CEGX-QC
User Documentation v0.2

CEGX Bioinformatics Team

Cambridge Epigenetix (CEGX), Babraham Research Campus, Cambridge, CB22 3AT

Contact email: technical@cegx.co.uk

www.cegx.co.uk

cegcQC: https://bitbucket.org/cegx-bfx/cegxqc

11/02/2015

Contents

1 What is cegxQC?

CEGX has developed a post-sequencing program, cegxQC, a customised version of FastQC, to output a set of summary documents and QC reports based on the conversion performance of the sequencing spike-in controls.

cegxQC is designed to perform quality control analysis of fastq files from bisulfite (BS-Seq) and oxidised bisulfite (oxBS-Seq) sequencing.

2 Download cegxQC

cegxQC is freely available as a pre-compiled binary

`https://bitbucket.org/cegx-bfx/cegxqc/downloads`

Or as source code

`https://bitbucket.org/cegx-bfx/cegxqc/src`

3 Installation Instructions

An INSTALL.txt file is provided in the cegxQC software package. Below is a brief summary of the steps required to install and run cegxQC.

1. cegxQC, like its parent program FastQC, is a java application. In order to run it needs your system to have a suitable Java Runtime Environment (JRE) installed. A JRE is available from `http://www.java.com`
2. Unzip the downloaded cegxQC package

As cegxQC is a customised version of FastQC the installation and running of the program are almost identical. We therefore recommend reading the FastQC documentation

`http://www.bioinformatics.babraham.ac.uk/projects/fastqc`.

4 Running cegxQC

cegxQC can be run in either an interactive graphical mode or as part of a pipeline using the command line interface.

Running cegxQC Interactively

Windows: Double click on the `run_cegxqc` bat file.

MacOSX/Linux: A wrapper script, `cegxqc`, is the simplest way to launch the program.

Running cegxQC as part of a pipeline

To run cegxQC non-interactively use the `cegxqc` wrapper script and specify a list of files to process on the command line

```
cegxqc afastqfile.fastq anotherfastqfile.fq
```

As many fastq files can be specified on the command line as required. We recommend only launching two or three to avoid running out of memory. If no fastq files are specified cegxQC will launch in graphical mode, which could cause errors if the display isn't set on e.g. a cluster.

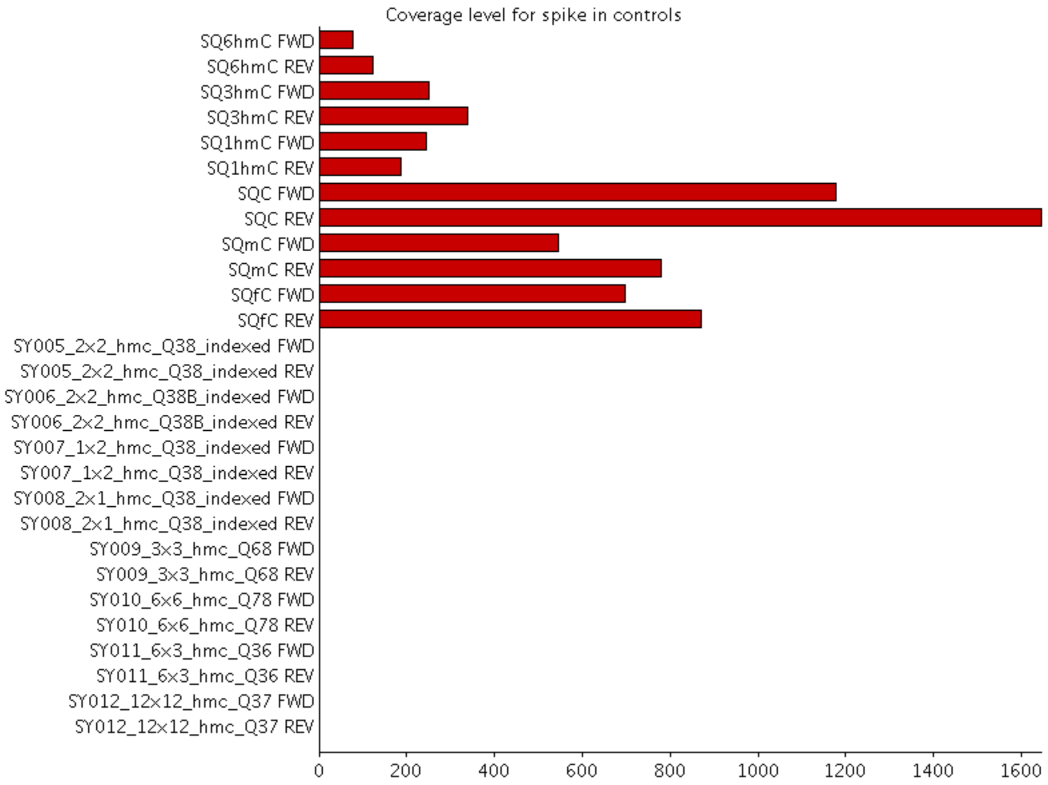
5 CEGX Specific Analysis Modules

As cegxQC is a customised version of FastQC we recommend reading the FastQC documentation for the more general non-control sequence features

`http://www.bioinformatics.babraham.ac.uk/projects/fastqc`.

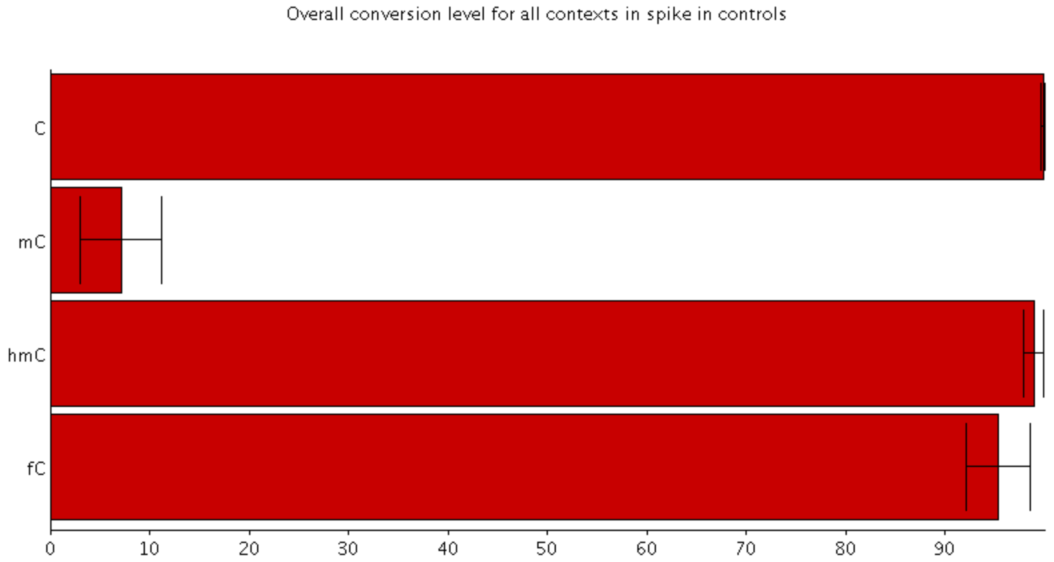
5.1 CEGX Control Coverage

Coverage levels are plotted for each of the spike in controls in both forward and reverse orientations.



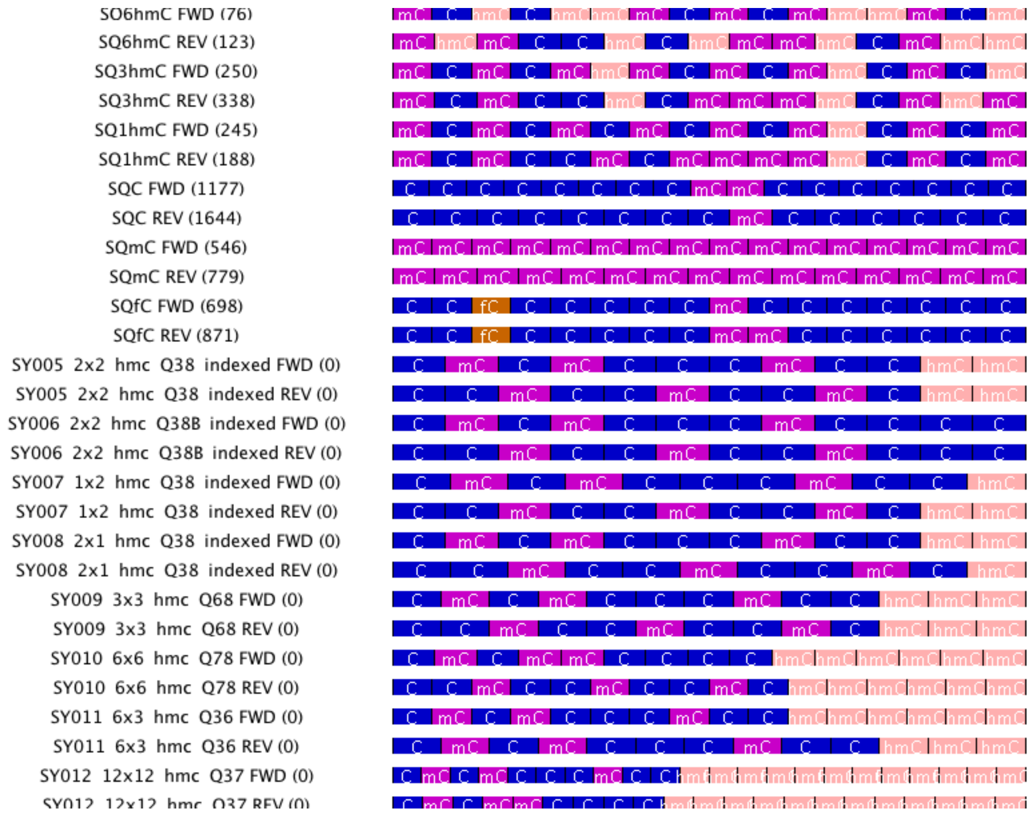
5.2 CEGX Conversion Summary

The conversion levels are summarised and plotted across all of the spike-in-controls for cytosines, methyl cytosines, hydroxymethyl cytosines and formyl cytosine.



5.3 CEGX Per Control Conversion

The modified cytosine positions in each of the controls are displayed as coloured blocks. The height of the blocks are proportional to the conversion rate of the specific cytosine position.



Note: This section is undergoing development and we hope to release an updated version soon.

6 Feature Requests and Bug Reports

We are always very happy to hear about any features you think should be added to cegxQC. We are actively developing the program and plan to add new features.

Bug reports or any issues using the software can be reported on our software repository site in the issue tracking section

`https://bitbucket.org/cegx-bfx/cegxqc/issues`

7 Other Recommended Packages

7.1 FastQC

The parent program for cegxQC

`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`

7.2 bsExpress

bsExpress is designed to perform quality control bisulfite (BS-Seq) and oxidised bisulfite (oxBS-Seq) libraries using ad hoc control sequences where cytosine modification are known. However, the pipeline is not limited to control sequences but is also suitable for processing (ox)BS-Seq data from raw fastq files to genome-wide methylation calls.

https://code.google.com/p/oxbs-sequencing-qc/wiki/bsExpressDoc

7.3 Bismark

Short read aligner for bisulfite (BS-Seq) and oxidised bisulfite (oxBS-Seq) sequencing data

http://www.bioinformatics.babraham.ac.uk/projects/bismark

8 GPLv3 License