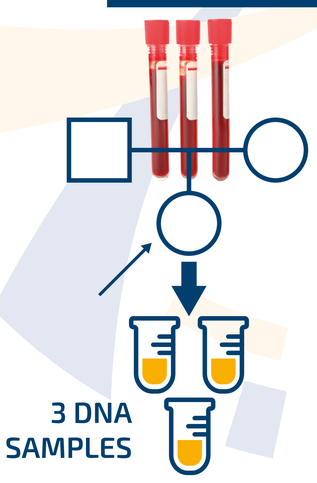
Rapid Whole Genome Sequencing and Analysis Using AI Emedgene

RNDr. Martin Kašný, Ph.D., Ing. Mgr. Přemysl Bartoš, Ph.D., Mgr. Alžbeta Mičúchová,Bc. Veronika Bůžková, Mgr. Laura Ambrozová, Ph.D., Mgr. Karolína Stivínová, Mgr. Irem Mertová, Mgr. Alžběta Hamplová, Mgr. Eliška Hladíková, Mgr. Nela Tomko, Mgr. Markéta Kovalová, Ph.D., prof. MUDr. Milan Macek DrSc., RNDr. Petr Kvapil

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

Abstract: Whole genome sequencing (WGS) has become a standard method in clinical diagnostic laboratories, where it represents a reliable analytical tool due to its excellent coverage quality, enabling deep insights into the genome. The PCR-free protocol offers a superior solution for traditionally difficult-to-sequence DNA regions, such as GC-rich regions, promoters, and repetitive content, while also reducing library bias and gaps. In addition, WGS can be used to make accurate diagnoses of very rare disorders that would otherwise require harmful and invasive diagnostic procedures. This protocol enables the rapid preparation of high-quality, sequencing-ready indexed whole-genome sequencing libraries, sequencing, and bioinformatic analysis within 40 hours at a low cost.

Material and methods



3 BLOOD SAMPLES

2 samples = proband's parents
1 sample = proband (daughter) diagnosed with Mayer-Rokitansky-Küster-Hauser syndrome with rare congenital disorder that affects the female reproductive system (HPO: Hypoplasia of the vagina, Aplasia of the uterus)
Isolation of gDNA from 200 µl blood - EZ1&2 DNA Blood 350 µl Kit (Qiagen, USA)

Quality Control of DNA

- Sample purity (NanoPhotometer) and Sample quantity (electrophoresis , Qubit 2.0)
- All samples showed good purity and integrity (Figure 1).
- The concentration of DNA was from 62 to 86 ng/µl, based on Qubit measurement (Table 1).

Sample ID	NanoPhotometer Results		Qubit Results		
	A260/280	c [ng/µl]	c [ng/µl]	Volume [µl]	DNA Amount [ng]
Proband	1.9	125	85,8	38	3260
Mother	2,0	110	61,6	43	2649
Father	1,9	168	83,3	59	4915

Table 1. The selected QC metrics of the DNA isolates.

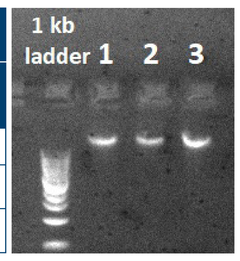


Figure 1. Gel electrophoresis

THE WHOLE ANALYSIS WITHIN 40 HOURS

library preparation

- TruSeq DNA PCR-Free High Throughput Library Prep Kit (Illumina, USA)
- Input **1050 ng** illumina

Library Preparation Results

- All 3 WGS libraries were successefully prepared.
- The concentration of the NGS libraries was ca 4 ng/µl (8 nM), based on Qubit measurement (Table 2).

- The average fragment length of the NGS libraries ranged from 755 to 917 bp, based on Bioanalyzer 2100 calculation (Table 2, Figure 3).

Proband **Mother** **Father**

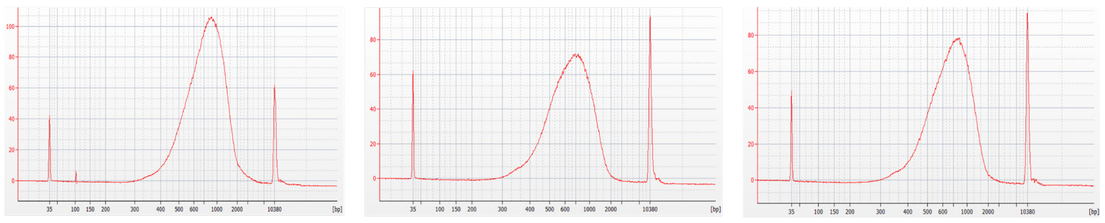


Figure 3: The electropherograms of the NGS libraries.

Sample ID	Library preparation		Qubit Results	qPCR Results	Bioanalyzer 2100 Results		
	Input [µl]	Input [ng]	c [ng/µl]	c [ng/µl]	c [ng/µl]	Average Fragment Size [bp]	C [nM]
Proband	12,2	1050	3,3	1,6	3,2	917	5,5
Mother	17,0	1050	4,3	1,8	3,4	755	8,7
Father	12,6	1050	5,2	2,1	3,4	801	9,9

Table 2: The selected QC metrics of WGS NGS libraries.

Sequencning

- **NovaSeq 6000 (Illumina, USA)**
  - All samples were successfully sequenced on NovaSeq 6000 S1 200 chemistry.
  - The yield of the data related to sequencing run was 464,92 Gbp; QC30% 94,66%; CPF 85,00%; PhiX 0,53%. The required yield 800 million PE reads per sample was achieved for all 3 NGS libraries.

Data Analysis with Emedgene SW

- **Automated analysis from FASTQ files:**
- Mapping and variant calling with DRAGEN 4.0 using hg38 reference genome
- Quality control, gender validation and relatedness control of samples
- Emedgene AI variant prioritization and correlation with HPO terms - Hypoplasia of the vagina, Aplasia of the uterus.

Results of Data Analysis

emedgene an Illumina company

- Compete data analysis in Emedgene (from FASTQs of the family to annotated AI prioritized variants) took 17 hours.
- All three samples met passed targeted QC metrics (table 3).

Sample	Average Cov.	% Bases With Cov. >20X	Gender Validation
Mother	36,7	92,7	PASS
Father	35,7	90,3	PASS
Proband	35,8	92,6	PASS

Table 3: Results of Data Analysis with Emedgene SW.

- Applying Emedgene AI SW 7 Candidate potentially pathogenic variants related to proband were defined (table 5).

Conclusion

All 3 WGS libraries were successfully prepared, sequenced, and prioritized annotated variants were provided in an efficient time of less than 40 hours. The innovative rWGS workflow, based on Illumina technologies, takes personalized diagnostics to the next level and paves the way for broader use in clinical practice. The presented results demonstrate that rWGS and Emedgene analysis offer a solution for rapid and accurate diagnostics, especially due to the swift detection of causal variants in genetic disorders, even in challenging cases where other diagnostic tools fall short.

Rapid Whole Genome Sequencing and Analysis Using AI Emedgene

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

	Sanger	Targeted NGS*	PCR*	CMA*	WES*	WGS*
SNVs	√√	√√	√√		√√	√√
Indels	√√	√√	√√	√	√√	√√
CNVs		√	√√	√√	√	√√
Expansion of repetitions			√		√	√√
Structural variants				√		√√
Mitochondria	√√	√√			√√	√√
Paralogs	√√		√√			√√

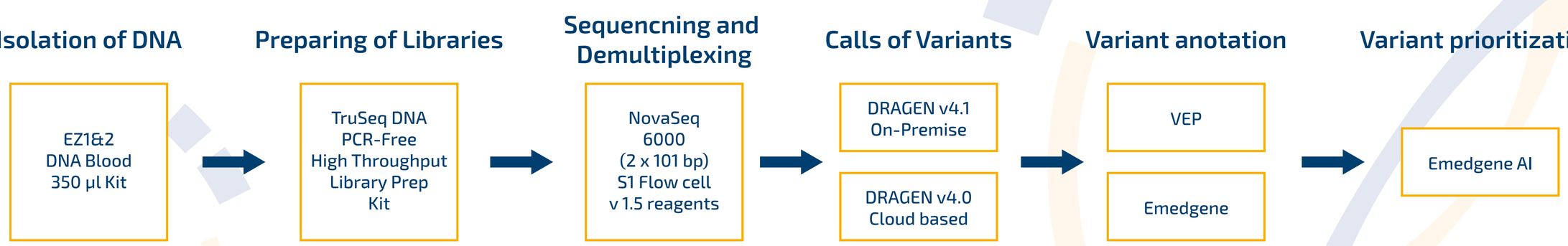
Supplementary Table 4. Comparison of genetic testing approaches based on variant detection ability.

**Variant detection may vary by laboratory and test offering. √√ Detection; √ Limited detection.*

Gen	Position	Variant	Classification	Proband	Mother	Father
NXN	chr17	819516 2.85(Kb) DEL	Aplasia of the uterus, Hypoplasia of the vagina	√	-	-
RNASEL	chr1	182586014C>A,rs74315364	Aplasia of the uterus, Hypoplasia of the vagina	√	-	√
CNTN2	chr1	205058319C>T	Aplasia of the uterus, Hypoplasia of the vagina	√	-	-
NAIP	chr5	71010576 3.392(Kb) DEL	-	√	-	-
RP1L1	chr8	10608139G>A,rs200846354	Aplasia of the uterus, Hypoplasia of the vagina	√	√	-
COL7A1	Chr3	48583584G>A,rs79378857	Hypoplasia of the vagina	√	-	√
PDGFRA	chr4	54265030G>A	Hypoplasia of the vagina	√	√	-

Supplementary Table 5. Causal variants of genetic disorders. Annotation cloud system DRAGEN v4.0 and AI software tool Emedgene

**Detection of variants. √ Detection; - No detection.*



Supplementary Figure 1. Workflow of rapid whole-genome sequencing (rWGS) for variant identification. The workflow from sample preparation to obtaining results using the local Dynamic Read Analysis for GENomics (DRAGEN) server version 4.1, followed by variant annotation using the publicly available Variant Effect Predictor (VEP) tool in 38.2 hours, or using the DRAGEN cloud server version 4.0 integrating artificial intelligence Emedgene to identify and prioritize potentially relevant variants in 39 hours and 54 minutes.