CPE 366-01
Lab4 Document

PolyDunkers

Team Members:
  Aaron Soo
  Chun Hui Pek
  Elush Shirazpour
  Ziyang Tan
**Entities:**

**Experiment (weak)**
- ID (Primary Key)
- eDate
- cell type
- from state
- to state

**Gene**
- gene name
- gene abbreviation (Primary Key)
- species (Primary Key)
- begin site
- end site
- chromosome

**Promoter Sequence**
- sequence

**Transcription Factor**
Name (Primary Key)

**T Number**
Num (Primary Key)

**Regulatory elements (weak)**
Beg (Primary Key)
Len (Primary Key)
Seq (Primary Key)
Sns (Primary Key)
MID (Primary Key)
MAC (Primary Key)
La
La/
L d
L q
L qv
S m
S c
S pv
P pv

**Promoter Sequence (weak)**
Promoter Sequence (Primary Key)

**Job Parameters**
JobNum (Primary Key)
Hierarchies
Aggregation: Gene and Experiment form an aggregation, and they are connected to job parameters by done_by.

Relationship Sets
done_by
experiment_on
binds
associate
ER Diagram

Change Log
- For the regulatory elements table, the elements gene_abbr, species, ex_id, beg, len, seq, and sns now have the UNIQUE constraint.
- Tables PSG and label have been removed
- Table transcription_factor has been created with element name
- Table t_number has been created with element num
- Table binds has been created with elements tf_name, t_num, gene_abbr, species, ex_id, beg, len, seq, sns, mid, mac.
Database Constraints

- A regulatory sequence string must match the substring of the promoter sequence for the same gene that starts at the same position indicated, and goes in the direction specified by the sense attribute.
- A regulatory sequence cannot start at a position if the position value is greater than the length of the promoter sequence.
- The sum of the position and the regulatory sequence cannot be greater than the length of the promoter sequence.
- The gene species must match the species the experiment was performed on.
- Chromosome value must be valid.
- FACs are only attribute that can be NULL.

DB Redesign Tables:

remark PolyDunkers

CREATE TABLE gene (  
gene_abbr VARCHAR2(10) UNIQUE,  
name VARCHAR2(50),  
species VARCHAR2(15) UNIQUE,  
Begin_site int,  
End_site int,  
chromosome int,  
PRIMARY KEY (gene_abbr, species) );

CREATE TABLE promoter_sequence (  
promoter_sequence VARCHAR2(2000) PRIMARY KEY,  
species VARCHAR2(15),  
gene_abbr VARCHAR2(10),  
constraint sid FOREIGN KEY (species) REFERENCES gene(species),  
constraint gid FOREIGN KEY (gene_abbr) REFERENCES gene(gene_abbr) );

CREATE TABLE experiment (  
id VARCHAR2(20) PRIMARY KEY,  
e_date DATE,  
cell_type VARCHAR2(20),  
species VARCHAR2(15),  
gene_abbr VARCHAR2(10),  
from_state VARCHAR2(25),  
to_state VARCHAR2(25),  
constraint ex_sid FOREIGN KEY (species) REFERENCES gene(species),  
constraint ex_gid FOREIGN KEY (gene_abbr) REFERENCES gene(gene_abbr) );

CREATE TABLE experiment_ongene_abbr VARCHAR2(10) PRIMARY KEY) REFERENCES gene(gene_abbr) );
gene_abbr VARCHAR2(10) UNIQUE,
species VARCHAR2(15) UNIQUE,
ex_id VARCHAR2(20) UNIQUE,
regulation VARCHAR2(10) UNIQUE,
constraint eo_gid FOREIGN KEY (gene_abbr) REFERENCES gene(gene_abbr),
constraint eo_sid FOREIGN KEY (species) REFERENCES gene(species),
constraint eo_eid FOREIGN KEY (ex_id) REFERENCES experiment(id),
PRIMARY KEY(gene_abbr, species, ex_id)
);

CREATE TABLE job_parameter (  
gene_abbr VARCHAR2(10),  
species VARCHAR2(30),  
ex_id VARCHAR2(20),  
JobNum VARCHAR2(30),  
storeLength VARCHAR2(20),  
email VARCHAR2(30),  
TRANSFACString INT,  
MyStieString INT,  
selected_1 INT,  
TRANSFAC_Matrices INT,  
IMD_Matrices INT,  
CBIL INT,  
JASPAR INT,  
myWeight INT,  
selected_2 INT,  
combine VARCHAR2(20),  
fac_attri VARCHAR2(20),  
matches VARCHAR2(20),  
user_core INT,  
tmm FLOAT,  
ts_a FLOAT,  
tw FLOAT,  
ta FLOAT,  
group_selection_1 VARCHAR2(20),  
td FLOAT,  
tc FLOAT,  
tm FLOAT,  
deficit INT,  
threshold FLOAT,  
Selected_3 INT,  
pseudo FLOAT,  
group_selection_2 VARCHAR2(20),  
distri VARCHAR2(20),  
bass VARCHAR2(20),  
constraint j_gid FOREIGN KEY (gene_abbr) REFERENCES experiment_on(gene_abbr),  
constraint j_sid FOREIGN KEY (species) REFERENCES experiment_on(species),  
constraint j_eid FOREIGN KEY (ex_id) REFERENCES experiment_on(ex_id),  
PRIMARY KEY(JobNum)
);
CREATE TABLE done_by (  
j_num VARCHAR2(30),
gene_abbr VARCHAR2(10),
species VARCHAR2(15),
ex_id VARCHAR2(20),
constraint db_jid FOREIGN KEY (j_num) REFERENCES job_parameter(JobNum),
constraint db_gid FOREIGN KEY (gene_abbr) REFERENCES experiment_on(gene_abbr),
constraint db_sid FOREIGN KEY (species) REFERENCES experiment_on(species),
constraint db_eid FOREIGN KEY (ex_id) REFERENCES experiment_on(ex_id),
PRIMARY KEY(j_num, gene_abbr, species, ex_id)
);

CREATE TABLE regulatory_elements (  
gene_abbr VARCHAR2(10) UNIQUE,
species VARCHAR2(15) UNIQUE,
ex_id VARCHAR2(20) UNIQUE,
beg int UNIQUE,
genlen int UNIQUE,
seq VARCHAR2(50) UNIQUE,
sns VARCHAR2(1) UNIQUE,
mid VARCHAR2(15) UNIQUE,
mac VARCHAR2(15) UNIQUE,
l_a FLOAT,
l_a2 FLOAT,
l_q FLOAT,
l_qv FLOAT,
sc FLOAT,
sm FLOAT,
spv FLOAT,
ppv FLOAT,
constraint r_gid FOREIGN KEY (gene_abbr) REFERENCES experiment_on(gene_abbr),
constraint r_sid FOREIGN KEY (species) REFERENCES experiment_on(species),
constraint r_eid FOREIGN KEY (ex_id) REFERENCES experiment_on(ex_id),
PRIMARY KEY(gene_abbr, ex_id, beg, len, seq, sns, mid, mac)
);

CREATE TABLE transcription_factor (  
name VARCHAR2(20) PRIMARY KEY
);

CREATE TABLE t_number (  
um VARCHAR2(20) PRIMARY KEY
);

CREATE TABLE associate (  
t_num VARCHAR2(20),
gene_abbr VARCHAR2(10),
species VARCHAR2(15),
constraint db_jid FOREIGN KEY (j_num) REFERENCES job_parameter(JobNum),
constraint db_gid FOREIGN KEY (gene_abbr) REFERENCES experiment_on(gene_abbr),
constraint db_sid FOREIGN KEY (species) REFERENCES experiment_on(species),
constraint db_eid FOREIGN KEY (ex_id) REFERENCES experiment_on(ex_id),
PRIMARY KEY(j_num, gene_abbr, species, ex_id)
);
```sql
CREATE TABLE regulatory_elements(
    ex_id VARCHAR2(20),
    beg int,
    len int,
    seq VARCHAR2(50),
    sns VARCHAR2(1),
    mid VARCHAR2(15),
    mac VARCHAR2(15),
    constraint a_mid FOREIGN KEY (mid) REFERENCES regulatory_elements(mid),
    constraint a_mad FOREIGN KEY (mac) REFERENCES regulatory_elements(mac),
    constraint a_gid FOREIGN KEY (gene_abbr) REFERENCES regulatory_elements(gene_abbr),
    constraint a_sid FOREIGN KEY (species) REFERENCES regulatory_elements(species),
    constraint a_eid FOREIGN KEY (ex_id) REFERENCES regulatory_elements(ex_id),
    constraint a_beg FOREIGN KEY (beg) REFERENCES regulatory_elements(beg),
    constraint a_len FOREIGN KEY (len) REFERENCES regulatory_elements(len),
    constraint a_seq FOREIGN KEY (seq) REFERENCES regulatory_elements(seq),
    constraint a_sns FOREIGN KEY (sns) REFERENCES regulatory_elements(sns),
    constraint a_num FOREIGN KEY (t_num) REFERENCES t_number(num),
    PRIMARY KEY(t_num, gene_abbr, ex_id, beg, len, seq, sns, mid, mac)
);

CREATE TABLE binds (    tf_name VARCHAR2(20),
    gene_abbr VARCHAR2(10),
    species VARCHAR2(15),
    ex_id VARCHAR2(20),
    beg int,
    len int,
    seq VARCHAR2(50),
    sns VARCHAR2(1),
    mid VARCHAR2(15),
    mac VARCHAR2(15),
    constraint b_mid FOREIGN KEY (mid) REFERENCES regulatory_elements(mid),
    constraint b_mad FOREIGN KEY (mac) REFERENCES regulatory_elements(mac),
    constraint b_gid FOREIGN KEY (gene_abbr) REFERENCES regulatory_elements(gene_abbr),
    constraint b_sid FOREIGN KEY (species) REFERENCES regulatory_elements(species),
    constraint b_eid FOREIGN KEY (ex_id) REFERENCES regulatory_elements(ex_id),
    constraint b_beg FOREIGN KEY (beg) REFERENCES regulatory_elements(beg),
    constraint b_len FOREIGN KEY (len) REFERENCES regulatory_elements(len),
    constraint b_seq FOREIGN KEY (seq) REFERENCES regulatory_elements(seq),
    constraint b_sns FOREIGN KEY (sns) REFERENCES regulatory_elements(sns),
    constraint b_tf FOREIGN KEY (tf_name) REFERENCES transcription_factor(name),
    PRIMARY KEY(tf_name, gene_abbr, ex_id, beg, len, seq, sns, mid, mac)
);
remark PolyDunkers

drop table experiment cascade constraints;
drop table gene cascade constraints;
drop table promoter_sequence cascade constraints;
drop table job_parameter cascade constraints;
drop table done_by cascade constraints;
drop table regulatory_elements cascade constraints;
drop table experiment_on cascade constraints;
drop table transcription_factor cascade constraints;
drop table t_number cascade constraints;
drop table associate cascade constraints;
drop table binds cascade constraints;