Supplemental Table 1.  Sequencing coverage metrics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gene | Full gene coverage (x) mean (stdev) | | Samples with >95% of sites at 20X | | Sites of Interest^ coverage (x) mean (stdev) | |
| Group 1 | ADMEseq | WGS | ADME-seq^^ | WGS | ADMEseq | WGS |
| *CYP2C8* | 113.4 (43.5) | 39.4 (4.1) | 82% | 99% | 120.5 (44.9) | 38.9 (4.6) |
| *CYP2C9* | 91.4 (37.2) | 38.8 (4.4) | 49% | 98% | 126.6 (45.2) | 37.8 (4.3) |
| *CYP2C19* | 81.4 (31.0) | 38.0 (4.3) | 36% | 99% | 126.6 (42.1) | 37.5 (4.2) |
| Groups 2+4^^^ | PGx-seq | WGS | PGx-seq^^^^ | WGS | PGx-seq | WGS |
| *CYP2C8* | 526.1 (89.4) | 38.8 (3.2) | 100% (100%) | 99% | 529.4 (90.5) | 38.2 (3.8) |
| *CYP2C9* | 464.0 (79.3) | 38.2 (3.5) | 0% (100%) | 100% | 510.7 (88.2) | 37.1 (3.2) |
| *CYP2C19* | 477.3 (81.8) | 38.0 (3.0) | 1% (100%) | 100% | 592.0 (102.6) | 36.7 (3.1) |
| Group 3 | PGRN-seq v1 | WGS-2 | PGRN-seq v1 | WGS-2 | PGRN-seq v1 | WGS-2 |
| *CYP2C8* | 271.9 (67.6) | 39.7 (7.5) | 100% | 97% | 377.6 (164.9) | 38.5 (10.1) |
| *CYP2C9* | 336.6 (83.6) | 39.9 (8.4) | 100% | 97% | 335.1 (104.4) | 39.1 (10.5) |
| *CYP2C19* | 318.6 (78.9) | 39.6 (8.2) | 97% | 98% | 344.6 (109.7) | 39.4 (10.7) |
|  |  |  |  |  |  |  |
| ^ Any site that is used to define a star allele | | | |  |  |  |
| ^^ Some sites of interest in the 5'-flanking region were not covered at 20X contributing to the lower percentages  ^^^ Groups 2 and 4 used the same data | | | | | |  |
|  |
| ^^^^ Number in brackets represents the percent after the removal of sites of interest that are not covered by PGx-seq | | | | | |  |

For each gene of interest and data sets used, three coverage metrics have been applied for both targeted-capture and whole-genome sequenced samples using Picard CollectWgsMetrics (see [https://gatk.broadinstitute.org/hc/en-us/articles/360037269351-CollectWgsMetrics-Picard](https://gatk.broadinstitute.org/hc/en-us/articles/360037269351-CollectWgsMetrics-Picard-), accessed 4/16/2021, for details). The metrics reported are the average and standard deviation of 1) the Full Gene Coverage, which is calculated by taking the mean coverage across all exonic regions of each gene for each sample, 2) the samples with >95% Coverage at 20x, calculated by determining the percentage of samples that have a minimum depth of 20x coverage across more than 95% of the exonic regions of each gene and 3) the Sites of Interest Coverage, which is a determination of the mean coverage for each variant site that is necessary to define the star alleles described in this report. Data sets used by each group are detailed in Table 1. Coverage differences reported for WGS data the groups are due to different bioinformatics pipelines used to process the data.