Comparison of Genome Sequencing and Clinical Genotyping for Pharmacogenes

Wenjian Yang, PhD

Lab of Mary V. Relling, PharmD
St Jude Children's Research Hospital, Memphis, TN

Clinical Pharmacology & Therapeutics

Explore this journal >

Research

Comparison of genome sequencing and clinical genotyping for pharmacogenes

W Yang, G Wu, U Broeckel, CA Smith, V Turner, CE Haidar, S Wang, R Carter, SE Karol, G Neale, KR Crews, JJ Yang, CG Mullighan, JR Downing, WE Evans, MV Relling

First published: 18 August 2016 Full publication history

DOI: 10.1002/cpt.411 View/save citation

http://www.ncbi.nlm.nih.gov/pubmed/27311679

Background

- St Jude Children's Research Hospital has been among the first to implement preemptive genomic testing to incorporate pharmacogenetics results in the medical record to assist in patient care.
- Recent St Jude protocol "*PGEN4Kids*" has implemented pharmacogenetics testing using pharmacogene-directed arrays such as the Affymetrix DMET plus array.
- Recently next generation sequencing (NGS) technology has experienced great advances, with lower cost and higher accuracies.
- Many NGS data have been generated at St Jude as part of research projects, such as Pediatric Cancer Genome Project (PCGP).

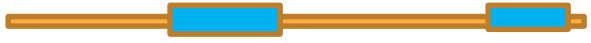
Objective

1. To examine the interrogation from genome sequencing technology for actionable pharmacogenes.
2. To compare the concordance between genotypes generated by genome sequencing and our clinical array-based genotyping results.

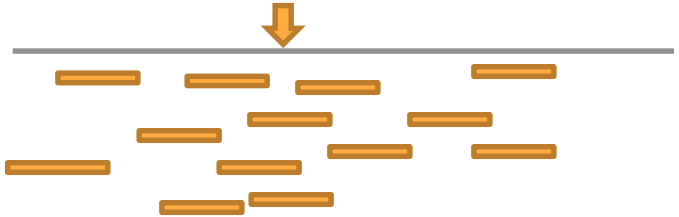
Basic Introduction of Next Generation Sequencing (NGS):

Whole Genome Sequencing (WGS)

1.



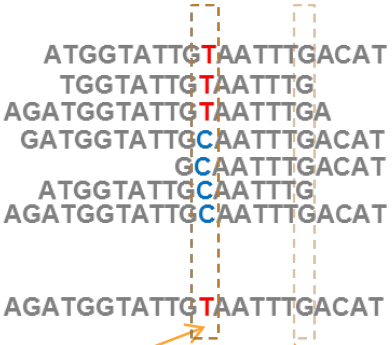
2.



Generate sequence reads (e.g. 100bps)

3.

Reads



Reference Genome

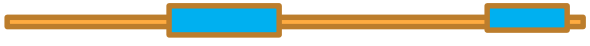
Ref/Alt -> C/T

Ref/Ref -> G/G

Align with reference genome and generate genotyping calls

Whole Exome Sequence (WES)

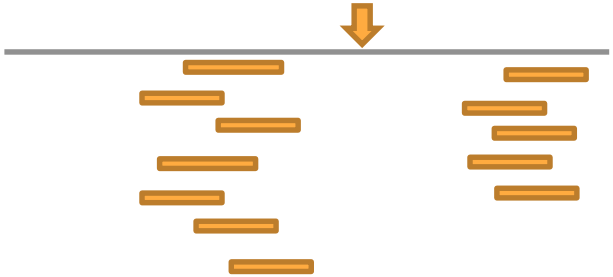
1.



2a.

Targets (exome) captured by hybridization

2b.



Generate sequence reads (e.g. 100bps)

Basic Quality Controls metrics in NGS:

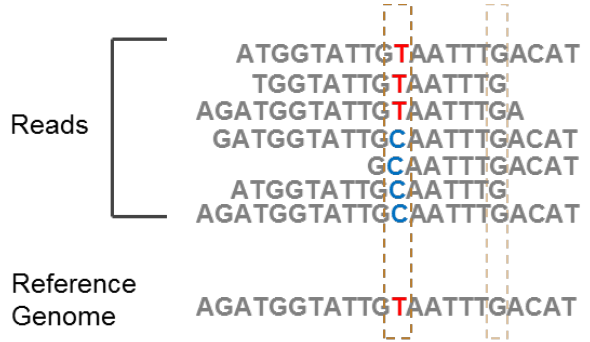
Coverage (read depth)

- Average WGS, 30X
- Average WES, 60X
- Read depth < 10X, considered “NoCall”

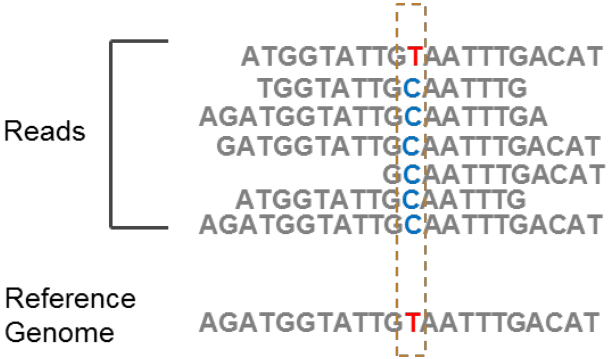
minor allele fractions (MAFrac)

- Heterzyous genotypes should be close to 0.5 (50%)
- Low allele fraction is questionable, suggesting contamination, sequencing error, etc.

Other QC: strand bias, Base quality, etc

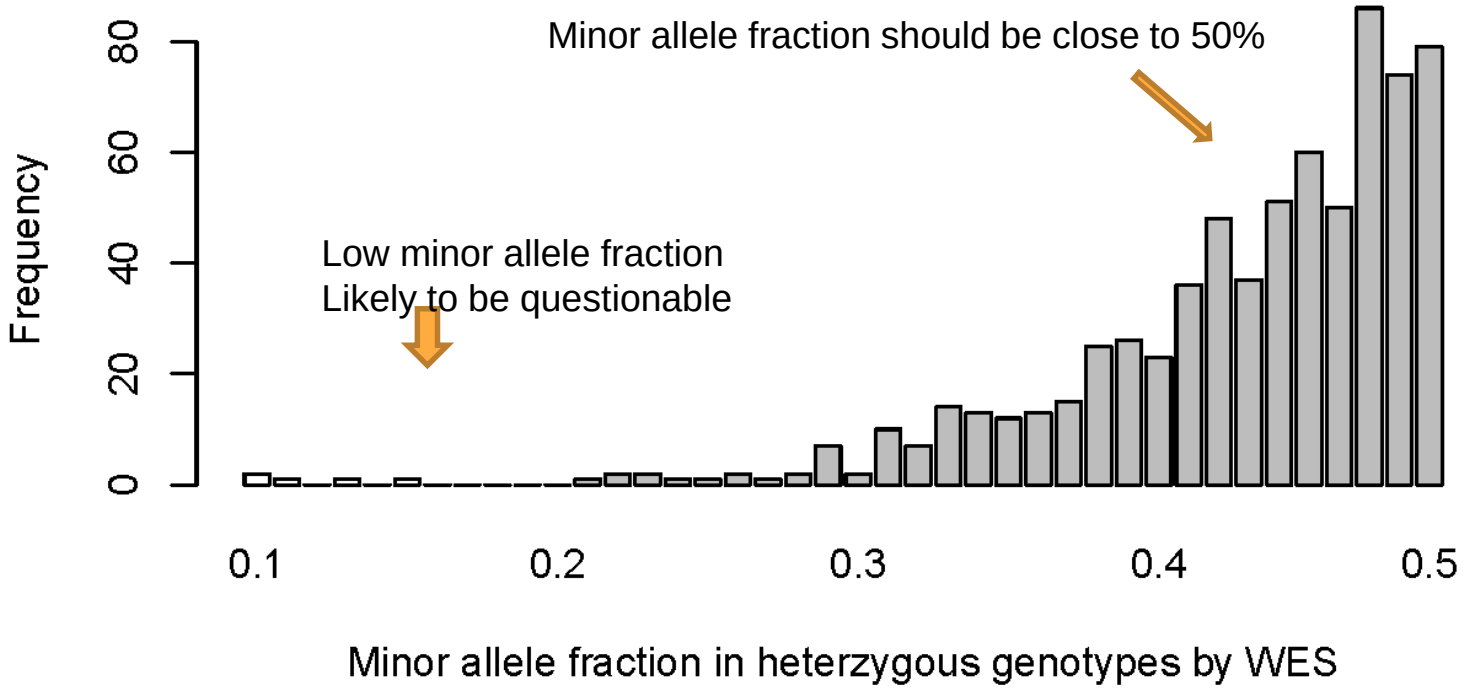


Read depth: 7
MAFrac: 3/7 = 42.9%



Read depth: 7
MAFrac: 1/7 = 10.2%

Distribution of minor allele fraction of heterozygous calls in NGS



Patient Data

Clinical Genotyping (Affymetrix DMET Plus Array v1)

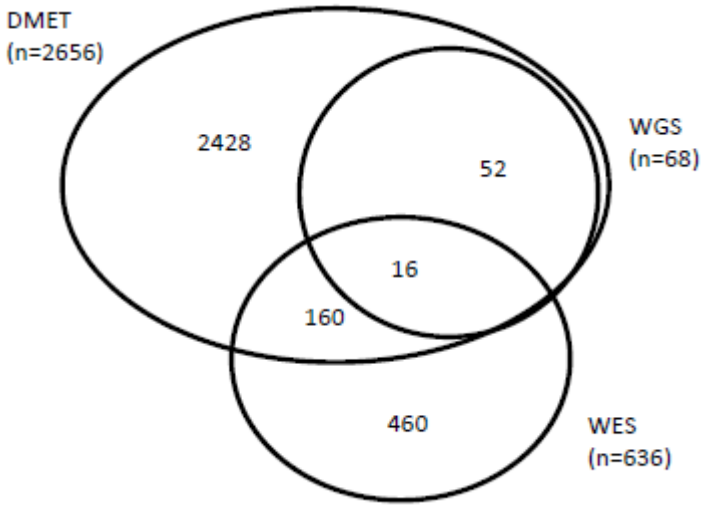
- N = 2656 (1319 whites, 998 blacks, 232 Hispanics)

Whole Genome Sequencing (WGS)

- N = 68 (44 whites, 18 blacks)
- all 68 patients have both DMET array and WGS

Whole Exome Sequencing (WES)

- N=636 (396 whites, 95 blacks, 86 Hispanics)
- 176 patients have both DMET array and WES



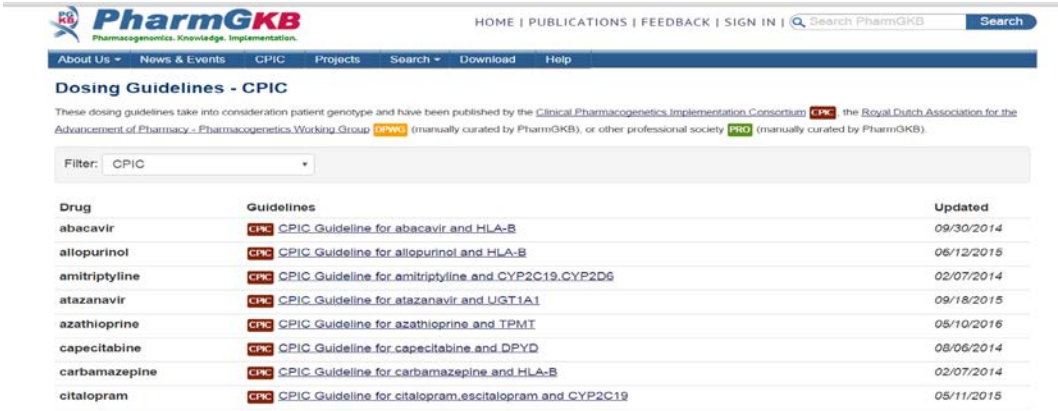
CPIC Important Genes and Variants

CPIC Important Genes: (n=13)

CFTR, CYP2C19, CYP2C9, , CYP2D6, CYP3A4, DPYD, G6PD,

HLA-B, IFNL3, SLCO1B1, TPMT, UGT1A1, VKORC1

(collected from https://www.pharmgkb.org/view/dosing-guidelines.do?source=CPIC as of 07/01/2015)



CPIC important variants

Based on gene activities associated with variants from supplemental table of published CPIC guidelines

- Variants associated with **increase/decreased/no-function** were considered important.
- Exclude variants with **unknown** function and **normal** functions.

SUPPLEMENTAL TABLE S2. ASSOCIATION BETWEEN ALLELIC VARIANTS[A] AND CYP2D6 ENZYME ACTIVITY

Functional Status (2, 7)	Activity Value[c,d]	Alleles
Increased function	>1	**1xN, *2xN, *35xN, *45[g]xN*
Normal or Increased function	1 or >1[h]	**9xN, *10xN, *17xN, *29xN, *41xN*
Normal function[b]	1	**1[e], *2, *27, *33, *34[f], *35, *39[f], *45[g], *46[g], *48, *53*
Decreased function	0.5	**9, *10, *14B, *17, *29, *41, *49, *50, *54, *55, *59, *72*
No-function	0	**3, *3xN, *4, *4xN, *5, *6, *6xN, *7, *8, *11, *12, *13, *14A, *15, *18, *19, *20, *21, *31, *36, *36xN, *38, *40, *42, *44, *47, *51, *56, *57, *62, *68, *69, *92, *100, *101*
Unknown	N/A	**22, *23, *24, *25, *26, *28, *30, *32, *37, *43, *43xN, *52, *58, *60, *61, *63, *64, *65, *70, *71, *73, *74, *75, *81, *82, *83, *84, *85, *86, *87, *88, *89, *90, *91, *93, *94, *95, *96, *97, *98, *102, *103, *104, *105*

* CPIC Guideline for codeine and *CYP2D6*

CPIC Important Variants: (n=127)

- 103 Single Nucleotide Variation (SNV) (95 exonic)
- 21 Indels/repeats (20 exonic)
- two structural variants (*CYP2D6*), Copy Number Variation (CNV) and *CYP2D6/2D7* hybrid
- one haplotype (*HLA-B*)

Gene	Number of CPIC important variants			
	SNV (exonic)	Indel (exonic)	Other	Total
CFTR	10 (10)	2 (2)		12
CYP2C19	8 (7)	0		8
CYP2C9	10 (10)	2 (2)		12
CYP2D6	26 (24)	13 (13)	2 structural variations	41
CYP3A5	2 (1)	1 (1)		3
DPYD	10 (10)	2 (2)		12
G6PD	7 (7)	0		7
HLA-B	0	0	1 haplotype	1
IFNL3	2 (0)	0		2
SLCO1B1	12 (11)	0		12
TPMT	15 (15)	0		15
UGT1A1	0	1 (0)		1
VKORC1	1 (0)	0		1
Total	103 (95)	21 (20)	3	127

Analysis pipelines used to generate genotypes

Affymetrix DMET Plus Array v1 (231 genes, 1936 variants)

- DMET Console software from Affymetrix

Whole Exome and Whole Genome Sequencing

- GATK v3.4 for SNVs and Indels, following best practice guideline, with recommended parameters and quality control steps.
- XHMM and CONSERTING for CNV estimation.
- Polysolver and OptiType for inferring *HLA-B* alleles.

1.CFTR

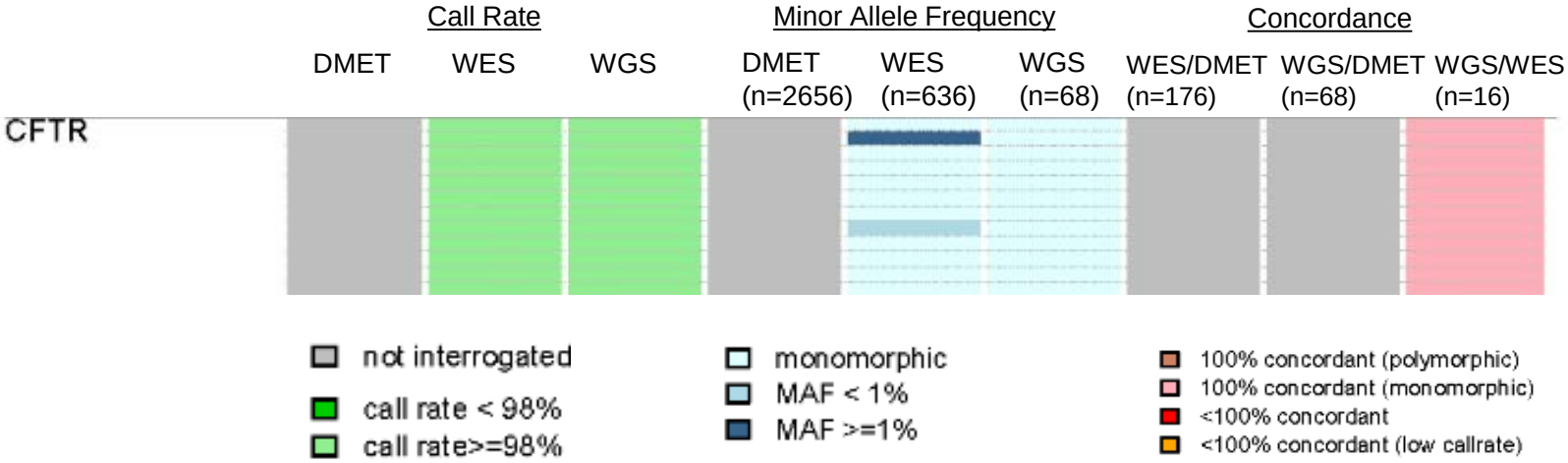
DMET: not interrogated
WES: good coverage
WGS: good coverage

No discordant genotypes between WGS and WES

Drug: ivacaftor

CPIC important Variants (n=12):

- 10 exonic SNV
- 2 exonic indels



G178R
F508del(TCT)
F508del(CTT)
S549R(A>C)
S549N
S549R(T>G)
G551S
G551D
G1244E
S1251N
S1255P
G1349D

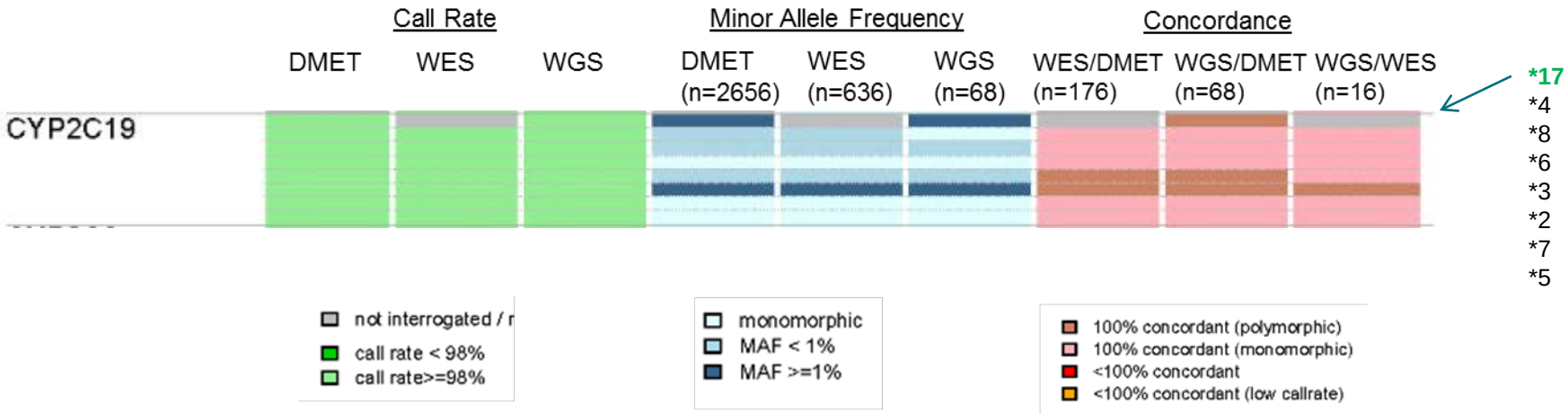
2.CYP2C19

DMET: good coverage
WES: missing important intronic variant
CYP2C19 *17, associated with increase activity
WGS: good coverage

No discordant genotypes were observed

Drug: Clopidogrel, Amitriptyline, citalopram, clomipramine, doxepin, imipramine, setraline, trimipramine

CPIC Important Variants (n=8):
- 8 SNV (7 exonic)



3. CYP2C9

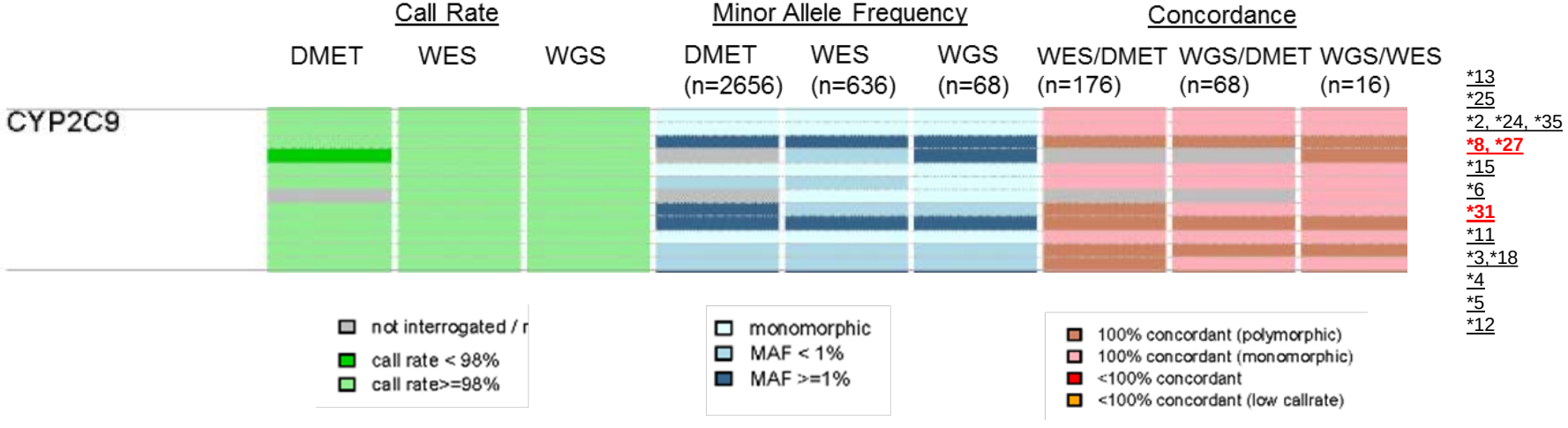
DMET: low call rate on R150H (*8) and not interrogated very rare variant I327T (*31); both “possible decreased activity”
WES: good coverage
WGS: good coverage

No discordant genotypes were observed.

Drug: Warfarin, Phenytoin

Important Variants (n=12):

- 10 exonic SNVs
- 2 exonic Indels



4. *CYP2D6*

Drug: amitriptyline, clomipramine, codeine, desipramine, doxepin, fluvoxamine, imipramine, nortriptyline, paroxetine, trimipramine.

CPIC Important Variants (n=41):

- 26 SNV (24 exonic)
- 13 exonic Indels
- 2 structural variations (CNV, *CYP2D6/2D7* hybrid)

Clinical Genotyping:

- Affymetrix **DMET** interrogated 23 SNV/Indels.
- CNV and *CYP2D6/2D7* were interrogated by add-on qPCR assay.

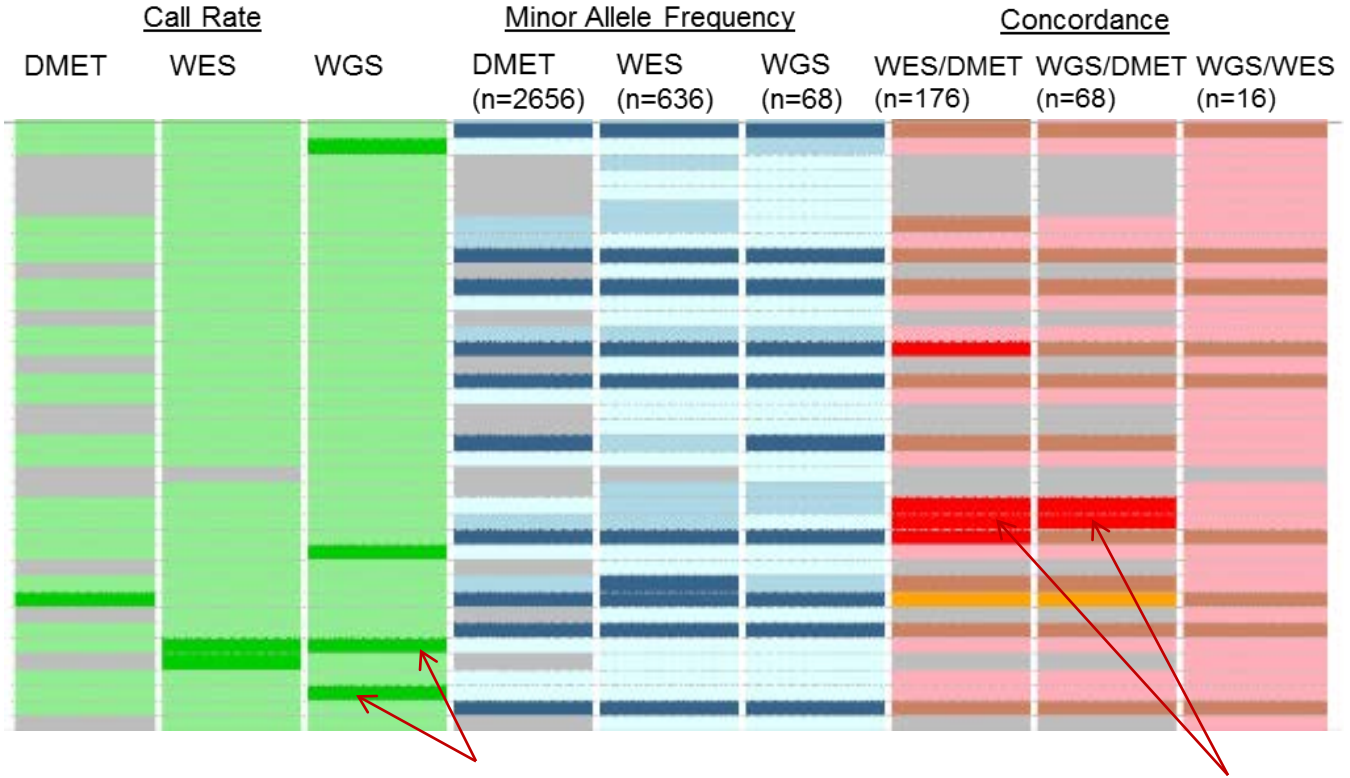
WES:

- interrogated 36 SNV/indels.
- CNV can be inferred, *CYP2D6/2D7* not interrogated.

WGS:

- Interrogated 35 SNV/Indels.
- CNV can be inferred, *CYP2D6/2D7* not interrogated.

CYP2D6: SNVs and Indels (n=39)



lower call rate from WGS likely due to CNV

discordant calls were observed in 7 WES and 3 WGS genotypes.

CYP2D6 discordant genotyping calls between DMET and WES

Gene	Allele	dbSNP	DMET Call	WES Call	Reference Allele Count	Alternative Allele Count	**Minor Allele Fraction**	Comment
CYP2D6	**20* (1973insG)	rs72549354	T/T	T/TC	364	57	**13.5%**	WES low minor allele fraction
CYP2D6	**20* (1973insG)	rs72549354	T/T	T/TC	278	51	**15.5%**	WES low minor allele fraction
CYP2D6	**4* (1846G>A)	rs3892097	T/T	C/T	10	83	**10.8%**	WES low minor allele fraction
CYP2D6	**4* (1846G>A)	rs3892097	T/T	C/T	15	133	**10.1%**	WES low minor allele fraction
CYP2D6	**2* (R296C)	rs16947	A/A	A/G	71	9	**11.3%**	WES low minor allele fraction
CYP2D6	**40* (1863_1864ins)	rs72549356	-/18bps	-/-	138	0	0.0%	Reason for discrepancy unclear
CYP2D6	**40* (1863_1864ins)	rs72549356	-/18bps	-/-	292	0	0.0%	Reason for discrepancy unclear

5 out of 7 discordant calls have low WES MAFraction, suggesting WES results may be suspect.

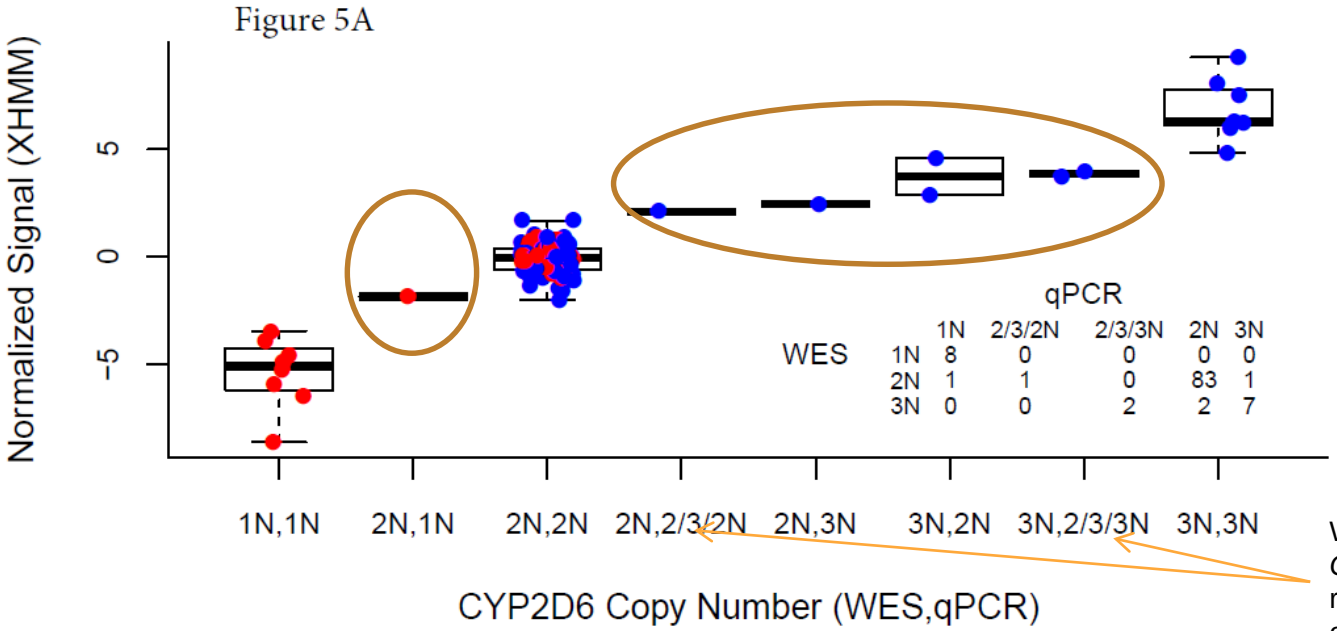
CYP2D6 discordant genotyping calls between DMET and WGS

Gene	Allele	dbSNP	DMET Call	WGS Call	Reference Allele Count	Alternative Allele Count	**Minor Allele Fraction**	Comments
CYP2D6	**20* (1973insG)	rs72549354	T/T	T/TC	42	6	12.5%	WGS low MAFraction
CYP2D6	**40* (1863_1864ins)	rs72549356	-/18bps	-/-	30	0	0.0%	Reason for discrepancy unclear
CYP2D6	**40* (1863_1864ins)	rs72549356	-/18bps	-/-	32	0	0.0%	Reason for discrepancy unclear

CYP2D6 Copy number can be inferred by WES

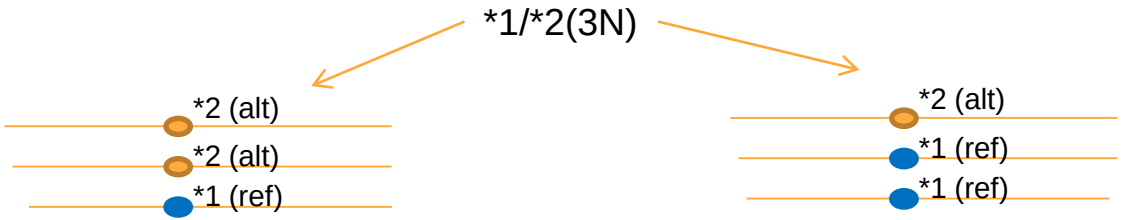
DMET CNV was inferred by qPCR add-on assay
WES CNV was inferred by XHMM

Concordance: 98/105 (93.3%), 3 of 7 discordant calls are possibly *CYP2D6/2D7* hybrid



WES can not infer *CYP2D6/2D7* hybrid, represented by qPCR 2/3/3N, or 2/3/2N, etc.

Haplotype composition for *CYP2D6* (3N) can be inferred by WES



Alt/Ref read depth ratio = 2
Log2(Alt/Ref read depth ratio) = 1.0
Inferred haplotypes: *1/*2/*2

Alt/Ref read depth ratio = 0.5
Log2(Alt/Ref read depth ratio) = -1.0
Inferred haplotypes: *1/*1/*2

patient	chr22:42526694 (P34S, *4)	chr22:42524947 (1846G>A, *4)	chr22:42522613 (S486T, *2, *4)	chr22:42523943 (R296C, *2)	chr22:42524178 (2615delAAG, *9)	WES CNV	qPCR CNV	Haplotype composition	Comment
1	Hom_Ref	Hom_Ref	1.093	0.813	Hom_Ref	3N	3N	*1/*2/*2	
2	Hom_Ref	Hom_Ref	1.052	1.222	Hom_Ref	3N	3N	*1/*2/*2	
3	Hom_Ref	Hom_Ref	-1.141	-0.955	Hom_Ref	3N	3N	*1/*1/*2	
4	-0.781	-1.188	Hom_Alt	0.595	Hom_Ref	3N	3N	*2/*2/*4	
5	-0.933	-1.322	-1.000	Ref	Hom_Ref	3N	3N	*1/*1/*4	

Constellation for CYP2D6 (Twist GP, et al, Genomic Medicine 2016; Gaedigk GA, ASHG 2015)

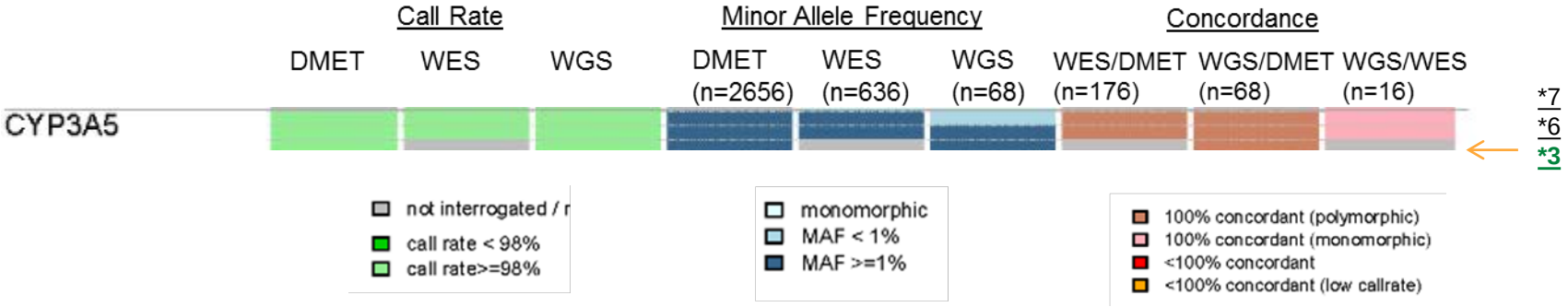
5. CYP3A5

DMET: Good Coverage
WES: missing important intronic variant CYP3A5*3
WGS: Good coverage

No discordant calls were observed

Drug: tacrolimus

Important Variants (n=3):
- 2 SNV (1 exonic)
- 1 exonic Indel



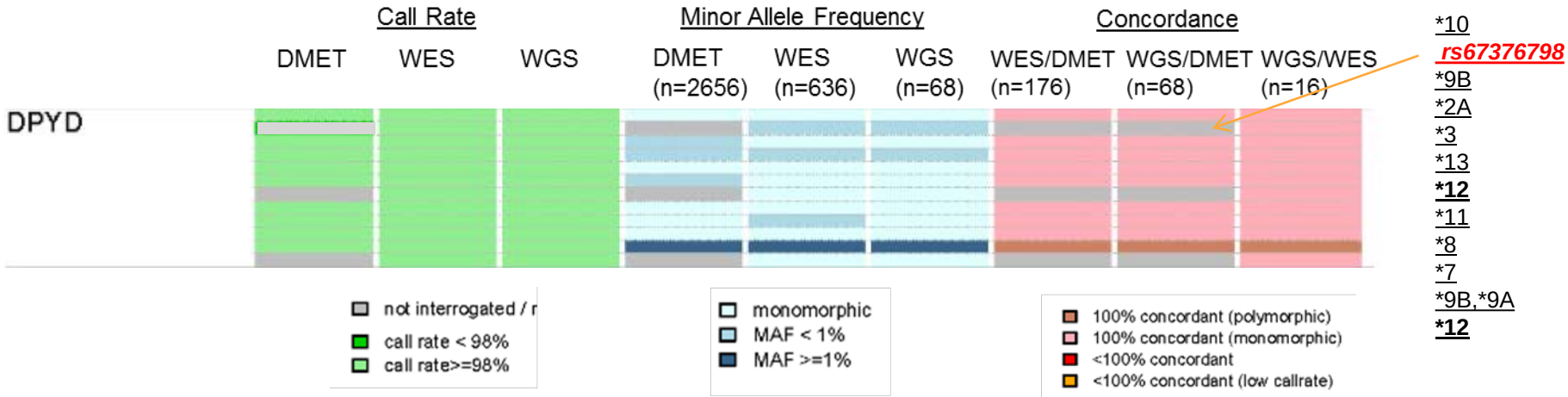
6. *DPYD*

Drug: capecitabine, fluorouracil, tegafur

DMET: not interrogating rs67376798 (Important) and two rare variants *12
WES: good coverage
WGS: good coverage

No discordant genotypes were observed.

Important Variants (n=12):
- 10 exonic SNV
- 2 exonic Indel



7. *G6PD*

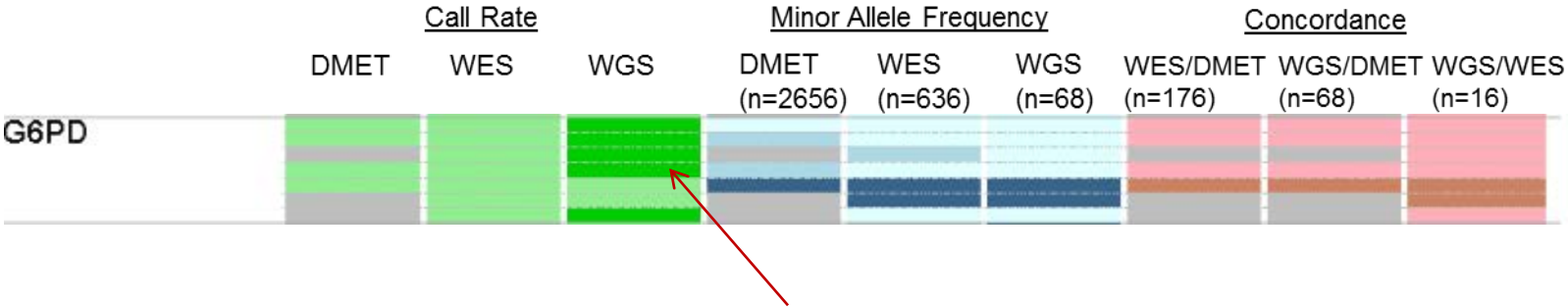
DMET: missing important variants, e.g. common variants Asahi; and other rare variants
WES: Good coverage
WGS: lower call rate in many positions due to lower coverage

No discordant calls were observed

Drug: rasburicase

Important Variants (n=7):

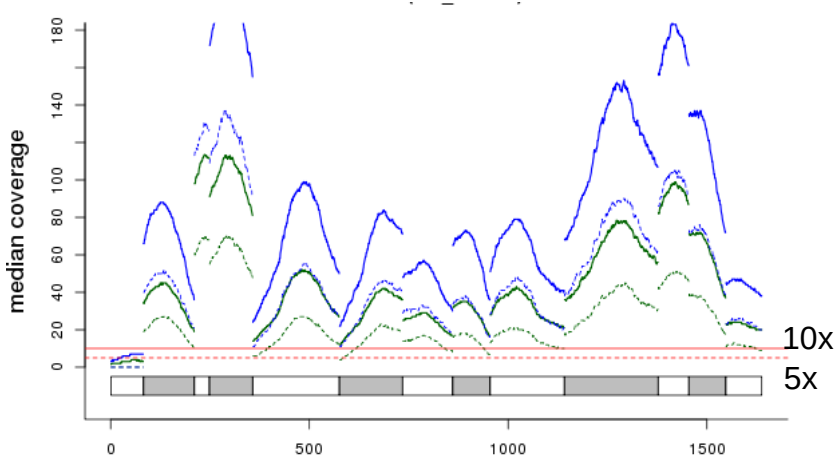
- 7 exonic variants (PharmGKB 2015)



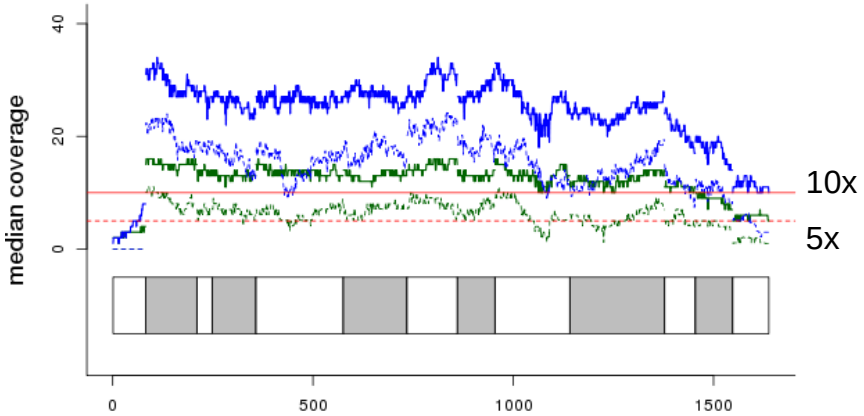
Canton
Chatham
Kalyan
Mediterranean
A-
Asahi
Orissa

Coverage of *G6PD* by Gender

WES (n=636)



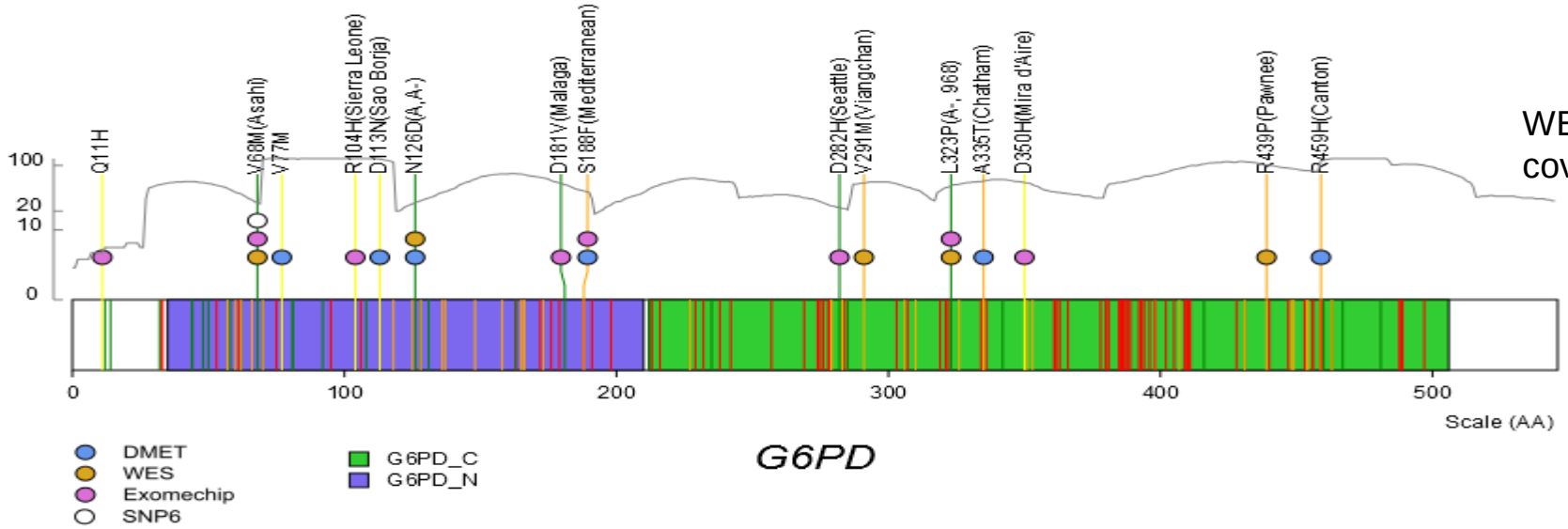
WGS (n=68)



Blue: Females; Green: Males

Solid line: median coverage
Dashed line: 5% patients have coverage below the dashed line

G6PD SNPs in Public Database and interrogated in SNPCHIPs



WES coverage

WHO class I
WHO class II
WHO class III
WHO class IV or unknown

Over 100 important rare variants (WHO class I/II).
DMET Plus v1 only interrogates six variants.

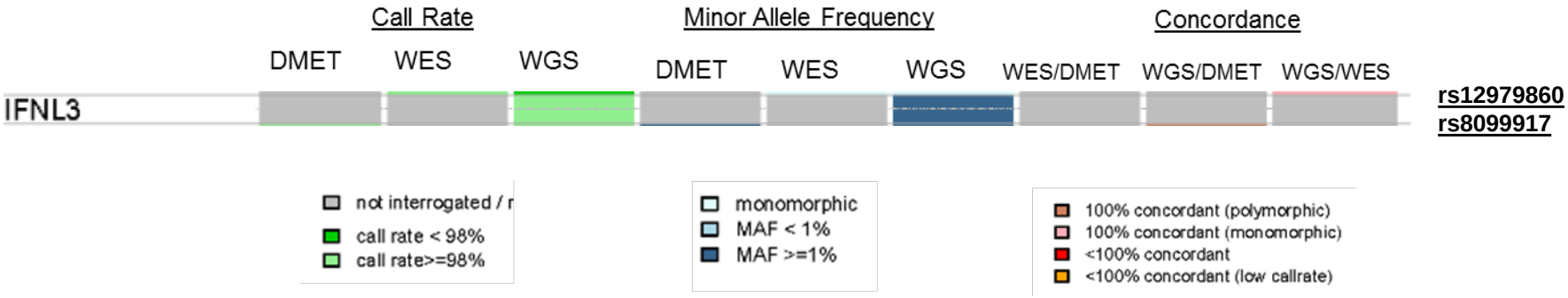
8. IFNL3

DMET: not on DMET array;
WES: upstream variants not targeted
WGS: Good coverage

Drug: peginterferon alfa-2, ribavirin

Important Variants (n=2):

- 2 variants upstream of the gene



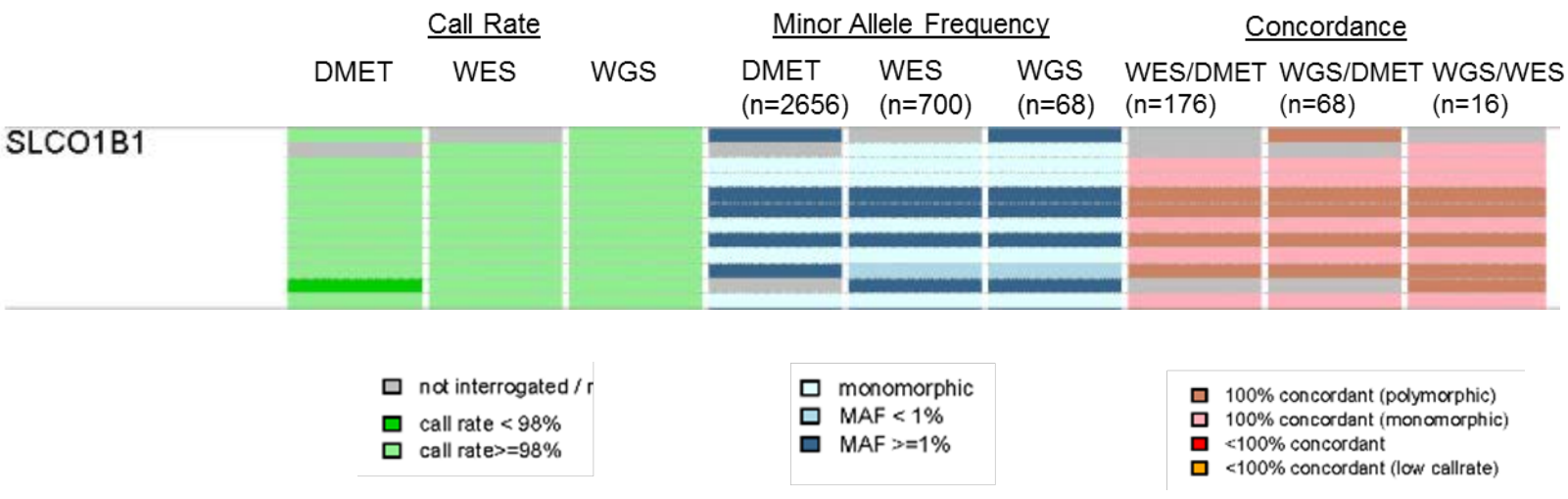
9. *SLCO1B1*

DMET: missing rare variant *23; low call rate at *35
WES: missing promoter SNP *SLCO1B1*17
WGS: good coverage

Drug: simvastatin

Important Variants (n=12):
- 12 SNVs (11 exonic)

No discordant genotypes were observed.



*17
*23
*2
*3
*5, *14, *15, *17
*14
*3
*5,*15,*17
*6
*31
*35
*10

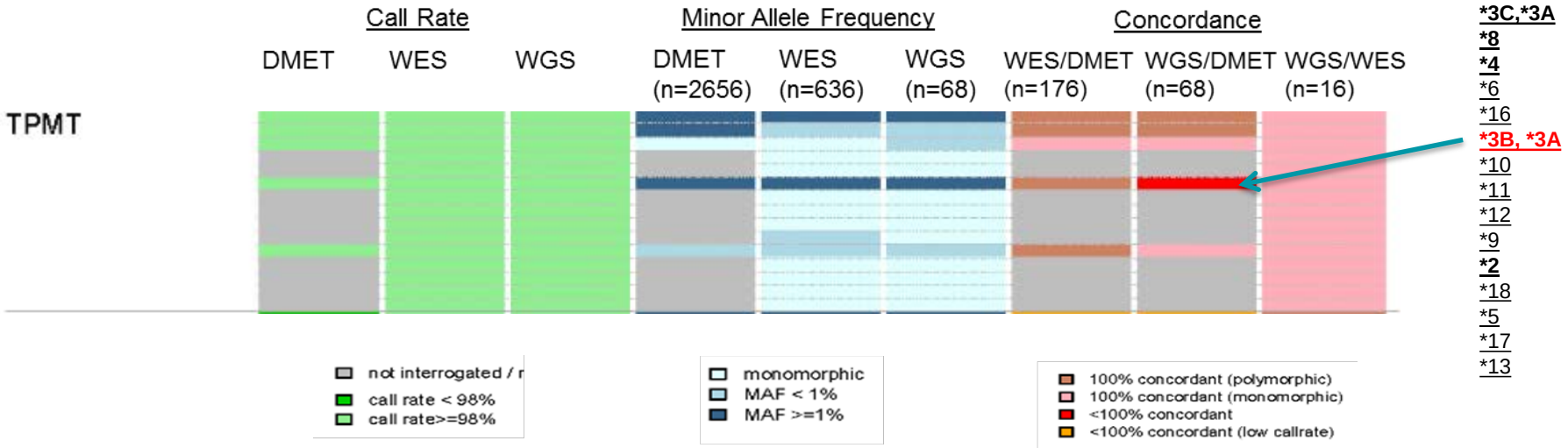
10. *TPMT*

DMET: interrogates most common variants; rare variants not interrogated
WES: good coverage
WGS: good coverage

One discordant genotype observed between WGS and DMET

Drug: azathioprine, mercaptopurine, thioguanine

Important Variants (n=15)
- 15 exonic SNVs



Only one *TPMT* discordant genotyping call between DMET and WGS

Allele	dbSNP	DMET Call	WGS Call	Read Count (Reference Allele, C)	Read Count (Alternative Allele, T)	Minor Allele Fraction
**3B,*3A* (A154T)	rs1800460	C/C	C/T	24	25	49.0%

WGS genotype has good quality: high coverage (24+25) and good minor allele fraction (49.0%).

Orthogonal PCR-RFLP method confirmed WGS genotype for this patient.

Affymetrix DMET Plus v1 result for rs1800460 can be erroneous, add-on reflex tested has been included as part of the clinical testing.

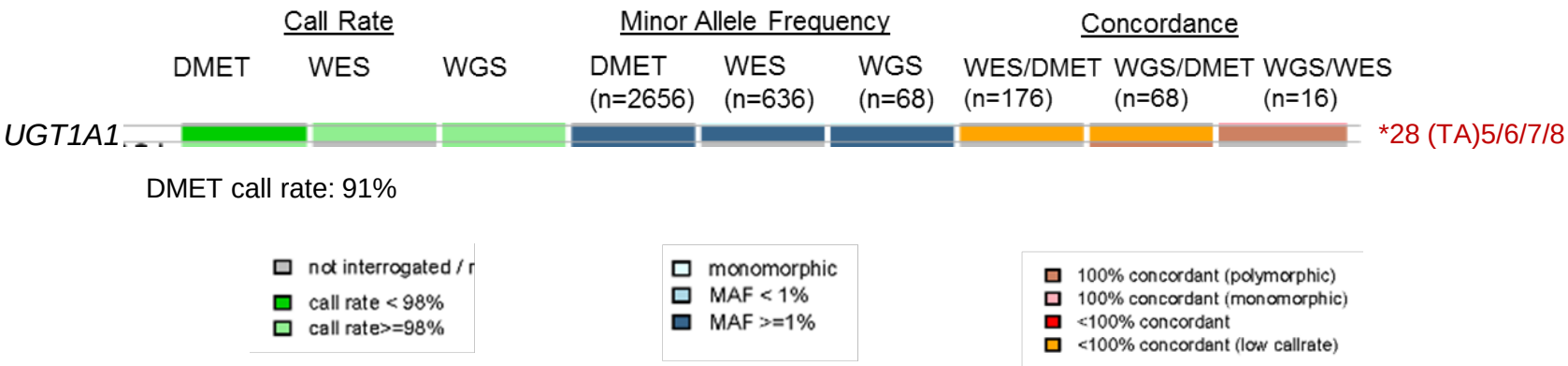
11. *UGT1A1*

Drug: atazanavir

DMET: low call rate for UGT1A1*28
WES: good coverage
WGS: good coverage

No discordant genotypes between WES and WGS

Important Variants (n=1):
- 1 repeat (promoter)



UGT1A1 concordance between locus-specific PCR and WES/WGS

	WES (n=240)		
PCR	(TA)5or6/(TA)5or6	(TA)5or6/(TA)7or8	(TA)7or8/(TA)7or8
(TA)5or6/(TA)5or6	103	0	0
(TA)5or6/(TA)7or8	0	102	0
(TA)7or8/(TA)7or8	0	**3***	32

- Discordant WES genotypes (*) have minor allele fractions lower than 15%, suggesting that WES calls are suspect in these cases.
- Possibility to improve WES genotyping accuracy by introduce additional minor allele fractions cutoff.
- Only 6 patients have both PCR and WGS, all genotypes concordant.
- WES and WGS have all concordant genotypes.

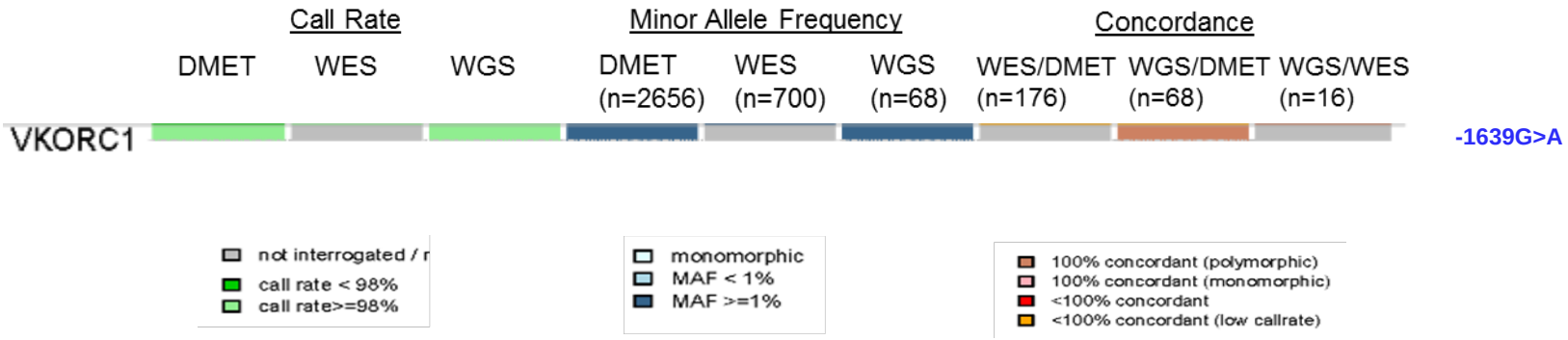
12. VKORC1

Drug: warfarin

DMET: good coverage
WES: missing the important variant
WGS: good coverage

No discordant genotypes were observed.

Important Variants (n=1):
- 1 promoter SNV



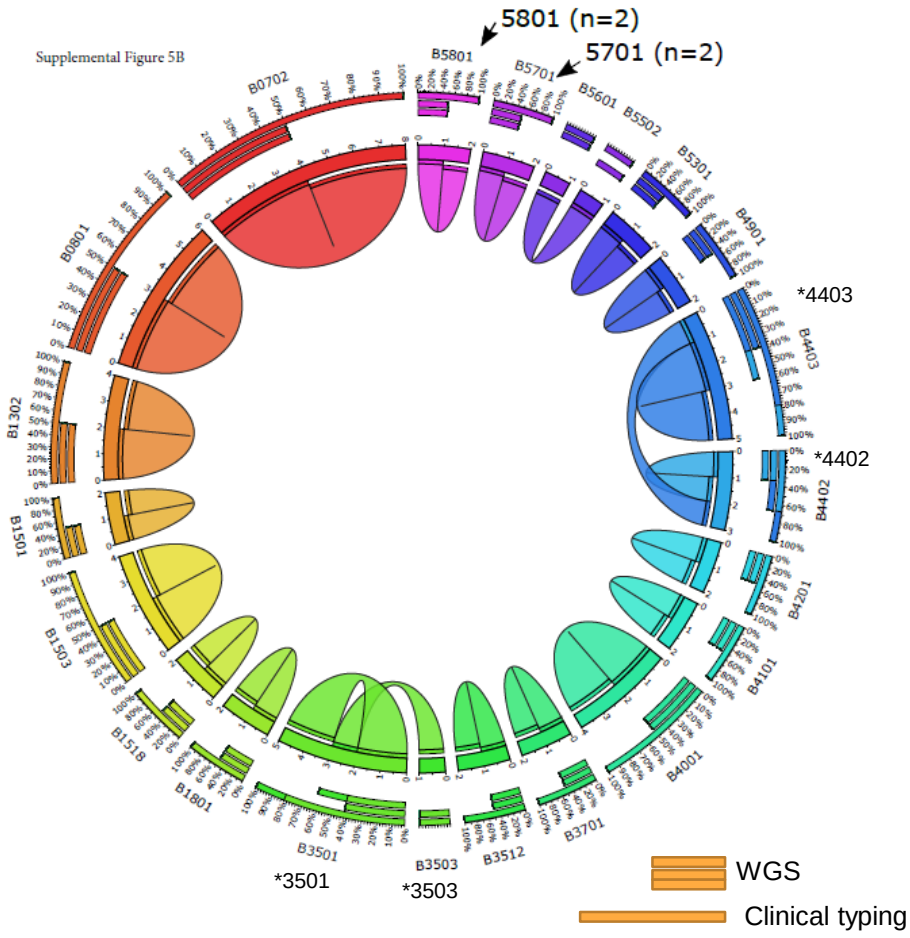
13a. *HLA-B* Haplotyping

Drug: abacavir, allopurinol, carbamazepine, phenytoin

Not interrogated on DMET plus V1

WGS (Optitype)

- Comparing with Clinical HLA typing (n=16)
- 4-digit (29 out of 32 haplotypes)
- 2-digit (31 out of 32 haplotypes)
- *HLA-B*5701* and *HLA-B*5801* were inferred correctly

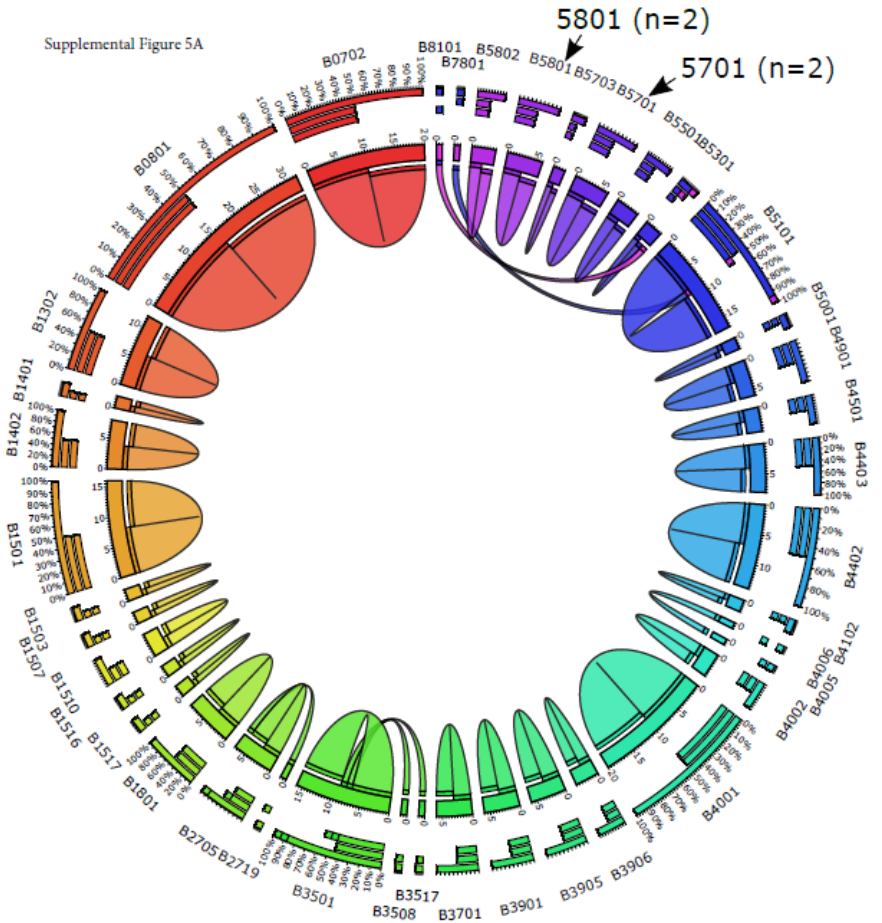


13b. HLA-B haplotyping

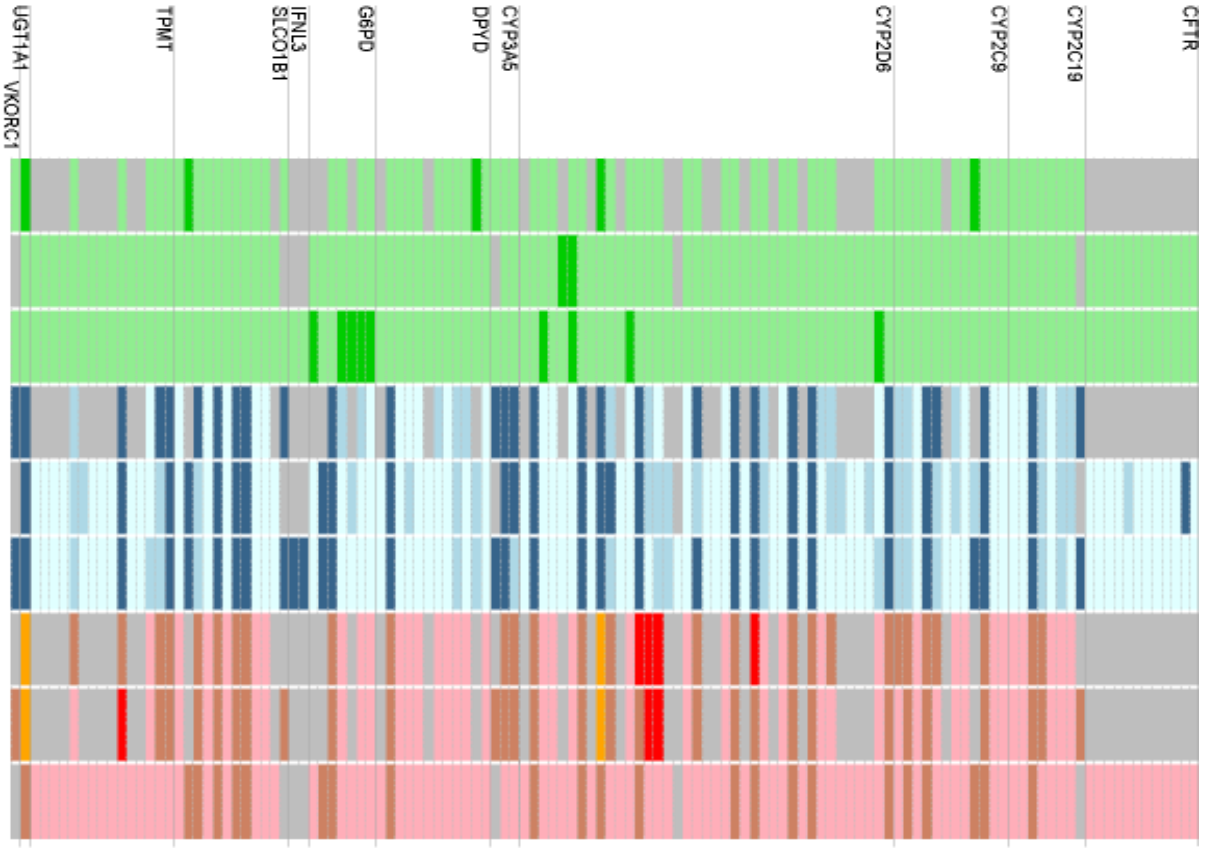
Not interrogated on DMET plus V1

WES (Polysolver)

- Comparing with Clinical HLA typing (n=66)
- 4 digits (126 out of 132 haplotypes)
- 2 digits (130 out of 132 haplotypes)
- *HLA-B*5701* and *HLA-B*5801* were inferred correctly



Overall Comparison of Variants Across Platforms



Not including *CYP2D6* structural variations and *HLA-B* haplotyping

Summary of Performance by Gene

Gene	Affymetrix DMET and add-on assays	Whole exome sequencing	Whole Genome sequencing
CFTR	Not interrogated	Good	Good
CYP2C19	Good	missing *17	Good
CYP2C9	Good	Good	Good
CYP2D6	Good	missing 2D6/2D7 hybrid	missing 2D6/2D7 hybrid
CYP3A5	Good	Missing important variants	Good
DPYD	9 (Missing important variants)	Good	Good
G6PD	Missing important variants	Good	Good; lower callrate due to CNV
HLA-B	Not interrogated	Good	Good
IFNL3	Not interrogated	Missing important variants	Good
SLCO1B1	Good, missing *23,*35	Good, missing *17	Good
TPMT	Good with add-on for *3B	Good	Good
UGT1A1	Low Call rate	Good	Good
VKORC1	Good	Missing important variants	Good

Additional coding variants discovered by NGS

Nonsense:

- WES (n=636): 9 nonsense variants, (2 CFTR, 1 CYP2D6, 3 CYP3A5, 1 DPYD and 2 SLCO1B1)
- WGS (n=68): 2 nonsense variants (1 CYP2C9 and 1 CYP3A5)

Missense variants:

- WES: 153 missense variants
- WGS: 66 missense variants

Most the variants were reported in public exome database.(ExAC http://exac.broadinstitute.org)

Function consequences are not clear.

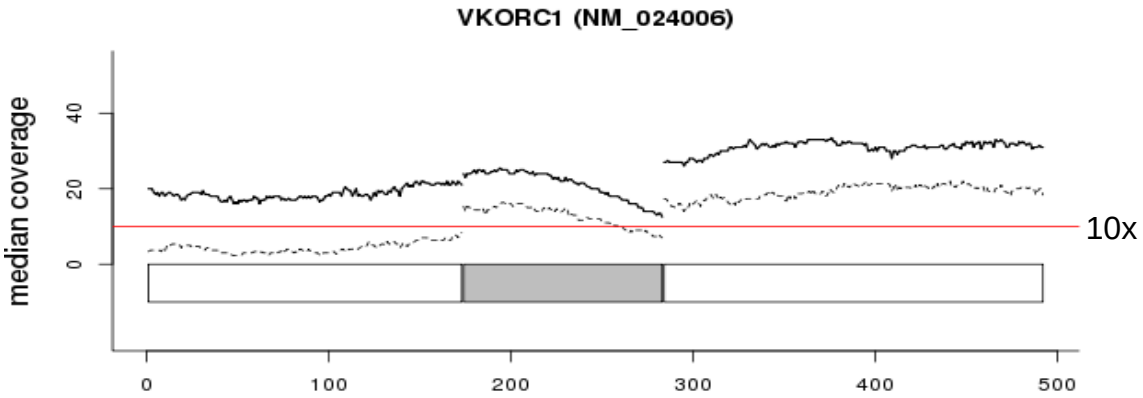
Coverage of coding region of CPIC genes by NGS

	WES		**WGS**	
Gene	Average Read Depth (n=636)	% of exonic region well covered *	Average Read Depth (n=68)	% of exonic region well covered *
CFTR	54	96.7%	37.4	100%
CYP2C19	56	99.0%	36.2	100%
CYP2C9	57	98.7%	36.6	100%
CYP2D6	123.5	98.3%	24.9	75.2%
CYP3A5	55	98.0%	37.0	100%
DPYD	59	99.3%	35.4	98.7%
G6PD	56	90.3%	15.7	41.5%
HLA-B	78	92.0%	19.7	44.5%
IFNL3	136	100%	25.5	84.4%
SLCO1B1	42	93.1%	37.5	98.2%
TPMT	60	100%	40.2	100%
UGT1A1	67	86.0%	33.9	99.4%
VKORC1	61	77.6%	23.2	64.8%

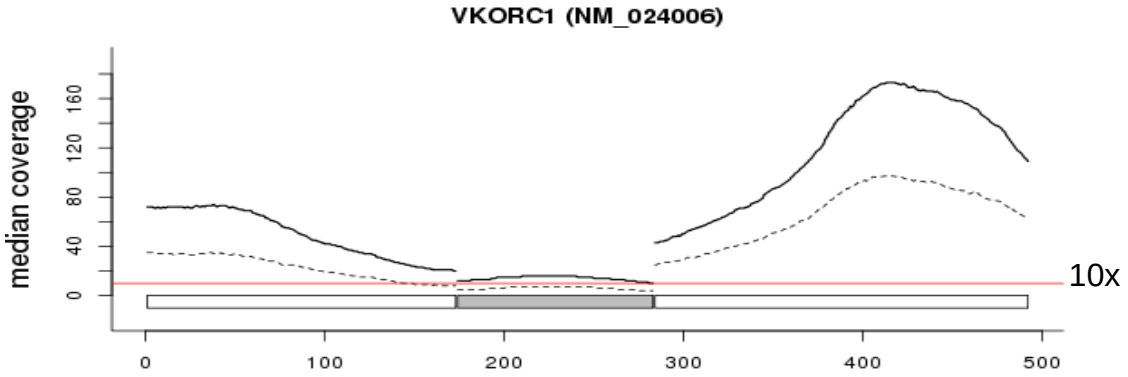
* A genomic position is well covered if the 95% of patients have read depth higher than 10x at the position

Coverage of VKORC1

WGS



WES



Summary

- Both WES and WGS provide high quality genotyping calls using standard pipeline (e.g. GATK).
- WES is missing important variants in several genes due to lack of interrogation, including *VKORC1*, *IFNL3*, *CYP3A5*3*, *CYP2C19*17*
- WGS has lower call rates in genes involved in CNV, including *CYP2D6* and *G6PD*. For *G6PD*, gender specific QC/calls can help to improve call rate.
- Additional adjustment on standard pipeline QC (e.g. MAFraction threshold) can further improve the accuracy of WES and WGS.

Limitations

- WES and WGS were not performed in a clinical lab setting.
- NGS were not performed on standard samples with known genotypes.
- Sensitivity and specificity is difficult to estimate due to relatively small number of patients. Especially for rare variants, it is difficult to establish the accuracy.

Future

- Targeted sequencing using NGS technology would be more cost effective in the implemetation of pharmacogenomics. E.g. PGRNseq (*Rasmussen-Torvik LJ, et. al CPT 2014*)
- Tailored algorithms can provide better results, e.g. Constellation for *CYP2D6 (Twist GP, et al, Genomic Medicine 2016; Gaedigk GA, ASHG 2015)*
- Needs to establish informatic pipeline to interpret NGS into action alleles. PharmCAT effort by PharmGKB to provide tools to interpret standard NGS output files (VCF) to starred alleles, e.g TPMT*3A, which can be used in downstream clinical decision making. (*https://github.com/PharmGKB/PharmCAT*)
- New version Affymetrix DMET array will be introduced soon, and could address some of the limitation of DMET array v1.

Acknowledgement

St Jude Children’s Research Hospital

Mary V. Relling, PharmD

William E. Evans, PharmD

Jun Yang, PhD

Charles Mullighan, MD

Gang Wu, PhD

Colton Smith, PhD

Medical College of Wisconsin

Ulrich Broeckel, MD

Pediatric Cancer Genome Project