**Table S3 Summary of Effective Data for exome sequencing**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exon Capture** | **IV:3** | **IV:5** | **IV:17** | **IV:31** |
| Initial bases on target: | 37563132 | 37563132 | 37563132 | 37640396 |
| \*Initial bases near target: | 57118957 | 57118957 | 57118957 | 57236802 |
| Initial bases on or near target: | 94682089 | 94682089 | 94682089 | 94877198 |
| \*\*Total effective reads: | 35368793 | 33997483 | 34180944 | 34441160 |
| Total effective yield(Mb): | 3050.76 | 2951.33 | 2932.59 | 2932.04 |
| Average read length(bp): | 86.26 | 86.81 | 85.8 | 85.13 |
| Effective sequence on target(Mb): | 1982.61 | 1888.21 | 1969.75 | 1912.5 |
| Effective sequence near target(Mb): | 322.97 | 328.79 | 301.08 | 283.94 |
| Effective sequence on or near target(Mb): | 2305.58 | 2217 | 2270.83 | 2196.44 |
| Fraction of effective bases on target: | 65.00% | 64.00% | 67.20% | 65.20% |
| Fraction of effective bases on or near target: | 75.60% | 75.10% | 77.40% | 74.90% |
| **Average sequencing depth on target:** | **52.78** | **50.27** | **52.44** | **50.81** |
| Average sequencing depth near target: | 5.65 | 5.76 | 5.27 | 4.96 |
| Mismatch rate in target region: | 0.35% | 0.33% | 0.38% | 0.41% |
| Mismatch rate in all effective sequence: | 0.38% | 0.36% | 0.40% | 0.45% |
| Base covered on target: | 37126868 | 37101900 | 37117803 | 37172045 |
| **Coverage of target region:** | **98.80%** | **98.80%** | **98.80%** | **98.80%** |
| Base covered near target: | 34863681 | 35792077 | 33083974 | 32670715 |
| Coverage of flanking region: | 61.00% | 62.70% | 57.90% | 57.10% |
| Fraction of target covered with at least 20X | 74.00% | 73.00% | 73.50% | 72.10% |
| Fraction of target covered with at least 10X | 88.30% | 87.90% | 88.10% | 87.20% |
| Fraction of target covered with at least 4X | 95.80% | 95.70% | 95.80% | 95.40% |
| Fraction of flanking region covered with at least 20X | 7.70% | 7.80% | 7.10% | 6.50% |
| Fraction of flanking region covered with at least 10X | 16.50% | 17.00% | 15.30% | 14.40% |
| Fraction of flanking region covered with at least 4X | 32.30% | 33.40% | 30.10% | 28.90% |

\* The region near target refers to flanking region within 200bp of target regions.

\*\* Total effective reads is the same meaning as the unique mapped reads which was stated in the pipeline above. Here the effective reads consist of two parts: i) the reads have only one best hit in the alignment. These reads comes from the unique region of genome ii) the reads have multiple best hits on the genome (the number of hits between 1 and 20), and they were randomly aligned onto the target regions. These reads mainly come from low complex genomic region, such as repetitive sequences, and account for about 4% of total effective reads.

\*\*\* Target regions used here refer to genomic regions that the Exome array actually covered. The aggregate length of target is about 37.8Mb.