The GTEx (Genotype-Tissue Expression) eQTL database will be populated by tissue-specific eQTL information as the GTEx project progresses, but is currently limited in navigability. SNPexp (Holm et al., 2010) provides the database for users to investigate a specified region, but contains a limited number of eQTL datasets. Despite the current attention to eQTL datasets, the need remains for a powerful and versatile eQTL database to easily investigate regions, loci and transcripts of interest.

Here we introduce seeQTL, a new database of human eQTL associations. It is based on the Gbrowse2 platform, which is more powerful and customizable than the original Gbrowse. Most of the studies represented in seeQTL (Supplementary Material) were re-analyzed using our own pipeline, combining quality control, population stratification control, association testing and false discovery rate (FDR) control (Fig. 1) (Benjamini and Hochberg, 1995). In addition, we performed a meta-analysis to obtain a consensus association score for each eQTL across the HapMap studies and populations. Here we use the terms ‘cis eQTL’ for local eQTLs (within 1 Mb of a gene) and ‘trans eQTL’ for more distant eQTLs. Cis associations are displayed using either segment plots (Fig. 2) or FDR q-value association Manhattan plots and trans associations using Manhattan plots.

1 INTRODUCTION

The association of RNA expression traits with DNA variation, including from single-nucleotide polymorphisms (SNPs) and copy-number variants (CNVs), has been a subject of active inquiry in recent years, shedding light on fundamental biological processes underlying transcription. Here we use the generic term ‘expression quantitative trait loci’ (eQTLs) to describe these DNA variants and their associated expression traits (Feuk et al., 2006). A large number of studies has been published on HapMap lymphoblastoid cell lines and other human tissues, covering several continental-level populations (Choy et al., 2008; Dimas et al., 2009; Grundberg et al., 2009; Montgomery et al., 2010; Myers et al., 2007; Pickrell et al., 2010; Price et al., 2008; Schadt et al., 2008; Spielman et al., 2007; Stranger et al., 2007; Zeller et al., 2010). In addition to the general importance of dissecting transcriptional regulation, eQTL analysis may also provide a window into the mechanisms underlying transcription-mediated disease (Consoli et al., 2002).

Several online databases are available which report eQTL associations based on published datasets. SCAN (Gamazon et al., 2009) is a large-scale database of genetic and genomic data, which allows users to search for eQTLs by querying multiple genes or SNPs/CNVs, but is not designed primarily as an eQTL database. The eQTL Browser (Pickrell et al., 2010) is based on the Gbrowse platform (Donlin, 2009) and displays results from multiple studies and allows navigation throughout the genome. The GTEx (Genotype-Tissue Expression) eQTL database will be populated by tissue-specific eQTL information as the GTEx project progresses, but is currently limited in navigability. SNPexp (Holm et al., 2010) provides the database for users to investigate a specified region, but contains a limited number of eQTL datasets. Despite the current attention to eQTL datasets, the need remains for a powerful and versatile eQTL database to easily investigate regions, loci and transcripts of interest.

Here we introduce seeQTL, a new database of human eQTL associations. It is based on the Gbrowse2 platform, which is more powerful and customizable than the original Gbrowse. Most of the studies represented in seeQTL (Supplementary Material) were re-analyzed using our own pipeline, combining quality control, population stratification control, association testing and false discovery rate (FDR) control (Fig. 1) (Benjamini and Hochberg, 1995). In addition, we performed a meta-analysis to obtain a consensus association score for each eQTL across the HapMap studies and populations. Here we use the terms ‘cis eQTL’ for local eQTLs (within 1 Mb of a gene) and ‘trans eQTL’ for more distant eQTLs. Cis associations are displayed using either segment plots (Fig. 2) or FDR q-value association Manhattan plots and trans associations using Manhattan plots.

1.1 Datasets and analysis

We collected 14 human eQTL datasets, including unrelated HapMap lymphoblastoid cell lines (Choy et al., 2008; Dimas et al., 2009; Montgomery et al., 2010; Pickrell et al., 2010; Price et al., 2008; Spielman et al., 2007; Stranger et al., 2007), human cortical samples (Myers et al., 2007) and monocytes (Zeller et al., 2010). The gene expression data were downloaded from NCBI GEO, and genotype data were downloaded from HapMap or the authors’ website (Supplementary Material). We excluded a sample for low expression quality and excluded SNPs with low minor allele frequency (MAF). Detail of eQTL calculations and FDR control are provided in the Supplementary Material and Figure 1. Summarized results of the datasets are provided in Supplementary Table S1. Additional datasets will be added as data are made available. We soon anticipate loading results from the ‘godot’ study, an eQTL evaluation of peripheral blood gene expression in ~800 monozygotic and 750 dizygotic twin pairs.

1.2 Consensus method

The HapMap lymphoblastoid cell line data consist of multiple expression datasets and cover several continental-level populations. Separate analyses can be performed within each dataset. However, as the data are all from the same tissue source, the availability...
1.3 Usage features

The seeQTL browser is navigable using text-searches for genes and SNPs, presenting a table view of features containing these text strings. Alternatively and subsequently, seeQTL is navigable by clicking and zooming. These browser features allow maximum flexibility in focusing on specific genes and genomic regions. As described above, cis-associations are displayed using segment plots, which are useful to display the ‘connection’ between genes and associated SNPs, as well as Manhattan plots. For the SNPs in a region, all the genes to which these SNPs exhibit significant association can also be displayed in Manhattan plots. Tracks based on individual datasets are shown separately, as well as the consensus HapMap track. Comparisons of seeQTL features and advantages are shown in Supplementary Table S2.

ACKNOWLEDGEMENTS

We thank Guanhua Chen for advice and help with the database development.

Funding: Gillings Innovation Laboratory in Statistical Genomics (NIMH 5RC2MH089951 and R01MH090936).

Conflict of Interest: None declared.

REFERENCES


Fig. 1. The pipeline of eQTL analysis. PC, principal component.

Fig. 2. A screenshot of the seeQTL browser, with various features highlighted: eQTLs can be viewed as segment plots for cis-associations, showing SNPs as ‘connected’ to associated genes, or as Manhattan plots for both cis and trans associations.

of a single consensus meta-analysis would greatly facilitate eQTL analysis of HapMap samples. We applied a standard meta-analysis approach to obtain a consensus score for each transcript and each SNP with study-specific weights chosen to maximize power (Supplementary Material).