Poster Presentation Number: P15.069.B

EVALUATION OF QUALITY OF WGS AND WES RESULTS USING VALIDATED PROTOCOLS

Katerina Kvapilova[1,3], Pavol Misenko[2], Jan Radvanszky[2], Ondrej Brzon[1], Veronika Buzkova [4], Jaroslav Budis[2], Juraj Gazdarica[2], Ondrej Pos[2], Martin Kasny[1], Tomas Szemes[2], Petr Kvapil[1]

[1] Institute of Applied Biotechnologies, Sluzeb 3056/4, 108 00 Prague, Czech Republic
Correspondence: kvapilova@iabio.eu; Tel.: +420 605 200 674
(All affiliations in detail are presented in the poster footer)

Whole exome sequencing (WES) and whole genome sequencing (WGS) are standards in human clinical diagnostics and population genomics (POPGEN). We evaluated the quality and genotype concordance of single nucleotide variants (SNVs) and small insertions/deletions (small-indels) in paired blood and saliva genomic DNA (gDNA) isolates for WES and WGS protocols using reference-validated protocols.

Material and methods

NIST standard Coriel NA12878 (RS NA12878) was repeatedly sequenced to validate WGS and WES protocols by comparing RS NA12878 data calls to the truth dataset published by the Genome in a Bottle Consortium (Ref. 1)*.

In the presented study, the total number of variants (DRAGEN) as well as genotype concordance (Emedgene) for SNVs and small-indels in paired blood-saliva samples was analysed, in accordance with the study design (Figure 1). The study is focused onto ACMG SF v 3.2. gene set in two settings: exonic regions only (WES; WGS-E) and complete genes (WGS-C).

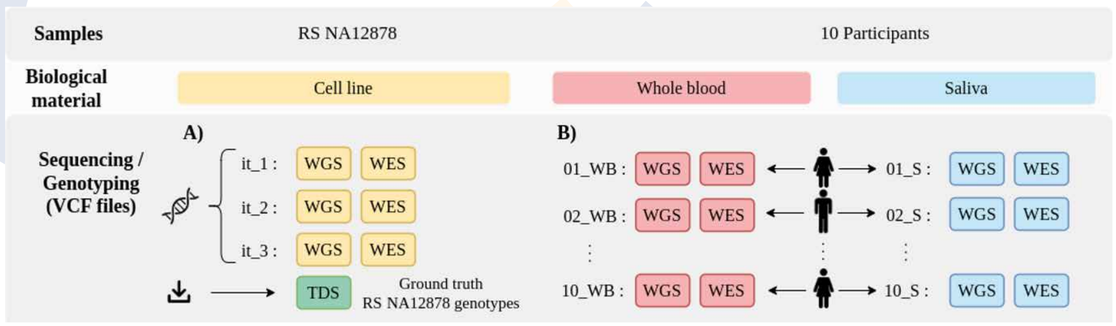

Figure 1. Graphical overview of the study design. A) Reference samples used in the study; B) Data analyses of paired blood-saliva samples performed during the study.

Results

First, variant concordance comparison of paired blood-saliva samples was determined by F1 scores (Ref. 1.)*. For the WES protocol, the median value of F1 score for ten paired blood-saliva samples was 0.9858 for SNVs and 0.9076 for small-indels. For the WGS protocol, the median value of F1 was 0.9761 for SNVs and 0.9511 for small-indels. The F1 score distribution of paired blood-saliva samples copied the distribution of F1 score RS NA12878 iterations for both, the whole genome and whole exome analysis, respectively.

The average values of SNVs variant calls concordance in ACMG SF gene set were 98,73 % for WGS-C, 99,81 % for WGS-E only and 99,54 % for WES. Concordance of indels calls reach on average 92,88 % for WGS-C samples, 97,77 % for WGS-E and 70,51 % for WES (Table 1).

For the tertiary analysis of paired blood-saliva samples, the Emedgene software package was employed to depict pathogenic or likely pathogenic variants using VCF files. The variants depicted by Emedgene is specified in Table 2.

	WGS-C		WES		WGS-E	
paired samples	blood	saliva	blod	saliva	blood	saliva
	5 062 690	5 087 100	33 358	33 353	33 056	33 202
RS NA12878	5 114 383		33 878		33 486	

Table 3. Number (median) of found variants by DRAGEN in the studied blood and saliva samples over the whole genome, the whole exome and the exonic part of the whole genome.

Conclusion

The same relevant variants were found in Emedgene software at paired blood-saliva samples for both the WES and WGS approaches in ACMG SF v3.2 list of genes. The resulting data shows a clear advantage of the WGS approach over WES; especialy for small-indels, while confirming no significant differences between blood and saliva samples for SNV´s and for small-indels analysis.

Conflict of interests: The authors declare no conflict of interests.

Ref. 1* Scan the QR code for Validated WGS and WES protocols proved saliva-derived gDNA as an equivalent to blood-derived gDNA for clinical and population genomic analyses
https://doi.org/10.1186/s12864-024-10080-0

Affiliations:
1 Institute of Applied Biotechnologies a.s., Prague, Czech Republic, 2 Geneton s.r.o., Bratislava, Slovakia, 3 Faculty of Science, Charles University, Czech Republic, 4 GeneTiCA s.r.o., Prague, Czech Republic

Isolation of gDNA and RNA

- QIAamp® DNA Blood Mini Kit (QIAGEN, Germany) with a sample origin--dependent protocol within 24 hours after sample collection.

NGS library preparation

- **WGS** - TruSeq DNA PCR-Free kit (Illumina, USA); gDNA input: 1 µg.
- **WES** - Illumina DNA Prep with Enrichment kit (Illumina, USA) with the Alliance VCGS Exome panel and Mitochondrial DNA panels (both Twist Bioscience, USA), gDNA input: 100 ng.

Sequencing

- NovaSeq 6000 (Illumina, USA), S4 chemistry with XP 4-Lane kit.

Bioinformatic analysis

- Reference mapping (GRCh38 reference genome) and variant calling by DRAGEN v3.10 (Illumina).
- Tertiary data analysis by Emedgene (Illumina) to study ACMG set of genes in paired blood-saliva samples for WES and WGS data was performed employing the internal Emedgene settings (Figure 1B).

WGS-C SNVs													
samples	1	2	3	4	5	6	7	8	9	10	RS NA12878		
blood unique variants	50	39	54	29	56	61	45	47	41	63	30	27	30
saliva unique variants	62	34	60	73	81	53	49	44	90	32	61	37	51
variant intersect	8408	8831	8125	8582	8069	8474	8073	7850	8140	8293	8125	8128	8135

WES SNVs													
samples	1	2	3	4	5	6	7	8	9	10	RS NA12878		
blood unique variants	0	1	1	0	1	0	0	0	1	0	0	0	0
saliva unique variants	2	2	0	2	0	0	0	2	2	0	0	0	0
variant intersect	341	368	326	305	285	320	321	255	268	338	287	287	287

WGS-E													
samples	1	2	3	4	5	6	7	8	9	10	RS NA12878		
blood unique variants	0	0	0	0	2	0	0	0	0	1	1	0	0
saliva unique variants	1	0	0	0	0	1	1	0	0	0	0	1	0
variant intersect	343	370	327	308	286	321	321	258	272	338	289	289	289

WGS-C small-indels													
samples	1	2	3	4	5	6	7	8	9	10	RS NA128787		
blood unique variants	71	68	72	64	108	68	71	80	81	72	68	67	63
saliva unique variants	102	85	91	88	106	89	82	83	85	84	76	50	88
variant intersect	2172	2276	2157	2299	2105	2228	2067	2038	2080	2150	2166	2167	2154

WES small-indels													
samples	1	2	3	4	5	6	7	8	9	10	RS NA12878		
blood unique variants	4	4	4	9	6	6	7	7	4	7	7	5	4
saliva unique variants	3	4	4	7	6	4	5	6	5	4	5	9	6
variant intersect	25	27	25	25	22	23	26	25	27	25	19	19	22

WGS-E small-indels													
samples	1	2	3	4	5	6	7	8	9	10	RS NA12878		
blood unique variants	0	0	0	0	2	0	1	0	0	0	1	0	0
saliva unique variants	0	0	1	0	1	0	2	0	0	0	0	1	0
variant intersect	28	27	26	28	27	28	32	33	29	29	23	23	23

Table 1. Variant concordance of paired blood-saliva samples in ACMG SF v3.2 list of genes. Unique variants represent variants found in only one type of the sample type, intersect represents variants found in both types of samples

samples	blood		saliva	
	WGS-C	WES	WGS-C	WES
1	NM_000410.4 c.187C>G	NM_000410.4 c.187C>G	NM_000410.4 c.187C>G	NM_000410.4 c.187C>G
2	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C
3	NM_000545.8 c.79A>C	NM_000545.8 c.79A>C	NM_000545.8 c.79A>C	NM_000545.8 c.79A>C
4	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C
5	NM_000551.4 c.74C>T NM_024675.4 c.2993G>A NM_000152.5 c.2065G>A	NM_000551.4 c.74C>T NM_024675.4 c.2993G>A NM_000152.5 c.2065G>A	NM_000551.4 c.74C>T NM_024675.4 c.2993G>A NM_000152.5 c.2065G>A	"NM_000551.4 c.74C>T NM_024675.4 c.2993G>A NM_000152.5 c.2065G>A"
6	/	/	/	/
7	NM_000410.4 c.187C>G NM_000410.4 c.845G>A	NM_000410.4 c.187C>G NM_000410.4 c.845G>A	NM_000410.4 c.187C>G NM_000410.4 c.845G>A	NM_000410.4 c.187C>G NM_000410.4 c.845G>A
8	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_000545.8 c.79A>C
9	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_000410.4 c.187C>G NM_001304717.5 c.10G>A NM_000545.8 c.79A>C
10	NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_001304717.5 c.10G>A NM_000545.8 c.79A>C	NM_001304717.5 c.10G>A NM_000545.8 c.79A>C

Table 2. Emedgene Analysis. Found pathogenic and likely-pathogenic variants in ACMG SF v3.2 list of genes detected by Emedgene software in paired blood and saliva samples.