Supplementary Figures

We applied Qualimap to data from the paired-end sequencing of C elegans DNA samples with DNA repair mutations (Sequence Read Archive (SRA): ERR089819). The reads were mapped to the C. elegans genome using bowtie with default parameters.

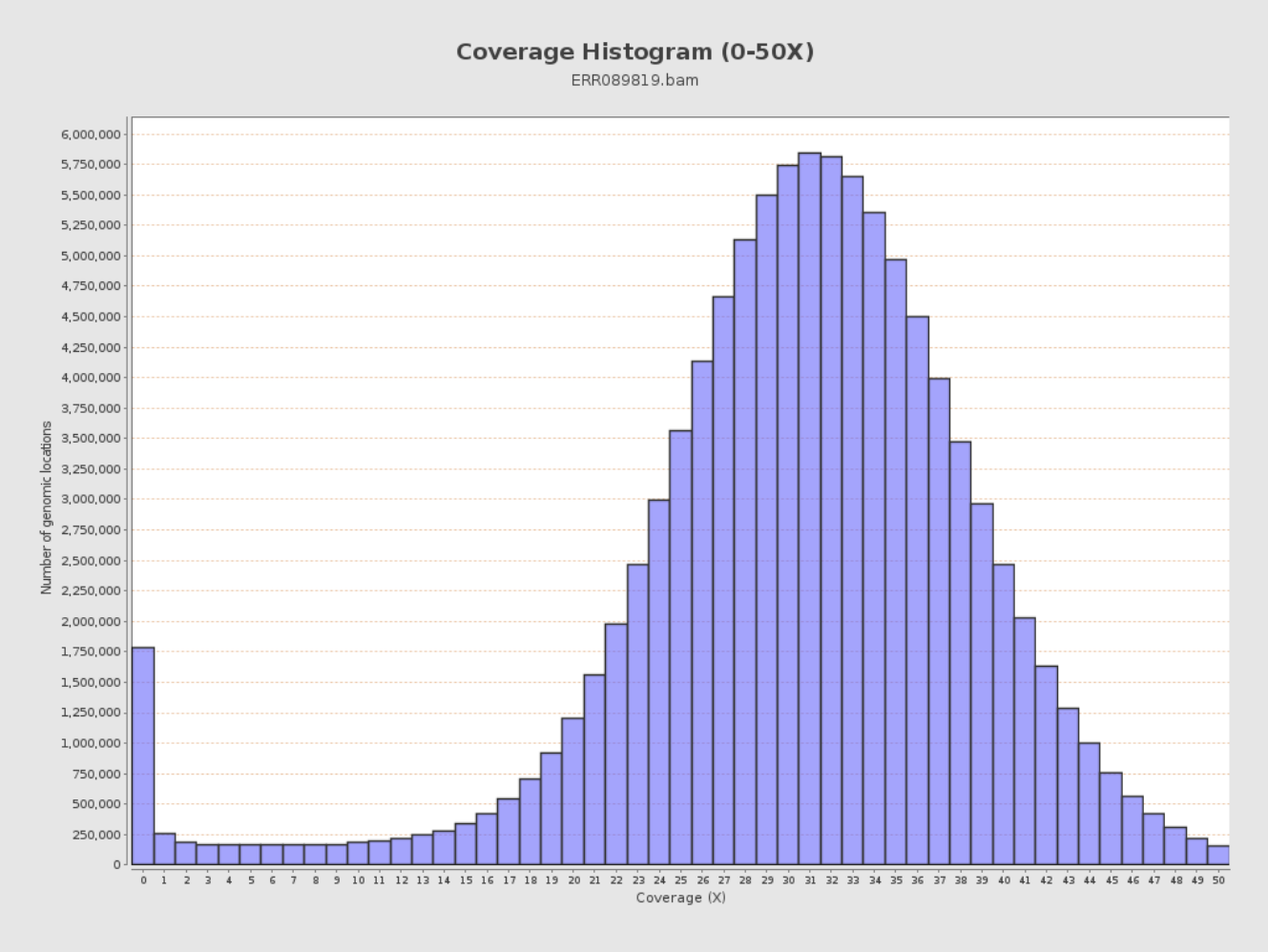

Figure S1. Genome-wide sequencing of a *C. elegans* sequencing . Coverage histogram is normal-distributed, centered at 31X.

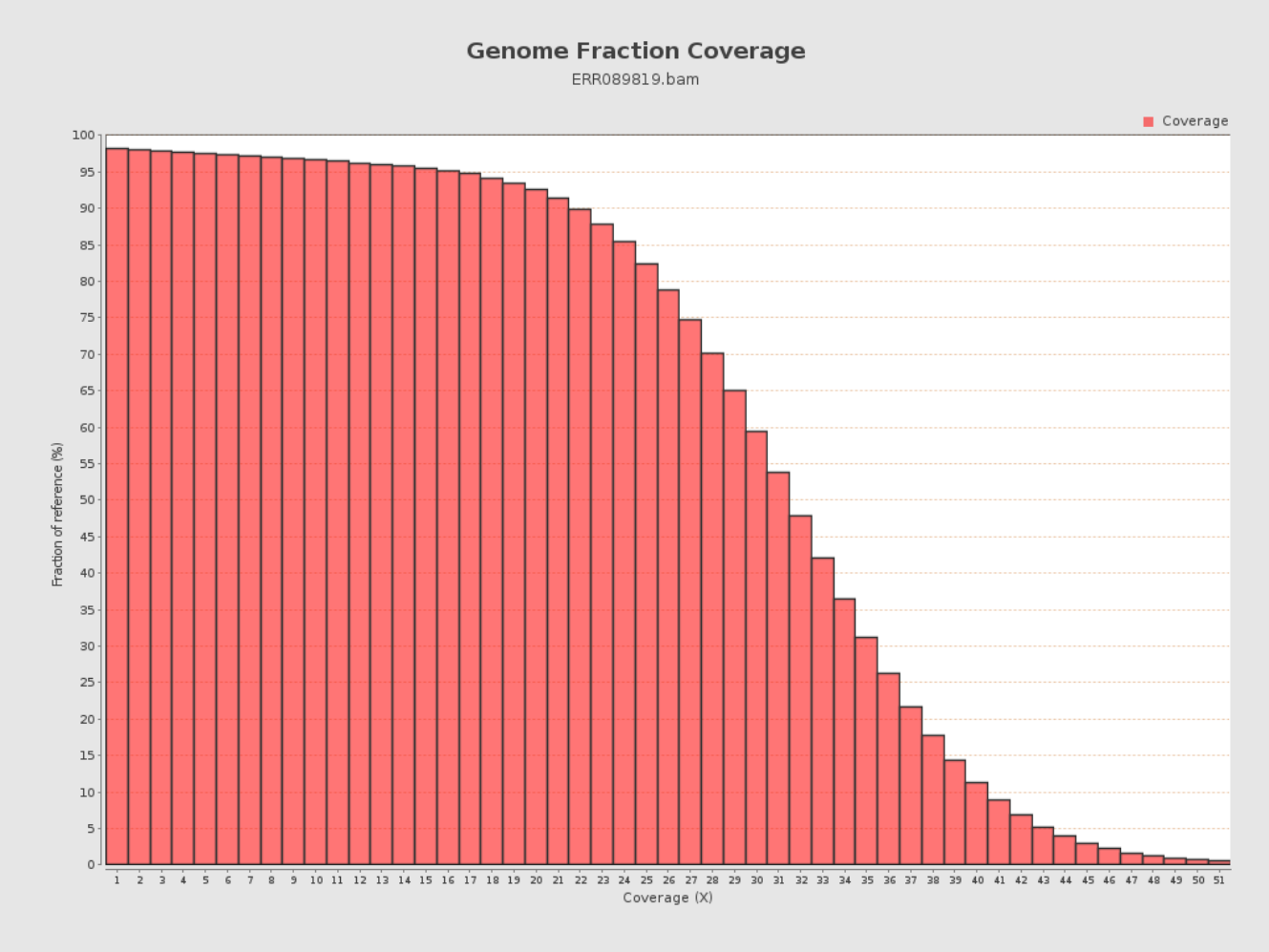

Figure S2. Genome-wide sequencing of *C. elegans*. Genome fraction per coverage. If a minimum coverage of 20X is selected, 93% of the reference genome would be included.

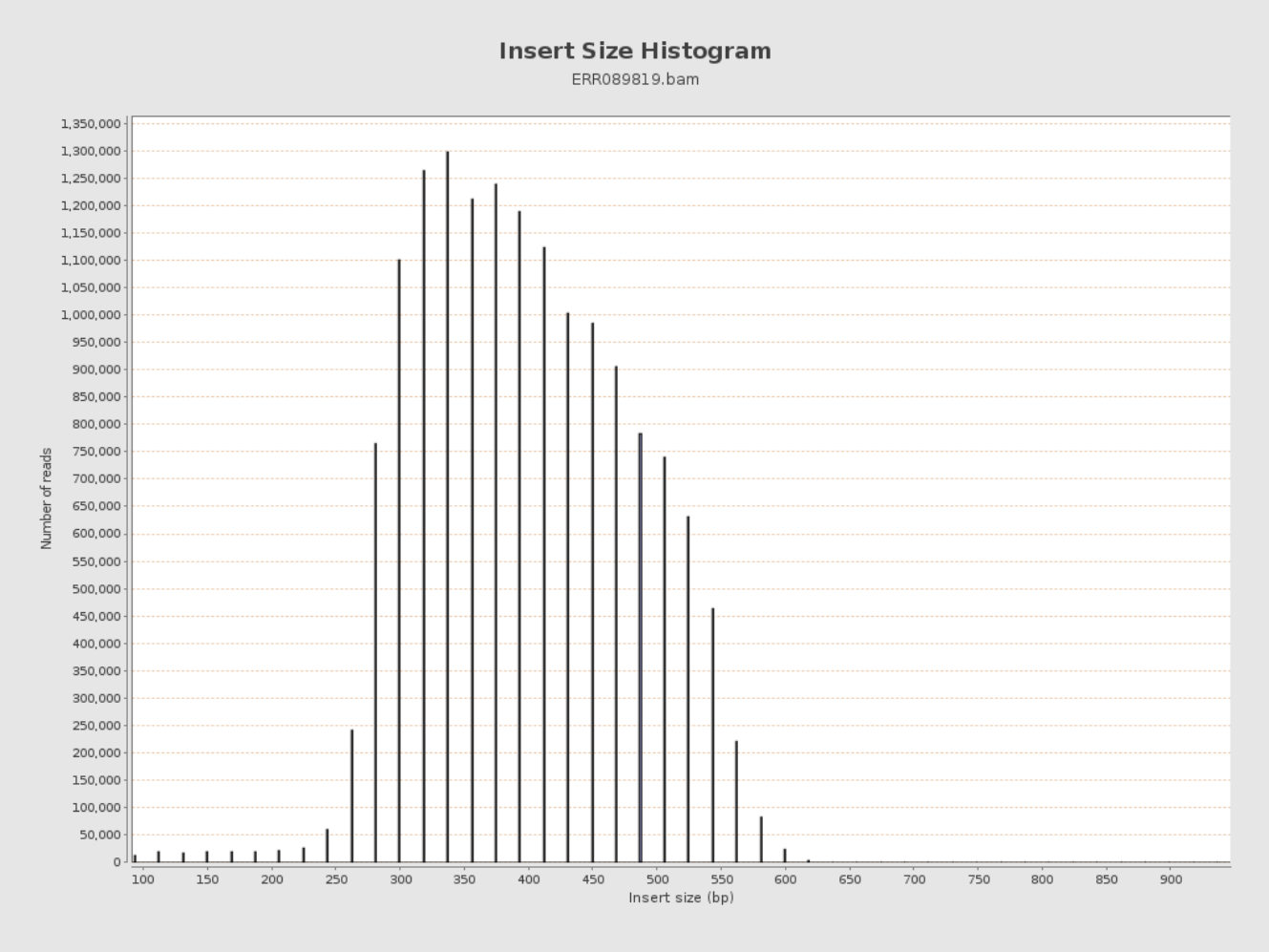

Figure S3. Genome-wide sequencing of *C. elegans*. Insert size distribution presents a certain deviation from that expected in the data description (374 bp with a standard deviation of 95).

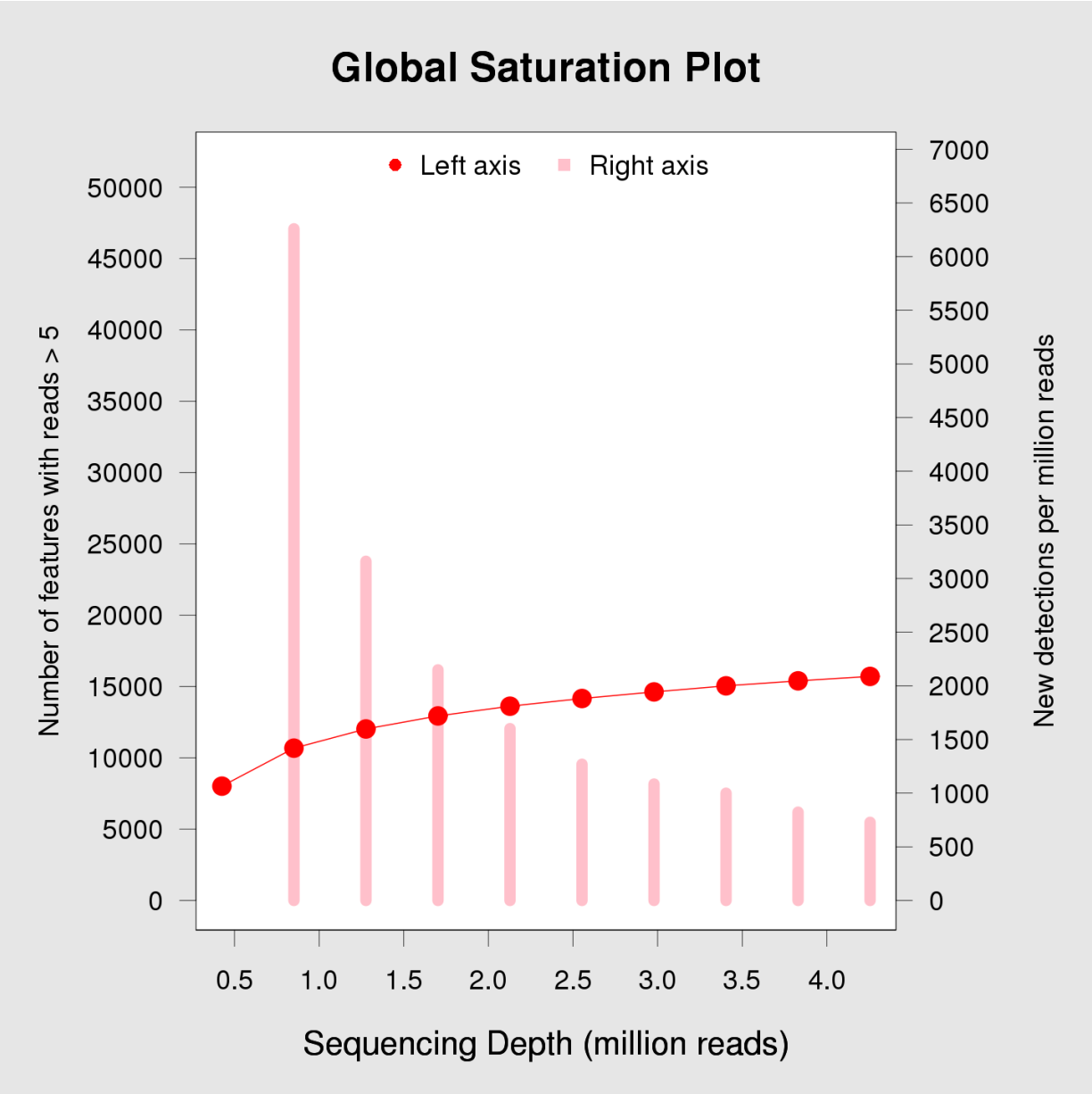

Figure S4. Human liver RNA-seq data from Marioni et al. (2008). Saturation plot of detected genes. As the sequencing depth grows there is a continual increase in the number of features detected, it would,therefore, be prudent to increase sequencing efforts.

Table S1. Comparison of quality control tools for NGS alignment data

	Picard tools	RNA-SeQC	RSeQC	**Qualimap**
Scope	Various NGS applications	RNA-seq	RNA-seq	Various NGS applications
User interface	Command-line	GUI + Command-line	Command-line	GUI + Command-line
Input formats	Alignment: BAM; annotations: RefSeq; reference: FASTA	Alignment: BAM; annotations: GTF; reference: FASTA	Alignment: BAM; annotations: BED.	Alignment: BAM; annotations: GFF/GTF or BED
Output	Raw data and plots. Different for each tool	Summary report including both html reports and raw data	Raw data and plots. Different for each tool	Summary report (HTML or PDF) and raw data
Alignment statistics	Yes	Partially	Partially	Yes
Overall Summary	No	Yes	No	Yes
Coverage analysis	Genome-wide or genes	Genes and transcripts	Genes and transcripts	Genome-wide or arbitrary regions
Average gene 5'-3' coverage plot	No	Yes	Yes	No
Counts computation	No	Yes	Yes	Yes
Insert size estimation	Yes	Yes	Yes	Yes
Sequence quality	No	Yes	Yes	Yes
Multiple samples support	No	Yes	No	2 samples
Homopolymer Indels	No	No	No	Yes
5’ / 3’ Bias	Yes	Yes	No	Yes
Strand specificity	Yes	Yes	Yes	Yes
Counts saturation	No	No	Yes	Yes
Counts RPKM	No	Yes	Yes	No
Expression correlation	No	Yes	No	2 samples
Clipping profile	No	No	Yes	Yes
Duplication rate	Yes	Yes	Yes	Yes
Clustering of coverage profiles	No	No	No	Yes

Table S1. Comparison of quality assignment tools. As can be observed, Qualimap provides capabilities similar to state-of-the-art existing tools, together with some additional unique features such as. Another main advantage of Qualimap is the incorporation of multiple quality assessment methods into a single pipeline, which outputs a comprehensive and easy-to-interpret summary. It is also noteworthy the availability of a user-friendly GUI interface and the support for annotation data in most common formats: GFF/GTF and BED. This comparison doesn’t include a popular FastQC toolkit (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) since it is designed for the analysis of the quality of sequencing reads. Therefore, Qualimap and other discussed tools are complementary to FastQC.

Analyzed tools

Piccard: *http://picard.sourceforge.net*
RNA-SeQC: *https://confluence.broadinstitute.org/display/CGATools/RNA-SeQC*
RseQC: *http://code.google.com/p/rseqc/*
Qualimap: *http://qualimap.org*