Evaluation of WGS and WES protocols based on dried bood spot sample-derived gDNA

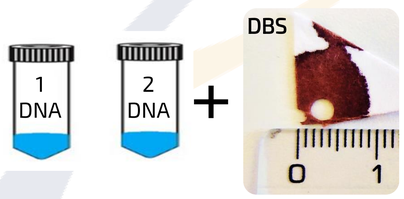
Martin Kašný [1*], Laura Ambrozová [1], Karolína Stivínová [1], Eliška Hladíková [1], Nela Chalupníková [1], Alžběta Hamplová [1], Ondřej Brzoň [1], Petr Kvapil [1]

1 Institute of Applied Biotechnologies, Služeb 3056/4, 108 00 Prague, Czech Republic

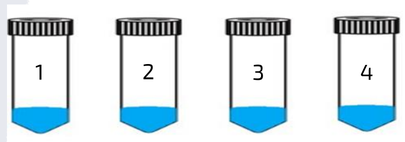
* Correspondence: kasny@iabio.eu; Tel.: +420 739 394 364

Whole exome sequencing (WES) and whole genome sequencing (WGS) have become a standard method in human clinical genetics. Blood-derived gDNA is routinely used in the clinical environment, presenting difficulties associated with invasive sampling. Dried venous blood spots (DBS) represent a convenient gDNA source in respect to low amount of collected material and simple sample transport and storage. However, the use of DBS-derived gDNA for NGS applications has not been analyzed in detail. To address this point, IAB validated both WGS and WES NGS protocols based on DBS-derived isolates.

Material and methods



4 biological samples
- 2x DNA from fresh blood (sample 1 and 2)
- 2x DNA from dried blood spot (DBS)
  - 10 years old DBS from leukemia patient
  - Whatmann 903™ Protein saver card, ca. 1 cm in diameter



Isolation of gDNA from 2x DBS
RNeasy Plus Micro Kit (Qiagen, Germany)

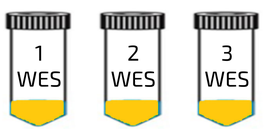
Sample 1 and 2: DNA from fresh blood were delivered as isolates
Sample 3 and 4: DNA from DBS isolated by IAB

Quality control of DNA
- Purity – NanoPhotometer
- Quantity – Qubit 2.0 (Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit)
- Quality – Agarose gel electrophoresis

The concentration of DNA isolated from DBS was ca 3 ng/µl, based on Qubit measurement (Table 1).

Sample ID	Original Sample Type	NanoPhotometer Results		Qubit Results		
		A260/268	c [ng/µl]	c [ng/µl]	Volume [µl]	DNA Amount [ng]
1	DNA from fresh blood	1,7	37,5	48,6	21	1021
2	DNA from fresh blood	1,7	72,5	96,0	22	2112
3	DNA from DBS	1,6	4,0	2,9	135	392
4	DNA from DBS	2,0	4,5	5,0	60	300

Table 1. The selected QC metrics of the DNA isolates



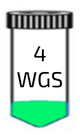
NGS library preparation

1. WES (Samples 1, 2 and 3)
Illumina DNA prep with Enrichment kit (Illumina, USA) with Twist Alliance VCGS panel (Twist Bioscience, USA)

Input: 100 ng

Library plex of 12 different samples (3 above mentioned + 9 other samples) was prepared.

+



2. WGS (Sample 4)
TruSeq DNA PCR-Free High Throughput Library Prep Kit (Illumina, USA)

Input: 300 ng, due to the limited amount of DNA isolate the input was reduced to maximum of available sample instead of original Illumina protocol requirement of 1000 ng

Sequencing
- Illumina NovaSeq 6000, S4 chemistry with XP 4-Lane kit

Bioinformatic analysis and reference mapping
- Sequence quality control – FastQC v. 0.11.9
- Aligning to the hg38 reference genome, variant calling – Dragen 4.0
- Vcf files comparison – vcf-compare

Library Preparation Results

- Preparation of the plex containing WES libraries was successful (Table 2):
  - Concentration: 19,6 ng/µl; 79,4 nM
  - Average fragment length: 370 bp (Table 2, Figure 1)
  - No adapter dimers were recorded

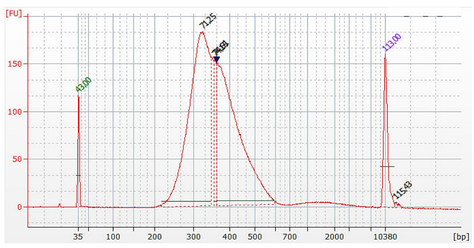


Figure 1. The electropherogram of the plex containing NGS libraries for WES.

Sample ID	Type of library preparation	Qubit Results	Bioanalyzer 2100 Results		
		c [ng/µl]	c [ng/µl]	Average Fragment Length [bp]	Dimers [%]
1, 2, 3	WES	19,6	17,9	370	0
4	WGS	2,3	1,8	979	0

Table 2. The selected QC metrics of WES and WGS NGS libraries

- Preparation of the WGS library was successful (Table 2):
  - Concentration: 2,3 ng/µl; 35,2 nM
  - Average fragment length: 979 bp (Table 2, Figure 2)
  - No adapter dimers were recorded

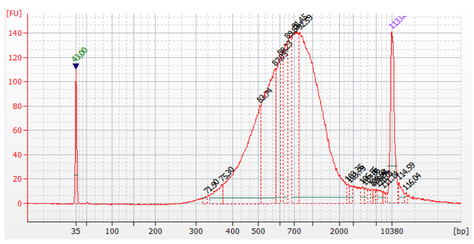


Figure 2. The electropherogram of the plex containing NGS libraries for WGS.

Sequencing Results

- All samples were successfully sequenced. Average coverage for WES samples was more than 128 in each sample, 97 % of target region was covered at least 20x. For WGS sample, average coverage of 31 was achieved, 92 % of genome was covered more than 20x. Full secondary analysis QC and results are shown in Supplementary Tables 1 and 2.

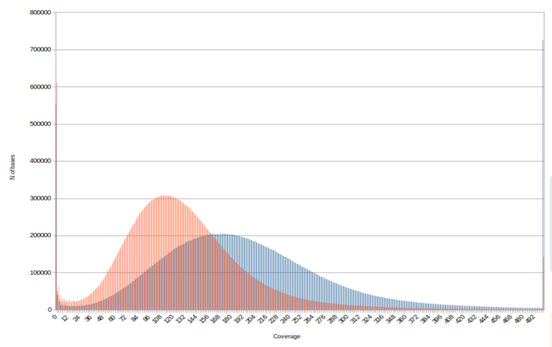


Figure 3. Uniformity of coverage in WES for NA12878 (blue) and our samples (red)

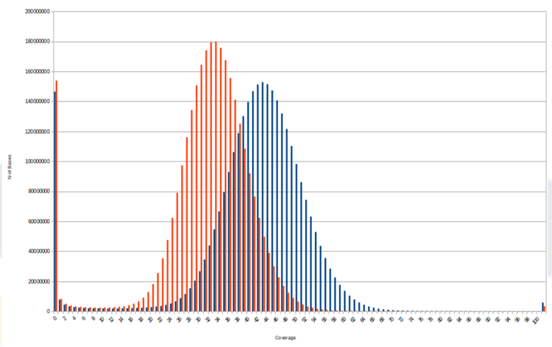


Figure 4. Uniformity of coverage in WGS for NA12878 (blue) and our samples (red)

Number of Variants for Standard and Samples	Coriell NA12878	1,2,3,4
WGS only	1 066	1 122
WES only	1 275	773
WES + WGS	32 615	31 909

Table 3. Concordance of variant calls between WES and WGS data in WES target regions. The analysis was done using the vcf-compare tool. Row WES+WGS contains total number of identical variants found in both WGS and WES. Rows WGS and WES only contain numbers of variants identified only in WGS/WES protocol.

Conclusion

The DNA isolation from DBS followed by 1 WES/1 WGS library preparation and sequencing was successful. All sequencing data had high quality regardless of the DNA source. There is no significant difference in data quality metrics (on-target, coverage uniformity) between DNA isolates from fresh blood and blood spot sample. Data obtained from Coriell NA12878 reference standard showed slightly better performance than fresh blood and DBS samples in uniformity of coverage and concordance of variant calls between WES and WGS. The above results showed that a dried venous blood spot is a sample form which can be effectively used for generating high quality WES and WGS data, simplifying specimen transport and long-term storage.

Evaluation of WGS and WES protocols based on dried bood spot sample-derived gDNA

Martin Kašný [1*], Laura Ambrozová [1], Karolína Stivínová [1], Eliška Hladíková [1], Nela Chalupníková [1], Alžběta Hamplová [1], Ondřej Brzoň [1], Petr Kvapil [1]

1 Institute of Applied Biotechnologies, Služeb 3056/4, 108 00 Prague, Czech Republic
* Correspondence: kasny@iabio.eu; Tel.: +420 739 394 364

WES	Coriell NA12878	1	2	3
Total input reads	112 769 322	78 106 480	75 504 900	76 720 272
Duplicates [%]	11,18	9,95	9,10	9,05
Mapped reads [%]	99,92	99,77	99,80	99,80
Insert length: mean	206	201	200	194
Average coverage	201,43	133,74	128,37	133,44
Uniformity of coverage (% > 0.2*mean) [%]	97,21	96,71	96,77	96,67
Uniformity of coverage (% > 0.4*mean) [%]	92,71	93,09	93,01	92,73
Aligned bases in target region [%]	60,60	57,36	56,73	58,13
Aligned reads in target region [%]	87,71	85,13	84,16	85,89
PCT of target with coverage > 100x [%]	87,97	69,10	66,49	68,31
PCT of target with coverage > 50x [%]	96,69	93,98	93,50	93,69
PCT of target with coverage > 20x [%]	97,89	97,14	97,17	97,10
PCT of target with coverage > 10x [%]	98,15	97,71	97,78	97,68
PCT of target with coverage > 1x [%]	98,62	98,49	98,60	98,47
Average mitochondrial coverage	8 003	1 107	1 078	1 052

Supplementary Table 1. The comparison of selected WES sequencing and mapping metrics with Coriell NA12878 standard. Analysis was performed using DRAGEN 4.0 (Illumina) on hg38 human reference genome.

WGS	Coriell NA12878	4
Total input reads	1 069 743 600	880 672 988
Duplicates [%]	16,49	22,73
Mapped reads [%]	99,73	99,30
Insert length: mean	389	407
Average coverage	42,24	31,53
Uniformity of coverage (% > 0.2*mean)	96,13	94,04
PCT of target with coverage > 50x [%]	19,73	1,05
PCT of target with coverage > 20x [%]	94,94	92,08
PCT of target with coverage > 10x [%]	96,05	93,79
PCT of target with coverage > 1x	97,22	94,90
Average mitochondrial coverage	19 287	916

Supplementary Table 2. The comparison of selected WGS sequencing QC and mapping metrics with Coriell NA12878 WGS reference standard. Analysis was performed using DRAGEN 4.0 (Illumina) on hg38 human reference genome.