

出國報告（出國類別：研究）

感染性病原基因體組分析研習與技術交流

服務機關：衛生福利部疾病管制署

姓名職稱：林智暉 副研究員

派赴國家/地區：新加坡

出國期間：民國108年11月27至11月30日

報告日期：民國108年12月13日

摘要

近年來全球化趨勢及氣候環境變化，使得各種傳染病病原體變異及傳播速度更勝以往，包括新興病原浮現、病毒性病原的變異、人畜共通病毒基因重組以及細菌性病原的抗藥性變異等，均有可能造成大規模甚至全球性疫情的發生。本次出國研習新型感染性病原基因體組分析技術，透過可攜式的奈米測序芯片檢測平台來獲得及分析解讀病原基因組成，不需增殖複製或培養，可直接自臨床檢體中獲得感染性病原體基因體組資訊，快速獲取相關病原基因變異及抗藥性等疫情資訊。此技術未來可應用於我國傳染病原檢測工作上，並可支援當疫情發生時之在地化即時檢測，以有效提升檢驗時效，迅速建構防疫網絡。

目錄

壹、 目的	4
貳、 過程	5
參、 研習內容重點	6
肆、 心得與建議	14
伍、 附錄	16

壹、目的

近年來全球化趨勢及氣候環境變化，使得各種傳染病病原體變異及傳播速度更勝以往，包括新興病原浮現、病毒性病原的變異、人畜共通病毒基因重組以及細菌性病原的抗藥性變異等，有可能造成大規模甚至全球性疫情的發生。國內現有的檢驗技術對於已知病原均可透過已建立之標準檢驗流程進行病原分離與確認，但在檢驗時效上仍有待提升；目前病原體檢測方法大多仍仰賴實驗室利用精密或大型檢測儀器進行檢驗，且操作人員須接受專業儀器操作訓練，此外患者採樣檢體必須低溫運送至實驗室方能進行檢測，在運送過程中常因天候、車況、溫控等各種突發狀況而導致檢驗確認時間上的延遲或者因檢體保存狀況異常而無法獲得正確有效的檢驗結果。

本次出國研習新型感染性病原基因體組分析技術，透過可攜式的奈米測序芯片基因定序技術，自臨床檢體中獲得病原基因體組資訊，對於我國現行傳染病原檢測技術上有顯著改進，在一日生活圈的台灣，此技術亦可立即支援特殊或發生嚴重疫情地區的在地化即時檢測，有效提升檢驗效能，達到快速防疫目的。

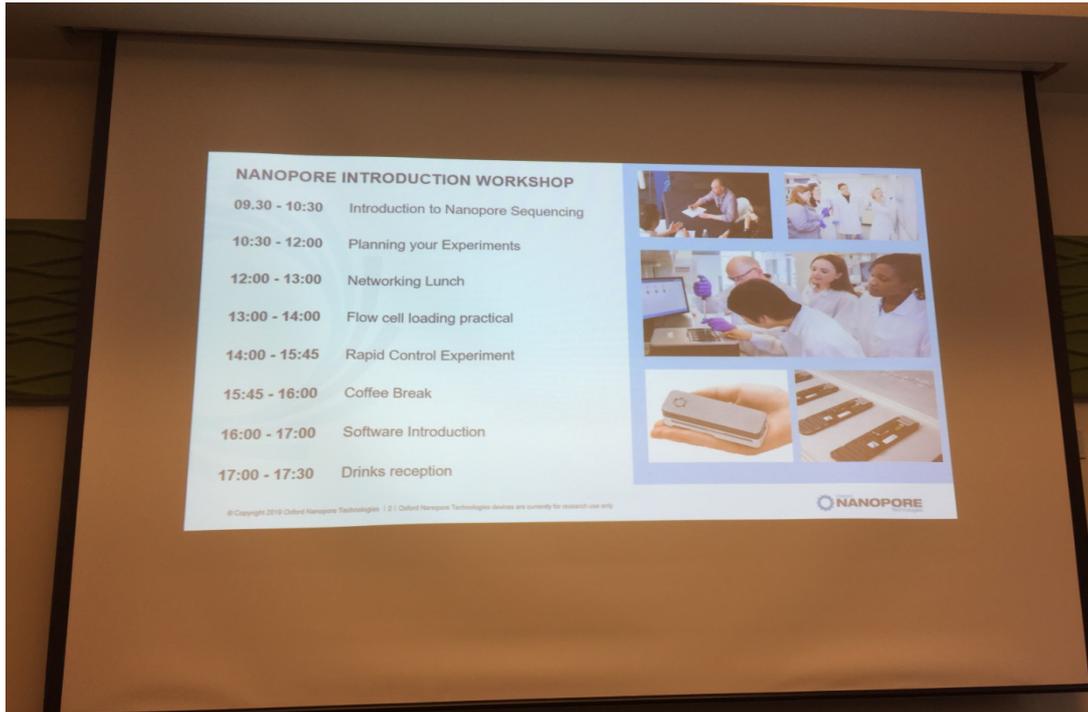
貳、過程

日期	地點	工作紀要
11月27日	台北→新加坡	啟程
11月28-29日	新加坡基因體組研究所	研習
11月30日	新加坡→台北	回程

參、研習內容重點

(一) NANOPORE INTRODUCTION WORKSHOP

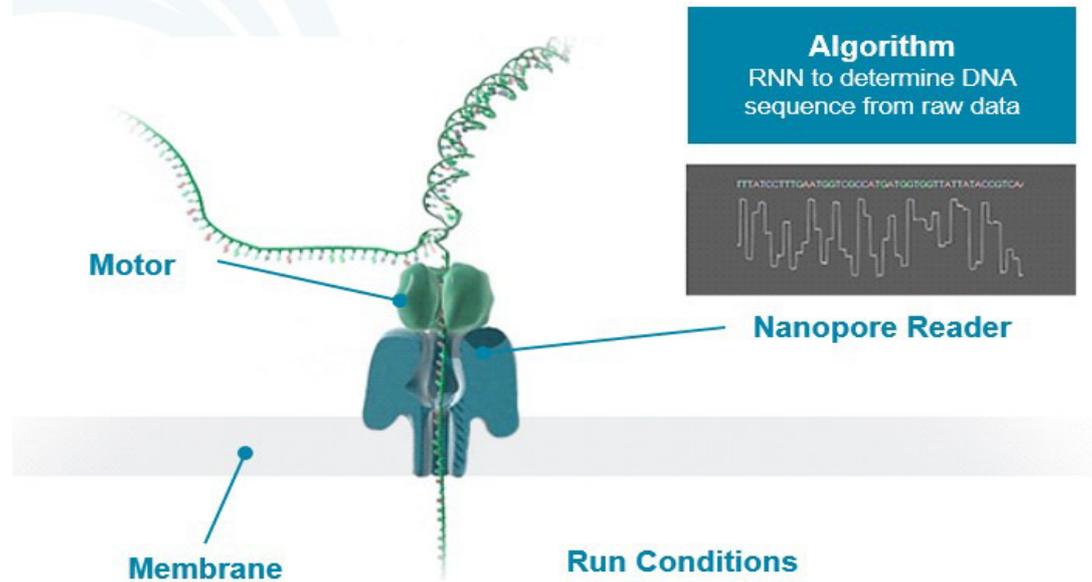
本次研習地點為新加坡基因體組研究所(Gemone Institute of Singapore, GIS) , 研習第一天課程為 NANOPORE INTRODUCTION WORKSHOP , 主要介紹技術原理以及實機操作。



Nanopore 為一新型基因體解讀技術，該技術的原理是藉由一個啟動蛋白 (motor protein) 攜帶 DNA 或 RNA 基因體通過穿膜蛋白 (membrane protein) 中的奈米孔，藉由紀錄不同鹼基通過奈米孔時造成的電流擾動訊號，回推出鹼基，進而解讀受測基因體組序列。

Nanopore DNA sequencing – replace slide

How does it work?



圖一、Nanopore 技術原理。

透過此項技術可以直接讀取基因體原序列片段，不需經過複製增幅，因此可用於 RNA 定序，也可避免 DNA 複製過程中所發生的變異。Nanopore 與其他定序技術最主要的差異在於不需合成，目前大多數的定序技術都是一邊進行合成一邊解讀鹼基。而 Nanopore 定序技術則是藉由鹼基序列通過奈米孔時，造成的連續電流變化來解讀序列，不需經過序列放大的步驟，因此定序速率更快，Nanopore 定序平台定序一個鹼基僅需 0.002 秒，相較 Illumina 定序平台的 2-20 秒快速許多；而 Nanopore 一個測序芯片(flow cell) 上有 2048 個奈米孔，每 4 個孔洞對應到一個積體電路(integrated circuit, IC)，總共有 512 個 IC。Nanopore 定序的另一特色為長序列片段(long reads) 讀取，其可讀取的序列片段平均高達 6-8 kb，遠大於現行 Sanger 定序的 800-1000 以及 Illumina 的 150bp 左右。長序列片段讀取技術的優勢包括(1) 可更容易找出結構變異，包括細菌具有

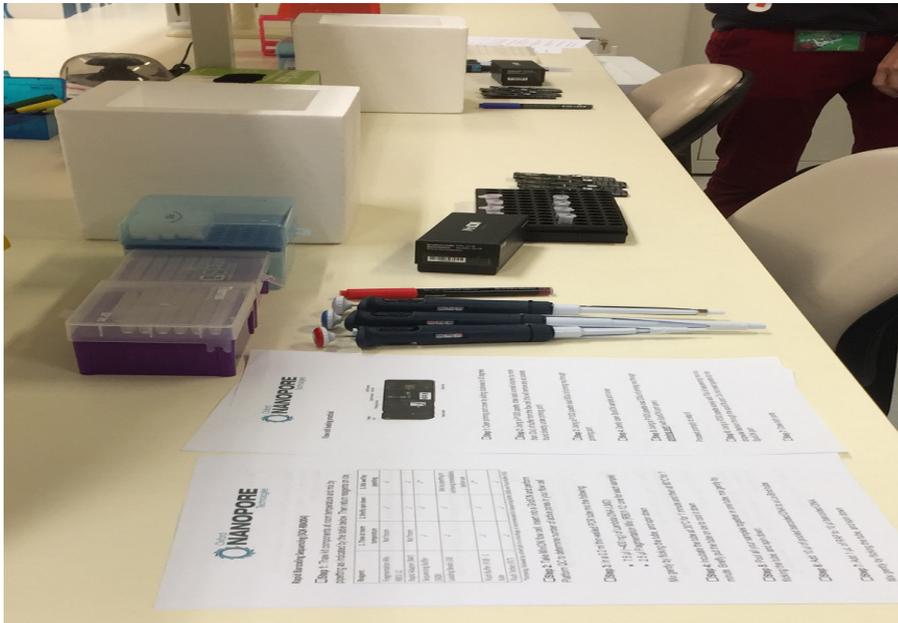
特殊的基因體變異，使用短序列片段定序只能看到破碎片段的結果，但若是用 Nanopore 的長序列片段讀取技術，則較容易看到這些特殊的結構變異。(2) 直接判斷偵測基因位置，短序列片段定基因都是 100 多 bp，就獲得的序列資料不易判斷該基因是位在質體 (plasmid) 或染色體 (chromosome) 上，以及該短片段序列應在基因體上的正確位置，若用 Nanopore 長片段定序就無此問題，可以確定序列在基因體上的相對位置。(3) 精準確認拷貝複製數 (copy number) 的數量，許多基因有不同的拷貝數。使用短序列片段定序時常要猜測有幾個拷貝數，若使用 Nanopore 技術定序，可藉由讀取完整基因體 (genome) 進而精準確認拷貝數量。(4) Nanopore 可以直接建庫，避免因 PCR 過程引子與模板配對錯誤或放大不均造成的誤差 (bias)。

在操作訓練課程方面以 GridION 儀器(如下圖)進行，GridION 為高通量設計，可同步進行5個測序芯片的基因解讀作業，透過軟體操作先確認可解讀基因序列的每個測序芯片品質(flow cell check)，之後依照檢驗方法將欲分析的 DNA

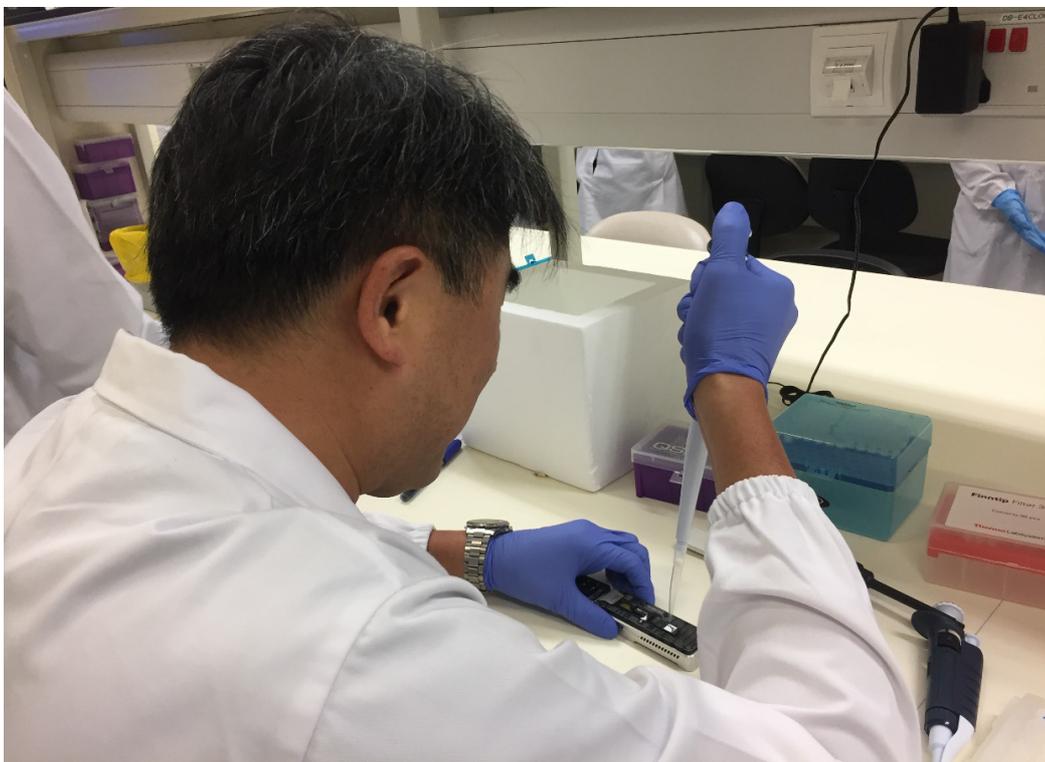


圖二、以高通量奈米檢測儀(GridION)同步進行基因解讀。

或 DNA/RNA 基因體經過編碼 (Barcoding)，再接上 adaptor 後，將檢體注入測序芯片，即可進行6-72小時的基因體定序解讀。



圖三、依照標準作業程序進行上機檢體配製。



圖四、實際操作將待測檢體注入測序芯片。

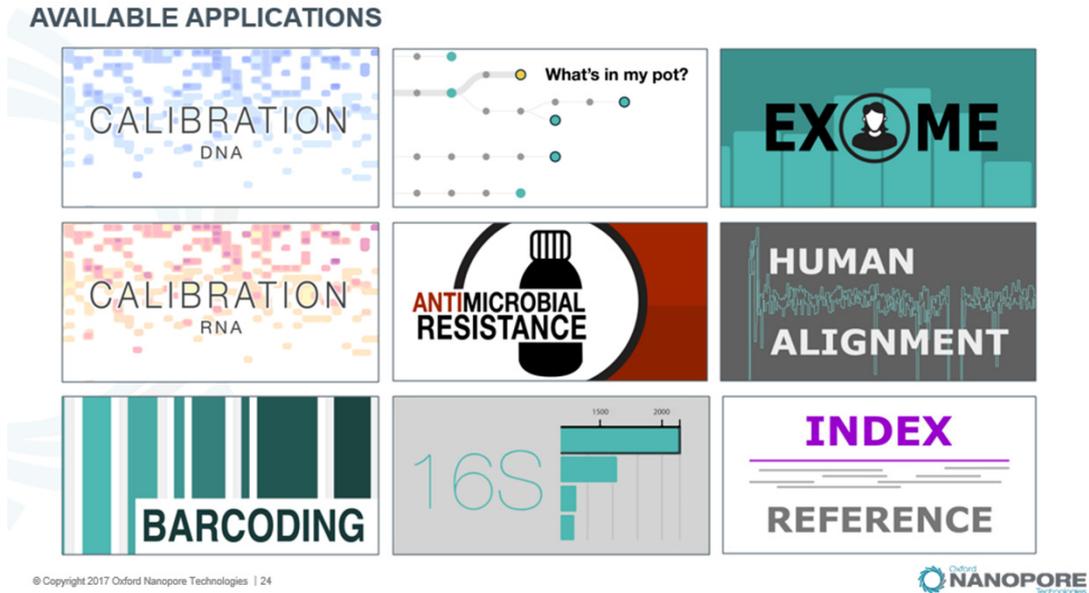
(二) DATA ANALYSIS WORKSHOP

研習第二天課程為 NANOPORE INTRODUCTION WORKSHOP，主要介紹上機完成獲得序列資料的分析與組裝。除了定序平台外，定序後的生物資訊分析也是重要的一環，正確的分析方法才能獲得最正確的基因資訊。將前一天上機所獲得的資料，利用 MinKnow 軟體將電流資料轉化為基因序列資料檔。



圖四、電流擾動訊號，回推出鹼基以解讀受測基因體組序列。

Nanopore 應用的資料分析軟體包括 (1) basecalling 有 Albacore、Scrapie、Guppy 等；(2) Reads Cleaning 有 Porechop、NanoFilter 等軟體可選擇；(3) alignment 有 NGMLR、GraphMap、marginAlign 等應用軟體，(4) Consensus & polish 則可利用 Pilon、Racon 或 Nanopolish。其中有些軟體包括 Guppy、Nanoplot、Nanopolish 等操作均須透過寫指令(command line)來執行，因此在操作的過程中需要熟悉各項指令的邏輯。



圖五、基因解碼資料後續相關分析軟體。

Oxford NANOPORE technologies

STEP : BASECALLING WITH GUPPY

Now we will use Guppy, our standard basecaller package, to basecall a Lambda dataset. But first we will need to download some raw data from Dropbox:

- Let's make a new directory for our Lambda data and download it:


```
ubuntu@ip-10-11-12-14:/home/ubuntu$ mkdir lambda
ubuntu@ip-10-11-12-14:/home/ubuntu$ cd lambda
ubuntu@ip-10-11-12-14:/home/ubuntu/lambda$
wget https://www.dropbox.com/s/t81ej91351d6fo/lambda.tar.gz?dl=0 -O lambda.tar.gz
```
- Let's extract this tar file:

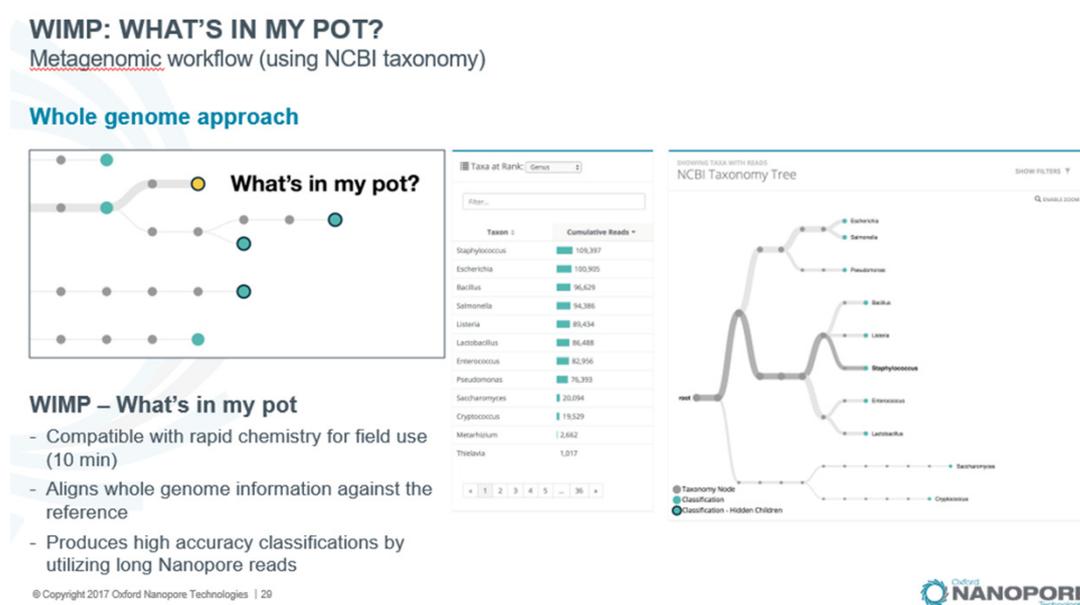

```
ubuntu@ip-10-11-12-14:/home/ubuntu/lambda$ tar -xf lambda.tar.gz
```

This should unzip and give us 2 individual multi-read fast5 files. To double-check that the extraction completed successfully, list the contents of the directory:

```
ubuntu@ip-10-11-12-14:/home/ubuntu/lambda$ ls
FAL06504_4dd5f26403973742429ebeec317fec22caaed8da_0.fast5
FAL06504_4dd5f26403973742429ebeec317fec22caaed8da_1.fast5
```

圖六、以下 command line 方式進行基因體解讀分析。

此外系統也提供即時分析 (real-time analysis)軟體包括 MinKNOW 以及 EPI2ME 等。在感染性病原體的偵測上，透過長片段基因定序以及完整的基因體序列解讀，可短時間之內獲得近乎全基因序列資料。未來可應用在我國傳染病原的即時偵測，而接近全長 (99%) 的基因體解讀除了可提供較為完整感染病原資訊，亦可快速分析病原體基因變異情形以及是否具有抗藥基因的偵測，並可支援疫情發生之在地即時檢測(real-time sequencing)，以有效提升檢驗效能，達到快速防疫目的。



圖七、全基因體解讀鑑定感染病原

Nanopore 數據分析本身就是一個全球各項研究的熱門重點，目前幾乎每個月甚至每周都有新的工具發表，透過加入 Oxford Nanopore Technology 的社群討論，可以即時獲取最新的研究資訊與分析方法的進展。由於 Nanopore 定序的資料量非常龐大，數據處理和運算對網路和計算硬體的性能要求都相當高，因

此在設計應用 Nanopore 技術的基因體組分析或研究時，也要考慮數據處理的硬體設備，以能達到最佳效能。

肆、心得及建議

1. 本次研習之新型基因體組解讀技術突破過往短序列片段讀取的限制，能夠直接一次讀取 6~8 KB 以上的基因片段，且由於不須透過核酸增幅過程，因此能直接應用在臨床檢體上去偵測病原，再配合後續基因解讀訊號技術，能夠更迅速分析感染病原體基因資訊，有效提升檢驗效能。
2. 近年來由全球化趨勢及氣候變遷加劇，使得各種傳染病病原體變異及傳播速度更勝以往，包括新興病原浮現、病毒性病原的變異、人畜共通病毒基因重組以及細菌性病原的抗藥性變異等，均有可能造成大規模甚至全球性疫情的發生。現有的檢驗技術與方法受限於大多仍需在實驗室裡利用精密儀器進行檢驗，操作人員亦須受過專業儀器操作訓練以及檢體必須運送至檢驗實驗室，在運送過程得考慮天候、車況、溫控等各種突發狀況而導致檢驗確認時間上的延遲或者因檢體保存狀況異常而無法進行有效的檢驗。利用此次研習之可攜式即時檢測基因體組解讀技術可支援疫情發生地區之在地即時檢測即定序 (rea-time sequencing)，確定病原體及瞭解其基因組成，協助防疫網絡的建置與防疫策略的實施。
3. 本次研習地點為新加坡基因體組研究所的研究大樓 Genome (Genome Institute of Singapore)，建築物位於新加坡的生醫園區 (Biopolis Biomedical Research Hub, Buona Vista)。新加坡政府為了促進科技研究興盛發展規劃 Biopolis 生醫園區，其中包含 5 個生醫研究所、地區性和國際跨國生醫公司、藥廠等研發單位，包括 Institute of Chemical and Engineering Sciences (Helios)、Institute of Bioengineering and Nanotechnology (Nanos)、Bioinformatics Institute (Matrix)、Institute of Molecular and Cell Biology (Proteos) 以及

Institute of Medical Biology (Neuro&Immunos)等，園區內提供生活、工作、學習和娛樂的完善環境，讓性質相近的研究單位或性質不同但能相互合作的各領域科技人才能夠近距離交流，以提升研發量能。這樣的規劃值得我國政府借鏡及學習，提供完善資源並集中研發量能有助於生技產業發展。

4. 傳染病的防治工作是屬於全球性的防疫戰線，透過持續的國際合作及研究交流，不但能彰顯我國在公共衛生領域實力與能見度，也能藉由交流合作強化我國自己的防疫能力。本次研習夥伴來自新加坡、菲律賓以及泰國等新南向政策指標國家，未來可以透過技術及經驗交流建立研究合作管道，有助於在防疫檢驗方面更密實合作，共同提升防疫成果。



Overview

The workshop will be available at different Oxford Nanopore locations around the world. Oxford Nanopore experts will provide attendees (up to 20) with a nanopore sequencing overview and case study discussion.

A nanopore sequencing demonstration as well as a rapid sequencing control experiment will allow each participant to gain practical hands-on experience of running the MinION/GridION.

The workshop will cover the following:

- ✓ Comprehensive technology overview
- ✓ Library preparation in groups
- ✓ Library loading practice
- ✓ Running the MinION/GridION
- ✓ Introduction to basecalling and EPI2ME analysis workflow

Agenda

- 08:45 - Registration opens
- 09:30 - Introduction to nanopore sequencing
- 10:30 - Planning your experiments
An overview of applications, kits and protocols. Nanopore sequencing experts will also discuss hints & tips on how to maximise your sequencing runs.
- 12:00 - Networking lunch
- 13:00 - Flow cell loading practical
An Oxford Nanopore specialist will demonstrate the correct handling and loading of a flow cell. Participants will then take part in a MinION Flow Cell loading practical.
- 14:30 - Rapid control experiment
An Oxford Nanopore specialist will perform a live Rapid control experiment. Participants will then complete a Rapid barcoding experiment to gain practical hands-on experience of running the MinION/GridION.
- 16:00 - Software introduction
Our technical specialist will provide an introduction to real-time sequencing including basecalling, MinKNOW feedback and EPI2ME analysis workflows.
- 17:00 - Drinks reception
- 17:30 - Workshop closes



Data Analysis Workshop

A one day workshop with Oxford Nanopore experts dedicated to data analyses

Overview

The workshop will be available at different Oxford Nanopore locations around the world. Oxford Nanopore experts will provide attendees (up to 20) with a comprehensive data analysis overview followed by practical hands-on experience running different analysis workflows.

This workshop will give participants experience with real-time analysis tools as well as in-depth analysis workflows with command line tools.

The workshop will cover the following:

- ✓ In depth overview of EPI2ME for real-time end-to-end analysis
- ✓ Command line tutorial
- ✓ Guppy (command, parameters and folder structure)
- ✓ Overview of fast5 file format
- ✓ Quality and read length distribution plot and filtering
- ✓ Mapping to a reference
- ✓ De novo assembly and consensus correction
- ✓ Methylation detection

Agenda

- 08:45 - Registration opens
- 09:15 - Introduction to nanopore sequencing
(optional for Introduction workshop attendees)

Real-time analysis of nanopore reads

- 10:00 - Nanopore file formats
- 10:30 - Introduction to EPI2ME

Analysis workflows with command line tools

- 11:00 - Introduction to practical sessions
- 11:15 - A brief Linux tutorial
- 12:00 - Networking lunch
- 13:00 - Practical session I
Basecalling with Guppy
Generating run metrics and summaries with Nanoplot
Data filtering with Filtlong
Reference mapping using Pomoxis (Minimap2)
Visualizing Alignments with Tablet
- 14:30 - Coffee Break
- 14:40 - Practical session II
De novo assemblies with Pomoxis (Miniasm/Racon)
Error correction with Medaka
Methylation detection
- 16:30 - Workshop closes



Certificate of attendance

This certifies that

Jih-Hui Lin

Has successfully completed

**Introduction Workshop
and
Data Analysis Overview**

Signed:  . *Andy Davies*
Senior Director of Global Services Date: 29-11-2019