JOURNAL ARTICLE

HTSeq—a Python framework to work with high-throughput sequencing data

Simon Anders ✉, Paul Theodor Pyl, Wolfgang Huber Author Notes

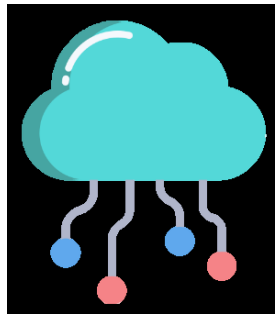
Bioinformatics, Volume 31, Issue 2, 15 January 2015, Pages 166–169,
https://doi.org/10.1093/bioinformatics/btu638
Published: 25 September 2014 **Article history** ▾

HTSeq – a Python framework

- facilitates rapid development of NGS data analysis scripts
- initial release in 2010, latest update published 2022
- allows users with moderate Python knowledge to create scripts, shields more advanced internals from the user

- **HTSeq-count** – a tool for preprocessing RNA-Seq data for differential expression analysis

SAM/BAM
GTF/GFF



HTSeq-count

how many aligned reads overlap exon/union of exons

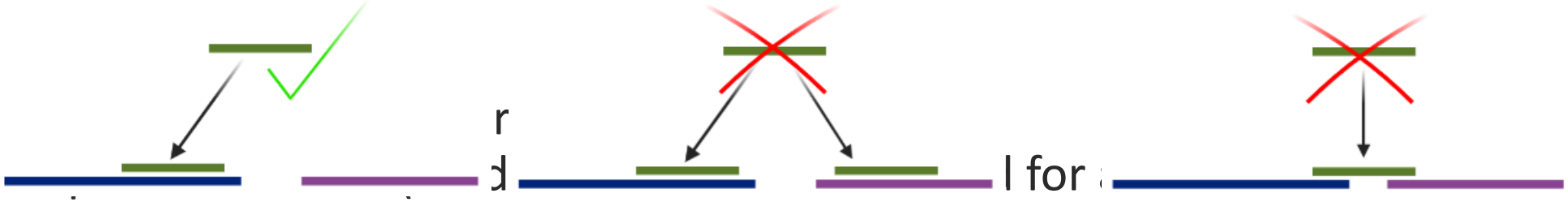
designed for **differential expression analysis**

default setting counts **unambiguously mapped** reads only

discards reads mapping to **multiple positions** or **overlapping** multiple features

for pair-end counts **fragments**, not reads

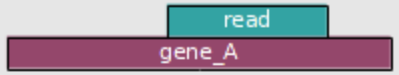
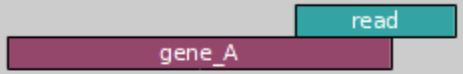
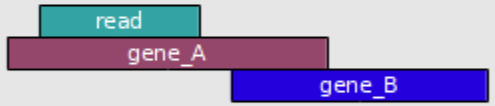
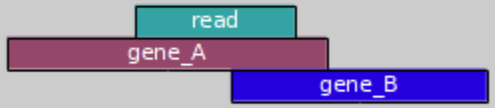
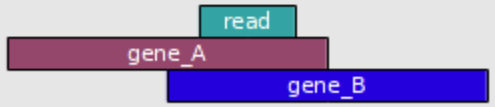
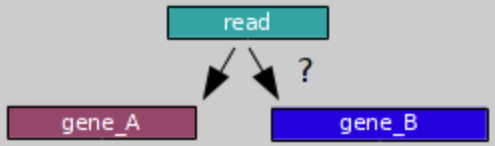
ta



r

d

l for

	union	intersection_strict	intersection_nonempty
	gene_A	gene_A	gene_A
	gene_A	no_feature	gene_A
	gene_A	no_feature	gene_A
	gene_A	gene_A	gene_A
	gene_A	gene_A	gene_A
	ambiguous (both genes with --nonunique all)	gene_A	gene_A
	ambiguous (both genes with --nonunique all)		
	alignment_not_unique (both genes with --nonunique all)		

If a read maps on multiple features:

--nonunique **none** (default)

--nonunique **all**

--nonunique **fraction**

--nonunique **random**