

# QualiMap 2.0: quality control of high throughput sequencing data

Konstantin Okonechnikov<sup>1</sup>, Ana Conesa<sup>2</sup>, Fernando García-Alcalde<sup>1,3</sup>

<sup>1</sup> Department of Molecular Biology, Max Planck Institute for Infection Biology, D-10117, Berlin, Germany. Email: [okonechnikov@mpiib-berlin.mpg.de](mailto:okonechnikov@mpiib-berlin.mpg.de)

<sup>2</sup> Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46012, Valencia, Spain

<sup>3</sup> Roche Pharma Research and Early Development, Roche Innovation Center Basel, Switzerland

**Project Website:** <http://qualimap.org/>

**Source Code:** <http://bitbucket.org/kokonech/qualimap/>

**License:** GNU General Public License v2

## Main Text of Abstract

Detection of random errors and systematic biases is a crucial step of a robust pipeline for processing of high throughput sequencing (HTS) data. There are bioinformatics software tools capable of performing this task. Some of them are suitable for general analysis of HTS data while others are targeted to a specific sequencing technology. QualiMap 2.0 represents a next step in the QC analysis of HTS data. It is a multiplatform user-friendly application with both graphical user and command line interfaces.

QualiMap includes four analysis modes: **BAM QC**, **Counts QC**, **RNA-seq QC** and **Multi-sample BAM QC**. Based on the selected type of analysis, users provide input data in the form of a BAM/SAM alignment, GTF/GFF/BED annotation file and/or read counts table. The results of the QC analysis are presented as an interactive report from GUI, as a static report in HTML or PDF format and as a plain text file suitable for parsing and further processing. The latter two analysis modes are first introduced in version 2.0. **Multi-sample BAM QC** allows combined quality control estimation for multiple alignment files. For this purpose QualiMap combines **BAM QC** results from multiple samples and creates a number of plots summarizing the datasets. **RNA-seq QC** performs computation of metrics specific for RNA-seq data, including per-transcript coverage, junction sequence distribution and reads genomic localization.

In addition, a large number of fixes and enhancements were implemented since the first version of QualiMap was released. Most of the bugs were reported by the users. Additionally due to open-source access to the code several fixes were implemented by the users and accepted in the main repository.