

# Implementation of a copy number analysis pipeline for shallow sequencing in Galaxy framework

Van Lieshout S., Scheinin I., Sie D., Fijneman R.J.A., Ylstra B.

Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands

## Introduction

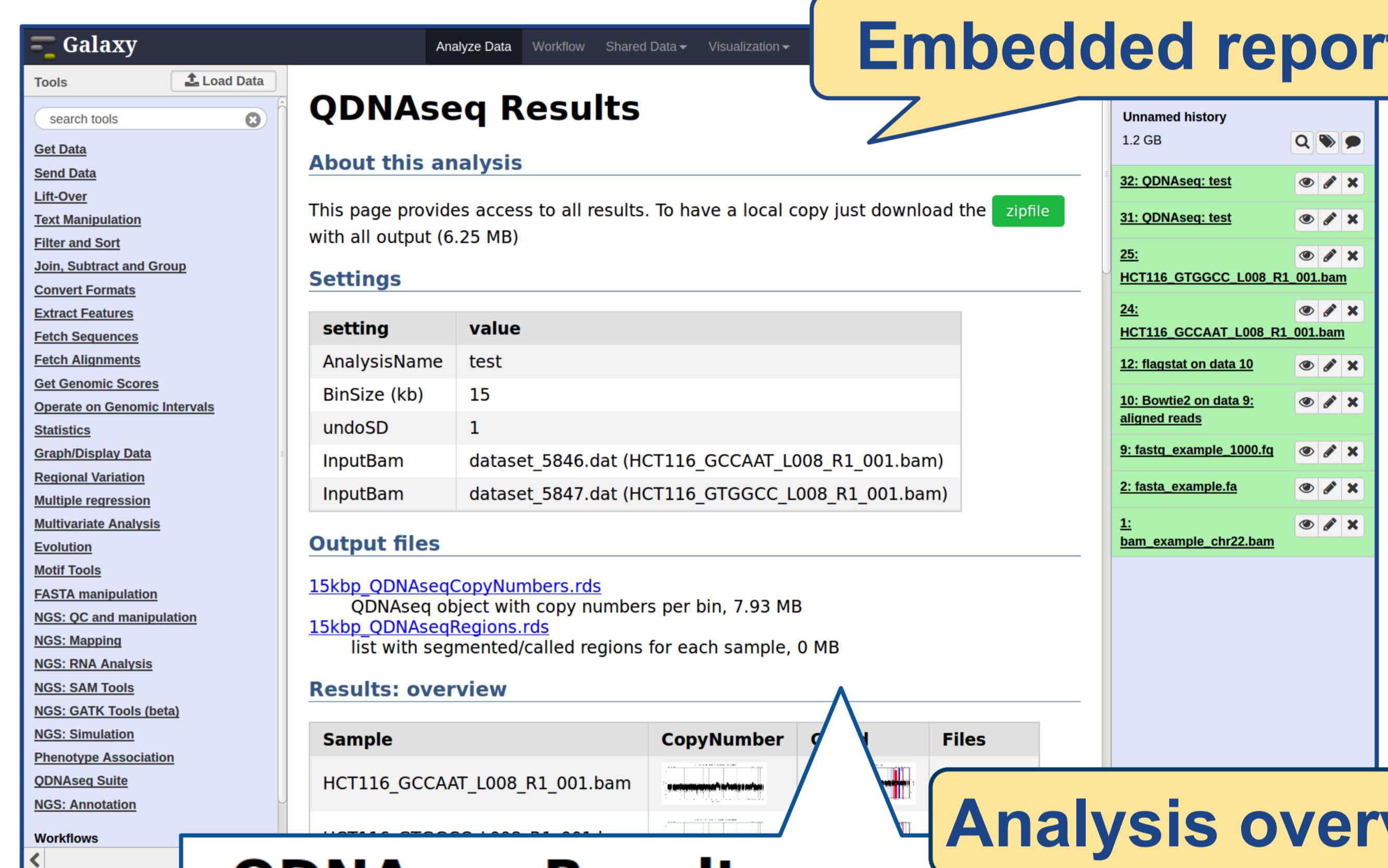
DNA copy number aberrations are a hallmark of cancer and can be quantified by **shallow whole-genome sequencing** (WGS). A robust method has been developed <sup>(1)</sup> that detects copy number aberrations. This method is currently being integrated into the **Galaxy platform** <sup>(3)</sup> within the Center for Translational Molecular Medicine - Translational Research IT (CTMM TraIT) project <sup>(4)</sup>.

- ✓ **highly concordant with array CGH** at considerably lower cost <sup>(1)</sup>
- ✓ also available as a **Bioconductor** package, QDNAseq <sup>(2)</sup>
- ✓ the popular **BAM format** as input
- ✓ reports results in a clear and concise **HTML based view** within Galaxy itself
- ✓ **R data structure** file for downstream analysis available
- ✓ Zipped archive with all output

## Methods

Sequence reads are **binned and counted in non overlapping windows** of 15kb and a combined LOESS correction for mappability and GC content is applied. Then a **comprehensive filtering** excludes genomic regions from both ENCODE project blacklists and a novel blacklist we have developed based on sequence depth of 38 individuals from the 1000 Genomes project. Procedures for copy number detection have been **optimized for** use in combination with DNA isolated from **formalin-fixed paraffin-embedded** (FFPE) samples.

## Embedded report



**QDNAseq Results**

About this analysis: This page provides access to all results. To have a local copy just download the [zipfile](#) with all output (6.25 MB)

Settings:

setting	value
AnalysisName	test
BinSize (kb)	15
undoSD	1
InputBam	dataset_5846.dat (HCT116_GCCAAT_L008_R1_001.bam)
InputBam	dataset_5847.dat (HCT116_GTGCC_L008_R1_001.bam)

Output files:

- 15kbp\_QDNAseqCopyNumbers.rds: QDNAseq object with copy numbers per bin, 7.93 MB
- 15kbp\_QDNAseqRegions.rds: list with segmented/called regions for each sample, 0 MB

Results: overview

Sample	CopyNumber	Files
HCT116_GCCAAT_L008_R1_001.bam		<a href="#">bedGraph</a>
HCT116_GTGCC_L008_R1_001.bam		<a href="#">bedGraph</a>

## Analysis overview

### QDNAseq Results

#### About this analysis

This page provides access to all results. To have a local copy just download the [zipfile](#) with all output (6.25 MB)

#### Settings

setting	value
AnalysisName	test
BinSize (kb)	15
undoSD	1
InputBam	dataset_5846.dat (HCT116_GCCAAT_L008_R1_001.bam)
InputBam	dataset_5847.dat (HCT116_GTGCC_L008_R1_001.bam)

#### Output files

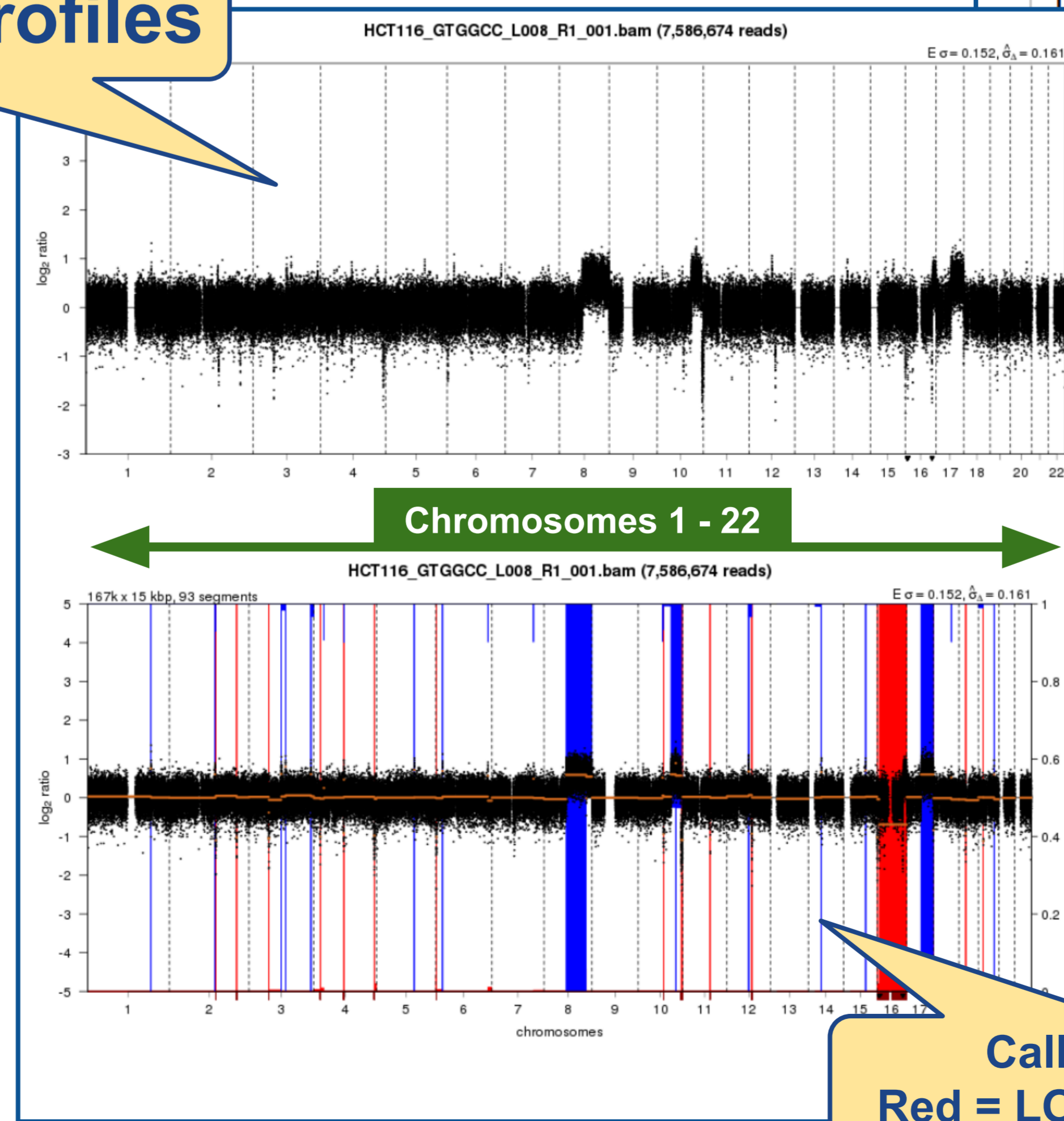
- [QDNAseqCopyNumbers.rds](#): QDNAseq object with copy numbers per bin, 7.93 MB
- [QDNAseqRegions.rds](#): list with segmented/called regions for each sample, 0 MB

#### Results: overview

Sample	CopyNumber	Called	Files
HCT116_GCCAAT_L008_R1_001.bam			<a href="#">bedGraph</a>
HCT116_GTGCC_L008_R1_001.bam			<a href="#">bedGraph</a>

Figures: Galaxy HTML output with links to output files and copy number profile images

## Copy Number Profiles



## References

1. I. Scheinin, D. Sie et al. DNA copy number analysis of fresh and formalin-fixed specimens by whole-genome sequencing: improved correction of systematic biases and exclusion of problematic regions (submitted).
2. <http://www.bioconductor.org/>
3. <http://galaxyproject.org>
4. <http://www.ctmm-trait.nl/>