

Fall 12-31-2018

PolyA DB3: a database cataloging polyadenation sites(pas) across different species and their conservation

Ram Mohan Nambiar
New Jersey Institute of Technology

Follow this and additional works at: <https://digitalcommons.njit.edu/theses>



Part of the [Bioinformatics Commons](#), and the [Computer Sciences Commons](#)

Recommended Citation

Nambiar, Ram Mohan, "PolyA DB3: a database cataloging polyadenation sites(pas) across different species and their conservation" (2018). *Theses*. 1636.
<https://digitalcommons.njit.edu/theses/1636>

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at Digital Commons @ NJIT. It has been accepted for inclusion in Theses by an authorized administrator of Digital Commons @ NJIT. For more information, please contact digitalcommons@njit.edu.

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

POLYA_DB3: A DATABASE CATALOGING POLYADENATION SITES(PAS) ACROSS DIFFERENT SPECIES AND THEIR CONSERVATION

**by
Ram Mohan Nambiar**

Polyadenation is an important process occurring in the messenger RNA that involves cleavage of 3' end nascent mRNAs and addition of poly(A) tails. For this thesis, I present PolyA_DB3, a database cataloging cleavage and polyadenylation sites (PASs) in several genomes specifically for human, mouse, rat and chicken. This database is based on deep sequencing data. PASs are mapped by the 3' region extraction and deep sequencing (3'READS) method, ensuring unequivocal PAS identification. Large volume of data based on diverse biological samples is used to increase PAS coverage and provide PAS usage information. Strand-specific RNA-seq data were used to extend annotated 3' ends of genes to obtain more thorough annotations of alternative polyadenylation (APA) sites. The database also has information regarding conservation of PAS between these species. Similar analysis has also been done on the PASs identified from frog samples and the identification of conservation of the PASs.

**POLYA_DB3:
A DATABASE CATALOGING POLYADENATION SITES(PAS)
ACROSS DIFFERENT SPECIES AND THEIR CONSERVATION**

by
Ram Mohan Nambiar

**A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Bioinformatics**

Department of Computer Science,NJIT,Newark

Dec 2018

Blank Page

APPROVAL PAGE

**POLYA_DB3:
A DATABASE CATALOGING POLYADENATION SITES(PAS)
ACROSS DIFFERENT SPECIES AND THEIR CONSERVATION**

Ram Mohan Nambiar

Dr. Bin Tian, Thesis Advisor Date
Associate Professor, Department of Biochemistry & Molecular Biology, NJMS

Dr. Usman Roshan, Committee Member Date
Associate Professor, Department of Computer Science, NJIT

Dr. Zhi Wei, Committee Member Date
Associate Professor, Department of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Ram Mohan Nambiar

Degree: Master of Science

Date: Dec 2018

Undergraduate and Graduate Education:

- Master of Science in Bioinformatics,
New Jersey Institute of Technology, Newark, NJ, 2018

Major: Bioinformatics

Presentations and Publications:

Wang, R., Nambiar, R., Zheng, D., and Tian, B., "PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes," *Nucleic Acids Research.*,

*Om Asato Maa Sad-Gamaya
Tamaso Maa Jyotir-Gamaya
Mrtyor-Maa Amrtam Gamaya
Om Shaantih Shaantih Shaantih*

(Lead me from the Temporal to the Eternal.
Lead me from darkness to light.
Lead me from death to immortality
Let there be Peace, Peace and Peace everywhere.)

Brhadaranyaka Upanishad

ACKNOWLEDGMENT

At the very outset, I would like to express my respect and sincere gratitude to my guide, Dr. Bin Tian, Associate Professor, Department of Biochemistry & Molecular Biology, NJMS who has given me the opportunity to work under him. It was a great learning experience to assist him in his research and to see him work diligently, with attention to minute details was truly inspiring.

I hereby, unequivocally acknowledge my abiding appreciation to my mentor, Dr. Usman Roshan, Associate Professor, Department of Computer Science, NJIT for his continuous encouragement and unremitting support throughout the tenure of my graduate degree program at NJIT.

A special thanks to the committee members, Dr. Bin Tian (Associate Professor, Department of Biochemistry & Molecular Biology, NJMS), Dr. Usman Roshan (Associate Professor, Department of Computer Science, NJIT) and Dr. Zhi Wei (Associate Professor, Department of Computer Science, NJIT) for taking the time off from their busy schedule for the presentation of my defence of this thesis

I would like to thank my peers Ruijia Wang and Aysegul Guvenek for their invaluable help.

Finally, I would like to thank my Parents for all their Blessings and for standing by me at all times.

TABLE OF CONTENTS

| Chapter | Page |
|---|------|
| 1 INTRODUCTION | 1 |
| 1.1 Objective | 1 |
| 1.2 Background | 1 |
| 2 METHODS | 3 |
| 2.1 Identification of PASs with 3READS Data | 3 |
| 2.2 PAS Annotation | 4 |
| 2.3 Conservation of PASs | 6 |
| 2.4 PAS Usage Levels | 6 |
| 2.5 Database Construction | 6 |
| 2.6 Identification and Conservation of PAS in <i>Xenopus Tropicalis</i> (Western clawed frog) | 7 |
| 3 RESULTS | 10 |
| 3.1 Data Access And Website Interface | 10 |
| APPENDIX A SCRIPT FOR INDEX.PHP | 12 |
| APPENDIX B SCRIPT FOR CHICKEN.PHP | 16 |
| APPENDIX C SCRIPT FOR MOUSE.PHP | 22 |
| APPENDIX D SCRIPT FOR HUMAN.PHP | 28 |
| APPENDIX E SCRIPT FOR RAT.PHP | 34 |
| BIBLIOGRAPHY | 40 |

LIST OF TABLES

| Table | Page |
|---|------|
| 2.1 Summary of PolyA_DB version 3 | 5 |
| 2.2 Summary of initial statistics for <i>Xenopus Tropicalis</i> (Western clawed frog) | 9 |

LIST OF FIGURES

| Figure | | Page |
|---------------|--|-------------|
| 2.1 | Schematic of PAS identification and presentation in PolyA_DB Version 3 | 3 |
| 2.2 | Relational Schema of PolyA_DB3 | 8 |
| 3.1 | An example of search result from PolyA_DB 3. | 11 |

CHAPTER 1

INTRODUCTION

1.1 Objective

The main objective of this thesis is to present PolyA_DB3 ,a database containing important information regarding the polyadenylation sites of different species such as mouse, chicken, rat and human. The PolyA_DB3 database also contains information regarding the conservation of polyadenylation sites among these species. In addition to that, similar analysis is done on frog and its Poly(A) conservation with human and mouse is identified.

1.2 Background

Polyadenylation is the addition of a poly(A) tail to a messenger RNA. The poly(A) tail consists of multiple adenosine base chain. The 3'-most segment of the pre-mRNA is cleaved off by some RNA binding proteins. These proteins then synthesize the poly(A) tail at the RNA's 3' end. The poly(A) tail is important for the nuclear export, translation, and stability of mRNA. The polyA site (PAS), is defined by multiple surrounding regulatory cis elements. In vertebrates, the cis elements include the PAS hexamer (AAUAAA, AUUAAA, or their variants), UGUA motif and U-rich motifs, all located upstream of the PAS, and downstream U-rich and UGUG motifs. PAS cis elements vary in lower species. In some genes these proteins add a poly(A) tail at one of several possible sites. Therefore, polyadenylation can produce more than one transcript from a single gene (alternative polyadenylation). Most APA sites are located in 3 untranslated regions (3UTRs) of mRNAs, resulting in isoforms with different 3UTR lengths and, consequently, distinct mRNA metabolisms. In addition, a sizable fraction of the sites are embedded in introns, influencing both coding and non-coding regions of gene transcripts. APA greatly increases the diversity of transcriptome

encoded by a genome, and is highly regulated across tissues and cell types . In addition, global regulation of the APA profile has been shown in cell proliferation, differentiation, and development , and in cells responding to environmental cues .

Given the critical role PAS play in termination of transcription,it is important to create a robust and accurate catalog of PASs in different genomes .Previous databases were based on cDNA and expression sequence tags. However this method is not comprehensive due to the limited number of cDNA/EST sequences available in public databases. In addition, internal A-rich sequences of transcripts often lead to poly(A/T) sequences in cDNAs, resulting in false identification of PAS. In the last few decades, a number of sequencing methods have been developed to specifically interrogate the 3 end of transcripts.This has led to identification of a lot of PASs. However,while 3 end sequencing methods have greatly facilitated PAS identification genome-wide, priming at internal A-rich sequences is still an issue leading to false identification of PASs .For this reason,the new database PolyADB3 is based on large volumes of data generated by 3READS which is a 3 end sequencing method that is not affected by internal A-rich sequences. The database also contains information regarding the conservation of PASs across human,mouse,rat and chicken genomes. In addition to the four genomes mentioned,PAS identification for common frog and its conservation with PASs identified in human was performed

CHAPTER 2

METHODS

Figure 2.1 gives the basic procedure involved in PAS identification and presentation in PolyA_DB version 3.

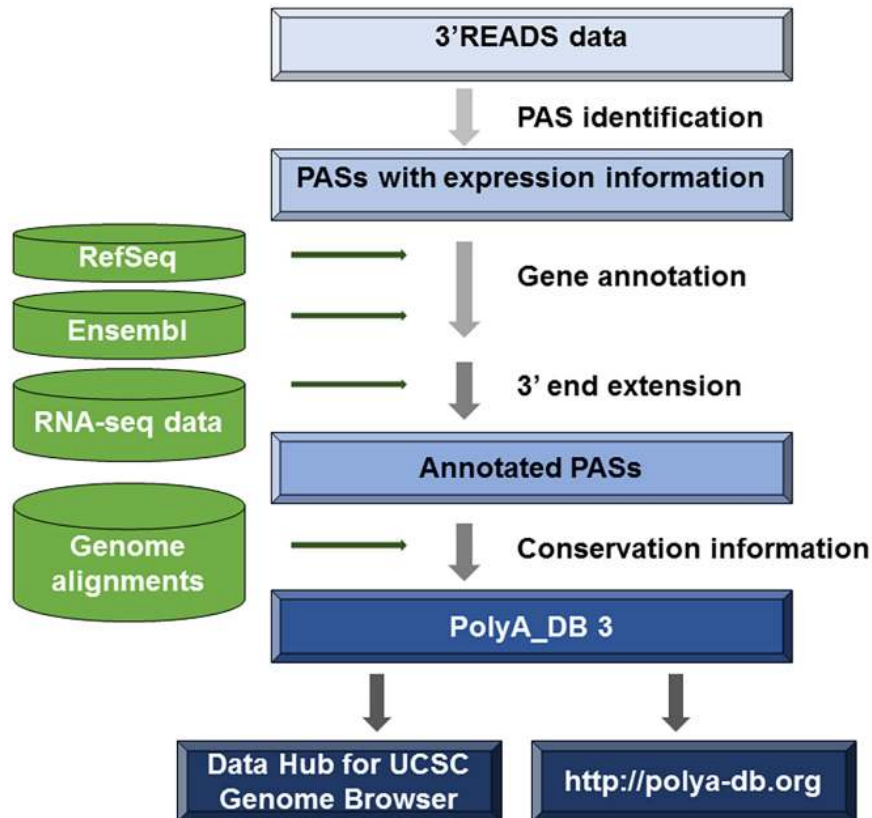


Figure 2.1 Schematic of PAS identification and presentation in PolyA_DB version 3. The data flow is indicated by arrowed lines.

2.1 Identification of PASs with 3READS Data

As described previously, the polyADB3 uses 3READS deep sequencing method to interrogate 3end of poly(A)+ transcripts. This method makes use of a chimeric oligo containing DNA and RNA (or locked nucleic acid) to retain the 5 end region of poly(A) tail in the cDNA. First, 959 3READS samples per species from diverse tissues

and cell lines of human, mouse, rat and chicken, totaling 23150 million PAS-containing reads per species were collected. Once the reads were collected, they were aligned to the corresponding genomes obtained from UCSC genome (mm9 for mouse, hg19 for human, rn5 for rat and galGal4 for chicken) for identification of PASs using bowtie2 (version 2.2.9). Before the reads were mapped onto the genome, the adapters of the reads were removed. Reads with a mapping quality score (MAPQ) 10 were kept for further analysis. Reads with 2 non-genomic 5Ts after alignment were identified as PAS reads. For each sample, the PASs within 24 nt from each other were clustered to address heterogeneous cleavage in PAS usage. Only the PASs with at least two reads in at least two samples were considered as genuine PASs.

2.2 PAS Annotation

Once the PASs were identified, they were assigned to genes based on RefSeq database (Release 83) and Ensembl database (release 75 for human, release 67 for mouse, release 79 for rat and release 85 for chicken). Because RefSeq and Ensembl gene annotations often miss PASs at the 3' end of genes, strand-specific, poly(A)⁺ RNA-seq datasets were used to extend the 3' ends defined by RefSeq and Ensembl. Continuous coverage of RNA-seq reads in the extended region, with a minimum of five reads at each position is required when RNA-seq dataset is used. Also, 3' end extension should not exceed the transcription start site of the downstream gene on the same strand. PASs are then annotated by their intron/exon location obtained from RefSeq and Ensembl databases i.e. 5'-most exon, internal exon, 3'-most exon, single exon and intron. This step was carried out for both mRNA and ncRNA genes. When a gene was annotated in both RefSeq and Ensembl databases, preference was given to RefSeq.

mRNA genes were further classified into four types based on coding information derived from the representative RefSeq or Ensembl sequence, including 5'UTR, CDS, 3'UTR and intron. Because most 3'UTRs harbor multiple PASs, 3'UTR were further

Table 2.1 Summary of PolyA_DB version 3

| | <i>Species</i> | <i>Human</i> | <i>Mouse</i> | <i>Rat</i> | <i>Chicken</i> |
|-----|-----------------------------------|--------------|--------------|------------|----------------|
| (1) | No. of samples used | 24 | 59 | 11 | 9 |
| (2) | No. of PAS reads used | 59,090,907 | 153,989,213 | 23,616,600 | 29,104,491 |
| (3) | No. of PASs | 108,042 | 202,426 | 61,905 | 65,909 |
| (4) | No. of genic PASs | 85,275 | 121,163 | 36,941 | 45,116 |
| (5) | No. of genes listed | 20,998 | 21,588 | 14,529 | 12,292 |
| (6) | No. of genes with 3 end extension | 8,962 | 12,027 | 8,302 | 8,352 |
| (7) | Median 3 end extension size (nt) | 758 | 469 | 617 | 1062 |
| (8) | No. of mRNA genes | 15,977 | 17,846 | 14,077 | 12,130 |
| (9) | No. of ncRNA genes | 5,021 | 3,742 | 452 | 162 |

classified into 3UTR first, middle and last PASs, based their relative locations. If a gene had a single 3UTR PAS, it was called single PAS. PAS hexamer sequence for each PAS were also annotated, using the 40-nt upstream region of the PAS. These include 5 types: AAUAAA, AUUAAA, Other (AGUAAA UAUAAA CAUAAA GAUAAA AAUAUA AAUACA AAUAGA AAAAAG ACUAAA), A-rich (AAAAAA) and None.

2.3 Conservation of PASs

In order to obtain the conservation of PASs between genomes, pair-wise genome alignment chain files from the UCSC Genome Bioinformatics Site to obtain conserved region between genomes. Two PASs from two species were considered to be orthologous when they were within 24-nt from one another in whole genome alignment. A particular site is said to be conserved if PAS is conserved between at least 2 of the 3 mammals (human, rat and mouse).

2.4 PAS Usage Levels

To evaluate the usage level of each PAS, two metrics were developed, percentage of samples expressed (PSE) and mean RPM (reads per million), based on the samples used for 3READS. The PSE of a PAS was calculated as $N_{\text{Expressed}}/N_{\text{Total}}$, where $N_{\text{Expressed}}$ is the number of samples in which the usage of PAS was detected (2 reads per sample), and N_{Total} is the total number of samples used. The mean RPM of each PAS is averaged RPM value across all the samples in which its usage was detected (2 reads). The RPM value of a PAS in each sample is the number of reads for the PAS normalized to the total number of reads mapped to the genome.

2.5 Database Construction

The PolyADB3 is based on relational database and the relational schema is shown in the figure 2.2. The PolyADB3 consists of 8 relational tables. The Chicken_PAS, Mouse_PAS, Human_PAS and Rat_PAS relational table gives infor-

mation regarding the PASs for each species. The unique PASid for each of these tables are used as primary keys to identify the PAS. They also contain various attributes that gives important information regarding the polyA sites such as the chromosome(chr) where the PAS is present, the position of the PAS (Pos) ,strand where the PAS is located, the PAS type based on the genic location(PAS type), the RefSeqID and EnsambleID of the PAS, the mean RPM (reads per million) ,the PSE(percentage of sample expressed) and the conservation of PAS across species. The PolyADB3 also contains information regarding the genes as seen in the relational tables Chicken_gene, Mouse_gene, Human_gene and Rat_gene with RefSeqID and EnsID used as primary keys to denote each gene. The RefSeqID and EnsID in the PAS table are foreign keys used to link the gene table to the PAS table. The gene table contains useful information regarding the genes such as chromosome(chr) where the gene is present, the strand ,the starting position of the gene(Start) ,Ending position of the gene, annotated gene symbol and the gene name. The gene table also contains the position of the last PAS identified for that species. The database also contains an ortholog table that consists of information regarding genes present in different species. Each row in this table is identified by the HID and TaxonomyID as primary keys. The HID is used to identify the Gene where each gene is given a unique HID and the TaxonomyID is used to identify the species the Gene belongs to .The gene tables are linked to the Ortholog table by their RefSeqID which is used as a foreign key pointing to the GeneID of the ortholog table.

2.6 Identification and Conservation of PAS in *Xenopus Tropicalis* (Western clawed frog)

Similar to the identification of PASs in human, mouse, chicken and rat, PAS identification for frog was performed using 3READS obtained from 4 samples of the frog at different stages represented as Stage 7, Stage 12.5, Stage 32 and Stage 40. The adapters of these samples are trimmed and mapped onto the *Xenopus Tropicalis*

genome (xenTro9) obtained from UCSC genome. Only those reads with mapping quality score (MAPQ) 10 and reads with 2 non-genomic 5Ts after alignment were considered as PAS reads. There was no sufficient data available for gene annotation of PASs for frog. For this purpose, PASs were not annotated. Those PASs in frog that were 24nt away from the whole genome alignment of human and frog pairwise alignment obtained from UCSC were considered to be conserved. Using the PASs information in human from PolyADB3, the conservation of PASs in other three species (rat, mouse and chicken) and frog is identified.

Table 2.2 Summary of initial statistics for *Xenopus Tropicalis* (Western clawed frog)

| | <i>Sample</i> | <i>No. of Raw reads</i> | <i>No. of mapped PAS</i> | <i>Percentage of mapped PAS(%)</i> | <i>No. of PASS reads</i> | <i>Percentage of PASS reads(%)</i> |
|-----|---------------|-------------------------|--------------------------|------------------------------------|--------------------------|------------------------------------|
| (1) | Stage12.5 | 22,194,474 | 17,166,086 | 77.6 | 6,201,808 | 37 |
| (2) | Stage32 | 23,436,343 | 17,800,299 | 76.3 | 6,354,724 | 34.8 |
| (3) | Stage40 | 25,622,024 | 20,739,594 | 81.3 | 5,437,896 | 26.2 |
| (4) | Stage7 | 23,568,245 | 18,158,461 | 77.3 | 5,258,578 | 29 |

CHAPTER 3

RESULTS

3.1 Data Access And Website Interface

The database can be accessed through a web interface (URL: <http://www.polya-db.org/v3>). The web interface consists on an query page where the user can query the database. It consists of a drop down button specifying the type of species and a text box where the user can type in the gene name or RefSeqID/EnsembleID of the gene. The below figure shows the index page. Once the user clicks the OK button, the website displays the information page as shown in Figure 3.1. The information page gives the user information regarding the gene and the PAS present in the gene. The gene table contains a summary of the queried gene, including gene symbol, gene ID (both RefSeq and Ensembl), gene name, gene type, genome version and annotated transcription start site and the last PASs based on RefSeq or Ensembl and PolyA_DB. It also contains orthologous genes present in other species in PolyA_DB based on the HomoloGene database from NCBI. The gene table also contains a link to UCSC genome browser for visualization of the gene and for integration with other public genomic data. The PolyA Site View table lists all the PASs assigned to the queried gene. For each PAS, information about its genomic location (also used as ID for the PAS, or PAS_ID), intron/exon location (5-most exon, 3-most exon, internal exon and intron), PAS type (5UTR, CDS, 3UTR and Intron), PSE, mean RPM and conservation in mammals. A link to UCSC genome browser is also provided for each PAS. The PolyA_DB data can also be viewed on UCSC genome browser through a custom track. The URL for the PolyA_DB track hub is <http://www.polya-db.org/v3/hub/>. As in PolyA_DB, each PAS is identified by its PAS_ID with conservation information (C for conserved, N for non-conserved). The mean RPM of all samples can also be displayed. In addition, batch download of data

in a tabular format is available at <http://www.polya-db.org/v3/download> The source code for the website is available in the Appendix.

| A | | B | | | | | |
|--------------------------------------|---|----------------------------------|-------------------------|------------|-------|----------|----------|
| Gene Summary | | PolyA Site Summary | | | | | |
| Species: | Mouse | PAS_ID | PAS type | PAS Signal | PSE | Mean RPM | ConserV. |
| Official Gene Symbol | Cstf3 | chr2:104442519+ | Intron | AUUAAA | 6.8% | 4.5 | Yes |
| RefSeq Gene ID | 228410 | chr2:1044439286+ | Intron | AUUAAA | 88.1% | 10.9 | Yes |
| Ensembl Gene ID | ENSMUSEG000000027176 | chr2:104445748+ | Intron | Other | 5.1% | 1.2 | No |
| Gene Name | coverage stimulator factor 3' psi-RNA subunit 3 | chr2:104451242+ | Intron | AUUAAA | 22.0% | 1.9 | No |
| Genome Version | mm9 | chr2:104451372+ | Intron | AAUAAA | 15.3% | 1.1 | Yes |
| Gene Type | protein-coding | chr2:104452239+ | Intron | AUUAAA | 10.2% | 1.6 | Yes |
| Chromosome | chr2 | chr2:104453382+ | Intron | AAUAAA | 11.9% | 1.3 | Yes |
| Transcription start site (NM_145529) | 194,439,841 | chr2:104453905+ | Intron | Other | 10.2% | 2.6 | No |
| Last polyA site (NM_145528) | 134,505,582 | chr2:104454030+ | Intron | AUUAAA | 3.4% | 3.9 | No |
| Last polyA site in PolyA_DB | 134,505,582 | chr2:104525204+ | 3'-most exon, 3'UTR (F) | None | 3.4% | 0.8 | No |
| Orthologs | Chick Human Rat | chr2:104525473+ | 3'-most exon, 3'UTR (M) | None | 11.9% | 1.0 | No |
| Links | UCSC Genome Browser | chr2:104525382+ | 3'-most exon, 3'UTR (L) | AAUAAA | 66.1% | 11.8 | Yes |

Figure 3.1 An example of search result from PolyA.DB 3.(A) Gene view. Mouse gene *Cstf3* is used as an example. The output includes a summary table of the gene as well as a link to UCSC genome browser. (B) PolyA SiteView. This table contains information of all individual PASs assigned to the queried gene and their links to UCSC genome browser.

APPENDIX A

SCRIPT FOR INDEX.PHP

```
< ?php include 'log.php'; ?>
<!DOCTYPE html>
<html>
<head>

    <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" >
<meta name="description" content="polyadenylation database" >
<meta name="keywords" content="polyadenylation, poly(A), polyA, database,
polyadb, polyA_DB, PAS, 3.UTR" >
<meta charset="utf-8" />
<title>PolyA_DB V.3</title>
<link rel="stylesheet" href="css/polya.style.css" type="text/css" />
<script>

function myFunction()
var x = document.getElementById("species").selectedIndex;
var y = document.getElementById("species").options;
var form = document.getElementById("form");
if(y[x].value == "Human")
form.action = "Human.php";

    else if (y[x].value == "Mouse")
form.action = "Mouse.php";

    else if (y[x].value == "Rat")
form.action = "Rat.php";
    else if (y[x].value == "Chicken")
form.action = "Chicken.php";

    return true;

</script>
</head>

    <body background="images/a.jpg" text="#000000" vlink="#660099" link="#3300ff"
bgcolor="#ffffff" >
<center>

    <table border=0 cellpadding=0 cellspacing=0, width="600" >
<tr>
<td valign=middle align=middle ><br><a href="http://polya-db.org"></a></td> </tr>
<tr><td align="middle" width="100%"><br><font size="2" color="#663300"
face="Verdana"><b>Release 3.1 (Sept 2017)</b></font>
```

```
</table>
```

```
<br>
```

```
<br>
```

```
<table border=1 cellpadding=6 cellspacing=0 width="600" bordercolor="#99CCFF"
align="middle">
```

```
<tr>
```

```
<td align="center">
```

```
<font color="#000066" size="2" face="Verdana"><b>Search PolyA_DB</b></font>
```

```
<br><br>
```

```
<form method="post" id="form" action="" onsubmit="myFunction()">
```

```
<table cellpadding="5">
```

```
<tr>
```

```
<td align="center">
```

```
<label style="padding-left:25px; padding-right:25px;"> <B><font SIZE=2 FACE="VERDANA">
Species:</font></B></label>
```

```
<select id="species" name="species" style="width:50%;height:5%;">
```

```
<option default_value="Human"><font SIZE=2 FACE="VERDANA">Human</font></option>
```

```
<option value="Mouse"><font SIZE=2 FACE="VERDANA">Mouse</font></option>
```

```
<option value="Rat"><font SIZE=2 FACE="VERDANA">Rat</font></option>
```

```
<option value="Chicken"><font SIZE=2 FACE="VERDANA">Chicken</font></option>
```

```
</select>
```

```
</td>
```

```
</tr>
```

```
<tr>
```

```
<td align="center">
```

```
<label><B><font SIZE=2 FACE="VERDANA">Enter Gene Symbol (e.g. 'Cstf3')
```

```
or Ensembl Gene ID (e.g. 'ENSG00000176102') or RefSeq Gene ID (e.g. '1479'):</font></B></label>
```

```
</td>
```

```
</tr>
```

```
<tr>
```

```
<td align="center">
```

```
<input type="text" name="Gene" id="Gene" value="Cstf3" style="width:50%;height:5%;text-align:center;" />
```

```
</td>
```

```
</tr>
```

```
</table>
```

```
<br>
```

```
<table align="center" cellpadding="10">
```

```
<tr>
```

```
<td align="center"><input style="min-width:200;" type="submit" value="OK"
name="submit" /></td>
```

```
<td align="center"><input style="min-width:200;" type="reset" class="res_but"
name="Reset" value="Reset" /></td>
```

```
</tr>
```

```
</table>
```

```

        </form>
</td>
</tr>

    <tr>

        <td>
<font color="#000066" size="2" face="Verdana"><b>Description:</font> </b>
<br>
<p><font SIZE=2 FACE="VERDANA">
PolyA_DB version 3.1 catalogs polyA sites (PASs) in genomes using deep sequencing
data. Gene annotations are mainly based on NCBI Gene and RefSeq databases, and
supplemented by the Ensembl database. Strand-specific RNA-seq data are used to
extend the 3' end of genes. PASs are identified using data from the 3' end region
extraction and sequencing (3'READS) method. Syntenic regions across human,
mouse and rat are used to annotate conserved PASs in mammals.
</font>
</p>
<table>
<tr>
<td><a href="misc/stat.php"><font SIZE=2 FACE="VERDANA">Database
Statistics</font></a></td>
</tr>
<tr>
<td><a href="misc/download.php"><font SIZE=2 FACE="VERDANA">Download
data</font></a></td>
</tr>

    <tr>
<td><a href="misc/new_release.php"><font SIZE=2 FACE="VERDANA">Information
about updates</font> </a></td>
</tr>

</table>

</td>
</tr>

    <tr>
<td>
<font color="#000066" size="2" face="Verdana"><b>References:</font><br><br>
<font SIZE=2 FACE="VERDANA">Wang, R., Nambiar, R., Zheng, D., & Tian,
B. (2017). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep
sequencing in multiple genomes. </font><a href="https://www.ncbi.nlm.nih.gov/pubmed/290694
SIZE=2 FACE="VERDANA">Nucleic Acids Research.</i></a></font><br>
<font SIZE=2 FACE="VERDANA">Hoque M*, Ji Z*, Zheng D, Luo W, Li
W, You B, Park JY, Yehia G, Tian B. (2013). Analysis of alternative cleavage
and polyadenylation by 3' region extraction and deep sequencing. </font><a
href="https://www.ncbi.nlm.nih.gov/pubmed/23241633"><i><font SIZE=2 FACE="VERDANA
Methods.10:133-9.</i></a></font><br>

```

Zheng D, Liu X, Tian B. (2016). 3'READS+, a sensitive and accurate method for 3' end sequencing of polyadenylated RNA. RNA.22:1631-9. <https://www.ncbi.nlm.nih.gov/pubmed/27512124>

Please contact rw479@rutgers.edu Wang / btian@rutgers.edu Bin Tian for questions/comments/suggestions

</BODY>
</HTML>

APPENDIX B

SCRIPT FOR CHICKEN.PHP

```
<?php include 'config/config.php';?>
<?php include 'Library/Database.php';?>
<?php include 'log.php'; ?>
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Chicken</title>
<link rel="stylesheet" href="css/polya_style.css" type="text/css" />

</head>
<body background="images/a.jpg" text="#000000" vlink="#660099" link="#3300ff"
bgcolor="#ffffff" >

<table border=0 cellpadding=0 cellspacing=0, width="600" >
<tr><td valign=top align=left ><a href="http://polya-db.org"></a></td></tr>
<tr><td align="left" valign="top" width="100%" ><br><font size="2" color="#663300"
face="Verdana"><b>Release 3.1 (Sept 2017)</b></font></td></tr> </table>
<br>
<a href="index.php"><font size="2" face="Verdana">Go back to the search
page</font></a>

<br>
<hr align=left width=725>
<br>
<?php
$db=new Database();

if (isset($_POST['species']))
$species= $_POST['species'];

if (isset($_POST['Gene']))
$gene= $_POST['Gene'];
if($gene=='NA'——$gene=='na')
die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></b>");

$sql = 'Select c1.PASid PASid,c1.chr chr,c1.Pos Pos,c1.strand strand,c1.pA_type
pA_type,c1.GeneSymbol GeneSymbol,c1.ext ext,c1.DB_conservation_level DB_conservation_level,c1.
genic_loc,c1.PAS_signal PAS_signal,c1.explable explable,c1.SumRPM SumRPM,c1.GeneName,c1.G
Gene_anno,c2.TXID,c2.start start,c2.end end,c2.TSS,c2.TES,c3.LastPAS FROM (chicken_pas2
c1 JOIN chicken_gene c2 ON c1.GeneID=c2.GeneID AND c1.EnsGeneID=c2.EnsGeneID)JOIN
chicken_lastPAS c3 ON c2.GeneID=c3.GeneID AND c2.EnsGeneID=c3.EnsGeneID
```

```
WHERE ';
```

```
    $sql_Conv='Select o.species species,o.GeneID GeneID,o.GeneSymbol GeneSymbol
from ortholog o WHERE HID = (SELECT distinct ortholog.HID FROM ortholog
JOIN chicken_gene ON (CASE WHEN chicken_gene.GeneID=NA THEN ortholog.GeneSymbol=chic
ELSE ortholog.GeneID=chicken_gene.GeneID END) WHERE ';
```

```
    if ( strval($gene) != strval(intval($gene)) )
    $str1=strpos($gene,'ENS',0);
    $str2=strpos($gene,'NM_',0);
    $str3=strpos($gene,'NR_',0);
```

```
    if($str1 === false && $str2===false && $str3===false)
    $sql=$sql.'c1.GeneSymbol='.$gene.'';
    $sql_Conv=$sql_Conv.'chicken_gene.GeneSymbol='.$gene.'');
```

```
    else if($str1 == 0)
    $sql=$sql.'c1.EnsGeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'chicken_gene.EnsGeneID='.$gene.'';
    else if($str2 == 0 ——— $str3 == 0)
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'chicken_gene.GeneID='.$gene.'');
```

```
    else if ( strval($gene) == strval(intval($gene)) )
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'chicken_gene.GeneID='.$gene.'');
```

```
    $result = $db->select($sql);
    $result_conv=$db->select($sql_Conv);
```

```
    ?>
    <?php if(mysqli_num_rows($result) > 0)
    while($r = mysqli_fetch_assoc($result))
    $arrow[]=$r;
    $len=count($arrow);
    else
    die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></B></font>");
    ?>
```

```
    <?php if(mysqli_num_rows($result_conv) > 0)
    while($r = mysqli_fetch_assoc($result_conv))
    $arrow_conv[]=$r;
```

```
    $len_conv=count($arrow_conv);
    ?>
```

```
    <font color="#000066" size=3 face="Verdana"><b>Gene Summary</b></font>
    <table id="displaygen" border=0 cellpadding=0 cellspacing=0 width=725 bordercolor="#99CCFF"
    <tr>
    <th align="left" width=33%><font FACE="VERDANA" SIZE=3>Species:</font></th>
```

```

<td align="left" width=67%><font FACE="VERDANA" SIZE=3>Chicken</font></td>
</tr>
<?php
$row=$arrow[0];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
_PADB3&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMo
$TSS=$row['TSS'];
$TES=$row['TES'];
$start=$row['start'];
$end=$row['end'];
$chr=$row['chr'];
$str=$str.$chr.'%3A'.$start.'-'.$end;
$genesymbol=$row['GeneSymbol'];
$genename=$row['GeneName'];
$geneID=$row['GeneID'];
$genetype=$row['Gene_anno'];
$txid=$row['TXID'];
$ensgeneid=$row['EnsGeneID'];
$lastPAS=$row['LastPAS'];
?>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Official Gene Symbol</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><i><?php echo $genesymbol;?></i></f
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>RefSeq Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $geneID;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Ensembl Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $ensgeneid;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Name</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genename;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Genome Version</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3>galGal4</font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Type</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genetype;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Chromosome</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $chr;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Transcription start site<br><?php
echo $txid;?></font></th>

```

```

<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TSS);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site<br>(<?php
echo $txid;?>)</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TES);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site in PolyA_DB</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($lastPAS);
?></font></td>

</tr>

<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Orthologs</font></th>
<td align="left">
<table id="pls">
<tr>

<?php for($x=0;$x < $len_conv;$x++)
$row=$arrow_conv[$x];
$spec=$row['species'];
if($spec=='human')?>
<td class="noBorder">
<form method="post" action="Human.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Human">
</form></td>
<?php ?>
<?php
if($spec=='mouse')?>
<td class="noBorder"> <form method="post" action="Mouse.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Mouse">
</form></td>
<?php ?>
<?php
if($spec=='rat')?>
<td class="noBorder">
<form method="post" action="Rat.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Rat">
</form></td>
<?php
?>
</tr>
</table>
</td>

```



```

</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Links</font></th>
<td align="left"><a href=?php echo $str?> target='_blank' ><font FACE="VERDANA"
SIZE=3>UCSC Genome Browser<font></a>
</td>
</tr>

</table>

<br>
<hr align=left width=725>
<br>

<font color="#000066" size=3 face="Verdana"><b>PolyA Site Summary</b></font>
<div style="overflow-x:auto;">
<table id="display" border=0 border-color="#99CCFF" cellpadding=0 cellspacing=0
width="725">
<tr>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS_ID</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Intron/exon
location</font></th>-->
<th width="30%" align="center"><font FACE="VERDANA" SIZE=3>PAS type</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Extension</font></th>
-->
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS Signal</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PSE</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Mean RPM</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Conserv.</font></th>
</tr>

<?php for($x=0;$x < $len;$x++)
$row_def=$arrow[$x];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
$start=$row_def['Pos']-200;
$end=$row_def['Pos']+200;
$str=$str.$row_def['chr']. '%3A'.$start.'-'. $end;
$pas=$row_def['PASid'];
?>
<tr>
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo " <a href=$str
target='_blank'> $pas </a>";?></font></td>
<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo $row_def['genic_loc'];?
-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php if($row_def['genic_loc']==$row._
echo $row_def['pA_type']; else echo $row_def['genic_loc'].', '.$row_def['pA_type'] ?></font></td>

<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo
$row_def['ext'];?></font></td>-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo $row_def['PAS_signal'];?></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $FOS=number_format(round(($row_
$FOS.'%';?></font></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $mean_RPM=number_format(rou
echo $mean_RPM;?></font></td>

```

<td align="center"><?php echo '-';?></td>
 </tr>
 <?php ?>
 </table>
 </div>
 <p>
 Click PAS.ID to view in UCSC Genome Browser</p>

 <hr align=left width=725>
 <div style="width:725px;overflow:auto">
 <p>Data Description</p>
 <p>PAS.ID: ID of each PAS,
 shown as chromosome:Position:Strand

 <!--Intron/exon location: PAS location relative to the splicing configura-
 tion, including 5'-most exon, internal exon, 3'-most exon, and intron. If a gene
 has only one exon, it is called single exon.
-->
 PAS type: PAS location in mRNA, including 5'UTR, CDS, and 3'UTR.
 For PASs in 3'UTRs, they are further divided into First (3'UTR (F)),Middle (3'UTR
 (M)), and Last (3'UTR (L)). If there is only one PAS in 3'UTR, it is called 3'UTR
 (S).

 <!--Extension: Whether PAS is located on the extended 3' end region,
 YES/NO
-->
 PAS Signal: The PAS signal is located within 40-nt upstream from the
 PAS, including
 AAUAAA, AUUAAA, Other AGTAAA, TATAAA, CATAAA, GATAAA, AATATA,
 AATACA, AATAGA,AAAAAG,ACTAAA), A-rich (AAAAAA), and None.

 PSE: Percentage of samples with expression (of all samples).

 Mean RPM: The average reads per million of PAS reads across all
 samples

 Conserv.: Whether the PAS is conserved in at least two mammals (human,
 mouse and rat).

 </p>
 </div>

 <hr align=left width=725>
 Please contact Ruijia Wang</a href="mailto:btian@rutgers.edu">Bin
 Tian for questions/comments/suggestions.
 </p>

 </body>
 </html>

APPENDIX C

SCRIPT FOR MOUSE.PHP

```
<?php include 'config/config.php';?>
<?php include 'Library/Database.php';?>
<?php include 'log.php'; ?>
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Mouse</title>
<link rel="stylesheet" href="css/polya_style.css" type="text/css" />

</head>
<body background="images/a.jpg" text="#000000" vlink="#660099" link="#3300ff"
bgcolor="#ffffff" >

<table border=0 cellpadding=0 cellspacing=0, width="600" >
<tr><td valign=top align=left ><a href="http://polya-db.org"></a></td></tr>
<tr><td align="left" valign="top" width="100%" ><br><font size="2" color="#663300"
face="Verdana"><b>Release 3.1 (Sept 2017)</b></font></td></tr> </table>
<br>
<a href="index.php"><font size="2" face="Verdana">Go back to the search
page</font></a>

<br>
<hr align=left width=725>
<br>
<?php
$db=new Database();

if (isset($_POST['species']))
$species= $_POST['species'];

if (isset($_POST['Gene']))
$gene= $_POST['Gene'];
if($gene=='NA'——$gene=='na')
die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></b>");

$sql = 'Select c1.PASid PASid,c1.chr chr,c1.Pos Pos,c1.strand strand,c1.pA_type
pA_type,c1.GeneSymbol GeneSymbol,c1.ext ext,c1.DB_conservation_level DB_conservation_level,c1.
genic_loc,c1.PAS_signal PAS_signal,c1.explable explable,c1.SumRPM SumRPM,c1.GeneName,c1.G
Gene_anno,c2.TXID,c2.start start,c2.end end,c2.TSS,c2.TES,c3.LastPAS FROM (mouse_pas2
c1 JOIN mouse_gene c2 ON c1.GeneID=c2.GeneID AND c1.EnsGeneID=c2.EnsGeneID)JOIN
mouse_lastPAS c3 ON c2.GeneID=c3.GeneID AND c2.EnsGeneID=c3.EnsGeneID
```

```
WHERE ';
```

```
    $sql_Conv='Select o.species species,o.GeneID GeneID,o.GeneSymbol GeneSymbol
from ortholog o WHERE HID = (SELECT distinct ortholog.HID FROM ortholog
JOIN mouse_gene ON (CASE WHEN mouse_gene.GeneID=NA THEN ortholog.GeneSymbol=mouse
ELSE ortholog.GeneID=mouse_gene.GeneID END) WHERE ';
```

```
    if ( strval($gene) != strval(intval($gene)) )
    $str1=strpos($gene,'ENS',0);
    $str2=strpos($gene,'NM_',0);
    $str3=strpos($gene,'NR_',0);
```

```
    if($str1 === false && $str2===false && $str3===false)
    $sql=$sql.'c1.GeneSymbol='.$gene.'';
    $sql_Conv=$sql_Conv.'mouse_gene.GeneSymbol='.$gene.'');
```

```
    else if($str1 == 0)
    $sql=$sql.'c1.EnsGeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'mouse_gene.EnsGeneID='.$gene.'';
    else if($str2 == 0 ——— $str3 == 0)
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'mouse_gene.GeneID='.$gene.'');
```

```
    else if ( strval($gene) == strval(intval($gene)) )
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'mouse_gene.GeneID='.$gene.'');
```

```
    $result = $db->select($sql);
    $result_conv=$db->select($sql_Conv);
```

```
    ?>
    <?php if(mysqli_num_rows($result) > 0)
    while($r = mysqli_fetch_assoc($result))
    $arrow[]=$r;
    $len=count($arrow);
    else
    die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></B></font>");
    ?>
```

```
    <?php if(mysqli_num_rows($result_conv) > 0)
    while($r = mysqli_fetch_assoc($result_conv))
    $arrow_conv[]=$r;
```

```
    $len_conv=count($arrow_conv);
    ?>
```

```
    <font color="#000066" size=3 face="Verdana"><b>Gene Summary</b></font>
    <table id="displaygen" border=0 cellpadding=0 cellspacing=0 width=725 bordercolor="#99CCFF"
    <tr>
    <th align="left" width=33%><font FACE="VERDANA" SIZE=3>Species:</font></th>
```

```

<td align="left" width=67%><font FACE="VERDANA" SIZE=3>mouse</font></td>
</tr>
<?php
$row=$arrow[0];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
_PADB3&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMo
$TSS=$row['TSS'];
$TES=$row['TES'];
$start=$row['start'];
$end=$row['end'];
$chr=$row['chr'];
$str=$str.$chr.'%3A'.$start.'-'.$end;
$genesymbol=$row['GeneSymbol'];
$genename=$row['GeneName'];
$geneID=$row['GeneID'];
$genetype=$row['Gene_anno'];
$txid=$row['TXID'];
$ensgeneid=$row['EnsGeneID'];
$lastPAS=$row['LastPAS'];
?>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Official Gene Symbol</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><i><?php echo $genesymbol;?></i></f
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>RefSeq Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $geneID;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Ensembl Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $ensgeneid;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Name</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genename;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Genome Version</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3>mm9</font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Type</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genetype;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Chromosome</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $chr;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Transcription start site<br><?php
echo $txid;?></font></th>

```

```

<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TSS);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site<br>(<?php
echo $txid;?>)</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TES);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site in PolyA_DB</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($lastPAS);
?></font></td>

</tr>

<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Orthologs</font></th>
<td align="left">
<table id="pls">
<tr>

<?php for($x=0;$x < $len_conv;$x++)
$row=$arrow_conv[$x];
$spec=$row['species'];
if($spec=='human')?>
<td class="noBorder">
<form method="post" action="Human.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Human">
</form></td>
<?php ?>
<?php
if($spec=='chicken')?>
<td class="noBorder"> <form method="post" action="chicken.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="chicken">
</form></td>
<?php ?>
<?php
if($spec=='rat')?>
<td class="noBorder">
<form method="post" action="Rat.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Rat">
</form></td>
<?php
?>
</tr>
</table>
</td>

```

```

</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Links</font></th>
<td align="left"><a href=?php echo $str?> target='_blank' ><font FACE="VERDANA"
SIZE=3>UCSC Genome Browser<font></a>
</td>
</tr>

</table>

<br>
<hr align=left width=725>
<br>

<font color="#000066" size=3 face="Verdana"><b>PolyA Site Summary</b></font>
<div style="overflow-x:auto;">
<table id="display" border=0 border-color="#99CCFF" cellpadding=0 cellspacing=0
width="725">
<tr>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS_ID</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Intron/exon
location</font></th>-->
<th width="30%" align="center"><font FACE="VERDANA" SIZE=3>PAS type</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Extension</font></th>
-->
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS Signal</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PSE</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Mean RPM</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Conserv.</font></th>
</tr>

<?php for($x=0;$x < $len;$x++)
$row_def=$arrow[$x];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
$start=$row_def['Pos']-200;
$end=$row_def['Pos']+200;
$str=$str.$row_def['chr']. '%3A'. $start. '-'. $end;
$pas=$row_def['PASid'];
?>
<tr>
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo " <a href=$str
target='_blank'> $pas </a>";?></font></td>
<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo $row_def['genic_loc'];?
-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php if($row_def['genic_loc']==$row._
echo $row_def['pA_type']; else echo $row_def['genic_loc'].', '. $row_def['pA_type'] ?></font></td>

<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo
$row_def['ext'];?></font></td>-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo $row_def['PAS_signal'];?></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $FOS=number_format(round(($row_
$FOS.'%';?></font></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $mean_RPM=number_format(round
echo $mean_RPM;?></font></td>

```

```

<td align="center"><font FACE="VERDANA" SIZE=3><?php echo '-';?></font></td>
</tr>
<?php ?>
</table>
</div>
<p><font SIZE=1 FACE="VERDANA">
Click PAS.ID to view in UCSC Genome Browser</font></p>
<br>
<hr align=left width=725>
<div style="width:725px;overflow:auto">
<p><font color="#000066" size=3 face="Verdana"><b>Data Description</b></font></p>
<p><font SIZE=2 FACE="VERDANA"><b>PAS.ID:</b> ID of each PAS,
shown as chromosome:Position:Strand<br>
<!--<b>Intron/exon location:</b> PAS location relative to the splicing configu-
ration, including 5'-most exon, internal exon, 3'-most exon, and intron. If a gene
has only one exon, it is called single exon.<br>-->
<b>PAS type:</b> PAS location in mRNA, including 5'UTR, CDS, and 3'UTR.
For PASs in 3'UTRs, they are further divided into First (3'UTR (F)),Middle (3'UTR
(M)), and Last (3'UTR (L)). If there is only one PAS in 3'UTR, it is called 3'UTR
(S).<br>
<!--<b>Extension:</b> Whether PAS is located on the extended 3' end region,
YES/NO<br>-->
<b>PAS Signal:</b> The PAS signal is located within 40-nt upstream from the
PAS, including
AAUAAA, AUUAAA, Other AGTAAA, TATAAA, CATAAA, GATAAA, AATATA,
AATACA, AATAGA,AAAAAG,ACTAAA), A-rich (AAAAAA), and None.<br>
<b>PSE:</b> Percentage of samples with expression (of all samples).<br>
<b>Mean RPM:</b> The average reads per million of PAS reads across all
samples<br>
<b>Conserv.:</b> Whether the PAS is conserved in at least two mammals (human,
mouse and rat).<br>
</font>
</p>
</div>
<br>
<hr align=left width=725>

<font color="#333300" size="1" face="Verdana"><b>Please contact <a
href="mailto:rw479@rutgers.edu">Ruijia Wang</a></a href="mailto:btian@rutgers.edu">Bin
Tian</a> for questions/comments/suggestions.</b></font>
</p>
<br>

</body>
</html>

```


APPENDIX D

SCRIPT FOR HUMAN.PHP

```
<?php include 'config/config.php';?>
<?php include 'Library/Database.php';?>
<?php include 'log.php'; ?>
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Human</title>
<link rel="stylesheet" href="css/polya_style.css" type="text/css" />

</head>
<body background="images/a.jpg" text="#000000" vlink="#660099" link="#3300ff"
bgcolor="#ffffff" >

<table border=0 cellpadding=0 cellspacing=0, width="600" >
<tr><td valign=top align=left ><a href="http://polya-db.org"></a></td></tr>
<tr><td align="left" valign="top" width="100%" ><br><font size="2" color="#663300"
face="Verdana"><b>Release 3.1 (Sept 2017)</b></font></td></tr> </table>
<br>
<a href="index.php"><font size="2" face="Verdana">Go back to the search
page</font></a>

<br>
<hr align=left width=725>
<br>
<?php
$db=new Database();

if (isset($_POST['species']))
$species= $_POST['species'];

if (isset($_POST['Gene']))
$gene= $_POST['Gene'];
if($gene=='NA'——$gene=='na')
die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></b>");

$sql = 'Select c1.PASid PASid,c1.chr chr,c1.Pos Pos,c1.strand strand,c1.pA_type
pA_type,c1.GeneSymbol GeneSymbol,c1.ext ext,c1.DB_conservation_level DB_conservation_level,c1.
genic_loc,c1.PAS_signal PAS_signal,c1.explable explable,c1.SumRPM SumRPM,c1.GeneName,c1.G
Gene_anno,c2.TXID,c2.start start,c2.end end,c2.TSS,c2.TES,c3.LastPAS FROM (human_pas2
c1 JOIN human_gene c2 ON c1.GeneID=c2.GeneID AND c1.EnsGeneID=c2.EnsGeneID)JOIN
human_lastPAS c3 ON c2.GeneID=c3.GeneID AND c2.EnsGeneID=c3.EnsGeneID
```

```
WHERE ';
```

```
    $sql_Conv='Select o.species species,o.GeneID GeneID,o.GeneSymbol GeneSymbol
from ortholog o WHERE HID = (SELECT distinct ortholog.HID FROM ortholog
JOIN human_gene ON (CASE WHEN human_gene.GeneID=NA THEN ortholog.GeneSymbol=hum
ELSE ortholog.GeneID=human_gene.GeneID END) WHERE ';
```

```
    if ( strval($gene) != strval(intval($gene)) )
    $str1=strpos($gene,'ENS',0);
    $str2=strpos($gene,'NM_',0);
    $str3=strpos($gene,'NR_',0);
```

```
    if($str1 === false && $str2===false && $str3===false)
    $sql=$sql.'c1.GeneSymbol='.$gene.'';
    $sql_Conv=$sql_Conv.'human_gene.GeneSymbol='.$gene.'');
```

```
    else if($str1 == 0)
    $sql=$sql.'c1.EnsGeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'human_gene.EnsGeneID='.$gene.'';
    else if($str2 == 0 ——— $str3 == 0)
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'human_gene.GeneID='.$gene.'');
```

```
    else if ( strval($gene) == strval(intval($gene)) )
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'human_gene.GeneID='.$gene.'');
```

```
    $result = $db->select($sql);
    $result_conv=$db->select($sql_Conv);
```

```
    ?>
    <?php if(mysqli_num_rows($result) > 0)
    while($r = mysqli_fetch_assoc($result))
    $arrow[]=$r;
    $len=count($arrow);
    else
    die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></B></font>");
    ?>
```

```
    <?php if(mysqli_num_rows($result_conv) > 0)
    while($r = mysqli_fetch_assoc($result_conv))
    $arrow_conv[]=$r;
```

```
    $len_conv=count($arrow_conv);
    ?>
```

```
    <font color="#000066" size=3 face="Verdana"><b>Gene Summary</b></font>
    <table id="displaygen" border=0 cellpadding=0 cellspacing=0 width=725 bordercolor="#99CCFF"
    <tr>
    <th align="left" width=33%><font FACE="VERDANA" SIZE=3>Species:</font></th>
```

```

<td align="left" width=67%><font FACE="VERDANA" SIZE=3>human</font></td>
</tr>
<?php
$row=$arrow[0];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
_PADB3&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMo
$TSS=$row['TSS'];
$TES=$row['TES'];
$start=$row['start'];
$end=$row['end'];
$chr=$row['chr'];
$str=$str.$chr.'%3A'.$start.'-'.$end;
$genesymbol=$row['GeneSymbol'];
$genename=$row['GeneName'];
$geneID=$row['GeneID'];
$genetype=$row['Gene_anno'];
$txid=$row['TXID'];
$ensgeneid=$row['EnsGeneID'];
$lastPAS=$row['LastPAS'];
?>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Official Gene Symbol</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><i><?php echo $genesymbol;?></i></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>RefSeq Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $geneID;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Ensembl Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $ensgeneid;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Name</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genename;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Genome Version</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3>hg19</font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Type</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genetype;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Chromosome</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $chr;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Transcription start site<br><?php
echo $txid;?></font></th>

```

```

<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TSS);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site<br>(<?php
echo $txid;?>)</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TES);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site in PolyA_DB</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($lastPAS);
?></font></td>

</tr>

<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Orthologs</font></th>
<td align="left">
<table id="pls">
<tr>

<?php for($x=0;$x < $len_conv;$x++)
$row=$arrow_conv[$x];
$spec=$row['species'];
if($spec=='mouse')?>
<td class="noBorder">
<form method="post" action="mouse.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="mouse">
</form></td>
<?php ?>
<?php
if($spec=='chicken')?>
<td class="noBorder"> <form method="post" action="chicken.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="chicken">
</form></td>
<?php ?>
<?php
if($spec=='rat')?>
<td class="noBorder">
<form method="post" action="Rat.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="Rat">
</form></td>
<?php
?>
</tr>
</table>
</td>

```

```

</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Links</font></th>
<td align="left"><a href=?php echo $str?> target='_blank' ><font FACE="VERDANA"
SIZE=3>UCSC Genome Browser<font></a>
</td>
</tr>

</table>

<br>
<hr align=left width=725>
<br>

<font color="#000066" size=3 face="Verdana"><b>PolyA Site Summary</b></font>
<div style="overflow-x:auto;">
<table id="display" border=0 border-color="#99CCFF" cellpadding=0 cellspacing=0
width="725">
<tr>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS_ID</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Intron/exon
location</font></th>-->
<th width="30%" align="center"><font FACE="VERDANA" SIZE=3>PAS type</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Extension</font></th>
-->
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS Signal</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PSE</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Mean RPM</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Conserv.</font></th>
</tr>

<?php for($x=0;$x < $len;$x++)
$row_def=$arrow[$x];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
$start=$row_def['Pos']-200;
$end=$row_def['Pos']+200;
$str=$str.$row_def['chr']. '%3A' . $start . '-' . $end;
$pas=$row_def['PASid'];
?>
<tr>
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo " <a href=$str
target='_blank'> $pas </a>";?></font></td>
<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo $row_def['genic_loc'];?
-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php if($row_def['genic_loc']==$row._
echo $row_def['pA_type']; else echo $row_def['genic_loc'].', '$row_def['pA_type'] ?></font></td>

<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo
$row_def['ext'];?></font></td>-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo $row_def['PAS_signal'];?></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $FOS=number_format(round(($row_
$FOS.'%';?></font></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $mean_RPM=number_format(round
echo $mean_RPM;?></font></td>

```

<td align="center"><?php echo '-';?></td>
 </tr>
 <?php ?>
 </table>
 </div>
 <p>
 Click PAS.ID to view in UCSC Genome Browser</p>

 <hr align=left width=725>
 <div style="width:725px;overflow:auto">
 <p>Data Description</p>
 <p>PAS.ID: ID of each PAS,
 shown as chromosome:Position:Strand

 <!--Intron/exon location: PAS location relative to the splicing configu-
 ration, including 5'-most exon, internal exon, 3'-most exon, and intron. If a gene
 has only one exon, it is called single exon.
-->
 PAS type: PAS location in mRNA, including 5'UTR, CDS, and 3'UTR.
 For PASs in 3'UTRs, they are further divided into First (3'UTR (F)),Middle (3'UTR
 (M)), and Last (3'UTR (L)). If there is only one PAS in 3'UTR, it is called 3'UTR
 (S).

 <!--Extension: Whether PAS is located on the extended 3' end region,
 YES/NO
-->
 PAS Signal: The PAS signal is located within 40-nt upstream from the
 PAS, including
 AAUAAA, AUUAAA, Other AGTAAA, TATAAA, CATAAA, GATAAA, AATATA,
 AATACA, AATAGA,AAAAAG,ACTAAA), A-rich (AAAAAA), and None.

 PSE: Percentage of samples with expression (of all samples).

 Mean RPM: The average reads per million of PAS reads across all
 samples

 Conserv.: Whether the PAS is conserved in at least two mammals (human,
 human and rat).

 </p>
 </div>

 <hr align=left width=725>
 Please contact Ruijia Wang</a href="mailto:btian@rutgers.edu">Bin
 Tian for questions/comments/suggestions.
 </p>

 </body>
 </html>

APPENDIX E

SCRIPT FOR RAT.PHP

```
<?php include 'config/config.php';?>
<?php include 'Library/Database.php';?>
<?php include 'log.php'; ?>
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8" />
<title>Rat</title>
<link rel="stylesheet" href="css/polya_style.css" type="text/css" />

</head>
<body background="images/a.jpg" text="#000000" vlink="#660099" link="#3300ff"
bgcolor="#ffffff" >

<table border=0 cellpadding=0 cellspacing=0, width="600" >
<tr><td valign=top align=left ><a href="http://polya-db.org"></a></td></tr>
<tr><td align="left" valign="top" width="100%" ><br><font size="2" color="#663300"
face="Verdana"><b>Release 3.1 (Sept 2017)</b></font></td></tr> </table>
<br>
<a href="index.php"><font size="2" face="Verdana">Go back to the search
page</font></a>

<br>
<hr align=left width=725>
<br>
<?php
$db=new Database();

if (isset($_POST['species']))
$species= $_POST['species'];

if (isset($_POST['Gene']))
$gene= $_POST['Gene'];
if($gene=='NA'——$gene=='na')
die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></b>");

$sql = 'Select c1.PASid PASid,c1.chr chr,c1.Pos Pos,c1.strand strand,c1.pA_type
pA_type,c1.GeneSymbol GeneSymbol,c1.ext ext,c1.DB_conservation_level DB_conservation_level,c1.
genic_loc,c1.PAS_signal PAS_signal,c1.explable explable,c1.SumRPM SumRPM,c1.GeneName,c1.G
Gene_anno,c2.TXID,c2.start start,c2.end end,c2.TSS,c2.TES,c3.LastPAS FROM (rat_pas2
c1 JOIN rat_gene c2 ON c1.GeneID=c2.GeneID AND c1.EnsGeneID=c2.EnsGeneID)JOIN
rat_lastPAS c3 ON c2.GeneID=c3.GeneID AND c2.EnsGeneID=c3.EnsGeneID WHERE
```

```

';

    $sql_Conv='Select o.species species,o.GeneID GeneID,o.GeneSymbol GeneSymbol
from ortholog o WHERE HID = (SELECT distinct ortholog.HID FROM ortholog
JOIN rat_gene ON (CASE WHEN rat_gene.GeneID=NA THEN ortholog.GeneSymbol=rat_gene.Gen
ELSE ortholog.GeneID=rat_gene.GeneID END) WHERE ' ;

    if ( strval($gene) != strval(intval($gene)) )
    $str1=strpos($gene,'ENS',0);
    $str2=strpos($gene,'NM_',0);
    $str3=strpos($gene,'NR_',0);

    if($str1 === false && $str2===false && $str3===false)
    $sql=$sql.'c1.GeneSymbol='.$gene.'';
    $sql_Conv=$sql_Conv.'rat_gene.GeneSymbol='.$gene.'';

    else if($str1 == 0)
    $sql=$sql.'c1.EnsGeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'rat_gene.EnsGeneID='.$gene.'';
    else if($str2 == 0 ——— $str3 == 0)
    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'rat_gene.GeneID='.$gene.'';

    else if ( strval($gene) == strval(intval($gene)) )

    $sql=$sql.'c1.GeneID='.$gene.'';
    $sql_Conv=$sql_Conv.'rat_gene.GeneID='.$gene.'';

    $result = $db->select($sql);
    $result_conv=$db->select($sql_Conv);

    ?>
    <?php if(mysqli_num_rows($result) > 0)
    while($r = mysqli_fetch_assoc($result))
    $arrow[]=$r;
    $len=count($arrow);
    else
    die("<B><FONT=5 FACE='VERDANA'>Please enter correct Gene and/or GeneID</FONT></B></font>");
    ?>

    <?php if(mysqli_num_rows($result_conv) > 0)
    while($r = mysqli_fetch_assoc($result_conv))
    $arrow_conv[]=$r;

    $len_conv=count($arrow_conv);
    ?>

    <font color="#000066" size=3 face="Verdana"><b>Gene Summary</b></font>
    <table id="displaygen" border=0 cellpadding=0 cellspacing=0 width=725 bordercolor="#99CCFF"
    <tr>
    <th align="left" width=33%><font FACE="VERDANA" SIZE=3>Species:</font></th>

```



```

<td align="left" width=67%><font FACE="VERDANA" SIZE=3>rat</font></td>
</tr>
<?php
$row=$arrow[0];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
_PADB3&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMo
$TSS=$row['TSS'];
$TES=$row['TES'];
$start=$row['start'];
$end=$row['end'];
$chr=$row['chr'];
$str=$str.$chr.'%3A'.$start.'-'.$end;
$genesymbol=$row['GeneSymbol'];
$genename=$row['GeneName'];
$geneID=$row['GeneID'];
$genetype=$row['Gene_anno'];
$txid=$row['TXID'];
$ensgeneid=$row['EnsGeneID'];
$lastPAS=$row['LastPAS'];
?>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Official Gene Symbol</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><i><?php echo $genesymbol;?></i></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>RefSeq Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $geneID;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Ensembl Gene ID</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $ensgeneid;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Name</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genename;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Genome Version</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3>rn5</font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Gene Type</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $genetype;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Chromosome</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo $chr;?></font></td>
</tr>
<tr>
<th align="left"><font FACE="VERDANA" SIZE=3>Transcription start site<br><?php
echo $txid;?></font></th>

```

```

<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TSS);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site<br>(<?php
echo $txid;?>)</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($TES);
?></font></td>
</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Last polyA site in PolyA_DB</font></th>
<td align="left"><font FACE="VERDANA" SIZE=3><?php echo number_format($lastPAS);
?></font></td>

</tr>

<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Orthologs</font></th>
<td align="left">
<table id="pls">
<tr>

<?php for($x=0;$x < $len_conv;$x++)
$row=$arrow_conv[$x];
$spec=$row['species'];
if($spec=='mouse')?>
<td class="noBorder">
<form method="post" action="mouse.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="mouse">
</form></td>
<?php ?>
<?php
if($spec=='chicken')?>
<td class="noBorder"> <form method="post" action="chicken.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="chicken">
</form></td>
<?php ?>
<?php
if($spec=='human')?>
<td class="noBorder">
<form method="post" action="human.php">
<input type="hidden" id="species" name="species" value=<?php echo $row['species']?>>
<input type="hidden" name="Gene" id="Gene" value=<?php echo $row['GeneSymbol']?>>
<input type="submit" value="human">
</form></td>
<?php
?>
</tr>
</table>
</td>

```

```

</tr>
<tr >
<th align="left"><font FACE="VERDANA" SIZE=3>Links</font></th>
<td align="left"><a href=?php echo $str?> target='_blank' ><font FACE="VERDANA"
SIZE=3>UCSC Genome Browser<font></a>
</td>
</tr>

</table>

<br>
<hr align=left width=725>
<br>

<font color="#000066" size=3 face="Verdana"><b>PolyA Site Summary</b></font>
<div style="overflow-x:auto;">
<table id="display" border=0 border-color="#99CCFF" cellpadding=0 cellspacing=0
width="725">
<tr>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS_ID</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Intron/exon
location</font></th>-->
<th width="30%" align="center"><font FACE="VERDANA" SIZE=3>PAS type</font></th>
<!--<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Extension</font></th>
-->
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PAS Signal</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>PSE</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Mean RPM</font></th>
<th width="10%" align="center"><font FACE="VERDANA" SIZE=3>Conserv.</font></th>
</tr>

<?php for($x=0;$x < $len;$x++)
$row_def=$arrow[$x];
$str='http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=
$start=$row_def['Pos']-200;
$end=$row_def['Pos']+200;
$str=$str.$row_def['chr']. '%3A'.$start.'-'. $end;
$pas=$row_def['PASid'];
?>
<tr>
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo " <a href=$str
target='_blank'> $pas </a>";?></font></td>
<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo $row_def['genic_loc'];?
-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php if($row_def['genic_loc']==$row._
echo $row_def['pA_type']; else echo $row_def['genic_loc'].', '.$row_def['pA_type'] ?></font></td>

<!--<td align="center"><font FACE="VERDANA" SIZE=3><?php #echo
$row_def['ext'];?></font></td>-->
<td align="center"><font FACE="VERDANA" SIZE=3><?php echo $row_def['PAS_signal'];?></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $FOS=number_format(round(($row
$FOS.'%';?></font></td>
<td align="right"><font FACE="VERDANA" SIZE=3><?php $mean_RPM=number_format(round
echo $mean_RPM;?></font></td>

```

```

<td align="center"><font FACE="VERDANA" SIZE=3><?php echo '-';?></font></td>
</tr>
<?php ?>
</table>
</div>
<p><font SIZE=1 FACE="VERDANA">
Click PAS.ID to view in UCSC Genome Browser</font></p>
<br>
<hr align=left width=725>
<div style="width:725px;overflow:auto">
<p><font color="#000066" size=3 face="Verdana"><b>Data Description</b></font></p>
<p><font SIZE=2 FACE="VERDANA"><b>PAS.ID:</b> ID of each PAS,
shown as chromosome:Position:Strand<br>
<!--<b>Intron/exon location:</b> PAS location relative to the splicing configu-
ration, including 5'-most exon, internal exon, 3'-most exon, and intron. If a gene
has only one exon, it is called single exon.<br>-->
<b>PAS type:</b> PAS location in mRNA, including 5'UTR, CDS, and 3'UTR.
For PASs in 3'UTRs, they are further divided into First (3'UTR (F)),Middle (3'UTR
(M)), and Last (3'UTR (L)). If there is only one PAS in 3'UTR, it is called 3'UTR
(S).<br>
<!--<b>Extension:</b> Whether PAS is located on the extended 3' end region,
YES/NO<br>-->
<b>PAS Signal:</b> The PAS signal is located within 40-nt upstream from the
PAS, including
AAUAAA, AUUAAA, Other AGTAAA, TATAAA, CATAAA, GATAAA, AATATA,
AATACA, AATAGA,AAAAAG,ACTAAA), A-rich (AAAAAA), and None.<br>
<b>PSE:</b> Percentage of samples with expression (of all samples).<br>
<b>Mean RPM:</b> The average reads per million of PAS reads across all
samples<br>
<b>Conserv.:</b> Whether the PAS is conserved in at least two mammals (rat,
rat and rat).<br>
</font>
</p>
</div>
<br>
<hr align=left width=725>

<font color="#333300" size="1" face="Verdana"><b>Please contact <a
href="mailto:rw479@rutgers.edu">Ruijia Wang</a></a href="mailto:btian@rutgers.edu">Bin
Tian</a> for questions/comments/suggestions.</b></font>
</p>
<br>

</body>
</html>

```

BIBLIOGRAPHY

- [1] Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nature Reviews. Molecular Cell Biology* 18(1), 18–30. <http://doi.org/10.1038/nrm.2016.116>
- [2] Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (New York, N.Y.)* 352(6291), aad9926. <http://doi.org/10.1126/science.aad9926>
- [3] Tian, B., & Graber, J. H. (2012). Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdisciplinary Reviews. RNA*, 3(3), 385–396. <http://doi.org/10.1002/wrna.116>
- [4] Jan, C. H., Friedman, R. C., Ruby, J. G., & Bartel, D. P. (2011). Formation, Regulation and Evolution of *Caenorhabditis elegans* 3UTRs. *Nature*, 469(7328), 97–101. <http://doi.org/10.1038/nature09616>
- [5] Haenni, S., Ji, Z., Hoque, M., Rust, N., Sharpe, H., Eberhard, R., Furger, A. (2012). Analysis of *C. elegans* intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3-end-seq. *Nucleic Acids Research*, 40(13), 6304–6318. <http://doi.org/10.1093/nar/gks282>
- [6] Graber, J. H., Cantor, C. R., Mohr, S. C., & Smith, T. F. (1999). In *silico* detection of control signals: mRNA 3-end-processing sequences in diverse species. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), 14055–14060.
- [7] Liu, X., Hoque, M., Larochelle, M., Lemay, J.-F., Yurko, N., Manley, J. L., Tian, B. (2017). Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*. *Genome Research*, 27(10), 1685–1695. <http://doi.org/10.1101/gr.222331.117>
- [8] Shi, Y. (2012). Alternative polyadenylation: New insights from global analyses. *RNA*, 18(12), 2105–2117. <http://doi.org/10.1261/rna.035899.112>
- [9] Tian, B., Pan, Z., & Lee, J. Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Research*, 17(2), 156–165. <http://doi.org/10.1101/gr.5532707>
- [10] Zhang, H., Lee, J. Y., & Tian, B. (2005). Biased alternative polyadenylation in human tissues. *Genome Biology*, 6(12), R100. <http://doi.org/10.1186/gb-2005-6-12-r100>
- [11] Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Burge, C. B. (2008). Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature*, 456(7221), 470–476. <http://doi.org/10.1038/nature07509>

- [12] Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S., & Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes & Development*, 27(21), 2380–2396. <http://doi.org/10.1101/gad.229328.113>