
seqinspector Documentation

Release 1.0 stable

Marcin Piechota

January 25, 2014

1 Indices and tables

3

Table of contents

1.1 Seqinspector

seqinspector.cremag.org

1.1.1 What is seqinspector?

The seqinspector tool was designed to provide a computational service that utilizes up-to-date ChIP-seq data. Seqinspector allows to study the functional enrichments of user-defined DNA regions. This includes putative regulators of co-expressed genes.

1.1.2 Contents

- *Tutorials*
- *User interface*
- *How to use it?*
- *FAQ (Frequently asked questions)*

1.1.3 Tutorials

Seqinspector SRF tutorial

Contents

- *Background*
- *List of genes*
- *Submit the list of genes*
- *Inspect results*

Background

Rodriguez-Parkitna et al. (2010) analyzed the consequences of SRF loss on the regulation of activity-dependent transcription in the striatum of Srf D1Cre animals after treatment with cocaine, a powerful activator of dopamine signaling. Using microarray profiling, they found that loss of SRF caused a specific and complete lack of induction of Egr1, Egr2, and Egr4 transcripts.

Figure 1.1: Heatmap presenting genes which induction is affected by Srf knock-out (Rodriguez-Parkitna et al., (2010)

List of genes

Here is the list of genes selected from the article. These genes are suspected to be regulated by SRF transcription factor.

```
Egr1
Egr2
Egr4
Arl4d
Rgs2
```

Submit the list of genes

You can copy/paste the above list of genes into the input area and press “submit” or just use the link: [submit query](#)

Inspect results

To inspect results press the “statistics” button. A table with results will be shown below input field. The most over-represented track is “SRF_02”. The average coverage in selected gene promoters is 0.74, while in the reference it is 0.078. This is 9.5 time more in query set then genome average. T test p value is 8.8E-63. Therefore, the result is very significant (even after Bonferroni correction, 1.9E-60). This means that SRF is highly over-represented in your query set. Therefore, this transcription factor might be key controlling factor of this list of genes. This is in agreement with expectations.

You can press “description” button in the first result line. You will see GEO submission number (GSM530190). You can get more information about this track there. Some other basic informations are also included (e.g. cell type).

If you press “stack plot” button, you will see distribution of coverages in reference and your query set.

It is also possible too see coverages for particular genes, by pressing the “show genes” button.

1.1.4 User interface

User interface consist of four parts. On the top of the page there is a navigation bar with some internal links to help or other tools. On the left there is an input part in which you can select genome assembly, input your query set and submit it. Right from it there is a list of submitted query sets. One of them is selected as background for statistics. On the bottom of the page is table with results sorted by p-value.

The screenshot shows the Seqinspector web interface. At the top, there is a navigation bar with links for Home, Seqinspector-one, Help, Contact, and News. Below this is a header area with the Seqinspector logo and a dropdown menu for the genome assembly, currently set to 'Mus Musculus (mm9)'. A green 'Submit' button is next to the dropdown. Below the dropdown is a text area for entering queries, containing several genomic coordinates. To the right of the text area is a list of query sets, each with a 'statistics' button. Below the text area are three buttons: 'Options', 'Example', and 'Help'. The main content area is a table with 11 columns: Track name, Query, Background, Fold diff, P value, Bonferroni, Stack plot, Histogram, Genes, and Description. The table contains 8 rows of data. Red circles with numbers 1 through 6 are overlaid on the image to highlight specific UI elements: 1 points to the genome assembly selector, 2 to the query input field, 3 to the Submit button, 4 to the 'statistics' button in the query sets list, 5 to the navigation bar, and 6 to the first row of the results table.

Track name	Query	Background	Fold diff	P value	Bonferroni	Stack plot	Histogram	Genes	Description
SRF_02	1.8	0.078	23.0	5.0E-163	1.1E-160	Stack plot	Histogram	Show genes	Description
SRF_01	1.1	0.076	15.0	7.1E-148	1.6E-145	Stack plot	Histogram	Show genes	Description
SRF_04	1.7	0.074	22.0	8.3E-137	1.8E-134	Stack plot	Histogram	Show genes	Description
SRF_03	1.2	0.07	17.0	7.8E-132	1.7E-129	Stack plot	Histogram	Show genes	Description
SRF_05	1.3	0.042	30.0	4.6E-109	1.0E-106	Stack plot	Histogram	Show genes	Description
E2F1_03	0.98	0.26	3.8	4.8E-13	1.1E-10	Stack plot	Histogram	Show genes	Description
NFKB1_03	1.2	0.39	3.0	2.9E-7	6.4E-5	Stack plot	Histogram	Show genes	Description

Figure 1.2: Seqinspector user interface. (1) genome assembly selector, (2) text field for query input, (3) submit button, (4) list of query sets, (5) navigation bar with internal links, (6) results table

1.1.5 How to use it?

Step 1: Prepare your list

You should prepare your input list in one of the following formats.

bed

```
chr16      30254029      30255503
chr4 49268883      49272093
chr1 145968637      145969897
...
```

genomic coordinates

```
chr10:66999617-67001617
chr18:35019861-35021861
chr14:70476252-70478252
...
```

gene symbols

```
Egr1
Egr2
Fos
...
```

ensembl transcript ids

```
ENSMUST00000165033
ENSMUST00000145936
ENSMUST00000140525
...
```

refseq mRNA ids (no subversions)

```
NM_007913
NM_010118
NM_010234
...
```

Step 2: Choose genome assembly

You are allowed to choose between Homo sapiens (hg19) and Mus musculus (mm9, mm10)

Step 3: Insert your query list

The optimal list length is about 50 items. Seqinspector was tested with up to 1000 items lists.

Step 4: Submit and wait

Press the submit button. Seqinspector will automatically convert gene symbols or transcripts into genomic intervals 2000 bp long around transcription start sites. If a gene has more than one start site, all will be used. If some genomic intervals are overlapping they are merged. Seqinspector will then compute coverage from all available tracks for your query set of genomic intervals. Calculation progress will be shown on the screen.

Step 5: Inspect list of query sets

After pressing the submit button your query will appear on the list under unique name (“Set_1” if it is your first submission). Additionally there is already preloaded reference set (1000 random promoters). Statistics is always computed in comparison to reference. Therefore, one of your uploaded queries might be set as a reference by pressing “set as reference” button. You can remove or rename any of your queries.

Step 6: Calculate statistics

After pressing the “statistics” button table with results will be presented on the bottom of the screen. The tracks are sorted by p-value of significance. The columns are:

- Track name - it is internal id of a track, that contains short name of a transcription factor
- Query - average coverage of query set
- Background - average coverage of reference set
- Fold diff - fold difference between query and reference
- P value - significance of difference between query and reference datasets (calculated by t-test)
- Bonferroni - Bonferroni corrected p-value
- Stack plot - heat stacked plots presenting distribution of coverages in all query sets with respective p-values
- Histogram - visualisation of average coverage (2000 bp around center of genomic interval) for all query sets
- Genes - genomic intervals in query sets, symbols for nearest genes and coverage for these intervals
- Description - description of a track

Step 7: Inspect your results

You can visualize your results by pressing “stack plot” or “histogram” buttons. You can also inspect which genes have the highest coverage of a particular track by pressing “show genes” button. From “show genes” dialog you can navigate to seqinspector-one tool to inspect individual gene.

Step 8: Change options

Change database. In seqinspector there are two databases: (1) Mus musculus and (2) Homo sapiens. You can inspect Human tracks with your murine genomic coordinates. Mouse coordinates will be translated into Human coordinates using liftover tool. Extend query range. If 1000 bp upstream and downstream from tss is too small for you, it is possible to expand query range by using this parameter. It is only possible for queries using gene symbols and transcript ids.

1.1.6 FAQ (Frequently asked questions)

1.2 Seqinspector-one

seqone.cremag.org

1.2.1 What is seqinspector-one?

The seqinspector-one was designed to provide a computational service that utilizes up-to-date ChIP-seq data. Seqinspector-one allows to study the functional enrichment of single user-defined gene or DNA region.

1.2.2 Contents

- *User interface*
- *How to use it?*
- *FAQ (Frequently asked questions)*

1.2.3 User interface

User interface consist of four parts. On the top of the page there is a navigation bar with some internal links to help or other tools. On the left there is an input part in which you can select genome assembly, input your gene name or genomic range and submit it. Right from it there is a list of transcription start sites if a gene name was submitted. On the bottom of the page is table with results sorted by p-value.



Seqinspector-one user interface. The first row consists of: genome assembly selector, gene symbol (or genomic range) input, number of transcription start sites infobox, transcription start site selector. The second row consists of text area with analysed genomic range, number of significant overrepresentation tracks and number of tracks with tendency. The third row consists of manipulation buttons: move left, zoom out, zoom in, move right, options, submit (calculate), move to bottom (to statistics table). Plot in the middle shows selected (by default top 10 sorted by p-value) ChIP-seq tracks for selected promoter or genomic range. On the bottom there is statistics table.

1.2.4 How to use it?

Step 1: Prepare your list

You should prepare your input in one of the following formats.

bed

```
chr16 30254029 30255503
```

genomic coordinates

```
chr10:66999617-67001617
```

gene symbols

```
Egr1
```

ensembl transcript ids

```
ENSMUST00000165033
```

refseq mRNA ids (no subversions)

```
NM_007913
```

Step 2: Choose genome assembly

You are allowed to choose between Homo sapiens (hg19) and Mus musculus (mm9, mm10)

Step 3: Insert your query

If you will choose to insert gene symbol there is autocomplete mechanism that starts to work after first two letters. If there are more than one transcription start sites for submitted gene user can choose between them in the selector box.

Step 4: Submit and wait

There is autosubmission mechanism in seqinspector-one. However, in case that something will not work you can press “compute” green button. Seqinspector will automatically convert gene symbols or transcripts into genomic intervals 2000 bp long around transcription start sites. Seqinspector will then compute coverage from all available tracks for your query and will compare to genome average.

Step 5: Inspect image

Image shows selected tracks and ensembl gene annotation in the selected query range. Selected tracks are presented as coverage histograms (with minimum value of 2 on the Y axis). You can choose tracks for visualisation in the results table (see below).

Step 6: Inspect results

After pressing the “Show statistics” button table with results will be presented on the bottom of the screen. The tracks are sorted by p-value of significance. The columns are:

- Track name - it is internal id of a track, that contains short name of a transcription factor
- Query - coverage for query range
- Background - average coverage of precomputed reference set
- Fold diff - fold difference between query and reference
- P value - significance of difference between query and reference datasets (calculated by z-score)
- Bonferroni - Bonferroni corrected p-value
- Stack plot - heat stacked plots presenting distribution of coverages in all query sets with respective p-values
- Description - description of a track

Step 7: Change options

Change database. In seqinspector there are two databases: (1) Mus musculus and (2) Homo sapiens. You can inspect Human tracks with your murine genomic coordinates. Mouse coordinates will be translated into Human coordinates using liftover tool.

1.2.5 FAQ (Frequently asked questions)

License

seqinspector is freely available under a GNU Public License (Version 2).

Contact

piechota . marcin [at] gmail . com