- Clustering and Alignment of ChIP profiles

User Manual

Revised June 20, 2011

by Fiona Nielsen

What is it?

CATCH is a cool new tool for clustering of ChIP profile patterns

Why would I use it?

To unravel the mysteries of epigenetics, you first have to identify the epigenetic patterns!

How do I use it?

1. Collect or generate **ChIP profiling data**

You probably already did this...

2. Identify your **genomic positions of interest**

Which sites do you want to analyse?

3. **CATCH** your ChIP profiles!

And let CATCH do alignment and clustering for you

The CATCH program

- **Unsupervised clustering** of ChIP profiles
- Input
  - A dataset of profiles, i.e. specific genome regions from your ChIP profiling experiment
    - ChIP signal (wiggle format)
    - Selected positions (bed format)
- Output
  - A clustering+alignment of all profiles from the dataset

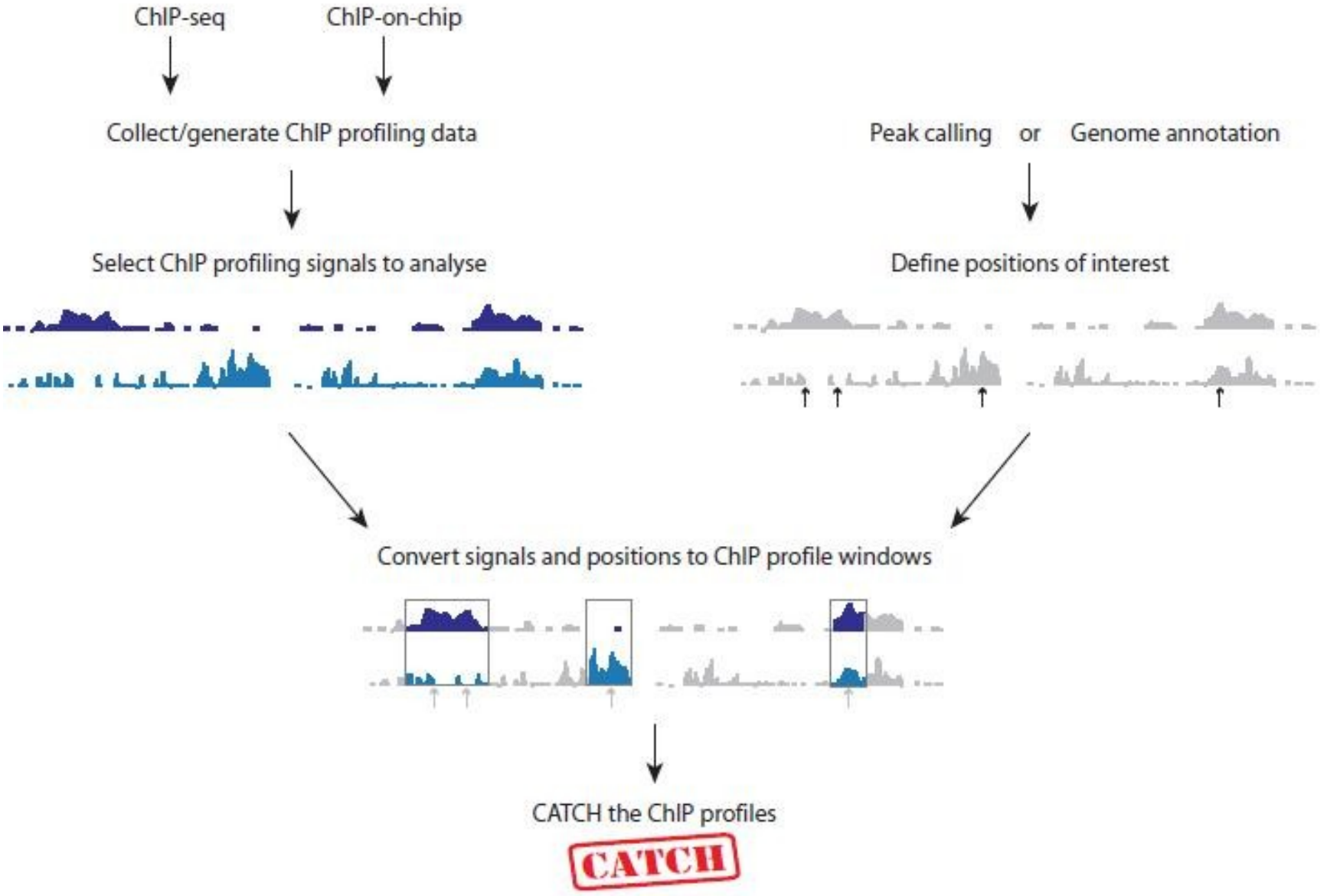
The CATCH workflow

(small datasets)

1) Import profiles
2) View profiles
3) Start CATCH algorithm
4) Browse patterns in clustering tree
5) Export interesting profiles for further analysis

The import procedure has more steps when working with large datasets (1000+ profile regions) ... →

The CATCH workflow



The CATCH workflow

(large datasets)

1) Generate profile wiggle files (pwig)
   - → Load and view dataset in CATCHprofiles
2) Make json input file from pwig files
3) Start CATCH algorithm from commandline
   - → Generating json output file
4) Load json result in CATCHprofiles

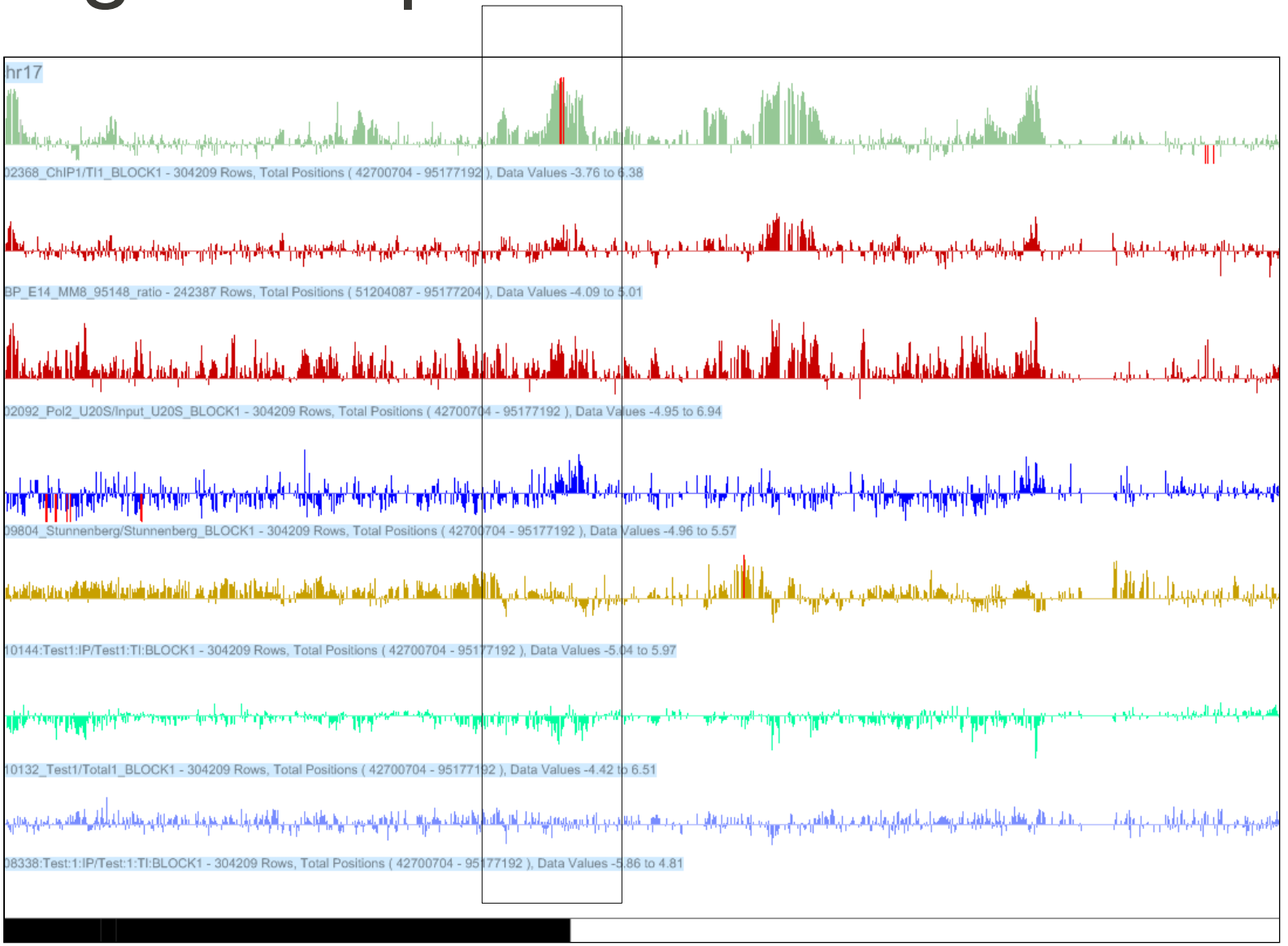
Profile regions

- You decide which areas are interesting to you:
  - Regions of annotation (TSS, CPG islands etc)
  - Peak regions
  - Other

Example workflow

1. make a bed file of your regions of interest

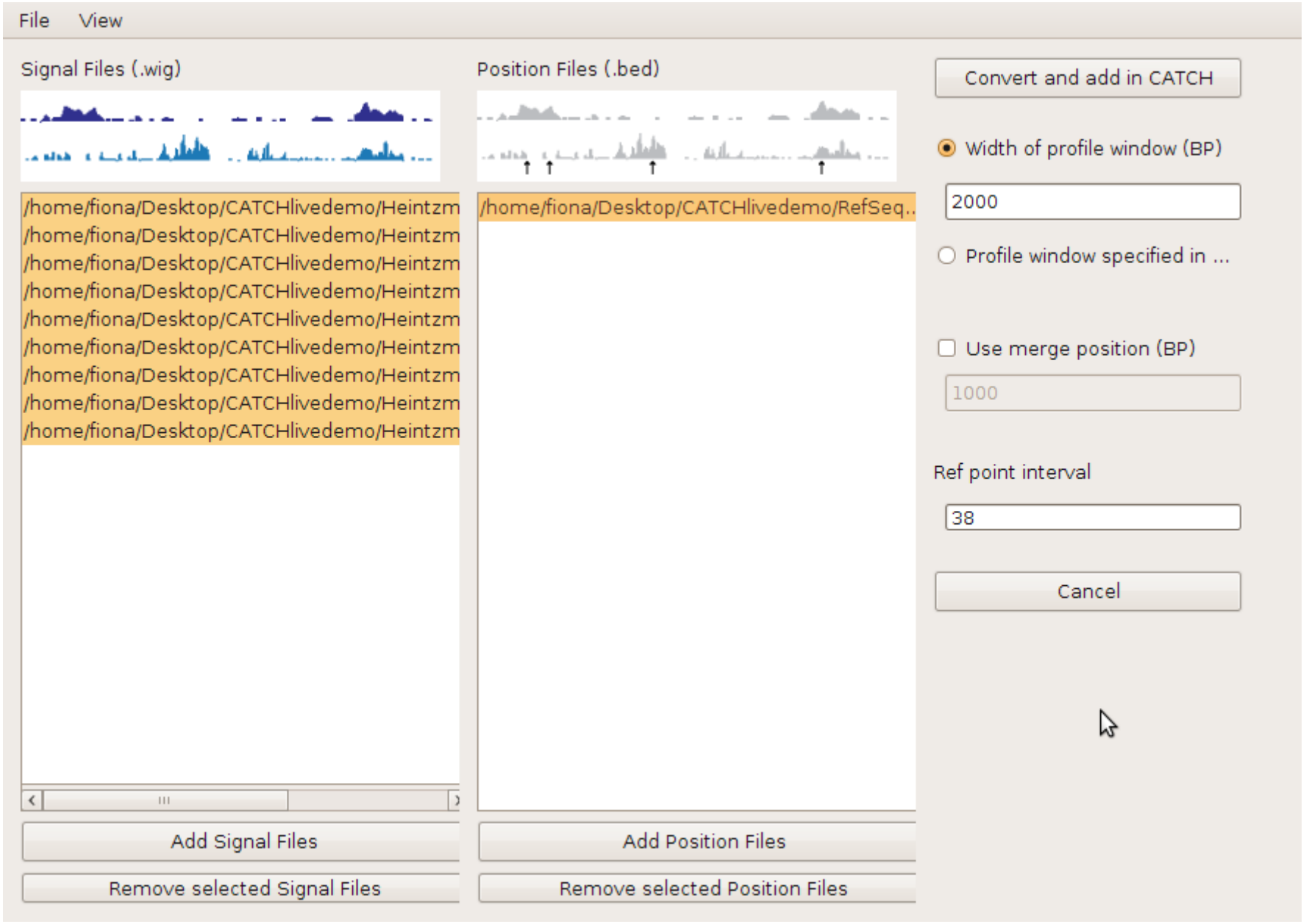
e.g. TSS profiles in ChIP-on-chip



TSS

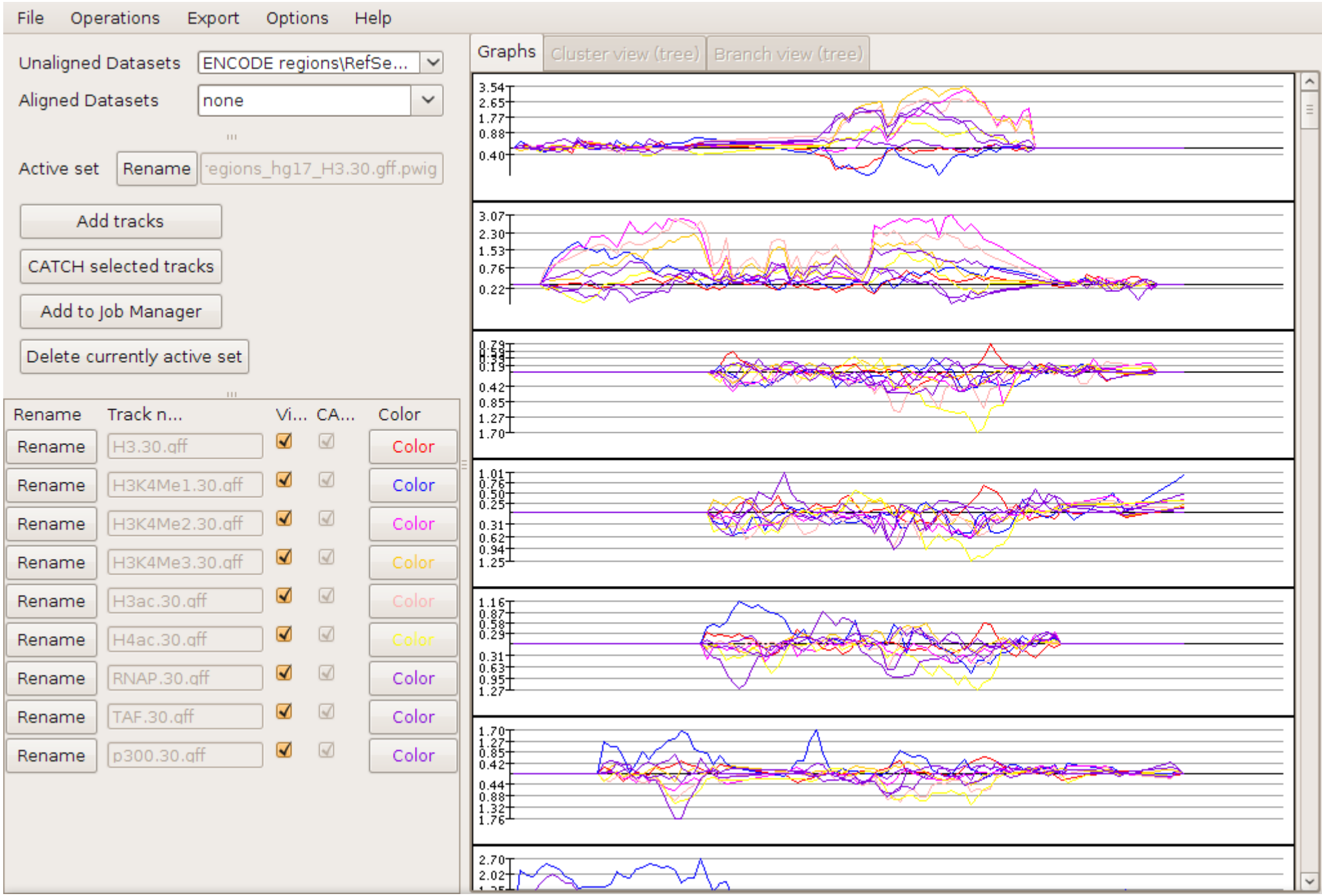
Import to CATCH

2. import your wiggle files and bed file using File → 'Import Profiles'



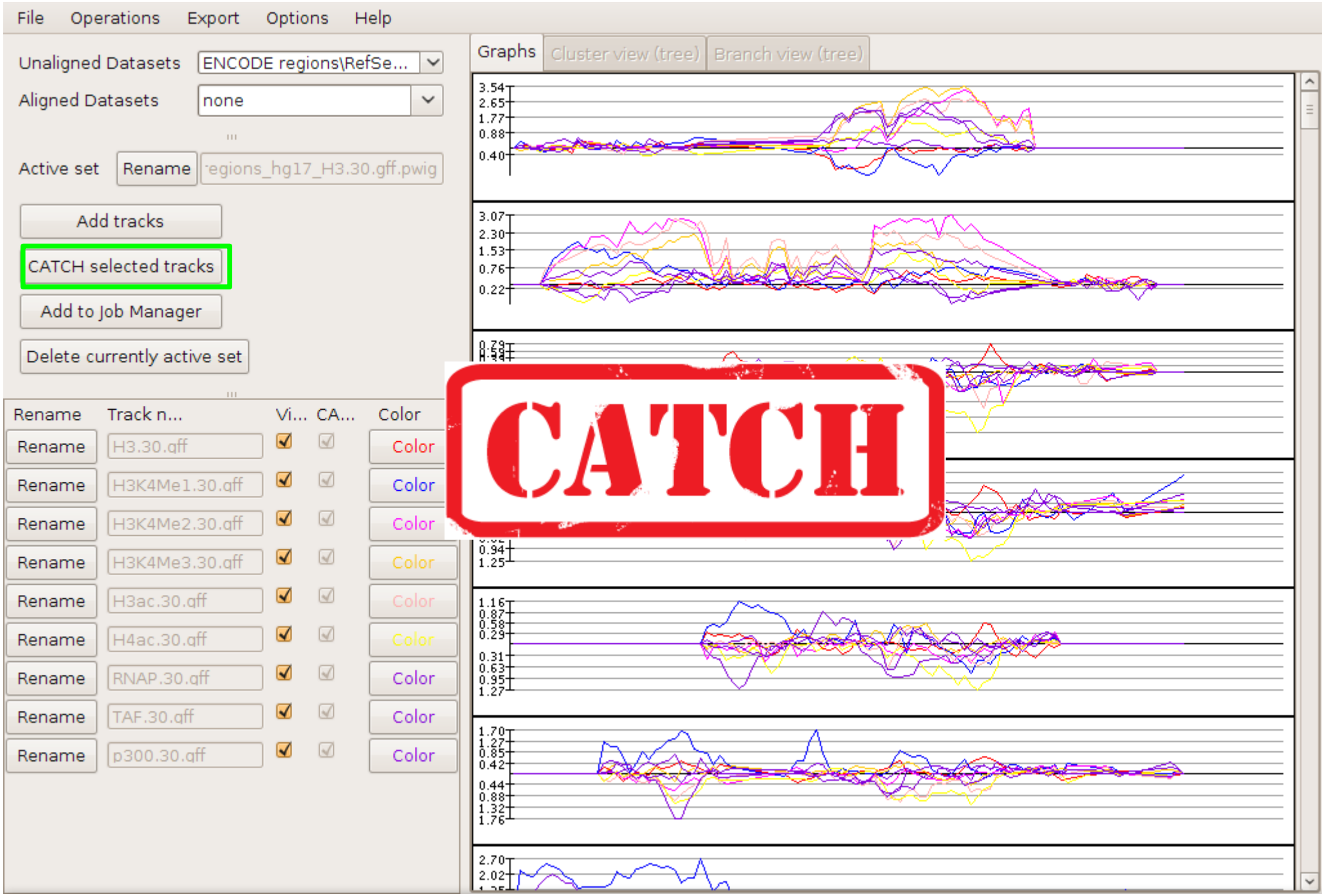
View profiles

3. Examine your data

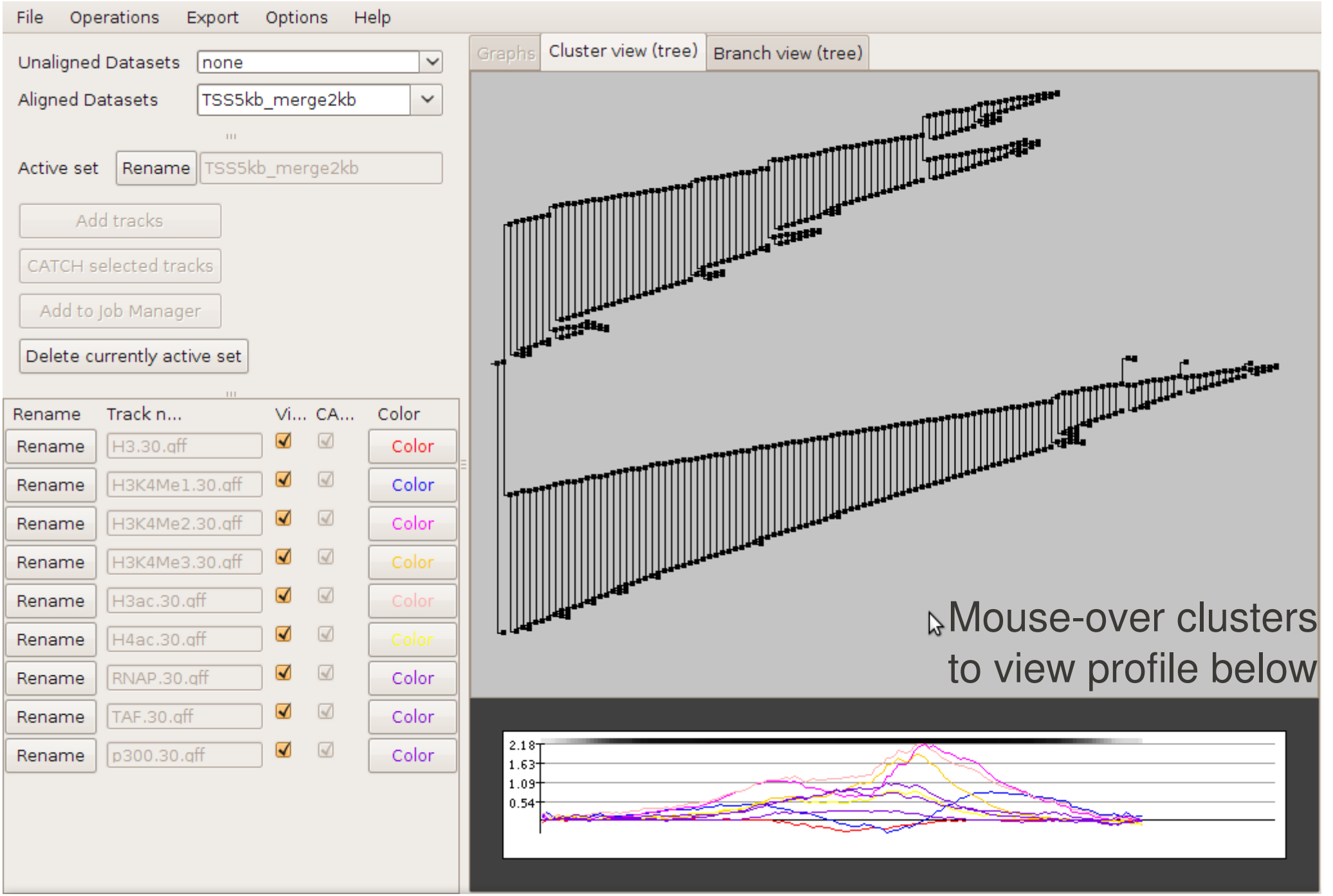


Start CATCH

4. Click 'CATCH selected tracks' and wait..



Examine profile clusters 4. Browse results interactively



interactive browsing

- Browse clustering tree:
  - Zoom: + and –
  - Collapse subtree: ctrl + leftclick
  - Highlight: leftclick
- Right-click to:
  - Open subtree
  - Export results

Save & Export

- CATCH clustering result (.catch)
  - To browse later or share results
- Use the export menu or right click to export:
  - Profile patterns (.csv)
  - Positions (.bed)
  - Cluster tree (.newick)

Quick Start Guide

1. Ensure JAVA 1.6 is installed
   (follow link from http://catch.cmbi.ru.nl)
2. Prepare your data
   - ChIP profiling tracks in wiggle format
   - Your chosen peaks/positions in .bed format
3. Start CATCH
4. Import files or load dataset
5. CATCH your ChIP profiles!

Java memory issues

- Run CATCH from the commandline:
  - `Java -Xmx1000m -Xss2000k -jar CATCH3v513.jar`
- If you get errors of insufficient Java memory:
  - Heapsize: increase -Xmx value
  - Stacksize: increase -Xss value
- If 'Import files' fails on big files:
  - Use script to generate datasets (.pwig)
  - Use script to generate CATCH input file (.json)

Big datasets?

- Use scripts to generate input files in JSON format
- Execute CATCH from the commandline
  - **`./execute.sh catch jobin.json jobout.json`**
  - Where second argument is the name of the executable

    win: `catch`

    linux: `catch-linux-static-x86`

    mac: `catch-macOSX`
- Load json result in CATCHprofiles

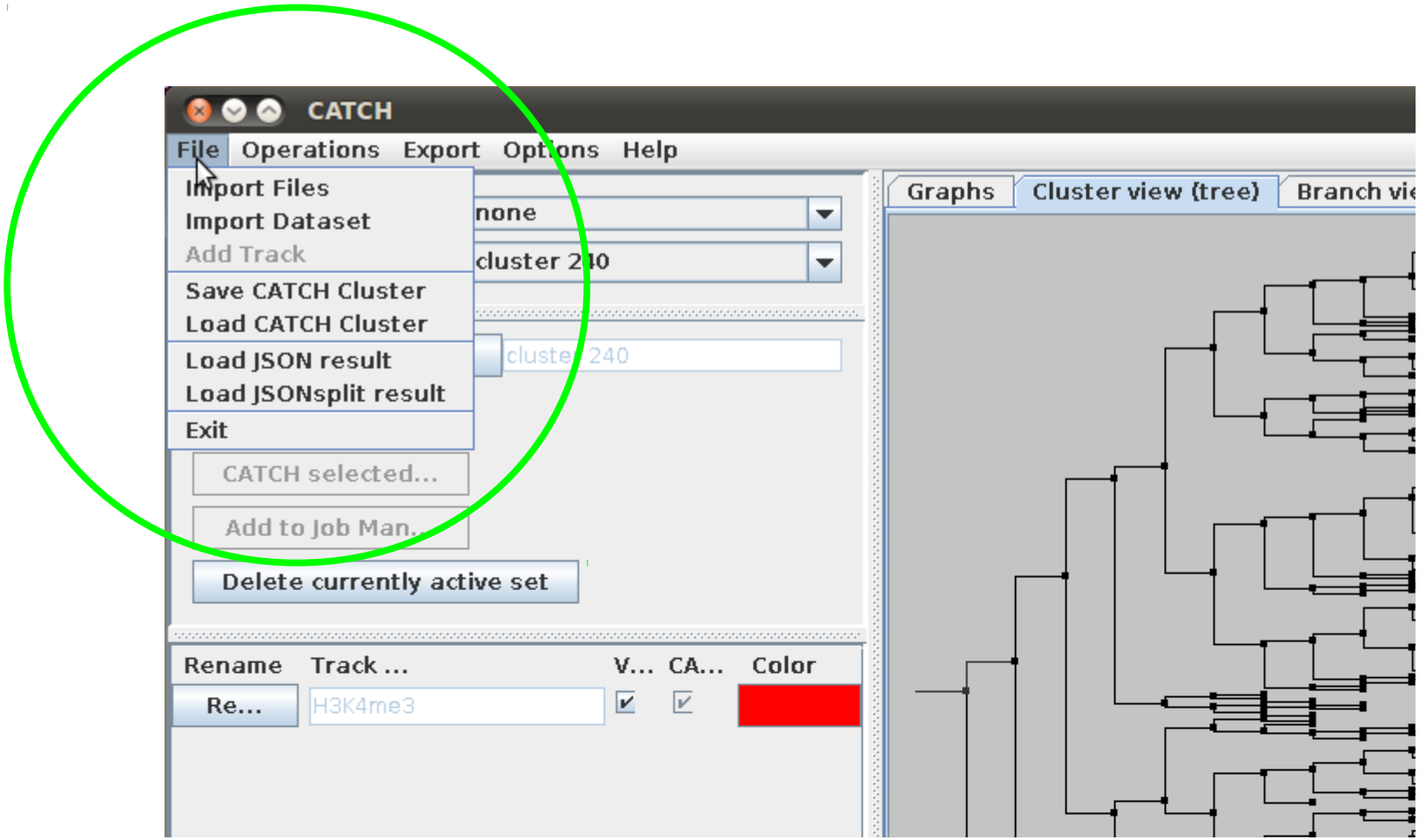
More information

- Go to the CATCH homepage:
  http://catch.cmbi.ru.nl
  - User manual, examples and links

- Updates?
  - Subscribe to: CATCH-updates@bioinformatics.org
- Questions?
  - Send them to: CATCH-users@bioinformatics.org

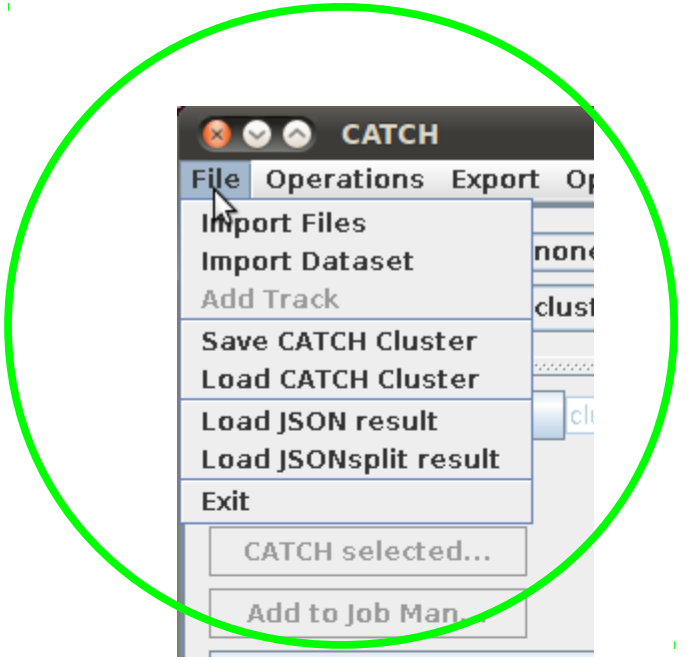
CATCH

Load files



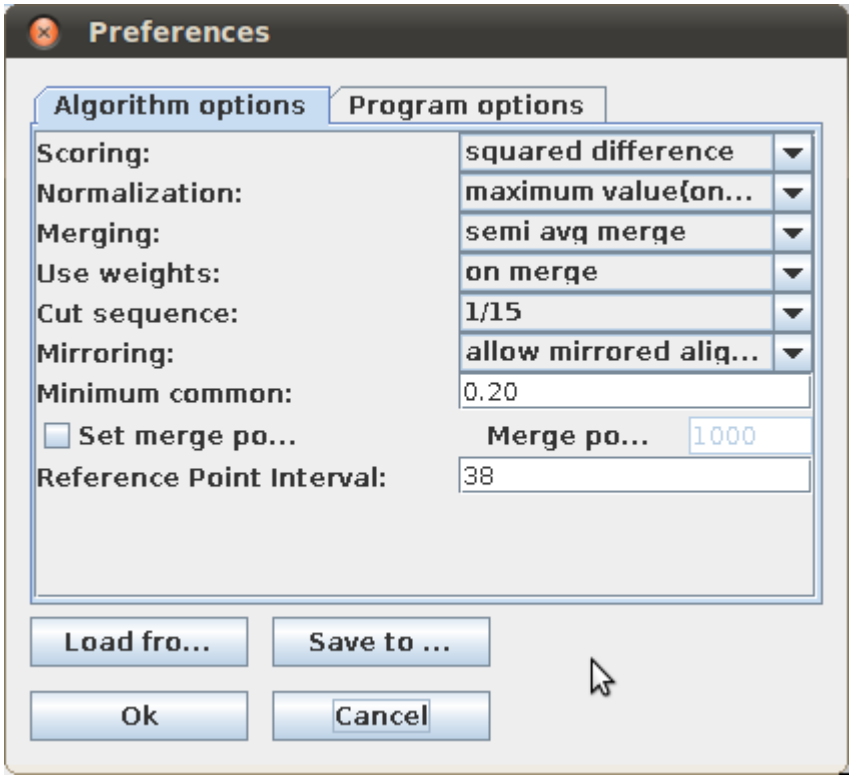
File formats

- Import files: **wiggle** and **bed**
- Import dataset: **pwig**
- CATCH cluster: **catch**
- JSON result: **json**



Change settings

- Check and save your default settings before starting the CATCH clustering
- Settings are written to the json-formatted input file before executing the CATCH core algorithm



Source code repository

- The CATCH source code is maintained in an SVN repository (http://subversion.tigris.org/) hosted by bioinformatics.org
- To browse the repository and the code, you can use the SVN web interface by pointing your browser to:

  http://www.bioinformatics.org/websvn/listing.php?repname=catch
- To download the source code from the CATCH svn, you can use anonymous checkout (after installing SVN on your system):

```
svn checkout svn://bioinformatics.org/svnroot/catch/branches/CATCH\
3/trunk/CATCH\ 3/
```

Compile executables

- The catch core algorithm (catch-engine) and the CATCHprofiles user interface are compiled separately
- To compile the catch core algorithm, download the “catch-engine” branch from svn and compile use the accompanying makefile:

```
catch:~/catch-engine> make

gcc-4.3 -static  catch.o JSON_parser.o ConvertUTF.o -lm -lgomp
-lpthread -o catch-linux-static-x86
```

- The CATCHprofiles Java GUI can be compiled using a development tool like Eclipse (http://www.eclipse.org/):
  - Import the “CATCH 3” directory from the “CATCH 3/trunk” branch from CATCH svn
  - Export the main program `programControllers.CATCH.java` as a “runnable jar file”

- For any further questions concerning CATCH development, please send your question to the CATCH mailinglist: CATCH-dev@bioinformatics.org

Acknowledgements

IMADA, Odense:

- **Kasper Markus**
  first implementation of CATCH
- **Lene Favrholdt**

Nijmegen:

- **Moniek Riemersma**
- GiP**CATCH** team
- **Maarten Kooyman**
  The PinkThing

Supervisors:

Henk Stunnenberg
Molecular Biology

Martijn Huynen
CMBI

Acknowledgements

CATCH users:

Hendrik Marks

Rasa Jurgelenaite

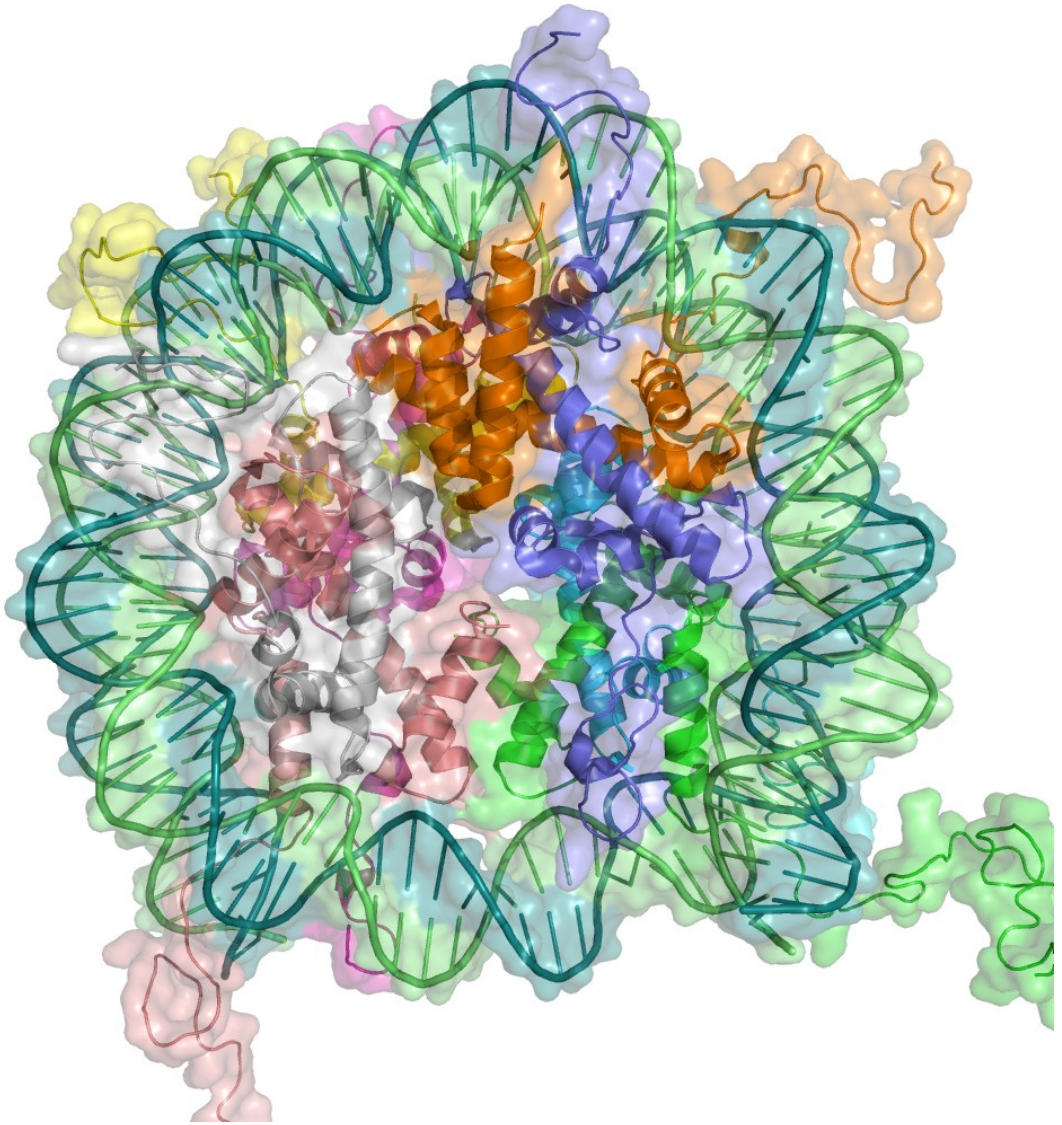
Copenhagen:

Rune Friborg

Funding:

HEROIC – mouse epigenetics

Thank you for using CATCHprofiles



http://catch.cmbi.ru.nl

fnielsen@cmbi.ru.nl