**Table S10: Variant-level quality metrics of high-quality variants identified from undetermined variants in the PSP dataset**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | No QC | ABHet | VQSR | ForestQC |
| Total SNVs | 3950305 | 779868 | 2746355 | 1711698 |
| Known SNVs | 2270658 | 655148 | 1748918 | 1103937 |
| Known SNVs (%) | 57.48% | 84.01% | 63.68% | 64.49% |
| Novel SNVs | 1679647 | 124720 | 997437 | 607761 |
| Novel SNVs (%) | 42.52% | 15.99% | 36.32% | 35.51% |
| Known Ti/Tv | 1.6060 | 1.8441 | 1.7801 | 1.9336 |
| Novel Ti/Tv | 1.1726 | 1.2524 | 1.2178 | 1.4236 |
| Total indels | 1596418 | 247622 | 819608 | 718606 |
| Known indels | 1009966 | 170389 | 687948 | 439129 |
| Known indels (%) | 63.26% | 68.81% | 83.94% | 61.11% |
| Novel indels | 586452 | 77233 | 131660 | 279477 |
| Novel indels (%) | 36.74% | 31.19% | 16.06% | 38.89% |
| Multi-allelic SNVs | 198530 | 541 | 165307 | 144688 |
| Multi-allelic SNVs (%) | 5.03% | 0.07% | 6.02% | 8.45% |
| Known multi-allelic SNVs | 174058 | 491 | 154111 | 137447 |
| Known multi-allelic SNVs (%) | 7.67% | 0.07% | 8.81% | 12.45% |
| Singletons in SNVs | 1276305 | 136018 | 1053352 | 715489 |
| Singletons in SNVs (%) | 32.31% | 17.44% | 38.35% | 41.80% |
| Singletons in indels | 300870 | 51339 | 139668 | 178152 |
| Singletons in indels (%) | 18.85% | 20.73% | 17.04% | 24.79% |

The undetermined variants in the PSP dataset are processed by four different methods, including no QC applied, ABHet approach, VQSR and ForestQC. There are 20 metrics in total, which are described in Material and Methods section in detail. “Known” stands for variants found in dbSNP. “Novel” stands for variants not found in dbSNP. The version of dbSNP is 150.