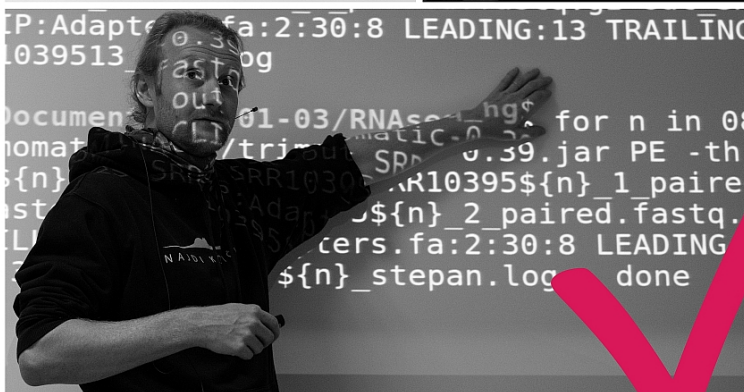
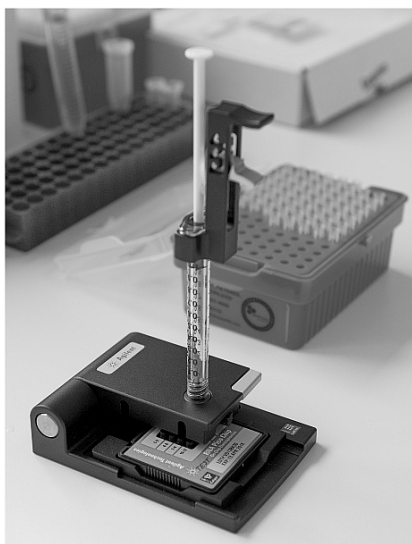


Zpráva o výsledcích Next-Generation sekvenování

Klient:	Vlasta Korenková
Číslo zakázky:	242114199
Datum tisku:	20.5.2025 11:52



Údaje o zakázce č. 242114199

Zpracování vzorků:	mRNA library (LP-060)
Počet vzorků:	19
Sekvenační jednotka:	PE150, 600 million reads / 90 Gb (DP600-PE150)
Počet sekvenačních jednotek:	1
Základní analýza:	Data handling services, price per order (BS-SB) + USB drive
Formát:	Zkumavky
Doručení:	Osobně
Vaše číslo objednávky:	24NA474

Moje soubory

ID	Název	Popis	Přípona	Velikost
242114199-1023	SEQme_NGS_Sample submission sheet v20241104_Korenkova_09122024.xlsx	Sample submission form. RNA byla izolována z jednobuněčných organismů, z trofozoitů, z kultury. Poznámka: RNA koncentrace byla měřena na nanodropu.	xlsx	217.8 kB
242114199-1032	SEQme_NGS_Sample submission sheet v20241104_Korenkova_02012025_UPRAVA.xlsx	Pro doplnění koncentrace vzorku 15, odebrání vzorku 16 a přidání vzorku 19.	xlsx	220.4 kB
242114199-1040	Popis experimentu pro analýzu SEQme 13012025.docx	Popis k analýze. V případě nejasností prosím volejte. Děkuji. VK	docx	18.8 kB
242114199-1081	ÚPRAVY názvů vzorků 09042025.docx	Přikládám ještě soubor o postupech ve změnách ve vzorcích, počty a přejmenování. Zatím je vzorek 15 stále špatně pojmenován (dne 9.4.2025). Potřebovali bychom si to opravit kvůli analýze. Díky VK	docx	86.7 kB

Vzorky

QC vzorek						QC knihovna
ID vzorku	Koncentrace (ng/ul)	Objem (ul)	Koncentrace (SEQme) (ng/ul)	RIN (pouze pro RNA)	Stav	Stav
S4199_001_1_BER_CK_1	840	20	1 250	10	Passed	Passed
S4199_002_2_BER_CM_1	133	20	150	10	Passed	Passed
S4199_003_3_40_CK_2	391	20	472	10	Passed	Passed
S4199_004_4_40_CM_2	68	25	62	9,5	Passed	Passed
S4199_005_5_41_CK_3	701	20	693	10	Passed	Passed
S4199_006_6_41_CM_3	112	20	127	9,7	Passed	Passed
S4199_007_7_24_CK_4	300	20	269	10	Passed	Passed
S4199_008_8_24_CM_4	48	25	55	10	Hraniční	Passed
S4199_009_9_BER_AK_5	686	20	744	10	Passed	Passed
S4199_010_10_BER_AM5	128	20	141	10	Passed	Passed
S4199_011_11_BER_CK5	1 157	20	821	10	Passed	Passed
S4199_012_12_BER_CM5	249	20	301	10	Passed	Passed
S4199_013_13_BER_CK6	1 171	20	556	10	Passed	Passed
S4199_014_14_BER_CM6	149	20	116	10	Passed	Passed
S4199_015_15_2_CK_7	885	20	35	10	Failed	Passed
S4199_016_16_2_CM_7	53	25	630	10	Passed	
S4199_017_17_8_CK_7	1 130	20	750	10	Passed	Passed
S4199_018_18_8_CM_7	144	20	130	10	Passed	Passed
S4199_019_19_2_CK_7U	530	20	530		Passed	Passed

QC samples

- Qubit RNA BR Assay Kit
- Agilent Bioanalyzer RNA 6000 Nano Kit

Komentář:

Vzorky mají velmi dobrou integritu. Bohužel u 2 vzorků (8 a 15) je celkové množství dodané RNA menší než jsou naše doporučení (1,5 ug total RNA). U vzorků 8 je množství hraniční, to nepředstavuje problém. U vzorku 15 je koncentrace ovšem pouze 35 ng/ul

(oproti udávané 885 ng/ul) a tudíž celkové množství dodané RNA je pouze 700 ng. I z tohoto množství jsme schopni knihovnu udělat, jen budeme při tvorbě knihovny vycházet z cca o 1/3 menšího množství RNA oproti ostatním vzorkům. Pokud máte vzorek ještě k dispozici, bylo by lépe nám jej dodat.

Dejte nám prosím vědět, zda budete chtít pokračovat v analýze za stávajících podmínek a nebo zda budete schopna dodat ještě alikvot alespoň 400 ng vzorku č. 15.

Update 18. 12. 2024

Email od zákaznice: Vzorek č. 15 - bude dodáno více materiálu, požadavek na přejmenování na 15_2_CM_7U. Vzorek č. 16 - vyřadit z analýzy a nahradit vzorkem č. 19 (19_2_CK_7U).

Library preparation:

- NEBNext Poly(A) mRNA Magnetic Isolation Module
- NEBNext Ultra II Directional RNA Library Prep Kit for Illumina
- NEBNext Multiplex Oligos for Illumina (Unique Dual Index Primer Pairs)

Library QC:

- Agilent Bioanalyzer 2100 High sensitivity DNA Kit
- Invitrogen Colibri Library Quantification Kit
- Qubit 1X dsDNA High-Sensitivity Assay Kit

Sekvenování

ID vzorku	Obsah GC (%)	Průměrná délka readů (b)	Celkový počet readů	Počet single-end readů
1_BER_CK_1	52	150	80 106 334	40 053 167
2_BER_CM_1	53	150	66 939 554	33 469 777
3_40_CK_2	53	150	66 819 890	33 409 945
4_40_CM_2	53	150	66 549 070	33 274 535
5_41_CK_3	53	150	60 332 352	30 166 176
6_41_CM_3	53	150	64 219 924	32 109 962
7_24_CK_4	53	150	63 401 282	31 700 641
8_24_CM_4	53	150	57 396 616	28 698 308
9_BER_AK_5	52	150	66 680 502	33 340 251
10_BER_AM5	54	150	64 196 516	32 098 258
11_BER_CK5	52	150	59 729 948	29 864 974
12_BER_CM5	52	150	74 261 554	37 130 777
13_BER_CK6	52	150	69 322 978	34 661 489
14_BER_CM6	52	150	71 446 868	35 723 434
15_2_CK_7	53	150	69 324 638	34 662 319
16_2_CM_7				
17_8_CK_7	52	150	63 891 542	31 945 771
18_8_CM_7	53	150	77 012 934	38 506 467
19_2_CK_7U	52	150	69 717 056	34 858 528
		Celkem:	1 211 349 558	605 674 779

Sequencing platform:

- NovaSeq X Plus - 2 x 150 bp

Analýza dat

Comment on data download:

The full dataset will be delivered on an external hard drive. Users are advised to copy the data to a secure local storage before working with it to ensure faster access and avoid accidental modification of the original files. It is strongly recommended to create a backup of the HDD contents, ideally on a separate physical drive or cloud storage, to prevent data loss.

Data Analysis

A comprehensive analysis of the data was performed to explore the different patterns of gene expression between several experimental groups (BER_CK, BER_CM, CK and CM). The analytical pipeline included several preprocessing steps and differential gene expression analysis using DESeq2.

Preprocessing

Quality control: to ensure data integrity and identify potential problems, the raw sequencing data was subjected to quality control using FastQC (v0.11.9) and MultiQC tools (v1.24).

Trimming: Trimmed Galore (v0.6.7) was employed to trim adapter sequences and low-quality bases from the raw reads, thereby improving the quality of the data for downstream analysis.

Alignment

Alignment to Reference Genome: The processed reads were aligned to five versions of *Giardia intestinalis* (A1, A2, P64, P392, P407) using HISAT2 (v2.2.1). Based on the mapping results, the reference sequence version A2 was selected, which achieved the highest number of mapped reads on average. Quality control of the sequencing data was performed using qualimap2 (v2.3).

Reference	A1		A2		P392		P407		P64	
Samples	mapped reads	%	mapped reads	%	mapped reads	%	mapped reads	%	mapped reads	%
S4199_001_1_BER_CK_1	67,107,097	84.14%	77,267,272	96.88%	77,099,444	96.66%	76,392,107	95.78%	73,545,798	92.21%
S4199_002_2_BER_CM_1	56,015,256	84.03%	64,377,629	96.58%	64,254,351	96.39%	62,325,653	93.5%	61,248,615	91.89%
S4199_003_3_40_CK_2	53,629,948	80.62%	62,976,637	94.67%	60,297,921	90.65%	59,283,511	89.12%	63,287,462	95.14%
S4199_004_4_40_CM_2	55,745,291	84.22%	62,922,877	95.06%	61,451,203	92.84%	58,489,519	88.36%	63,329,120	95.67%
S4199_005_5_41_CK_3	49,403,967	82.25%	54,446,996	90.64%	56,725,441	94.44%	54,621,588	90.93%	54,855,438	91.32%
S4199_006_6_41_CM_3	52,717,148	82.47%	57,702,903	90.27%	60,348,630	94.41%	56,093,593	87.75%	58,355,567	91.29%
S4199_007_7_24_CK_4	50,505,340	80.05%	59,169,114	93.78%	59,866,412	94.89%	58,364,365	92.51%	59,266,554	93.94%
S4199_008_8_24_CM_4	44,988,974	78.84%	53,190,969	93.22%	53,884,908	94.43%	52,021,781	91.17%	53,369,581	93.53%
S4199_009_9_BER_AK_5	55,416,836	83.58%	64,135,256	96.73%	64,035,944	96.58%	63,497,356	95.76%	60,999,523	92%
S4199_010_10_BER_AM5	53,963,595	84.44%	61,218,569	95.79%	61,358,831	96.01%	57,232,394	89.55%	59,212,267	92.65%
S4199_011_11_BER_CK5	49,556,398	83.43%	57,455,077	96.73%	57,343,682	96.54%	56,855,935	95.72%	54,518,108	91.78%
S4199_012_12_BER_CM5	61,561,387	83.23%	71,402,550	96.54%	71,162,280	96.21%	69,825,467	94.4%	67,905,969	91.81%
S4199_013_13_BER_CK6	58,091,433	84.34%	66,702,245	96.84%	66,635,074	96.74%	65,960,269	95.77%	63,404,979	92.06%
S4199_014_14_BER_CM6	59,748,006	84.01%	68,719,483	96.62%	68,621,707	96.48%	67,596,206	95.04%	65,591,200	92.22%

S4199_015_15_2_CM_7U	58,887,986	85.31%	64,793,316	93.87%	65,305,878	94.61%	63,353,389	91.78%	66,010,229	95.63%
S4199_017_17_8_CK_7	53,258,669	83.78%	60,855,965	95.73%	61,633,864	96.96%	60,970,129	95.91%	58,386,251	91.85%
S4199_018_18_8_CM_7	64,821,057	84.56%	72,669,130	94.80%	73,616,094	96.04%	71,002,335	92.63%	70,730,514	92.27%
S4199_019_19_2_CK_7U	57,385,188	82.77%	64,239,010	92.66%	64,503,393	93.04%	66,224,397	95.52%	66,378,519	95.74%

Quantification

Feature Counts: Gene-level quantification was performed using FeatureCounts (v2.0.3), which assigned reads to genomic features such as genes, allowing for the creation of a count matrix that represents the expression levels of genes across samples.

Status	Assigned	Unassigned_Unmapped	Unassigned_MultiMapping	Unassigned_NoFeatures	Unassigned_Ambiguity
S4199_001_1_BER_CK_1_A2	70901052	2492338	4525648	3240884	1129730
S4199_002_2_BER_CM_1_A2	58500695	2280257	5184045	2676445	1199421
S4199_003_3_40_CK_2_A2	57287481	3543277	4885594	2642730	1135982
S4199_004_4_40_CM_2_A2	56863520	3270955	6826548	2556544	1118999
S4199_005_5_41_CK_3_A2	48779720	5619632	5405171	2283979	1061908
S4199_006_6_41_CM_3_A2	50982546	6218803	7647060	2531035	1233505
S4199_007_7_24_CK_4_A2	52762604	3922354	5822469	2928510	1012044
S4199_008_8_24_CM_4_A2	47248400	3870819	5574125	2688596	991600
S4199_009_9_BER_AK_5_A2	58668952	2170536	3743329	2870448	937265
S4199_010_10_BER_AM5_A2	54953704	2691311	8246300	2404906	1116468
S4199_011_11_BER_CK5_A2	52509174	1945265	3293401	2617416	875355
S4199_012_12_BER_CM5_A2	64801686	2561986	4921010	3269293	1333036
S4199_013_13_BER_CK6_A2	61160853	2174965	3901210	2857429	967518
S4199_014_14_BER_CM6_A2	62701312	2403517	4461818	2970770	1180564
S4199_015_15_2_CM_7U_A2	57683239	4233672	7732349	3033112	1311607
S4199_017_17_8_CK_7_A2	55379100	2713187	4087328	2702060	981198
S4199_018_18_8_CM_7_A2	65712643	3983716	6179785	3172505	1405513
S4199_019_19_2_CK_7U_A2	58200022	5091860	5283775	2764248	1073056

Differential Gene Expression Analysis

DESeq2: Differential gene expression analysis was conducted using the DESeq2 (v1.44.0) package in R. DESeq2 employs a negative binomial distribution model to detect genes that are differentially expressed between experimental conditions, accounting for variability and batch effects.

Charts were created for each group:

Volcano Plot

A volcano plot is a graphical method used in differential gene expression analysis to visualize the relationship between fold change

(log2FC) and statistical significance (usually represented as $-\log_{10}$ p-values). Each point in the plot represents a gene, with the x-axis showing the log2FC between experimental conditions and the y-axis showing statistical significance. Genes with significant changes in expression are usually further from the center (representing zero log2FC) and higher on the y-axis (indicating lower p-values). The graph is named "volcano" due to its characteristic shape, where significant genes resemble erupting volcanoes.

MA plot

Another visualization tool commonly used in RNA-seq analysis to assess differential gene expression is the MA graph. It shows the relationship between the average expression level (M) and the logarithmic ratio of expression levels (A) between two conditions or groups. The M value represents the log2 fold change (log2FC) in gene expression between conditions, calculated as the difference in expression levels, while the A value represents the average expression level, calculated as the average expression across conditions. Each point on the plot represents a gene, with the x-axis showing the average expression (A) and the y-axis showing the log2FC (M). MA plots help identify systematic bias and assess variability in gene expression across conditions.

In addition, a gene expression heatmap for selected genes and a sample correlation heatmap were created.

Heatmap of Gene Expression for Selected Genes

The heatmap displays the expression levels of a subset of genes that met specific criteria, including a log2fold change greater than 2 or less than -2, with a p-value < 0.05 . These genes were selected for visualization based on their potential differential expression between experimental conditions.

The heatmap provides a visual representation of the relative expression levels of the selected genes across different samples or conditions. Each row in the heatmap corresponds to a gene, while each column represents a sample or experimental condition. The color intensity in each cell reflects the expression level of the corresponding gene in the respective sample, with higher intensity indicating higher expression and lower intensity indicating lower expression.

It is essential to acknowledge the limitations of the heatmap, including the exclusion of genes with lower log2fold change or higher p-values, which may also contribute to biological significance. Additionally, the heatmap does not consider the number of mapped reads or adjusted p-values, which are important factors in determining the statistical significance of differential gene expression. Therefore, further analysis and validation are required to confirm the significance of the observed expression patterns.

Results of analysis

Differential gene expression (DGE) analysis was performed to identify genes that show significant changes in expression under different conditions or treatments. Differentially expressed genes (DEGs) were according to defined criteria (log2 fold change $>1/ <-1$, p-value cut-off for FDR < 0.05 , minimum 1 CPM within at least 2 replicates to include the gene in the analysis). The final results with expression values are stored in files named XXX_DESeq2.DE_results.P0.05_C2.DE.xlsx.

Project Structure and File Descriptions

This dataset includes all files related to read preprocessing, alignment, differential gene expression (DGE) analysis, and functional enrichment. Below is a brief description of each folder and its contents:

242114199_data.tar.gz

- Raw data archive (likely containing original FASTQ files or metadata).

242114199_multiqc_report.html

- Comprehensive MultiQC report summarizing all preprocessing and alignment metrics across samples.

Alignment/

Contains BAM files and alignment statistics for each sample.

- *.bam: Aligned reads to the mouse genome (A2).
- *_stats: Alignment summary statistics per sample.

DGE/

Results of differential gene expression analysis.

- count_gene.matrix, count_gene.summary, count_gene.txt: Gene count matrix and summary statistics.

BER_CK_vs_CK/, CK_vs_CM/, etc.

Subfolders with results from specific pairwise comparisons.

- *.matrix: Filtered DGE matrices ($p < 0.05$).
- *.heatmap.pdf, *.cor_matrix.pdf: Visualization of gene expression and sample correlation.
- *.count_matrix*: Raw and normalized count data.
- *.DE_results*: Full and subset DESeq2 results (including MA plots, volcano plots, and Excel exports).
- *.UP.subset, *.UP.xlsx: Subsets of significantly upregulated genes in either condition.

TrimSeq/

Trimmed sequencing data and quality control reports.

- S4199_001_1_BER_CK_1_R1_val_1.fq.gz: Archive of trimmed R1 FASTQ file sample 1_BER_CK_1.
- S4199_001_1_BER_CK_1_R2_val_2.fq.gz: Archive of trimmed R2 FASTQ file sample 1_BER_CK_1.
-
- 242114199_TrimSeq_multiqc_report.html: MultiQC report summarizing trimming results.

Příloha

Údaje v této zprávě jsou platné ke dni jejího vytvoření / tisku (viz datum na str. 1).

Zpráva sestává z těchto částí: Údaje o zakázce | Moje soubory | Vzorky | Sekvenování | Analýza dat | Příloha (tato strana). Pokud některá z uvedených částí chybí nebo je prázdná, nebyly dané služby předmětem zakázky.

Část Údaje o zakázce:

- V této sekci jsou uvedeny parametry objednávky tak, jak byla klientem / klientkou zadána na stránkách společnosti SEQme.
- Pokud následně došlo k jakékoliv **úpravě objednaných služeb**, tyto změny se již v této zprávě v části Údaje o zakázce neprojeví, jsou ale uvedeny v potvrzení objednávky (pdf), které je dostupné v detailu objednávky na webu, a jsou zohledněny při fakturaci. Pokud v průběhu zpracování zakázky byly provedeny služby **nad rámec objednávky** (vícepráce), jsou tyto vícepráce buď rovněž uvedeny v potvrzení objednávky (pdf) nebo jsou evidovány a případně fakturovány samostatně.
- Detailní údaje o průběhu zpracování zakázky, úpravě objednaných služeb a případných vícepracích lze nalézt v detailu objednávky na webu, v této zprávě nebo jsou k dispozici na vyžádání.

Část Moje soubory:

- V této sekci je uveden seznam všech souborů, které se týkají dané zakázky, a které klient/ka nebo zaměstnanec SEQme nahráli prostřednictvím stránek www.seqme.eu do detailu dané zakázky.

Části Vzorky | Sekvenování | Analýza dat:

- V těchto sekcích je uveden seznam všech analyzovaných vzorků, jejich parametry a popis jejich zpracování. Tyto údaje se liší podle toho, jaké služby byly předmětem zakázky. Typicky se jedná o kontrolu kvality vzorků a údaje o přípravě sekvenačních knihoven, výstupy sekvenování, údaje o analýze dat, seznamy datových souborů, které jsou poskytnuty jako výsledky zakázky, apod.

Výsledky zakázky:

- Není-li domluveno jinak, jsou výsledky zakázky dostupné ke stažení ve vašem uživatelském profilu na našem webu.
- Přístup a bezpečnost: Protože jsou výsledky dostupné ve vašem uživatelském profilu, má přístup pouze uživatel, který zná přihlašovací údaje. Pokud chcete sdílet přístup k výsledkům se svými kolegy, nemusí se tito registrovat na našich webových stránkách a nemusíte s nimi sdílet své přihlašovací údaje. Místo toho můžete sdílet výsledky konkrétní zakázky sdílením přímého zabezpečeného odkazu na naše datové úložiště a hesla. Odkaz i heslo jsou k dispozici ve vašem uživatelském profilu. Můžete volně sdílet přístup k datovému úložišti s třetí stranou na vlastní riziko. Vezměte prosím na vědomí, že pro přístup do datového úložiště musí být v rámci vaší IT infrastruktury povolen port 5001. V případě potíží se obraťte na svého místního IT specialistu.
- Firma SEQme nesdílí přístup k datovému úložišti s třetí stranou, pokud to osoba, která objednávku zadává, písemně neschválí.
- Obsah datového úložiště – V datovém úložišti je struktura složek a souborů s výsledky poskytovanými na základě provedené laboratorní analýzy nebo analýzy dat. Soubory / složky lze zobrazit nebo stáhnout, všechny úpravy nebo mazání jsou zakázány.
- Stahování dat – Jakmile zakázku dokončíme, budete informováni emailem. K datovému úložišti lze přistupovat a data stahovat tolikrát, kolikrát je třeba.

DŮLEŽITÉ!

- Zaručujeme, že data jsou dostupná v datovém úložišti nejméně 4 týdny počínaje dnem dokončení zakázky.
- Po uplynutí této lhůty si vyhrazujeme právo smazat všechny nebo vybrané datové soubory bez předchozího upozornění!
- Stáhněte si proto data včas.
- Doporučujeme po stažení dat vytvořit jejich záložní kopii!

Děkujeme, že jste si nás vybrali jako poskytovatele sekvenačních služeb.