

README v1.1

11/1/2007

README v1.2

1/22/2008

2/8/2008

7/15/2008

2/23/2009

README v1.3

7/8/2009

By Everett Zhou

Changes: for README v1.2

- 1. Load Single Experiment: has to load all arrays for one experiment.**
- 2. Successfully loaded Files will be archived.**
- 3. Validation for a complete experiment has been implemented.**
- 4. Implementation of Retrieval of Raw Data Files function**
- 5. Implementation of Edit SOFT Content function**
- 6. Implementation of Export Metadata function**
- 7. Implementation of List Experiment function**

Changes: for README v1.3

- 1. Rearrangement of all functions.**
- 2. Implementation of ChIP-seq metadata registration and export.**

Loading experiments into UNC database includes two steps:

First step: To transfer files to the server incoming directory. Please see “Step-by-step instructions for uploading files to the UNC servers “at <http://wormencode.pbwiki.com/UNC+Database>

Second step: To load experiments into UNC database.

This document is for the second step.

To access the web site, please enter:

<https://genome.unc.edu:8443/nimblegen>

For more information, please refer to

<http://wormencode.pbwiki.com/UNC+Database>

We strongly recommend Firefox, rather than Microsoft IE, since IE has certificate validation issue.

The application functions can be grouped into five parts:
i.e. User Management, Loading Data, Experiment Management, Searching and Retrieving Data, and Submitting Data to GEO and DCC.

For each part, different roles have different sets of functions. We will discuss all functions based on the administration role.

- **User Management**

For admin, the nimblegen project has the following functions:

Create User Group, Create User Account, and Change User Password functions are very straightforward.

Admin Role

User Management

[Create User Group](#)

[Create User Role](#)

[Create User Account](#)

[Change User Password](#)

Loading Data

[Load Design File](#)

[Load Single Experiment](#)

[Load Batch Experiment](#)

[Replace GFF File](#)

[Load Single Expression Experiment](#)

Experiment Management

[Delete Experiment](#)

[List ChIP-chip Experiment](#)

Searching and Retrieving Data

[Retrieve Raw Data File](#)

[Export ChIP-chip Metadata](#)

[Export ChIP-seq Metadata](#)

Metadata Registration and Editing

[Edit ChIP-chip Metadata](#)

[Register ChIP-seq GEO Metadata for each Experiment](#)

[Register ChIP-seq GEO Metadata for Series](#)

- **Loading Data**

Data Loading functions include Load Design File, Load Single Experiment, and Replace GFF File.

Load Design File

Design File Already in DB: tells you how many design files you or someone else have loaded. It is composed of four parts, separated by “----”. The first part is design name, the second is design ID, then design file and finally the person who submitted.

UNC Design Name: * It is required. The conventional naming is to use part of the design file name, but without _1or _2 etc. so that design name is easy to be associated with a particular design.

Please **DO NOT** leave any space for **UNC Design Name**.

Design File: * It is required. **YOU HAVE** to load a complete design at one time.

There are four methods to create a SOFT file. If a SOFT file exists, then no need to create another one. This is the default choice. If a SOFT template file is already created, then just select one. The SOFT will copy the content, but is renamed as “design file name_SOFT.txt”. This rule applies to both cases where the SOFT file is located on either server machine or local PC.

If no SOFT file is available, you need to create a new one. After you click the button “Create SOFT file now”, a long list of items is displayed. For easy matching, the new SOFT file is named as “design file name_SOFT.txt”.

For a complete design, one SOFT file is needed.

Note: The displayed contents for selection are the abbreviated ones for nice looking. The content which is written into SOFT file is different from what you can see. To make sure what is exactly written into the file, please reference @
<http://wormencode.pbwiki.com/Protocols%20%28new%29?doneLogin=28e51030cff9d7d649a8fd04434fae44101818a1>

Load Design File.

Design Files Already in DB:

UNC Design Name: *

Design Files Available in You Home Directory: *

You need to select one of three methods to create SOFT format file.

No, a SOFT file already exists

Select SOFT file from server

Upload SOFT file from local Machine

Create SOFT file now

platform_manufacturer

platform_manufacture_protocol

platform_catalog_number

platform_web_link

platform_support

platform_coating

platform_description

platform_contributor

platform_pubmed_id

platform_geo_accession

platform_table_begin

platform_table_end

If you clicked Submit button, nothing is running, then you need to scroll down to see any warnings. Most likely, you have not selected or entered a required field.

If you are not familiar with SOFT, here is a brief introduction.

There are a few methods for data center to accept data. SOFT (Simple Omnibus Format in Text) is one of methods for GEO to store data. SOFT includes three parts: Platform, Sample and Series. Platform and sample correspond to our design and experiment (sample). Series is just to put things together. Since SOFT is straightforward (compared to MAGE-TAB, MAGE-ML) and supports MIAME-compliant data submissions, we chose it as the method to submit data to GEO. Please see the following link.

<http://www.ncbi.nlm.nih.gov/projects/geo/info/soft2.html>.

Currently, we only catch most of meta data related to SOFT and we need to edit and bundle samples and designs together for submission to GEO.

Load Single Experiment

This is both admin and user function.

Your Previously Submitted Experiments: how many experiments you have submitted. It is composed of four parts, separated by “---“. The first part is the experiment name, the second is design name in that experiment, and the third is the file name, and finally the person who submitted.

New Experiment Name: * It is required. An experiment defined here includes one to a few sub-arrays, each of which is part of a complete design

Here is the naming convention: [antibody name]_[strain]_[stage]_[replicate number]_[option]. Here is the example: AB1791_H3_GLP1TS_AD_1_EVERETT

Please **DO NOT** leave any space for the **Experiment Name**.

Pair Files: * It is required. You need to select files from two different channels.

Channel Comment: It is optional.

Dye Swap: Default is no.

GFF Files: * It is required. You need to select files which correspond to the selected pair file.

With regards to creation of SOFT file, the same rule applies as for Load Design File function. But the SOFT file is named as “experiment name_SOFT.txt”. For a complete experiment, one copy of SOFT file is needed.

Note: Please do not enter any “=” symbols for those metadata (SOFT) items.

After you click the button “Submit”, the single experiment info is loaded into the job queue. And it is displayed as the process is finished whatever it is successful or not.

The jobs in the queue will be loaded into data tables later and you will receive email for each experiment. The email informs you whether a particular experiment is loaded into data table successfully or not. If it fails, please forward the email content to us (either to informaticshelp@unc.edu or xlzhou@email.unc.edu).

Note: You have to select **ALL** arrays for a particular experiment. For example, design 2006-05-08_C_elegans_SXGA has two arrays: 2006-05-08_C_elegans_SXGA_1 and 2006-05-08_C_elegans_SXGA_2. Pair files 86659_532.pair and 86659_635.pair and GFF file 86659_635_ratio.gff correspond to the design 2006-05-08_C_elegans_SXGA_1; Pair files 89828_532.pair and 89828_635.pair and GFF file 89828_635_ratio.gff correspond to the design 2006-05-08_C_elegans_SXGA_2. Therefore, you need to select 86659_532.pair and 89828_532.pair from Green Channel; 86659_635.pair and 89828_635.pair from Red Channel, and 86659_635_ratio.gff and 89828_635_ratio.gff from GFF files.

Load Single Experiment.

Your Existing Experiment Files		All Existing Experiment Files	
<div style="border: 1px solid #ccc; padding: 2px;"><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_532.pair—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_635.pair—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_635_ratio.gff—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_2—89950_532.pair—xlzhou</p></div>		<div style="border: 1px solid #ccc; padding: 2px;"><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_532.pair—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_635.pair—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_1—86677_635_ratio.gff—xlzhou</p><p>UNC_LIEB_CWHITTLE_N2_EMBRYO_HTZ-1_CHIP_2—2006-05-08_C_elegans_SXGA_2—89950_532.pair—xlzhou</p></div>	
New Experiment Name: * <input type="text"/>			
Here is the naming convention: [location]_[lab]_[user]_[strain]_[condition (stage/temp)]_[experimental protein]_[experiment]_[replicate number].			
Here is the example: UNC_lieb_cwhittle_N2_embryo_HTZ-1_CHIP_1			
Green Channel		Red Channel	
Pair Files: *	<div style="border: 1px solid #ccc; padding: 2px;"><p>11553501_532.pair</p><p>86659_532.pair</p><p>86677_532.pair</p><p>86678_532.pair</p></div>		<div style="border: 1px solid #ccc; padding: 2px;"><p>11553501_635.pair</p><p>86659_635.pair</p><p>86677_635.pair</p><p>86678_635.pair</p></div>
Channel Comment:	<input type="text"/>		<input type="text"/>
Dye Swap: <input type="radio"/> Yes <input checked="" type="radio"/> No			
GFF Files: *	<div style="border: 1px solid #ccc; padding: 2px;"><p>11553501_635_ratio.gff</p><p>86659_635_ratio_new.gff</p><p>86677_635_ratio.gff</p><p>86678_635_ratio.gff</p></div>		

Replace GFF File

When user loads data into UNC DB first time, the GFF file could be generated by NimbleGen Inc. Later, the data analyst may generate a better normalized GFF file. This function comes into play to replace the old GFF file with the new GFF file.

You need to select a particular experiment first, then select its GFF file, select new GFF file, finally select the new protocol.

Please note that you ONLY can replace one GFF file each time. Therefore, if you have three GFF files for a particular experiment, you need to do three times.

- **Experiment Management**

Delete Experiment

Users only can delete his/her own experiments. It provides the confirmation and it is queued .

List ChIP-chip Experiment

It lists all ChIP-chip experiments in the UNC DB. It displays who, when each experiment was submitted, from which lab, the design associated with each experiment, and whether the experiment has been submitted to DCC.

The handy part of this function is that you can sort based on each column.

- **Searching and Retrieving Data**

Retrieve Raw Data File

The registered users can retrieve all experiments, including the user's own experiments and other user's experiments. It is your responsibility that you should let the owner of the experiment know that you download their experiments.

The metadata for the downloaded experiments will also be exported together.

Export ChIP-chip Metadata

For ChIP-chip experiments, if the user just wants to see the metadata for some experiments, then use this function.

Export ChIP-seq Metadata

There are two types of metadata for exporting, single experiment or series, which consists of one to many single experiments. Therefore, you need to select either experiment or series button first.

- Metadata Registration and Editing

Edit ChIP-chip Metadata

You need to click the button to select one experiment file first.

The first column is the SOFT item;

The second column is the protocol or comment you selected for the channel 1;

The third column is the available protocol you can select or an empty box for you to enter the content for the channel 1;

The fourth column is the protocol or comment you selected for the channel 2;

The fifth column is the available protocol you can select or an empty box for you to enter the content for the channel 2.

Note: Please do not enter any “=” symbols for those metadata (SOFT) items.

Register ChIP_seq GEO Metadata for each Experiment

This is the first step to register ChIP-seq GEO metadata.

Registration of ChIP-seq GEO metadata is modeled based on GEO data submission requirements. It is a lot of similar to the Load Single Experiment function. The form consists of three part: samples, protocols and platform. Users need to select all required fields. I have loaded the names of all sequence data sets acquired from UNC sequencing facility into UNC DB. Users only need to select a particular experiment. But there are three big differences between this function and Load Single Experiment function.

1. The data file is not loaded into UNC DB, only metadata is loaded into UNC DB;
2. The data file is not archived;
3. The job is not queued.

If the sequence data was acquired from other sequencing facilities, please let the project know this.

Register ChIP_seq GEO Metadata for Series

Based on GEO data submission requirements, all single experiments will be included in the series. A series consists of those experiments which share all the same platform parameters and share the same overall design. At this moment, we are not sure how many experiments can go into the same series, so that the network can tolerate.

For the regular users, the UNC nimblegen project has the less functions, but the ways to use those functions are exactly same as Admin Role.