Cross platform comparison of pharmacogenomics reporting from whole-genome sequencing

S. Garikipati[1], J. Arres[1], P. Kandavel[1], D. Sanchez[1], C. Minas[1], H. Wu[1], F. Aldhuhoori[1], J. Mafofo[1], , S. Elavalli[1], S. Chandrashekar[1], A. El-Khani[1], M. Shaikh[1], T. R. Magalhães[1], J. Quilez[1]

[1] M42

Introduction

- Pharmacogenomics (PGx) studies how genetic variations in ~40 genes affect drug response
- PGx profiling is usually based on microarrays targeting such genes and (less so) whole-exome sequencing (WES) with short read sequencing (SRS)
- Relative to such current PGx testing standards, WGS and long read sequencing (LRS) may present some advantages
- A key question is: **Can WGS and / or LRS produce high-accuracy PGx predictions?**

Results

- Star allele: ~85% precision in all three sequencing platforms, and >98% across-platforms concordance
- Drug response: ~95% precision in any platform and complete concordance across them (**Table 1**)
- While some variation between platforms exist, ONT showed performance at par with SRS
- *DPYD*, *UGT1A1*, and *SLCO1B1* are known to be challenging and performed worse in our analysis
- All three though had high concordance (~80–100%) across platforms so we will investigate this further

Table 1. Precision to detect (a) star alleles and (b) drug response profiles. For MGI, we used two variant calling pipelines (Sentieon and Zbolt) and compared the performance for both. Performance improved when excluding challenging *UGT1A1* and *SLCO1B1* genes.

a

		Per-sample precision			Per-gene precision		
Platform	**Gene set**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**
Illumina	All 14	66.7	85.2	100.0	28.6	86.0	100.0
	12 genes	75.0	87.1	100.0	28.6	87.9	100.0
MGI (Sentieon)	All 14	72.7	87.4	100.0	28.6	86.7	100.0
	12 genes	80.0	89.2	100.0	28.6	88.7	100.0
MGI (Zbolt)	All 14	72.7	87.4	100.0	28.6	86.7	100.0
	12 genes	80.0	89.2	100.0	28.6	88.7	100.0
ONT (R9)	All 14	72.7	86.2	100.0	28.6	86.0	100.0
	12 genes	80.0	88.0	100.0	28.6	87.9	100.0
ONT (R10)	All 14	63.6	84.9	100.0	28.6	85.0	100.0
	12 genes	70.0	86.6	100.0	28.6	86.7	100.0

b

		Per-sample precision			Per-gene precision		
Platform	**Gene set**	**Min**	**Mean**	**Max**	**Min**	**Mean**	**Max**
Illumina	All 14	77.8	94.5	100.0	60.0	95.1	100.0
	12 genes	87.5	95.5	100.0	60.0	95.5	100.0
MGI (Sentieon)	All 14	81.8	96.3	100.0	75.0	96.2	100.0
	12 genes	90.0	97.3	100.0	75.0	96.7	100.0
MGI (Zbolt)	All 14	81.8	96.3	100.0	75.0	96.2	100.0
	12 genes	90.0	97.3	100.0	75.0	96.7	100.0
ONT (R9)	All 14	81.8	95.2	100.0	60.0	95.1	100.0
	12 genes	90.0	96.0	100.0	60.0	95.5	100.0
ONT (R10)	All 14	81.8	95.2	100.0	60.0	95.1	100.0
	12 genes	90.0	96.0	100.0	60.0	95.5	100.0

Implemented analysis pipeline to **predict personalized response** to **126 drugs**.

Whole-genome sequencing (WGS) is highly **accurate** for detecting genetic differences that affect drug response.

Both **short**- and **long**-read sequencing had **similar performance** and high concordance.

DPYD, *UGT1A1*, and *SLCO1B1* were more difficult to analyze accurately across all platforms.

snleela@m42.ae

@m42_health

Methods

- Implemented an analysis pipeline to predict star alleles and drug response from WGS (**Figure 1**)
- Selected 7 Coriell reference samples with truth sets for 14/23 PGx genes in the pipeline
- Performed ~30X WGS for all 7 samples in SRS (Illumina and MGI) and LRS Oxford Nanopore Technologies (ONT) – both R9 and R10 chemistries
- Evaluated performance by comparing against truth sets and between sequencing technologies

Figure 1. M42 PGx analysis pipeline. We use WGS datasets (derived from any of our three sequencing technologies) to determine star alleles in 23 genes relevant for PGx, and such star alleles are mapped to drug response predictions.

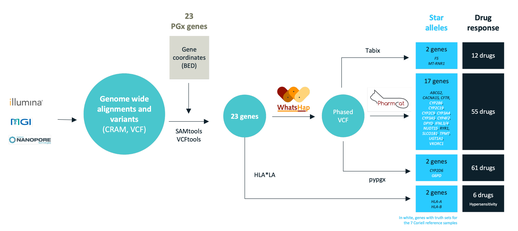

Table 2. Detailed performance in predicting drug response

		CYP2B6	CYP2C9	CYP2C19	CYP3A4	CYP3A5	CYP4F2	DPYD	SLCO1B1	TPMT	UGT1A1	VKORC1	NUDT15	G6PD	IFNL3	
Illumina	GM12878	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM07000	1	1	1	1	1	-	1	1	1	-	1	1	1	1	100.0
	GM12753	0	1	[2]	1	1	-	1	0	[2]	-	1	-	1	1	77.8
	GM18564	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM18855	1	1	1	1	0	1	1	1	1	1	1	-	1	1	92.3
	GM18992	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0
	GM19176	1	1	[2]	1	1	0	1	1	1	1	1	-	1	1	91.7
	Gene precision	85.7	100.0	100.0	100.0	100.0	60.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	**95.1**
MGI (Sentieon)	GM12878	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM07000	1	1	1	1	1	-	1	1	1	-	1	1	1	1	100.0
	GM12753	0	1	[2]	1	1	-	1	0	[2]	-	1	-	1	1	81.8
	GM18564	1	1	[2]	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM18855	1	1	1	1	0	1	1	1	1	1	1	-	1	1	92.3
	GM18992	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0
	GM19176	1	1	[2]	1	1	0	1	1	1	1	1	-	1	1	91.7
	Precision gene	85.7	100.0	100.0	100.0	100.0	75.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	**96.2**
MGI (Zbolt)	GM12878	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM07000	1	1	1	1	1	-	1	1	1	-	1	1	1	1	100.0
	GM12753	0	1	[2]	1	1	-	1	0	[2]	-	1	-	1	1	81.8
	GM18564	1	1	[2]	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM18855	1	1	1	1	0	1	1	1	1	1	1	-	1	1	92.3
	GM18992	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0
	GM19176	1	1	[2]	1	1	0	1	1	1	1	1	-	1	1	91.7
	Precision gene	85.7	100.0	100.0	100.0	100.0	75.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	**96.2**
ONT (R9)	GM12878	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM07000	1	1	1	1	1	-	1	1	1	-	1	1	1	1	100.0
	GM12753	0	1	[2]	1	1	-	1	0	[2]	-	1	-	1	1	81.8
	GM18564	1	1	[2]	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM18855	1	1	1	1	0	1	1	1	1	1	1	-	1	1	92.3
	GM18992	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0
	GM19176	1	1	[2]	1	1	0	1	1	1	1	1	-	1	1	92.3
	Precision gene	85.7	100.0	100.0	100.0	100.0	60.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	**95.1**
ONT (R10)	GM12878	1	1	1	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM07000	1	1	1	1	1	-	1	1	1	-	1	1	1	1	100.0
	GM12753	0	1	[2]	1	1	-	1	0	[2]	-	1	-	1	1	81.8
	GM18564	1	1	[2]	1	1	1	1	1	1	1	1	-	1	1	100.0
	GM18855	1	1	1	1	0	1	1	1	1	1	1	-	1	1	92.3
	GM18992	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100.0
	GM19176	1	1	[2]	1	1	0	1	1	1	1	1	-	1	1	92.3
	Precision gene	85.7	100.0	100.0	100.0	100.0	60.0	100.0	85.7	100.0	100.0	100.0	100.0	100.0	100.0	**95.1**
																Overall precision

1 Match
0 No Match
- No GT
? Deprecated allele
. No results available

Acknowledgements

This study used samples (GM12878, GM07000, GM12753, GM18564, GM18855, GM18992 and GM19176) from the Coriell Institute for Medical Research.

CORIELL INSTITUTE