# Outline

* A) 6 CPIC guidelines (high level view)
* Clinical decision support implementer
* B) WGS as input data (69 subjects with whole genome genome sequence)
* C) So what?

# Background

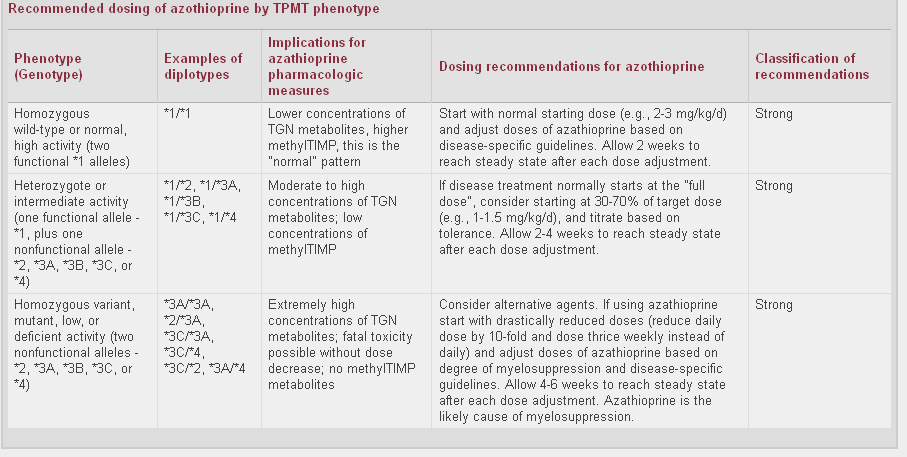
* <http://www.pharmgkb.org/page.action?key=cpicGeneDrugPairs>
* Pharmacogenes: TPMT, CYP2C19, CYP2C9, VKORC1, CYP2D6, HCP5, and SLCO1B1.
* Implementer:
* if <DRUG\_ORDER (shortlist)> and <G\_DATA> then EXECUTE GUIDELINE
* If <SIGNIFICANT> then <ALERT>
* Actions: 1.change drug, 2.keep the drug and change dose
* “Avoid codeine use due to lack of efficacy. Consider alternative analgesics such   
  as morphine or a nonopiod”
* Obtain genetic data 🡪 determine haplotypes (combine into diplotype)   
  🡪 assign “metabolizer type” 🡪 action (*recommendation table*)
* Pure implementer site (no own PGx research), not a formal member of PGRN
* Clinical informatics background (implement hypertension guideline, rheumatology, preventive care, drug-drug interactions)

# Decision tables (June 2012)

*Haplotypes, diplotype, tag allele* (tag variant - position that determines the haplotype)

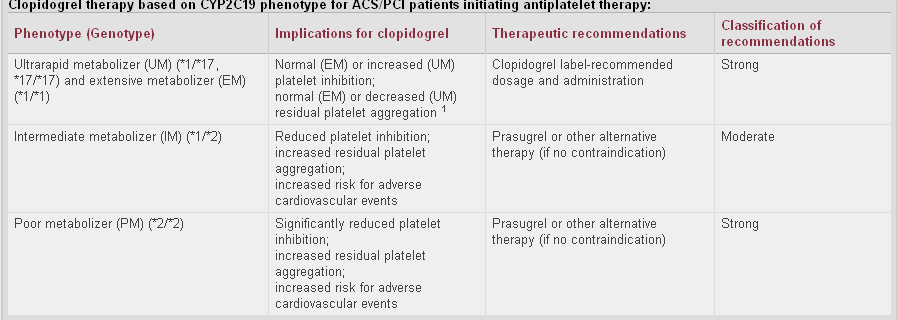
### TMPT

Normal 1 Dose 1 Change 1



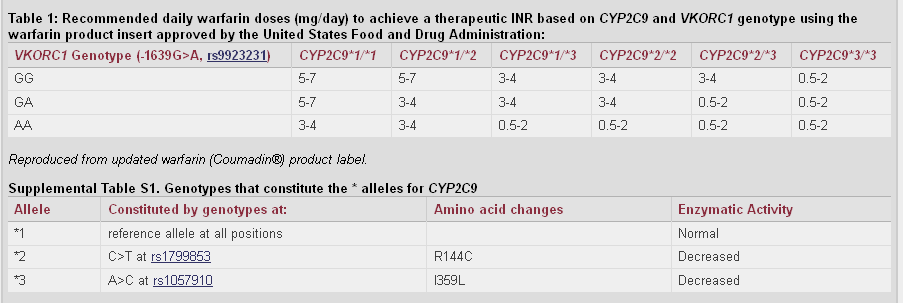
### Clopidrogel

Normal 1 Change 2



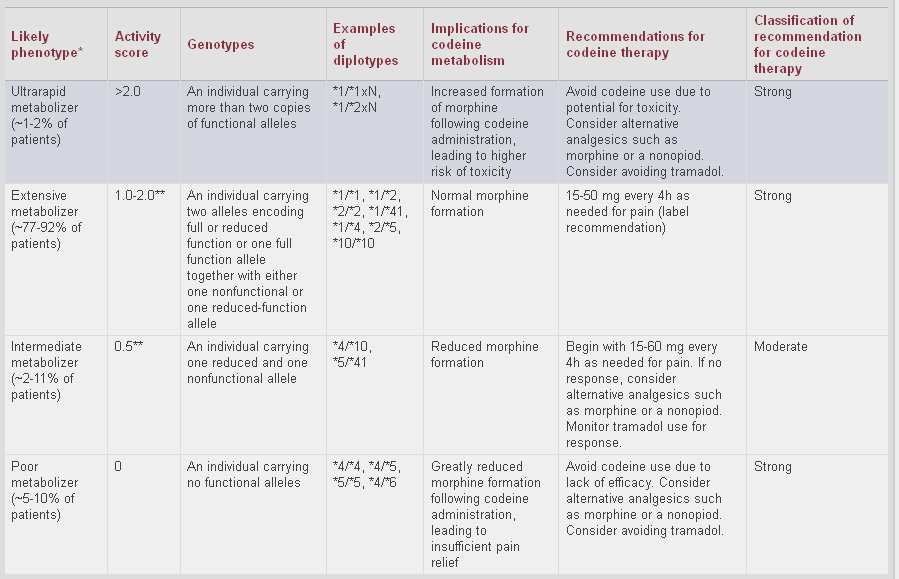
### Warfarin

Always adjust dose



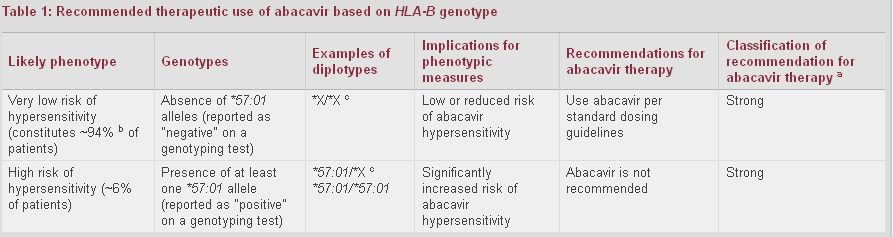
### Codeine

Normal 2 Change 2



### Abacavir

Normal 1 Change 1

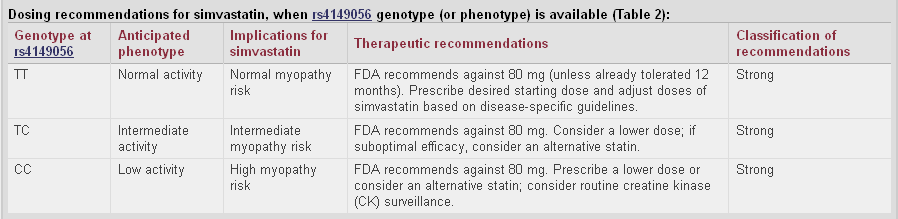


### Simvastatin

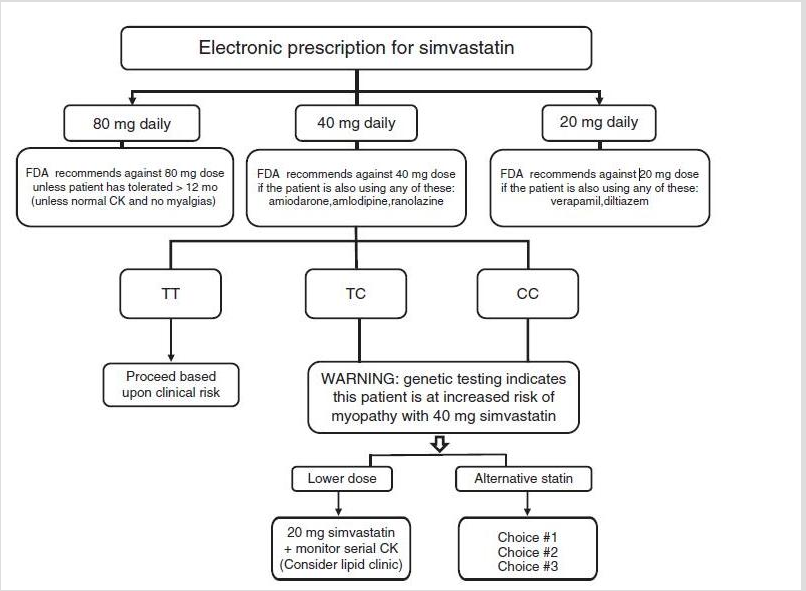
“ADE prevention guideline”

Kicks in only if dose is 40mg or greater

THEN Normal 1 Dose 1 Change 1

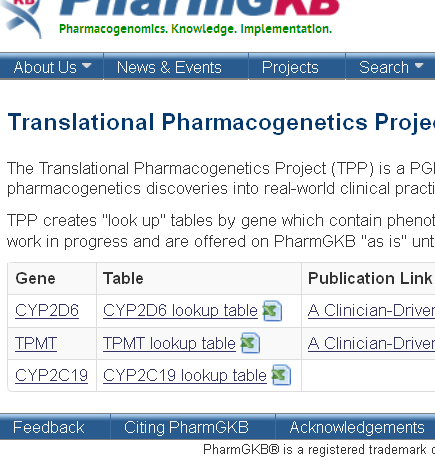


PLUS flowchart

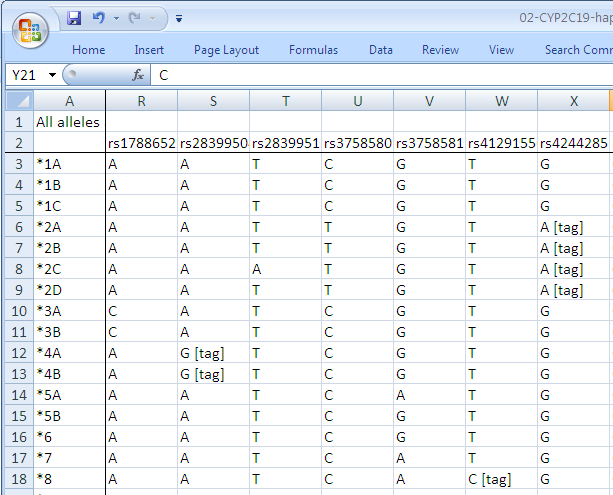


# Look up tables (September 2012)

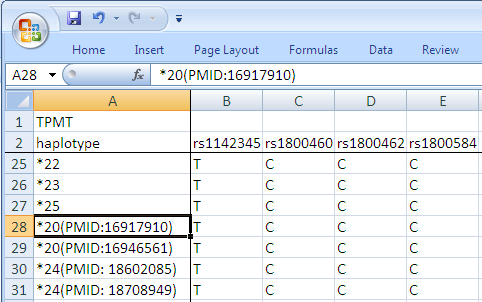
* <http://www.pharmgkb.org/page/tppTables>



# Tag alleles



## Surprise



Web vs. CSV file differences, emailed PharmGKB team

# Overview table

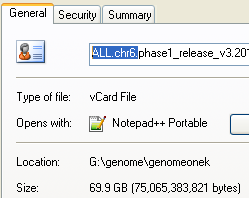
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Drug/drug class** | **Gene** | **EntrezGene ID** | **Chromosome** | **Strand** | **PDG: Number of tag variations with clinical recommendations** | **PGKB: count of haplotypes** | **PGKB: count of tag alleles** | **VCF: Count of variants** |
| thipurines | [TPMT](http://useast.ensembl.org/Homo_sapiens/geneview?gene=TPMT) | [7172](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=7172%5buid%5d) | [6](http://useast.ensembl.org/Homo_sapiens/mapview?chr=6) | -1 | 9 | 29 | 9 | 201 |
| clopidrogel | [CYP2C19](http://useast.ensembl.org/Homo_sapiens/geneview?gene=CYP2C19) | [1557](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=1557%5buid%5d) | [10](http://useast.ensembl.org/Homo_sapiens/mapview?chr=10) | 1 | 2 | 33 | 37 | 1392 |
| warfarin | [CYP2C9](http://useast.ensembl.org/Homo_sapiens/geneview?gene=CYP2C9) | [1559](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=1559%5buid%5d) | [10](http://useast.ensembl.org/Homo_sapiens/mapview?chr=10) | 1 | 3 | 34 | 20 | 481 |
| warfarin | [VKORC1](http://useast.ensembl.org/Homo_sapiens/geneview?gene=VKORC1) | [79001](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=79001%5buid%5d) | [16](http://useast.ensembl.org/Homo_sapiens/mapview?chr=16) | -1 | 13 | 10 | 30 |
| codeine | [CYP2D6](http://useast.ensembl.org/Homo_sapiens/geneview?gene=CYP2D6) | [1565](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=1565%5buid%5d) | [22](http://useast.ensembl.org/Homo_sapiens/mapview?chr=22) | -1 | 9 | 11 | 10 | 242 |
| abacavir | [HCP5](http://useast.ensembl.org/Homo_sapiens/geneview?gene=HCP5) | [10866](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=10866%5buid%5d) | [6](http://useast.ensembl.org/Homo_sapiens/mapview?chr=6) | 1 | 1 | 2 | 1 | 1370 |
| simvastatin | [SLCO1B1](http://useast.ensembl.org/Homo_sapiens/geneview?gene=SLCO1B1) | [10599](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=search&db=gene&term=10599%5buid%5d) | [12](http://useast.ensembl.org/Homo_sapiens/mapview?chr=12) | 1 | 1 | 34 | 25 | 1021 |

# CPIC guidelines implementation conclusions

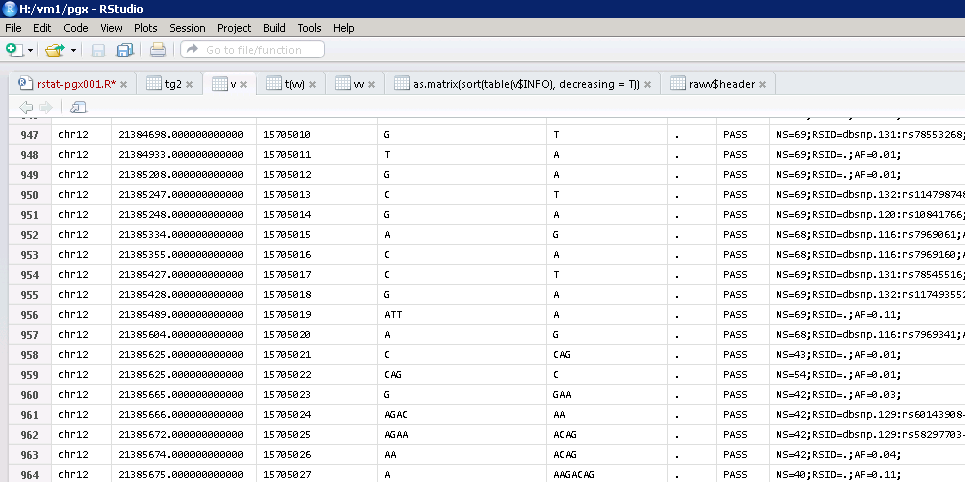
* Drug terminology codes
* if <DRUG\_ORDER (shortlist)>
* RxNorm codes (or drug classes: NDR-RT or ATC codes)
* CPIC guidelines differ in degree of variability addressed by the guideline
* Address all possible situations given a list of gene positions (tag variants)
* Lookup table
* Highlight and address only specific situations (simvastatin)
* “within the guideline” / “outside the guideline”
* Star notation system
* Curation authority (PharmgKB, cypalleles.ki.se, other?)
* Computable solution
* GetUpdatedHaplotypeTable(‘<http://www.pharmgkb.org/download/PA128?data=haplotype&format=csv>’)
* Long-term solution for determining haplotypes (in 2015)
* CPIC guidelines are developed with genotyping data in mind

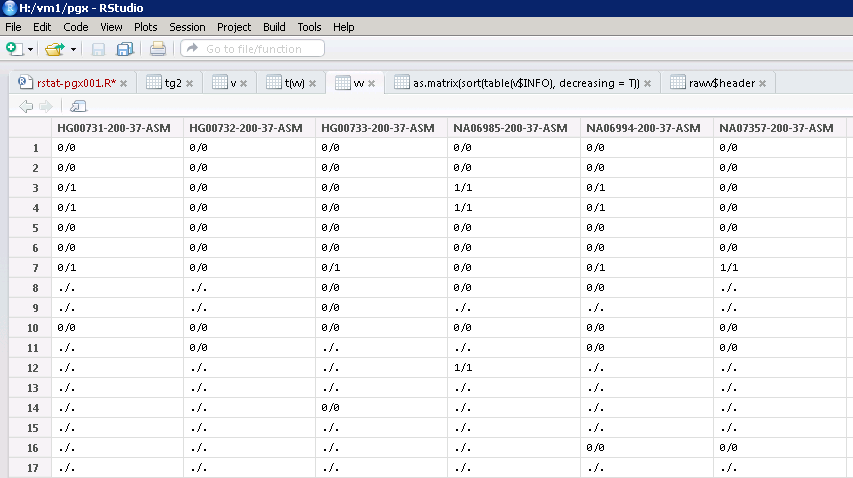
# B) WGS as input data

* Variant Call Files (VCF)
* Complete Genomics, 7.8GB (all chromosomes, 69 patients)
* 1000G (70GB, only chromosome 6) (1.1TB)
* Tabix utility to restrict it to 7 genes
* The filtered VCF file sizes ranged from 13kB to 500kB
* Two scenarios
* SIMPLE: VCF data (WGS) 🡪 genotyping data 🡪 PGx [dosing] recommendations
* COMPLEX: VCF data (WGS) 🡪 PGx [dosing] recommendations



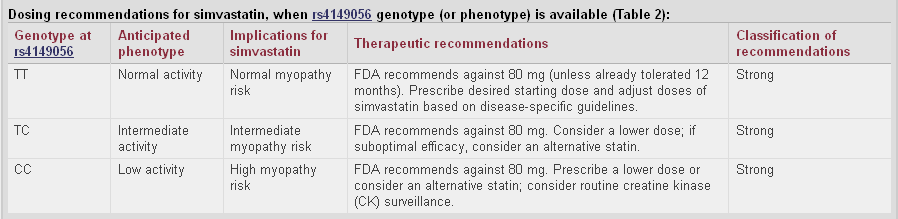
# VCF file example





# Generating recommendations

* Example: Simvastatin guideline
* Simple scenario: in 16% of patients (11/69): dose alteration (if >40mg simvastatin is being ordered)
* 1 tag variant, all scenarios are covered by the guideline, always “within the guideline”,   
  16% of patients get a clear message
* Complex scenario: in an average patient, there is 259 (SD:±76) other variations (other than rs4149056)
* of which 58.7 (SD:±10.2) are not present in dbSNP
* 1021 variants, every patient is “outside the guideline”  
  0% of patients get a clear message
* rare mutations (some of them missense (nonsynonymous))
* frameshift insertions or deletions?
* early stop codon (nonsense mutation)
* There will never be a guideline covering all possibilities (meta-guideline)



# WGS as input data – conclusions

* Existing CPIC guidelines were written with genotyping input data in mind
* When WGS data is simplified to emulate genotyping data, CPIC guidelines could be applied
* WGS shows large amount of additional mutations
* some of which cannot be ignored and probably can be   
  interpreted with current knowledge (e.g., early stop codon)
* WGS is the only method that can see insertions and deletions (unless a custom genotyping array is used)
* We have today 1000+ whole genomes at our institution (policy for incidental findings, cancer domain)
* Potential error:
* genotyping data predicts *gain of function* of one allele 🡪 lower dose (rs1234567)
* WGS shows *loss of function* due to early stop codon due to insertion (10 bp prior rs123467)

# So what?

* Pre-requisite fact: dealing with WGS sequencing data is an unsolved problem (outside PGx domain)
* Do you have similar WGS findings? (rare mutations, indels)
* Approach
* (1) Managing combinations of variants (star alleles)

vs.

* (1) managing individual variants and (2) combining pipeline
* Meta CPIC guideline?
* Instead of: genetic data 🡪 haplotypes/diplotypes 🡪 “metabolizer type” 🡪 action
* Interpretation pipeline
* “Inside the guideline”
* Known and well described variants
* General approach to situations “outside the guideline”
* A truly unknown variants (no clear message)
* Predictable biological phenomena with known impact   
   (synonymous rare mutation)
* What to say to the clinician when genetic variation exists and there is no clear clinical message can be generated
* Implementer’s decision (local) vs. expert concensus (CPIC meta-guideline)

The simvastating PDG does state that other mutations in the pharmacogene impact dosing but does not offer a

# Conlusions

* Part A) Lookup tables are important part of a published CPIC guideline (implementers)
* PharmgKB as single source of knowledge (lkup table integration)
* Part B)