**Supplemental Table 4. Novel *CYP2C8, CYP2C9* and *CYP2C19* alleles and suballeles identified in this investigation**

|  |  |  |  |
| --- | --- | --- | --- |
| **Coriell ID**†  | **Star allele** | **Variants in haplotype**‡ | **Method(s) used to establish haplotype** |
| ***CYP2C8*** NM\_000770.3: |
| NA12156 | *\*15.001* | **c.541G>A, p.V181I (rs41286886)** | 10X Linked-Reads |
| NA19917 | *\*16.001* | c.-86A>G (rs11572066)c.-6G>A (rs142144821)**c.992T>, p.I331T (rs146806199)** | Haplotype determined in trioNA19917 (mother), NA19916 (father) and NA19918 (child) |
| NA19213 | *\*17.001* | c.480G>A, p.K160= (rs11572081)**c.730A>G, p.I244V (rs11572102)** | 10X Linked-Reads |
| NA07048 | *\*18.001* | c.-411C>T (rs7912549)c.-271C>A (rs7909236)**c.1081C>T, p.L361F (rs45438799)** | Haplotype determined in trio NA07055 (mother), NA07034(father) and NA07048 (child) |
| NA19213 | *\*1.004* | c.-411C>T (rs7912549) | 10X Linked-Reads |
| NA23878^ | *\*1.005* | c.-411C>T (rs7912549)c.1059C>T, p.H352= (rs11188150) | Haplotype determined in HG03740 homozygous for both variants |
| HG00436 | *\*1.006* | c.-411C>T (rs7912549)c.-370T>G (rs17110453)c.\*24C>T (rs1058932) | Sample heterozygous for one variant¶ |
| HG00436 | *\*1.007* | c.-411C>T (rs7912549)c.-370T>G (rs17110453)c.1230C>T, G410= (rs148515896)c.\*24C>T (rs1058932) | Sample heterozygous for one variant |
| NA19819 | *\*1.008* | c.-411C>T (rs7912549) c.480G>A, K160= (rs11572081) | 10X Linked-Reads |
| HG00276 | *\*1.009* | c.-411C>T (rs7912549)c.-271C>A (rs7909236)c.936C>T, p.L242= (rs200154618)  | 10X Linked-Reads |
| NA12156, NA19122 | *\*1.010* | c.-411C>T (rs7912549)c.\*24C>T (rs1058932) | 10X Linked-Reads |
| NA07348 | *\*1.011* | c.-411C>T (rs7912549)c.\*53dup (rs761474363) | 10X Linked-Reads |
| NA19178, NA18855 | *\*1.012* | c.-411C>T (rs7912549)c.-248dup (rs11572065)c.\*89C>T (rs28399518) | 10X Linked-Reads |
| NA19178 | *\*1.013* | c.-411C>T (rs7912549)c.-133A>C (rs78041161) | 10X Linked-Reads |
| NA19122 | *\*1.014* | c.-411C>T (rs7912549)c.-248dup (rs11572065) | 10X Linked-Reads |
| NA18855 | *\*1.015* | c.-411C>T (rs7912549)c.-248dup (rs11572065)c.\*24C>T (rs1058932) | 10X Linked-Reads |
| NA19007, HG01256 | *\*1.016* | c.-411C>T (rs7912549)c.-271C>A (rs7909236) | 10X Linked-Reads |
| HG01256 | *\*1.017* | c.-122A>C (rs547978262) | Haplotype determined in trio HG01256 (mother), HG01257 (father) and HG01258 (child) |
| HG03272 | *\*1.018* | c.-411C>T (rs7912549)c.-86A>C (rs11572066)c.480G>A, K160= (rs11572081) | Sample heterozygous for one variant  |
| ***CYP2C9*** NM\_000771.4: |
| NA15245  | *\*71* | c.-643T>C (rs4918758)**c.815A>G, p.E272G (rs9332130)****c.1465C>T, p.P489S (rs9332239)** | 10X Linked-Reads |
| NA19239 | *\*1.007* | c.-1885C>G (rs9332093) c.-1537G>A (rs61604699)c.-1188T>C (rs4918758)c.-779C>T (rs112053016)c.-681C>T (rs112218573)c.-510G>A (rs541636827)c.390G>T, p.T130= (rs5031019)c.1389C>T, p.D463= (rs112354725) | 10X Linked-Reads |
| NA19239 | *\*1.008* | c.-633dup (rs9332099) | 10X Linked-Reads |
| NA20296 | *\*1.009* | c.1323C>T, p.A441= (rs2017319) | Sample heterozygous for one variant |
| NA18861§ | *\*1.010* | c.-1565C>T (rs9332096) c.-1188T>C (rs4918758)c.-375T>C (rs9332103) | Haplotype was determined in trio identified in the CMRI data warehouse |
| NA19917 | *\*1.011* | c.-1885C>G (rs9332093)c.-1537G>A (rs61604699)c.-1188T>C (rs4918758)c.-1079G>A (rs528228364)c.-981G>A (rs9332098)c.1425A>T, p.G475= (rs1057911) | 10X Linked-Reads |
| NA21781 | *\*1.012* | c.-1188T>C (rs4918758)c.-29G>T (rs369385517) | 10X Linked-Reads |
| NA19920 | *\*1.013* | c.-433T>C (rs143510067)c.1323C>T, p.A441= (rs2017319) | 10X Linked-Reads |
| NA19226 | *\*8.004* | **c.-1766T>C (rs9332094)**c.-1188T>C (rs4918758) c.-643G>T (rs185008625)**c.449G>A, p**.**R150H (rs7900194)**c.\*88C>T (rs9332241) | 10X Linked-Reads |
| NA12815 | *\*8.005* | c.-643T>C (rs4918758)**c.499G>A, p.R150H (rs7900194)** | Haplotype was determined using trio (Figure 2) |
| NA18966 | n/a | **c.1147A>T, p.K383X (no rsID)** | Sample heterozygous for one variant  |
| ***CYP2C19*** NM\_000769.4: |
| NA19213 | *\*39* | c.-1439T>C (rs17878739)c.-1418C>T (rs3814637)c.-887del (rs17880036)**c.55A>C, p.I19L (rs17882687)****c.365A>C, p.E122A (rs17885179)**c.390G>T, p.T130= (rs17882291)c.903A>G, p.T301= (rs17879239)**c.991A>G, p.I331V (rs3758581)** c.1059C>T, p.H353= (rs17882744)c.\*82T>C (rs17882796) | 10X Linked-Reads |
| NA18861 | *\*1.007* | c.-1418C>T (rs3814637) c.-1410T>G (rs11568730)c.-889T>G (rs11568732)**c.991A>G, p.I331V (rs3758581)** c.1251A>C, p.G417= (rs17886522) | 10X Linked-Reads |
| NA18966 | *\*1.008* | c.-775T>A (rs185375194)c.99C>T, p.P33= (rs17885098)**c.991A>G, p.I331V (rs3758581)** | Sample heterozygous for one variant |
| NA18526 | *\*1.009* | c.-1178del (rs771491246)c.99C>T, p.P33= (rs17885098)**c.991A>G, p.I331V (rs3758581)** | Sample heterozygous for one variant |
| NA19908 | *\*1.010* | c.-1960G>A (rs141540628)c.99C>T, p.P33= (rs17885098)**c.991A>G, p.I331V (rs3758581**) | 10X Linked-Reads |
| NA11832 | *\*2.011* | c.-98T>C (rs4986894)c.99C>T, p.P33= (rs17885098)**c.332-23A>G, splice defect (rs12769205)****c.681G>A, splice defect (rs4244285)**c.990C>T, p.V330= (rs3758580)**c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads |
| NA19917 | *\*2.012* | c.-98T>C (rs4986894)c.99C>T, p.P33= (rs17885098)**c.276G>C, p.E92D (rs17878459)****c.332-23A>G, splice defect (rs12769205)****c.681G>A, splice defect (rs4244285)**c.990C>T, p.V330= (rs3758580)**c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads |
| NA18564 | *\*3.004* | c.-1418C>T (rs3814637) c.-889T>G (rs11568732)**c.636G>A, p.W212X (rs4986893)****c.991A>G, p.I331V (rs3758581)** c.1251A>C, p.G417= (rs17886522) | 10X Linked-Reads |
| NA07029 | *\*8.002* | c.-1986C>T (rs117673124)c.-1418C>T (rs3814637)c.-906A>G (rs186694202)**c.358T>C, p.W120R (rs41291556)****c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads |
| [NA24008, NA24009∥ | *\*9.002* | c.-783C>T (rs11568729)c.99C>T, p.P33= (rs17885098)**c.431G>A, p.R144H (rs17884712)****c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads using a sample identified in the CMRI data warehouse |
| NA19239 | *\*13.002* | c.-1985G>T (rs114276550)c.-1418C>T (rs3814637)**c.991A>G p.I331V (rs3758581)****c.1228C>T, p.R410C (rs17879685)** | 10XLinked-Reads |
| NA19917 | *\*15.002* | c.-1904G>T (rs150732393)c.-1439T>C (rs17878739)c.-1418C>T (rs3814637)c.-1114G>C (rs140164087)c.-887del (rs17880036)c.-783C>T (rs11568729)**c.55A>C, p.I19L (rs17882687)****c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads |
| NA19122 | *\*35.002* | c.-913G>A (rs76822098) **c.55A>C, p.I19L (rs17882687)**c.99C>T, p.P33= (rs17885098)**c.332-23A>G, splice defect (rs12769205)****c.991A>G, p.I331V (rs3758581)** | 10X Linked-Reads |
| NA11832 | *\*38.002* | c.-1418C>T (rs3814637)c.-889T>G (rs11568732)c.-70T>C (rs17886301) | 10X Linked-Reads |
| NA11839 | *\*38.003* | c.-1418C>T (rs3814637)c.-889T>G (rs11568732) | 10X Linked-Reads |

† Samples used for PharmVar submission (note that some alleles may also have been found in other samples). In some instances, the novel haplotype was determined in a sample(s) that was not part of the GeT-RM project due to lack of data to unequivocally determine the phase of variants.

‡ Positions are shown for the transcript and protein level followed by their respective rsID. Core variants (nonsynonymous variants and variants causing a splice defect) are highlighted in bold.

§ A novel suballele was inferred to be present in NA18861. Since no 10X Linked-Read data or pedigree information was available for this sample, a trio from the CMRI data warehouse was utilized to establish the *CYP2C9\*10.010* haplotype.

^ A novel suballele was inferred to be present in NA23878. Since no 10X Linked-Read data or pedigree information was available for this sample, the haplotype was determined in HG03740.

∥ The novel suballele was inferred to be present in NA24008 and NA24009. Since no 10X Linked-Read data or pedigree information was available for this sample, a subject from the Children’s Mercy Research Institute (CMRI) data warehouse was utilized to establish the *CYP2C19\*9.002* haplotype.

¶ ‘Sample heterozygous for one variant’ indicates that there was either only one variant found or that all other variants were homozygous but one; these scenarios unequivocally inform the haplotypes found in a sample.

**Bold font** indicates the defining variant(s) for each allele