User Manual (RespiratoryGenomics.com)

Title: A database of annotated promoters of genes associated with common respiratory and related diseases

This database was created in an effort to elucidate transcriptional regulatory signatures of genes that are associated with common respiratory and related diseases (RRD). A large number of genes have been implicated in RRDs; however, their mechanisms of transcriptional regulation remain largely unknown. In this study, we annotated the promoter regions of these genes by analyzing them with a novel computational approach/pipeline that comprises an ensemble of some of the most state-of-the-art programs developed for promoter analysis. The results of our analyses are presented in the database in several different views that allow users easy querying and exploration of results.

To give a holistic view on transcription regulation features of RRD-associated genes, we have divided the database into the following five sections. Usage instructions for each section follow the descriptions.

System requirement
Internet Explorer ver. 8 or above; Firefox ver. 4 or above; and Chrome ver. 14.

Motif-centered view
This database section shows promoter region annotations for target disease genes. Annotations include putative transcription factor binding site (TFBS) motif signatures with their location in the promoter and strand. Putative TFBSs are also shown graphically in the context of known (possibly regulatory) genomic annotations including SNPs, histone modification sites, and CpG islands. We also provide a filter for the user to view promoter annotation diagrams restricted by putative TFBS motifs (i.e. shows promoter annotations that include a set of putative TFBS motifs selected by the user).

Putative TFBS motif models are presented as position weight matrices (PWMs) and shown ranked based on the statistical significance of their level of conservation in the target disease promoter group. These motif models could potentially be used by researchers in conducting other transcription regulation related studies such as i) finding other human genes that have promoter structures (i.e. motif patterns) similar to target disease promoters with the aim of identifying co-regulated genes; ii) characterizing and associating genome-wide regions with such motifs; and iii) identifying TFBS motif conservation between humans and other species. This section may also be useful for investigators interested in analyzing or using a particular disease-specific regulatory element or its model.

This section also presents transcription factors (TFs) that are likely to bind to the discovered TFBS motifs. In order to verify the association between the discovered TF and the target disease, we provide a link in this section to the user that displays evidence in always up-to-date PubMed literature for such an association (e.g. by automatically searching PubMed with a query construct as, ‘human’ AND ‘disease’ AND ‘TF gene name and its synonyms separated by OR’). The user can verify the evidence by manually reading the abstracts that are returned.

We have discovered a large number of known and yet to be discovered putative and novel TF-disease associations. In addition, we discovered a number of putative and novel TF-TFBS-disease associations. Such associations are potential candidates for further validation/experimental studies. The list of putative RRD-associated TFs can also be used in RRD-specific gene expression experiments to validate/confirm which of those TFs are actually differentially expressed and to determine which of those TFs form a functional unit of regulatory programs specific to a particular
gene-expression experiment or tissue condition. These results can also be used in planning protein-protein interaction experiments related to specific respiratory diseases.

**Paired-motif view**
TFBSs in related (e.g. co-regulated) promoters tend to form functional modules by pairing-up together. This section presents pairs of putative TFBS motifs that co-occur in the promoters of a given disease group. These pairs are shown ranked based on the statistical significance of the over-representation of the pair in a disease promoter group compared to that in the background (all human promoters together). Motif pairs that appear on top with higher significance (lower p-value) may be interesting candidates to explore further through experimental testing/validation.

**Promoter-centered view**
This section provides information similar to ‘Motif-centered view’ section above, except that the information provided here is more promoter-centric. This section has a filter for the user to view promoter annotation diagrams restricted based on promoter (i.e. show promoter annotations for a set of promoters selected by the user), unlike the motif filter based annotations in Motif-centered view. This section provides a link to the user that displays PubMed evidence for an association between putative TF-regulated gene pair in the context of the target human disease by searching PubMed with a query construct as, ‘human’ AND ‘disease’ AND ‘TF gene name and its synonyms separated by OR’ AND ‘TFBS motif containing (i.e. putatively TF-regulated) gene name and its synonyms separated by OR’. Further experimental studies are necessary to validate putative TFBSs that identified in different disease promoters in this study.

**Functional annotation view**
This database section provides a view that associates putative TF regulated (TFBS containing) gene pairs with known functional annotations such as, overlapping SNPs (dbSNP), and biological terms [e.g., pathways (KEGG, BioCarta, Reactome), protein domain (Interpro), Gene Ontology (GO, GO slim) and micro RNAs (miRBase)]. For assigning biological terms to the pair, we first identify terms that significantly enrich the genes in a given disease group before mapping the significant terms to the TF-gene pair. Statistically significant terms are shown in two ways: i) terms that are associated with either a target disease-group gene or a TF that is mapped to it; and ii) terms that are associated with a target disease-group gene and also with a TF that is mapped to it. This view provides insight into how regulatory elements may affect the function of a given gene of interest and a starting point for experimental validation.

**Meta-analysis view**
This section provides query functions to permit users to conduct meta-analyses on different disease groups. For example, the user can explore genes, TFs, GO-terms, pathways, SNPs or micro RNAs where a selected set of diseases converge or diverge. Understanding the similarities and differences in the regulatory mechanisms of respiratory related diseases may, for example, provide insight into the efficacy of a given treatment across this spectrum of related diseases.

It is our hope that this database proves to be a useful resource for respiratory researchers to gain deeper insight into gene transcription regulation mechanisms associated with RRD-associated target diseases.

More usage details on these five database sections are provided below. The user begins by selecting the view of interest from the choices on the left.

1. Motif-centered view
Step 1: After selecting “Motif-centered view” from the choices on the left, the user can choose a disease group from the drop down menu. Selecting a disease group and clicking ‘Submit’, the user can view TFBS motif models (along with mapped TFs) that are significantly associated with the group, as shown in Figure 1 below. Discovered motif models are ordered by increasing p-value (descending significance). Motif models are shown as position weight matrices (PWMs) that essentially represent the frequencies of A, C, T, and G at each position of the motif segment discovered in the promoters. The table can be sorted by clicking on the appropriate column header section. The magnifying glass icon indicates literature evidence for the association. Clicking on the magnifying glass icon links the user directly to the PubMed search. If on the search result page the user sees, 'No items found', it means that no PubMed abstract was found with an association.

Step 2: The user may then select individual TFBS motif models by checking the boxes on the left side of the table or may select all motif-models by checking the box that indicates ‘Select all motif-models’. By clicking ‘Submit’, the user will move to the next page, which shows signatures of the selected motif models in the promoter sequences (see Figure 2).

Step 3: By clicking the ‘Submit’ button, the tool will graphically show promoters with putative TFBS signatures, together with other known genomic annotations (e.g. histone modification sites, SNPs and CpG islands). For an example, see Figure 3 below.

Fig 1: Putative TFBS motif models shown with mapped TFs.
Fig 2: Signatures of selected TFBS motif models in disease gene promoters.
Fig 3: Motif diagram showing putative TFBS motifs with other genomic features.
2. Paired motif view

Step 1: After selecting “Paired Motif View” from the choices on the left, the user can select the disease group of interest and click ‘Submit’.

Step 2: Pairs of putative TFBS motif signatures that are over-represented in given disease-group promoters in a statistically significant manner are shown (Figure 4). Such pairs may possibly represent functional modules. Mapped TFs are also shown.

Fig 4: Putative TFBS motif pairs associated with disease genes.
3. Promoter-centered view

Step 1: After selecting “Promoter-centered view” from the choices on the left, the user can select the disease group of interest and click ‘Submit’.

Step 2: The tool will display a table with disease-associated promoters that the user can select for further analysis (Figure 5).

Step 3: Clicking the ‘Submit’ button allows the user to view putative TFBS motif signatures and known genomic annotations in the selected promoters (as shown in Figure 2).

Step 4: By clicking the ‘Submit’ button, the user can graphically view putative TFBS motif signatures and known genomic annotations (as shown in Figure 3).

Fig 5: Promoters of disease genes.
4. Functional annotation view

Step 1: After selecting “Functional Annotation View” from the menu on the left, the user can select the disease group of interest and click ‘Submit’.

Step 2: The system displays a page as shown in Figure 5, listing disease-associated promoters. The user may choose any number of associated disease genes and click the ‘Submit’ button.

Step 3: Statistically significant functional terms are shown (Figure 6). The last two columns show: i) terms that are associated with a target disease-group gene and also with a TF that is mapped to it and ii) terms that are associated with either a target disease-group gene or a TF that is mapped to it. In addition, SNPs that overlap with the discovered motifs are reported.

Fig 6: Functional terms mapped to of TF-regulated gene pairs in a disease group.
5. Meta-analysis view

Step 1: To conduct a meta-analysis, the user begins by selecting and number of different disease groups of interest (Figure 7). The user can explore genes, TFs, GO-terms, pathways, SNPs or micro RNAs where a selected set of diseases converge or diverge by selecting from the dropdown menu labeled “Select annotation type” and clicking ‘Submit’.

Step 2: Clicking the ‘Submit’ button allows the user to view how selected diseases converge (highlighted in yellow) or diverge (not highlighted) (Figure 8). The user can probe further by clicking on the links.

Fig 7: Meta Analysis of diseases on different annotation type
Fig 8: Meta Analysis for three diseases based on KEGG pathway information.