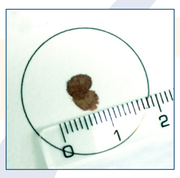
From dried blood spot to whole genome long-read sequencing

Martin Kašný [1*], Alžběta Hamplová [1], Karolína Stivínová [1], Eliška Hladíková [1], Nela Chalupníková [1], Markéta Kovalová [1], Irem Mertová [1], Alžbeta Mičůchová [1], Ondřej Janča [1], Natália Floreková [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

Dried venous blood spots (DBS) represent a convenient source of gDNA with respect to the small amount of material collected, ease of sample transport and storage. The limitations are mainly related to the difficulties in mapping the short reads to the specific regions of the reference genome and subsequent *de novo* assembly formation. Long-read sequencing is gaining ground in clinical applications because it overcomes this problem. Illumina Complete Long Reads technology makes long-read sequencing accessible to genomic laboratories by enabling comprehensive human whole genome analysis with both long and short reads data generated in the same sequencing experiment or in two independent experiments at different Illumina sequencing platforms.

Material and methods



1 biological sample
4x dried blood spot (DBS)
QIAcard™ FTA™ Classic, approx. 1 cm in diameter

Isolation of gDNA from 4 DBS
QIAamp DNA Investigator Kit (QIAGEN, Germany)

Quality control of DNA
- Quantity – Qubit 2.0 (Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit)
- Quality – Agilent 4200 TapeStation

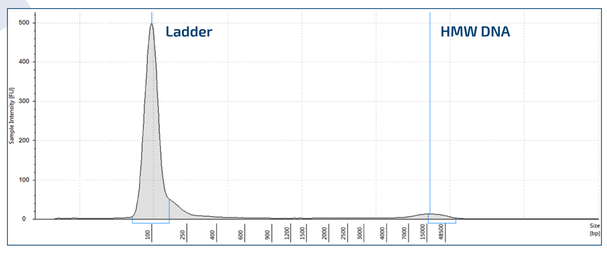


Figure 1. The electrophoretogram of the DNA sample

The concentration of DNA was 1.1 ng/µl, based on Qubit measurement and the sample showed very good integrity (Figure 1).

Long

NGS library preparation
1. Long Read
Illumina Complete Long Read Prep Kit, Human (Illumina, USA) with Illumina Unique Dual Indexes (Illumina, USA)
The input: 32 ng (protocol recommendation 50 ng)

\+

In terms of the protocol, short reads library was also prepared from DNA sample.

Short

2. Short Read
Illumina® DNA PCR-Free Prep, Tagmentation (Illumina, USA) with IDT® for Illumina® DNA/RNA UD Indexes, Tagmentation (Illumina, USA)
The input: 27 ng (protocol recommendation 50 ng)

Due to the limiting amount of DNA, the input was reduced to maximum sample available.

Sequencing
- Illumina NovaSeq X Plus, 10B Reagent Kit

Bioinformatic analysis
- Sequence quality control, secondary analysis: DRAGEN ICLR WGS (Illumina Complete Long Reads v1.1.2 Easy-to–use BaseSpace application for complete secondary analysis, hg38 human reference genome

Results

- Preparation of the NGS library for long reads was successful (Table 1):
  - Concentration: 8.5 ng/µl; 13.9 nM
  - Average fragment length: 928 bp (Table 1, Figure 2)
  - No adapter dimers were detected

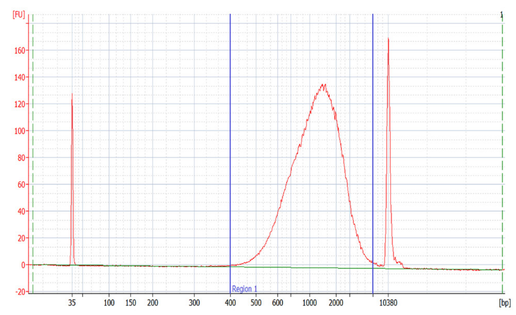


Figure 2. Example of electrophoretogram of NGS library for long reads.

- Preparation of the NGS library for short reads was successful (Table 1):
  - Concentration: 0.2 ng/µl; 0.6 nM

Table 1. Selected sequencing and analysis metrics

Short reads	
Number of reads	947 934 718
Median coverage	49
Insert size average	431.5
Percentage duplicated reads	28.38
ICLR	
N50	4 928
Total input reads (raw)	4 663 749 034
Number of reads (assembled)	18 845 587
Insert size average (raw)	510.3
Percentage duplicated reads (raw)	22.12
Median coverage	25
Percentage of reads mapped > 10kb	2.36
Percentage of bases in reads mapped > 10kb	7.59
Combination short reads + ICLR	
Number mapped reads	961 258 163
Median coverage	74
VCF metrics	
SNPs	4 092 632
Insertions	473 786
Deletions	479 524
Indels	25 642
Total Het/Hom ratio	1.47
Insertion/Deletion ratio	0.99
Total SVs	29 285
Phasing statistics	
Number biallelic, het variants phased	2 858 021
Percentage biallelic, het variants phased	97.70
Number of phase blocks	46 088
Phase block N50	138 673

Table 2. The selected quantitative and qualitative metrics related to the DNA isolate and NGS library.
* Since the NGS libraries for short reads are single stranded DNA (ssDNA), only Qubit™ ssDNA Assay Kit was used for quality control of the short read NGS library.

Sample ID	Volume [µl]	Qubit Results		Library Preparation			Qubit Results		Bioanalyzer 2100 Results	
		c [ng/µl]	DNA Amount [ng]	Type	Input [µl]	Input [ng]	c [ng/µl]	c [nM]	c [ng/µl]	Avg. Fragment Lenght [bp]
1	60	1.1	63	Long Read	30.0	32	8.5	13.9	3.9	928
				Short Read	28.4	27	0.2	0.6	–*	–*

Conclusion

The gDNA from dried venous blood spots was successfully isolated in sufficient quantity and quality. Two types of whole genome NGS libraries were prepared from one sample using two approaches - Illumina® DNA PCR-Free Prep, Tagmentation and Illumina Complete Long Read Prep Kit - and sequenced on NovaSeq X plus. Data met appropriate QC parameters and were analyzed using the Dragen 4.2 software tool along with the comprehensive Illumina Complete Long Reads v1.1.2 – a push-button long read analysis application. For the long reads, the average N50 value was 4.9 kb, with 7.6% of the data coming from reads longer than 10 kb. Combining short and long reads, the median coverage was 74% and the phased N50 value exceeded 138 kb.
Our work demonstrates that dried blood spots are not only a good source of gDNA for whole genome sequencing, but can also be used to prepare long read libraries, making this approach available not only for the use in clinical genetics, but also potentitally for whole genome *de novo* assembly.