|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total no. predictions | Avg. sensitivity | Avg. precision | Long read confirmed | CPU time | Citations | Year published |
| Retroseq | 2286 | 0.6 | 0.77 | 23% | 03:43:01 | 185 | 2013 |
| Melt | 638 | 0.48 | 0.81 | 13% | 03:23:05 | 232 | 2017 |
| Retroseq+ | 296 | 0.53 | 0.63 | 78% | 04:01:15 | NA | NA |
| Steak | 13770 | 0.13 | 0.9 | Non-ref: 41%  Ref + Non-ref: 77% | 04:48:35 | 18 | 2017 |
| ERVcaller | 439 | 0.77 | 0.9 | 16% | 14:17:22 | 13 | 2019 |

Supplementary table 3, aggregated results across each benchmarking test.

This table draws together key results of all benchmarking experiments to give an overview of tool performance. The HERV specific tools are highlighted in blue, generalist tools in orange.

Total no. predictions refers to the number of predictions given by each tool when applied to 50 short read whole genome sequences.

Avg. sensitivity and Avg. precision are obtained by applying each tool to 4 simulated short read genome sequences with known HML6 proviral insertions.

The long read column shows the proportion of predictions made by each tool in short read data that were validated in associated long read data using repeatmasker.

CPU time is the CPU hours used to run each tool on a single SR-WGS file.

The final two columns report the publication date and citation number to indicate popularity.