Notes on CPIC development

Most recent updates:

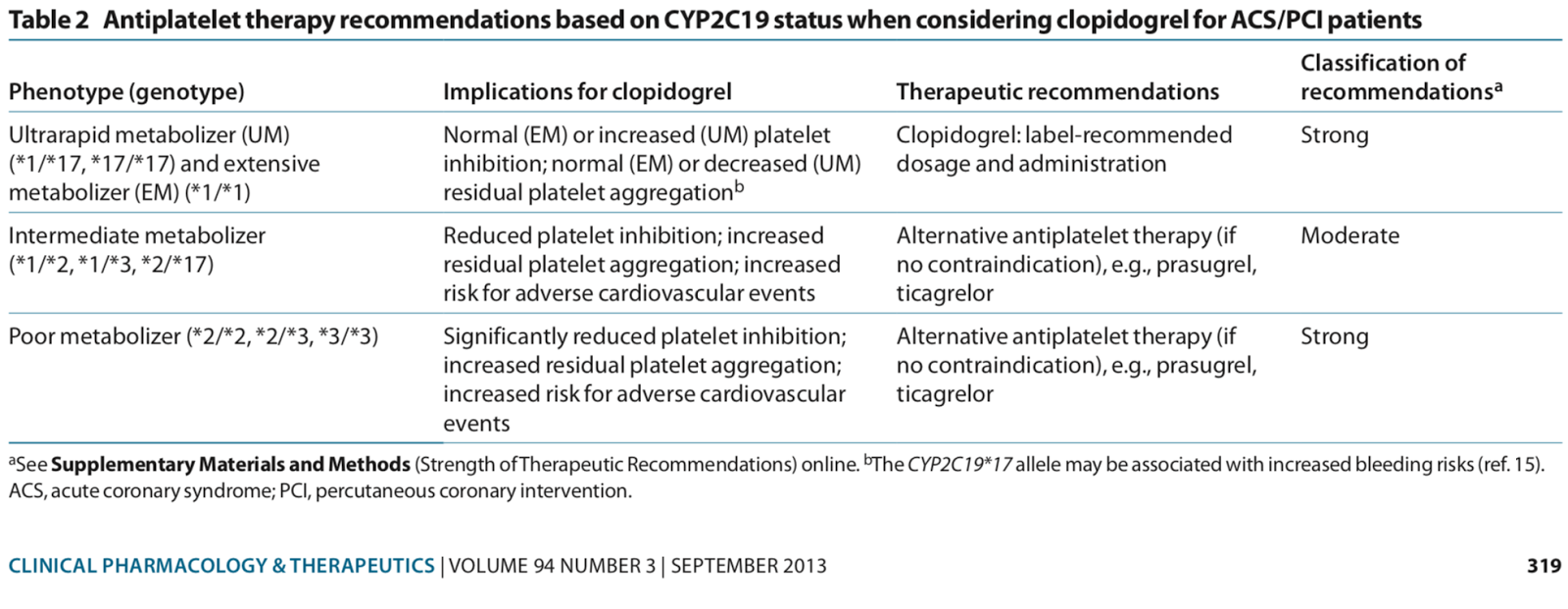
1. 1-31-2018 Improved CPIC\_GPWizard(v1.3).py released. Payload has been upgraded to skeleton\_geno\_pheno-base(v1.2).py
2. 1-31-2018 Some excel geno-pheno tables contain complex alleles. Created ‘Table\_tool\_kit’, a jupyter notebook file that can return all unique values contained in each column to get an overview of all possible values contained in the table.
3. 1-31-2018 Created skeleton\_geno\_pheno-base(v1.2).py to handle complex alleles containing non numerical characters. Previously, it could only handle numerical values, and the comparison step to reorder the alleles based on numerical value would cause an error.
4. Improved CPIC\_GPWizard(v1.2).py. Fixed bug that existed on payload which prevented users from inputting allele specifications without ‘\*’ format.
5. Removed “phenotype” dictionary key. Only diplotype, allele1 and allele2 retained
6. Edited Input xml to reflect reduction in parameters passed. Deleted phenotype and adjusted number of parameters from 4 🡪 3.
7. Edited Build.py in MakeKO file to be compatible for python3.
8. Created CPIC\_GPWizard(v1.1).py. This wizard has been updated to contain the latest updated skeleton code for the genotype to phenotype payload, as well as an updated inputxml.
9. Emailed James Hoffman about questions.
10. CPIC\_GPWizard(v1.1).py modified to generate ark ID using file name
11. Comments explaining how to use the skeleton geno-pheno (payload code) results in an error when running. REMOVED for the Wizard.
12. Early stage prototype phenotype to recommendation KO created for UGT1A1
13. Much reformatting needed for UGT1A1: Removed superscript “c” and “d”, and fixed “metazolizer” to “metabolizer”.

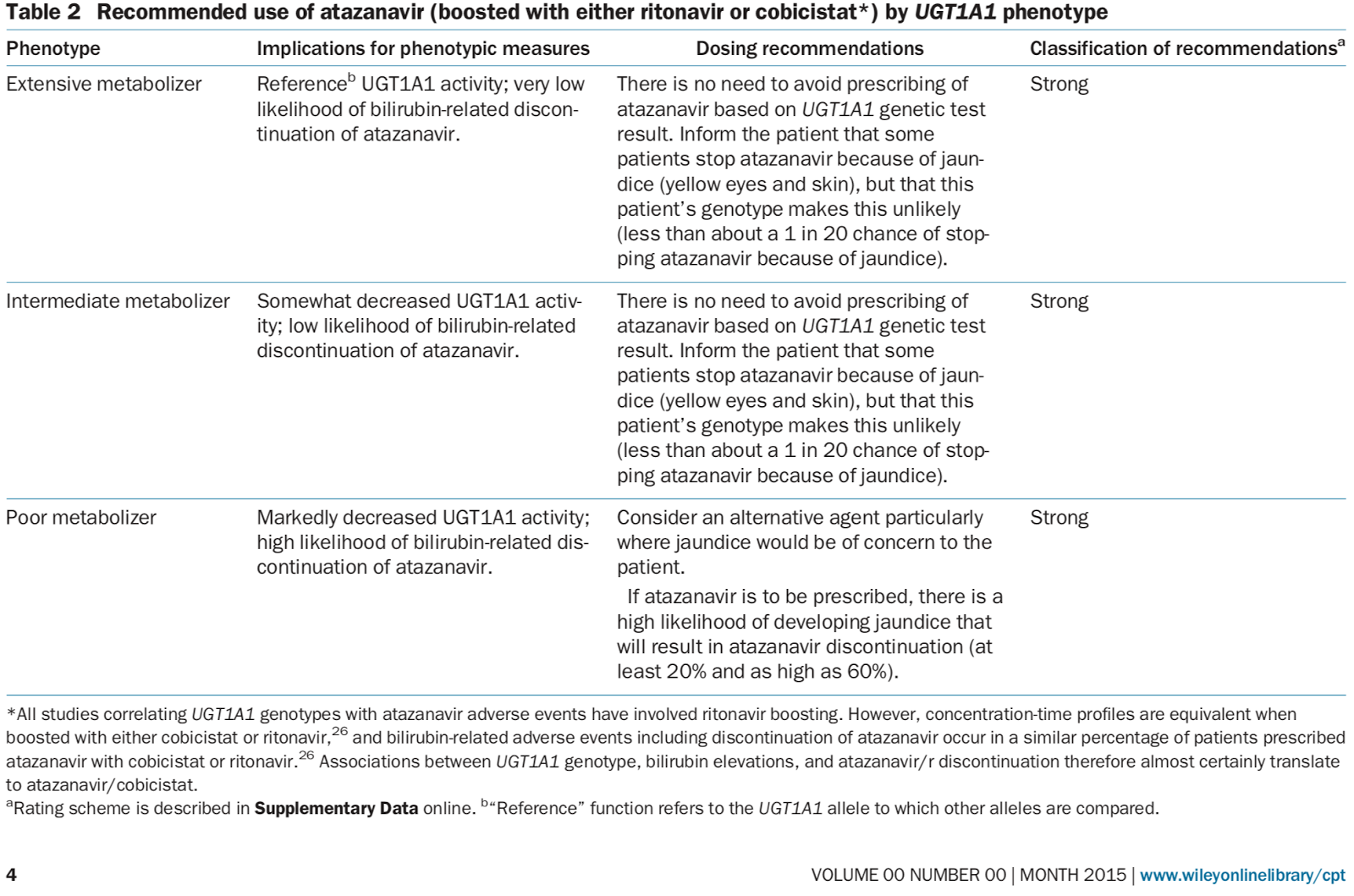
Current work:

1. When the gene folder is located in a different file, the wizard will still make the KO, but the naming will be off. Perhaps play around with regular expressions to solve this issue.
2. Create KO for CYP2C19 and CYP2D6. Generate altered CYP2D6 to contain “Indeterminate” entries.
3. Working to create a KO for phenotype to recommendation
   1. Begin with easy model UGT1A1 (DONE)
   2. CYP2C19 (ambiguities)
   3. CYP2D6 (ambiguities)
4. Email: James Hoffman for table data (cc Allen in email) – no response yet
5. Re-visited the Wizard. How can we customize this?
6. Testing KO
7. Testing: Have Tina follow READ\_ME for CPIC\_Wizard to see if it makes sense.

Questions/Concerns:

1. UGT1A1 Miss spelled “metazolizer”
2. Difference in formatting from the 2013 paper vs 2015 paper. Something to keep in mind as we create phenotype to recommendation KO because the header names are different. Need to set a standard.





1. The CYP2C19 table has rapid metabolizer, intermediate metabolizer, likely intermediate metabolizer (treat as intermediate according to CPIC people), likely poor metabolizer, and poor metabolizer. The paper on CYP2C19 published in 2013 contains extensive metabolizer which does NOT exist in the table.