scientific tools can be used for discovery-based eQTL research

ACKNOWLEDGEMENTS

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS