様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ９ 月 １１ 日現在

機関番号：１２６０１
研究種目：若手研究
研究期間：2018～2019
課題番号：１８Ｋ１６１７０

研究課題名（和文）Analysis of MHC class I as one of the host factors responsible for severe dengue infection by means of a portable sequencer

研究課題名（英文）Analysis of MHC class I as one of the host factors responsible for severe dengue infection by means of a portable sequencer

研究代表者
ＲＵＮＴＵＷＥＮＥ　ＬＵＣＫＹ（Runtuwene, Lucky）

東京大学・大学院新領域創成科学研究科・特任助教

研究者番号：４０７９３７８２

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：2018年よりインドネシア国のマナド市におけるデング熱の臨床検体は23個を回収した。去年の報告により、この臨床検体はHLA-A、HLA-B、とHLA-Cの遺伝子がMinIONという簡易型シークエンサーを用いてシークエンスを行った。Athlonというバイオインフォマティクスツールで、HLAタイプを明らかにした。2019年に同じ臨床検体はIlluminaでシークエンスを行い、結果の比較を行った。MinIONはIlluminaに一致するHLAタイプが少なく、一致率は36.5%であった。この結果により、MinIONはHLAタイピングとする方法がまだできない、デング熱の重症度のHLAは明らかにしてない。

研究成果の学術的意義や社会的意義
We think that MinION may have advantage over Illumina because the pipeline is able to detect MHC up to three fields which is more specific. Furthermore, with its relatively cheap investment overhead, MHC class I detection can be very valuable in developing countries, but it will still need refining.

研究成果の概要（英文）：Since 2018, we have collected 23 samples from Manado, Indonesia. Previously we have reported on 2019, that these samples were collected from hospitalized patients that were positive for dengue fever. We were able to sequence their HLA-A, HLA-B, and HLA-C genes and had analyzed the HLA type using a bioinformatics pipeline called Athlon.
Continuing on this project, in 2019 we have performed Illumina sequencing to validate the MinION results. Using the available data from 23 samples, we found that the results obtained from MinION+ Athlon and Illumina have 36.5% difference. We conclude that MinION is still not good for HLA typing and we were not able to determine the HLA correlated to dengue severity.
One of the reasons is because of the high error rate of the v9.4 flowcell. Therefore, using the new v10.3 flowcell that has lower error rate available in 2020, we will re-sequence and re-analyze the samples.

研究分野：Infectious diseases

キーワード：MinION　Dengue virus　Dengue hemorrhagic fever　Dengue fever　MHC class I

１．研究開始当初の背景

Dengue fever (DF) is a disease with great interest due to its wide spread in the tropical countries and the potential to cause severe disease. The high mortality rate of severe dengue is due to the difficulties in vector control, medicine unavailability, and vaccine ineffectiveness. There are many factors for the disease to advance into severe disease; host factor is one of them. Major histocompatibility complex (MHC) is claimed to have a role in this process. Depending on the MHC molecules, some individuals are susceptible to severe disease; probably in combination with the virus strain itself. Using a recently available portable sequencer, MinION, we will sequence MHC genes of the subject patients. We will compare MHC alleles of non-severe to severe dengue patients and describe which alleles are correlated to severe dengue. We will also sequence the complete genome of the virus and associate which combination of MHC and virus genome that is highly probable to cause severe dengue.

２．研究の目的

This research aims to describe the MHC class I (that is cell surface proteins essential for the immune system to recognize foreign molecule) that has a correlation with the severity of dengue fever using MinION portable sequencer. This research will benefit the treatment of dengue fever in these aspects:

1. We will be able to predict which patient will develop into severe dengue (that is dengue hemorrhagic fever or dengue shock syndrome) based on the MHC class I molecule they possess.
2. By predicting the possible course of the disease, we could prepare the treatment necessary for severe dengue beforehand, such as platelet transfusion.
3. The diagnosis technique by means of MinION portable sequencer is not costly to be employed in the hospitals in the developing countries.
4. This technology will be an alternative to serologic HLA typing that is not routinely used for infectious diseases.
5. Pathogen identification can be performed in parallel with HLA typing.

３．研究の方法

The site of research was in Manado, North Sulawesi, Indonesia. The country is endemic to dengue fever, especially in the rainy season and North Sulawesi Province is categorized by the Indonesian Ministry of Health as high risk for the disease. The province had the highest risk for dengue fever for five consecutive years (2005-2010), while the bordering provinces had low to medium risk. The provincial capital city also has the highest dengue fever cases compared to other areas in the province.

Since 2018, we collected samples from Budi Mulia Hospital in North Sulawesi, Indonesia. This research included the collaboration of the local physician. The criteria of inclusion were:

1. Body temperature > 38°.
2. DENV NS1 antigen (+) or DENV IgM antibody (+) and/or DENV IgG antibody (+).
3. Progression to severe dengue as shown by signs of hemorrhagic manifestation, thrombocytopenia, increased vascular permeability, or shock.

Patients that only had signs 1 and 2 until recovery were included in DF group. Patients that had all signs of the above were included in DHF group.

In total, we were able to collect 23 blood samples but only 16 were included for sequencing. Eight samples (50%) were diagnosed as severe dengue and other eight samples (50%) were diagnosed as dengue fever. About 5 ml of blood samples were collected. Half was transferred to PAX Gene DNA tube (BD BioSciences) and the remaining half was transferred to PAX Gene RNA tube (BD BioSciences). Samples were kept in -20°C until subsequent processes.
DNA was extracted using MagAttract HMW DNA Kit (Qiagen) and using three sets of primers, each for HLA-A, HLA-B, and HLA-C genes we amplified the respective genes. For multiplex sequencing, we also added an additional 20 nucleotides in the primers as barcode. Amplification was performed with KAPA HiFi Hotstart ReadyMix DNA polymerase (KAPA Biosystems) under these PCR conditions: 95°C for 3 minutes; 30 cycles of 98°C for 30 seconds, 55°C for 15 seconds, 72°C for 90 seconds; and 72°C for 3 minutes for HLA-A and HLA-C genes. For HLA-B genes the conditions were: 95°C for 3 minutes; 30 cycles of 98°C for 30 seconds,

58°C for 15 seconds, 72°C for 90 seconds; and 72°C for 3 minutes.

Amplicons were purified with AMPure XP (Beckman Coultier) and confirmed with Bioanalyzer (Agilent). The amplicons length was around 3 kb. Sequencing library was created with 1D Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies). Sequencing was performed with MinION flowcell FLO-MIN107 (Oxford Nanopore Technologies) for 48 hours. In total we performed seven rounds of sequencing.
Raw data was converted and demultiplexed using Guppy software with standard parameters. The resulting FAST5 for each sample was subjected to Athlon pipeline (Liu Chang et al. Journal of Molecular Diagnostics. 2018).

Confirmation with Illumina was done at the Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo. Samples were enriched for HLA genes using SureSelect XT HL with bait v5+MHC (Agilent). Sequencing was performed with HiSeq 3000 (Illumina). Data analysis was performed with OptiType (Szolek A et al. Bioinformatics. 2014).

4．研究成果

Table 1 and table 2 shows the statistics of MinION sequencing before demultiplexing. Table 2 shows the number of reads mapped to each of HLA-A, HLA-B, and HLA-C genes for every sample.

	Seq #1	Seq #2	Seq #3		Seq #4			Seq #5			Seq #6			Seq #7		
	AM	AN	#3	#4	#8	#14	#15	#12	#13	#16	#17	#18	#20	#21	#22	#23
1D reads #	1.5 mio	673,388	2,503,502		1,523,261			2,643,369			2,515,503			2,028,549		
Average 1D length	2,674	3,312	1,948		2,118			1,861			2,245			2,540		
1D average score	7	7	7		7			7			7			7		
Demultiplex reads #	N/A	N/A	278,408 (11.12%)	233.974 (9.35%)	303,754 (19.94%	98,531 (6.47%)	136,244 (8.94%	192,765 (7.29%(	154,948 (5.86%)	258,985 (9.79%)	320,988 (12.76%)	102,301 (4.07%)	320,835 (12.75%)	75,053 (3.7%)	45,544 (2.26%)	88,169 (4.35%)

Table 1. MinION sequencing before demultiplexing.

	Seq #1	Seq #2	Seq #3		Seq #4			Seq #5			Seq #6			Seq #7		
	AM	AN	#3	#4	#8	#14	#15	#12	#13	#16	#17	#18	#20	#21	#22	#23
Total reads	1.5 mio	673,388	278,408	233.974	303,754	98,531	136,244	192,765	154,948	258,985	320,988	102,301	320,835	75,053	45,544	88,169
HLA-A	72,277	31,237	3,826	4,451	3,146	1,784	2,333	9,040	8,725	11,585	4,784	2,141	6,497	1,769	1,615	2,630
HLA-B	445,820	160.112	68,443	58,859	131,984	38,000	53,024	59,741	35,995	66,197	127,448	41,193	124,107	28,988	16,364	33,537
HLA-C	55,386	40,519	15,371	14,626	6,011	3,514	4,200	7,558	9,722	13,258	10,032	5,663	15,468	2,836	2,554	2,996

Table 2. The number of reads mapped to each HLA-A, HLA-B, and HLA-C genes.

There are discrepancies in the sequencing for each HLA gene with a bias towards HLA-B. The minimum input reads for the subsequent pipeline is 1,000 reads. Although there were discrepancies, all samples were sufficient for subsequent procedure to Athlon. Table 3 shows the HLA type for each sample. Figure 1 summarizes the distribution of HLA typing. It is obvious that some alleles are found in high frequency in the small number of populations.

	Seq #1	Seq #2	Seq #3		Seq #4			Seq #5			Seq #6			Seq #7		
	AM	AN	#3	#4	#8	#14	#15	#12	#13	#16	#17	#18	#20	#21	#22	#23
HLA-A	24:02:87 02:01:29	02:01:29 24:02:87	24:232N:XX 26:01:09	24:02:02 -	11:50Q:XX 24:02:60	24:02:60 11:50Q:XX	26:01:09 34:01:01	26:01:09 30:02:03	11:01:40 24:02:62	24:02:87 -	11:50Q:XX -	11:01:40 -	11:01:40 11:50Q:XX	24:02:93 24:232N:XX	11:01:08 26:01:09	24:02:87 11:01:40
HLA-B	27:95:XX 18:27:XX	67:02:XX 54:09:XX	27:95:XX 27:62:XX	27:95:XX 27:62:XX	27:90:03 07:225:XX	27:95:XX 07:225:XX	27:90:03 07:197:XX	27:90:03 07:225:XX	27:90:03 -	27:95:XX 27:62:XX	27:90:03 07:225:XX	27:90:03 07:225:XX	27:90:03 07:225:XX	27:95:XX 27:62:XX	27:90:03 07:225:XX	27:05:06 07:225:XX
HLA-C	07:02:56 06:147:XX	07:02:56 12:03:36	04:01:11 14:02:02	04:01:03 14:02:04	04:01:11 03:04:28	04:01:03 16:85:XX	04:03:03 14:02:11	03:04:28 04:03:01	07:04:10 04:01:11	07:04:10 08:01:19	04:01:11 07:02:56	01:02:04 08:01:19	04:01:11 14:02:04	04:01:11 14:02:02	16:67:XX 04:01:03	16:67:XX 04:01:39

Table 3. The output of Athlon pipeline showing the type of HLA-A, HLA-B, and HLA-C for each sample.

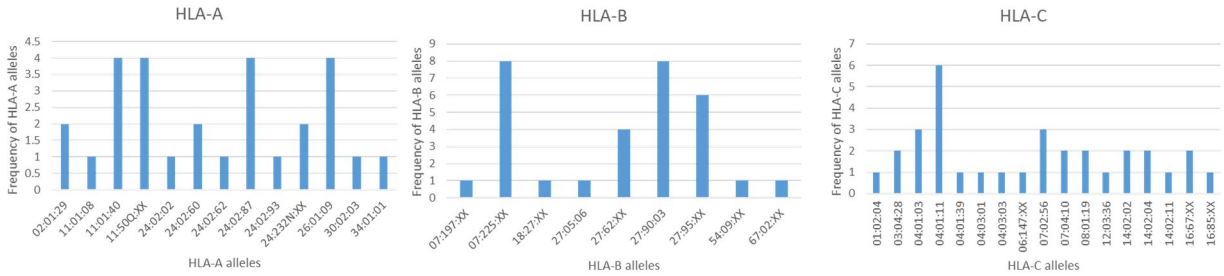

Figure 1. The distribution of each HLA alleles. Some alleles are found up to eight times in a small population.

In 2019, we sequenced the same samples using Illumina HiSeq 3000. The sequencing and data analysis were outsourced to the Computational Biology and Medical Science Department, Graduate School of Frontier Sciences, The University of Tokyo.
The comparison of HLA-typing between two sequencing platforms is summarized in Table 4.

	A1		A2		B1		B2		C1		C2	
	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore	Illumina	Nanopore
Healthy	A*02:01	A*02:01:29	A*02:01	A*30:02:19	B*35:01	B*35:186:XX	B*40:01	B*40:01:10	C*03:03	C*03:04:28	C*07:02	C*07:02:56
AM	A*02:01	A*02:01:29	A*24:07	A*24:02:87	B*15:35	B*27:95:XX	B*18:01	B*18:27:XX	C*07:02	C*07:02:56	C*07:04	C*06:147:XX
AN	A*02:03	A*02:01:29	A*24:07	A*24:02:87	B*38:02	B*67:02:XX	B*55:02	B*54:09:XX	C*07:02	C*07:02:56	C*12:03	C*12:03:36
3	A*24:07	A*24:232N:XX	A*34:01	A*26:01:09	B*15:21	B*27:95:XX	B*35:05	B*27:62:XX	C*04:01	C*04:01:11	C*04:03	C*14:02:02
4	A*24:02	A*24:02:02	A*24:02	A*24:02:02	B*40:01	B*27:95:XX	B*51:01	B*27:62:XX	C*04:01	C*04:01:03	C*14:02	C*14:02:04
8	A*11:01	A*11:50Q:XX	A*24:07	A*24:02:60	B*27:06	B*27:90:03	B*35:05	B*07:225:XX	C*03:04	C*03:04:28	C*04:01	C*04:01:11
12	A*33:03	A*26:01:09	A*34:01	A*30:02:03	B*15:21	B*27:90:03	B*58:01	B*07:225:XX	C*03:02	C*03:04:28	C*04:03	C*04:03:01
13	A*11:01	A*11:01:40	A*24:07	A*24:02:62	B*35:05	B*27:90:03	B*51:01	B*27:90:03	C*04:01	C*04:01:11	C*14:02	C*07:04:10
14	A*11:01	A*11:50Q:XX	A*24:07	A*24:02:60	B*15:35	B*27:95:XX	B*35:05	B*07:225:XX	C*04:01	C*04:01:03	C*07:02	C*16:85:XX
16	A*24:02	A*24:02:87	A*24:02	A*24:02:87	B*15:13	B*27:95:XX	B*18:01	B*27:62:XX	C*07:04	C*07:04:10	C*08:01	C*08:01:19
17	A*11:01	A*11:50Q:XX	A*24:07	A*11:50Q:XX	B*07:05	B*27:90:03	B*35:05	B*07:225:XX	C*04:01	C*04:01:11	C*07:02	C*07:02:56
18	A*11:01	A*11:01:40	A*24:02	A*11:01:40	B*15:02	B*27:90:03	B*15:02	B*07:225:XX	C*01:02	C*01:02:04	C*08:01	C*08:01:19
20	A*11:01	A*11:01:40	A*24:07	A*11:50Q:XX	B*35:05	B*27:90:03	B*51:01	B*07:225:XX	C*04:01	C*04:01:11	C*14:02	C*14:02:04
21	A*24:07	A*24:02:93	A*24:07	A*24:232N:XX	B*35:05	B*27:95:XX	B*35:05	B*27:62:XX	C*04:01	C*04:01:11	C*04:01	C*14:02:02
22	A*11:01	A*11:01:08	A*34:01	A*26:01:09	B*39:01	B*27:90:03	B*51:01	B*07:225:XX	C*04:01	C*04:01:03	C*07:02	C*16:67:XX
23	A*11:01	A*11:01:40	A*24:02	A*24:02:87	B*15:21	B*27:05:06	B*15:35	B*07:225:XX	C*04:03	C*04:01:39	C*07:02	C*16:67:XX

Table 4. Comparison between MinION and Illumina sequencing platforms. Yellow colors indicate the similarity between the two platforms. Note that Illumina can only type up to two fields while MinION is more specific with three-fields typing.

The comparison in table 4 shows that there are 35 HLA types that are similar between the two platforms. Therefore, the similarity is only 36.5%. Furthermore, even the sequencing bias is toward the HLA-B gene, but the sequencing similarity is the lowest.

There is still a disadvantage for MiniON in HLA typing, mainly because the high sequencing error of the v9.4 flowcell that we used. As shown in table 4, there is a high dissimilarity, especially in the HLA B gene.

This is intriguing because MinION sequencing has a sequencing bias towards HLA-B gene. This dissimilarity prevents us to further analyze the data to define which HLA is correlated to severe dengue using MinION sequencing platform. However, since the beginning of 2020, a new version of flowcell (v10.3) has been available. The change in the pore structure to longer barrel and dual reader head increases the sequencing accuracy to 99.995% single molecule consensus accuracy. We will need to resequence and re-analyze the samples using this new flowcell.

We think that MinION may have advantage over Illumina because the pipeline is able to detect MHC up to three fields, while Illumina pipeline is only able to detect up to two fields. Furthermore, with its relatively cheap investment overhead, MHC class I detection can be very valuable in developing countries, so we will keep on working to refine the results.

５．主な発表論文等

〔雑誌論文〕　計0件

〔学会発表〕　計3件（うち招待講演　2件／うち国際学会　2件）

１．発表者名 Lucky Runtuwene
２．発表標題 An international collaborative effort for infectious disease analyses using MinION
３．学会等名 London Calling 2019（招待講演）（国際学会）
４．発表年 2019年

１．発表者名 Lucky Runtuwene
２．発表標題 MinION applications in human infectious diseases
３．学会等名 4th MinION Technology Seminar in Hokkaido University（国際学会）
４．発表年 2019年

１．発表者名 Lucky Runtuwene
２．発表標題 On-site sequencing
３．学会等名 MinION Technology Training in Gadjah Mada University（招待講演）
４．発表年 2019年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考