All current allele definition tables used downloaded from PharmGKB on 1/12/19 except DPYD gene that was downloaded again on 20/2/20 since the 1/12/19 version was so wrong even the bases on Reference allele. This DPYD new version has less alleles (from 108 alleles to 97 alleles), also less columns.

CYP2B6

1) merge "g.41016810C>A" and "g.41016810C>T" to "g.41016810C>A/T" (and in the same column merging “g.30512C>A” and “g.30512C>T” to "g.41016810C>A/T")

2) delete column "CYP2B7/CYP2B6 hybrid (crossover in intron 4; partial deletion allele)\*\*" and "CYP2B6/CYP2B7 hybrid (crossover in intron 4; duplicated fusion allele)\*\*" because no HGVS or position. and delete \*29,\*30 that are related to these two columns

CYP2C9

1) merge "g.94942213\_94942222delAGAAATGGAA" and "g.94942216A>G" to "g.94942213\_94942222delAGAAATGGAA;g.94942216A>G"

Then delete ;g.94942216A>G to have only "g.94942213\_94942222delAGAAATGGAA” (in order to have only one position range, no semicolon) for easier parsing of positions in the next step

2) edit “g.94942234G>A; g.94942234G>T” to “g.94942234G>A/T” (and in the same column edit “g.8577G>A; g.8577G>A” to “g.8577G>A/T”

3) edit “g.94942249C>T;g.94942249C>G” to “g.94942249C>T/G” (and in the same column edit “g.8592C>T;g.8592C>G” to “g.8592C>T/G”

4) edit “g.94942309G>A;g.94942309G>T” to “g.94942309G>A/T” (and in the same column edit “g.8652G>A;g.8652G>T” to “g.8652G>A/T”

CYP2C19

1) Delete \*36 (full gene deletion) difficult to check if no info for that gene or full gene deletion

2) Delete \*37 due to partial deletion (Need other solutions)

3) Delete column “\*37 Partical Gene Deletion” no need to use

DPYD (use version downloaded on 20/2/20)

1) edit column “g.97740411\_97740414ATGA”

rs72549309 = g.97740411- g.97740418 ATGAATGA/ATGA

<https://www.ncbi.nlm.nih.gov/snp/rs72549309>

wild type should be ATGA 2 copies from g.97740411 to g.97740418, variant has one copy of ATGA. Since we are not sure if it is delATGA from g.97740411 to g.97740414 or delATGA from g.97740415 to g.97740418 (may caused by shifting of aligned positions of repeat sequences between these two locations)

Therefore, we edit column “g.97740411\_97740414ATGA” to “g.97740411\_97740414delATGA” and add one more column of “g.97740415\_97740418delATGA”

Add one more row of allele “c.295\_298delTCAT (\*7)” to separate the cases of " g.97740411\_97740414delATGA" and "g.97740415\_97740418delATGA "

G6PD

1) merge "g.154534419\_154534421delGAG" and "g.154534419G>A" to "g.154534419\_154534421delGAG;g.154534419G>A" Then delete “;g.154534419G>A” to have only "g.154534419\_154534421delGAG” (in order to have only one position range, no semicolon) for easier parsing of positions in the next step

2) merge "g.154533015\_154533038delGGGGTCGTCCAGGTACCCTTTGGT" and "g.154533031C>T" and "g.154533029A>G" and "g.154533025A>G" and "g.154533016G>T" to "g.154533015\_154533038delGGGGTCGTCCAGGTACCCTTTGGT;g.154533031C>T;g.154533029A>G;g.154533025A>G;g.154533016G>T” Then delete ;g.154533031C>T;g.154533029A>G;g.154533025A>G;g.154533016G>T” to have only " g.154533015\_154533038delGGGGTCGTCCAGGTACCCTTTGGT”(in order to have only one position range, no semicolon) for easier parsing of positions in the next step

3) merge "g.154532753\_154532770delGGCCTTGCGCTCGTTCAG" and "g.154532765G>C/T" and "g.154532758T>C" to "g.154532753\_154532770delGGCCTTGCGCTCGTTCAG;g.154532765G>C/T;g.154532758T>C" Then delete ”;g.154532765G>C/T;g.154532758T>C” to have only " g.154532753\_154532770delGGCCTTGCGCTCGTTCAG”(in order to have only one position range, no semicolon) for easier parsing of positions in the next step

ต้องลอง check more ว่า มี del ช่วงไหนที่ยังมี SNP OVERLAP อยู่หรือไม่

NUDT15

1)edit "48037801\_48037802insGAGTCG" to "g.48037801\_48037802insGAGTCG"

2)replace "ref" to "del" for insertion variants (g.48037801\_48037802insGAGTCG, g.48037826\_48037827insCGGG, g.48041103\_48041104insG)

3) edit “g.48040982delA” \*18 from “del” to “delA”

4) edit “g.48037783\_48037788delGGAGTC” \*9 from “del” to “delGGAGTC”

RYR1

1) separate “g.38499646\_38499648delGAG or g.38499649\_38499651delGAG” to 2 columns "g.38499646\_38499648delGAG" and "g.38499649\_38499651delGAG"

And add one more rows of allele “c.7039\_7041delGAG or c.7042\_7044delGAG” to separate the cases of "g.38499646\_38499648delGAG" and "g.38499649\_38499651delGAG"

Edit “del” to “delGAGA”

TPMT

1) edit “g.18138969C>T; g.18138969C>G” to “g.18138969C>T/G” (and in the same column edit “g.21175G>A; g.21175G>C” to “g.21175G>A/C”

2) edit “g.18130687T>C; g.18130687T>G” to “g.18130687T>C/G” (and in the same column edit “g.29457A>G; g.29457A>C” to “g.29457A>G/C”

UGT1A1

1) repeat at promoter

\*1 (TA 6 copies) chr2: CHR2:233760233-233760247 CA(TA)6TAA

\*28 (TA 7 copies) CA(TA)7TAA

\*37 (TA 8 copies) CA(TA)8TAA

\*36 (TA 5 copies) CA(TA)5TAA

Genotypes shown in VCF for all possible combinations of repeat alleles

|  |  |  |  |
| --- | --- | --- | --- |
| Variant type | diplotype | Genotype in VCF | call |
| -/- | \*1/\*1 | 33 C C | no variant |
| -/ins | \*1/\*28 | 33 C CAT | 0/1 |
| -/del | \*1/\*36 | 33 CAT C | 0/1 |
| -/ins | \*1/\*37 | 33 C CATAT | 0/1 |
| ins/del | \*28/\*36 | 33 CATAT C | 0/1 |
| ins/ins | \*28/\*37 | 33 CAT CATAT ?? | 0/1 |
| del/ins | \*36/\*37 | 33 CATATAT C | 0/1 |
| ins/ins | \*28/\*28 | 33 C CAT ?? | 1/1 |
| del/del | \*36/\*36 | 33 CAT C | 1/1 |
| ins/ins | \*37/\*37 | 33 C CATAT | 1/1 |
|  |  |  |  |

* Diplotypes with \*1 🡪 can be shown in allele def table as ins/del compared with \*1
* Combinations between ins/del (ex. \*28/\*36) shown as relative difference between two alleles
* Combinations between ins/ins of del/del 🡪 how will it be shown in VCF?

IFNL3

* Edited Position at NC\_000019.10 (Homo sapiens chromosome 16, GRCh38.p2) to Position at NC\_000019.10 (Homo sapiens chromosome 19, GRCh38.p2)

CFTR

* Update CFTR allele definition (Download on 8 Dec 2020)