The Phenogen website (https://phenogen.org) has been operating for the last 16 years providing a resource for quantitative transcriptome data for RI panels of rats and mice. Recent development efforts have focused on providing data access through friendly tools on the web interface. Phenogen currently has more than 5TB of raw and 35TB of processed sequence data, deposited on public repositories such as SRA for 44% of the Hybrid Rattus Diversa Panel in whole brain, heart, liver, and kidney, much of which is accessible through the web interface. The representative state transfer application interface (REST API) will provide direct access to data and tools to make functions accessible for advanced users or enable programmatic access for other services. The REST API is the next step in the move to a new architecture, which along with the new domain name, is now hosted in the cloud to provide robust performance and more flexible data access. The API is implemented using AWS Gateway connected to python AWS Lambda functions for modularity, scalability, and flexibility. As many programming languages support REST APIs, this will enable direct access to the RNA-seq and microarray data in other environments for reuse and further analysis. The goal is to provide a central resource for versions of the data set with access to various steps in the processing pipeline, and to provide programmatic access to Phenogen unique set of data and tools that are centered around analyzing differences in and regulation of the transcriptome across the HRPD and across tissues. The API will extend beyond data to provide meta-data related to documenting how versions of the dataset were generated, to generating images that summarize expression, QTLs, alternative splicing, among others. Phenogen is supported by NIAAA (R01AA103162, NIGMS (P01GM104423), Barsony Fund, and computing resources from the National Supercomputing Center & Dedicated Research Network, UNLV.

**Background**

The full HRPD when completed will be a combination of reconstituted inbred (HR/BH/HF and FXL/FEX) and classic inbred strains of rats that optimizes genetic mapping projects. It will provide a genomically stable and diverse population lending itself to multi-omics studies. As can be seen in table 1 the right, currently about half of the HRPD is made up of strains that have RNA-seq data, the remaining strains are from a sequencing of ribosomal depleted total RNA from two tissues. The Phenogen website (https://phenogen.org) provides data and tools to apply both phenotype-driven analyses and gene driven analyses to understand the contribution that a genetic network can make to a phenotype. The site provides graphics and tables to explore genetic predisposition of the HRPD population to any new phenotype measured on all or part of the population. The tools are used for this range from gene co-expression networks to mRNA target predictions to gene or transcript expression of novel and annotated named transcripts. A brain and liver transcriptome reconstruction has been completed across the currently available HRPD strains. The REST API will support retrieving these data and tools. Phenogen will be connected to python AWS Lambda functions for modularity, scalability, and flexibility.

**Example API function for Gene Level Data**

```bash
http://rest.phenogen.org/genes/idLookup
```

- Gene target databases including both validated and validated gene targets, Gene Ontology summaries, and other modules correlated to this gene.
- Retrieve normalized RNA-Seq/Array expression of gene data across strains in a tissue or for a specific tissue.
- Examine correlation of eQTL targets and gene expression data across a tissue.
- Retrieve multi-RM results for a gene or mRNA. MultiRM provides mRNA targets directly for 12 different mouse RI strain gene datasets including both validated and prediction sources. MultiRM is already available as an R package but this will provide functionality outside of R.

**Example API workflow for P2rx4**

1. Look up required Gene IDs using gene symbol P2rx4

```
http://rest.phenogen.org/genes/symbol
```


2. Get gene report

```
http://rest.phenogen.org/genes/report
```


3. Look at eQTLs for gene from RNA-Seq and/ or Arrays

```
http://rest.phenogen.org/genes/eQTLs
```

- Example eQTLs for an eQTL scan in a strain: “eneXID: "ID, "geneList": [ {gene:"ENSRNOG00000001300", exprVer:1}, {gene:"ENSRNOG00000001300", exprVer:2} ]

4. Look at eQTLs for gene from RNA-Seq and/ or Arrays

```
http://rest.phenogen.org/genes/eQTLs
```

**Example Table 1:**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Tissue</th>
<th>Sex</th>
<th>Number of Biologically Relevant Genes</th>
<th>Number of Single-End RNA-Seq and Total RNA</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
<th>Number of Biologically Relevant Gene Lists</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
<th>Number of Biologically Relevant Gene Lists</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHR and B6X</td>
<td>Brain</td>
<td>male</td>
<td>3</td>
<td>465 million</td>
<td>56 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Brain</td>
<td>female</td>
<td>4</td>
<td>982 million</td>
<td>297 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Liver</td>
<td>male</td>
<td>3</td>
<td>583 million</td>
<td>342 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Liver</td>
<td>female</td>
<td>4</td>
<td>750 million</td>
<td>300 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B6 and FVB</td>
<td>Brain</td>
<td>male</td>
<td>3</td>
<td>17.0 billion</td>
<td>5.0 billion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B6 and FVB</td>
<td>Liver</td>
<td>male</td>
<td>4</td>
<td>4.0 billion</td>
<td>5.1 billion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Background**

- The Phenogen website (https://phenogen.org) provides a quick summary of data related to the gene.
- Look at WGCNA modules containing the gene of interest. Then look at connectivity between genes in the module, module Gene Ontology (GO) terms, mRNA targets of contained genes, Gene Ontology summaries, and other modules correlated to this gene.
- Retrieve normalized RNA-Seq/Array expression of gene data across strains in a tissue or for a specific tissue.
- Examine correlation of eQTL targets and gene expression data across a tissue.
- Retrieve multi-RM results for a gene or mRNA. MultiRM provides mRNA targets directly for 12 different mouse RI strain gene datasets including both validated and prediction sources. MultiRM is already available as an R package but this will provide functionality outside of R.

**Table 1: Brain and Liver RNA-Seq Data Available**

- SHR and B6X
- B6X and FVB
- SHR and B6X
- B6 and FVB

**Example Table 2: Summary of RNA-Seq Data Collected**

<table>
<thead>
<tr>
<th>Strains</th>
<th>Tissue</th>
<th>Sex</th>
<th>Number of Biologically Relevant Genes</th>
<th>Number of Single-End RNA-Seq Strains</th>
<th>Number of Biologically Relevant Gene Lists</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
<th>Number of Biologically Relevant Gene Lists</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
<th>Number of Biologically Relevant Gene Lists</th>
<th>Number of Single-End Published RNA-Seq Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHR and B6X</td>
<td>Brain</td>
<td>male</td>
<td>3</td>
<td>465 million</td>
<td>56 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Brain</td>
<td>female</td>
<td>4</td>
<td>982 million</td>
<td>297 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Liver</td>
<td>male</td>
<td>3</td>
<td>583 million</td>
<td>342 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHR and B6X</td>
<td>Liver</td>
<td>female</td>
<td>4</td>
<td>750 million</td>
<td>300 million</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B6 and FVB</td>
<td>Brain</td>
<td>male</td>
<td>3</td>
<td>17.0 billion</td>
<td>5.0 billion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B6 and FVB</td>
<td>Liver</td>
<td>male</td>
<td>4</td>
<td>4.0 billion</td>
<td>5.1 billion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Conclusion**

Phenogen.org provides extensive data on the HRPD transcriptome, which at the gene level is available through the REST API. Continued development will make remaining data and tools available to access programmatically in addition to the web interface. The API will be generating some of the graphics available through the web interface or creating gene lists from other function calls and proceeding to analyze the gene lists. This will provide a valuable resource for a variety of users developing their own tools/pipelines.

**Phenogen.org**

- Provides access to a wide range of data on the HRPD and across tissues.
- Accessible through friendly tools on the web interface.
- The REST API provides programmatic access to the data.

**Acknowledgments**

This project was supported by NIAAA (K AAA103162, NIGMS (P01GM104423), and Barsony Fund.

**Contact Information**

Boris Tabakoff, PhD
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of Colorado Denver Anschutz Medical Campus
E-mail: Boris.Tabakoff@CUAnschutz.edu

**Phenogen.org**

- Provides access to a wide range of data on the HRPD and across tissues.
- Accessible through friendly tools on the web interface.
- The REST API provides programmatic access to the data.