**Aligner Benchmark Report**

**Target:**

Benchmarking all the steps from all the four aligners till the creation of sample.sam file.

**Resource Settings:**

1. CPU = 10 cores
2. Vmem = 20G

**Aligner – BWA:**

1. step1: create read1.sai file -> 113m2.911s
2. step2: create read2.sai file -> 343m12.921s
3. step3: create sample.sam file from reads sai file -> 28m43.922s

Total time from reads to sam file -> ~485min

**Aligner – BWA-MEM:**

1. step1: create sample.sam file from reads fasta file -> 26m12.405s

Total time from reads to sam file -> ~27min

**Aligner – BOWTIE2:**

1. step1: create sample.sam file from reads fasta file -> 71m42.881s

Total time from reads to sam file -> ~72min

**Aligner – GEM:**

1. step1: create read1.gem map file -> 8m23.311s
2. step2: create read2.gem map file -> 10m3.591s
3. step3: create sample.both.map file -> 2m5.657s
4. step4: create sample.both.qual.map file -> 24m57.273s
5. step5: create sample.sam file -> 34m23.234s

Total time from reads to sam file -> ~80m

**Variant Calls:**

**Aligner – BWA:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 9518 | 71.51% | 1.33 | 1.46 | 1.05 |
| GATK-MPILEUP | 17072 | 91% | 1.84 | 1.9 | 1.37 |
| GATK-SHORE | 3839 | 93.25% | 1.32 | 1.36 | 0.88 |
| Intersection | 111459 | 98.66% | 2.36 | 2.36 | 2.04 |
| MPILEUP | 2663 | 82.13% | 1.46 | 1.54 | 1.14 |
| MPILEUP-SHORE | 2582 | 97.75% | 2.04 | 2.06 | 1.42 |
| SHORE | 11171 | 77.26% | 1.3 | 1.89 | 0.35 |
| None | 158304 | 94.27% | 2.06 | 2.18 | 0.92 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 5346 | 74.39% | 1.34 | 1.52 | 0.93 |
| GATK-MPILEUP | 8996 | 94.01% | 2.12 | 2.16 | 1.57 |
| GATK-SHORE | 2499 | 93.80% | 1.20 | 1.25 | 0.68 |
| Intersection | 92225 | 98.98% | 2.40 | 2.41 | 2.08 |
| MPILEUP | 700 | 89.14% | 1.73 | 1.79 | 1.38 |
| MPILEUP-SHORE | 993 | 99.09% | 2.14 | 2.14 | 2.00 |
| SHORE | 3586 | 59.26% | 0.87 | 1.75 | 0.28 |
| None | 114345 | 96.02% | 2.19 | 2.29 | 0.82 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 0 bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 2279 | 77.23% | 1.24 | 1.40 | 0.82 |
| GATK-MPILEUP | 4042 | 95.25% | 2.21 | 2.24 | 1.67 |
| GATK-SHORE | 694 | 94.96% | 1.07 | 1.13 | 0.40 |
| Intersection | 46834 | 99.04% | 2.55 | 2.56 | 2.20 |
| MPILEUP | 232 | 90.52% | 1.55 | 1.59 | 1.20 |
| MPILEUP-SHORE | 234 | 99.57% | 2.24 | 2.27 | 0.00 |
| SHORE | 865 | 48.44% | 0.67 | 1.41 | 0.29 |
| None | 55180 | 96.98% | 2.35 | 2.43 | 0.90 |
| **INDELs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 1228 | 62.54% | 24.51% |  |  |
| GATK-MPILEUP | 910 | 73.74% | 35.38% |  |  |
| GATK-SHORE | 2784 | 91.74% | 50.00% |  |  |
| Intersection | 8478 | 93.45% | 73.10% |  |  |
| MPILEUP | 2838 | 66.24% | 32.73% |  |  |
| MPILEUP-SHORE | 940 | 89.68% | 56.81% |  |  |
| SHORE | 2039 | 87.54% | 48.65% |  |  |
| Clindel | 22277 | 93.97% | 62.35% |  |  |
| None | 19217 | 85.47% | 55.51% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 904 | 66.70% | 28.87% |  |  |
| GATK-MPILEUP | 627 | 76.08% | 39.55% |  |  |
| GATK-SHORE | 2442 | 92.22% | 52.91% |  |  |
| Intersection | 7279 | 94.09% | 76.15% |  |  |
| MPILEUP | 1759 | 67.65% | 34.85% |  |  |
| MPILEUP-SHORE | 614 | 90.23% | 62.38% |  |  |
| SHORE | 1075 | 86.23% | 48.74% |  |  |
| Clindel | 12530 | 93.16% | 67.52% |  |  |
| None | 14700 | 87.43% | 60.30% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 0bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 273 | 71.06% | 28.94% |  |  |
| GATK-MPILEUP | 222 | 80.63% | 41.89% |  |  |
| GATK-SHORE | 681 | 91.19% | 54.77% |  |  |
| Intersection | 2993 | 94.32% | 78.18% |  |  |
| MPILEUP | 348 | 55.46% | 21.26% |  |  |
| MPILEUP-SHORE | 105 | 90.48% | 55.24% |  |  |
| SHORE | 169 | 83.43% | 46.15% |  |  |
| Clindel | 4078 | 93.23% | 71.55% |  |  |
| None | 4791 | 88.62% | 64.60% |  |  |

**Aligner – BWA-MEM:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 8174 | 76.07% | 1.53 | 1.61 | 1.30 |
| GATK-MPILEUP | 13937 | 87.14% | 1.69 | 1.78 | 1.23 |
| GATK-SHORE | 4677 | 93.54% | 1.37 | 1.41 | 0.99 |
| Intersection | 118883 | 98.42% | 2.31 | 2.32 | 1.91 |
| MPILEUP | 3409 | 81.72% | 1.34 | 1.37 | 1.25 |
| MPILEUP-SHORE | 3270 | 96.51% | 1.78 | 1.79 | 1.65 |
| SHORE | 29447 | 69.16% | 1.40 | 1.54 | 1.13 |
| None | 181797 | 91.34% | 1.97 | 2.07 | 1.24 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 3750 | 82.16% | 1.71 | 1.82 | 1.31 |
| GATK-MPILEUP | 5994 | 93.04% | 2.09 | 2.15 | 1.46 |
| GATK-SHORE | 2819 | 94.54% | 1.32 | 1.36 | 0.83 |
| Intersection | 95473 | 98.88% | 2.39 | 2.40 | 2.07 |
| MPILEUP | 577 | 90.64% | 1.89 | 1.84 | 2.38 |
| MPILEUP-SHORE | 1029 | 98.35% | 1.95 | 1.94 | 2.40 |
| SHORE | 8305 | 59.72% | 1.49 | 1.94 | 1.04 |
| None | 117947 | 95.15% | 2.23 | 2.30 | 1.24 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 0 bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 1482 | 81.98% | 1.68 | 1.76 | 1.38 |
| GATK-MPILEUP | 2478 | 92.62% | 2.08 | 2.14 | 1.54 |
| GATK-SHORE | 821 | 94.52% | 1.22 | 1.28 | 0.55 |
| Intersection | 48706 | 98.97% | 2.54 | 2.55 | 2.25 |
| MPILEUP | 234 | 90.60% | 1.72 | 1.62 | 3.40 |
| MPILEUP-SHORE | 273 | 98.17% | 1.84 | 1.82 | 4.00 |
| SHORE | 57074 | 59.35% | 1.63 | 1.91 | 1.30 |
| None | 55180 | 96.01% | 2.39 | 2.45 | 1.48 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 860 | 61.63% | 28.14% |  |  |
| GATK-MPILEUP | 564 | 73.05% | 39.72% |  |  |
| GATK-SHORE | 2865 | 48.34 | 48.34% |  |  |
| Intersection | 8565 | 93.08% | 71.48% |  |  |
| MPILEUP | 2672 | 70.21% | 33.76% |  |  |
| MPILEUP-SHORE | 1117 | 53.45 | 53.45% |  |  |
| SHORE | 3035 | 71.53% | 35.78% |  |  |
| Clindel | 24297 | 91.07% | 58.34% |  |  |
| None | 19678 | 84.03% | 53.65% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 465 | 72.26% | 38.92% |  |  |
| GATK-MPILEUP | 388 | 78.87% | 44.85% |  |  |
| GATK-SHORE | 2470 | 91.74% | 51.66% |  |  |
| Intersection | 7253 | 93.92% | 75.38% |  |  |
| MPILEUP | 1562 | 74.58% | 39.37% |  |  |
| MPILEUP-SHORE | 726 | 88.43% | 59.78% |  |  |
| SHORE | 1515 | 74.98% | 39.54% |  |  |
| Clindel | 13252 | 91.25% | 64.82% |  |  |
| None | 14379 | 88.07% | 60.82% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 0bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 146 | 72.60% | 35.62% |  |  |
| GATK-MPILEUP | 118 | 79.66% | 52.54% |  |  |
| GATK-SHORE | 715 | 91.05% | 51.61% |  |  |
| Intersection | 3059 | 94.05% | 77.25% |  |  |
| MPILEUP | 312 | 66.03% | 28.21% |  |  |
| MPILEUP-SHORE | 119 | 89.08% | 50.42% |  |  |
| SHORE | 313 | 66.77% | 32.91% |  |  |
| Clindel | 4349 | 91.42% | 68.43% |  |  |
| None | 4782 | 88.85% | 64.76% |  |  |

**Aligner – BOWTIE2:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 2574 | 74.98% | 1 | 1.29 | 0.46 |
| GATK-MPILEUP | 6067 | 95.55% | 2.01 | 2.09 | 0.91 |
| GATK-SHORE | 2449 | 96.33% | 1.41 | 1.44 | 0.76 |
| Intersection | 89650 | 98.83% | 2.43 | 2.43 | 1.99 |
| MPILEUP | 1083 | 95.84% | 2.09 | 2.16 | 1.14 |
| MPILEUP-SHORE | 1862 | 98.55% | 2.19 | 2.20 | 1.70 |
| SHORE | 54371 | 83.48% | 1.67 | 1.78 | 1.21 |
| None | 158056 | 92.97% | 2.05 | 2.14 | 1.19 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 2062 | 72.41% | 0.91 | 1.19 | 0.43 |
| GATK-MPILEUP | 4537 | 95.97% | 2.04 | 2.15 | 0.65 |
| GATK-SHORE | 1888 | 96.19% | 1.28 | 1.33 | 0.53 |
| Intersection | 81479 | 96.33% | 2.44 | 2.45 | 1.99 |
| MPILEUP | 463 | 89.14% | 1.88 | 1.97 | 0.55 |
| MPILEUP-SHORE | 958 | 99.06% | 1.94 | 1.94 | 2.00 |
| SHORE | 21158 | 83.62% | 1.86 | 2.07 | 1.12 |
| None | 112545 | 95.46% | 2.22 | 2.31 | 1.07 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 0 bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 758 | 67.55% | 0.49 | 0.63 | 0.27 |
| GATK-MPILEUP | 1591 | 96.23% | 2.08 | 2.18 | 0.76 |
| GATK-SHORE | 456 | 95.61% | 1.10 | 1.18 | 0.18 |
| Intersection | 42074 | 99.06% | 2.61 | 2.61 | 2.05 |
| MPILEUP | 137 | 96.35% | 1.58 | 1.54 | 4.00 |
| MPILEUP-SHORE | 250 | 98.80% | 1.80 | 1.76 | 3.00 |
| SHORE | 9738 | 86.18% | 1.98 | 2.11 | 1.40 |
| None | 55004 | 96.23% | 2.37 | 2.44 | 1.21 |
| **INDELs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 813 | 74.78% | 40.47% |  |  |
| GATK-MPILEUP | 398 | 82.66% | 55.53% |  |  |
| GATK-SHORE | 2370 | 91.90% | 53.16% |  |  |
| Intersection | 6578 | 94.44% | 78.99% |  |  |
| MPILEUP | 1502 | 78.83% | 47.60% |  |  |
| MPILEUP-SHORE | 685 | 93.14% | 70.66% |  |  |
| SHORE | 4567 | 79.13% | 37.09% |  |  |
| Clindel | 22331 | 91.89% | 60.14% |  |  |
| None | 16913 | 87.29% | 58.53% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 711 | 74.82% | 41.07% |  |  |
| GATK-MPILEUP | 345 | 82.61% | 54.78% |  |  |
| GATK-SHORE | 2190 | 92.01% | 54.57% |  |  |
| Intersection | 6068 | 94.73% | 79.93% |  |  |
| MPILEUP | 1171 | 77.28% | 47.57% |  |  |
| MPILEUP-SHORE | 550 | 92.00% | 70.55% |  |  |
| SHORE | 2397 | 81.14% | 40.84% |  |  |
| Clindel | 12573 | 91.62% | 65.88% |  |  |
| None | 13432 | 88.86% | 62.91% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 0bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 202 | 72.77% | 44.06% |  |  |
| GATK-MPILEUP | 105 | 84.76% | 55.24% |  |  |
| GATK-SHORE | 651 | 89.86% | 54.38% |  |  |
| Intersection | 2616 | 94.92% | 81.35% |  |  |
| MPILEUP | 231 | 66.23% | 32.47% |  |  |
| MPILEUP-SHORE | 115 | 93.04% | 73.04% |  |  |
| SHORE | 684 | 80.56% | 40.64% |  |  |
| Clindel | 4198 | 91.64% | 69.41% |  |  |
| None | 4604 | 89.38% | 66.59% |  |  |

**Aligner – GEM:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 45769 | 51.70% | 1.37 | 1.39 | 1.35 |
| GATK-MPILEUP | 65301 | 67.76% | 1.64 | 1.64 | 1.64 |
| GATK-SHORE | 10429 | 95.71% | 2.08 | 2.12 | 1.44 |
| Intersection | 92066 | 99.03% | 2.40 | 2.40 | 1.97 |
| MPILEUP | 9148 | 69.69% | 1.25 | 1.28 | 1.18 |
| MPILEUP-SHORE | 2024 | 99.31% | 2.15 | 2.15 | 1.80 |
| SHORE | 8619 | 74.78% | 1.27 | 2.11 | 0.26 |
| None | 233356 | 78.80% | 1.81 | 1.96 | 1.37 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 17465 | 49.74% | 1.62 | 1.73 | 1.51 |
| GATK-MPILEUP | 26062 | 68.87% | 2.05 | 2.09 | 1.96 |
| GATK-SHORE | 7471 | 97.56% | 2.09 | 2.13 | 1.02 |
| Intersection | 80772 | 99.02% | 2.41 | 2.41 | 1.96 |
| MPILEUP | 1444 | 69.18% | 1.61 | 1.64 | 1.53 |
| MPILEUP-SHORE | 937 | 99.47% | 1.97 | 1.99 | 0.67 |
| SHORE | 3341 | 57.65% | 0.84 | 1.87 | 0.24 |
| None | 137492 | 85.65% | 2.12 | 2.26 | 1.49 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNPs variant report from enriched regions + plus 0 bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **titv\_all** | **titv\_known** | **titv\_novel** |
| GATK | 8465 | 50.05% | 1.63 | 1.67 | 1.58 |
| GATK-MPILEUP | 11805 | 72.65% | 2.14 | 2.18 | 2.05 |
| GATK-SHORE | 3240 | 98.24% | 2.30 | 2.36 | 0.63 |
| Intersection | 41312 | 99.05% | 2.57 | 2.57 | 2.10 |
| MPILEUP | 542 | 64.21% | 1.45 | 1.32 | 1.73 |
| MPILEUP-SHORE | 188 | 99.47% | 1.98 | 2.02 | 0.00 |
| SHORE | 743 | 36.34% | 0.47 | 1.27 | 0.23 |
| None | 66295 | 87.07% | 2.26 | 2.39 | 1.58 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from all regions (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 7262 | 16.25% | 5.22% |  |  |
| GATK-MPILEUP | 3125 | 35.39% | 16.03% |  |  |
| GATK-SHORE | 3144 | 80.44% | 52.64% |  |  |
| Intersection | 6529 | 83.37% | 67.59% |  |  |
| MPILEUP | 5515 | 31.21% | 14.69% |  |  |
| MPILEUP-SHORE | 628 | 75.48% | 57.01% |  |  |
| SHORE | 1442 | 80.24% | 51.11% |  |  |
| Clindel | 17938 | 87.40% | 62.49% |  |  |
| None | 27645 | 49.23% | 32.02% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 150bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 3016 | 23.44% | 8.99% |  |  |
| GATK-MPILEUP | 1785 | 37.76% | 18.43% |  |  |
| GATK-SHORE | 2700 | 81.07% | 55.11% |  |  |
| Intersection | 5903 | 82.82% | 68.36% |  |  |
| MPILEUP | 3311 | 28.27% | 14.98% |  |  |
| MPILEUP-SHORE | 456 | 69.52% | 54.82% |  |  |
| SHORE | 814 | 71.62% | 44.47% |  |  |
| Clindel | 10723 | 82.20% | 62.55% |  |  |
| None | 17985 | 57.24% | 40.21% |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **INDELs variant report from enriched regions + plus 0bp (Nimblegen kit)** | | | | | |
| **Tool** | **#ofVariants** | **Known% in dbSNP** | **Known% in Mills** |  |  |
| GATK | 1159 | 21.57% | 8.20% |  |  |
| GATK-MPILEUP | 629 | 38.79% | 18.12% |  |  |
| GATK-SHORE | 912 | 81.58% | 58.22% |  |  |
| Intersection | 2349 | 80.72% | 67.69% |  |  |
| MPILEUP | 984 | 15.24% | 6.50% |  |  |
| MPILEUP-SHORE | 66 | 51.52% | 40.91% |  |  |
| SHORE | 147 | 61.22% | 29.93% |  |  |
| Clindel | 3569 | 79.71% | 62.82% |  |  |
| None | 6246 | 54.56% | 39.47% |  |  |