Subject:	CCR-SF Data Delivery for mRNA-Seq Samples Sequenced on Flowcell AAAMV57M5
Date:	Friday, October 8, 2021 at 3:18:19 PM Eastern Daylight Time
From:	Chen, Vicky (NIH/NCI) [C]
To:	Ambs, Stefan (NIH/NCI) [E], Panigrahi, Gatikrushna (NIH/NCI) [E], Tang, Wei (NIH/NCI) [V]
CC:	CCRSF_IFX, SFILLUMINALAB
Attachments:	StefanAmbs_CS029103_24RNA_081221_AAAMV57M5.pdf, StefanAmbs_CS029103_24RNA_081221_AAAMV57M5.xlsx, StefanAmbs_CS029103_24RNA_081221_AAAMV57M5_multiqc.html, MainPcaPlot.html, RawCountFile_rsemgenes.txt, InstructionHPCDME_CCRSF_Y21.pdf, image001.png, image002.png, image003.png, image004.png, image005.png, image006.png, image007.png, image008.png

Dear Drs. Ambs and all,

We are pleased to inform you that sequencing and primary analysis of 24 mRNA-Seq samples are completed here at the Sequencing Facility. The details are:

NAS: CS029103
Flowcell: AAAMV57M5

Sample ID	Sample Yield (Mbases)	Percent of (PF) Bases >= Q30	Total Reads (PF)	Total Reads After Trimming	Percent Total Reads after Trimming	Total Mapped Reads (Trimmed)	Percent Total Mapped Reads (Trimmed)	Uniquely Mapped Reads (Trimmed)
10_C2	4328	94.77	42850860	42523936	99.237	41460052	97.498	38745414
11_C3	3770	94.69	37330252	37052278	99.255	36160574	97.593	33821884
12_C4	3690	94.97	36531398	36321668	99.426	35477714	97.676	33154764
13_D1	4013	94.85	39731118	39471764	99.347	38510282	97.564	36162304
14_D2	3876	94.87	38377376	38163070	99.442	37338230	97.839	35099046
15_D3	3906	94.62	38671784	38279132	98.985	37234738	97.272	34833374
16_D4	3606	94.15	35701242	35048204	98.171	33500064	95.583	31416758
17_E1	4031	94.77	39910874	39551516	99.1	38495720	97.331	35785276
18_E2	4362	94.79	43185366	42898558	99.336	41919318	97.717	38982620
19_E3	4074	94.82	40337758	40034426	99.248	38982136	97.372	36359758
1_A1	3823	94.65	37847180	37453368	98.959	36440526	97.296	34053668
20_E4	3818	94.77	37800250	37537632	99.305	36623760	97.565	34157402
21_F1	3971	94.11	39317436	38623172	98.234	36897256	95.531	34188220
22_F2	4056	94.75	40161740	39884246	99.309	38933262	97.616	36029444
23_F3	3682	94.68	36453952	36174542	99.234	35280176	97.528	32691028
24_F4	3617	92.86	35815066	34301400	95.774	31246262	91.093	29160260
2_A2	3818	94.76	37806872	37527234	99.26	36658268	97.684	34291324
3_A3	3819	94.6	37806960	37471862	99.114	36436964	97.238	34071724
4_A4	4102	94.67	40611742	40250430	99.11	39070644	97.069	36518922
5_B1	4363	94.58	43198138	42783460	99.04	41592290	97.216	38819546
6_B2	4092	94.64	40516916	40109528	98.995	38975228	97.172	36427062
7_B3	4117	94.75	40762148	40397058	99.104	39352226	97.414	36747236

8_B4	4069	94.45	40288686	39760752	98.69	38543802	96.939	36027230
9_C1	3960	94.79	39204112	38918008	99.27	37959040	97.536	35507768

Run Comments

24 mRNA-Seq samples were pooled and sequenced on NextSeq 2000 P2 using TruSeq Stranded mRNA Prep and paired-end sequencing. The samples have 35 to 43 million pass filter reads with more than 92% of bases above the quality score of Q30. Reads of the samples were trimmed for adapters and low-quality bases using Cutadapt before alignment with the reference genome (hg38) and the annotated transcripts using STAR. The average mapping rate of all samples is 97%. Unique alignment is above 85%. There are 2.16 to 8.91% unmapped reads. The mapping statistics are calculated using Picard software. The samples have 0.00% ribosomal bases. Percent coding bases are between 60-64%. Percent UTR bases are 30-34%, and mRNA bases are between 93-95% for all the samples. Library complexity is measured in terms of unique fragments in the mapped reads using Picard's MarkDuplicate utility. The samples have 69-75% non-duplicate reads. In addition, the gene expression quantification analysis was performed for all samples using STAR/RSEM tools. Both the normalized count and the raw count are provided as part of the data delivery.

Please find attached further sequencing and analysis details. The data has been transferred to a shared folder on HPCDME and can be accessed at the following link:
https://hpcdmeweb.nci.nih.gov/collection?action=view&path=/FNL_SF_Archive/PI_Lab_Stefan_Ambs/Project_StefanAmbs_CS029103_24RNA_081221&source=browse&init

Please let us know if you have any questions.

Thank you,
Vicky and CCR-SF Staff

Frederick National Laboratory

Vicky Chen | Bioinformatics Analyst III
CCR-SF Bioinformatics Group
Advanced Biomedical and Computational Science (ABCS)
Biomedical Informatics and Data Science (BIDS) Directorate

✉ vicky.chen@nih.gov *[Contractor]*
📍 Post Office Box B, Frederick, MD 21702
🔗 frederick.cancer.gov